



**HAL**  
open science

# Le clone épidémique "Bourg-en-Bresse" de l'espèce *Burkholderia cenocepacia*: origine, positionnement phylétique et phénomènes génétiques liés à son émergence

Arnault Graindorge

► **To cite this version:**

Arnault Graindorge. Le clone épidémique "Bourg-en-Bresse" de l'espèce *Burkholderia cenocepacia*: origine, positionnement phylétique et phénomènes génétiques liés à son émergence. Sciences agricoles. Université Claude Bernard - Lyon I, 2009. Français. NNT: 2009LYO10217. tel-00586252

**HAL Id: tel-00586252**

**<https://theses.hal.science/tel-00586252>**

Submitted on 15 Apr 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

présentée

devant l'UNIVERSITE CLAUDE BERNARD - LYON 1

pour l'obtention

du DIPLOME DE DOCTORAT

(arrêté du 25 avril 2002)

présentée et soutenue publiquement le 25 Novembre 2009

par

Arnault GRAINDORGE

**Le clone épidémique "Bourg-en-Bresse" de l'espèce *Burkholderia cenocepacia* : origine, positionnement phylétique et phénomènes génétiques liés à son émergence**

Directeur de thèse :

Benoit COURNOYER

JURY :

T. COENYE

A. HARTMANN

J. FRENEY

A.C. VERGUNST

B. COURNOYER

Rapporteur

Rapporteur

Examineur

Examineur

Examineur

# UNIVERSITE CLAUDE BERNARD - LYON 1

## **Président de l'Université**

Vice-président du Conseil Scientifique

Vice-président du Conseil d'Administration

Vice-président du Conseil des Etudes et de la Vie Universitaire

Secrétaire Général

**M. le Professeur L. Collet**

M. le Professeur J-F. Mornex

M. le Professeur G. Annat

M. le Professeur D. Simon

M. G. Gay

## ***COMPOSANTES SANTE***

Faculté de Médecine Lyon Est – Claude Bernard

Faculté de Médecine Lyon Sud – Charles Mérieux

UFR d'Odontologie

Institut des Sciences Pharmaceutiques et Biologiques

Institut des Sciences et Techniques de Réadaptation

Département de Formation et Centre de Recherche en Biologie Humaine

Directeur : M. le Professeur J. Etienne

Directeur : M. le Professeur F-N. Gilly

Directeur : M. le Professeur D. Bourgeois

Directeur : M. le Professeur F. Locher

Directeur : M. le Professeur Y. Matillon

Directeur : M. le Professeur P. Farge

## ***COMPOSANTES SCIENCES ET TECHNOLOGIE***

Faculté des Sciences et Technologies

UFR Sciences et Techniques des Activités Physiques et Sportives

Observatoire de Lyon

Institut des Sciences et des Techniques de l'Ingénieur de Lyon

Institut Universitaire de Technologie A

Institut Universitaire de Technologie B

Institut de Science Financière et d'Assurance

Institut Universitaire de Formation des Maîtres

Directeur : M. Le Professeur F. Gieres

Directeur : M. C. Collignon

Directeur : M. B. Guiderdoni

Directeur : M. le Professeur J. Lieto

Directeur : M. le Professeur C. Coulet

Directeur : M. le Professeur R. Lamartine

Directeur : M. le Professeur J-C. Augros

Directeur : M R. Bernard

## Remerciements.

*Enfin ce moment tellement attendu est là, je m'étais toujours dis que les remerciements signeraient la fin de cette thèse et l'aboutissement de nombreuses années d'études parsemées de moments aussi bien heureux que malheureux. Tous ces événements m'ont conduit, on ne sait pas vraiment pourquoi, à faire une thèse. Mais ils m'ont surtout permis de devenir la personne que je suis aujourd'hui... Durant ce périple, j'ai eu l'occasion de rencontrer de nombreuses personnes qui m'ont permis d'avancer, de grandir et d'en arriver là. Je vais donc essayer de n'oublier personne dans ces remerciements et d'avancer plus ou moins chronologiquement dans ma vie afin de citer les étapes clefs qui ont fait de moi, après 365,33 Lunes passées sur cette terre, un Docteur en Microbiologie.*

*Je tiens donc tout d'abord à remercier mes parents de m'avoir conçu aux Baléares durant l'été 1979, me permettant ainsi de voir le jour le 13 mai 1980 à Echirolles. Je les remercie pour toutes les années de bonheur insouciant que j'ai pu vivre à leurs cotés quand la flamme de leur amour ne s'était pas encore éteinte. Je les remercie pour les étés où nous partions en Algérie, retrouver notre famille de l'autre cote de la méditerranée. J'en profite pour remercier Kader, la Houaria, Belkâsem, Aicha, Momo, et toutes les personnes que j'ai rencontrées durant ces nombreuses vacances dans un pays où l'accueil n'a d'égal que la chaleur des journées. Je tiens aussi à remercier mes grands parents paternels pour les nombreux noëls que nous passions en famille. Un merci tout particulier à mon grand père Guy, qui s'en est allé à un moment de déchirement dans notre famille. Tu ne parlais pas beaucoup mais tes silences en disaient long... Merci pour les dernières années où tu m'a fais découvrir le château de Versailles et Eurodisney!!!!!! Je tiens également à remercier mes grands parents maternels, je regrette sincèrement de ne pas vous avoir plus connue, et je me souviendrai toujours de vous, de votre appartement dans le XIIIème et de vos convictions qui sont aujourd'hui les miennes. Je te remercie, Pépé, pour les gènes que tu m'as transmis et qui m'ont donné le gout de l'ovalie, si tu savais comme je regrette que tu n'ais pas eu le temps de venir me voir jouer à ce sport que tu aimais tant...*

*Je tiens aussi à remercier mon père, je me souviens des bons moments que nous passions en Algérie et surtout d'un moment particulier où tu me faisais marcher au plafond dans notre*

*appartement à Voiron. Les traces de mes pieds sont restées très longtemps inscrites sur ce plafond et font partie d'un des plus beaux souvenirs que j'ai de nous. Je regrette beaucoup le manque de communication qui nous a amené à de nombreux déchirements dans notre relation... Mais même si l'expression des sentiments n'est pas la chose que tu maîtrises le plus, j'espère juste qu'un jour tu pourras transmettre aux gens que tu aimes, toutes les choses que tu ressens.*

*Je tiens à remercier ma mère qui a toujours été là pour moi et que j'ai malheureusement usé toutes ces années... Je pourrais écrire des paragraphes sur toutes les choses que tu as faites pour moi et sur les sentiments que j'ai pour toi. Je veux juste que tu saches que même si tu crois que certains de tes choix n'étaient pas toujours les meilleurs, peu de personnes auraient réussi à s'occuper de nous comme tu la fais. Ton investissement au niveau du travail, des amis et de tes deux garçons fait de toi la personne la plus altruiste que je connaisse. Tu as su me transmettre des valeurs et même si je ne te le dis pas assez, je t'aime... J'aimerais maintenant que tu t'occupes de toi afin que je puisse te garder à mes côtés le plus longtemps possible.*

*Je tiens aussi à remercier mon frère, les moments que nous avons partagés aussi bien dans la souffrance que dans le bonheur m'ont permis de connaître ce qu'était l'amour fraternel. J'espère que la face sombre de ton être ne prendra pas le dessus dans le futur, et que tu arriveras à sortir de cette spirale infernale qui te détruit un peu plus chaque jour... Je sais que la vie d'artiste est toujours oscillante mais je souhaite sincèrement que ton génie soit reconnu de ton vivant !!!! Je te souhaite toute la réussite et tout le bonheur que tu mérites.*

*Merci à mon oncle et ma tante (et toute la famille Gau) de m'avoir épaulé dans les moments difficiles, de m'avoir accueilli quand il fallait que je me ressource... Mireille et Patrick, je ne vous remercierai jamais assez d'être venu me soutenir lors de nombreux épisodes difficiles de mon existence... J'en profite pour remercier mes cousins, Benjamin & Emi, Marie & Simon et Antoine & Helene avec qui je partage de plus en plus de choses au fur et à mesure des années. Un bisou particulier à Victor qui vient de naître... Merci aussi à mon autre tante, Madeleine, qui nous a précipitamment quitté il y a quelques années...*

*Merci aux enseignants que j'ai rencontré durant toute ma scolarité et qui ont su me gérer dans des moments où j'étais ingérable, qui ont été énormément patients avec moi et qui ont su m'intéresser à leur cours, je pense à Mme Masson, Mr Graz, Mme Moreau, Mr Sorita, Mr Medan, Mme Puig, Mr Curcy et Mme Hugon. Merci à toute l'équipe pédagogique du collège la*

*Garenne de m'avoir donné quelques limites trop peu souvent respectées: Mr Perrin, Mr Raffin et Mr Paquien. Merci à Mme Lutz de m'avoir donné une deuxième chance et permis de réintégrer le lycée Edouard H. Merci à Mr Zaccai de m'avoir accueilli dans son laboratoire de recherche durant l'été 1997, me permettant ainsi de connaître le monde de la recherche. Merci à tous mes profs d'IUT de m'avoir enseigné la rigueur du travail de laboratoire. Merci à Bruno Franzetti de m'avoir accueilli à l'IBS pour mon stage de fin d'étude et d'avoir ainsi confirmé mon envie de faire de la recherche.*

*Merci à tous les amis que j'ai rencontrés durant ces deux années dans le Var (Département où le Var ne passe pas, ahhhh les gens du Sud...), Flo dit le Tarin, Djak, AsbaTriiiiiiaaa, Pamela la fameuse témoin de Jeova, Yonel Tarrago, Benoît et bien sur mon Fiston, Vincent que j'aime de tout mon cœur et avec qui j'adore partager des séjours à NY et à Paris !!!!! Vincent, mon Fils soleil, j'espère que nous ne nous quitterons jamais !!!!*

*Merci à toutes les personnes que j'ai rencontrées au cours de mon cursus à l'UCBL, en particulier Nico L'indien, Jeremy, Seb la Pu., mon Patou adoré et bien sûr mon Pierral avec qui je partage toujours des moments inoubliables. Je vous remercie tous de m'avoir supporté durant toutes ces années sur les bancs de la fac et ailleurs.*

*Merci à toutes les personnes du club très fermé de l'ADJ: JJ pour les nombreuses discussions, les pauses café et surtout pour les formidables repas que tu nous as concoctés, Chleichhhhh pour son poil luisant, Carbo pour son sourire et sa bonne humeur, Andrew (qui faisait partie du club ???), Laurette pour tous ces bon repas, petits déjeuners et dîners, les clopes, pour les quelques footings, pour l'organisation de notre voyage au Maroc, Peuz pour toutes les soirées que nous avons faites, les poker et les matchs de foot, même si mon foi lui, ne te remercie pas, et bien sûr DarkG pour les mille et une clopes, pour les nombreuses soirées arrosées ou non, pour tes succulents makj, pour nos discussions philosophiques et pour ton accueil quand je n'allais pas bien... Je remercie également Julien pour m'avoir supporté pendant la formation pour l'expérimentation animale, Anne-So pour sa gentillesse et ses fous rires mémorables, Karine pour sa tortue qui assure et qui est morte malheureusement, mais ne t'inquiète pas elle est au paradis des tortues..., et à Rico mon ami suisse américain (c'est pas banal !!!!) nos multiples échanges lors des retours de soirée sur les routes respectives de nos domiciles.*

*Je tiens aussi à remercier tou(te)s mes amis de longue date, rencontrés par l'intermédiaire du Rugby et/ou de ma vie à Voiron. Ferdi qui sait tout faire avec ses mains, et autres extrémités, le fameux pâtissier qui monte sa propre entreprise de piscine PPMC (traduction censurée), Momo pour tous ces moments passés au collège, au stade et ailleurs, Mouss pour nos vacances à Hyères et toutes les escapades que nous avons réalisé ensemble, Thomas, mon frère en quelque sorte, merci pour notre vie "commune", nos vacances en Espagne et tant d'autre expériences telles que des gardes à vue remplies de bonheur... Merci à Sylvie de m'avoir fait manger tous les midis durant les années collège et ta succulente glace à la vanille... Merci à Christian pour les discussions devant les retransmissions sportives, au bord des terrains ou au bord des routes de Rallye. Merci à Guigui pour les étés passés au Verney et les hivers passés chez ma mère, pour les fêtes et les boumes, les matchs et les mille et une expériences que nous avons partagées. Merci à Willy et Souz pour les dimanches après midi et pour ce formidable jour de l'an dans cette belle maison de campagne. Merci à Adeline pour sa voix stridente qui réveillerait les morts, merci à Grigrouche de m'avoir supporté au collège et au lycée, merci à Linda pour son franc parler et sa bonne humeur. Un merci tout particulier à Mél qui m'a accompagné pendant huit ans, qui m'a permis d'évoluer et qui m'a appris à mieux me connaître. J'espère que ta vie future sera remplie de bonheur et que tu me pardonneras tous les mauvais moments que je t'ai fait subir... Merci à toute ta famille qui m'a accueilli comme si c'était la mienne, avec une pensée particulière pour tes grands parents maternels, pour Gérard et Sylvana, qui ne quitteront jamais mon cœur... Merci à toute la famille Charlot pour m'avoir accompagné durant ces 15 années de rugby, à la famille Garcia qui ne cesse de s'agrandir et qui a toujours été là pour moi, un gros bisou à Juju et Jean Marc. Merci aussi à Ivan de m'avoir entraîné au rugby mais aussi aux soirées. Merci aussi à mes deux enfants du rugby qui sont bien plus fous que moi, Colon et Cotton !!! Merci à Jacques Insardi, Jean Marc Celestine, Domi, Charly, Beber, Alain (Merci de m'avoir fait jouer 2eme ligne contre Rumilly...) et merci à tous les autres entraîneurs qui ont essayé de me gérer aussi bien pendant les matchs que les entraînements. Merci à toute la famille du rugby qui m'a appris à partager et à vivre en groupe (et à boire aussi), à organiser des soirées de folie et des événements aussi importants que les moules frites du 11 Novembre à Voiron... Merci également à COS et Mathieu Pantaloni pour les soirées à l'internat.*

*Merci à René Bally de m'avoir accueilli au sein du Laboratoire d'Ecologie Microbienne de Lyon. Merci à Benoit de m'avoir encadré durant ces années, d'avoir trouvé le financement qui m'a permis de réaliser cette thèse, merci pour ta patience et ta capacité à répondre à mes questions inopinées. Merci de m'avoir emmené aux différents colloques IBCWG et pour cette discussion qui m'a permise de mieux gérer mon stress la veille de mon oral à Ann Harbor. Merci pour le temps que tu as pris pour corriger mon manuscrit et merci également pour tes hamburgers qui sont vraiment trop bon. Merci à mon mentor, mon ami et mon maître Burkho, Aymeric et toute sa famille, Anne, Guislain, Paul et Philomène qui m'ont souvent invité à manger et à jouer avec les petits. Merci Aymeric, c'est un peu grâce à toi que j'ai pu réaliser cette thèse... d'ailleurs, cette thèse est la suite de tes travaux, je te la dédie tout particulièrement. Merci à toutes les personnes de mon jury d'avoir accepté d'évaluer mon manuscrit. Merci à Stéphanie (ou Step, Crtd'n) pour ton aide lorsque la rédaction devenait un calvaire, pour ta bonne humeur, tes post-it réconfortants, ton accent du sud et surtout pour les chansons qu'on chantonait ensemble dans le bureau. Merci à Claire pour la précieuse aide logistique et les petites attentions du quotidien, merci à toi Sabine pour tout plein de choses : les TPs, les discussions sur nos vécus, ta bonne humeur, tes envolées pour l'état lamentable du labo et pour ton soutien durant la rédaction de ce manuscrit... Merci à Babeth de m'avoir formé à la PCR lors de mon arrivée au laboratoire. Merci à Sylvie de m'avoir ouvert sa porte lors de mon stage de maîtrise et surtout merci de m'avoir aidé à finir cette thèse sereinement et pas au chômage !!!!! Merci à Bruno Tilly pour nos longues journées et nos nombreuses « bastons » de pissette d'alcool dans le P3 de l'école véto... Merci à tous les autres membres de l'équipe qui m'ont encouragé, qui ont dû subir mes moments de folie passagère et mes présentations en réunions, souvent incompréhensibles... Un grand merci à CyberValérie pour les milles et un cafés offert, les repas du CROUS et les tarots. Merci au staff « technique » du labo qui nous permet d'avancer : Christine, Dominique, Danis, Corinne, Gigi2, Frank, Stéphane et Gigi pour sa pièce sympa, ses barbecues et sa grande gueu.. Gigi, tu nous manques tellement... Merci aux habitants passés et présents du bureau 505 d'avoir su partager mon désordre organisé. Vous avez eu du courage, ce n'était pas tous les jours facile et surtout merci d'avoir supporté mes chants un peu faux.. Merci également à toutes les personnes de l'IBCWG qui m'ont permis de progresser tant au niveau scientifique que dans la langue de Snoop Dog.*



*Un grand merci aux membres du Jury qui ont accepté d'évaluer ce travail de thèse de 267 pages...*

*Merci à tous les étudiants du laboratoire d'écologie microbienne, mon ami Denis avec qui j'ai partagé de nombreux moments aussi bien heureux que tristes, avec qui j'aime jouer au squash, boire de la chartreuse© et faire du vélo, aller au salon des vins, dans le Vercors etc..., merci à Eléonore et Pierre pour tous ces repas, ces soirées et ces sorties en ville, merci à Kratouffffff pour ta musique qui assure grave, merci à Béné et Loic pour leur somptueux mariage et tous les BBQ et autres soirées Wii/réseau, j'espère que vous saurez transmettre votre force et votre bonne humeur à toutes les personnes que vous croiserez. Merci à Chouchou (BlackBB ou le Serpent Noir et autres surnoms bizarres) pour ces bons moments passés ensemble, ces réunions d'équipe et ces BBQ qui nous ont permis de nous éclater ensemble. Merci à Jeepee pour les soirées réseaux et les voyages aux urgences après une bonne sortie vélo. Merci à Ludo pour les poses clopes et nos ragots sur les gens du labo !!!! Merci aux trois frères pour leur multiples invitations quai Saint Antoine et ces bons moments autour d'une bonne bière... Merci à Martina de m'avoir fait découvrir le Slibovitséééé, à la branche armée d'Al-Qaïda du labo (Karima, Amel, Amel2, Lamia et autres...), à Marie-Lara (qui est trop bien, rien à dire de plus...), à Emeline (qui est complètement folle et qui WAAAZZZAAAAA avec moi au tel), à Manuelle (qui mange trop de lapin), à Tatiana (qui tire trop souvent son lait en publique), à Hervé qui paraît calme comme ça mais en faite c'est grave un guedin !!!!!!! Merci également à Thomas et Mickael pour les sorties VTT. Merci à Flo W. et à Céline Lavire pour les TPs in et outdoor. Merci à Sandrine, David Chapulliot, Pierre Pujic, Frank Poly, Daniel, et bien sûr merci à tous les autres étudiants en thèse ou en master ; c'est grâce à vous que le laboratoire peut vivre et que la recherche peut avancer, j'espère que quand je ne serai plus là vous continuerez à pourrir les boites mail comme il se doit et à chanter dans les couloirs...*

*Enfin, un grand merci à toi ma chérie, Marie-Croupette-Chupita Beaume, pour ton aide précieuse tant au niveau matériel (et ca c'est pas le plus important) que moral (et ca c'est très important !!!), merci pour nos fous rires, nos pleurs, nos moments de pelotonnage durant les nuits d'hiver... et d'été, pour nos discussions sur l'oreiller et pour m'avoir soutenu tout au long de cette rédaction même dans les moments les plus durs. Merci pour nos moments au pays de Mickey, même quand tu avais mal aux... nos moments à la montagne et ailleurs... Passer ces*

*instants avec toi me remplit de bonheur. Merci aussi à madame nature pour tes BB et merci pour ton amour au quotidien. Je te souhaite cependant bonne chance parce que des fois, il paraît que je suis un peu LOURD. Love You Darling.*

*Merci également à toutes les personnes qui liront cette thèse et pas uniquement les remerciements et à toutes les personnes que j'ai oubliées.*

# Table des matières

INTRODUCTION GENERALE .....	9
<b>1. Description du complexe <i>Burkholderia cepacia</i> .....</b>	<b>12</b>
1.1. Taxonomie et phylogénie moléculaire.....	12
1.2. Les <i>Burkholderia</i> du "complexe <i>cepacia</i> " .....	12
1.3. Identification des souches.....	13
1.4. Epidémiologie et structure génétique du complexe <i>Burkholderia cepacia</i> .....	14
1.5. Les clones épidémiques de l'espèce <i>Burkholderia cenocepacia</i> .....	16
<b>2. <i>Burkholderia</i> du complexe <i>cepacia</i> et environnement.....</b>	<b>17</b>
2.1. Eaux.....	18
2.2. Sols .....	19
2.3. Associations.....	21
2.3.1 Interactions bénéfiques des Bcc dans l'environnement.....	21
2.3.2 Interactions néfastes des Bcc dans l'environnement .....	24
2.4. Les gènes "environnementaux".....	26
<b>3. <i>Burkholderia</i> du complexe <i>cepacia</i> chez l'homme .....</b>	<b>27</b>
3.1. Infections communautaires.....	27
3.1.1. Cas de la mucoviscidose.....	29
3.2. Infections nosocomiales.....	34
3.3. Prophylaxie.....	35
<b>4. Mécanismes de la Pathogénèse.....</b>	<b>35</b>
4.1. Adhérence aux cellules épithéliales.....	37
4.2. Invasion des cellules épithéliales .....	38
4.3. Translocation au travers de l'épithélium.....	39
<b>5. Déterminants de la virulence.....</b>	<b>40</b>
5.1. Lipopolysaccharide .....	42
5.2. Systèmes de sécrétion .....	43
5.3. Acquisition du Fer .....	43
5.4. Résistance aux antibiotiques .....	45
5.5. Autres facteurs.....	46
5.6. Modèles d'étude.....	48
<b>6. Adaptation aux contraintes environnementales .....</b>	<b>49</b>
6.1. Acclimatation.....	49
6.2. Génération de diversité .....	51
6.2.1. Hypermutabilité .....	51
6.2.2. Séquences d'insertion .....	53
6.2.3. Réarrangements génétiques .....	53
6.3. Acquisition d'information génétique .....	54
6.3.1. Phages .....	54

6.3.2. Plasmides .....	55
6.3.3. Ilots génomiques .....	55
<b>7. Conclusion.....</b>	<b>56</b>

Caractérisation moléculaire et distribution environnementale du clone B&B de l'espèce <i>Burkholderia cenocepacia</i> .....	57
---	----

Epidemiology and molecular characterization of a clone of <i>Burkholderia cenocepacia</i> responsible for nosocomial pulmonary tract infections in a French intensive care unit.....	59
--	----

<b>1. Introduction .....</b>	<b>59</b>
<b>2. Materials and methods.....</b>	<b>60</b>
2.1. <i>Sampling of the ICU, patients, and outdoor environments</i> .....	60
2.2. <i>Isolation of bacterial strains</i> .....	62
2.3. <i>Phenotypic analyses and antibiotic resistances</i> .....	62
2.4. <i>DNA manipulations</i> .....	62
<b>3. Results .....</b>	<b>63</b>
3.1. <i>Description of the outbreak</i> .....	63
3.2. <i>Environmental sources of Bcc-like strains</i> .....	63
3.3. <i>Molecular characterization of the Bcc-like isolates</i> .....	64
3.3.1. <i>recA</i> PCR-RFLP analysis.....	64
3.3.2. <i>Phylogenetic analyses</i> .....	64
3.3.3. <i>PFGE-SpeI analysis</i> .....	65
3.3.4. <i>Virulence genes and antibiotic resistance profiles</i> .....	66
<b>4. Discussion.....</b>	<b>66</b>
<b>5. Acknowledgments.....</b>	<b>68</b>
<b>6. References .....</b>	<b>68</b>

Description des échantillonnages réalisés au niveau du centre hospitalier	71
---	----

<b>Conclusion du chapitre .....</b>	<b>76</b>
-------------------------------------	-----------

Description de facteurs transcriptionnels alternatifs primaires et de remaniements génomiques des gènes codant des facteurs ECF au sein de la famille $\sigma^{70}$ des <i>Burkholderia</i> du complexe <i>cepacia</i> .....	77
--	----

<b>1. Préambule .....</b>	<b>77</b>
---------------------------	-----------

# Architecture of Burkholderia cepacia complex $\sigma^{70}$ gene family: evidence of alternative primary and clade-specific factors, and genomic instability

<b>1. Background</b> .....	<b>81</b>
<b>2. Results and Discussion</b> .....	<b>82</b>
2.1. Detection and number of $\sigma^{70}$ gene sequences in Bcen-J2315 genome.....	82
2.2. Classification of $\sigma^{70}$ gene sequences of Bcen-J2315 into functional categories .....	82
2.3. Molecular phylogeny of Bcen-J2315 primary, stationary, and stress response/cellular differentiation/flagellin synthesis $\sigma^{70}$ gene sequences .....	85
2.3.1. The primary and alternative primary $\sigma^{70}$ factors phylogenetic radiation (group 1) .....	86
2.3.2. The stationary phase $\sigma^{70}$ factors phylogenetic radiation (group 2) .....	87
2.3.3. The stress response/cellular differentiation/flagellin synthesis $\sigma^{70}$ factors phylogenetic radiation (group 3) .....	88
2.4. Molecular phylogeny of Bcen-J2315 extracytoplasmic functions $\sigma^{70}$ gene sequences (group 4).....	91
2.4.1. <i>ecfA</i> : evidence of an ET12 clone-specific duplication event.....	91
2.4.2. <i>ecfB</i> : involvement in temperature-dependent processes .....	92
2.4.3. <i>ecfC</i> and <i>ecfI</i> : ECF gene sequences involved in iron-uptake.....	92
2.4.4. <i>ecfD</i> , <i>ecfE</i> , <i>ecfG</i> , <i>ecfH</i> , <i>ecfL</i> and <i>ecfM</i> loci: unknown physiological implications.....	95
2.4.5. <i>ecfF</i> and <i>ecfJ</i> : evidence of Burkholderia-specific lineages.....	97
2.5. Genomic rearrangements of $\sigma^{70}$ genetic loci among Bcc sequenced genomes.....	97
2.5.1. Loss of $\sigma^{70}$ gene sequences in the Bcc: cases of <i>ecfC</i> , <i>ecfG</i> , <i>ecfH</i> , <i>ecfK</i> , <i>sigI</i> , and <i>sigJ</i> loci.....	98
2.5.2. Loss of <i>ecfE</i> in <i>B. vietnamiensis</i> : involvement of phage DNA.....	98
2.5.3. Reorganized $\sigma^{70}$ genetic loci in the Bcc: cases of <i>ecfI</i> and <i>ecfE</i> loci .....	99
<b>3. Conclusion</b> .....	<b>100</b>
<b>4. Methods</b> .....	<b>101</b>
4.1. <i>B. cenocepacia</i> genome sequence, and $\sigma^{70}$ BLAST and ORF searches .....	101
4.2. Molecular phylogenetic analyses .....	101
4.3. Bacterial strains, DNA extractions, PCR amplification and DNA blot hybridizations.....	101
<b>5. Acknowledgements</b> .....	<b>101</b>
<b>6. References</b> .....	<b>103</b>

## Analyses complémentaires de la structure génétique de la famille des facteurs $\sigma^{70}$ chez les Bcc: mise à jour et cas du clone B&B de *B. cenocepacia* .....

<b>1. Introduction</b> .....	<b>107</b>
<b>2. Résultats et discussion</b> .....	<b>108</b>
2.1. Détection de nouveaux facteurs $\sigma^{70}$ .....	108
2.2. Classification fonctionnelle des nouveaux $\sigma^{70}$ .....	110

2.3. Phylogénie moléculaire des facteurs $\sigma 70$ non-ECF .....	110
2.4. Phylogénie moléculaire des facteurs $\sigma 70$ ECF .....	113
2.5. Distribution et environnement génétique de gènes <i>ecf</i> chez les <i>Bcc</i> .....	115
2.5.1. <i>ecfA</i> .....	115
2.5.2. <i>ecfC</i> .....	118
2.5.3. <i>ecfE</i> .....	118
2.5.4. <i>ecfI</i> .....	121
2.5.5. <i>ecfK</i> .....	121
2.5.6. <i>ecfB</i> , <i>ecfD</i> , <i>ecfG</i> , <i>ecfH</i> , <i>sigI</i> et <i>sigJ</i> .....	122
<b>3. Conclusions .....</b>	<b>123</b>
<b>4. Matériels et méthodes .....</b>	<b>128</b>
4.1. Génome de <i>Burkholderia</i> , $\sigma^{70}$ BLAST et recherches d'ORF.....	128
4.2. Analyses par phylogénie moléculaire.....	128
4.3. Souches bactériennes, extraction d'ADN, amplifications PCR et Southern blot.....	128
<b>5. Bibliographie (partie B du chapitre 2) .....</b>	<b>131</b>
<b>6. Conclusion de chapitre.....</b>	<b>135</b>

Relation entre émergence de clones épidémiques de <i>Burkholderia cenocepacia</i> et amplification d'éléments répétés de type IS .....	136
--	-----

Genomic organization of <i>Burkholderia cenocepacia</i> IS elements and its relation with the emergence of epidemic clones .....	140
--	-----

<b>1. Introduction .....</b>	<b>140</b>
<b>2. Materials and Methods .....</b>	<b>142</b>
2.1. Strain used, DNA extractions, PCR amplification and DNA blot hybridization.....	142
2.2. Identification of IS and insertion sites among sequenced genomes .....	143
2.3. Molecular phylogenetic analyses .....	145
<b>3. Results and Discussion .....</b>	<b>147</b>
3.1. IS numbers among <i>Burkholderia cenocepacia</i> genomes.....	147
3.2. IS family sub-groups broadly distributed among <i>B. cenocepacia</i> .....	150
3.2.1. IS3 .....	150
3.2.2. IS110.....	168
3.2.3. IS256.....	179
3.2.4. IS21 .....	185
3.3. IS family sub-groups with a narrow distribution in <i>B. cenocepacia</i> .....	188
3.3.1. IS5 .....	188
3.3.2. IS66.....	190
3.3.3. ISL3.....	192
3.3.4. IS481 .....	192
3.4. IS and large scale genomic rearrangements .....	194

<b>4. Conclusions .....</b>	<b>196</b>
-----------------------------	------------

Ilots génomiques et éclosions épidémiques chez les <i>Burkholderia cenocepacia</i> : une association étroite .....	198
--	-----

Genomic islands and IS elements: a significant association among the <i>Burkholderia cenocepacia</i> ET12 and B&B genomes.....	201
--	-----

<b>1. Introduction .....</b>	<b>201</b>
------------------------------	------------

<b>2. Material and methods .....</b>	<b>203</b>
--------------------------------------	------------

2.1. Strains, DNA extractions, PCR amplification and sequencing .....	203
---	-----

2.2. B&B clone genomic library.....	204
-------------------------------------	-----

2.3. DNA blot analyses .....	206
------------------------------	-----

2.4. Comparative genomic analyses .....	206
---	-----

<b>3. Results and discussion.....</b>	<b>207</b>
---------------------------------------	------------

3.1. DNA regions specific of the J2315 genome.....	207
--	-----

3.2. <i>B. cenocepacia</i> J2315 GI harboring high number of IS elements.....	210
---	-----

3.3. B&B clone screening of the six GI detected among J2315 genome.....	218
---	-----

3.4. Detection of GI specific to the B&B clone genome .....	221
---	-----

<b>4. Conclusion.....</b>	<b>224</b>
---------------------------	------------

Conclusion et perspectives .....	226
----------------------------------	-----

Bibliographie .....	231
---------------------	-----

Annexes .....	242
---------------	-----

# Liste des Figures

## Introduction bibliographique

Fig. 1 : Arbre phylogénétique ADNr 16S des <i>Burkholderia</i> .	11
Fig. 2 : Phénotypes de pathogénie chez la plante.	17
Fig. 3 : Cascade physiopathologique des maladies pulmonaires chez les patients atteints de mucoviscidose.	28
Fig. 4 : Prévalence des infections bactériennes des voies respiratoires, selon les groupes d'âge chez les patients CF.	30
Fig. 5 : Diagramme de Venn montrant les co-infections des voies respiratoires de patients CF.	30
Fig. 6 : Schéma de cellules épithéliales du poumon et leurs interactions avec les Bcc.	36
Fig. 7 : Schéma d'une séquence d'insertion.	52
Fig. 8 : Schéma d'un îlot génomique.	52

## Chapitre 1

Fig. 1: Photos du clone B&B de <i>B. cenocepacia</i> en microscopie électronique à balayage.	58
--	----

### Partie A

Fig. 1: Outline of the ICU considered in this study	60
Fig. 2: Synoptic table of the Bcc outbreak.	63
Fig. 3: Neighbor-joining phylogenetic tree of concatenated <i>recA</i> and <i>ecfB</i> gene sequences from the outbreak and representative strains of the Bcc.	65
Fig. 4: <i>SpeI</i> restriction patterns of total genomic DNA from all Bcc strains reported in this work.	66
Fig. 5: Global geographic distribution of B&B-like strains of <i>B. cenocepacia</i> .	67

### Partie B

Fig. 1: Situation géographique de l'hôpital Fleyriat et lieux d'échantillonnage.	73
Fig. 2: Prises de vue des différents prélèvements positifs obtenus au cours des échantillonnages.	75

## Chapitre 2

Fig. A: Régions conservées des $\sigma^{70}$ et fonctions liées à la liaison à l'ADN.	78
---	----

### Partie A

Fig. 1: Allocation of <i>B. cenocepacia</i> strain J2315 $\sigma^{70}$ deduced factors into the main $\sigma^{70}$ gene family phylogenetic groups.	85
Fig. 2: Conserved regions among <i>B. cenocepacia</i> strain J2315 $\sigma^{70}$ factors.	86
Fig. 3: Neighbor-Joining phylogenetic tree of non-ECF $\sigma^{70}$ factors.	89
Fig. 4: Neighbor-Joining phylogenetic tree of ECF $\sigma^{70}$ factors.	94



Fig. 5: Synteny analysis of <i>B. cenocepacia</i> strain J2315 <i>ecf</i> genomic regions showing rearrangements in other Bcc sequenced genomes.	96
Fig. 6: The <i>ecfA</i> ( <i>rpoE</i> ) gene duplication.	97
Fig. 7: Arbre phylogénétique en Neighbor-Joining des facteurs $\sigma$ 70 non-ECF.	112
Fig. 8: Arbre phylogénétique en Neighbor-Joining des facteurs $\sigma$ 70 de type ECF.	114
Fig. 9 Crible PCR de la région <i>ecfA2</i> à partir d'extrait d'ADN dilué de <i>B. cenocepacia</i> .	116
Fig. 10: Analyse de la synténie de la région <i>ecfC</i> entre <i>B. cenocepacia</i> strain J2315 et <i>B. multivorans</i> ATCC17616.	120
Fig. 11: Analyse de la synténie de la région <i>ecfE</i> entre <i>B. cenocepacia</i> J2315 et <i>B. multivorans</i> ATCC17616.	124
Fig. 12: Analyse de la synténie de la région de <i>ecfI</i> entre <i>B. cenocepacia</i> strain J2315 et différentes souches de <i>Burkholderia</i> .	125
Fig. 13: Analyse de la synténie de la région de <i>ecfK</i> entre <i>B. cenocepacia</i> J2315 et différents génomes de <i>Burkholderia</i> .	126
Fig. 14: Voie de dégradation du benzoate chez <i>B. multivorans</i> ATCC17616.	127

### Chapitre 3

Fig. 1a : Différents modes de transposition d'IS.	137
Fig. 1b: Gene classes of disrupted ORF among <i>B. cenocepacia</i> strains investigated.	149
Fig. 2: Main features of IS3 elements detected among <i>B. cenocepacia</i> strains.	151
Fig. 3: Neighbor-Joining phylogenetic tree of IS3 family insertion sequences.	154
Fig. 4: Main features of DNA regions showing particular transposase CDS organization among <i>B. cenocepacia</i> strains.	157
Fig. 5: Genomic distribution of IS elements among <i>B. cenocepacia</i> J2315 genome.	160
Fig. 6: Neighbor-Joining phylogenetic tree of IS407 transposase DNA sequences.	162
Fig. 7 : Example of tripartite comparisons of <i>B. cenocepacia</i> conserved DNA regions showing presence or absence of an IS element, using the ACT graphic tool.	164
Fig. 8: Cladogram illustrating the conservation of <i>B. cenocepacia</i> DNA loci colonized by the IS407 among the <i>B. cenocepacia</i> J2315 genome.	166
Fig. 9: Cladogram illustrating the relation between the number of IS407 and the strains of <i>B. cenocepacia</i> .	166
Fig. 10: Neighbor-Joining phylogenetic tree of transposases of IS110 family.	169
Fig. 11: Neighbor-Joining phylogenetic trees of IS <i>Bcen</i> 8, IS <i>Bma</i> 3, and IS <i>Bcen</i> 2 DNA sequences from <i>B. cenocepacia</i> and closely related species.	171
Fig. 12: Main features of IS110 elements found among <i>B. cenocepacia</i> strains.	173
Fig. 13: NJ phylogenetic trees of IS <i>Bcen</i> 4, from <i>B. cenocepacia</i> J2315, B&B clone and closely related species.	177
Fig 14. Neighbor-Joining phylogenetic tree of IS256 family transposases.	181
Fig. 15: Neighbor-Joining phylogenetic trees of IS1356 (cluster I) and IS <i>Bcen</i> 18 (cluster II) DNA sequences encoding transposases of the IS256 family.	184
Fig. 16: Main features of IS21 and IS256 families elements found among <i>B. cenocepacia</i> strains.	186
Fig 17: IS21 family Neighbor-Joining phylogenetic trees.	187

Fig. 18: Main features of IS elements with a narrow distribution among <i>Burkholderia cenocepacia</i> strains.	189
Fig. 19: IS5 family Neighbor-Joining phylogenetic tree using transposase CDS DNA sequences from the <i>Burkholderia</i> .	191
Fig. 20: <i>Burkholderia</i> IS <i>Bma1</i> Neighbor-Joining phylogenetic tree using CDS DNA sequences of their respective transposase.	193
Fig. 21: <i>Burkholderia</i> IS481 family Neighbor-Joining phylogenetic tree using CDS DNA sequences of their respective transposase.	193
Fig. 22: Global synteny analysis of the <i>B. cenocepacia</i> AU1054 and HI2424 genomes.	195

## Chapitre 4

Fig. A: Caractéristiques générales des îlots génomiques.	200
Fig. 1: The 92.8 kb Genomic Island 10 of <i>B. cenocepacia</i> J2315 and its high number of IS elements.	211
Fig. 2: The 34.2 kb Genomic Island 11 of <i>B. cenocepacia</i> J2315 and its high number of IS elements.	212
Fig. 3: The 121.9 kb Genomic Island 20 of <i>B. cenocepacia</i> J2315 and its high number of IS elements.	213
Fig. 4: The 40.7 kb Genomic Island 26 of <i>B. cenocepacia</i> J2315 and its high number of IS elements.	214
Fig. 5: The 46.2 kb Genomic Island 41 of <i>B. cenocepacia</i> J2315 and its high number of IS elements.	216
Fig. 6: The 91.6 kb Genomic Island 57 of <i>B. cenocepacia</i> J2315 and its high number of IS elements.	217
Fig. 7: The 142.2 kb Genomic Island 29 harbouring the <i>Cenocepacia</i> island (BCESM).	221
Fig. 8. Methodology used for GI screening among non-sequenced genomes; case of GI10.	223
Fig. 9: Genomic islands detected among the B&B clone.	225

## Liste des tableaux

### Introduction bibliographique

Tableau 1 : Incidence des espèces (Genomovars) parmi les Bcc isolés chez les patients CF.	32
Tableau 2 : Liste des facteurs de virulence présents chez les Bcc, distribution et fonction.	41

### Chapitre 1

#### Partie A

Table 1. Epidemiologic features of the ICU Bcc outbreak.	61
Table 2. Strains of the Bcc used or isolated in this study.	61
Table 3. PCR DNA primers used in this study.	62

#### Partie B

Tableau 1. Sources de carbones utilisées par différentes souches de <i>Burkholderia</i> .	74
---	----

### Chapitre 2

#### Partie A

Table 1. Distribution of <i>B. cenocepacia</i> J2315 strain $\sigma^{70}$ factors among other <i>Burkholderia</i> genome sequences and <i>P. aeruginosa</i> .	84
Table 2. <i>Burkholderia cepacia</i> complex strains used in this study.	90
Table 3. PCR primers used in this study.	102
Tableau 4. Distribution des facteurs $\sigma^{70}$ au sein des génomes de <i>Burkholderia</i> séquencés et chez <i>P. aeruginosa</i> .	109
Tableau 5. Souches de <i>Burkholderia</i> utilisées dans cette étude.	117
Tableau 6. Amorces PCR utilisées dans ce travail.	130

### Chapitre 3

Table 1. Bacterial strains used in this study.	142
Table 2. PCR primer used in this study	144
Table 3. IS elements among <i>Burkholderia</i> genomes	146
Table 4. Characterization of IS direct DNA repeats among <i>B. cenocepacia</i> J2315 genome.	155
Table 5. Characterization of IS407 flanking ORFs in the J2315 genome.	160
Table 6. PCR screenings of selected IS407 insertion sites among a collection <i>B. cenocepacia</i> strains.	165
Table 7. Characterization of IS407 insertion sites containing no IS element.	167
Table 8. Characterization of IS <i>Bcen8</i> flanking ORFs.	174
Table 9. Characterisation of IS <i>Bcen2</i> flanking ORFs.	178
Tables 10. Characterisation of IS <i>I356</i> flanking ORFs among Bcc genomes.	183

## Chapitre 4

Table 1. Bacterial strains used in this study.	203
Table 2. PCR Primers used in this study.	205
Table 3. DNA regions only detected in J2315 genome when performing comparisons with the PHDC genomes.	208-209
Table 4. PCR screenings of GI identified in <i>B. cenocepacia</i> J2315 genome.	219

## INTRODUCTION GENERALE

Les *Burkholderia* sont des bacilles gram négatifs de la famille des  $\beta$ -protéobactéries. La première espèce de ce genre a été décrite par Burkholder (1950) qui la classa dans les pseudomonades : *Pseudomonas cepacia*. Yabuuchi *et al.* (1992), sur la base d'une étude de l'ADNr 16S, ont proposé la création du nouveau genre *Burkholderia* en y rattachant les espèces *Pseudomonas caryophylli*, *Pseudomonas cepacia*, *Pseudomonas gladioli*, *Pseudomonas mallei*, *Pseudomonas pickettii*, *Pseudomonas pseudomallei* et *Pseudomonas solanacearum*. A la suite d'analyses phylogénétiques, *Burkholderia pickettii* et *B. solanacearum* ont ensuite été transférées dans le genre *Ralstonia* (Yabuuchi *et al.*, 1995). On compte actuellement soixante quatre espèces de *Burkholderia*.

Les bactéries au sein de ce genre présentent une étonnante diversité, tout d'abord due à leurs interactions avec les eucaryotes pouvant aller de la symbiose racinaire au parasitisme. Ensuite, en raison de leurs propriétés métaboliques ; certaines sont des parasites obligatoires (*B. mallei*), d'autres peuvent se développer sur plus de deux cents sources de carbone différentes (*B. xenovorans*). Nous retrouvons également cette diversité dans les écosystèmes colonisés : eau douce, sol, surface des feuilles ou bien les poumons de l'homme. Les traces de ces adaptations et de leur évolution sont, sans aucun doute, retrouvées dans les génomes de ces bactéries. En effet, au sein du genre, les génomes de *Burkholderia* divergent fortement de part leur taille, le nombre de réplicons, leur organisation et leur composition.

La définition du genre *Burkholderia* repose avant tout sur des considérations génomiques et sur la composition des lipides cellulaires (Henry *et al.*, 2001). En ce qui concerne les principaux caractères phénotypiques, notons que ce sont des bacilles droits, à Gram négatif, accumulant généralement des granules de poly-bêta-hydroxybutyrate, mobiles grâce à un ou à plusieurs flagelles polaires (*Burkholderia mallei* est cependant dépourvue de flagelles et est donc immobile), aérobies strictes, catalase positive, oxydase variable selon les espèces, capables de croître en utilisant comme unique source de carbone le glucose, le glycérol, l'inositol, le galactose, le sorbitol, le mannose (à la différence des *Ralstonia sp.* qui n'utilisent pas ce dernier) et le mannitol. Les différentes espèces du genre peuvent être discriminées par les résultats des auxanogrammes (137 substrats ont été utilisés dans l'étude de Gillis *et al.*, 1995). Les espèces du genre *Burkholderia* ont pour principal habitat le sol (Salles *et al.*, 2002, Compant *et al.*, 2008), la rhizosphère des plantes (Sfalanga *et al.*, 1999, Balandreau *et al.*, 2001, Richardson *et al.*, 2002, Han *et al.*, 2009, Kyselkova *et al.*, 2009) mais quelques espèces sont pathogènes stricts et/ou opportunistes pour l'homme et l'animal (Mahenthalingam *et al.*, 2008).

Différentes approches sont actuellement menées pour appréhender la diversité au sein du genre et progresser sur deux voies importantes : la lutte contre les maladies liées à certaines espèces de *Burkholderia*, et l'exploitation des effets bénéfiques en agriculture ou dans l'industrie de la dépollution par certaines espèces de ce genre. Ce double jeu des *Burkholderia* n'est pas sans poser problème : la frontière (si elle existe) entre espèces ou même souche ayant des effets bénéfiques ou pathogènes reste mal définie à l'heure actuelle.

Nous aborderons ces différents aspects en établissant tout d'abord un bilan des caractéristiques du genre *Burkholderia*, et ce en nous focalisant plus particulièrement sur un groupe d'espèces : le complexe *cepacia* (Bcc). Nous analyserons sa systématique, et les outils permettant son étude. L'écologie des bactéries du complexe *cepacia* sera ensuite détaillée pour appréhender leur croissance dans l'environnement abiotique (eaux et sols) en étudiant les différents types d'associations présents au sein du Bcc (symbiose, phytopathogène, amibes, protozoaires) et le support génétique ou "gènes environnementaux" permettant la colonisation de niches aussi diversifiées. Nous nous intéresserons ensuite aux différentes interactions connues entre les Bcc et l'homme en nous concentrant sur les différents types d'infections connues (infections communautaires, avec le cas particulier des patients atteints de mucoviscidose et infections nosocomiales). Nous essayerons de comprendre les processus impliqués dans leur transmission nosocomiale en étudiant leur épidémiologie et les principales mesures d'hygiène pouvant enrayer cette transmission. Nous tenterons ensuite d'appréhender les raisons de leur émergence, de leur pathogénèse et les supports génétiques de leur virulence. Enfin, nous essayerons de mettre en évidence, en utilisant notamment les annotations des génomes des *Burkholderia*, les phénomènes génétiques liés à leur haut pouvoir adaptatif. La description d'un nouveau clone épidémique, le clone B&B, et l'accès à de nombreux génomes séquencés au sein de ce genre nous a permis de développer des travaux de génomique comparative permettant d'apprécier une partie de l'histoire évolutive de ce nouveau clone, et d'entrevoir les effets de certains processus génétiques sur son évolution et son adaptation au milieu clinique.

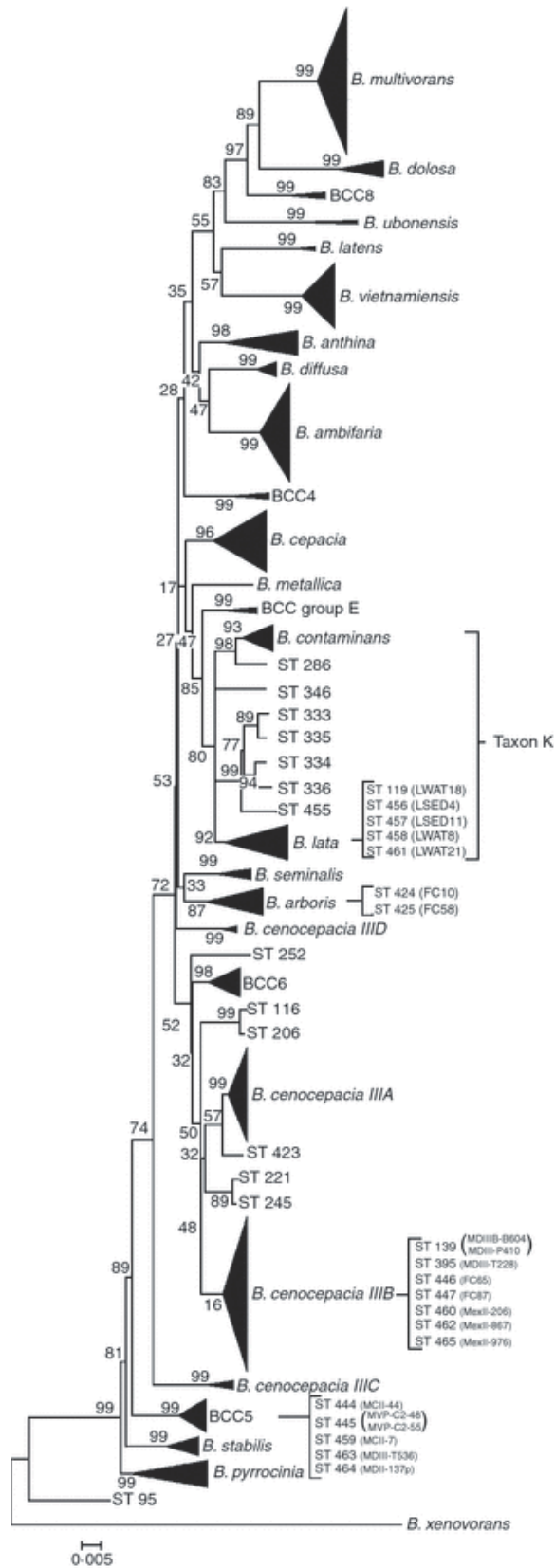


Fig 1: Arbre phylogénétique MLST montrant la position de l'ensemble des espèces du *Burkholderia cepacia* complex. La barre d'échelle représente 0.5% de divergence. D'après Cesarini *et al.*, 2009.

## 1. Description du complexe *Burkholderia cepacia*

### 1.1. Taxonomie et phylogénie moléculaire

La taxinomie ou taxonomie est la science ayant pour objet de décrire les organismes vivants et de les regrouper en entités appelées taxons afin de pouvoir les nommer et les classer. Les plus récentes méthodes utilisées incluent une nouvelle approche conceptuelle de la classification: les méthodes phylogénétiques et cladistiques. La phylogénie a pour but d'établir les liens de parenté entre différents organismes vivants de même espèce ou d'espèces différentes. A l'échelle de la bactérie, la phylogénie permet l'étude très fine des "liens de parentés et de clonalité" entre différentes souches. La phylogénie moléculaire revêt une grande importance avec le développement des techniques de séquençage disponible aujourd'hui et financièrement accessibles. Celles-ci nous permettent d'étudier les relations phylétiques entre bactéries et ainsi de déterminer leur ancêtre commun. Ces techniques peuvent aussi permettre de comparer des bactéries qui seraient considérées comme identiques par les méthodes d'étude phénotypiques classiques (réactions biochimiques, morphologie, résistances aux antibiotiques etc.). La taxonomie, et notamment la phylogénie moléculaire, sont devenues des méthodes incontournables pour la caractérisation de clones épidémiques (cf. chap.1).

### 1.2. Les *Burkholderia* du "complexe *cepacia*"

Basé sur une étude polyphasique (analyse électrophorétique des protéines totales, composition en acides gras, hybridation ADN/ADN), Vandamme *et al.* (1997) proposent de regrouper les bactéries du genre *Burkholderia* en cinq génomovars. Ce terme de génomovar a été introduit par Ursing *et al.* (1995) pour décrire des groupes de souches phénotypiquement très proches mais génétiquement trop éloignés pour être de la même espèce. En effet, ils montrent des valeurs d'hybridations ADN/ADN supérieures à 65% au sein des mêmes génomovars, alors qu'entre génomovars différents, ces valeurs sont comprises entre 20 et 45%. Ils incluent *B. vietnamiensis* (génomovar V), déjà décrite par ailleurs (Gillis *et al.*, 1995) dans ce complexe d'espèces, et décrivent une nouvelle espèce *B. multivorans* (génomovar II). Le nom d'espèce *B. cepacia* est conservé pour le génomovar I. L'étude de l'ARN 16S qui peut être utilisée en taxonomie bactérienne n'est dans ce cas pas assez discriminante, étant donné les très fortes valeurs d'identité (98 à 100%) rencontrées au sein de ce complexe. Le



terme de "complexe *cepacia*" a été proposé pour faire référence à ces espèces génomiques lors du troisième Congrès de l'International *Burkholderia cepacia* working group (IBCWG) en 1997. Différents articles ont contribué à la description, en tant qu'espèces, des différents génomovars de ce complexe : I : *B. cepacia* (Vandamme et al., 1997), II : *B. multivorans* (Vandamme et al., 1997), III : *B. cenocepacia* (Vandamme et al., 2003), IV : *B. stabilis* (Vandamme et al., 2000), V : *B. vietnamiensis* (Gillis et al., 1995), VI : *B. dolosa* (Vermis et al., 2004), VII : *B. ambifaria* (Coenye et al., 2001), VIII : *B. anthina* (Vandamme et al., 2002), IX : *B. pyrrocinia* (Vandamme et al., 2002). Une dixième espèce, *B. ubonensis*, qui avait été évaluée pour son appartenance à ce complexe (Payne et al., 2005) a récemment été confirmée comme appartenant au Bcc (Vanlaere et al., 2008). Cette étude a également permis de définir cinq nouvelles espèces comme appartenant au Bcc : *Burkholderia latens*, *Burkholderia diffusa*, *Burkholderia arboris*, *Burkholderia seminalis*, et *Burkholderia metallica*. Récemment, une nouvelle étude a défini deux nouvelles espèces à partir du taxon K : *Burkholderia lata*, et *Burkholderia contaminans* (Vanlaere et al., 2009). Ce complexe, en perpétuel remaniement, contient donc dix-sept espèces à ce jour (Fig. 1). Un trait commun, cependant, à l'ensemble des espèces de ce complexe est qu'elles ont toutes été retrouvées, à des fréquences variables, chez des individus atteints de mucoviscidose. Au sein du genre *Burkholderia*, la phylogénie moléculaire basée sur l'étude de l'ADNr 16S montre que le Bcc est proche des espèces *B. pseudomallei*, *B. mallei* et *B. thailandensis* (Coenye and Vandamme 2003). Ces deux premières espèces sont des pathogènes primaires de l'homme et de l'animal, alors que *B. thailandensis* n'a jamais été associée à une maladie : la position phylétique des espèces de *Burkholderia* ne permet donc pas de déduire une relation avec leur pathogénie.

### 1.3. Identification des souches.

Malgré les avancées effectuées dans la description des différentes espèces du Bcc, il s'avère délicat de réaliser l'identification spécifique et rapide de nouvelles souches du fait de leur proximité phénotypique et génétique. Les analyses basées sur les caractères biochimiques conduisent à trop de mauvaises identifications pour permettre de différencier les espèces du Bcc en routine (Rabkin et al., 1989). En conséquence, différents outils basés sur l'analyse de fragments génétiques ont été développés. Mahenthiralingam et al. (2000) proposent l'utilisation d'une PCR-RFLP sur le gène *recA* pour discriminer les espèces du Bcc. Après amplification avec des amorces spécifiques du gène *recA* des bactéries du Bcc, le produit PCR peut être digéré par deux enzymes *HaeIII* ou *MnII*, et ainsi conduire à une classification dans

un des cinq génomovars du complexe connu à cette date. A noter que cette technique également de distinguer deux sous-groupes A et B au sein du génomovar III. Par la suite, cette approche a largement été reprise et un second article de Vermis et al. (2002) a évalué la fiabilité de cet outil pour l'identification des neuf espèces du Bcc décrites alors. Cette dernière se révèle correct pour le Genomovar III, *B. ambifaria*, alors que *B. multivorans*, *B. cepacia* génomovar I et *B. pyrrocina* étaient souvent mal discriminées. Parallèlement à l'utilisation de cette PCR-RFLP *recA*, des amorces pour le gène *recA*, spécifiques de chaque espèce, ont été employées. Il semble cependant que, pour certaines espèces, la spécificité des amorces soit problématique, notamment pour des souches environnementales (Dalmastri et al., 2005). Une nouvelle approche PCR *recA*, permettant l'identification de la majeure partie des espèces du genre *Burkholderia*, a été développée par Payne et al., en 2005. Cette technique a facilité l'identification des souches de *B. cepacia*, *B. stabilis* et *B. ambifaria*. En 2005, Baldwin et al., ont publié le schéma MLST (Multi Locus Sequence Typing) du Bcc. Cet outil permet la discrimination des espèces moyennant le séquençage de sept loci différents. La phylogénie reconstruite avec chacun de ces sept loci ne permet pas une identification complète de l'ensemble des espèces et des sous-espèces. En revanche, en utilisant des séquences concaténées, une phylogénie complète et robuste peut être obtenue. Cependant, certaines souches du Bcc ne produisaient pas d'amplification en utilisant les amorces décrites. De nouvelles amorces ont récemment été dessinées, améliorant la performance du schéma MLST pour le Bcc, et permettant l'amplification des souches non amplifiables jusqu'alors (Spilker et al., 2009). Une autre étude récente, comparant l'identification par séquence *recA* et par MLST, a démontré que 65.4% et 88.5% des isolats de Bcc testés pouvaient être identifiés respectivement par séquençage *recA* et par MLST (Cesarini et al., 2009). Ces travaux ayant été réalisés en utilisant les anciennes amorces MLST, l'efficacité des nouvelles amorces ne peut donc pas être appréciée. L'avantage indéniable de la MLST réside dans son caractère aisément transposable d'un laboratoire à l'autre par l'intermédiaire d'une base de données disponible en ligne. De plus, cet outil peut être aussi bien être utilisé pour l'identification d'espèces que pour réaliser des études épidémiologiques (e. g. Chapitre 1 et section suivante).

#### 1.4. Epidémiologie et structure génétique du complexe *Burkholderia cepacia*

Au cours des 30 dernières années, la discrimination entre les isolats du Bcc a été évaluée par diverses méthodes. Le ribotypage a été une des premières méthodes utilisée pour l'épidémiologie des Bcc. Cette méthode consiste en une amplification du gène codant le 16S

suivie d'une restriction permettant une discrimination, bien que faible, de souches provenant de différents patients (LiPuma et al., 1988). Quelques années plus tard, Kostman et al., (1992) développèrent une variante de cette technique, l'AFLP (Amplified Fragment Length Polymorphism) sur la base d'une amplification de la région intergénique comprise entre le 16S rRNA et le 23S rRNA. La longueur variable de cette région peut être utilisée pour discriminer différents isolats. Une autre technique, la MLEE (Multi-Locus Enzyme Electrophoresis) a été employée dans les années 90s pour différencier certains isolats provenant de différents environnements (patients CF, patients non-CF et sol) (Johnson et al., 1994, Wise et al., 1995). Cette technique, alliée au ribotypage, a permis de mettre en évidence un clone épidémique, le clone ET12 qui représente à ce jour la lignée la plus dommageable de l'espèce *B. cenocepacia* (e. g. section suivante). Le typage par RAPD (Random amplified polymorphic DNA) a ensuite été développé ; ce dernier présente un fort pouvoir discriminant grâce aux nombreux profils différents par génomovar (Mahenthalingam et al., 1996, Segonds et al., 1997). Cette technique est cependant plus délicate à mettre en œuvre. Toutes ces techniques présentent de nombreux avantages mais restent difficilement reproductibles entre laboratoires. Actuellement, la macro-restriction du génome couplée à l'électrophorèse en champ pulsé (PFGE) est la méthode de référence car elle demeure la plus robuste et la plus discriminante (Agodi et al., 2001, Lambiase et al., 2007). Cependant, sa mise en œuvre est lourde et nécessite en outre du matériel spécialisé. La plupart de ces méthodes sont appropriées pour la comparaison d'un ensemble d'isolats obtenus lors d'un même échantillonnage ; leur utilité pour répondre aux questions concernant l'épidémiologie à long terme est cependant limitée (Enright and Spratt 1999). Une revue décrivant ces diverses techniques, leurs avantages et leurs inconvénients a été publiée par Coenye and LiPuma en 2003. Une technique considérée comme alternative à la MLEE et la MLST a été développée par Coenye et LiPuma en 2002 : la MLRT (MultiLocus Restriction Typing). Cette technique, développée pour l'espèce *B. cenocepacia*, consiste à amplifier 5 loci puis à effectuer une restriction sur chacun de ces amplifias. Cette méthode a montré des résultats en accord avec ceux obtenus en PFGE. Cependant, l'accès aux technologies de séquençage étant désormais possible tant au niveau financier qu'au niveau technologique, la MLST semble être le meilleur compromis pour réaliser des études épidémiologiques à long terme. D'autre part, cette technique présente une très bonne reproductibilité, un pouvoir discriminant important, une facilité d'utilisation, une rapidité de mise en œuvre, et se détache des autres techniques par la disponibilité des données en ligne.

### 1.5. Les clones épidémiques de l'espèce *Burkholderia cenocepacia*

L'émergence de *Burkholderia* dans les cas d'infections de patients atteints de mucoviscidose (patients CF) date de la fin des années 70s, début des années 80s (Laraya-Cuasay et al., 1977, Nolan et al., 1982, Isles et al., 1984). Diverses espèces, telles que *B. pseudomallei* et *B. gladioli*, ont été ponctuellement retrouvées chez ces patients CF (Jones et al., 2001, O'Carroll et al., 2003). Cependant, les Bcc sont actuellement les *Burkholderia* les plus fréquemment isolées chez ces patients (Govan et al., 2007, Lynch 2009, e.g. section 3.1.1). Au sein de ce complexe, les espèces *B. multivorans* et *B. cenocepacia* prédominent, tant chez les patients CF que non CF (Reik et al., 2005). L'émergence ainsi que l'augmentation des infections dues à *Burkholderia* a ainsi motivé de nombreuses études aboutissant à la caractérisation de trois clones épidémiques majeurs. Tous trois, isolés à partir de patients CF, appartiennent à l'espèce *B. cenocepacia* ; leur différenciation, dans un premier temps, résida principalement sur des critères géographiques. Tout d'abord, le clone Midwest, isolé pour la première fois au niveau de 5 centres de patients atteints de mucoviscidose dans le Michigan, est responsable de nombreuses infections dans le Middle West des États-Unis (Kumar et al., 1997, Coenye and LiPuma 2002). D'autre part, le clone PHDC représente la majorité des isolats obtenus à partir d'infections à Bcc chez les patients CF de la côte Atlantique des États-Unis (Chen et al., 2001). Différentes souches appartenant à la lignée PHDC ont de plus été isolées au niveaux de sols agricoles et de patients CF en Europe, suggérant un réservoir potentiel de ce clone dans l'environnement (Coenye et al., 2004, Campana et al., 2005). Le dernier clone épidémique décrit, le clone ET12, a été identifié pour la première fois en 1994 par Johnson et al. Ce clone est prédominant en Ontario et au Royaume-Uni où il représente plus de la moitié des isolats de Bcc chez les patients CF (Pitt et al., 1996). Il a été démontré que la propagation de cette dernière lignée épidémique au niveau de ces différents pays avait résulté d'infections croisées entre patients CF lors d'un camp d'été (Govan et al., 1993). De plus, le clone ET12 est caractéristique du "syndrome *cepacia*" qui consiste en une pneumonie nécrotique foudroyante à l'origine de nombreux décès chez les patients CF (Isles et al., 1984). Différentes études semblent cependant démontrer que ce syndrome peut également être provoqué par les espèces *B. multivorans* et *B. dolosa* (Zahariadis et al., 2003, Blackburn et al., 2004, Kalish et al., 2006). Par ailleurs, d'autres clones épidémiques ont été isolés chez des patients CF en Italie et en République Tchèque mais leur prévalence au niveau mondiale est plus restreinte (Manno et al., 2004, Drevinek et

al., 2005). Il existe de plus des cas documentés d'épidémies nosocomiales liées aux *Burkholderia cenocepacia* (e.g. section 3.2).

## **2. *Burkholderia* du complexe *cepacia* et environnement**

Les bactéries sont ubiquitaires et sont présentes dans tous les types de biotopes rencontrés sur Terre. Elles peuvent être isolées des eaux douces ou marines, du sol, des profondeurs océaniques, de l'air, sur la peau, du système digestif des animaux et sont même capable de coloniser des milieux particulièrement extrêmes tel que les déchets radioactifs (Fredrickson et al., 2004). Il y a environ 1 milliard de cellules bactériennes dans un gramme de sol et  $10^{10}$  million de cellules bactériennes dans un millilitre d'eau douce. En 1998, on estimait qu'il y avait quatre à six quintillions ( $4 \times 10^{30}$  à  $6 \times 10^{30}$ ) de bactéries dans le monde, représentant ainsi une grande partie de la biomasse mondiale (Whitman et al., 1998). Dans l'environnement, les bactéries sont en interaction avec de nombreux organismes aussi bien microscopiques que macroscopiques. Cette proximité peut engendrer une compétition pour les ressources du milieu, contraignant ainsi les bactéries à s'adapter aux changements du milieu. Ces changements peuvent être causés par différents processus, notamment l'activité des organismes eux-mêmes (variation du pH, diminution des nutriments etc...) mais aussi des facteurs extérieurs comme les variations d'hygrométrie, l'alternance jour/nuit, les saisons, etc. Les capacités d'adaptation des bactéries à un milieu, conduisant à une variation phénotypique, vont ainsi déterminer leur présence in fine dans le milieu considéré. En effet, malgré un apport exogène régulier, une espèce bactérienne peut être incapable de se maintenir dans un environnement donné. L'étude de la distribution d'une espèce ou d'un genre bactérien dans différents milieux permet d'étudier leur potentiel adaptatif et leur capacité à évoluer dans diverses conditions, ce qui s'avère fondamental pour les micro-organismes pathogènes. Dans ce contexte, l'identification de réservoirs est un élément clé de l'étude d'un pathogène. Ces observations visent ainsi à développer nos moyens de contrôle et de surveillance, ainsi qu'à identifier puis développer de nouveaux agents de lutte.

Bien que la première espèce du genre *Burkholderia* ait été identifiée de l'environnement (Burkholder 1950), peu d'études aboutissant à la mise en évidence des réservoirs environnementaux de ces pathogènes opportunistes ont été réalisées. Cependant, les espèces du Bcc ont été retrouvées classiquement niveau de divers environnements, tels que la rhizosphère du maïs et du blé (Vandamme et al., 2003), les racines du pois (Coenye et al., 2001), les rivières, les eaux, les lacs et divers produits industriels (Baldwin et al., 2007). Ceci

suggère une large distribution des espèces du Bcc dans l'environnement ; étonnamment, peu d'études ont identifié des souches de patients dans l'environnement (LiPuma et al., 2002, Payne et al., 2005, Baldwin et al., 2007).

### 2.1. Eaux

Les *Burkholderia* colonisent classiquement le milieu eau douce et sont éventuellement retrouvées en milieu marin. En 2004, une étude remarquable publiée dans Science a permis de reconstituer un génome appartenant à une souche du Bcc à partir du séquençage d'ADNs métagénomiques extraits de la mer des Sargasses (Venter et al., 2004). Néanmoins, des analyses supplémentaires tendent à démontrer une contamination des échantillons (Mahenthiralingam et al., 2006). La prévalence des Bcc dans le milieu marin est peu étudiée, quelques souches seulement ont pu être isolées à partir de sédiments marins, et des tests de survie réalisés sur des espèces *B. multivorans* et *B. cenocepacia* dans ce milieu montre la survie de l'espèce *B. multivorans* uniquement, sous forme non-cultivable (VBNC) (Moore et al., 2007, Wang et al., 2008). Des études plus approfondies sont nécessaires pour connaître la réelle distribution des Bcc dans le milieu marin qui représente plus de 70% de la surface terrestre.

La plupart des souches appartenant au Bcc ayant une origine hydrique proviennent de prélèvements d'eaux douces. La quasi-totalité de l'eau douce est stockée sous forme solide au niveau des pôles et des glaciers, le reste étant réparti entre les eaux souterraines (nappes phréatiques) et les eaux de surface (cours d'eaux, lacs, étangs). L'anthropisation a désormais des répercussions considérables sur ces eaux, contribuant à une modification non négligeable des populations bactériennes présentes. Différents travaux ont démontré la présence de Bcc dans des cours d'eau peu anthropisés établissant un accroissement de la population en automne. Il semblerait que ce pic corresponde au transfert des bactéries présentes sur les feuilles des végétaux lors de leur chute dans les cours d'eau (Lemke and Leff 1999, Olapade and Leff 2005). Par ailleurs, le suivi de ces populations au sein de cours d'eau chimiquement pollués démontre que cette modification du milieu naturel ne semble pas affecter l'abondance bactérienne des populations de Bcc (Lemke and Leff 1999). En revanche, une comparaison de cours d'eau ayant des paramètres physico-chimiques différents (température, pH, turbidité, concentration en carbone organique dissous et en Nitrate), établit une corrélation positive entre l'abondance des Bcc et la concentration en nitrates et en carbone organique dissous (Olapade et al., 2005). Ces paramètres semblent donc influencer sur la fitness bactérienne, la capacité à entrer

en compétition avec les autres micro-organismes présents dans le milieu, et à résister à la prédation. La distribution des différentes espèces du Bcc au niveau des eaux douces reste peu connue. Très peu d'isolats environnementaux ont été obtenus, mais l'espèce *B. multivorans* a pu être détectée au niveau de sources naturelles. Les analyses phylogénétiques ont de plus démontré la proximité de ces souches avec d'autres souches isolées de patient CF au Portugal (Baldwin et al., 2008). Enfin, Vermis et al. en 2003 ont identifié diverses espèces du Bcc, notamment *B. vietnamiensis*, *B. multivorans* *B. cenocepacia* au sein d'eaux de rivières. En conclusion, il semble, malgré une littérature faiblement documentée à ce propos, que les eaux douces représentent un véritable réservoir en ce qui concerne les espèces du Bcc

## 2.2. Sols

Les sols sont caractérisés par une grande hétérogénéité et une importante complexité ; cependant, leurs accessibilités plutôt aisées font d'eux des matériels de choix pour étudier la diversité bactérienne. Ainsi, l'étude des sols représente une partie importante des travaux en écologie microbienne. Dans ce sens, la recherche de pathogènes opportunistes présents dans les sols représente un challenge important ayant des répercussions sérieuses en santé publique. De nombreuses études ont été réalisées à la fin des années 90s, début des années 2000, afin d'estimer les populations de Bcc présentes dans divers sols et ainsi identifier leur réservoir potentiel. Les résultats obtenus sont très contrastés : dans certains cas elles sont retrouvées fréquemment alors que dans d'autres elles sont très rares (Parke and Gurian-Sherman 2001). Cette variabilité semble s'expliquer par les méthodes utilisées, notamment l'isolement et les moyens d'identification des espèces ainsi que par l'évolution rapide de la taxonomie des Bcc. En effet, le perpétuel remaniement du Bcc qui comptait 5 genomovars dans les années 90s et qui abrite maintenant plus de 17 espèces, rend très délicate l'identification. De plus, une difficulté demeure puisque l'utilisation de milieux sélectifs, pouvant contenir des antibiotiques, plutôt performants pour l'isolement de souches cliniques, peut inhiber les souches environnementales. Ceci a été confirmé par l'emploi de milieux peu sélectifs ou minimum permettant de retrouver des bactéries en plus grande quantité. Néanmoins, la méthode culturale en tant que telle contient de nombreux biais et, pour s'en affranchir, des méthodes sans culture préalable ont été utilisées afin de ne pas négliger les bactéries viables non cultivables. Les résultats ont montré une efficacité cinq fois plus importante dans la détection de souches par approche PCR spécifique de l'ADNr 16S sur de l'ADN extrait de ces sols (Miller et al., 2002). Ces résultats soulignent la faiblesse des méthodes culturales pour

identifier l'ensemble des Bcc, même si, par ailleurs, le développement des méthodes d'identification directes s'est accompagné de nouveaux biais concernant principalement la sensibilité et la spécificité.

Cette évolution des techniques d'isolement et le développement de méthodes plus fines pour d'identification des Bcc ont permis la réalisation de nombreux travaux durant les dernières années. Ceux-ci ont principalement été menés pour déterminer les espèces du Bcc présentes dans les sols sous cultures agricoles ou en forêt. Une étude de Bevivino et al., (2002) a documenté les prévalences respectives des espèces du Bcc dans des sols cultivés de maïs et dans des prélèvements issus de patients atteints de mucoviscidose. Dans les sols sous maïs, *B. cenocepacia* est majoritaire (53.4%), suivi par *B. ambifaria* (37,3%), *B. pyrrocinia*, *B. cepacia*, quant à elle, est retrouvée dans moins de 10% des cas. En parallèle, l'étude identifiait *B. cenocepacia* dans 86,4% des prélèvements cliniques, les autres espèces étant présentes dans moins de 10% des cas. Ces résultats ont été confirmés par l'étude de Zhang and Xie (2007) qui indique une abondance variable des espèces du Bcc dans la rhizosphère du maïs et du riz, avec une prévalence de l'espèce *B. vietnamiensis* pour la rhizosphère du riz et de *B. cenocepacia* au niveau de la rhizosphère du maïs. Cette hétérogénéité des espèces de Bcc au sein de plantes d'un même champs avait préalablement été mise en évidence par Ramette et al., en 2005, et a encore été enrichie par Dalmastri et al., 2007 qui ont démontré que l'espèce *B. ambifaria* était majoritaire dans un sol sous maïs. Ces différences dans la prévalence des Bcc montrent que les abondances relatives des espèces de Bcc sont extrêmement variables entre les différentes variétés de plantes et les localisations géographiques. Une avancée majeure dans la recherche de réservoirs des Bcc fut l'identification d'une souche clinique de la lignée PHDC au niveau d'un champ d'oignons (LiPuma et al., 2002). Ce résultat suggère que les isolats humains ne sont pas nécessairement distincts de ceux de l'environnement. Ainsi, l'acquisition de nouvelles souches de Bcc par les patients CF pourrait être due à des contaminations par des souches environnementales. Vanlaere et al., (2005) ont mis en évidence diverses souche de Bcc, notamment des souches *B. cenocepacia* et *B. multivorans*, au niveau de sols et de points d'eau, et ceci au sein même de maisons de patients CF. La grande diversité des isolats obtenus leur a permis de conclure que le sol et l'eau jouaient le rôle de réservoirs des souches présentes chez les patients plutôt que de confirmer les hypothèses inverses. Les récentes et dernières autres études permettant de relier des souches environnementales et cliniques ont été réalisées en utilisant le schéma MLST (Baldwin et al., 2007, Springman et al., 2009). Cette technique présente l'avantage de



pouvoir réaliser des analyses épidémiologiques à grande échelle, mais les isolats regroupés sous une même ST (séquence type) ne représentent pas forcément la même souche.

L'identification des réservoirs environnementaux de souches cliniques semble délicate ; cependant, les sols sous cultures apparaissent comme de bons "candidats". Les plantes semblent aussi jouer un rôle important dans leur capacité à structurer les communautés microbiennes en favorisant l'émergence ponctuelle d'une espèce particulière. Il apparaît important, de plus, de prendre en compte les flux hydriques afin d'évaluer la dispersion de ces bactéries dans d'autres environnements.

### 2.3. Associations

La présence des Bcc dans l'environnement implique de nombreuses interactions avec les différents organismes présents à proximité. Ces interactions peuvent être considérées sous différents aspects en fonction de la résultante de l'interaction. On parlera de symbiose quand les deux partenaires bénéficient directement de l'association, de commensalisme quand l'un retire un bénéfice sans effet sur l'autre et de parasitisme quand le bénéfice de l'un se fait au détriment de l'autre. Mahenthalingam et al., (2005) ont fait un résumé des aspects positifs et négatifs d'interactions impliquant les Bcc.

#### 2.3.1 Interactions bénéfiques des Bcc dans l'environnement

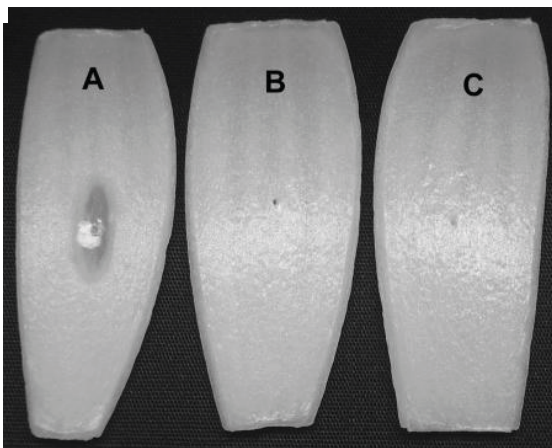
Différents types d'interactions bénéfiques ont été mis en évidence entre les bactéries et d'autres organismes micro- ou macroscopiques. A titre d'exemple, deux grands types d'interactions positives sont observés entre les bactéries et les plantes : la symbiose stricte avec formation d'une structure particulière favorisant l'échange entre les deux partenaires (le nodule), et la symbiose associative qui se fait sans structure particulière, au niveau de la rhizosphère ou dans les tissus de la plante.

Dans ce premier cas, la symbiose est établie dans un but principalement métabolique puisque, d'une part, la bactérie trouve avantage à utiliser les composés carbonés fournis par la plante et que d'autre part, la plante, incapable de métaboliser l'azote atmosphérique, tire bénéfice de sa fixation par la structure nodulaire. Cette capacité à former des nodules est restreinte à quelques genres bactériens : Les *Rhizobia*, *Sinorhizobia*, et les *Mesorhizobia* ( $\alpha$ -protéobactéries) avec les Légumineuses et *Frankia* avec diverses plantes. Moulin et al., (2001) ont mis en évidence la capacité de certaines *Burkholderia* et *Cupriavidus* à former des

nodules racinaires avec des Légumineuses. Ces résultats ont été confirmés quelque années plus tard par Chen et al., (2005). Cette capacité à former des nodules a été découverte chez les Bcc pour la première fois en 2005 par Rasolomampianina et al., qui ont identifié une souche du Bcc (STM1424) capable de former des nodules avec une légumineuse (*Dalbergia spp.*) présente sur l'île de Madagascar. L'espèce à laquelle appartient cette souche n'a pas pu être formellement identifiée mais l'analyse phylogénétique du gène codant pour l'ADNr 16S la classe dans le Bcc.

Dans le contexte des symbioses associatives, aucune structure particulière n'est mise en place. L'interaction avec la plante peut alors se situer au niveau de la rhizosphère, sur les parties aériennes (épiphytes) ou dans les tissus (endophytes). L'apport des bactéries à la plante peut se faire sous différentes formes : fixation libre d'azote atmosphérique, synthèse de phytohormones, modification de la balance hormonale, solubilisation de composés du sol, ou encore minéralisation plus rapide de composés du sol améliorant la nutrition végétale. Quand ces interactions prennent place au niveau de la racine, on parle alors de bactérie PGPR (Plant Growth Promoting Rhizobacteria). Tran Van et al., (2000) ont mené des études aux champs afin d'évaluer la capacité de souches de *B. vietnamiensis* à augmenter la croissance et le rendement du riz (nombre de grains par plant). Il s'avère que *B. vietnamiensis* peut conduire à une augmentation significative de la récolte, et bien que les mécanismes impliqués n'aient pas formellement été identifiés, la capacité de *B. vietnamiensis* à fixer l'azote pourrait jouer un rôle majeur (Gillis et al., 1995). La littérature démontre peu d'exemples d'interactions positives des Bcc avec les plantes, cependant cette capacité à promouvoir le développement d'organisme ne s'arrête pas seulement aux plantes. En effet, une étude originale de Santos et al., (2004) a permis d'identifier une souche de l'espèce *B. cenocepacia* en symbiose tripartite avec la fourmi *Atta sexdens rubripilosa* et le champignon *Leucoagaricus gongylophorus*. Dans ce cadre, les trois partenaires retirent donc un bénéfice : (i) les fourmis élèvent les *L. gongylophorus* afin qu'ils prédigèrent les feuilles dont elles se nourrissent, (ii) la bactérie (qui produit par ailleurs des composés antifongiques n'ayant pas d'effets sur *L. gongylophorus*, mais protégeant la fourmi et le champignon d'autres champignons néfastes) dispose d'un apport non négligeable en substances nutritives ; (iii) le champignon, quant à lui, bénéficie d'une niche protégée et du glucose fourni par les fourmis. Ces travaux démontrent la capacité des Bcc à interagir de façon positive avec des organismes très différents et soulignent la nécessité de considérer les interactions dans un cadre multipartite plutôt que bilatéral, ce afin d'enrichir la littérature, encore peu développée à ce jour sur ce sujet.

a



b

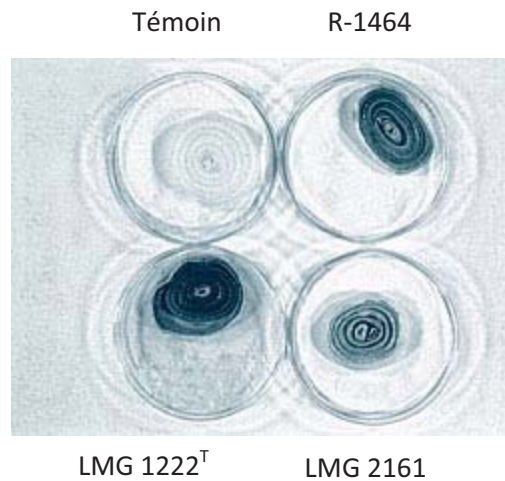


Fig. 2 : Phénotypes de pathogénie chez la plante

a : Phénotype de pourriture molle par piqure d'oignon. Seule la souche possédant le système de sécrétion de type IV (A) présente le phénotype de pourriture molle. La souche mutante pour ce système (C) ou une piqure avec de l'eau (B) stérile ne conduisent pas au phénotype. b : Infections de tranches de bulbe d'oignon par différentes souches de *B. cepacia*. La maladie causée par *B. cepacia* est la carie de l'oignon. Photos tirées du BCCM (BELGIAN CO-ORDINATED COLLECTIONS OF MICRO-ORGANISMS)

### 2.3.2 Interactions néfastes des Bcc dans l'environnement

Les effets néfastes des bactéries sur un organisme donné et dans un environnement donné, sont toujours le fruit de multiples interactions positives, négatives ou neutres entre les différents organismes présents dans l'écosystème considéré. Au sein des milieux complexes comme les sols, de nombreux acteurs (bactéries, champignons, protozoaires, nématodes, etc) peuvent interagir, c'est la résultante de toutes ces interactions qui va déterminer l'effet engendré sur l'un des organismes composant ce milieu.

Chez les Bcc, les interactions néfastes dans l'environnement ont tout d'abord été décrites au niveau de l'oignon, particulièrement pour l'espèce *B. cepacia*, puis d'autres travaux ont montré la capacité des Bcc à interagir de façon négative sur les amibes, certains champignons et les arthropodes. Dans le cas de *B. cepacia*, comme l'avait initialement décrit Burkholder, la bactérie peut causer d'importants dégâts aux cultures en détruisant les tissus de bulbes d'oignons (Burkholder 1950, Fig. 2). Dans ce contexte, deux mécanismes de virulence ont été découverts : le premier concerne une hydrolase de pectine portée par un plasmide qui conduit à la digestion de constituants de la paroi des cellules végétales (Gonzalez et al., 1997). Le deuxième concerne le phénotype de pourriture molle qui apparaît chez les souches dépourvues du plasmide décrit ci-dessus. Ce second phénotype implique un système de sécrétion de type IV qui libère une protéine provoquant la plasmolyse de protoplastes de plantes (Engledow et al., 2004). Ce phénotype est retrouvé également chez d'autres espèces du Bcc : *B. cenocepacia* et *B. vietnamiensis*. Un deuxième système de sécrétion de type IV a récemment été décrit (bc-VirB/D4) ; contrairement au premier, celui-ci n'a pas été jugé utile dans la pathogénèse de plantes, mais se serait révélé impliqué dans la mobilisation de plasmides (Zhang et al., 2009). L'éventuelle action de ces systèmes sur des cellules de mammifère reste à déterminer. Enfin, les autres espèces du Bcc n'ont pas été décrites comme phytopathogènes.

Les Bcc interagissent également avec des micro-eucaryotes unicellulaires tels que les amibes. Ces eucaryotes unicellulaires forment un groupe hétérogène d'organismes qui partagent la même morphologie et qui sont largement retrouvés dans le sol ainsi qu'en milieu hydrique. Longtemps considérés comme inoffensifs, ces micro-eucaryotes sont maintenant pris en considération dans certaines pathologies animales et humaines (Winiecka-Krusnell and Linder 2001, Pritt and Clark 2008). Leur source de nutrition comprend de nombreux micro-organismes, notamment les bactéries, qu'elles phagocytent et lysent dans des vacuoles spécialisées. Dans certains cas, les bactéries peuvent cependant résister à la lyse et persister

dans l'amibe. C'est en 1999, que Marolda et al., ont décrit les interactions entre des Bcc et des amibes. Différentes souches du Bcc ont été testées sur trois espèces d'amibes du genre *Acanthamoeba*. Les interactions apparaissaient très variables entre les espèces du Bcc et entre souches d'une même espèce. Pour comprendre le devenir des cellules phagocytées par les amibes, Lamothe et al., (2004) ont, par la suite, utilisé une série de marqueurs fluorescents pour suivre les souches H111 de *B. cenocepacia* et CEP040 de *B. vietnamiensis* chez l'amibe *Acanthamoeba polyphaga*. Ces travaux attestent que cette dernière est capable de phagocyter les bactéries, qui cependant restent présentes dans des vacuoles acides distinctes du lysosome et ainsi ne subissent pas la destruction engendrée par l'organel. Les amibes abritant des souches persistantes de *B. cenocepacia* pourraient jouer un rôle de réservoir non négligeable de cette espèce dans l'environnement. De plus, un point reste étonnant mais remarquable : cette survie intracellulaire de souches de *B. cenocepacia* est aussi retrouvée au niveau des macrophages humain. Des études complémentaires ont identifiées et démontrées le rôle d'un système de sécrétion de type VI et d'un gène régulateur de type kinase-senseur (*atsR*) dans la survie intracellulaire de deux souches de *B. cenocepacia* (Lamothe et al., 2007, Aubert et al., 2008).

Egalement présents dans le sol et l'eau, les nématodes se nourrissent aussi de bactéries avec une certaine sélectivité, contribuant ainsi à la régulation des populations bactériennes. En retour, certaines bactéries sont capables de perturber le développement des nématodes par des procédés antagonistes. *Acrobeloides sp* et *Pristionchus lheritieri*, sont deux espèces de nématodes pouvant se développer en consommant des *B. cepacia* (Rabkin et al., 1989). Carta (2000) a mené une étude visant à évaluer la diversité des interactions entre des nématodes et des Bcc. Il en ressort une grande hétérogénéité : certains nématodes peuvent ingérer toutes les souches testées tandis que d'autres sont infectés puis tués par certaines souches. Ces interactions pourraient être valorisées par une utilisation de nématodes pour contrôler les populations de Bcc ou, à l'inverse, utilisation des Bcc pour contrôler les nématodes parasites de plantes. Un autre nématode, *Caenorhabditis elegans* est retrouvé dans de nombreuses études notamment pour son utilisation comme organisme modèle dans l'identification de gènes liés à la virulence (Cardona et al., 2005).

Ces interactions entre bactéries et micro-eucaryotes apparaissent le plus souvent comme ciblées, même espèce voire même souche spécifique. Ceci révèle l'importance du dialogue moléculaire entre les deux protagonistes. Ces spécificités d'interactions pourraient être mises à profit pour limiter la prolifération d'agents considérés comme nuisibles notamment en agriculture.

#### 2.4. Les gènes "environnementaux"

La présence des Bcc dans l'environnement implique certaines "aptitudes" permettant une survie dans des écosystèmes très hétérogènes. Un large panel de processus biochimiques peut ainsi être mis en œuvre par ces bactéries pour métaboliser les molécules plus ou moins complexes de l'environnement dans le but de couvrir leurs besoins en azote, en carbone, en phosphore, en métaux, etc. Ceci suppose un pool de gènes pouvant être catalogués comme environnementaux, permettant le support de cette plasticité génomique relative aux divers milieux rencontrés. Certains d'entre eux semblent avoir une large distribution et sont retrouvés dans la majorité des espèces du Bcc. A titre d'exemple, nous pouvons citer le cas du gène bactérien *acdS* qui code une ACC désaminase ayant un effet positif sur la croissance des plantes (Onofre-Lemus et al., 2009), du cluster *arsRCB* impliqué dans la résistance à l'arsenic, et du cluster de gènes *iac* reliés au catabolisme d'une hormone végétale, l'AIA (Leveau and Gerards 2008). Cette liste non-exhaustive de gènes étaye clairement la présence d'un support génétique permettant aux espèces du Bcc de coloniser différents environnements. Toutefois, quelques travaux seulement ont testé l'activité de ces gènes, concluant que la plupart d'entre eux ne sont pas forcément actifs, bien que présents. Désormais, les récentes études ne se contentent plus d'évaluer la présence ou l'absence de certains gènes mais enrichissent ces travaux par des aspects de transcriptomique voire même de protéomique, à large échelle (Yoder-Himes et al., 2009). En effet, la présence d'un gène n'implique pas obligatoirement son expression. La régulation de l'expression génique permet ainsi un contrôle fin des phénotypes associés. Divers systèmes sont à ce jour identifiés tels que les systèmes à double composantes intervenant dans le quorum-sensing ou bien les petits ARNs non codants. Evaluer le transcriptome permet dorénavant d'estimer la part de gènes effectivement exprimés dans un environnement donné et de mesurer la régulation de l'expression des gènes par ce milieu.

D'autre part, d'autres gènes ou clusters de gènes ayant une distribution plus restreinte au sein du Bcc, ont pu être mis en évidence. Le cluster *nif*, par exemple, impliqué dans la fixation d'azote atmosphérique, a été détecté uniquement chez l'espèce *B. vietnamiensis* (Estrada-De Los Santos et al., 2001). De plus, le génome de *B. vietnamiensis* contient de nombreux gènes permettant de dégrader un large panel de composés tels que les hydrocarbures aromatiques ou chlorés, certains composés phénoliques, et de nombreux xénobiotiques (O'Sullivan et al., 2007, Van Nostrand et al., 2007, Govindarajan et al., 2008). Cette espèce apparaît être l'espèce du Bcc la plus apte à coloniser l'environnement. Sa capacité à fixer l'azote ainsi qu'à

métaboliser divers composés représente de réels atouts pour cette bactérie, pouvant expliquer sa distribution élevée au niveau des sols.

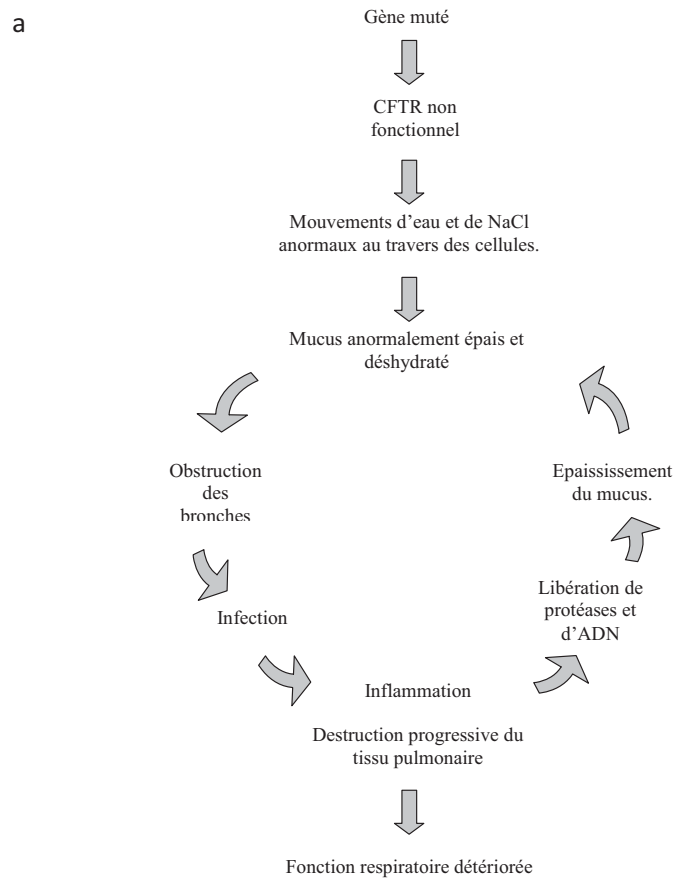
Tous ces travaux démontrent une présence des Bcc au sein de milieux naturels très hétérogènes ; cependant l'identification des réservoirs de Bcc retrouvés en milieux cliniques est un challenge qui ne fait que débiter. Les investigations réalisées dans l'environnement ont certes permis l'identification de nombreux isolats de Bcc, mais peu d'entre elles ont formellement mis en évidence des souches cliniques ou trouvées des niches avec des populations élevées de Bcc qui seraient au niveau des doses infectantes. Cette difficulté est probablement due à de nombreux paramètres liés aux techniques utilisées, aux sites explorés ou tout simplement à une réelle absence de ces souches dans les environnements prospectés. En perspective, de nombreuses études doivent être menées afin de compléter notre connaissance relative à l'écologie des Bcc dans le milieu naturel.

### **3. *Burkholderia* du complexe *cepacia* chez l'homme**

Les espèces du Bcc sont classées dans la catégorie des pathogènes "opportunistes", ceci signifie qu'elles affectent principalement des personnes immunodéprimées ou fragilisées. Dans ce contexte, les principales populations humaines sensibles aux bactéries du Bcc sont les patients atteints de mucoviscidose (CF). Néanmoins, il existe des cas documentés d'infections nosocomiales et d'infections de patients présentant une granulomatose septique chronique (CGD) (van den Berg et al., 2009).

#### *3.1. Infections communautaires*

Les infections communautaires correspondent à toutes les infections survenues en dehors des hôpitaux, impliquant des contaminations croisées directes ou indirectes, et déclarées dans les 48h premières heures suivant l'hospitalisation. La principale infection communautaire impliquant des Bcc est l'infection de patients CF décrite dans la section suivante. Il existe par ailleurs quelques cas infectieux se rapprochant d'infections communautaires impliquant des solutions de bain de bouche ou de matériel contaminé. Cependant, si l'on écarte les études concernant les patients CF ou les infections nosocomiales, aucun autre travail ne présente de transmission de Bcc entre individu. Etant donné que la frontière entre



b

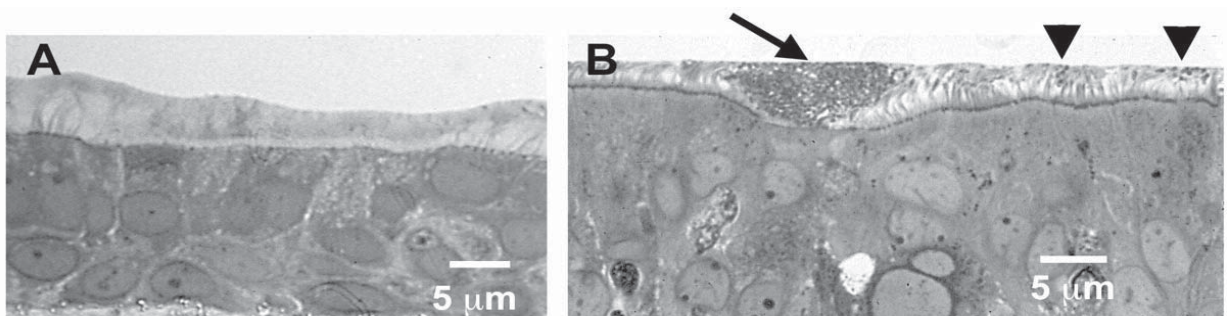


Fig. 3 : Cascade physiopathologique des maladies pulmonaires chez les patients atteints de mucoviscidose. a : successions des étapes conduisant à la dégradation de la fonction respiratoire lors des infections bactériennes. (D'après Boas et al., 2000). b : Infections de cellules épithéliales du poumon par *B. cenocepacia*. A : cellules épithéliales en culture. B : cellules épithéliales infectées par une souche de *B. cenocepacia*. Les bactéries sont seules (triangles) ou organisées en biofilm (flèche) (D'après Schwab et al., 2002).



infections nosocomiales et infections communautaires est complexe à déterminer et que les travaux réalisés ne renseignent pas toujours la chronologie exacte des événements, nous ne traiterons par conséquent que des aspects impliquant des contaminations par l'intermédiaire de composés industriels (section 3.3).

### 3.1.1. Cas de la mucoviscidose

La mucoviscidose est une maladie génétique, affectant les épithéliums glandulaires de nombreux organes. C'est, de nos jours, la maladie la plus fréquente dans les populations de type caucasienne, tandis qu'elle est très rare dans les populations africaines et asiatiques. Elle affecte de nombreux organes, toutefois les atteintes respiratoires sont prédominantes, représentant l'essentiel de la morbidité. La mucoviscidose est associée à des mutations du gène *cftr* sur le chromosome 7. L'altération de la protéine CFTR provoque une augmentation de la viscosité du mucus et son accumulation dans les voies respiratoires et digestives, conduisant ainsi à un environnement privilégié pour la colonisation bactérienne (Fig.3). Les personnes déficientes sont sensibles à différentes bactéries, particulièrement *Pseudomonas aeruginosa*. Suivant leur âge, les patients CF sont préférentiellement colonisés par différents types de bactéries. Chez les plus jeunes, les bactéries majoritairement retrouvées sont *Staphylococcus aureus*, *Haemophilus influenzae* tandis que, chez les plus âgés *Pseudomonas aeruginosa* devient prédominant ; à l'âge adulte, 80% des patients sont colonisés par cette dernière (Harrison 2007 et Fig. 4). Les premiers cas d'infections impliquant des Bcc chez les patients CF ont été répertoriés entre la fin des années 1970 et le début 1980 (LiPuma 1998). Consécutivement, la première description clinique détaillée fut publiée par Isles et al., en 1984. Une augmentation de l'incidence de ces infections dans différents centres de patients CF a été démontrée en Europe et au nord des Etats-Unis (Tablan et al., 1985, Thomassen et al., 1985). La succession bactérienne observée au cours du vieillissement des patients CF reste méconnue ; dans le cas d'infections à *P. aeruginosa*, on observe initialement des infections transitoires par différentes souches évoluant en infection chronique par une souche dominante. Dans le cas de Bcc, ce scénario semble être similaire bien qu'il existe des cas d'infections transitoires avec plusieurs souches et le remplacement de l'une d'entre elles dans la phase chronique par une autre (Coenye and Vandamme 2003, Lipuma 2005). Les Bcc sont plutôt retrouvées chez des patients âgés et notamment en co-infection avec d'autres bactéries (Fig. 4 et 5). Les différentes étapes d'une infection par les Bcc commencent à être décryptées mais de nombreux points restent mal compris. Présentes au niveau des poumons, les bactéries

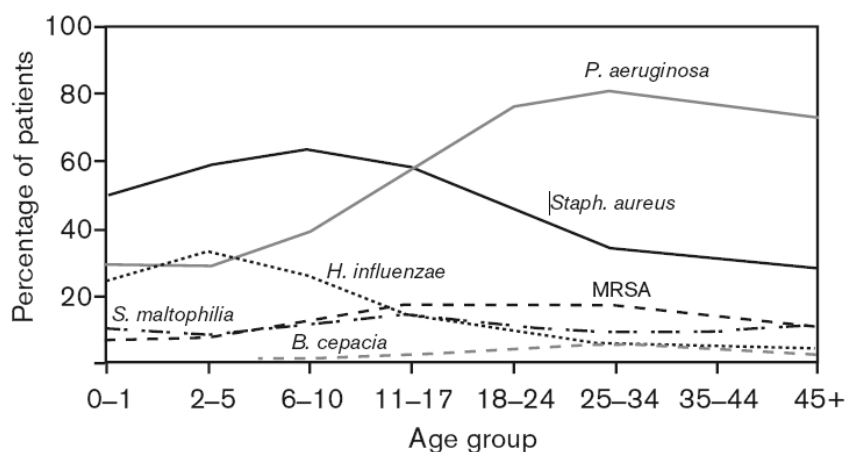


Fig. 4. Prévalence des infections bactériennes des voies respiratoires, selon les groupes d'âge chez les patients CF. D'après Harrison (2007).

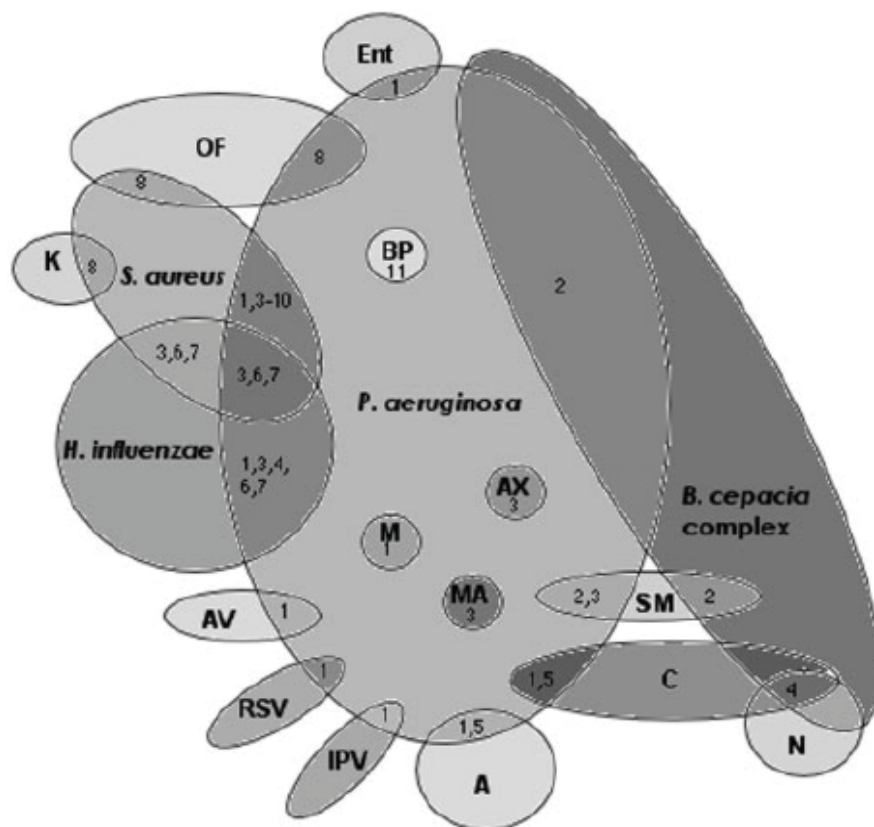


Fig. 5. Diagramme de Venn montrant les co-infections des voies respiratoires de patients CF. (NB: la co-infection ne signifie pas nécessairement l'interaction directe entre les espèces.) A, *Aspergillus* spp.; AV, adenovirus; AX, *A. xylosoxidans*; BP, bacteriophage; C, *Candida* spp.; Ent, enterobacteria; IPV, influenza and/or parainfluenza virus; K, *Klebsiella* spp.; M, mycoplasma; MA, *Mycobacterium abscessus*; N, *Neisseria* spp.; OF, oropharyngeal flora; RSV, respiratory syncytial virus; SM, *S. maltophilia*. Les chiffres indiqués correspondent aux références bibliographiques e. g. Harrison (2007).

expriment différents facteurs de virulence régulés en partie par quorum-sensing (Uehlinger et al., 2009). En conséquence, il survient une réponse inflammatoire importante conduisant à la lésion des tissus pulmonaires. Une persistance de la bactérie viable dans les cellules immunitaires est liée aux phénomènes d'apparente rémission (cultures négatives) mais peut conduire à une rechute avec les mêmes souches (Lamothe and Valvano 2008, Saldias and Valvano 2009). Dans certaines conditions (mal déterminées), la bactérie peut devenir invasive et provoquer une septicémie (Gautam et al., 2008).

Dans les années 1980, différents clones épidémiques ont été identifiés au sein de l'espèce *B. cenocepacia* (section 1.5). Lipuma *et al.* (1990) ont mis en évidence la possibilité de transmission inter-individuelle de souches du Bcc. Des mesures de prophylaxie sanitaire ont été adoptées afin de limiter les contacts entre patients, et des mesures d'isolement ont été mises en place dans les services. Ces mesures semblent avoir engendré une diminution des cas d'infection à *B. cenocepacia* dans les pays où elles étaient prédominantes. Cette réduction se fait malheureusement au profit de l'augmentation des cas d'infections à *B. multivorans*, tout en maintenant le nombre de cas. (Reik et al., 2005, Lynch 2009). Les actions menées semblent donc avoir permis de contenir *B. cenocepacia*, démontrant ainsi que cette espèce possède une bonne capacité de transmission inter-individus. Les cas groupés à *B. multivorans* semblent moins fréquents. Les infections à *B. multivorans* seraient principalement le résultat d'acquisitions isolées depuis l'environnement (Baldwin et al., 2008).

Les co-infections impliquant les Bcc et *P. aeruginosa* sont fréquemment rencontrées au niveau des poumons de patients atteints de mucoviscidose (Lambiase et al., 2006). Cette proximité implique de multiples interactions entre ces genres bactériens. Weaver et Kolter en 2004 ont réalisé un criblage des gènes de *P. aeruginosa* induits en présence de Bcc. Par l'intermédiaire de cette technique, ils ont identifié un sidérophore, l'ornibactine, sécrété par l'ensemble des *Burkholderia* et induisant une forte réponse chez *P. aeruginosa*. Une étude supplémentaire démontrant la sécrétion de sidérophores retrouvés dans les expectorations de patient CF, semble indiquer l'importance d'une compétition pour le fer dans l'environnement des poumons de patients CF (Haas et al., 1991). Pour comparer la réponse initiale de l'organisme à ces deux germes, Worgall *et al.* (2005) ont évalué l'impact de ces bactéries sur les macrophages alvéolaires. Une modification similaire de l'expression des gènes de ces deux pathogènes a été observée et les interactions précoces hôtes-pathogènes semblent donc identiques dans ce cas là. L'évolution différente des infections à *P. aeruginosa* et à *B. cenocepacia* semble donc se faire plus tardivement. Il a également été démontré que *P. aeruginosa* PAO1 est capable d'augmenter la pathogénèse de *B. cenocepacia* H111 via le

Tableau 1. Incidence des espèces (Genomovars) parmi les Bcc isolés chez les patients CF.

Country (no. isolates)	Genomovar							
	I (%)	II (%)	III (%)	IV (%)	V (%)	VI (%)	VII (%)	Other (%)
Australia <sup>60</sup> (n=28)	28.6	21.4	39.3	0	3.6	0	7.1	0
Australia <sup>61</sup> (n=116)	11.2	29.3	45.7	0	5.2	0.8	5.2	2.6
Brazil <sup>65</sup> (n=41)	3.1	37.5	53.1	0	6.2	0	0	0
Brazil <sup>56</sup> (n=59)	0	10	67	0	15	0	7	0
Canada <sup>37</sup> (n=461)	0.2	9.3	80	3.8	1.6	0	0	5.3
Canada <sup>14</sup> (n=62)	0	30	74	2	2	0	0	2
Czech Republic <sup>64</sup> (n=61)	0	5	92	1.6	0	0	0	1.6
Slovakia <sup>64</sup> (n=24)	0	0	41	54	0	0	0	5
France <sup>79</sup> (n=153)	0	53	45	1.3	0.7	0	0	1.3
Italy <sup>78</sup> (n=48)	0	0	70.8	0	0	0	6.3	22.9
Italy <sup>158</sup> (n=31)	3	0	90.3	3	3	0	0	0
Italy <sup>98</sup> (n=60)	8.3	5	90	3.3	0	0	0	0
Italy <sup>11</sup> (n=75)	0	0	90	6.7	0	0	0	0
Italy <sup>41</sup> (n=208)	7.7	6.3	61.1	6.3	1	0	0.5	17.3
Portugal <sup>83</sup> (n=22)	36.4	9.0	52.2	18.2	0	0	0	0
Portugal <sup>82</sup> (n=94)	24.5	9.5	50	16	0	0	0	0
UK <sup>17</sup> (n=50)	0	32	66	0	0	0	0	2
UK, Europe <sup>77</sup> (n=131)	0.8	17.6	78.6	0.8	3	0	0	0
USA <sup>51</sup> (n=1218 CF)	3.1	38.7	45.6	0.3	5.9	3.8	0.8	1.8
USA <sup>74</sup> (n=606)	2.6	37.8	50.0	0.2	5.1	2.0	0.7	1.6

Les numéros en exposant correspondent aux différentes références bibliographiques citées par Lynch 2009.

quorum-sensing, conduisant ainsi à une surexpression des facteurs de virulence chez H111 (Riedel et al., 2001). Cette communication a cependant été démontrée comme unilatérale, les molécules secrétées par H111 n'ayant pas d'effet sur *P. aeruginosa*. Il apparaît clairement que de nombreuses interactions se produisent lors de co-infections et que celles-ci jouent un rôle majeur dans l'évolution des pathologies.

La distribution des différentes espèces du Bcc chez les patients CF a été le sujet de nombreuses études. Cependant, la taxonomie du Bcc est en perpétuelle évolution ce qui rend délicate l'étude de la proportion relative des différentes espèces. Il apparaît toutefois que la prévalence des espèces est fortement hétérogène, variant en fonction du temps et des zones géographiques considérées. Deux espèces, *B. cenocepacia* (genomovar III) et *B. multivorans* (genomovar II) représentent 80 à 93% des isolats 2000 de patients CF aux Etats-Unis, au Canada, en Italie, et en Europe (Tableau 1, Lynch 2009). Le genomovar I, *B. cepacia*, compte pour 0,2 à 8,3% des infections à Bcc chez les patients CF en Amérique du Nord et en Europe. Toutefois, des taux élevés ont été retrouvés en Australie (11,2 à 28,6%) et au Portugal (50% des isolats de Bcc de 1995 à 2006). Cette haute incidence au Portugal reflète la prédominance de deux clones de *B. cepacia*, probablement impliqués dans la contamination de solutions salines (Cunha et al., 2007, Leitao et al., 2008). Les autres genomovars (IV-IX) comptent pour un total inférieur à 10% des isolats de Bcc dans la plupart des centres de patients CF. Néanmoins, certains centres présentent de fortes disparités dans la distribution de ces espèces, par exemple, *B. stabilis*, est retrouvé dans 54% des cas en Slovaquie (Drevinek et al., 2003). Ceci souligne les différences qui règnent entre les organismes d'accueil pour patients CF, les pays et le temps. Ainsi, aux États-Unis *B. multivorans* et *B. cenocepacia* représentaient une proportion supérieure à 83% des Bcc isolées au cours de deux grandes études (Tableau 1, LiPuma et al., 2001, Reik et al., 2005). Parmi les 556 *B. cenocepacia* isolées sur des patients CF aux Etats-Unis, 27,2% appartenaient au groupe RFLP-*recA* IIIA, 72,8% au groupe IIIB et aucun isolat n'appartenait aux groupes IIIC et IIID. Contrairement aux États-Unis, le sous-groupe IIIA de l'espèce *B. cenocepacia* prédomine au Canada et en Europe, reflétant ainsi la propagation clonale de la souche ET12 (Mahenthalingam et al., 2008). Dans une vaste étude de 208 isolats provenant de 18 centres CF en Italie, *B. cenocepacia* représentait 60,1% et *B. multivorans*, 6,3% des souches (Campana et al., 2005). Des variations similaires ont été retrouvées pour les centres de patients CF au Royaume Uni et de Belgique. Globalement, ces études ne considèrent pas ou peu l'existence de clones épidémiques, conduisant ainsi à un biais dans les résultats observés. En conséquence, l'évaluation de la morbidité associée à une espèce dans son ensemble ne peut pas toujours clairement être établie. En France, les espèces

prédominantes parmi les patients CF appartenant au Bcc et leur répartition est la suivante : *B. cenocepacia* (45,8%) ; *B. multivorans* (41,4%) ; parmi les autres espèces, *B. vietnamiensis* vient en tête (3,3%), puis *B. stabilis*, *B. dolosa* et *B. pyrrocinia* (0,8%). Dans ce contexte, certains patients ont été colonisés puis infectés à la fois par *B. cenocepacia* et *B. multivorans* et deux patients ont été atteints par des souches appartenant à des genomovars non décrits à cette époque. Enfin, parmi les espèces n'appartenant pas au complexe *Burkholderia cepacia*, *B. gladioli* est la plus fréquemment isolée (6,7%) puis *B. thailandensis* (0,8%). Le clone *B. cenocepacia* ET12 n'a jamais été retrouvé en France, en revanche il existe 5 clones transmissibles, dont 2 appartiennent à l'espèce *B. multivorans* et 3 à l'espèce *B. cenocepacia*, du genomovar IIIA (1/3), IIIB (2/3) (Segonds C., Observatoire National *Burkholderia cepacia* ; Bilan épidémiologique 2004, Abalain et al., 2009).

### 3.2. Infections nosocomiales

Les bactéries du Bcc peuvent être retrouvées dans le cadre d'infections nosocomiales et peuvent provoquer des épidémies localisées dans des services hospitaliers. Dans plusieurs cas rapportés, ces épidémies ont pu durer jusqu'à plusieurs mois avec un nombre de cas groupés réduits. La gravité des infections à Bcc est souvent due au fait qu'elles peuvent évoluer en septicémie (Huang et al., 2001). L'origine de ces épidémies a pu être identifiées, fréquemment, au niveau du réseau d'eau (douche, robinet, siphon...), du matériel médical, ou de solutions contaminées (Panlilio et al., 1992, van Laer et al., 1998, Loukil et al., 2003, Nasser et al., 2004, Ghazal et al., 2006, Lo Cascio et al., 2006, Alvarez-Lerma et al., 2008). Ces bactéries ont une versatilité métabolique importante leur donnant la capacité de se développer sur une grande variété de sources carbonées, voire même dans des solutions de Bétadine® (Panlilio et al., 1992). Dans les cas cités, les épidémies ont pu être soit très brèves (1 semaine), soit très longues (35 mois). Les méthodes de typage moléculaire utilisées n'ont pas démontré de différence entre les isolats d'une même épidémie. Il est intéressant de noter que certains clones isolés de patients CF ont été retrouvés comme agent causal d'infections nosocomiales (Vandamme et al., 1997). De nombreuses autres études relatant des épidémies nosocomiales à Bcc ont été publiées ; cependant les outils employés pour la caractérisation des souches ne permettaient pas une identification formelle des espèces impliquées. La prévalence de clones épidémiques de l'espèce *B. cenocepacia* au niveau d'infections nosocomiales semble donc sous-estimée et il apparaît nécessaire d'investiguer l'incidence de cette espèce au niveau de ce type d'épidémie.

### 3.3. Prophylaxie

De nombreuses précautions ont été mises en place depuis l'émergence des Bcc, dans le but de lutter contre les infections de patients CF. La ségrégation des patients infectés au Bcc semble avoir été la mesure la plus adaptée, permettant une réduction non négligeable des infections au niveau de la communauté CF. Un exemple phare est l'étude de Thomassen et al. réalisée en 1986, qui met en évidence une transition de 8.2% de patient infectés/colonisés en 1983 à 1.7% en 1984, après mise en place de cette ségrégation. De même, un travail réalisé en 1992 dans un centre pour patients CF de Manchester montrait une diminution de plus de 28 % de l'incidence des Bcc après une ségrégation stricte de tous les patients CF (France et al., 2008). Ces observations ont été confirmées par l'identification d'une augmentation de la transmission de souches du Bcc entre patients dans des centres où la ségrégation n'avait pas été établie (Paul et al., 1998, Chen et al., 2001). Bien que très contraignante pour les malades, cette ségrégation semble être la seule alternative permettant une réduction notable de la transmission de Bcc aux patients CF.

Les Bcc apparaissent donc comme des agents pouvant provoquer de nombreuses infections, principalement pulmonaires, chez l'Homme. L'origine des souches est souvent mal connue, rendant délicate la mise en place de mesures de prophylaxie sanitaire supplémentaire à la ségrégation. Un enrichissement des connaissances de l'écologie de ces bactéries ainsi que des processus impliqués dans leur virulence apparaît donc comme fondamental.

## 4. Mécanismes de la Pathogénèse

La pathogénèse est l'étude du ou des mécanismes responsables du déclenchement et du développement d'une maladie. Elle fait intervenir souvent plusieurs mécanismes pathologiques pour une seule maladie, dans lesquels les gènes, l'environnement, le conditionnement, l'individu et de très nombreux autres facteurs jouent des rôles importants. La pathogénèse se focalise donc sur le déroulement et les processus mis en jeu par l'organisme au cours d'une maladie. De nombreux éléments semblent jouer un rôle dans la colonisation,

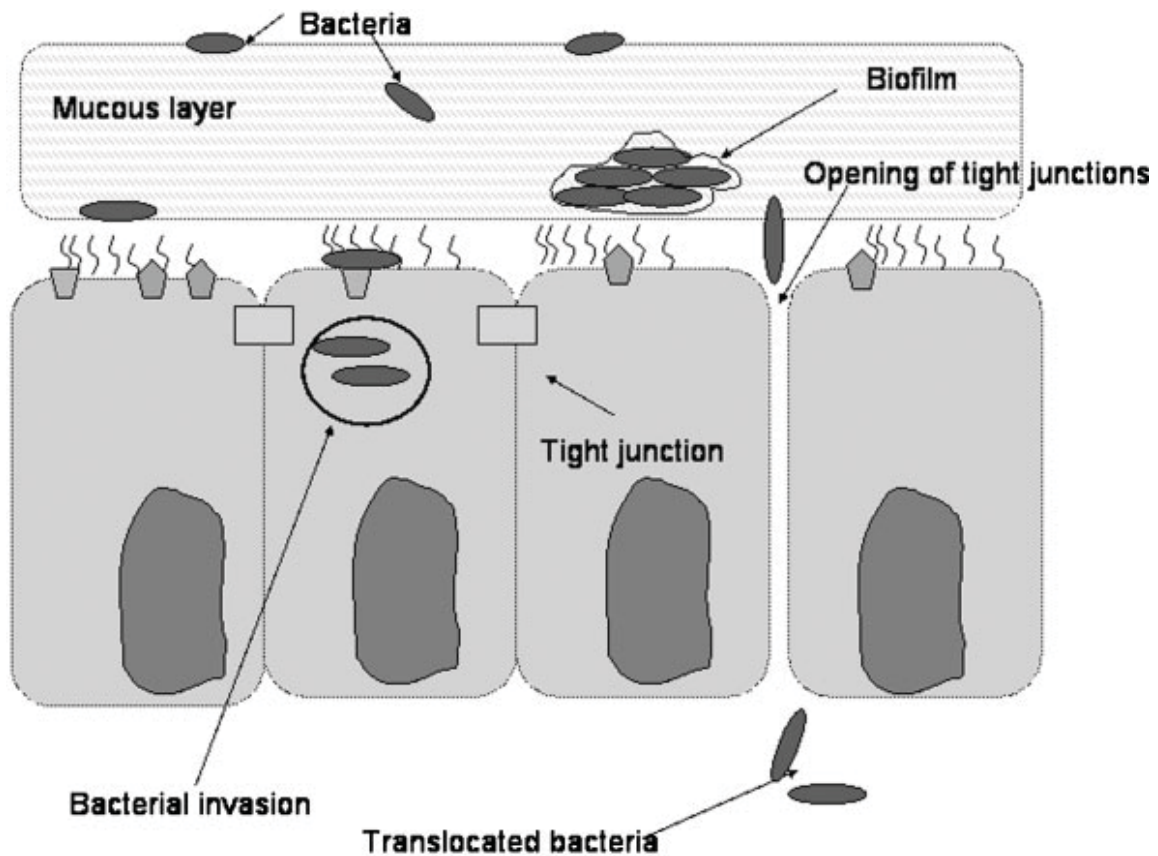


Fig. 6. Schéma de cellules épithéliales du poumon et leurs interactions avec les Bcc. Les bactéries se fixent à des récepteurs sur les cellules épithéliales du poumon et peuvent soit l'envahir par l'intermédiaire de vacuoles intracellulaires ou par translocation à travers l'épithélium. Les Bcc peuvent aussi former des biofilms, qui ont une interaction différente avec des cellules épithéliales du poumon. D'après McClean and Callaghan 2009.



l'invasion et l'établissement des souches de Bcc dans les poumons de patients CF (Fig. 6). Ces mécanismes, communs aux cellules épithéliales, comprennent la fixation aux cellules, l'invasion, la survie intracellulaire, l'acquisition du fer et la régulation des gènes de virulence (Finlay and Falkow 1997).

#### 4.1. Adhérence aux cellules épithéliales

L'adhérence du micro-organisme à l'épithélium de surface est la première étape dans l'interaction avec la cellule hôte. Les Bcc adhèrent aux cellules épithéliales par l'intermédiaire de récepteurs protéiques, glycolipidiques, et via la mucine sécrétée par la cellule hôte. Sajjan et al., (1992) ont démontré la capacité des Bcc à se lier à des mucines purifiées ; toutefois, l'identification exacte des espèces auxquelles appartenaient ces isolats n'a pas été réalisée. Néanmoins, il a été montré que ces souches exprimaient le pilus câble, caractéristique de la lignée ET12 de l'espèce *B. cenocepacia*. Le récepteur de la cellule épithéliale pour le câble pilus, lui-même associé à une adhésine de 22 kDa, a été identifié comme étant la cytokératine 13, une protéine de 55 kDa présente en grande quantité au niveau des cellules épithéliales de patients CF (Sajjan et al., 1992). Cependant, les interactions entre la cytokératine 13 et l'adhésine de 22 kDa semblent moins évidentes. En effet, des études ultérieures ont démontré que les souches de *B. cenocepacia*, ne présentant pas le pilus câble à leur surface, pouvaient également se lier à la cytokératine 13. En dépit de cela, l'adhésine de 22 kDa impliquée dans l'attachement reste, tout de même, un acteur potentiel dans la pathogenèse, puisque le blocage de son activité par des anticorps avant l'infection de cellules épithéliales par la souche LMG18826 de *B. cenocepacia* réduit à la fois la cytotoxicité et la sécrétion d'IL-8 dans des cultures primaires de cellules épithéliales bronchiques (Sajjan et al., 2002). En définitive, l'expression combinée de l'adhésine de 22 kDa et du pilus câble est nécessaire pour une adhésion optimale à la cytokératine 13 (Urban et al., 2005). Fait intéressant, ce mécanisme semble limité à certaines souches de *B. cenocepacia*, et reste peu répandu chez les souches de la lignée ET12. Plus récemment, une étude a démontré que la souche LMG18826 pouvait également interagir avec un second récepteur, le TNF1 (TNFR1), protéine de 55 kDa. Cette interaction, n'utilisant pas la protéine de 22 kDa, est considérée comme partiellement responsable de la puissante réponse inflammatoire provoquée par la présence de souches virulentes du Bcc (Sajjan et al., 2008b). Il reste cependant à déterminer l'importance du TNFR1 pour les autres membres de l'espèce *B. cenocepacia*, et du Bcc.

La deuxième classe de récepteurs nécessaires pour l'adhésion bactérienne et présents au niveau de la cellule hôte concerne les récepteurs glycolipides. Krivan et al. ont démontré en 1988 que les membres du Bcc présentaient des glyco-sphingolipides de surface contenant des motifs permettant la liaison aux aGM1 et aGM2 (asialoganglioside GM1 et 2) des cellules épithéliales. Cette dernière a été évaluée comme plus faible chez les souches ne présentant pas le câble pilus, suggérant ainsi que l'absence ou la présence de ce pilus affecte l'affinité des souches pour ces récepteurs glycolipides (Sylvester et al., 1996). Toutefois, le rôle ultime de ces récepteurs dans la pathogénèse des Bcc n'a pas été exploré. Le glycolipide aGM1 est situé à la surface apicale des cellules épithéliales pulmonaires, il a été identifié comme permettant la fixation des flagelles chez *P. aeruginosa* et semble disponible/accessible par les bactéries présentes dans le poumon (Adamo et al., 2004). Une étude étonnante a démontré que, bien que *P. aeruginosa* et les Bcc puissent se lier aux aGM1, aucune compétition pour la liaison ne semblait s'opérer entre ces deux agents pathogènes ; au contraire, l'attachement des souches de Bcc est renforcée en présence de surnageant de *P. aeruginosa* (Saiman et al., 1990). Malgré le fait que le mécanisme moléculaire ne soit pas identifié, il apparaît évident que ces deux pathogènes interagissent, et cela au détriment du patient CF. Dans les épithéliums polarisés, les récepteurs aGM1 sont généralement présents occasionnellement sur la surface apicale des cellules saines intactes ; cependant, la quantité de ces récepteurs augmente considérablement suite à une lésion ou dans les épithéliums de patients CF. Ceci pourrait en partie expliquer la prédisposition de cette population aux agents pathogènes (Saiman and Prince 1993). Ces récepteurs semblent également exploités par d'autres agents pathogènes comme *Shigella* et *E. coli*, et pourraient jouer un rôle non négligeable dans la réponse inflammatoire (Samuel et al., 1990, Stapleton et al., 1998, Riethmuller et al., 2006).

#### 4.2. Invasion des cellules épithéliales

De nombreuses études ont démontré que les Bcc sont capables d'envahir et de survivre dans les cellules épithéliales tant *in vitro* que *in vivo* (Martin and Mohr 2000, Cieri et al., 2002, Duff et al., 2006). Différentes voies d'invasion de l'épithélium pulmonaire ont été observées selon les espèces de Bcc. Ces processus peuvent comprendre notamment la formation de biofilm et/ou le réarrangement du cytosquelette, entraînant la destruction de la cellule et ainsi la pénétration à travers l'épithélium par paracytose (Schwab et al., 2002). Des récepteurs situés à plusieurs endroits des cellules épithéliales semblent aussi intervenir dans les mécanismes d'invasion utilisés par les espèces du Bcc (Duff et al., 2006). Les processus de

réarrangement de l'actine par l'espèce *B. cenocepacia* ont été confirmés, et un point particulièrement intéressant met en évidence ces réarrangements à partir de bactéries viables et non viables (Sajjan et al., 2006). Il a également été démontré que des souches de *B. multivorans* étaient capables de provoquer des disruptions du réseau de filaments d'actine, permettant ainsi une translocation dans les cellules épithéliales du poumon. En revanche, aucune réorganisation des filaments d'actine ne semble requise pour l'entrée de *B. multivorans* par l'intermédiaire de biofilm. Outre ces réarrangements, des flagelles semblent également impliqués dans l'invasion. Dans ce contexte, des mutants non mobiles de *B. cenocepacia* J2315 présentant une inactivation des gènes *fliG* ou *fliI* ont, tout deux, montré une réduction dans l'invasion de cellule en culture (Tomich et al., 2002). De plus, les lipases produites en grande quantité chez les *Burkholderia*, semblent impliquées dans l'invasion ; cependant, les mécanismes associés ne sont pas connus (Mullen et al., 2007). D'autre part, le devenir intracellulaire des souches du Bcc a été exploré par la mise en contact de la souche *B. cenocepacia* K56-2 avec des lignées de cellules épithéliales de cellules de patients CF (lignée IB3) (Sajjan et al., 2006). Il en résulte que les bactéries vivantes paraissent éviter la digestion lysosomale par leur présence dans le réticulum endoplasmique dérivé de l'autophagosome, tout en conservant leur capacité de division. En revanche, dans cette étude, les cellules non-viables de K56-2 étaient digérées dans les 6 h après internalisation dans les cellules épithéliales. Il semble donc que l'invasion des cellules épithéliales par des souches du Bcc soit un processus complexe faisant intervenir de nombreux mécanismes plus ou moins caractérisés chez les Bcc. Récemment, un système de sécrétion de type IV a été décrit comme majeur dans ce processus d'invasion et de multiplication intracellulaire (Sajjan et al., 2008a).

#### 4.3. Translocation au travers de l'épithélium

Une des raisons principales de la mortalité des patients CF colonisés par les Bcc repose sur le développement d'une bactériémie chez certains d'entre eux, conduisant à un déclin rapide. Dans ce contexte, les Bcc doivent pénétrer à travers l'épithélium pulmonaire, et ainsi accéder au système sanguin. Différents travaux ont démontré la capacité des Bcc à perturber les jonctions serrées intercellulaires, leur permettant ainsi une translocation de la partie apicale à la partie basolatérale de cellules épithéliales *in vitro*. Ce processus a été identifié chez quatre espèces différentes du Bcc (*B. cepacia*, *B. cenocepacia*, *B. multivorans* et *B. stabilis*) à des niveaux comparables de translocation, justifiant ainsi le fait que la "capacité" d'engendrer une septicémie n'est pas seulement limitée aux deux espèces les plus

virulentes (Kim et al., 2005, Duff et al., 2006). Dans ces deux études, les protéines des jonctions serrées se sont révélées être perturbées, et le surnageant seul paraissait contribuer partiellement à cet effet. La perturbation des jonctions serrées a également été reliée aux mécanismes de translocation chez d'autres bactéries, comprenant les espèces *Salmonella enterica*, *E. coli* et *P. aeruginosa* (Jepson et al., 2000, Simonovic et al., 2001, Zulianello et al., 2006).

Le choix des isolats et des souches utilisées dans ces études est la clé de la réussite pour l'élucidation des mécanismes de pathogenèse des Bcc chez les patients CF. La caractérisation des mécanismes au niveau d'une souche est importante, mais il est ainsi alors nécessaire d'élargir ces investigations en incluant plusieurs isolats cliniques pour confirmer les tendances. Différents facteurs liés à l'hôte doivent également être considérés. A titre d'exemple, lors d'une épidémie, une même souche engendre des syndromes variés au sein des patients. Les résultats cliniques des infections à Bcc resteront délicats à prédire au sein des patients CF ; néanmoins la caractérisation des différents processus impliqués dans la pathogenèse pourra permettre l'identification de nouvelles cibles, et par voie de fait contribuera au développement de traitements préventifs ou curatifs indépendants des "cocktails d'antibiotiques".

## **5. Déterminants de la virulence**

La virulence désigne l'aptitude d'un germe pathogène à se multiplier dans un organisme vivant et à y entraîner des manifestations pathologiques. Ses caractéristiques multifactorielles et polygéniques contraignent les études à ce propos qui ne peuvent en aucun cas résoudre ce mystère en ne considérant qu'un seul gène. De nombreux facteurs potentiellement impliqués dans la virulence ont été décrits chez les Bcc, et une revue sur ce sujet a été publiée par Mahenthiralingam *et al.* en 2005. Elle décrit l'ensemble des facteurs identifiés à ce jour ainsi que les modèles d'interactions utilisés pour leur évaluation. La liste des facteurs de virulence identifiés a été reprise dans le tableau 2. Il apparaît qu'au sein du Bcc, la distribution de ces facteurs est fortement hétérogène. Ceci doit être mis en relation avec la pathogénie variable des espèces et des souches. Une brève description de ces éléments est présentée dans les sections suivantes. La distribution des facteurs listés ici n'est pas nécessairement connue pour l'ensemble du Bcc, certaines études s'étant focalisées sur une ou deux espèces.

Tableau 2: Liste des facteurs de virulence présents chez les Bcc, distribution et fonctions. D'après Mahenthiralingam *et al.* (2002).

Facteurs de virulence	Occurrence chez les Bcc	Caractéristiques
Résistance aux antibiotiques	Toutes les espèces	Résistance intrinsèque forte aux antibiotiques et aux désinfectants
Formation de biofilm	Etudié pour les espèces I à V (valable pour toutes ?).	Les espèces <i>B.cenocepacia</i> et <i>B.multivorans</i> forment le plus de biofilm. Soumis à régulation du quorum sensing
Cable pili et 22kDa adhesin	<i>B.cenocepacia</i> (homologues chez <i>B.cepacia</i> )	Nécessaire pour l'adhérence aux cellules épithéliales, les isomutants qui n'ont pas de cable pili s'auto-agrègent.
Catalase et SOD	Toutes les espèces	La résistance aux formes réactives de l'oxygène pourrait aider à la survie intracellulaire.
Ilot de pathogénie de <i>cenocepacia</i>	Seulement <i>B.cenocepacia</i>	Premier ilot de pathogénie identifié, comprend le BCESM, code un second système de quorum-sensing, nécessaire pour la persistance et l'inflammation dans le modèle d'infection du poumon du rat.
Exopolysaccharide	<i>B.cepacia</i> et <i>B.cenocepacia</i>	Produit par <i>B.cepacia</i> in vitro et induit après passage sur le modèle murin chez <i>B.cenocepacia</i> .
Protéase extracellulaire	<i>B.cepacia</i> , <i>B.cenocepacia</i> , <i>B.stabilis</i> , <i>B.ambifaria</i> , <i>B.pyrricinia</i>	La métalloprotéase à zinc de <i>B.cenocepacia</i> est bien caractérisée, nécessaire pour la virulence de la souche K56-2 mais pas pour Pc715j
Flagelle	Toutes les espèces	Le gène de la sous unité majeur de tubuline est polymorphe, nécessaire pour l'invasion cellulaire mais pas l'adhérence de <i>B.cenocepacia</i> .
Hémolysine	Identifié chez <i>B.cenocepacia</i>	Induit la dégranulation et la mort cellulaire de phagocytes humains. Activité hémolytique identifiée chez les espèces <i>B.cepacia</i> , <i>B.cenocepacia</i> , <i>B.stabilis</i> , <i>B.ambifaria</i> , <i>B.pyrricinia</i>
Invasion et survie intracellulaire	Variable au sein des espèces testées	<i>B.cenocepacia</i> est hautement invasive chez les cellules A549, <i>B.cenocepacia</i> , <i>B.multivorans</i> et <i>B.stabilis</i> pénètrent les cellules épithéliales respiratoires humaines. <i>B.dolosa</i> et <i>B.vietnamiensis</i> survivent dans des macrophages 5 jours.
LPS	Toutes les espèces	Structure du lipide A inhabituelle, le cœur à une structure KDO-KO inhabituelle
Mélanine	<i>B.cepacia</i> , <i>B.multivorans</i> , <i>B.cenocepacia</i>	Piège les radicaux libres <i>in vitro</i>
Quorum sensing	Toutes les espèces, étudiées systématiquement chez les 5 premières.	Toutes les espèces ont un système classique de type LuxIR qui permet la communication de cellule à cellule, la virulence, régule l'expression de toxines, protéases, lipase et sidérophores, module la mobilité et la formation de biofilm. <i>B.cenocepacia</i> possède un autre système sur un ilot de pathogénie.
Production de sidérophores	Toutes les espèces	Quatre types produits : acide salicylique, ornibactine, pyochelone et cepabactin. La synthèse et l'entrée de l'ornibactine sont essentielles pour la virulence de <i>B.cenocepacia</i> dans le modèle d'infection du rat.
Système de sécrétion de type III	Absent de <i>B.cepacia</i>	Les clusters de <i>B.multivorans</i> et <i>B.cenocepacia</i> ont la même organisation, nécessaire pour la virulence de <i>B.cenocepacia</i> dans le modèle d'infection murin.
Système de sécrétion de type IV et géli-faction	<i>B.cepacia</i> , <i>B.cenocepacia</i> , <i>B.vietnamiensis</i>	La macération de l'oignon a été observé en premier chez <i>B.cepacia</i> , sous la dépendance d'une pectine hydrolase plasmidique. Le phénotype de pourriture molle de <i>B.cenocepacia</i> est dépendant d'un système de sécrétion de type IV plasmidique. Le rôle du système porté par le chromosome 2 est inconnu.

### 5.1. Lipopolysaccharide

Les lipopolysaccharides sont des complexes macromoléculaires toxiques présents de manière constitutive dans la membrane externe de toutes les bactéries à Gram négatif. Le LPS est constitué d'un lipide A au niveau basal, et est relié à un antigène O de surface par l'intermédiaire d'un "core" constitué d'un enchaînement de différents sucres. Le LPS des Bcc a une structure originale qui le différencie des autres LPS de bactéries Gram-négatives (Silipo et al., 2007). Il joue un double rôle dans la pathogénie : il contribue à la résistance aux peptides antimicrobiens, et entraîne une réaction inflammatoire (Bamford et al., 2007). En effet, le lipide A est doué de propriétés endotoxiques qui agissent de manière massive lors de sa libération au cours de la lyse bactérienne. Les endotoxines exercent leurs effets de manière indirecte, en activant des molécules circulantes du sang ou en se liant sur des récepteurs cellulaires. Le LPS de *B. cenocepacia* K56-2 semble également jouer un rôle dans la survie intracellulaire de cette souche. Les gènes *hldA* et *hldD*, codant des enzymes impliquées dans la modification des sucres de l'oligosaccharide, semblent requis pour cette survie (Loutet et al., 2006). Les interactions du LPS ne sont cependant pas limitées aux interactions avec les cellules immunitaires de l'hôte ; le LPS de *B. cenocepacia* peut aussi induire la sécrétion d'interleukine 8 par les cellules épithéliales des voies respiratoires. Il a été démontré, de plus, que cette induction est plus importante chez *B. cenocepacia* que chez *P. aeruginosa*. L'antigène O, situé à la surface du LPS, est spécifique d'une espèce à une autre, et est largement utilisé pour le sérogroupage des souches bactériennes et pour le développement de vaccins. Cependant, la souche J2315 de l'espèce *B. cenocepacia* présente une insertion au niveau du cluster de gènes impliqués dans la synthèse de l'antigène O, conduisant à une absence de celui-ci et ainsi à une impossibilité de développer un vaccin contre cette souche (Ortega et al., 2005). Différents gènes impliqués dans la synthèse du LPS ont été décrits chez les Bcc ; le gène *wbiI*, par exemple, a été identifié comme marqueur génétique de la virulence. Son utilisation en RFLP permet une discrimination efficace des différents sérogroupes (Vinion-Dubiel et al., 2004). Le LPS semble également capable d'être modifié en fonction des conditions environnementales, permettant ainsi une adaptation des souches à un milieu donné. Dans ce contexte, Ierano et al. (2008) ont observé des changements majeurs au niveau de la région interne du lipide A chez une souche isolée pré et post-transplantation de poumons chez un patient CF. Ces variations structurales au niveau du LPS et les nombreuses applications thérapeutiques potentielles issues de ces observations justifient l'importance des études du LPS pour la compréhension de la virulence des Bcc.

## 5.2. Systèmes de sécrétion

Il existe à ce jour sept systèmes de sécrétion différents identifiés chez les procaryotes. Ces systèmes sont impliqués dans le transport passif ou actif de composés simples comme les acides aminés, les petits peptides, les ions, mais peuvent aussi permettre le passage de plus grosses molécules telles que des fragments d'ADN, des protéines, ou des composés chimiques divers. Leur rôle dans la virulence est déterminé par le type de composé secrété. Chez les Bcc, les systèmes de sécrétion de type III, IV et VI ont été décrits comme impliqués dans la virulence.

Le système de sécrétion de type III peut être comparé à une seringue moléculaire par laquelle une bactérie peut injecter des protéines dans les cellules eucaryotes juxtaposées. Ce système est impliqué dans la virulence de nombreuses espèces dont *Salmonella*, *Shigella*, *Yersinia*, *Vibrio*. Concernant les Bcc, Tomich et al., (2003) ont montré, chez la souris, une atténuation de la virulence de souches mutées dans le cluster de gènes du système de sécrétion de type III. Des études complémentaires ont démontré la présence de ce système chez toutes les espèces du Bcc, sauf *B. cepacia* (Tableau 2, Glendinning et al., 2004).

Le système de sécrétion de type IV présente des homologues avec la machinerie de conjugaison bactérienne et permet le transport d'ADN et/ou de protéines. Chez *Legionella pneumophila*, il permet le transfert de nombreuses protéines effectrices vers son hôte eucaryote, et chez *Agrobacterium tumefaciens*, il est impliqué dans le transfert d'ADN bactérien dans la plante (Cascales and Christie 2003, Christie et al., 2005). Il a été étudié chez la souche K56-2 de l'espèce *B. cenocepacia*. Ces travaux ont mis en évidence plusieurs fonctions de ce système. Il se révèle ainsi être responsable de la sécrétion de protéines cytotoxiques chez la plante conduisant à une altération de celle-ci, et il semble aussi contribuer à la survie et à la réplication intracellulaire de la souche K56-2 dans les cellules épithéliales et les macrophages (Engledow et al., 2004, Sajjan et al., 2008a). Un deuxième système de sécrétion de ce type a récemment été décrit par Zhang et al., 2009 ; celui-ci jouerait un rôle dans la mobilisation de plasmides.

## 5.3. Acquisition du Fer

Le fer est essentiel à la croissance bactérienne. Il est notamment utilisé dans la synthèse des cytochromes, et est un cofacteur des enzymes et des protéines transporteurs d'électrons. La disponibilité du fer est extrêmement variable en fonction des environnements.

Dans les poumons de patients CF, par exemple, le fer est sous forme chélatée au niveau des hématies. Les poumons constituent donc un environnement présentant peu de fer disponible, ce qui implique la nécessité d'adaptation des bactéries afin de survivre et se développer dans ce milieu à forte pression de sélection. Pour cela, une des stratégies développées est l'utilisation de sidérophores. Il s'agit d'éléments chimiques sécrétés par les micro-organismes et capables de chélater l'ion  $Fe^{3+}$ . Chez les Bcc, il existe au moins quatre types de sidérophores : ornibactine, pyocheline, cepabactine et cepaciacheline. De plus, les membres du Bcc sont capables d'utiliser certains sidérophores exogènes qu'ils ne sont pas en mesure de synthétiser. En outre, les Bcc possèdent des mécanismes supplémentaires aux sidérophores leur permettant d'acquérir du fer à partir d'hème ou de ferritine (Thomas 2007). La distribution de ces quatre sidérophores au sein du Bcc semble hétérogène. Dans le cas de la pyocheline, deux études cliniques ont démontré qu'environ 50% des isolats de Bcc testés (comprenant la souche *B. cenocepacia* K56-2) ne produisaient pas ou peu de pyocheline (Sokol 1986, Darling et al., 1998). D'autre part, le séquençage du génome de J2315 révèle une mutation dans le cadre de lecture du gène *pchF* conduisant à la production d'une pyocheline non fonctionnelle. Chez *B. vietnamiensis*, aucune souche n'a montré la capacité de produire de la pyocheline (Meyer et al., 1995). L'ornibactine, quant à elle, semble être le sidérophore le plus répandu parmi les *Burkholderia*. Cependant, une étude surprenante montre que l'ornibactine et ses dérivées ne semblent pas utilisables par *P. aeruginosa*. La cepabactine a été isolée pour la première fois chez la souche type de l'espèce *B. cepacia*. Ce composé présente toutes les caractéristiques d'un sidérophore ; cependant, seules quelques souches de *B. cenocepacia* semblent capables de produire ce composé indépendamment de la production d'autres types de sidérophore (Darling et al., 1998). Néanmoins, les souches ne produisant pas la cepabactine pourraient être capables de l'utiliser, leur conférant un réel avantage adaptatif. Les autres sidérophores décrits chez les Bcc semblent spécifiques et moins étudiés, comme c'est le cas pour la cepaciacheline, qui n'a été retrouvée que chez une souche de *B. ambifaria* ; ce sidérophore a été déduit à partir d'homologies de séquences avec les autres sidérophores. Il existe par ailleurs d'autres systèmes utilisés par les Bcc pour acquérir du fer exogène. *B. cenocepacia* J2315 peut ainsi utiliser l'hème oxydé ou la ferritine comme source de fer (Whitby et al., 2006). Le mécanisme permettant l'utilisation de la ferritine semble exiger une dégradation de celle-ci par des sérines protéases sécrétées dans le milieu. Les concentrations en ferritine étant particulièrement élevées dans les poumons des patients CF, cette capacité semble être un réel avantage pour *B. cenocepacia* lors d'infections de ce type de patient. De nombreux systèmes permettent la régulation de ces facteurs tel que des systèmes à deux



composantes, des facteurs transcriptionnels de type ECF mais aussi certains acteurs du quorum-sensing.

#### 5.4. Résistance aux antibiotiques

Une caractéristique importante des Bcc concerne leur résistance élevée à de nombreux antimicrobiens, tels que les antibiotiques ou les biocides (Govan et al., 2007). Les espèces du Bcc sont intrinsèquement résistantes à la polymyxine B et aux aminosides. Ceci implique la mise en place d'antibiothérapies basées sur des combinaisons de différentes molécules chez les patients CF. Nzula et al., (2002) ont testé la sensibilité aux antibiotiques de 65 souches représentant six espèces du Bcc ; ils ont ainsi démontré une hétérogénéité importante dans la résistance aux antibiotiques à l'exception de la résistance innée à la polymyxine B. D'autre part, ils ont pu mettre en évidence que l'ensemble des souches cliniques et environnementales évaluées présentaient une sensibilité similaire aux triméthoprime et ceftazidime. Ces molécules sont par conséquent des antibiotiques de choix contre les Bcc. De récents travaux ont testé la résistance des Bcc en fonction de leurs "mode de vie", état planctonique versus état de cellules en monocouches type biofilm. En théorie, les bactéries sous forme de biofilm devraient être moins sensibles aux antibiotiques, puisque cette structure conduit à une protection passive des cellules bactériennes. Cette étude a cependant démontré une activité bactériostatique et bactéricide similaire entre les souches du Bcc testées, qu'elles soient sous forme libre ou en biofilm (Peeters et al., 2008). Cependant, ces similitudes dans les profils de résistance ne reflètent pas rigoureusement la réalité. En effet, certains antibiotiques connus pour leurs actions *in vivo* ne montrent pas les mêmes performances *in vitro*. Ainsi, la mise en place et l'utilisation de techniques permettant de se rapprocher au maximum des conditions *in vivo* sont indispensables. Il semble néanmoins que la résistance aux antibiotiques est moins variable entre des cellules planctoniques ou en biofilm. Par ailleurs, les Bcc ont été décrits comme hautement résistants aux biocides, et certaines souches ont été impliquées dans de nombreux cas de contamination de solutions désinfectantes (Jimenez et al., 2000). Les bases moléculaires de cette résistance sont peu étudiées. Toutefois, les raisons de cette prévalence dans les solutions concernées semblent être justifiées par leur importante capacité de résistance, ainsi que par l'emploi par les industriels de composés désinfectants peu actifs ou présents à de trop faibles concentrations. Une étude sur ce sujet a récemment été réalisée sur un large panel de souches représentant l'ensemble des espèces du Bcc. Les concentrations minimales inhibitrices de composés biocides ont été testées sur ces souches, puis ces données

ont ensuite été corrélées aux résistances à différents antibiotiques (Rose et al., 2009). Ce travail a révélé que les souches de *B. cepacia*, *B. multivorans*, *B. cenocepacia* et *B. dolosa* présentaient le plus fort niveau de résistance aux composés antimicrobiens. De plus, au cours de ces expériences, certaines souches de *B. cenocepacia* (35%), *B. multivorans* (41%) et *B. vietnamiensis* (11%) ont montré l'apparition de résistances spontanées aux bêta-lactames, suggérant une évolution rapide des profils de résistance. Ces données offrent un éventail de profils de résistance des Bcc pour certains biocides et conduiront peut-être à la mise en place de mesures permettant la réduction des infections de ce type dans le futur.

Les processus impliqués dans cette résistance accrue sont multiples. Des systèmes tels que des pompes à efflux, la sécrétion de bêta-lactamases ou l'imperméabilité de la membrane externe des souches de Bcc semblent être impliqués dans la majeure partie des cas (Mahenthiralingam et al., 2005). Il apparaît primordial d'approfondir les connaissances relatives aux mécanismes de résistance en utilisant des modèles se rapprochant au maximum des conditions *in vivo*, mais aussi de déterminer les processus impliqués dans l'évolution et la transmission de ces résistances entre souches.

### 5.5. Autres facteurs

D'autres déterminants impliqués dans la virulence des souches du Bcc ont été décrits. A titre d'exemple, le BCESM (*Burkholderia cepacia* epidemic strain marker) est un îlot de pathogénie spécifique à certaines souches de *B. cenocepacia*. Il avait été initialement identifié par Mahenthiralingam et al., en 1997. Il possède d'une part des fonctions liées à la virulence – notamment un système de régulation du quorum-sensing de la bactérie – et d'autre part des fonctions relatives au métabolisme. Une caractérisation plus précise de cet îlot est présentée dans la section 6.3.3.

Outre cet îlot, certains gènes codant des métalloprotéases bactériennes peuvent également jouer un rôle dans la pathogénèse d'agents pathogènes pulmonaires tels que les Bcc (Corbett et al., 2003). La métalloprotéase ZmpA, par exemple, a été identifiée dans cinq espèces du Bcc dont *B. cenocepacia*, mais n'est pas présente chez d'autres espèces telles que *B. multivorans* ou *B. dolosa* (Gingues et al., 2005). Il existe d'autres métalloprotéases distribuées de façons hétérogènes dans le Bcc jouant des rôles divers dans la virulence. Elles permettent notamment la dégradation du collagène, de la fibronectine, et permettent de lutter contre certains membres du système immunitaire comme les neutrophiles (Kooi et al., 2006).

Ces protéases sont retrouvées chez beaucoup d'autres agents pathogènes comme *Vibrio cholerae* ou *P. aeruginosa*.

Il existe de multiples études démontrant la fonction de virulence de composés extrêmement variés comme les lipases, les exopolysaccharides, les adhésines etc. Cependant, chez *Burkholderia*, la plupart des gènes impliqués dans la virulence paraît être régulée par *quorum-sensing* (QS). Ce système inductible en fonction de la densité bactérienne a été identifié chez la plupart des espèces du Bcc, et semble mis en place tant en conditions infectieuses que lors du développement saprophyte ou symbiotique. La régulation par QS engendre des phénotypes variés plus ou moins liés à la virulence des souches. A titre d'exemple, la macération de l'oignon, la production de protéases, la mobilité, la formation de biofilms, la production de fongicides, sont tous des phénomènes régulés par QS (Eberl 2006). Par ailleurs, la plupart des études décrivant le ou les rôles du QS ont été réalisées sur l'espèce *B. cenocepacia*, et ceci en utilisant les différents modèles animaux ou végétaux décrits dans la section suivante. Une étude particulièrement intéressante visant à tester différents gènes de virulence chez *B. cenocepacia* a été publiée par Uehlinger et al., en 2009. Lors de ce projet, ils ont inactivé un large panel de gènes décrits comme impliqués dans la virulence, puis ont testé les mutants obtenus sur la plupart des modèles utilisés pour les Bcc (incluant des mammifères, des nématodes et des plantes). Par le biais de ce travail original, les auteurs ont démontré que la plupart des déterminants de virulence connus étaient spécifiques à l'un ou l'autre des modèles d'infection, et que seulement trois facteurs semblaient essentiels pour la pathogénie chez les différents modèles: (i) le quorum-sensing, (ii) la production de sidérophores, et (iii) la biosynthèse du LPS. L'inactivation de chacun de ces facteurs conduit à une atténuation de la virulence au niveau d'au moins trois des modèles d'infection testés. Ces travaux ont montré le rôle central du QS dans la virulence des souches du Bcc et ont également confirmé le rôle de deux autres déterminants majeurs de la virulence. Ces études ouvrent évidemment des perspectives intéressantes vers le développement de nouveaux anti-infectieux ciblant le QS.

## 5.6. Modèles d'étude

Différents modèles d'hôtes alternatifs ont été mis au point pour identifier et caractériser les processus de l'infection. Chaque modèle a son importance puisque chacun n'évalue qu'une partie des processus mis en œuvre dans une infection humaine. Ils ont ainsi permis d'identifier des facteurs de virulence divers. Les modèles *in vivo*, utilisant des souris ou des rats, permettent à la fois d'étudier les infections aiguës et chroniques. Dans le premier cas, les bactéries sont injectées dans le péritoine et retrouvées dans les poumons, le foie et la rate. Elles sont rapidement éliminées des poumons et du foie, mais peuvent persister plus longtemps dans la rate. Des souris CftR déficientes ou neutropéniques peuvent être infectées par aérosol ; cependant, par ce biais, les bactéries sont moins persistantes.

Pour les infections chroniques, les bactéries sont enrobées dans des billes d'agar et insérées dans la trachée de souris ou de rats. Ainsi, les bactéries peuvent être suivies dans les poumons ou la rate jusqu'à 21 jours. Alternativement, les souris peuvent être immunodéprimées avec de la cyclophosphamide et infectées par voie nasale. Dans ce cas, les bactéries peuvent être observées pendant 16 jours à l'exception de *B. cenocepacia* qui, paradoxalement, est éliminée plus rapidement.

Ces modèles étant lourds et coûteux à mettre en place, des modèles *in vitro* ont également été développés et validés. Différentes lignées cellulaires humaines d'individus CF ou non, ainsi que des lignées murines ont été utilisées pour évaluer les capacités des *Burkholderia* pour l'attachement, l'invasion, et la résistance aux antibiotiques.

D'autres modèles alternatifs ont été utilisés, comme l'inoculation de pelures d'oignons avec suivi des concentrations bactériennes et du phénotype de pourriture molle. La luzerne a également été testée pour évaluer la pathogénie de souches de *Burkholderia*. Les résultats obtenus ont été congruents avec ceux du modèle murin. Enfin, un dernier modèle d'interactions avec des nématodes a été développé. Le nématode *Caenorhabditis elegans* ne possède pas d'immunité adaptative, et est par voie de fait un bon modèle pour étudier la réponse immunitaire innée à des bactéries pathogènes (Aballay and Ausubel 2002). Initialement utilisé pour étudier *P. aeruginosa*, le processus d'infection a été décrit par O'Quinn et al. (2001) pour les *B. pseudomallei*. Kothe et al. (2003) ont pu mettre en évidence que la souche H111 de *B. cenocepacia* tuait ce nématode de deux façons différentes en fonction du milieu de culture utilisé. A forte osmolarité, les nématodes étaient rapidement détruits (quelques heures) mais, étonnamment, aucune bactérie n'était retrouvée dans le tube digestif du ver. Une toxine était en fait impliquée dans la mort du nématode. Sur milieu

standard de croissance pour nématode, la bactérie est plutôt ingérée par le nématode et est retrouvée dans l'appareil digestif en grande quantité. Le ver meurt alors en deux ou trois jours. Par l'utilisation de ce modèle, la nécessité du système de quorum sensing CepI/CepR dans la virulence a été confirmée.

Cette étude a été complétée par Cardona *et al.* (2005) pour évaluer la pathogénie des différentes espèces du Bcc sur *C. elegans*. Il est apparu que certaines souches pouvaient tuer les nématodes tandis que d'autres avaient des effets négligeables voir nuls. De plus, les souches du clone épidémique ET12 ont présenté différents phénotypes sur ce modèle. En outre, il n'y a pas eu de corrélation évidente entre l'origine (environnementale ou clinique) des souches et leur capacité à tuer *C. elegans*.

## 6. Adaptation aux contraintes environnementales

### 6.1. Acclimatation

L'acclimatation est le fait, pour un organisme vivant, de s'adapter à un changement durable de son environnement, notamment climatique (température, humidité, ressources, etc.). Elle se déroule sur une période relativement courte – au plus égale à la durée de vie de l'organisme – et correspond à des changements phénotypiques réversibles, ce qui la différencie de l'adaptation. Au sein d'une même espèce, de nombreux caractères phénotypiques sont soumis à variation. Cette diversité phénotypique va conduire à un changement dans le profil d'expression des gènes et ainsi engendrer une réponse correspondant au stimulus reçu. De manière globale, l'acclimatation est associée à des modifications transcriptionnelles et traductionnelles, tandis que l'adaptation concerne les changements relatifs au contenu génomique. Différents mécanismes de régulation existent chez les bactéries ; ils impliquent des gènes dits "régulateurs", dont les principales familles sont les systèmes à deux composantes (SDC), le quorum-sensing (QS) et les facteurs transcriptionnels de la famille des sigma70 ( $\sigma_{70}$ ).

La plupart des études réalisées au niveau du Bcc portent sur la régulation par quorum-sensing. Ceci peut être expliqué par les implications de ce système de régulation dans la virulence des souches (e. g. section 5.5.). Il existe cependant quelques études portant sur la régulation de l'expression de gènes par des systèmes à deux ou plusieurs composantes. Par exemple, une étude visant à caractériser le rôle du gène *cblS* chez *B. cenocepacia* a permis de mettre en évidence une régulation complexe de la synthèse du pilus câble (Tomich and Mohr 2004). Ainsi, deux senseurs distincts ont pu être identifiés pour un même système de

régulation à deux composantes. La plupart du temps, dans ces systèmes, il n'existe qu'un seul "senseur" ou capteur. Quelques rares exceptions ont pu être identifiées, comme par exemple chez *Bacillus subtilis*, où cinq senseurs sont présents pour un unique régulateur. Par ailleurs, le séquençage du génome de la souche *B. cenocepacia* J2315 a permis d'identifier des gènes ayant des homologies avec des SDC d'autres souches (Holden et al., 2009). Cependant, ce type de régulation est très fréquemment utilisé par les bactéries, et ce de manière très étendue, ce qui rend difficile la prédiction de leurs rôles chez J2315.

D'autres régulateurs tels que les facteurs  $\sigma 70$  peuvent intervenir dans les réponses à différentes perturbations de l'environnement. Les facteurs  $\sigma 70$  permettent la reconnaissance spécifique des différents types de promoteurs au sein d'un génome, leur activation ainsi que la transcription des gènes en aval. Ces facteurs sont impliqués dans différents phénomènes d'acclimatation suivant leurs classes : phase exponentielle de croissance, réponse au stress de température, carence en nutriments, variation de pH, entrée en phase stationnaire, etc. La majorité d'entre eux est impliquée dans des fonctions extracytoplasmiques (ECF). Chez les Bcc, peu d'études spécifiques ont été réalisées sur les  $\sigma 70$  ; cependant l'équipe du Dr. Valvano a mis en évidence en 2008 un facteur  $\sigma 70$  impliqué dans la formation de biofilm et permettant de retarder la fusion phagolysomales dans le macrophage (Saldias et al., 2008). Ce retard permettrait de prolonger la survie des bactéries intracellulaires dans les macrophages et pourrait jouer un rôle dans la virulence de la souche K56-2 de *B. cenocepacia*. Ce travail montre une des multiples implications des  $\sigma 70$  dans la réponse aux stress. D'autre part, un travail particulièrement intéressant est paru en début d'année dans la revue PNAS. Les auteurs ont examiné la réponse globale de deux souches de *B. cenocepacia* (AU1054 et HI2424) cultivées dans deux environnements différents (sol et crachat de patient CF) par séquençage haut débit (Yoder-Himes et al., 2009). Cette évaluation du transcriptome à large échelle ainsi que la multitude de données obtenues ont permis, entre autres, de démontrer non pas une activation de gènes différents selon les conditions, mais un changement global de la régulation de ces gènes. De plus, les auteurs ont identifié une série de petits ARNs non codant (ncRNA) présents uniquement lors de la croissance en condition sol. Depuis quelques années, la communauté scientifique explore les fonctions potentielles de ces ncRNA. Leur rôle effectif dans la régulation de l'expression génétique est désormais admis. (Vecerek et al., 2007, Waters and Storz 2009). L'ensemble de ces données apporte un éclairage conséquent sur l'importance et la complexité de tous ces mécanismes de régulation pour la survie et le développement des bactéries dans un environnement donné. De plus, tous ces systèmes peuvent interagir les uns avec les autres, aboutissant à des cascades de régulations

extrêmement complexes et difficilement identifiable. Il paraît donc primordial de déterminer le rôle de chacun de ces régulateurs, que ce soit lors de la virulence ou de la symbiose, afin de mieux comprendre l'écologie des bactéries, et ce qu'elles soient pathogènes ou non.

## 6.2. Génération de diversité

La génération de diversité a été classiquement perçue comme un phénomène progressif d'accumulation de mutations ponctuelles conduisant à l'émergence de nouvelles propriétés et à l'isolement sexuel des espèces. Chez les bactéries, ce processus est largement complété par d'autres mécanismes, comme les réarrangements de grandes portions des génomes, la fusion de réplicons, etc. Ces différents mécanismes ont permis une diversification rapide des bactéries, rendant possible une réponse rapide aux pressions environnementales et la colonisation d'une grande variété de niches écologiques. De plus, l'absence de phase diploïde dans le cycle de développement des bactéries favorise la fixation des modifications génétiques. En conséquence, l'évolution des génomes bactériens doit être abordée à différentes échelles allant de la modification locale d'un gène à l'organisation du (ou des) chromosome(s).

### 6.2.1. Hypermutableté

L'hypermutableté chez les bactéries concerne des souches mutantes altérées dans le système de réparation des mésappariements (*mutHLS*). Une augmentation significative du taux de mutations est alors observée ( $10^{-6}$  voire  $10^{-5}$  mutation par nucléotide incorporé pour les souches hypermutables, contre  $10^{-8}$  chez les souches normales). Ce phénomène génère de la diversité génétique pouvant conduire à un changement phénotypique. Dans le cas des infections pulmonaires chroniques de patients atteints de mucoviscidose, des variations significatives dans les phénotypes de *P. aeruginosa* ont été observées et ont été reliées à des phénomènes d'hypermutableté (Mena et al., 2008). Ces mutations peuvent aboutir à un changement de phénotype. Ainsi, certaines souches de *P. aeruginosa* deviennent mucoïde, perdent et/ou changent de mobilité et présentent une résistance accrue aux antibiotiques. Bien qu'aucun mutant hypermutable n'ait été identifié chez les Bcc, il apparaît plausible que ce phénomène puisse exister, notamment lors de la phase chronique des infections de patient CF.

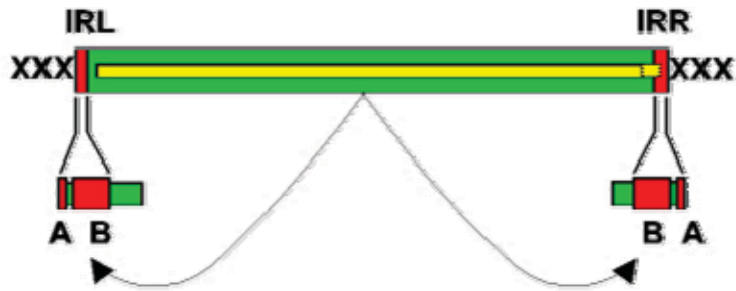
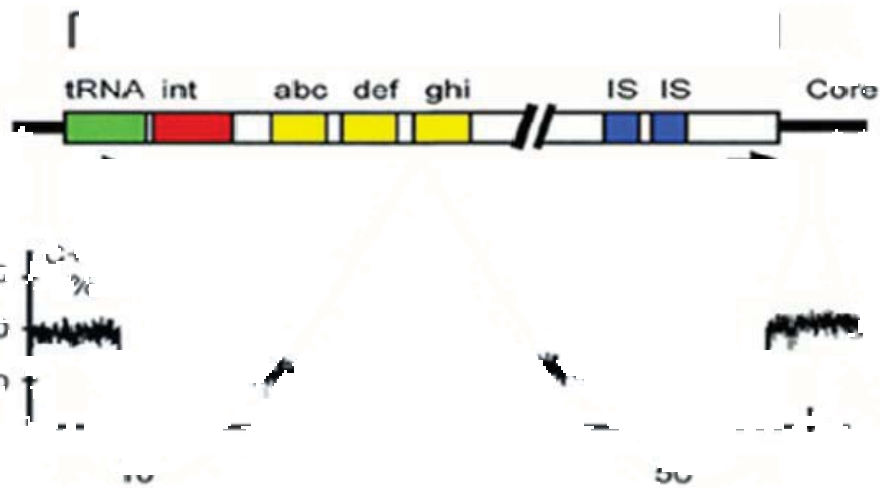


Fig. 7 : Schéma d'une séquence d'insertion. IRR = régions inversées répétées droite, IRL = région inversée répétée gauche, XXX = séquences dupliquées lors de l'intégration de l'IS.



EF180 reports of 2 no 5 pp376-381 2001

Fig. 8 : schéma d'un îlot génomique.  
IS = séquence d'insertion, abc = gènes présents sur l'IG, DR = séquences répétées.



### 6.2.2. Séquences d'insertion

Les séquences d'insertion (IS) sont présentes dans la plupart des génomes bactériens et sont génétiquement compactes (Fig. 7). Elles ne codent généralement pas de fonction autre que celle impliquée dans leur mobilité (Mahillon and Chandler 1998). Les IS peuvent s'intégrer au niveau d'une séquence codante (CDS) conduisant à l'inactivation du gène. De plus, les IS sont impliquées dans des recombinaisons homologues conduisant à des réarrangements au sein des génomes bactériens (Kumar et al., 1997). Il est intéressant de souligner que certaines IS présentent des séquences promotrices pouvant modifier l'expression des gènes en aval (Hubner and Hendrickson 1997, Woods et al., 2004). Chez les Bcc, différents travaux ont permis de déterminer que le nombre d'IS varie en fonction des souches et des espèces. Certaines d'entre elles ont été plus particulièrement étudiées en raison des changements de phénotype qu'elles pouvaient engendrer. DeShazer et al., (2001) ont montré que l'insertion d'une IS407 bloquait la synthèse de la capsule chez *B. mallei* et provoquait une réduction du pouvoir pathogène. Holden et al., 2009 ont suggéré que la présence de nombreux pseudogènes (notamment due à des insertions d'IS) pourrait jouer un rôle dans l'adaptation d'une souche à un biotope particulier. D'autres études ont démontrées que les IS pouvaient être des marqueurs de certains clones épidémiques ; ainsi, l'IS1363 n'a été détectée que chez les clones appartenant à la ligné PHDC (Liu et al., 2003). D'autre part, les génomes du Bcc sont composés de plusieurs chromosomes, et la plupart d'entre eux sont riche en IS. Ces deux caractéristiques semblent jouer un rôle clé dans l'évolution de nouvelles fonctions de dégradation et dans la versatilité des bactéries du Bcc (Lessie et al., 1996). Les IS semblent ainsi contribuer de façon non négligeable à l'augmentation de la diversité et à l'évolution des phénotypes. Ces différents points seront plus amplement discutés dans le chapitre 3 de cette thèse qui porte sur l'étude des séquences d'insertion.

### 6.2.3. Réarrangements génétiques

Les évènements de réarrangement du génome, comprenant l'acquisition ou la perte de fragments d'ADN, les inversions et les translocations, sont des phénomènes fréquemment observés chez les bactéries. Cependant, l'impact de tels évènements sur la diversité fonctionnelle n'est encore pas clairement définie à ce jour.

### 6.3. Acquisition d'informations génétiques

Chez la plupart des eucaryotes, la reproduction sexuée permet le maintien d'une diversité génétique au sein des populations en assurant le brassage génétique. De leur côté, les bactéries ont développé d'autres mécanismes leur permettant d'acquérir de l'information génétique. Il existe 3 mécanismes de transfert de gènes entre les bactéries : la transformation, la transduction, et la conjugaison. Au cours de la transformation, un ADN bicaténaire, libre, nu et en solution est introduit dans une bactérie réceptrice ; lors d'une transduction, le transfert d'ADN est réalisé par l'intermédiaire d'un phage. La conjugaison, quant à elle, nécessite un contact des cellules donneuses et receveuses afin d'effectuer le transfert par l'intermédiaire d'une structure spécialisée, le pili sexuel. L'ADN étranger peut alors être intégré dans le génome par recombinaison ou rester dans la cellule sous forme de plasmide ou bien être dégradé.

#### 6.3.1. Phages

Chez les *Burkholderia*, le potentiel évolutif généré par les phages a été peu pris en compte dans la littérature. Langley *et al.* (2003) ont fait un bilan des phages présents dans des souches du Bcc et évalué leurs capacités à infecter les différentes espèces du Bcc. Ils montrent que les phages retrouvés ne sont en aucun cas spécifiques d'une espèce. Ainsi, ils peuvent contribuer aux transferts horizontaux de gènes au sein du Bcc ou même, pour certains, vers *B. pseudomallei* ou *P. aeruginosa*. Nzula *et al.* (2000) ont montré qu'un phage permet le transfert de gènes de résistance aux antibiotiques entre souches de *B. vietnamiensis*. Par ailleurs, Hens *et al.* (2006) ont mis en évidence un phage de *B. cepacia* capable de transférer des multirésistances à *Shigella flexneri*. D'autre part, le phage BcepMu semble largement distribué au sein de la lignée ET12 de l'espèce *B. cenocepacia* et pourrait jouer un rôle dans la pathogénie de *B. cenocepacia* (Summer *et al.*, 2004). Lors de coinfections avec d'autres espèces bactériennes (notamment *P. aeruginosa*) au niveau des poumons des patients atteints de mucoviscidose, les échanges génétiques par le biais de phages pourraient être facilités. Par ailleurs, de nombreux phages ont été retrouvés dans des sols colonisés par des *Burkholderia* ; cet environnement semblerait ainsi jouer le rôle de réservoir potentiel de phénomène de transduction. De plus, certains phages paraissent spécifiques d'espèces voire de souches du Bcc, et pourraient représenter une alternative thérapeutique pour lutter contre les infections à Bcc. Dans ce sens, Seed and Dennis (2009) ont testé cette alternative sur modèle murin. Sous

certaines conditions, la thérapie par phage a montré des résultats positifs sur la souche *B. cenocepacia* K56-2 ; cette alternative à l'antibiothérapie semble donc prometteuse.

### 6.3.2. Plasmides

Le transfert d'ADN entre bactéries peut également se faire par le biais de plasmides. Dans de nombreux cas, les caractères transférés concernent la résistance à des antibiotiques, ou l'acquisition de voies métaboliques particulières. Peu d'études sont disponibles quant à la présence de plasmides chez les *Burkholderia*. Lennon et Decicco (1991) rapportent avoir identifié au moins un plasmide dans 84% des 37 souches de *Burkholderia* qu'ils ont étudié. La taille des plasmides identifiés variait de 2,7 à 222 kb. Le transfert conjugatif d'un de ces plasmides a été montré entre *B. cepacia* et *P. aeruginosa*. D'autre part, Engledow et al., (2004) ont identifié des gènes codant pour un système de sécrétion de type IV au niveau d'un méga-plasmide de la souche *B. cenocepacia* K56-2. Ce plasmide est impliqué dans l'infection de l'oignon, cependant son origine et les mécanismes impliqués dans son acquisition ne sont pas connus.

### 6.3.3. Ilots génomiques

Les îlots génomiques (IG) sont des zones du chromosome bactérien possédant des caractéristiques différentes du reste du génome ; le pourcentage en bases G et C peut notamment être différent (Figure 8). Ces îlots sont acquis par une souche lors d'une transformation, d'une transduction ou d'une conjugaison puis sont intégrés de manière plus ou moins stable dans le génome bactérien. Certains îlots génomiques ne sont associés à aucun phénotype connu, mais la plupart d'entre eux codent pour des informations biochimiques ou des facteurs de pathogénie (îlots de pathogénie). Au sein du genre *Burkholderia*, plus d'une cinquantaine de génomes ont été séquencés ; ceci permet l'analyse en génomique comparative de ces replicons. En utilisant cette approche, de nombreux fragments d'ADN spécifiques de souches, d'espèces ou de complexes d'espèces peuvent être identifiés et reliés à des îlots génomiques en fonction de leurs caractéristiques. La fonction écologique de ces différents îlots est souvent difficile à définir. Au sein du Bcc, un des premiers îlots décrits fut le BCESM. Sa caractérisation a permis de démontrer son rôle dans la virulence de souches de *B. cenocepacia* (Baldwin et al., 2004). Depuis, de nombreux autres IG ont été identifiés chez les *Burkholderia*, notamment chez *B. pseudomallei* et *B. mallei* (Tuanyok et al., 2008). Ces phénomènes de transfert d'ADN peuvent jouer un rôle important dans l'émergence de

souches hypervirulentes (Croucher et al., 2009). Une description plus précise de la distribution et des rôles potentiels des IG chez *B. cenocepacia* est présentée dans le chapitre 4 de ce manuscrit de thèse.

## 7. Conclusion

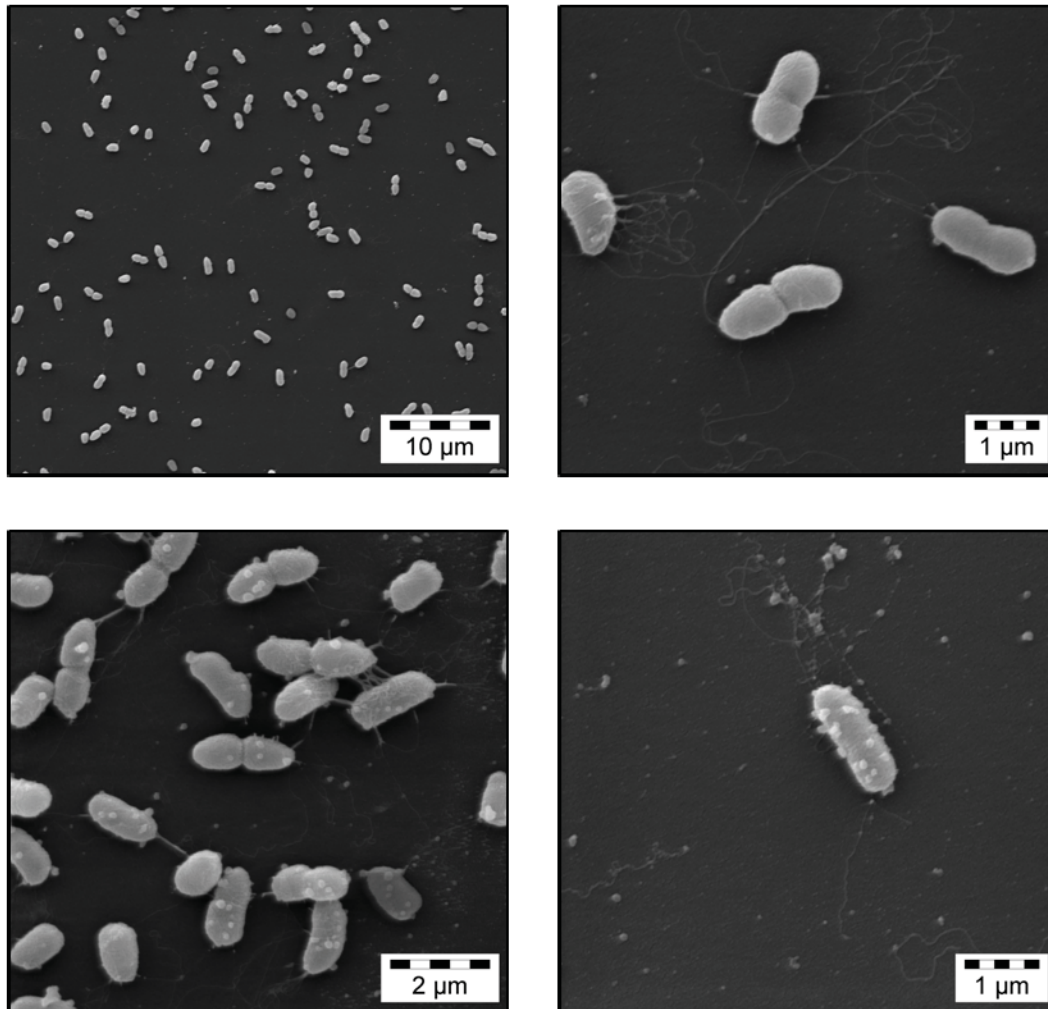
Les *Burkholderia* présentent une importante diversité tant au niveau génomique que phénotypique. Au sein de ces espèces, des souches issues de milieux très différents et possédant des capacités de virulence hétérogènes sont isolées. Faire le lien entre cette diversité phénotypique et génétique est actuellement le sujet de nombreuses études. Un des buts importants de ces études est d'identifier et de comprendre les caractères déterminants des interactions entre ces bactéries et leur environnement. Ces études peuvent aujourd'hui bénéficier des nombreux génomes séquencés de *Burkholderia*. Ces données sont indispensables pour la compréhension du complexe *cepacia* qui montre une variabilité génomique très étendue.

Le travail de thèse présenté dans ce document s'est focalisé sur la caractérisation et l'étude d'un clone épidémique de l'espèce *B. cenocepacia*, le clone B&B. Quatre chapitres ont été développés. Le premier chapitre décrit l'origine du clone B&B et sa classification au sein du Bcc. Le deuxième chapitre décrit la structure génétique des facteurs transcriptionnels de la famille sigma70 au sein du Bcc et du clone B&B ; cette approche permet d'étudier une partie de la plasticité des processus de régulation globaux des génomes et notamment le rôle des facteurs extracytoplasmiques indispensables à la virulence bactérienne. Les deux derniers chapitres sont dédiés aux processus génétiques impliqués dans les réarrangements de génome, en se focalisant sur les séquences d'insertions (chapitre 3) et les îlots génomiques (chapitre 4). L'implication de ces réarrangements dans l'émergence de clones épidémiques chez l'espèce *B. cenocepacia* est discutée.

## Caractérisation Moléculaire et distribution environnementale du clone B&B de l'espèce *Burkholderia cenocepacia*

Les *Burkholderia* du complexe *cepacia* sont des bactéries pathogènes opportunistes principalement retrouvées dans des infections pulmonaires de patients atteints de mucoviscidose (patients CF) et associées à d'autres infections comme les otites et kératites (Hobson, et al. 1995, Mahenthiralingam, et al. 1996, Mahenthiralingam, et al. 2001, Vandamme, et al. 2002, Matoba 2003, Mahenthiralingam, et al. 2008). Trois clones épidémiques majeurs de l'espèce *B. cenocepacia* ont été décrits chez les patients CF. Le clone ET12 est principalement retrouvé en Europe de l'Ouest et en Amérique du Nord, il provoque le "syndrome cepacia" qui consiste en une pneumonie nécrotique foudroyante qui est à l'origine de nombreux décès chez les patients CF (Isles, et al. 1984). Deux autres clones ont également été décrits chez les patients CF, les clones PHDC et MidWest qui sont principalement retrouvés en Amérique du Nord (Chen, et al. 2001, LiPuma, et al. 2002, Vandamme, et al. 2003). Il existe par ailleurs des cas documentés d'infections nosocomiales liées aux *Burkholderia* du complexe *cepacia*. La durée des épidémies nosocomiales est très variable et l'origine des souches est le plus souvent liée à un matériel ou une solution contaminée (Nasser, et al. 2004, Ghazal, et al. 2006, Lo Cascio, et al. 2006, Alvarez-Lerma, et al. 2008). Des clones isolés de patients CF ont été retrouvés dans ces infections (Holmes, et al. 1999). En 2004, le service de soins intensifs de l'hôpital Fleyriat à Bourg-en-Bresse a été confronté à une épidémie impliquant des *Burkholderia* du complexe *cepacia*. Les différents travaux réalisés au sein notre équipe ont permis de caractériser le clone responsable de cette épidémie. Ce clone a été nommé le clone B&B pour son origine (Bourg-en-Bresse), une photographie en microscopie électronique est présentée dans la Figure 1. La virulence et la forte transmissibilité de ce clone nous ont amené à étudier son origine, sa proximité génétique avec les clones virulents du Bcc, et à rechercher les éléments et innovations génétiques ayant favorisé son émergence. Ce clone a été dernièrement observé comme étant l'agent principal des infections à *B. cenocepacia* chez les patients CF en France métropolitaine (Abalain, et al. 2009). Les travaux concernant la caractérisation de ce clone ont fait l'objet d'une publication dans le journal Diagnostic Microbiology and Infectious Disease en Aout 2009, cet article est présenté dans la section ci-dessous. En complément de cette publication, une description des différentes campagnes d'échantillonnages réalisées à proximité et au sein de l'Hôpital Fleyriat, est présentée dans la partie annexe de ce chapitre. Les démarches expérimentales

utilisées ainsi que les résultats obtenus sont également présentées dans la partie annexe. Les autres chapitres de cette thèse sont dédiés aux phénomènes génétiques ayant pu contribuer à l'émergence de ce clone.



**Fig. 1:** Photos du clone B&B de *B. cenocepacia* en microscopie électronique à balayage. (S. Favre-Bonté, 2008).



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

ScienceDirect

Diagnostic Microbiology and Infectious Disease 66 (2010) 29–40

DIAGNOSTIC  
MICROBIOLOGY  
AND INFECTIOUS  
DISEASE

[www.elsevier.com/locate/diagmicrobio](http://www.elsevier.com/locate/diagmicrobio)

## Epidemiology and molecular characterization of a clone of *Burkholderia cenocepacia* responsible for nosocomial pulmonary tract infections in a French intensive care unit

Arnault Graindorge<sup>a,b,c</sup>, Aymeric Menard<sup>a,b,c</sup>, Manuelle Neto<sup>a,b,c</sup>, Claude Bouvet<sup>d</sup>, Roger Miollan<sup>d</sup>, Sandrine Gaillard<sup>d</sup>, Henri de Montclos<sup>d</sup>, Frédéric Laurent<sup>a,b,c,d,1</sup>, Benoit Cournoyer<sup>a,b,c,\*</sup>

<sup>a</sup>Université de Lyon, Caserne Sergent Blandan, 37 Rue du Repos, 69361 Lyon, France

<sup>b</sup>Research group on «Bacterial Opportunistic Pathogens and Environment», UMR5557 Ecologie Microbienne, Université Lyon 1, CNRS, 69 622 Villeurbanne cedex, France

<sup>c</sup>Ecole Nationale Vétérinaire de Lyon, 1 Avenue Bourgelat, 69280 Marcy l'Etoile, France

<sup>d</sup>Centre Hospitalier de Bourg-en-Bresse, 900 Route de Paris, 01012 Bourg-en-Bresse, France

Received 3 March 2009; accepted 8 June 2009

### Abstract

Clustered cases of nosocomial pulmonary infections were observed in a French intensive care unit. Biochemical tests showed the etiologic agents to be part of the Bcc (Bcc). *recA* polymerase chain reaction-restriction fragment length polymorphism analysis and molecular phylogeny positioned the isolates into *Burkholderia cenocepacia*. Their *recA* sequences were found identical to those of ET12 strains responsible of necrotic pneumonia in cystic fibrosis patients. Analyses of a multi locus sequence typing genes set confirmed this proximity and suggested a wide distribution among occidental countries but could not resolve their phylogenetic position unambiguously. A novel marker, *ecfB*, indicated a significant phylogenetic divergence from ET12 strains. Pulse field gel electrophoresis analysis of *SpeI*-restricted total genomic DNA of the strains showed a unique profile indicative of a clonal outbreak. Environmental hospital screenings indicated cross-contamination between staff and patients. Bcc strains from outdoor environments were not related to this clone but indicated the presence of *Burkholderia multivorans* and *Burkholderia vietnamiensis*.

© 2010 Elsevier Inc. All rights reserved.

**Keywords:** Nosocomial outbreak; *Burkholderia cenocepacia*; *B. cepacia*; *B. vietnamiensis*; *B. multivorans*; Pneumonia; Cystic fibrosis

### 1. Introduction

*Burkholderia* is a bacterial genus grouping about 60 species of  $\beta$ -proteobacteria that can be found in a wide range of habitats (Coenye and Vandamme, 2003; Compant et al., 2008). It includes a taxonomic entity named the *Burkholderia cepacia* complex (Bcc). This complex groups the *B. cepacia* species that is involved in onion diseases and 16 other species isolated from outdoor or clinical samples.

From the 1980s, strains of these species were found as the etiologic agents of human infections (Hobson et al., 1995), and an increased incidence of cepacia infections was reported among cystic fibrosis (CF) patients (Isles et al., 1984). One of the most dramatic outcome was related to a highly transmissible clone of *B. cenocepacia* named ET12 (Johnson et al., 1994), which spread among CF patients of Canada, England, and Scotland, and led to the so-called cepacia syndrome, a necrotizing pulmonary infection with a 62% fatality rate (Isles et al., 1984). Strains of the ET12 clonal complex were found to have unusual properties like a cable pilus encoded by the *cblA* gene (Sajjan et al., 1995; Sun et al., 1995). Other *B. cenocepacia* epidemic clones among CF cohorts have been reported and showed properties and genomes with significant differences from the ones of the ET12 clone (Chen et al., 2001; LiPuma et al., 1988).

\* Corresponding author. Université Lyon 1, UMR5557 Ecologie Microbienne, 69 622 Villeurbanne cedex, France. Tel.: +33-472431495.

E-mail address: [cournoye@biomserv.univ-lyon1.fr](mailto:cournoye@biomserv.univ-lyon1.fr) (B. Cournoyer).

<sup>1</sup> Present address: Laboratoire de bactériologie, Inserm U851, Hôpital de la Croix Rousse, Hospices Civils de Lyon, France.

More than 20 papers have described nosocomial infections involving the Bcc since the 1990s. These outbreaks have affected up to several hundreds of patients and have lasted a few months to several years. In France, the prevalence of Bcc species among the overall number of infectious agents involved in nosocomial infections was estimated to be of 0.1% (Savey and Tressière, 2007). A survey of the literature and various databases showed the Bcc species to be one of the most frequent pathogens infecting patients through contaminated medical solutions like disinfectants (Ghazal et al., 2006; Lo Cascio et al., 2006; Oie and Kamiya, 1996b), heparin vials (Vonberg and Gastmeier, 2007; Yan et al., 2008), and contaminated material like nebulizers (Nasser et al., 2004; Oie and Kamiya, 1996a). Bcc have also been introduced in medical units through moisturizing body milks and alcohol-free mouthwash solutions used in the hygienic care of hospitalized patients (Alvarez-Lerma et al., 2008; Matrician et al., 2000). Most reported nosocomial Bcc outbreaks have occurred among hospital intensive care units (ICUs) including neonatal ones, and hemodialysis units. These outbreaks were mainly due to a single clone, but polyclonal infections have been reported in a few instances (Magalhaes et al., 2003; Yan et al., 2008). In most cases, phenotypic analyses using Api20NE strips (BioMerieux, Marcy-l'Étoile, France), the VITEK2 (BioMerieux), or the MicroScan system (Dade Diagnostic MicroScan Division, West Sacramento, CA) were performed to identify the etiologic agents involved in these outbreaks. These methods allowed identification at the Bcc level but could not allocate the isolates to a particular species. In a few investigations, a *recA* polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) identification scheme (Mahenthalingam et al., 2000a) was applied and suggested *B. cenocepacia* (Shehabi et al., 2004; Siddiqui et al., 2001; Woods et al., 2004) and *B. stabilis* (Jacobson et al., 2006; Otag et al., 2005) to be the most prevalent species. Other species reported to have been involved in some of these outbreaks were *B. cepacia* (Jacobson et al., 2006; Kotsanas et al., 2008; Souza et al., 2004), *Burkholderia multivorans* (Woods et al., 2004) and *Burkholderia vietnamiensis* (Magalhaes et al., 2003; Woods et al., 2004). Rarely, the strains involved in these outbreaks were compared to CF isolates, but a large hospital outbreak reported in 1999 showed that both patient groups can be colonized by a same clone (Holmes et al., 1999). In such outbreaks, acquisition risk in the CF patients was associated with significantly increased mortality rates, suggesting that the most virulent CF epidemic clones could have emerged from the hospital environment.

Here, a Bcc nosocomial outbreak in a French ICU is described. This outbreak consisted of clustered cases of respiratory tract infections among mechanically ventilated patients. The taxonomic status of the strains involved in this outbreak was investigated and revealed a close phylogenetic proximity with the ET12 epidemic clone. However, the

*recA* and Bcc multi locus sequence typing (MLST) scheme genetic markers could not unambiguously resolve the phylogenetic relationships between these strains and the ET12 lineage. These investigations thus required the development of a novel genetic marker. The *ecfB* gene was selected and used to infer phylogenetic relationships among *B. cenocepacia* and a subgroup named IIIA according to *recA* PCR-RFLP profiles. The *ecfB* marker showed the outbreak strains to be significantly different from the ET12 ones. Nevertheless, all these data pointed toward a close proximity with the ET12 strains and a likely recent common ancestor.

## 2. Materials and methods

### 2.1. Sampling of the ICU, patients, and outdoor environments

The ICU concerned by the nosocomial outbreak reported in this study consists of 10 rooms (Fig. 1), each having a general aeration system, a sink, a bed, and a mechanical respiratory ventilator. During 2004, 504 patients have been admitted in this ICU, and 36 nosocomial infections, by any microorganisms, were recorded. The outbreak described in this paper is the only one to have led to the death of a patient at this hospital in 2004.

The main epidemiologic features of the outbreak are indicated in Table 1. This outbreak occurred between March and July 2004 and involved 7 mechanically ventilated patients. These patients were ventilated between 4 and 20 days. The main initial clinical features best describing these patients were respiratory distress for patient G, pleurisy

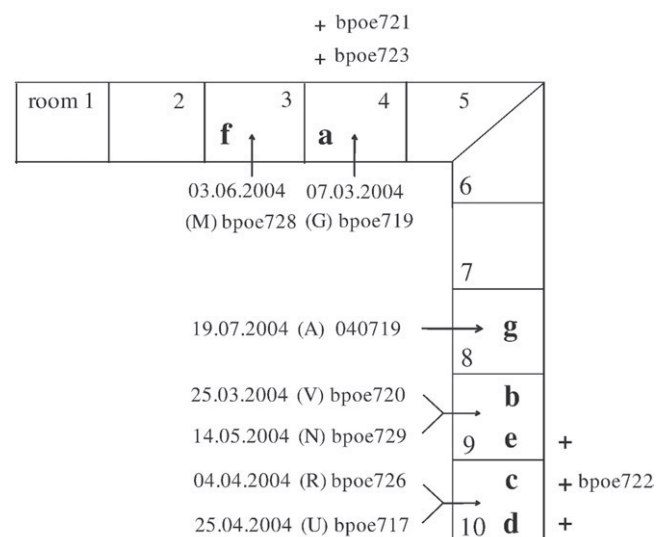


Fig. 1. Outline of the ICU considered in this study. Position of the Bcc-infected patients (capital letters and arrows) and positive environmental samples (+) are indicated. Small letters from a to g are used to illustrate the spread of the etiologic agents. For each patient, date of the first Bcc positive sample is indicated. bpoe collection codes of the Bcc strains are indicated. Strains are described in Table 2 and in the text.



Table 1  
Epidemiologic features of the ICU Bcc outbreak

Date of admission	Patient code	Sex	Age	Underlying illness	Bcc-positive sample	Length of stay before infection
07/03/04	G	M	63	Respiratory distress	Bronchial mucus	8 days
25/03/04	V	F	74	Pleurisy	Tracheal mucus	17 days
04/04/04	R	M	78	Meningitis	Tracheal mucus	9 days
25/04/04	U	F	nk	Pyelonephritis	Bronchoalveolar fluid	4 days
14/05/04	N	F	nk	Pneumopathy	Tracheal mucus	17 days
03/06/04	M	F	nk	Peritonitis	Tracheal mucus	7 days
19/07/04	A	M	22	Polytraumatism	Bronchial mucus	20 days

nk = not known.

Table 2  
Strains of the Bcc used or isolated in this study

Strains	Origin/description	Genomovar/RFLP group	Reference or author
LMG 16656 <sup>T</sup>	CF patient UK (ET12 lineage)	IIIA	Vandamme et al. (2003)
LMG 18826	CF patient Canada (ET12 lineage)	IIIA	Sajjan et al. (1995)
LMG 18863	CF patient Canada (ET12 lineage)	IIIA	Lewenza et al. (1999)
LMG 12615	CF patient UK (ET12 lineage)	IIIA	Vandamme et al. (1997)
H111	CF patient	IIIA	Romling et al. (1994)
LMG 18829	Midwest clone	IIIB	Vandamme et al. (2003)
AU1054	PHDC clonal	IIIB	Chen et al. (2001)
HI2424	PHDC lineage	IIIB	LiPuma et al. (2002)
LMG 16654	CF patient	IIIB	Vandamme et al. (2003)
LMG 18830	CF patient	IIIB	Vandamme et al. (2003)
LMG 19230	Kapunda Australia	IIIC	K. Ophel-Keller (1995) <sup>a</sup>
LMG 19240	Kapunda Australia	IIIC	K. Ophel-Keller (1995) <sup>a</sup>
LMG 21461	CF patient Argentina	IIID	Mahenthalingam et al. (2000b)
bpoe717	Bronchoalveolar fluid (U)	IIIA	This study
bpoe716	Tracheal mucus (R)	IIIA	This study
bpoe718	Respiratory tract (C)	IIIA	This study
bpoe719	Bronchial mucus (G)	IIIA	This study
bpoe720	Tracheal mucus (V)	IIIA	This study
bpoe721 <sup>b</sup>	Tap water of room 4	IIIA	This study
bpoe722	Tap water of room 10	IIIA	This study
bpoe723	Branches of the respiratory circuits (G) <sup>c</sup>	IIIA	This study
bpoe724	Water from the humidifier (O) <sup>c</sup>	IIIA	This study
bpoe725	Water from the humidifier (U) <sup>c</sup>	IIIA	This study
bpoe726	Respiratory tract (R)	IIIA	This study
bpoe727 <sup>b</sup>	Water from the humidifier (G) <sup>c</sup>	IIIA	This study
bpoe728	Tracheal mucus (M)	IIIA	This study
bpoe729 <sup>b</sup>	Tracheal mucus (N)	IIIA	This study
bpoe730 to 732 <sup>b</sup>	Water from inspirator (N) <sup>c</sup>	IIIA	This study
bpoe734	Reyssouze river, site A—near the hospital	II	This study
bpoe735	Reyssouze river, site A—near the hospital	II	This study
bpoe733	Reyssouze river, site B—near the hospital	V	This study
bpoe736	Sediment of a pond—near the hospital	V	This study
bpoe737	Sediment of a pond—near the hospital	V	This study
bpoe738	Pond surface water—near the hospital	V	This study
LMG 1222 <sup>T</sup>	Allium cepa	I	Burkholder (1949) <sup>a</sup>
LMG 17588	Soil with anthranilate	II	Lessie et al. (1996)
LMG 14294 <sup>T</sup>	CF patient Belgium	IV	LiPuma et al. (1999)
G4	Wastewater containing TCE	V	Nelson et al. (1987)
LMG 18941	CF patient	VI	J. LiPuma (1996) <sup>a</sup>
LMG 19182 <sup>T</sup>	Rhizosphere of pea	VII	Coenye et al. (2001)
LMG 20980 <sup>T</sup>	Soil rhizosphere USA	VIII	Vandamme et al. (2002)
LMG 14191 <sup>T</sup>	Soil	IX	Vandamme et al. (1997)

The letter code in parentheses for the bpoe strains indicates the code used to identify the patients or rooms of the ICU. See Fig. 1 for their position in the ICU.

T = type strain.

<sup>a</sup> No report available.

<sup>b</sup> Analyzed by MLST.

<sup>c</sup> Samples collected from an artificial respiratory device.

for patient V, meningitis for patient R, pyelonephritis for patient U, pneumopathy for patient N, peritonitis for patient M, and polytraumatism for patient A. No patient had CF. After several days in the ICU, these patients developed signs of pulmonary tract infections. Bronchial (patient G) or tracheal (patient V, N, and M) aspiration fluids, or bronchoalveolar washing fluids (patient U and A) were collected, and analyzed as described below.

After the first signs of clustered cases detected (April 2004), environmental samplings of the ICU were performed. Tap water, drain-off, expiratory filters of ventilators, and iodine bottles were sampled. Multidose nebulizer solutions and alcohol-free mouthwash liquids were also sampled. These samplings were performed every week until the end of the outbreak. Several months after the outbreak, swabbings of the general aeration system of the ICU, and of tap nozzle aerators of each sink were performed. Tap water samples were collected. The closest medical unit (nephrology) and outdoor sites (soils, river and pond) to the hospital building were also sampled.

## 2.2. Isolation of bacterial strains

Bronchial and tracheal aspiration fluids and bronchoalveolar washing fluids were liquefied using dithiothreitol. Dilutions of these samples were plated on Columbia agar with 5% sheep blood (AES laboratories, Rennes, France) and

incubated at 37 °C. ICU environmental samples were plated on Bromocresol Purple media (bioMérieux, Marcy l'Etoile, France) and incubated at 37 °C. Prior plating of outdoor samples, an enrichment procedure derived from Vermis et al. (2003) was applied. The enrichment broth included L-arabinose (0.1%), L-threonine (0.1%), polymyxin B (300 µg/mL), and 9-chloro-9-(4-diethylaminophenyl)-10-phenylacridan (C-390 at 3 µg/mL). After 5 days of incubation at 28 °C, the enrichment broths were diluted and plated on BCSA (Henry et al., 1997).

All strains used and isolated in this study are indicated in Table 2. Routine bacterial cultures were grown in LB (Bertani, 1951) broth at 28 °C for 24 h.

## 2.3. Phenotypic analyses and antibiotic resistances

Identification of bacterial strains by phenotypic analyses was performed using API 20NE strips® (BioMérieux sa) and/or the VITEK2® system (BioMérieux). Antibiotic resistance profiles were performed using the VITEK2® system and its Advanced Expert System™.

## 2.4. DNA manipulations

Bcc DNA extractions were performed according to Pitcher et al. (1989). PCR reactions were performed in 25 µL using 1U *Taq* polymerase, 200 µM of each dNTP, 1× PCR buffer, 5% DMSO, 1.5 mmol/L MgCl<sub>2</sub>, 0.6 µmol/L of each PCR

Table 3  
PCR DNA primers used in this study

Gene	Function	Product size (bp)	Sequence (5' to 3')	Annealing temperature (°C)	Reference
<i>ecfB</i>	ECF sigma factor	450	<i>ecfBF</i> CGCTTCGCGCTCTGGCTG <i>ecfBcenR</i> TCAGCCGCAAGGAAGGAGC	63	This study
<i>B. cenocepacia</i>	ECF sigma factor	460	<i>ecfBmultiF</i> ATGCCCTCAAACGCCCTCG <i>ecfBmultiR</i> CGACAGCCGCGACATCACC	58	This study
<i>B. multivorans</i>	ECF sigma factor	440	<i>ecfBccF</i> GGCTGCGACGCTTCGCGC <i>ecfBccR</i> CCGTCGCTGAGCTGGCGC	64	This study
<i>Bcc species</i>	SOS response	1043	BCR1 TGACCGCCGAGAAGAGCAA BCR2 CTCTTCTTCGTCCATCGCCTC	58	Mahenthalingam et al. (2000a)
<i>recA</i>	Cable pilus	237	cbIAF GCAGCTGTAGTGAACACG cbIAR TCTGACCGATCGACAGCG	58	Tomich and Mohr (2004)
BCESM	Epidemic marker	1418	BCESMF CCACGGACGTGACTAACA BCESMR CGTCCATCCGAACACGAT	62	Baldwin et al. (2004)
<i>cepI</i>	Quorum sensing	600	<i>cepIF</i> CCCTGTAAGAGTTACCAGTT <i>cepIR</i> GATATCGATCCAGCACGCCACGAC	56	Gotschlich et al. (2001)
<i>cepR</i>	Quorum sensing	450	<i>cepRF</i> CTGGATGGCGCACTACCAGGC <i>cepRR</i> ACGTGGAAGTTGACCGTCCGC	60	Gotschlich et al. (2001)
<i>fliG</i>	Flagellin	1000	<i>fliGF</i> ATGAACGCTGAAGGCTTGAC <i>fliGR</i> TCAGACATACGCGTCTTCC	56	This study
<i>llpE</i>	Efflux pump	625	<i>llpEF</i> GACCGGGTTGCCAGTAGTGC <i>llpER</i> CGTGCGTGATGTCCGGGTAG	55	This study
<i>zmpA</i>	Metalloprotease	1500	<i>zmpAF</i> CGCGGCAGACATCGACTAC <i>zmpAR</i> AGATGCCGTTGCGGTTGT	58	Subsin et al. (2007)
<i>wbiI</i>	LPS	542	<i>wbiIF</i> CCATGCGGCGGTACAAGCACGTGCC <i>wbiIR</i> TAGAGCTTCTCGCCCGGCGCAG	52	Vinion-Dubiel et al. (2004)
<i>bescV</i>	TTSS	379	<i>bescVF</i> GCCCTTCACGAACTTCATC <i>bescVR</i> GCTGCTGCTGTTACGAC	53	Parsons et al. (2001)

TTSS = type III secretion system.

DNA primer, and 10–50 ng of template DNA. PCR thermal cycles ( $n = 35$ ) were carried out on a Veriti™ 96-Well Cyclor (Applied Biosystems, Foster City, CA), and each cycle was as follows: 45 s at 95 °C, then 45 s at the recommended annealing temperature for the PCR primers (see Table 3), and finally 1 min at 72 °C. After these cycles, a final elongation step of 5 min at 72 °C was performed. PCR screenings of the *B. cepacia* epidemic strain marker (BCESM) genes were done according to Baldwin et al. (2004). DNA sequencing of the PCR products was performed by Cogenics® (Grenoble, France). GenBank accession numbers of the DNA sequences obtained in this work are as follows: FJ670544 to FJ670565 for the *recA*, FJ668290 to FJ668310 for *ecfB*. MLST GenBank accession numbers for strains bpoe721, bpoe727, bpoe729, and bpoe731 are FJ668311 to FJ668330 and FJ666039 to FJ666042.

*Hae*III RFLP analysis of *recA* PCR products were performed according to the recommendations of the manufacturer (FERMENTAS, ST Remy Lès Chevreuse, France). RFLP profiles were analyzed by agarose gel electrophoresis following Cournoyer and Normand (1994).

Pulse field gel electrophoresis (PFGE) of *Spe*I-restricted total genomic DNA was performed according to Romling et al. (1997). Electrophoresis was run for 11 h with pulses of 5–25 s then 13 h with pulses of 5–60 s. Voltage was kept at 5 V/cm. *Spe*I DNA fragments were visualized by ethidium bromide staining and UV exposure using a gel doc system (Biorad®, Marnes-la-Vallée, France).

### 2.5. Molecular phylogenetic analyses

Multiple alignments of DNA sequences used in this study were computed using CLUSTALW (Thompson et al., 1994). Distances between sequence pairs, the inferred phylogenetic trees, and bootstrap values were all computed through the Mega3 graphic tool (Kumar et al., 2004). Phylogenetic trees were built using the neighbor-joining method (Saitou and Nei, 1987). One thousand bootstrap replicates were performed. Gap-containing sites were not considered in the analyses. The Bcc MLST scheme was performed as described by Baldwin et al. (2005).

## 3. Results

### 3.1. Description of the outbreak

Between March 2004 and July 2004, 7 mechanically ventilated patients hospitalized in a French ICU developed respiratory tract infections. The main features and the synoptic table of the outbreak are presented, respectively, in Table 1 and Fig. 2. The index case, named G, was a 63-year-old man suffering of acute respiratory distress. One of the infected patient, named U, died of respiratory failure, 3 days after the nosocomial pulmonary infection was diagnosed.

Microbiological analysis of bronchial or tracheal aspiration fluids or bronchoalveolar washing fluids of these patients

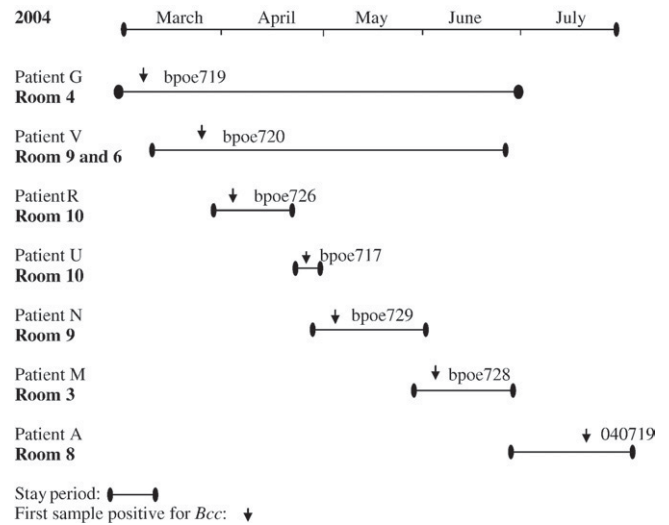


Fig. 2. Synoptic table of the Bcc outbreak.

was performed. The dominant bacterial colony morphotypes, found on Columbia agar with 5% sheep blood, and recovered on the dates indicated on Figs. 1 and 2, were identified as being strains of the Bcc by phenotypic analyses using VITEK2® or API 20NE strips. Before infection by a Bcc strain, patient V was found colonized by a bacterial strain of *Pseudomonas aeruginosa*, according to the VITEK2® system, and patient A by a strain of *Citrobacter freundii*.

### 3.2. Environmental sources of Bcc-like strains

Patients concerned by this outbreak occupied 5 of the 10 rooms of the ICU (Fig. 1). The index patient (G) occupied room 4, patient V room 9 then 6, patient N room 9 after the transfer of patient V to room 6, patient R room 10, patient U room 10 after patient R left the unit, and patient M room 3. Length of the stays in the ICU was quite variable from one patient to another (Fig. 2).

Environmental samplings were performed after infection of patient U. Tap water samples, tap nozzle aerators, drain-off, filters of ventilators, iodine bottles, multidose nebulizer solutions, and alcohol-free mouthwash liquids were sampled and analyzed for the presence of Bcc strains. Tap water samples, mechanical respiratory ventilators, and condensing humidifiers from the index patient's room and from the room of patient U were found to contain Bcc strains according to VITEK2® analyses. Following these results, disinfection of the taps, drain-offs, ventilators, and humidifiers were undertaken. Moreover, bottled water was used to prepare drugs administered by gastric tubes and to perform patients' hygienic care. Medical staff hand washings were reinforced.

Despite these measures, a fifth case of Bcc infection, patient N, was reported. Water, collected from the mechanical respiratory ventilator and humidifier of this patient, was found to contain Bcc-like strains. This patient was

occupying room 9, which had been used by patient V, the second patient to have been infected by a Bcc-like strain. About 3 weeks after this case, the sixth patient got infected. This patient was not in a room with a previous record of Bcc-like pulmonary infection. However, water from the ventilator, and bottles of soap and iodine of this room were found colonized by Bcc-like strains. Cleaning and disinfection measures were again undertaken. After this case, tap water samples from all rooms were analyzed, but none showed the presence of Bcc-like strains. Anyhow, a last case of Bcc-like infection occurred, in a room that had not received any of the previously infected patients.

This last non-point source contamination led us to further investigate the distribution of Bcc-like strains around the ICU and the hospital building. An enrichment step in the isolation procedure was introduced to increase the probability of isolating Bcc-like strains. Using this approach, water samples and swabs from the ICU medical unit and the closest neighboring unit (nephrology) failed to give any Bcc-like strain. Soils surrounding the hospital building were also found negative. However, surface water samples coming from the Reyssouze river, a large pond close to this river, and the sediment from the overflow channel of this pond and connected to the Reyssouze river, all located within 1 km from the hospital and near a wastewater treatment plant, gave bacterial strains ( $n = 11$ ) using the enrichment procedure. Six of these strains were identified by the VITEK2<sup>®</sup> system as Bcc, one as *Rhizobium radiobacter*, and one as *Ochrobactrum anthropi*. Three of these strains could not be identified with the VITEK2<sup>®</sup> database.

### 3.3. Molecular characterization of the Bcc-like isolates

#### 3.3.1. *recA* PCR-RFLP analysis

Eighteen Bcc-like strains isolated from the ICU patients, the ICU environment, and outdoor sites close to the hospital building were analyzed by molecular approaches to clarify their taxonomic status. The *recA* *Hae*III PCR profiles of the collected strains were compared with those of reference strains of most of the Bcc species. This analysis showed all hospital isolates to have a *B. cenocepacia*-like *recA* *Hae*III profile. However, the six Bcc strains coming from outdoor environments were found to have profiles matching those of *B. multivorans* (bpoe734 and bpoe735) and *B. vietnamiensis* (bpoe733, bpoe736–bpoe738). The *B. cenocepacia* *recA* *Hae*III profile of the hospital isolates was found matching the one of subgroup IIIA, a subgroup well known to contain the highly virulent and transmissible ET12 clone. Phylogenetic analyses and PCR screenings of virulence genes were performed to further investigate the proximity of these outbreak strains with those of the ET12 clone.

#### 3.3.2. Phylogenetic analyses

DNA sequencings of the full *recA* PCR products from strains of the outbreak clone ( $n = 14$ ) and the environmental Bcc strains ( $n = 6$ ) reported here were performed. These sequences were aligned with database sequences, and a

phylogenetic tree was built. All *recA* sequences from the outbreak clone were found identical. These sequences were also identical to those of several ET12 strains. *recA* phylogenetic analysis positioned these sequences with those of the *recA* PCR RFLP group IIIA sequences. These sequences were significantly apart from those of strains of the IIIB, IIIC, and IIID RFLP subgroups (data not shown). However, infra-specific *recA* divergences within the IIIA group could not be investigated because of the low number of substitutions between these sequences. Nevertheless, the *recA* phylogeny revealed reliable sublineages among the IIIB RFLP subgroup.

The *recA* sequences from Bcc strains of the surface water from the Reyssouze river, a large pond, and sediment from this pond overflow channel (connected to the river) were positioned in the *recA* phylogenetic tree. The *recA* sequences from strains with the *recA* RFLP profile of *B. vietnamiensis* were clustered with the *B. vietnamiensis* G4 *recA* sequence. Three of these strains (bpoe737, sediment of the pond overflow; bpoe738, pond; and bpoe733, river) were found to have identical sequences. The *recA* sequences of Bcc strains from the Reyssouze river surface water samples with a *recA* RFLP profile of the *B. multivorans* type were found identical and grouping with sequences from this last species. Alignment of the *recA* sequences is available at [http://www.ecomicro.univ-lyon1.fr/poe/recA\\_aln.txt](http://www.ecomicro.univ-lyon1.fr/poe/recA_aln.txt).

The inability to investigate infra-specific relationships among the PCR RFLP IIIA subgroup using the *recA* data set led us to consider other gene targets to resolve these relationships. Genes of the Bcc MLST scheme were considered. They were PCR amplified and sequenced for 4 of the *B. cenocepacia* ICU outbreak strains (bpoe721, bpoe727, bpoe729, and bpoe732). DNA sequences were compared with those of the ET12 clone (strain LMG16656), and 4 of 7 genes (*gltB*, *recA*, *lepA*, and *phaC*) were found identical, whereas 2 genes (*gyrB* and *atpD*) showed one nucleotide change and one (*trpB*) 2 nucleotide changes. The ICU outbreak strain could thus be classified into the Sequence Type 32 (ST32) of the MLST scheme. To complete this analysis, a phylogenetic tree was built from the concatenated gene sequences of the MLST scheme. This phylogenetic tree differentiated sequences according to their species of origin, with sequences from each *B. cenocepacia* *recA* PCR RFLP subgroups (IIIA, IIIB, IIIC, and IIID) being differentiated. However, infra-specific divergences could not be resolved inside the IIIA subgroup.

Another gene target was thus used to achieve this goal. Our previous studies showed that genes encoding extracytoplasmic function (ECF) factors can allow a good estimation of infra-specific relationships (Cournoyer et al., 1996) and that the *ecfB* gene was found distributed among all Bcc species (Menard et al., 2007a). PCR primers were designed to amplify this gene from all Bcc strains. *ecfB* PCR products were obtained and sequenced from the *B. cenocepacia*, *B. vietnamiensis*, and *B. multivorans* strains of this study. This set of *ecfB*

sequences was completed with sequences from the databases. Additional sequences were obtained in this work for the following strains: LMG21461, LMG12615, LMG18826, LMG18863, LMG16654, LMG18830, LMG19240, LMG19230, LMG1222, LMG14294, LMG14191, LMG20980, LMG18941, and LMG17588. A phylogenetic tree was computed from this *ecfB* data set (data not shown). Alignment of the *ecfB* sequences is available at [http://www.ecomicro.univ-lyon1.fr/poe/ecfB\\_aln.txt](http://www.ecomicro.univ-lyon1.fr/poe/ecfB_aln.txt). The *ecfB* sequences from the outbreak strains ( $n = 14$ ) were found identical but showed significant differences with the ET12 *ecfB* sequences (8 nucleotide changes; 97% identity). The main clusters of the *ecfB* phylogenetic tree were found matching those inferred from the *recA* and MLST data sets, but a better resolution of infra-specific divergences was obtained. This led us to combine the *ecfB* and *recA* data sets to obtain

the most informative phylogenetic tree (Fig. 3). This phylogenetic tree shows a clear differentiation of sequences of the outbreak clone and those of ET12 strains (LMG16656, LMG12615, LMG18863, and LMG18826) among the cluster of sequences from strains of the *B. cenocepacia* *recA* PCR RFLP IIIA subgroup.

### 3.3.3. PFGE-*SpeI* analysis

PFGE analysis of *SpeI*-restricted total genomic DNA (PFGE-*SpeI*) of the strains recovered in this study was performed. The 14 *B. cenocepacia* strains analyzed showed the same PFGE-*SpeI* profile regardless of their origin (patients, tap water, etc.) (Fig. 4). *B. vietnamiensis* strains of the surface water of the Reyssozou river, a large pond, and sediment from this pond overflow channel (connected to the river) with identical *recA* and *ecfB* sequences showed the same PFGE-*SpeI* profile. The *B. vietnamiensis*

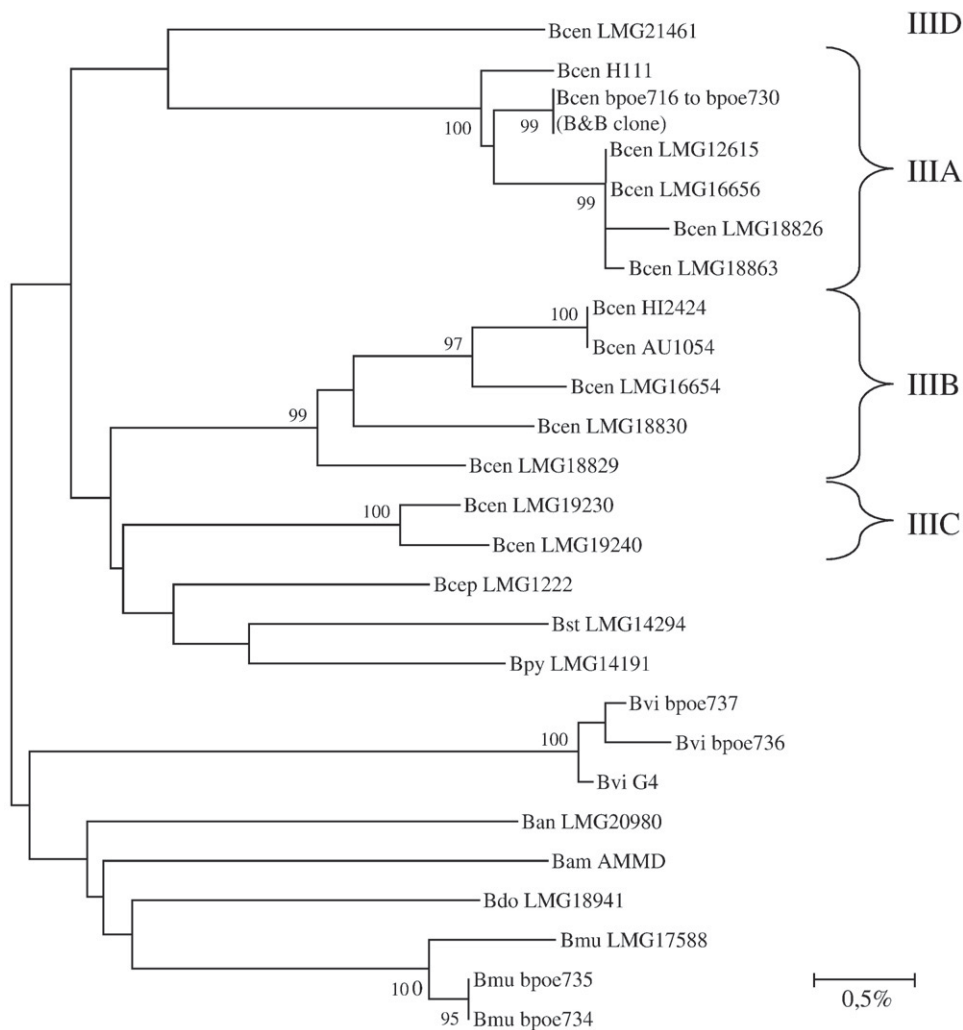


Fig. 3. Neighbor-joining phylogenetic tree of concatenated *recA* and *ecfB* gene sequences from the outbreak and representative strains of the Bcc. DNA sequences were either obtained in this work or retrieved from the GenBank database. A total of 654 sites were analyzed. Distances are proportional to evolutionary divergences expressed in substitutions per 100 sites. Bootstrap values above 90% are indicated on the tree branches. Bcen = *B. cenocepacia*; Bcep = *B. cepacia*; Bst = *B. stabilis*; Bpy = *B. pyrrocinia*; Bvi = *B. vietnamiensis*; Ban = *B. anthina*; Bam = *B. ambifaria*; Bdo = *B. dolosa*; Bmu = *B. multivorans*.

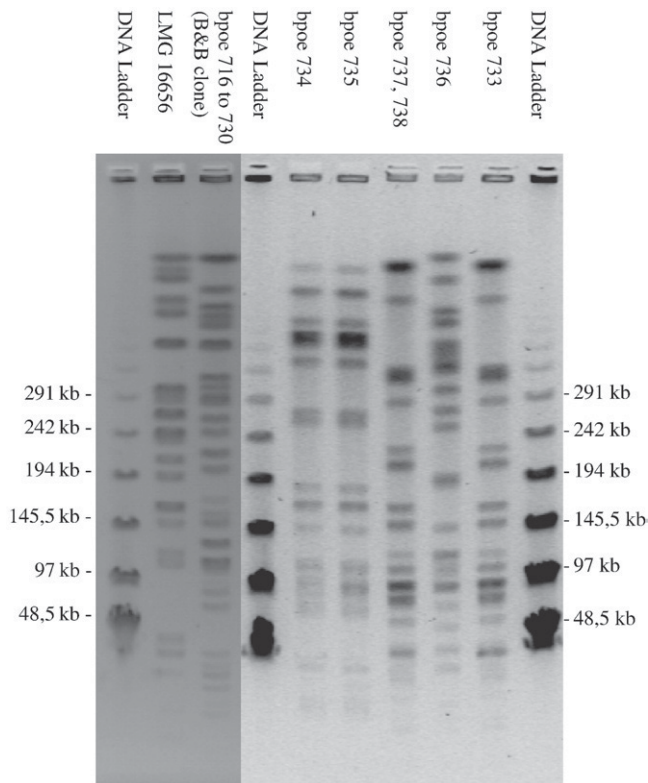


Fig. 4. *SpeI* restriction patterns of total genomic DNA from all Bcc strains reported in this work. PFGE conditions are described in the text. All outbreak strains (bpoe716 to bpoe730) of the ICU showed the same profile, and this profile is indicated under the name B&B clone. The Pulse Marker™ ladder (50–1000 kb) is shown. The ET12 profile (LMG 16656) is shown for comparative purposes. See Table 2 for description of the strains.

bpoe736 strain showed a distinct profile. *B. multivorans* river strains bpoe734 and bpoe735 showed the same PFGE-*SpeI* profile.

#### 3.3.4. Virulence genes and antibiotic resistance profiles

PCR screenings of 9 genetic determinants (*cbIA*, BCESM, *cepI*, *cepR*, *fliG*, *llpE*, *zmpA*, *wbiI*, *bcsC*) associated with virulence were performed on genomic DNA extracted from the *B. cenocepacia* outbreak strains. All screenings but one gave positive PCR products of the expected size. Only the *cbIA* gene could not be detected. This result was confirmed by *cbIA* DNA probing of genomic DNA blot (data not shown) of the outbreak strains. All genes belonging to the BCESM (*cci* island) were detected, by PCR, in the outbreak strains.

Antibiotic resistance profiles using the VITEK2® system showed the *B. cenocepacia* outbreak strains ( $n = 14$ ) to be resistant to aminosides, quinolones, and a large part of betalactamines. Results were as follow: (1) strains were resistant to ticarcillin (MIC > 128), ticarcillin plus acid clavulanic (MIC > 128), imipenem (MIC > 16), amikacin (MIC > 64), gentamicin (MIC > 16), isepamicin (MIC > 64), tobramycin (MIC > 16), ciprofloxacin (MIC > 4), pefloxacin (MIC = 8), minocycline (MIC > 16), colistin

(MIC > 16); and (2) were sensitive to piperacillin (MIC < 4), piperacillin plus tazobactam (MIC < 4), ceftazidim (MIC = 4), and trimethoprim/sulfamethoxazole (MIC < 20). All patients were treated with Fortum® (ceftazidim, Glaxo-SmithKline, Marly-le-Roi, France) and Bactrim® (trimethoprim/sulfamethoxazole, Roche, Paris, France).

*B. multivorans* river strains showed resistance to ticarcillin (MIC > 128), ticarcillin plus acid clavulanic (MIC > 128), imipenem (MIC > 16), amikacin (MIC > 64), gentamicin (MIC > 16), isepamicin (MIC > 64), tobramycin (MIC > 16), colistin (MIC > 16), and were sensitive to piperacillin (MIC < 4), piperacillin plus tazobactam (MIC < 4), ceftazidim (MIC = 2), trimethoprim/sulfamethoxazole (MIC < 20), ciprofloxacin (MIC = 1), pefloxacin (MIC = 1), and minocycline (MIC < 1). *B. vietnamiensis* strains showed resistance to gentamicin (MIC > 16) and colistin (MIC > 16) and were sensitive to piperacillin plus tazobactam (MIC < 4), ceftazidim (MIC = 2), ciprofloxacin (MIC < 1), minocycline (MIC < 1), and trimethoprim/sulfamethoxazole (MIC < 20).

## 4. Discussion

The ICU nosocomial outbreak of pulmonary tract infections reported here lasted a few months and had serious consequences on one patient, who died only 7 days after his admission. This dramatic outcome led us to investigate the environmental origin of the etiologic agents involved in this outbreak and to analyze their genetic proximity with previously reported highly transmissible infectious agents. Biochemical tests performed through the VITEK2® system identified the etiologic agents as being part of the Bcc. This identification was confirmed by various molecular approaches and was further investigated at the species and infra-specific levels. These analyses showed the etiologic agents to be clonal and closely related to a highly virulent and transmissible *Burkholderia cenocepacia* clone named ET12. However, significant differences between ET12 and the outbreak strains of this report led us to differentiate this clone by naming it B&B, according to its geographic origin (Bourg-en-Bresse, France).

The environmental origin of the B&B clone could only be partly resolved in this study. Tap water samples, water from mechanical respiratory ventilators and humidifiers, swabs of tap nozzle aerators, drain-offs, iodine bottles, multidose nebulizer solutions, and liquid from alcohol-free mouthwash were collected and analyzed. The phenotypic and molecular analysis showed a Bcc contamination of tap water and of the mechanical respiratory ventilator and humidifier of the room of the index case of this outbreak. The presence of the B&B clone in waters of mechanical respiratory ventilators and humidifiers in the rooms of the subsequently infected patients was also observed. This highlighted the likely involvement of the medical staff in its transmission. This possibility was further supported by the detection on the

neck of bottles of soap and iodine of isolates of the B&B clone. However, these bottles did not appear the source of this pathogen; unopened bottles of the same batches did not show any signs of these bacteria. It is noteworthy that several B&B-infected patients did not occupy contiguous rooms, adding further support for cross-contaminations between staff and patients. Nevertheless, these samplings did not identify the source of this clone in the ICU. Further environmental sampling of water samples and air conditioning systems did not lead to the identification of this source. Outdoor environments near the hospital were then investigated but did not yield any isolates of this clone. However, surface water samples coming from the Reyssouze river, a large pond close to this river, and the sediment from the overflow channel of this pond and connected to the Reyssouze river, all located within 1 km from the hospital and being close to a wastewater treatment plant, yielded isolates of 2 species of Bcc, after an enrichment procedure, that is, *B. multivorans* and *B. vietnamiensis*. These are the first French outdoor strains of these 2 species, but several isolates of these species from French cohorts of CF patients have been reported by the “Observatoire National *Burkholderia cepacia*” (C. Segonds and G. Chabanon, Hôpital Rangueil, France).

To get further insights about the source of the B&B clone and establish its proximity with virulent and transmissible *B. cenocepacia* strains, MLST and phylogenetic analyses were performed. Bcc *recA* molecular phylogeny confirmed the

*recA* RFLP typings, which related the B&B clone to the *B. cenocepacia* IIIA subgroup and ET12 clone. In fact, the *recA* sequences from the B&B clone and several strains of the ET12 clone were found identical. However, the MLST analysis grouped the B&B clone with strains of a sequence type with no record of ET12-like isolates. This sequence type was found to group strains isolated from radish in Mexico, a non-CF patient (UK), and CF patients from Belgium, and the Czech Republic (Baldwin et al., 2007). On the Bcc MLST web site (<http://www.pubmlst.org/bcc/>), this sequence type was further found to include strains isolated from several Canadian and US CF patients, with the oldest strain isolated in 1991. Fig. 5 shows the geographic repartition of B&B-like strains and suggests a wide distribution among occidental countries. Nevertheless, phylogenetic analysis of the concatenated set of MLST gene sequences could not resolve, with a high level of confidence, the relationships between all sequence types harboring ET12 strains and the one of the B&B clone. To clarify this point, an additional genetic marker of phylogenetic relationships among *B. cenocepacia* was used. A gene encoding an ECF sigma 70 factor, named *ecfB* (Menard et al., 2007a), was selected. *ecfB* gene sequences were obtained from reference strains of the Bcc and of the B&B clone. *ecfB* molecular phylogeny was then investigated and compared with the ones inferred from *recA* and the MLST genes sequences. The *ecfB* phylogenetic tree resolved infra-specific relationships among the *B. cenocepacia* IIIA *recA* PCR RFLP subgroup and showed the B&B

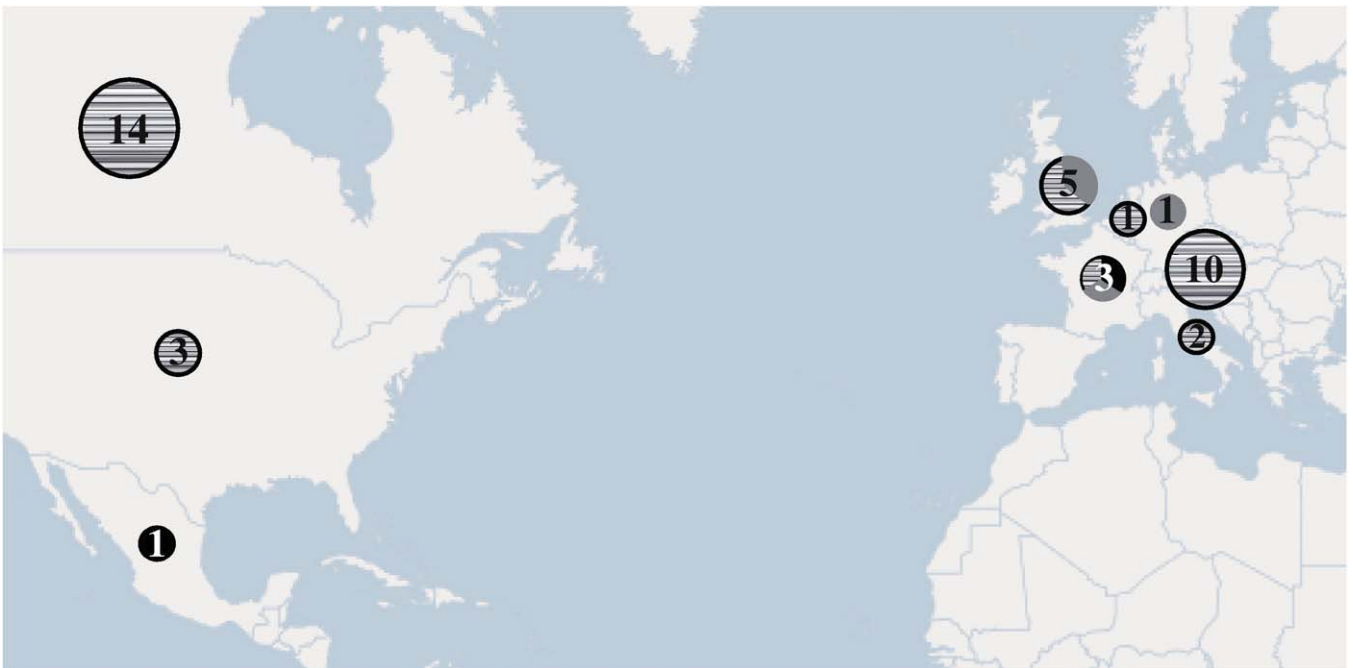


Fig. 5. Global geographic distribution of B&B-like strains of *B. cenocepacia*. Position of the circles indicates the country of origin of strains found to belong to ST32 based on the Bcc MLST scheme and data available at <http://www.pubmlst.org> (in May 2009). Size of the circles is proportional to the number of strains typed for each country, according to data on <http://www.pubmlst.org> and this study. Black circles indicate environmental strains, gray circles indicate non-CF strains, and circles with black lines those isolated from CF patients. The B&B clone was considered as an equivalent of one strain per context, that is, one strain for the environmental samples and one for the clinical ones.

clone and ET12 strains to be closely related but part of distinct lineages. PCR screenings of virulence genes among the B&B clone were in line with this separation between the B&B and ET12 strains. The B&B clone was found to harbor most of the virulence genes tested including the BCESM genomic island (Baldwin et al., 2004) but not the *cblA* gene, which is known to be mainly found among ET12 isolates. These data sets showed the MLST e-Burst analyses, *ecfB* molecular phylogeny, and virulence genes distribution analyses to be in agreement. The *ecfB* analysis was found to have the advantage of making use of a single genetic locus, reducing significantly the cost of such epidemiologic investigations. However, it would now need to be validated on a larger panel of bacterial strains to further confirm its reliability as a tool to investigate infra-specific divergences between other *B. cenocepacia* strains and clonal complexes. It is noteworthy that the *ecfB* gene has the potential to be used as a target for the development of a quantitative PCR approach for the Bcc. Quantitative PCR would be very useful to address issues such as the one of the viable but not culturable state of these bacteria in outdoor and clinical situations. About 97.5% of these bacteria found in river surface waters were considered none culturable by a fluorescent in situ hybridization method making use of a 16S rRNA DNA probe (Lemke and Leff, 1999).

Rarely Bcc strains involved in nosocomial outbreaks were compared to CF isolates, but a large hospital outbreak of pulmonary infections reported in 1999 showed that both patient groups can be colonized by a same clone (Holmes et al., 1999). In this outbreak, acquisition risk in the CF patients was associated with significantly increased mortality rates, suggesting that highly virulent CF epidemic clones could emerge from the hospital environment. In our investigation, the B&B clone was found related to several CF isolates from various countries. These data thus confirm the observations made by Holmes et al. (1999) that both patient groups can be infected by a same clone or clonal complex. However, this is not an evidence of similar bacterial weapons or virulence expression profiles leading to the observed pulmonary infections. Genomic refinements involving selection of adaptive mutations or of novel genes acquired by horizontal gene transfer events can occur and lead to major differences between closely related strains as reported among *B. vietnamiensis* (Menard et al., 2007b). Such genomic refinements would now need to be investigated between the nosocomial B&B clone and closely related isolates from CF patients to further improve our understanding of the differences between CF and non-CF pulmonary tract infections.

## Acknowledgments

A. Graindorge was supported by a PhD fellowship from the “Cluster Environnement” of the Rhône-Alpes Region (France) and by Université Lyon 1. We thank the CNRS,

Université Lyon 1, and Rhône-Alpes Region (France) for having funded parts of this work. This work was also partly funded by the Agence Nationale de la Recherche (ANR) SEST 2005 009-01, SEST 2007 018-01, and CESA 2008 022-01 projects, and the “Association Vaincre la Mucoviscidose”.

## References

- Alvarez-Lerma F, Maull E, Terradas R, Segura C, Planells I, Coll P, Knobel H, Vazquez A (2008) Moisturizing body milk as a reservoir of *Burkholderia cepacia*: outbreak of nosocomial infection in a multidisciplinary intensive care unit. *Crit Care* 12(1):R10.
- Baldwin A, Sokol PA, Parkhill J, Mahenthalingam E (2004) The *Burkholderia cepacia* epidemic strain marker is part of a novel genomic island encoding both virulence and metabolism-associated genes in *Burkholderia cenocepacia*. *Infect Immun* 72(3):1537–1547.
- Baldwin A, Mahenthalingam E, Thickett KM, Honeybourne D, Maiden MC, Govan JR, Speert DP, Lipuma JJ, Vandamme P, Dowson CG (2005) Multilocus sequence typing scheme that provides both species and strain differentiation for the *Burkholderia cepacia* complex. *J Clin Microbiol* 43(9):4665–4673.
- Baldwin A, Mahenthalingam E, Drevinek P, Vandamme P, Govan JR, Waite DJ, LiPuma JJ, Chiarini L, Dalmastrì C, Henry DA, Speert DP, Honeybourne D, Maiden MC, Dowson CG (2007) Environmental *Burkholderia cepacia* complex isolates in human infections. *Emerg Infect Dis* 13(3):458–461.
- Bertani G (1951) Studies on lysogeny. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol* 62(3):293–300.
- Chen JS, Witzmann KA, Spilker T, Fink RJ, LiPuma JJ (2001) Endemicity and inter-city spread of *Burkholderia cepacia* genomovar III in cystic fibrosis. *J Pediatr* 139(5):643–649.
- Coenye T, Vandamme P (2003) Diversity and significance of *Burkholderia* species occupying diverse ecological niches. *Environ Microbiol* 5(9):719–729.
- Coenye T, Mahenthalingam E, Henry D, LiPuma JJ, Laevens S, Gillis M, Speert DP, Vandamme P (2001) *Burkholderia ambifaria* sp. nov., a novel member of the *Burkholderia cepacia* complex including biocontrol and cystic fibrosis-related isolates. *Int J Syst Evol Microbiol* 51(Pt 4):1481–1490.
- Compant S, Nowak J, Coenye T, Clement C, Ait Barka E (2008) Diversity and occurrence of *Burkholderia* spp. in the natural environment. *FEMS Microbiol Rev* 32(4):607–626.
- Courmoyer B, Normand P (1994) Characterization of a spontaneous thiostrepton-resistant *Frankia alni* infective isolate using PCR-RFLP of *nif* and *glnII* genes. *Soil Biol Biochem* 26(5):553–559.
- Courmoyer B, Arnold D, Jackson R, Vivian A (1996) Phylogenetic evidence for a diversification of *Pseudomonas syringae* pv. *pisi* race 4 strains into two distinct lineages. *Phytopathology* 86:1051–1056.
- Ghazal SS, Al-Mudaimigh K, Al Fakihi EM, Asery AT (2006) Outbreak of *Burkholderia cepacia* bacteremia in immunocompetent children caused by contaminated nebulized sulbutamol in Saudi Arabia. *Am J Infect Control* 34(6):394–398.
- Gotschlich A, Huber B, Geisenberger O, Togl A, Steidle A, Riedel K, Hill P, Tummler B, Vandamme P, Middleton B, Camara M, Williams P, Hardman A, Eberl L (2001) Synthesis of multiple *N*-acylhomoserine lactones is wide-spread among the members of the *Burkholderia cepacia* complex. *Syst Appl Microbiol* 24(1):1–14.
- Henry DA, Campbell ME, LiPuma JJ, Speert DP (1997) Identification of *Burkholderia cepacia* isolates from patients with cystic fibrosis and use of a simple new selective medium. *J Clin Microbiol* 35(3):614–619.
- Hobson R, Gould I, Govan J (1995) *Burkholderia* (*Pseudomonas*) *cepacia* as a cause of brain abscesses secondary to chronic suppurative otitis media. *Eur J Clin Microbiol Infect Dis* 14(10):908–911.



- Holmes A, Nolan R, Taylor R, Finley R, Riley M, Jiang RZ, Steinbach S, Goldstein R (1999) An epidemic of *Burkholderia cepacia* transmitted between patients with and without cystic fibrosis. *J Infect Dis* 179(5): 1197–1205.
- Isles A, Macluskay I, Corey M, Gold R, Prober C, Fleming P, Levison H (1984) *Pseudomonas cepacia* infection in cystic fibrosis: an emerging problem. *J Pediatr* 104(2):206–210.
- Jacobson M, Wray R, Kovach D, Henry D, Speert D, Matlow A (2006) Sustained endemicity of *Burkholderia cepacia* complex in a pediatric institution, associated with contaminated ultrasound gel. *Infect Control Hosp Epidemiol* 27(4):362–366.
- Johnson WM, Tyler SD, Rozee KR (1994) Linkage analysis of geographic and clinical clusters in *Pseudomonas cepacia* infections by multilocus enzyme electrophoresis and ribotyping. *J Clin Microbiol* 32(4):924–930.
- Kotsanas D, Brett J, Kidd TJ, Stuart RL, Korman TM (2008) Disinfection of *Burkholderia cepacia* complex from non-touch taps in a neonatal nursery. *J Perinat Med* 36(3):235–239.
- Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5(2):150–163.
- Lenke MJ, Leff LG (1999) Bacterial populations in an anthropogenically disturbed stream: comparison of different seasons. *Microb Ecol* 38(3): 234–243.
- Lessie TG, Hendrickson W, Manning BD, Devereux R (1996) Genomic complexity and plasticity of *Burkholderia cepacia*. *FEMS Microbiol Lett* 144(2-3):117–128.
- Lewenza S, Conway B, Greenberg EP, Sokol PA (1999) Quorum sensing in *Burkholderia cepacia*: identification of the LuxRI homologs CepRI. *J Bacteriol* 181(3):748–756.
- LiPuma JJ, Mortensen JE, Dasee SE, Edlind TD, Schidlow DV, Burns JL, Stull TL (1988) Ribotype analysis of *Pseudomonas cepacia* from cystic fibrosis treatment centers. *J Pediatr* 113(5):859–862.
- LiPuma JJ, Dulaney BJ, McMenamin JD, Whitby PW, Stull TL, Coenye T, Vandamme P (1999) Development of rRNA-based PCR assays for identification of *Burkholderia cepacia* complex isolates recovered from cystic fibrosis patients. *J Clin Microbiol* 37(10):3167–3170.
- LiPuma JJ, Spilker T, Coenye T, Gonzalez CF (2002) An epidemic *Burkholderia cepacia* complex strain identified in soil. *Lancet* 359(9322): 2002–2003.
- Lo Cascio G, Bonora MG, Zorzi A, Mortani E, Tessitore N, Loschiavo C, Lupo A, Solbiati M, Fontana R (2006) A napkin-associated outbreak of *Burkholderia cenocepacia* bacteraemia in haemodialysis patients. *J Hosp Infect* 64(1):56–62.
- Magalhaes M, Doherty C, Govan JR, Vandamme P (2003) Polyclonal outbreak of *Burkholderia cepacia* complex bacteraemia in haemodialysis patients. *J Hosp Infect* 54(2):120–123.
- Mahenthalingam E, Bischof J, Byrne SK, Radomski C, Davies JE, Av-Gay Y, Vandamme P (2000a) DNA-Based diagnostic approaches for identification of *Burkholderia cepacia* complex, *Burkholderia vietnamiensis*, *Burkholderia multivorans*, *Burkholderia stabilis*, and *Burkholderia cepacia* genomovars I and III. *J Clin Microbiol* 38(9): 3165–3173.
- Mahenthalingam E, Coenye T, Chung JW, Speert DP, Govan JR, Taylor P, Vandamme P (2000b) Diagnostically and experimentally useful panel of strains from the *Burkholderia cepacia* complex. *J Clin Microbiol* 38(2): 910–913.
- Matrician L, Ange G, Burns S, Fanning WL, Kioski C, Cage GD, Komatsu KK (2000) Outbreak of nosocomial *Burkholderia cepacia* infection and colonization associated with intrinsically contaminated mouthwash. *Infect Control Hosp Epidemiol* 21(11):739–741.
- Menard A, de Los Santos PE, Graindorge A, Cournoyer B (2007a) Architecture of *Burkholderia cepacia* complex sigma70 gene family: evidence of alternative primary and clade-specific factors, and genomic instability. *BMC Genomics* 8(308).
- Menard A, Monnez C, Estrada de Los Santos P, Segonds C, Caballero-Mellado J, Lipuma JJ, Chabanon G, Cournoyer B (2007b) Selection of nitrogen-fixing deficient *Burkholderia vietnamiensis* strains by cystic fibrosis patients: involvement of *nif* gene deletions and auxotrophic mutations. *Environ Microbiol* 9(5):1176–1185.
- Nasser RM, Rahi AC, Haddad MF, Daoud Z, Irani-Hakime N, Almawi WY (2004) Outbreak of *Burkholderia cepacia* bacteremia traced to contaminated hospital water used for dilution of an alcohol skin antiseptic. *Infect Control Hosp Epidemiol* 25(3):231–239.
- Nelson MJ, Montgomery SO, Mahaffey WR, Pritchard PH (1987) Biodegradation of trichloroethylene and involvement of an aromatic biodegradative pathway. *Appl Environ Microbiol* 53(5):949–954.
- Oie S, Kamiya A (1996a) Bacterial contamination of commercially available ethacridine lactate (acrinol) products. *J Hosp Infect* 34(1):51–58.
- Oie S, Kamiya A (1996b) Microbial contamination of antiseptics and disinfectants. *Am J Infect Control* 24(5):389–395.
- Otag F, Ersoz G, Salcioglu M, Bal C, Schneider I, Bauernfeind A (2005) Nosocomial bloodstream infections with *Burkholderia stabilis*. *J Hosp Infect* 59(1):46–52.
- Parsons YN, Glendinning KJ, Thornton V, Hales BA, Hart CA, Winstanley C (2001) A putative type III secretion gene cluster is widely distributed in the *Burkholderia cepacia* complex but absent from genomovar I. *FEMS Microbiol Lett* 203(1):103–108.
- Pitcher DG, Saunders NA, Owen RJ (1989) Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Lett Appl Microbiol* 8: 154–156.
- Romling U, Fiedler B, Bosshammer J, Grothues D, Greipel J, von der Hardt H, Tummeler B (1994) Epidemiology of chronic *Pseudomonas aeruginosa* infections in cystic fibrosis. *J Infect Dis* 170(6): 1616–1621.
- Romling U, Schmidt KD, Tummeler B (1997) Large genome rearrangements discovered by the detailed analysis of 21 *Pseudomonas aeruginosa* clone C isolates found in environment and disease habitats. *J Mol Biol* 271(3):386–404.
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4(4):406–425.
- Sajjan US, Sun L, Goldstein R, Forstner JF (1995) Cable (cbl) type II pili of cystic fibrosis-associated *Burkholderia (Pseudomonas) cepacia*: nucleotide sequence of the *cblA* major subunit pilin gene and novel morphology of the assembled appendage fibers. *J Bacteriol* 177(4):1030–1038.
- Savey A, Tressière B (2007) Surveillance des infections nosocomiales en réanimation adulte - Réseau REA-Raisin - Résultats 2006. France: Institut de Veille Sanitaire. p. 28.
- Shehabi AA, Abu-Al-Soud W, Mahafzah A, Khuri-Bulos N, Abu Khader I, Ouis I, Wadstrom T (2004) Investigation of *Burkholderia cepacia* nosocomial outbreak with high fatality in patients suffering from diseases other than cystic fibrosis. *Scand J Infect Dis* 36(3):174–178.
- Siddiqui AH, Mulligan ME, Mahenthalingam E, Hebden J, Brewrink J, Qaiyumi S, Johnson JA, LiPuma JJ (2001) An episodic outbreak of genetically related *Burkholderia cepacia* among non-cystic fibrosis patients at a university hospital. *Infect Control Hosp Epidemiol* 22(7): 419–422.
- Souza AV, Moreira CR, Pasternak J, Hirata Mde L, Saltini DA, Caetano V, Ciosak S, Azevedo FM, Severino P, Vandamme P, Magalhaes V (2004) Characterizing uncommon *Burkholderia cepacia* complex isolates from an outbreak in a haemodialysis unit. *J Med Microbiol* 53(Pt 10):999–1005.
- Subsin B, Chambers CE, Visser MB, Sokol PA (2007) Identification of genes regulated by the *cepIR* quorum-sensing system in *Burkholderia cenocepacia* by high-throughput screening of a random promoter library. *J Bacteriol* 189(3):968–979.
- Sun L, Jiang RZ, Steinbach S, Holmes A, Campanelli C, Forstner J, Sajjan U, Tan Y, Riley M, Goldstein R (1995) The emergence of a highly transmissible lineage of cbl+ *Pseudomonas (Burkholderia) cepacia* causing CF centre epidemics in North America and Britain. *Nat Med* 1(7):661–666.
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22):4673–4680.

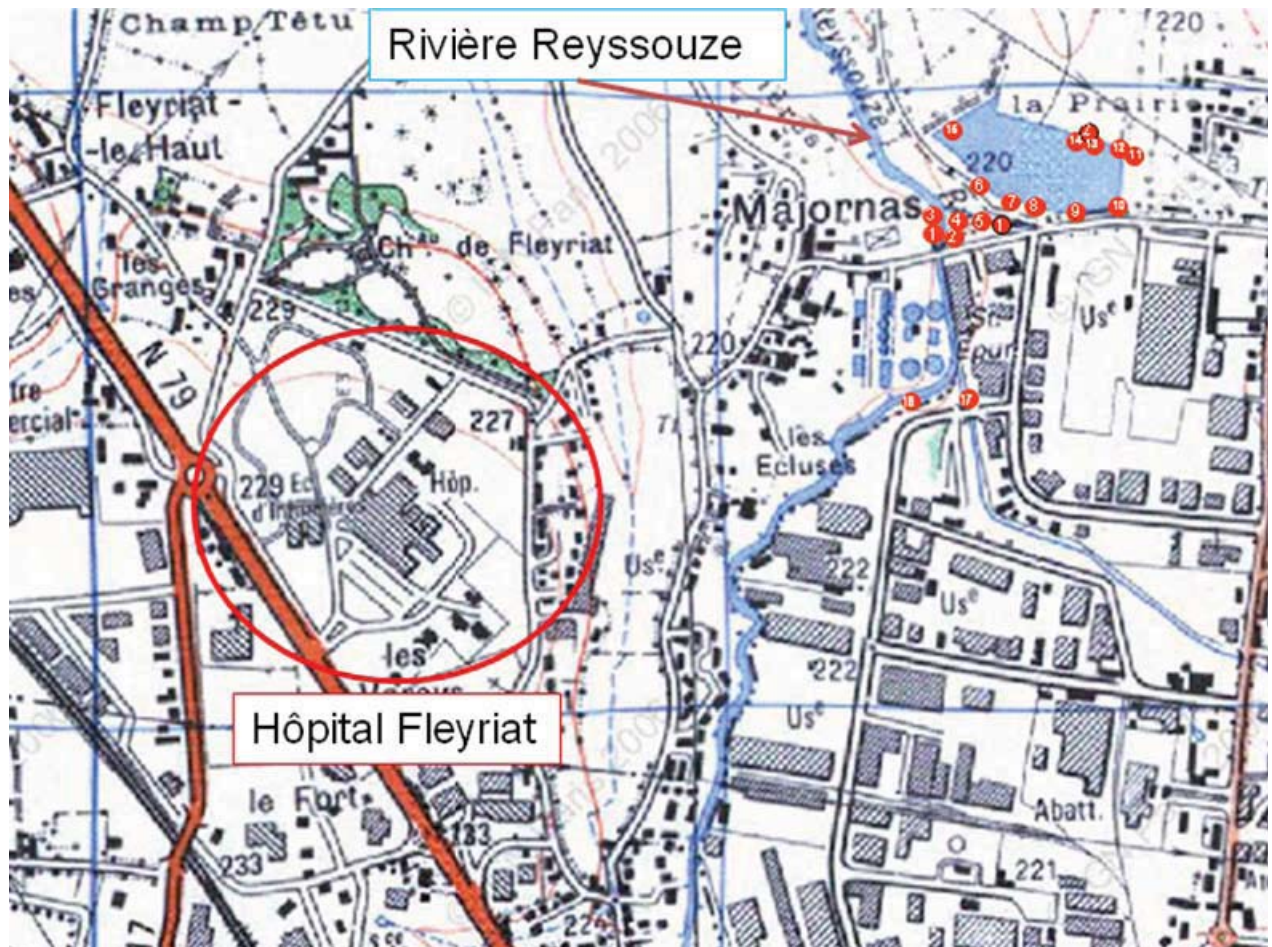
- Tomich M, Mohr CD (2004) Transcriptional and posttranscriptional control of cable pilus gene expression in *Burkholderia cenocepacia*. *J Bacteriol* 186(4):1009–1020.
- Vandamme P, Holmes B, Vancanneyt M, Coenye T, Hoste B, Coopman R, Revets H, Lauwers S, Gillis M, Kersters K, Govan JR (1997) Occurrence of multiple genomovars of *Burkholderia cepacia* in cystic fibrosis patients and proposal of *Burkholderia multivorans* sp. nov. *Int J Syst Bacteriol* 47(4):1188–1200.
- Vandamme P, Henry D, Coenye T, Nzula S, Vancanneyt M, LiPuma JJ, Speert DP, Govan JR, Mahenthiralingam E (2002) *Burkholderia anthina* sp. nov. and *Burkholderia pyrrocinia*, two additional *Burkholderia cepacia* complex bacteria, may confound results of new molecular diagnostic tools. *FEMS Immunol Med Microbiol* 33(2): 143–149.
- Vandamme P, Holmes B, Coenye T, Goris J, Mahenthiralingam E, LiPuma JJ, Govan JR (2003) *Burkholderia cenocepacia* sp. nov.—a new twist to an old story. *Res Microbiol* 154(2):91–96.
- Vermis K, Brachkova M, Vandamme P, Nelis H (2003) Isolation of *Burkholderia cepacia* complex genomovars from waters. *Syst Appl Microbiol* 26(4):595–600.
- Vinon-Dubiel AD, Spilker T, Dean CR, Monteil H, LiPuma JJ, Goldberg JB (2004) Correlation of *wbiI* genotype, serotype, and isolate source within species of the *Burkholderia cepacia* complex. *J Clin Microbiol* 42(9): 4121–4126.
- Vonberg RP, Gastmeier P (2007) Hospital-acquired infections related to contaminated substances. *J Hosp Infect* 65(1):15–23.
- Woods CW, Bressler AM, LiPuma JJ, Alexander BD, Clements DA, Weber DJ, Moore CM, Reller LB, Kaye KS (2004) Virulence associated with outbreak-related strains of *Burkholderia cepacia* complex among a cohort of patients with bacteremia. *Clin Infect Dis* 38(9):1243–1250.
- Yan H, Shi L, Alam MJ, Li L, Yang L, Yamasaki S (2008) Usefulness of Sau-PCR for molecular epidemiology of nosocomial outbreaks due to *Burkholderia cepacia* which occurred in a local hospital in Guangzhou, China. *Microbiol Immunol* 52(5):283–286.

## **Partie B : Description des échantillonnages réalisés au niveau du centre hospitalier.**

Afin de déterminer l'origine du clone B&B et/ou sa source environnementale, deux campagnes ont été réalisées à proximité du centre hospitalier Fleyriat: (A) au niveau du centre hospitalier (service de soin intensif, service de néphrologie, pelouses, eaux usées, lagunes d'épuration, air conditionnée...), (B) au niveau d'eaux provenant d'un bassin et d'une rivière, tous deux en aval de la station d'épuration reliée au CH-Fleyriat (Fig. 2). Ces différents prélèvements ont été traités selon une méthode adaptée du protocole de Vermis, et al. 2003. Afin de sélectionner l'espèce recherchée (*B. cenocepacia*) nous avons testé une dizaine de sources de carbone sur une trentaine de souches du *Bcc* (Tableau 1). Les résultats nous ont permis de démontrer que l'espèce *B. cenocepacia* utilise préférentiellement le L-arabinose et la L-thréonine. Nous avons donc utilisé le milieu sélectif de Vermis et al. (2003) avec comme seules source de carbone et d'azote ces deux composés. Quatre nouveaux géotypes de *Bcc* ont pu être identifiés au niveau du premier échantillonnage à proximité de l'hôpital (Annexe A1). L'analyse phylogénétique a permis de déterminer que ces souches étaient proches de l'espèce *B. cenocepacia* cependant les analyses RFLP *recA* et l'analyse MLST ne nous ont pas permis de conclure quant à leur espèce d'appartenance (Annexe A2 et A3). Ces résultats confirment la difficulté d'identifier des souches de *Burkholderia* avec les techniques classiquement utilisées. Au niveau du service de soin intensif du centre hospitalier, aucune souche n'a été isolée à partir des prélèvements réalisés. Ce résultat démontre l'efficacité de la procédure de décontamination mise en œuvre après la période de l'épidémie. La seconde campagne (B) a été réalisée à la fin du mois de juillet 2006 sur des eaux et des boues d'une rivière (Reyssouze) proche de l'hôpital (Fig. 2). Seize isolats, positifs au niveau des cribles PCR *recA* et *ecfB*, ont été obtenus à partir des échantillons d'eau N°2, N°4, N°8 et de la boue N°1. La situation géographique de ces prélèvements est présentée au niveau de la Figure 2 et les prises de vue des sites présentant des échantillons positifs sont présentées dans la Figure 3. Les analyses PFGE ont permis de déterminer que ces 16 isolats représentaient 5 souches différentes (Bpoe733 pour l'échantillon N°2 et N°8, Bpoe734 et Bpoe735 pour l'échantillon N°4, Bpoe736 et Bpoe737 pour l'échantillon de boue N°1, e. g. publication précédente). Les analyses RFLP *recA*, le séquençage du gène *recA* et l'identification par le Vitek2 compact® ont permis d'identifier ces clones comme appartenant à deux espèces distinctes : *B. vietnamiensis* (pour les souches Bpoe733, Bpoe736 et Bpoe737) et *B. multivorans* (pour les souches Bpoe734 et Bpoe735). Les

souches identifiées comme appartenant à l'espèce *B. vietnamiensis* ont été confirmées à l'aide de la PCR *nifH* développée dans notre équipe par Vlot, et al. 2002. Ces souches représentent les premiers isolats de *B. vietnamiensis* isolés de l'environnement sur le territoire français. Contrairement à l'espèce *B. vietnamiensis*, seulement une dizaine de souches de *B. multivorans* ont été isolées à ce jour, au niveau mondial, à partir de milieux non-hospitaliers. L'isolement de ces souches nous permettra donc d'augmenter le pool de *B. multivorans* provenant de l'environnement. Les travaux réalisés par Baldwin, et al. en janvier 2008 suggèrent que l'environnement serait un réservoir pour cette espèce, l'identification de ces deux souches pourrait nous aider à mieux comprendre l'écologie de cette espèce.

Les Bcc semblent être assez répandues dans les environnements à proximité de l'hôpital cependant leur diversité semble assez faible dans les sols et les eaux prospectés. Le peu de souches obtenues au cours de ces campagnes d'échantillonnage ne nous permet pas de conclure quant à la structure génétique de la population de Bcc mais l'analyse en PFGE de toutes ces souches et la comparaison de ces profils avec celui du clone B&B nous indique que ce clone est absent des échantillons traités.



**Fig. 2: Situation géographique de l'hôpital Fleyriat et lieux d'échantillonnage.**

Les lieux de prélèvements réalisés au niveau des eaux sont indiqués par des points rouges. Les deux prélèvements réalisés au niveau de boues sont représentés par des points rouge entourés de noir.

**Tableau 1** : Sources de carbones utilisées par différentes souches de *Burkholderia*.

souches	espèces	D-Fucose	L-Arabinose	L-Thréonine	D-Cellobiose
LMG 1222	<i>B. cepacia</i>	-	+	-	+/-
LMG 14294	<i>B. stabilis</i>	-	-	-	-
B&B clone	<i>B. cenocepacia</i>	-	+/-	+	+/-
B&B clone	<i>B. cenocepacia</i>	-	+	+/-	-
LMG 19240	<i>B. cenocepacia</i>	-	+	-	-
LMG 19230	<i>B. cenocepacia</i>	-	+	-	-
LMG 21461	<i>B. cenocepacia</i>	-	+	+	-
LMG 18832	<i>B. cenocepacia</i>	-	+	+	+
LMG 18830	<i>B. cenocepacia</i>	-	+	-	-
LMG 16654	<i>B. cenocepacia</i>	-	+	+	-
LMG 18826	<i>B. cenocepacia</i>	-	-	-	-
LMG 18828	<i>B. cenocepacia</i>	-	-	+	-
LMG 16656	<i>B. cenocepacia</i>	-	+	+	-
LMG 18829	<i>B. cenocepacia</i>	-	+	+	+
LMG 13010	<i>B. multivorans</i>	-	+	-	+
LMG 16660	<i>B. multivorans</i>	-	+	+	-
LMG 18822	<i>B. multivorans</i>	-	-	+	+
LMG 16665	<i>B. multivorans</i>	-	+	+	+
LMG 10929	<i>B. vietnamiensis</i>	-	-	+/-	+/-
ATCC 53617	<i>B. vietnamiensis</i>	-	-	-	-
Clo	<i>B. vietnamiensis</i>	-	-	-	-
LMG 16232	<i>B. vietnamiensis</i>	-	+	+/-	-
LMG 16228	<i>B. vietnamiensis</i>	-	-	-	-
LMG 17830	<i>B. vietnamiensis</i>	-	+	+	+
LMG 10823	<i>B. vietnamiensis</i>	-	+/-	+	-
AU 1458	<i>B. vietnamiensis</i>	-	-	+	-
AU 0913	<i>B. vietnamiensis</i>	-	+/-	+	-
AU 1160	<i>B. vietnamiensis</i>	-	-	-	-
AU 0749	<i>B. vietnamiensis</i>	-	-	-	-
AU 0829	<i>B. vietnamiensis</i>	-	+	+	+/-
LMG 18941	<i>B. dolosa</i>	-	-	+	+
LMG 14191	<i>B. pyrrocinia</i>	-	-	+	+

+ : croissance positive au bout de 48h, - : croissance négative au bout de 48h, +/- : croissance très faible (<0.3 DO) au bout de 48h. Utilisation d'un milieu modifié de Vermis, et al. 2003



**Врое̄ 33**



**Врое̄ 34, Врое̄ 35**



**Врое̄ 33**



**Врое̄ 36, Врое̄ 37**

Fig. 3: Prises de vue des différents prélèvements positifs obtenus au cours des échantillonnages. Les noms des souches obtenues au cours de l'échantillonnage sont indiqués sous chaque photo. Les numéros indiqués sur les photos correspondent aux numéros des échantillons de la figure 2

## Conclusions du chapitre

L'épidémie nosocomiale survenue au sein du service de soins intensifs de l'hôpital Fleyriat a démontré la haute pathogénie et haute transmissibilité du clone B&B de l'espèce *B. cenocepacia*. Les études moléculaires ont permis de détecter de nombreux facteurs génétiques pouvant être liés à la virulence de ce clone (gènes de résistance aux antibiotiques, gènes du quorum sensing, flagelles etc.). Les analyses réalisées par phylogénie moléculaire ont démontré une proximité du clone B&B avec le clone épidémique ET12 responsable de pneumonies nécrotiques chez les patients CF. L'analyse MLST a mis en évidence une répartition mondiale du clone B&B avec une prévalence de celui-ci au niveau des patients CF français. A l'heure actuelle, une cinquantaine de génomes sont disponibles au sein du genre *Burkholderia*, dont quatre appartenant à l'espèce *B. cenocepacia*. Deux d'entre eux représentent la lignée PHDC (souches HI2424 et AU1054), un, la lignée ET12 (souche J2315) et le dernier, la lignée Midwest (souche PC184). Ces génomes disponibles ont été utilisés comme outils d'analyse comparative, dans le but d'identifier les configurations génétiques originales impliqués dans l'émergence du clone B&B. La virulence étant un processus complexe, il est nécessaire de considérer tant les gènes directement impliqués dans celle-ci que leurs régulateurs. Les facteurs sigma70 ( $\sigma_{70}$ ) jouant un rôle majeur dans l'expression des gènes en fonction des conditions environnementales, un bilan de ces gènes au sein des génomes séquencés ainsi que chez le clone B&B a été réalisé. L'étude de la distribution des  $\sigma_{70}$  et les phénomènes de réarrangements génétiques expliquant cette distribution inégale ont été déterminés. Les résultats obtenus ainsi que la démarche expérimentale mise en œuvre sont présentés dans le chapitre suivant. Certains des réarrangements génétiques observés ont pu être reliés à la présence d'éléments génétiques répétés de la famille des séquences d'insertion (IS) et à certains phénomènes d'acquisition d'éléments génétiques mobiles de type îlot génomique (IG). L'implication de ces deux phénomènes dans l'émergence du clone B&B a ensuite été étudiée dans les chapitres suivants.



# Description de facteurs transcriptionnels alternatifs primaires et de remaniements génomiques des gènes codant des facteurs ECF au sein de la famille $\sigma^{70}$ des *Burkholderia* du complexe *cepacia*

## 1. Préambule :

Ce chapitre a été le fruit de travaux collaboratifs sous la direction de Benoit Cournoyer impliquant une post-doctorante, Paulina Estrada de los Santos et un étudiant en formation doctorale, Aymeric Menard. Cette étude a permis d'effectuer un travail fondateur sur la problématique des facteurs transcriptionnels sigma70 ( $\sigma^{70}$ ) chez les Bcc et a donné lieu à une publication dans la revue *BMC Genomics* en septembre 2007. Ce travail a été utilisé comme référence pour entreprendre des études comparatives de la structure génétique des  $\sigma^{70}$  au sein du Bcc dont les principaux clones de l'espèce *B. cenocepacia* et du clone B&B.

Au sein d'une même espèce, de nombreux caractères phénotypiques sont soumis à variation. Cette diversité phénotypique peut avoir plusieurs origines, elle peut impliquer des insertions d'éléments génétiques mobiles comme nous l'aborderons dans le chapitre suivant de ce travail de thèse mais peut également impliquer le gain ou la perte de gènes (facteurs de transcription, inhibiteurs) conduisant à un changement dans le profil d'expression des gènes impliqués. La structure génétique des familles de gènes régulateurs peut donner un aperçu des potentiels d'adaptation des microorganismes à diverses conditions de croissance. Une bactérie ayant une grande complexité génétique pour de telles familles est plus versatile et peut moduler plus rapidement l'expression de son patrimoine génétique. Les principales familles de régulateurs sont impliquées dans les étapes de transcription des déterminants génétiques e. g. système à deux composants,  $\sigma^{70}$ .

La transcription bactérienne est réalisée par l'ARN polymérase ADN dépendante constituée de deux sous unités  $\alpha$ , une  $\beta$ , une  $\beta'$ . Cependant, l'initiation de la transcription requière une cinquième unité, le facteur sigma ( $\sigma$ ), qui se lie à cet ensemble pour former l'holoenzyme. Les facteurs sigma sont composés de différents domaines protéiques [1] leur permettant d'interagir avec la polymérase, les séquences promotrices ou d'autres facteurs régulant la transcription (Fig. A). Les facteurs sigma permettent la reconnaissance spécifique des différents types de promoteurs au sein d'un génome. Ils sont divisés en deux grandes familles, celle des sigma 54 ( $\sigma^{54}$ ) et celle des sigma 70 ( $\sigma^{70}$ ). Les  $\sigma^{54}$  ont une fonction similaire aux  $\sigma^{70}$  mais une structure différente. Les  $\sigma^{54}$  montrent une faible diversité génétique. Leur

distribution est restreinte à certains groupes bactériens et il y a rarement plus d'une copie de ce gène par génome. Chez les  $\sigma^{70}$ , nous pouvons retrouver jusqu'à une soixantaine de paralogues par génome bactérien. L'abondance relative des différents facteurs sigma au sein d'une cellule, leur affinité pour l'ARN polymérase et leur disponibilité (liaison à une protéine bloquant leur action : l'anti sigma) déterminent les groupes de gènes pour lesquels la transcription sera possible. L'ensemble des gènes régulés par un facteur sigma donné forme un sigmulon.

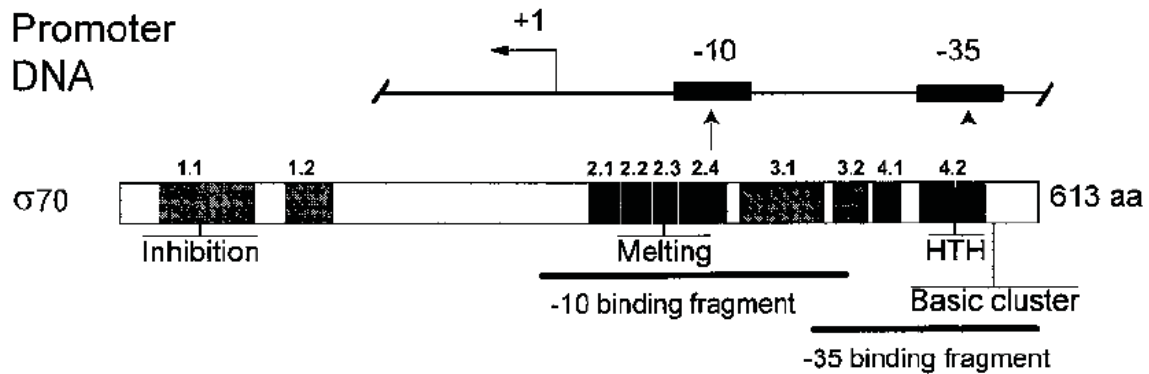


Fig. A: Régions conservées des  $\sigma^{70}$  et fonctions liées à la liaison à l'ADN. La région 1.1 inhibe la fixation du  $\sigma^{70}$  à l'ADN, cependant d'autres régions du facteur  $\sigma^{70}$  se lient aux boîtes -10 et -35 du promoteur. D'après Lonetto et al. (1998) [1].

Sur la base d'études phylogénétique et fonctionnelles les facteurs  $\sigma^{70}$  ont été classés en quatre groupes fonctionnels. Suivant les auteurs les frontières de ces groupes varient. Lonetto *et al.* [2] proposent de regrouper ensemble le facteur nécessaire pour la croissance bactérienne ou facteur primaire et les facteurs primaires alternatifs présents chez les actinobactéries qui viennent sous certaines conditions remplacer le facteur primaire. Dans un deuxième groupe sont regroupés les facteurs similaires au facteur RpoS d'*E. coli* notamment impliqués dans la vie en phase stationnaire. Le troisième groupe comprend les facteurs liés à la sporulation, la synthèse des flagelles et la réponse aux chocs thermiques. Enfin, le dernier groupe, de loin le plus important en nombre, comprend l'ensemble des facteurs impliqués dans les fonctions extracytoplasmiques (ECF). Les fonctions liées à ces facteurs sont très diverses et incluent notamment l'expression de facteurs de virulences [3, 4].

Les eubactéries possèdent au moins un facteur sigma, le facteur primaire (à l'exception de *Deinococcus radiodurans*, [5]). Cependant, les autres facteurs ne sont pas systématiquement présents. Le nombre de facteurs au sein d'un même génome est très variable pouvant aller jusqu'à 63 dans le cas de *S. coelicolor* [6]. Les études portant sur un grand nombre de bactéries supposent qu'il existe une relation entre le nombre de facteurs sigma présents dans un génome

et l'écologie des bactéries considérées. Les bactéries pathogènes possèderaient moins de facteurs sigma que les bactéries libres. L'hypothèse est qu'une bactérie pathogène est moins soumise à des modifications des contraintes environnementales qu'une bactérie libre et qu'elle possèderait ainsi moins de facteurs pouvant moduler de façon globale l'expression de ses gènes. Il s'agit cependant d'une tendance, aucun seuil n'a pu être fixé [5].

Une variabilité génétique au sein de la famille des facteurs  $\sigma^{70}$  peut s'observer au sein même d'une espèce [7]. Les facteurs ECF peuvent ainsi être des marqueurs d'une évolution récente des bactéries soumises à d'importantes variations de forces sélectives liées aux hôtes et environnements colonisés. Les Bcc pourraient donc avoir une structure génétique de cette famille qui serait variable et représentative des pressions rencontrées.

En introduction, nous indiquions qu'au sein du genre *Burkholderia*, près de 53 génomes ont été séquencés. Ceci offre la possibilité d'identifier les facteurs  $\sigma^{70}$  présents au sein de ce genre et de mener une étude de leur distribution au sein des espèces. Nous avons ainsi pu évaluer la diversité de cette famille et apporter des éléments concernant sa plasticité chez les Bcc. Ces travaux ont permis d'identifier les  $\sigma^{70}$  liés spécifiquement au Bcc ou au genre *Burkholderia*. En comparaison avec d'autres agents pathogènes, notamment *Pseudomonas aeruginosa*, cette étude a permis d'identifier des facteurs sigma potentiellement impliqués dans la virulence chez les Bcc. Un nouveau marqueur, *ecfB*, décrit dans ce chapitre a été utilisé dans le chapitre précédent pour étudier la phylogénie du clone B&B [8]. Ces travaux constituent une première étape vers l'étude des sigmulons et de leur rôle dans l'écologie des Bcc. L'article qui suit présente la structure de la famille  $\sigma^{70}$  chez les Bcc et certains aspects de son évolution. Il est suivi d'un travail complémentaire principalement dédié à la structure de cette famille chez le clone B&B. Dans cette section, certaines hypothèses émises dans l'article *BMC Genomics* ont été testées.

Research article BioMed Central

## Architecture of *Burkholderia cepacia* complex $\sigma^{70}$ gene family: evidence of alternative primary and clade-specific factors, and genomic instability

Aymeric Menard<sup>1,2</sup>, Paulina Estrada de los Santos<sup>1,2</sup>, Arnault Graindorge<sup>1,2</sup> and Benoit Cournoyer\*<sup>1,2,3</sup>

Address: <sup>1</sup>Université de Lyon, Lyon, France, <sup>2</sup>Research group on «Bacterial Opportunistic Pathogens and Environment», UMR5557 Ecologie Microbienne, Université Lyon 1, CNRS, and Ecole Nationale Vétérinaire de Lyon, France and <sup>3</sup>UMR CNRS 5557 Ecologie Microbienne, Mendel Bldg., 5th floor, Université Lyon 1, 69622 Villeurbanne Cedex, France

Email: Aymeric Menard -menard@biomserv.univ-lyon1.fr; Paulina Estrada de los Santos -Paulina.Estrada.De.Los.Santos@vub.ac.be; Arnault Graindorge -graindorge.umr5557@gmail.com; Benoit Cournoyer\* -cournoye@biomserv.univ-lyon1.fr

\* Corresponding author

Published: 4 September 2007

Received: 1 March 2007 Accepted: 4 September 2007

BMC Genomics 2007, 8:308 doi:10.1186/1471-2164-8-308

This article is available from: <http://www.biomedcentral.com/1471-2164/8/308>

© 2007 Menard et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** The *Burkholderia cepacia* complex (Bcc) groups bacterial species with beneficial properties that can improve crop yields or remediate polluted sites but can also lead to dramatic human clinical outcomes among cystic fibrosis (CF) or immuno-compromised individuals. Genome-wide regulatory processes of gene expression could explain parts of this bacterial duality. Transcriptional  $\sigma^{70}$  factors are components of these processes. They allow the reversible binding of the DNA-dependent RNA polymerase to form the holoenzyme that will lead to mRNA synthesis from a DNA promoter region. Bcc genome-wide analyses were performed to investigate the major evolutionary trends taking place in the  $\sigma^{70}$  family of these bacteria.

**Results:** Twenty  $\sigma^{70}$  paralogous genes were detected in the *Burkholderia cenocepacia* strain J2315 (*Bcen*-J2315) genome, of which 14 were of the ECF (extracytoplasmic function) group. Non-ECF paralogs were related to primary (*rpoD*), alternative primary, stationary phase (*rpoS*), flagellin biosynthesis (*fliA*), and heat shock (*rpoH*) factors. The number of  $\sigma^{70}$  genetic determinants among this genome was of 2,48 per Mb. This number is lower than the one of *Pseudomonas aeruginosa*, a species found in similar habitats including CF lungs. These two bacterial groups showed strikingly different  $\sigma^{70}$  family architectures, with only three ECF paralogs in common (*fecI*-like, *pvdS* and *algU*). *Bcen*-J2315  $\sigma^{70}$  paralogs showed clade-specific distributions. Some paralogs appeared limited to the ET12 epidemic clone (*ecfA2*), particular Bcc species (*sigI*), the *Burkholderia* genus (*ecfJ*, *ecfF*, and *sigJ*), certain proteobacterial groups (*ecfA1*, *ecfC*, *ecfD*, *ecfE*, *ecfG*, *ecfL*, *ecfM* and *rpoS*), or were broadly distributed in the eubacteria (*ecfI*, *ecfK*, *ecfH*, *ecfB*, and *rpoD*-, *rpoH*-, *fliA*-like genes). Genomic instability of this gene family was driven by chromosomal inversion (*ecfA2*), recent duplication events (*ecfA* and *RpoD*), localized (*ecfG*) and large scale deletions (*sigI*, *sigJ*, *ecfC*, *ecfH*, and *ecfK*), and a phage integration event (*ecfE*).

**Conclusion:** The Bcc  $\sigma^{70}$  gene family was found to be under strong selective pressures that could lead to acquisition/deletion, and duplication events modifying its architecture. Comparative analysis of Bcc and *Pseudomonas aeruginosa*  $\sigma^{70}$  gene families revealed distinct evolutionary strategies, with the Bcc having selected several alternative primary factors, something not recorded among *P. aeruginosa* and only previously reported to occur among the actinobacteria.

## 1. Background

*Burkholderia cepacia* complex (Bcc) bacterial species are  $\beta$ -proteobacteria that can be found in various environments including freshwater and plant rhizosphere. They are also the etiological agents of several human infections. This bacterial complex includes several closely related species or genomovars of *Burkholderia* that can be isolated from cystic fibrosis (CF) patients. These species can also be found in nosocomial infections. *B. cenocepacia* (*Bcen*) is a species of the Bcc that can lead to the "cepacia syndrome", a dramatic necrotizing pneumonia (for review see [1]). Three major *B. cenocepacia* epidemic clones have been described around the world: ET12, PHDC and Midwest [2-4]. Three strains of these clonal complexes (one of ET12 and two of PHDC) were sequenced by the Sanger Institute and the Joint Genome Institute, and are available on their respective web site. Several other Bcc and *Burkholderia* strains were sequenced or are in progress of getting sequenced (for a list see [5]).

The ability at colonizing several biotopes is a character observed in the Bcc but also among other bacterial species like *Pseudomonas aeruginosa* of the  $\gamma$ -proteobacteria. This versatility has been linked to an important diversity of regulatory processes including those of the  $\sigma^{70}$  gene family [6]. Transcriptional  $\sigma^{70}$  factors can bind reversibly to eubacterial DNA-dependent RNA polymerase (RNAP) to form the holoenzyme, and enable its promoter-specific attachment. This holoenzyme can drive synthesis of mRNA (but the process could be regulated by several transcriptional repressors [7] and activators [8]). Through evolution, selective pressures seem to have favoured the emergence of several paralogous lineages of  $\sigma^{70}$ , and these factors became a central component of bacterial adaptability to changing environments including host colonization. This resulted in the observation of some sort of relation between the lifestyle of bacteria and the number of genes encoding  $\sigma^{70}$  factors in their genome. For example, 65  $\sigma^{70}$  gene sequences were detected among the genome of *Streptomyces coelicolor* [9] which is ubiquitous and morphologically complex but only one in *Mycoplasma* spp. which are host- and disease-specific [10,11]. Among *Pseudomonas aeruginosa*, a species sharing several niches with the Bcc, 23  $\sigma^{70}$  gene sequences have been reported [6]. Generally speaking,  $\sigma^{70}$  factors can be divided into four groups according to sequence similarities and promoter-recognition specificities [12,13]: (1) the essential primary sigma factors, involved mainly in exponential growth, and the alternative actinomycetal sigma factors; (2) the stationary phase factors; (3) the factors involved in sporangium development, flagellin synthesis and heat shock response, and (4) the extracytoplasmic function sigma factors (termed ECF). All these factors are in competition for the same RNAP core enzyme [14]. Genes under the control of a same sigma factor are termed to be part of a same sigmulon.

Here, we report the first global analysis of the  $\sigma^{70}$  gene family among the *Burkholderia cepacia* complex. The first fully sequenced Bcc genome of *B. cenocepacia* (strain J2315 of the ET12 epidemic clone) was used as the reference sequence for this study.

## 2. Results and Discussion

### 2.1. Detection and number of $\sigma^{70}$ gene sequences in Bcen-J2315 genome

$\sigma^{70}$  TBLASTN searches of *B. cenocepacia* strain J2315 (*Bcen*-J2315) genome sequence were performed. First, searches making use of ten  $\sigma^{70}$  factors of the main groups of this family were carried out. Then, a second round of TBLASTN searches using all *B. cenocepacia* factors detected in the initial screening was done. This led to the identification of twenty  $\sigma^{70}$  gene sequences, thirteen on chromosome 1, seven on chromosome 2, and none on chromosome 3 (Table 1). All genomic regions harbouring  $\sigma^{70}$  coding sequences (CDS) were analysed by the testcode programme, to identify their corresponding ORF. In the case of the presence of multiple initiation codons, the ORF showing the longest CDS was selected. All the defined  $\sigma^{70}$  gene sequences and their position on the *Bcen*-J2315 genome are listed in Table 1. Given the genome size of *Bcen*-J2315 (8,056 Mb), an average of 2,48  $\sigma^{70}$  gene sequences per Mb was computed. This number of sigma genes is slightly lower than the number reported in *P. aeruginosa* (23  $\sigma^{70}$  in strain PA01, genome size of 6,26 Mb) for which an average of 3,67  $\sigma^{70}$  genes per Mb was calculated. It is also lower than the one of the genome of *Streptomyces coelicolor* strain A3(2) (8,67 Mb) which has a similar genome size. However, *S. coelicolor* has a very distinct morphology (being hyphal and producing spores) and ecology. This simplistic analysis suggests that bacteria like the Bcc and *P. aeruginosa*, which are sharing several habitats, have similar number of  $\sigma^{70}$  gene sequences. However, the composition of these families was found to be quite different (as described in the next sections).

### 2.2. Classification of $\sigma^{70}$ gene sequences of Bcen-J2315 into functional categories

Translated amino acid (AA) sequences from *Bcen*-J2315  $\sigma^{70}$  coding sequences were positioned among the main eubacterial  $\sigma^{70}$  lineages, using a decisional  $\sigma^{70}$  Neighbor-Joining (NJ) phylogenetic tree limited to key sequences (Fig. 1b). This phylogenetic tree divides the  $\sigma^{70}$  gene family into four groups (see [13]): (1) one grouping the primary (mainly RpoD) and alternative primary actinomycetal factors, (2) one grouping the stationary phase factors (mainly named RpoS), (3) one the heat shock (RpoH), flagellin (FliA), and cellular differentiation (WhiG) factors, and (4) one grouping factors involved in extracytoplasmic functions (named ECF factors). All these groups are supported by high bootstrap values. This phylogenetic tree allowed to allocate 14 of the *Bcen*-J2315  $\sigma^{70}$  gene sequences into the ECF (see Table 1). The ECF group is often the most important in number, and is the main group involved in bacterial responses toward changing environmental conditions. ECF sequences represent 66% of the  $\sigma^{70}$  gene sequences of the *Bcen*-J2315 genome, a number which is lower than the one observed in *P. aeruginosa* (82%) [6,15]. Two deduced AA sequences, EcfA1 and EcfA2, are 100% identical in *Bcen*-J2315, and their respective genes are positioned on two identical DNA strands of 40 kb that are diametrically opposed on chromosome 1 (see section on *ecfA* for further details). For the remaining 6 factors, three were allocated to group 1 (SigE, SigI and SigJ), one to group 2 (SigK), and two to group 3 (SigL and SigM). *Bcen*-J2315 deduced  $\sigma^{70}$  factors conserved domains were revealed by CDD analyses (Fig. 2). Four conserved domains (regions 1 to 4) were described [12] among the  $\sigma^{70}$  family, with region 1 being sub-divided into R1\_1 and R1\_2. *Bcen*-J2315 SigE, I and J contain all four regions. Regions 2 and 4 were identified in

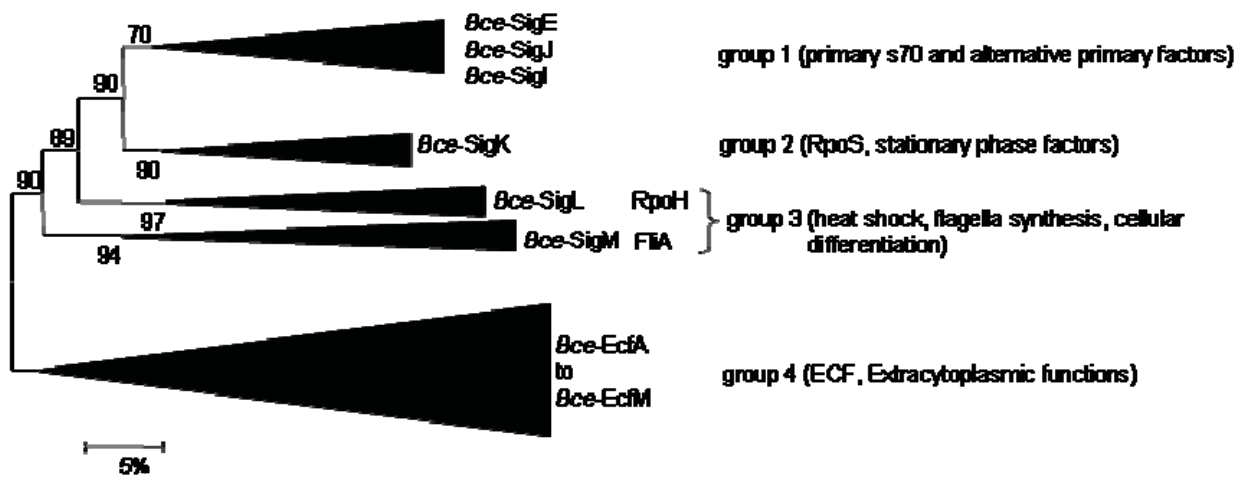
*Bcen*-J2315 deduced SigL, SigM and SigK sequences but only a full region 3 was found in SigM and SigK. Sub region 1\_1 was detected in SigK. Only regions 2 and 4 were detected among the ECF factors.

**Table 1.** Distribution of *B. cenocepacia* J2315 strain  $\sigma^{70}$  factors among other *Burkholderia* genome sequences and *P. aeruginosa*.

<i>Bcen</i> $\sigma^{70}$ factor	associated name	cds position in J2315 genome	locus tag	<i>Bcc</i>				<i>pseudomallei</i> lineage							
				<i>B. cenocepacia</i>	<i>B. vietnamiensis</i>	<i>B. sp. 383</i>	<i>B. ambifaria</i>	<i>B. dolosa</i>	<i>B. mallei</i>	<i>B. pseudomallei</i>	<i>B. thailandensis</i>	<i>B. xenovorans</i>	<i>P. aeruginosa</i>		
<i>Bcen</i> -SigE	RpoD	chr.2 1009395*-1011812	BCAM0918	4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	1/1	2/2
<i>Bcen</i> -SigI		chr.2 2666931-2668817	BCAM2371	4/4	1/1	0/1	1/1	1/1	1/1	0/4	0/6	0/1	0/1	0/1	0/2
<i>Bcen</i> -SigJ		chr.2 1383611-1382085	BCAM1259	3/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	0/1	0/2
<i>Bcen</i> -SigL	RpoH-like	chr.1 2088369-2089457	BCAL1892	4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	1/1	2/2
<i>Bcen</i> -SigM	Flia-like	chr.1 860723-859788	BCAL0787	4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	1/1	2/2
<i>Bcen</i> -SigK	RpoS	chr.1 168141-168872	BCAL0144	4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	1/1	2/2
<i>Bcen</i> -Ecfa1	RpoE	chr.1 1085335-1085934	BCAL0998	4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	1/1	2/2
<i>Bcen</i> -Ecfa2	RpoE	chr.1 153677-153078	BCAL2872	1/4	0/1	0/1	0/1	0/1	0/1	0/4	0/6	0/1	0/1	0/1	0/2
<i>Bcen</i> -Ecfb	PrtI-like	chr.1 3817661-3818182	BCAL3478	4/4	1/1	1/1	1/1	1/1	1/1	0/4	6/6	1/1	1/1	0/1	0/2
<i>Bcen</i> -EcfC	FecI-like	chr.1 1506330-1506836	BCAL1369	4/4	0/1	1/1	0/1	0/1	0/1	4/4	6/6	1/1	1/1	0/1	2/2
<i>Bcen</i> -EcfD		chr.1 2615164-2615727	BCAL2360	4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	1/1	0/2
<i>Bcen</i> -EcfE		chr.1 2728250-2727696	BCAL2462	4/4	0/1	1/1	0/1	0/1	1/1	0/4	0/6	0/1	1/1	1/1	0/2
★ <i>Bcen</i> -EcfF		chr.1 3441702-3441187	BCAL3151	4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	1/1	0/2
<i>Bcen</i> -EcfG		chr.1 2146519*-2145818	BCAL1947	4/4	0/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	1/1	0/2
<i>Bcen</i> -EcfH		chr.2 1846346-1845561	BCAM1661	4/4	0/1	1/1	1/1	1/1	0/1	0/4	0/6	0/1	1/1	1/1	0/2
<i>Bcen</i> -EcfI	OrbS	chr.1 1846454-1847122	BCAL1688	4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	1/1	2/2 •
★ <i>Bcen</i> -EcfJ		chr.2 230-829	BCAM0001	4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	0/1	0/2
<i>Bcen</i> -EcfK		chr.2 3109088-3109981	BCAM2748	4/4	0/1	1/1	0/1	0/1	1/1	4/4	6/6	1/1	1/1	0/1	0/2
<i>Bcen</i> -EcfL		chr.2 934738-936006	BCAM0849	4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	1/1	0/2
<i>Bcen</i> -EcfM		chr.1 3824313-3823432	BCAL3486	4/4	1/1	1/1	1/1	1/1	1/1	0/4	0/6	0/1	0/1	0/1	0/2

TBLASTN searches of *Bce*-J2315  $\sigma^{70}$  factors were performed on fully sequenced *Burkholderia* genomes. In accordance with phylogenetic inferences shown in Fig. 3 and 4, TBLASTN were considered to reveal homologs or orthologs of a same  $\sigma^{70}$  lineage when showing identity values above 50% and similarity higher than 63%, and when being part of a phylogenetic cluster supported by high bootstrap values. • *P. aeruginosa* OrbS homologs, named PvdS, showed 38% identity and 56% similarity with *Bcen*-J2315 homolog. Sequences were retrieved from [50-53]. Genomes of the following strains were analysed: *B. cenocepacia*, strains J2315, AU1054, HI2424, PC184 (negative for *Bcen*-SigJ); *B. vietnamiensis*, G4; *Burkholderia* sp., 383 ; *B. ambifaria*, AMMD, *B. dolosa*, AU0158; *B. xenovorans*, LB400; *B. mallei*, ATCC23344, 10229, GB8 horse 4, SAVP1; *B. pseudomallei*, 1710a, 1710b, S13, 1655, 668, K96243; *B. thailandensis*, E264; *P. aeruginosa*, PAO1, PA14. \* indicates an ORF different from the one of the annotated genome. ★ Distribution limited to *Burkholderia* genomes, see Fig. 4.



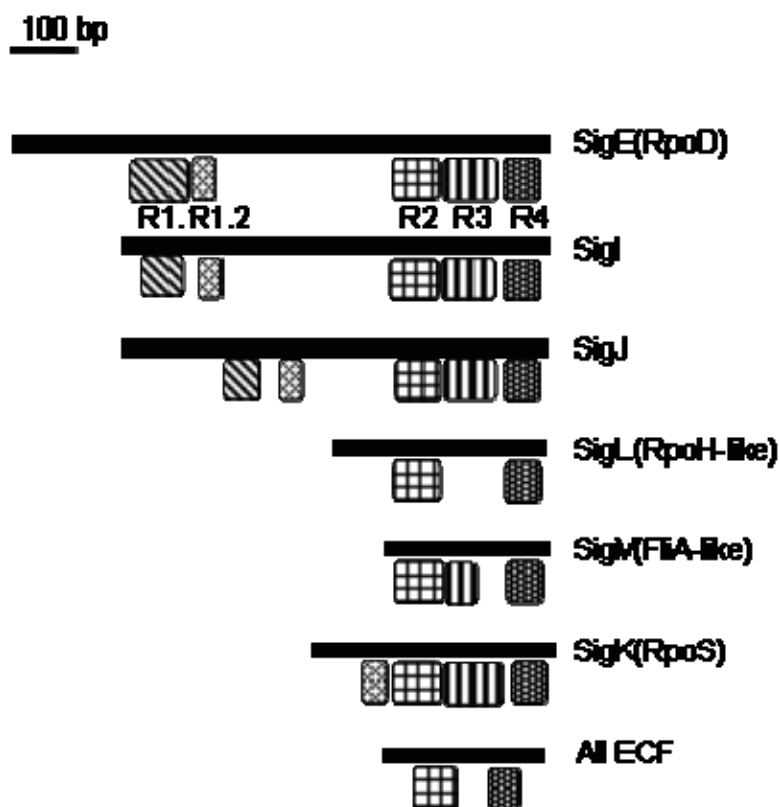


**Figure 1b:** Allocation of *B. cenocepacia* strain J2315  $\sigma^{70}$  deduced factors into the main  $\sigma^{70}$  gene family phylogenetic groups. All *Bcen*-J2315 deduced amino acids sequences from  $\sigma^{70}$  coding sequences were added to the frame of the Lavire et al. [13] multiple alignment. A Neighbor-joining phylogenetic tree was computed from this multiple alignment. Distances on this tree are proportional to evolutionary divergences expressed in substitutions per 100 sites. The *Bcen*-J2315 sequences were allocated to one of the following phylogenetic groups: group 1, the essential primary factors, and alternative primary factors of the actinobacteria; group 2, stationary phase factors of the RpoS type; group 3, factors involved in heat shock response, flagella synthesis or cellular differentiation, and group 4, factors involved in extracytoplasmic functions (ECF).

### 2.3. Molecular phylogeny of *Bcen*-J2315 primary, stationary, and stress response/cellular differentiation/flagellin synthesis $\sigma^{70}$ gene sequences

The NJ tree of Fig. 1 could not be used to assess accurately the phylogenetic relationships among each of the main  $\sigma^{70}$  phylogenetic groups. The phylogenetic resolution of the  $\sigma^{70}$  family appeared to be more robustly determined by splitting the dataset into two subsets corresponding to ECF and non-ECF (primary, stationary, and stress response/cellular differentiation) deduced  $\sigma^{70}$  AA sequences.

Reliable phylogenetic positioning of *Bcen*-J2315 deduced  $\sigma^{70}$  factors was performed by first extracting all AA sequences in the databases showing significant similarities with these factors. These sequences were then added to the multiple alignment obtained by Lavire et al. [13], and a NJ phylogenetic tree was inferred (Fig. 3). Names of lineages used in Lavire et al. [13] were kept. The NJ tree of non-ECF AA sequences is divided into three major clusters of paralogs, matching the non-ECF categories described in the above section (groups 1, 2 and 3). All these paralogous  $\sigma^{70}$  sequences appear to have emerged early in the evolution of the eubacteria. The SigE/RpoD (group 1), SigM/FliA and SigL/RpoH homologs (group 3) were detected among at least the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -proteobacteria, suggesting an emergence prior to the diversification of these bacterial groups. However, the group 2 factors (RpoS) were found restricted to  $\beta$ -proteobacteria and  $\gamma$ -proteobacteria.



**Figure 2:** Conserved regions among *B. cenocepacia* strain J2315  $\sigma^{70}$  factors. These regions were detected by CDD (Conserved Domain Detection [42]). R1.1: region 1.1 can modulate or prevent DNA binding by regions 2 and 4 when the sigma is unbound by the core RNA polymerase; R2: region 2 contains both the -10 promoter recognition helix and the primary core RNA polymerase binding domain. R3: region 3 is primarily involved in binding to the core RNA polymerase. R4: region 4 is involved in binding to the -35 promoter region.

### 2.3.1. The primary and alternative primary $\sigma^{70}$ factors phylogenetic radiation (group 1)

In the NJ tree of non-ECF  $\sigma^{70}$  factors (Fig. 3), the group of primary and alternative primary factors was found divided into 7 distinct and significant lineages: C, D and E corresponding to the essential eubacterial primary factors, G and H being representatives of the alternative primary factors lineages of the actinobacteria, and I and J being two novel lineages. *Bcen*-J2315 SigE and its *Burkholderia* orthologs are found to be part of cluster E, the RpoD lineage of essential sigma factors of the proteobacteria, and were thus renamed RpoD. A *B. dolosa* strain AU0158 RpoD-like sequence appeared distantly related to other *Burkholderia* deduced RpoD but a closer look at its gene structure showed a frameshift probably due to a sequencing mistake. An additional copy of the *rpoD* gene was detected in the *B. xenovorans* genome (Fig. 3). This copy shows a deduced AA sequence significantly different from the ones of the *sigE/rpoD Burkholderia* orthologs, suggesting either acquisition by an horizontal transfer event or an ancient duplication event prior to the diversification of the  $\beta$ -proteobacteria. *Bcen*-J2315 SigI and SigJ deduced sequences and their orthologs revealed two novel lineages of group 1 factors which are restricted to the *Burkholderia* genomes (Fig. 3), and were thus termed alternative *Burkholderia* primary factors. The *sigJ* coding sequence was not detected in *B. xenovorans* and *B. cenocepacia* PC184 genomes (Table 2 and see section on genomic rearrangements). However, *sigJ* DNA probings of Bcc genomic DNA blots showed a broad

distribution among the Bcc (including *B. cepacia*, *B. cenocepacia*, *B. stabilis*, *B. dolosa*, *B. vietnamiensis*, *B. multivorans* and *B. pyrrocinia*) (Table 2). *sigI* was not detected in the genomes of the Bcc strain 383, of *B. xenovorans*, and of *B. pseudomallei* and its closely related genomes (see section on genomic rearrangements). Nevertheless, Bcc DNA blot analyses using *sigI* DNA probes showed this gene largely distributed in the Bcc (Table 2). In *B. xenovorans* LB400 genome, *sigI* and *sigJ* were not detected but another lineage of alternative *Burkholderia* primary factors was found. The presence of coding sequences of such factors can thus be considered a common feature of *Burkholderia* genomes. Alternative primary factors were never detected in *P. aeruginosa*. In fact, this is the first report demonstrating the presence of alternative primary factors among bacteria outside the actinomycetes. The function of these alternative primary factors is unknown but they possess all the domains found among the RpoD factors i. e. -10 and -35 promoter recognition domains, RNA polymerase binding domains, etc (Fig. 2). These factors could be replacing the primary one under growth conditions inhibiting its expression or functioning (potentially certain antibiotics).

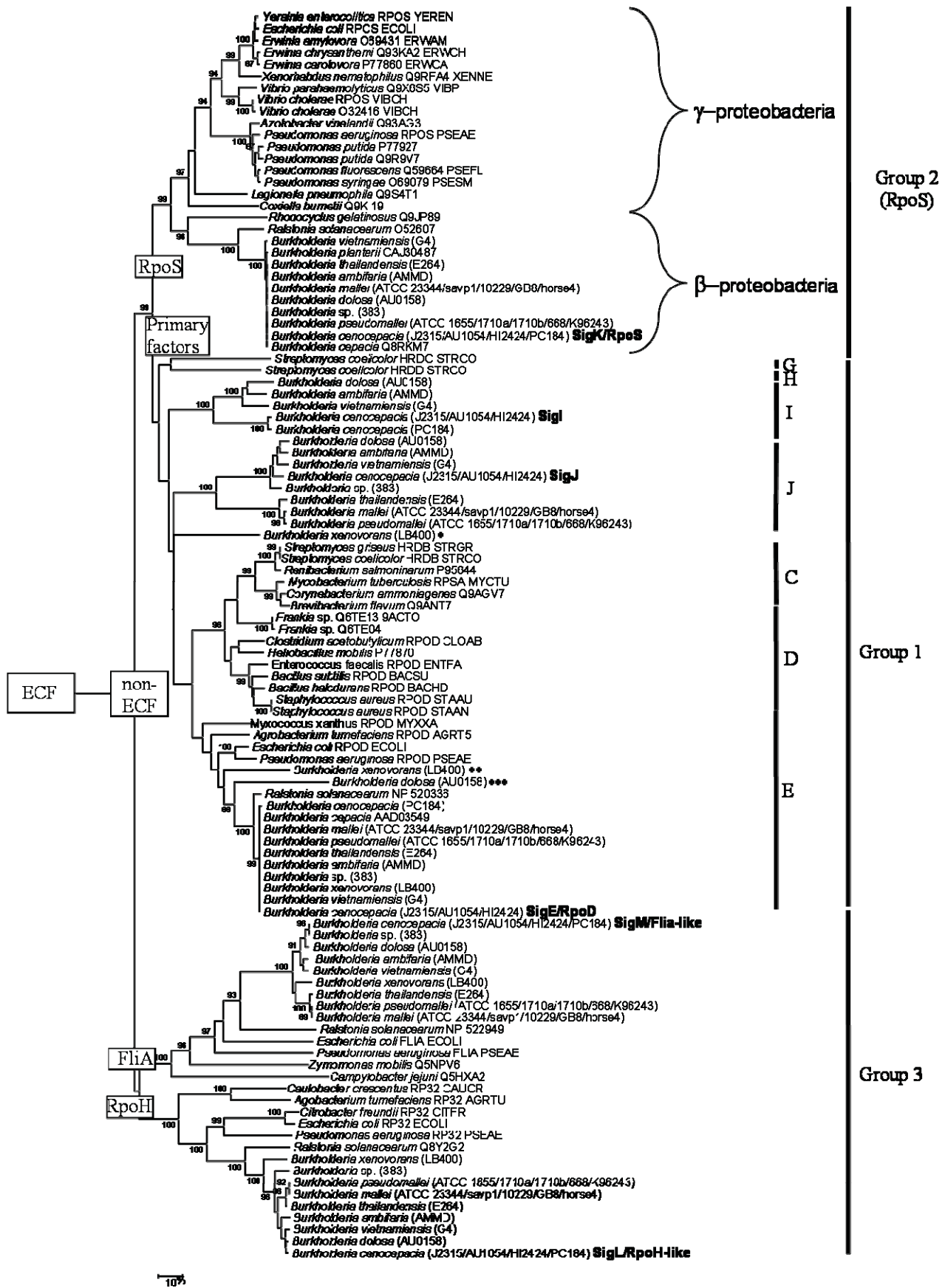
### 2.3.2. The stationary phase $\sigma^{70}$ factors phylogenetic radiation (group 2)

The group 2 phylogenetic radiation is divided into two clusters (supported by 96% and 97% of the bootstrap replicates, respectively) which match the diversification of the proteobacteria into the  $\beta$  and  $\gamma$  subgroups (Fig. 3). *Bcen-J2315 sigK* deduced amino acids sequence is clustered with the  $\beta$ -proteobacteria RpoS deduced sequences (and was thus renamed *rpoS*). In *B. cepacia* and *B. pseudomallei*, RpoS was shown to be implicated in oxidative stress response, to be induced by heat shock treatments, and upon carbon starvation [16,17]. No RpoS-like coding sequence was detected in any  $\alpha$ -proteobacterial or other eubacterial genomes. Among the actinomycetes, the stationary phase factors are known to be part of a fifth group of sigma factors which was not considered in this study i.e. the lineage of actinomycetal stress response and stationary phase factors (including sequences like *S. coelicolor* RpoX, *Mycobacterium tuberculosis* SigF, etc) [18,19]. This group is not closely related to group 2. Misleading automatic annotations of several RpoS gene sequences were observed in *Borrelia*, *Bacteroides*, *Blastopirellula*, *Rhodopirellula*, *Moorella* or *Aquifex* genome; their deduced factors did not group in the RpoS cluster (data not shown). To clarify this situation, we suggest the use of the name *rpoS* only for sequences of the group 2  $\sigma^{70}$  factors from the  $\beta$ - and  $\gamma$ -proteobacteria. The restricted distribution of RpoS gene sequences and their close phylogenetic proximity with Group 1 sequences (primary and alternative primary factors) suggest a likely emergence of *rpoS* from an ancestral group 1 sequence. A duplication of the highly conserved eubacterial *rpoD* gene probably led to this paralogous group. Alternative primary factors described in the above section might have evolved from similar but more recent duplication events.

The emergence of *rpoS* can be considered a key event in the evolution of the eubacteria. Molecular phylogenies of RpoD (group 1), and FliA-like and RpoH-like sequences (group 3) (Fig. 3) suggest an early diversification of the  $\alpha$ -proteobacterial sequences from those of the  $\beta$ - and  $\gamma$ -proteobacterial sequences. This can also be inferred from other datasets e.g. *rrs* and concatenation of essential proteins [20]. The emergence of *rpoS* matches the separation of the  $\alpha$ -proteobacteria (no *rpoS* – ancestral state) from the  $\beta$ - and  $\gamma$ -proteobacteria (harbouring *rpoS*).

2.3.3. *The stress response/cellular differentiation/flagellin synthesis  $\sigma^{70}$  factors phylogenetic radiation (group 3)*

Group 3 is polyphyletic and divided into two groups, one corresponding to FliA-like sequences, and one corresponding to RpoH-like ones (supported by 100% of the bootstrap replicates). *Bcen*-J2315 *sigM* deduced AA sequence belongs to the FliA subgroup, whereas *Bcen*-J2315 *sigL* deduced AA sequence groups with the RpoH-like sequences (Fig. 3). A *sigM* (*fliA*) homolog was confirmed to be involved in motility among the *Burkholderia* [21].



**Figure 3:** Neighbor-Joining phylogenetic tree of non-ECF  $\sigma^{70}$  factors.

The  $\sigma^{70}$  sequences were from [13] or were retrieved from sequenced *Burkholderia* genomes or the GenBank database. Within a single species only one ortholog per sigma lineage was kept. A total of 178 sites were analyzed (with global removal of gap-containing sites). Distances are proportional to evolutionary divergences expressed in substitutions per 100 sites. Bootstrap values higher than 85% are given. Sub-groups are defined in the text. (\*) *B. xenovorans* LB400 alternative primary factor. (\*\*) additional copy of *rpoD* found in *B. xenovorans* LB400 genome. (\*\*\*) Deduced *B. dolosa* AU0158 RpoD sequence that probably contains a frameshift due to a DNA sequencing mistake. Main phylogenetic groups of non-ECF factors are indicated on the figure: group 1, primary and alternative primary factors; group 2, stationary phase factors; and group 3, heat shock (RpoH), flagellin (FliA), and cellular differentiation factor

Table 2. *Burkholderia cepacia* complex strains used in this study, and summary of DNA blot analyses of these strains using *ecfB*, *ecfD*, *ecfE*, *sigI*, and *sigJ* DNA probes.

Strains	References	<i>ecfB</i>	<i>ecfD</i>	<i>ecfE</i>	<i>sigI</i>	<i>sigJ</i>
<i>B. cepacia</i>						
LMG 1222	[54]	+	+	+	+	+
AU0113	[55]	+	+	+	+	+
LMG 2161	[54]	+	+	+	+	+
<i>B. stabilis</i>						
LMG 14294	[56]	+	+	+	+	+
AU0244	[57]	+	+	+	+	+
<i>B. cenocepacia</i>						
J2315	[58]	+	+	+	+	+
H111	[59]	+	+	+	+	+
IST404	[60]	+	+	+	+	+
BCC020	this study	+	+	+	+	+
LMG 18828	[58]	+	+	+	+	+
LMG 18830	[58]	+	+	+	+	+
BCC021	this study	+	+	+	+	+
LMG 21440	[58]	+	+	+	+	+
LMG 18829	[58]	+	+	+	+	+
<i>B. multivorans</i>						
LMG 13010	[61]	+	+	-	-(+)*	+
LMG 16660	[61]	+	+	-	-(+)*	+
LMG 18822	[61]	+	+	-	-(+)*	+
C1962	[61]	+	+	-	-(+)*	+
LMG 16665	[61]	+	+	-	-(+)*	+
<i>B. vietnamiensis</i>						
LMG 10929	[62]	+	+	-	-(+)*	+
ATCC 53617	[62]	+	+	-	-(+)*	+
CL	[63]	+	+	-	-(+)*	+
LMG 16232	[62]	+	+	-	-(+)*	+
LMG 16228	[62]	+	+	-	-(+)*	+
LMG 17830	[62]	+	+	-	-(+)*	+
LMG 10823	[62]	+	+	-	-(+)*	+
<i>B. dolosa</i>						
LMG 18941	[64]	+	+	+	-(+)*	+
<i>B. pyrrocinia</i>						
LMG 14191	[65]	+	+	+	+	+

<sup>32</sup>P-random labelling of DNA probes were done as described in the Methods section. See Table 3 for position of the PCR products used for the synthesis of DNA probes. + : positive hybridization signal, - : no hybridization signal (<72% identity). \* : hybridization signals varied according to the *sigI* DNA probe used. (+): these positive signals were only obtained with the *Bvi-sigI* probe.

## 2.4. Molecular phylogeny of *Bcen*-J2315 extracytoplasmic functions $\sigma^{70}$ gene sequences (group 4)

Phylogenetic relationships among the ECF cluster are presented in Fig. 4. All database sequences closely related to *Bcen*-J2315 putative ECF were considered in this analysis. A first round of phylogenetic analysis allowed the identification of all database factors related to *Bcen*-J2315 ECFs. A second round of analyses was then performed by limiting the dataset to a set of selected sequences representative of the main phylogenetic trends observed. Fig. 4 shows the best NJ tree obtained. Its structure is not supported by high bootstrap values but all sub-groups of ECFs are supported by high ones. EcfA, EcfC and EcfI deduced AA sequences are the only *Bcen*-J2315 ECF factors having homologs in the *P. aeruginosa* genomes.

### 2.4.1. *ecfA*: evidence of an ET12 clone-specific duplication event

EcfA-like deduced sequences are found in several proteobacterial genomes but could not be detected in the ones of  $\alpha$ -proteobacteria (Fig. 4). Surprisingly, *B. vietnamiensis* and *Burkholderia* sp. 383 deduced EcfA sequences were found to group apart from other EcfA-like sequences belonging to the *Burkholderia* (Fig. 4); making these sequences closer relatives of a *Ralstonia* homologous sequence. *ecfA* encodes a factor highly similar to the well-characterized AlgU factor of *P. aeruginosa* which is involved in alginate biosynthesis [22]. However, the *B. cepacia* homolog of EcfA was recently shown not to be involved in exopolysaccharide biosynthesis and onion pathogenicity. It was found involved in the adaptation to heat stress and phosphate starvation [23]. Interestingly, the *ecfA* gene sequence was found in two copies in the *Bcen*-J2315 genome. These two copies are harboured by identical 40 kb DNA fragments delimited, on one hand, by an IS element.

To trace back the *ecfA* duplication event in *B. cenocepacia*, a PCR strategy was built. PCR primers were defined for each hand of the duplicated 40 kb regions containing the *ecfA* copies (Fig. 6a). PCR products from the B and C primers (Table 3), targeting the duplicated region containing the *ecfA2* allele, were obtained for all but the LMG 18863 strain of the ET12 clonal complex; a strain confirmed to be a member of the ET12 clone by PCR screening of the BCESM and *cblA* loci (data not shown, see [24]). PCR products amplified with the A and B primers (Table 3), targeting the end of the DNA region affected by the *ecfA* duplication and its linked chromosomal inversion, were obtained for this latter ET12 strain and the non-ET12 *B. cenocepacia* isolates. *B. cenocepacia* PHDC strains (AU1054 and HI2424) and *B. cenocepacia* PC184 genomes harboured only the *ecfA1* allele. The ET12 LMG 18863 strain, which was not found to harbour the 40 kb *ecfA*-duplicated region, could have either reverted to its previous genetic organization or could be a more ancestral representative of the ET12 clonal complex. These data suggest that duplication of the *ecfA* 40 kb locus could have contributed to the radiation of the ET12 complex.

*Bcen*-J2315 chromosome 1, was compared to chromosome 1 of *Burkholderia* sp. 383 using the Promer tool according to Kurtz *et al.* [25]. This tool provides large scale DNA sequence analysis. ORFs were automatically detected and similar ones were plotted on Fig. 6b. This analysis detected two inversion events on these chromosomes. To determine the role of *ecfA*-40 kb region in these duplication events, the chromosome 1 region was inverted

manually between the two 40 kb duplications and compared with *Burkholderia* sp. 383 chromosome 1 (Fig. 6c). In this way, only the first inversion is detected, indicating that the 40 kb duplication was implicated in the large scale inversion of *Bcen*-J2315 chromosome 1. One border of the 40 kb duplication was found to harbor an Insertion Sequence (IS) named ISBcen8 according to the IS Database [26]. This IS could have played a role in the duplication event that led to the inversion of an important part of chromosome 1 of ET12 strains. This insertion sequence is found (with at least 99% identity) in seven copies on chromosome 1, two copies on chromosome 2, and 3 copies on chromosome 3. The ORF encoding the transposase of this IS was identified with ORF finder and confirmed by GC skew analysis. IS492 from the IS110 family (ORF: 34% identity, 52% similarity) was found to be the closest phyletic neighbour of this IS.

In several species such as *P. aeruginosa*, *Salmonella typhi* or *Bordetella pertussis*, genome rearrangements were found associated with infectious outbreaks [27-29]. Hughes *et al.* [30] suggested that rearranged genomes could allow to overcome the immune system of the host. The *ecfA* 40 kb locus could thus have contributed to the emergence of the hyper-virulent ET12 infra-specific clonal complex by driving a large scale inversion of chromosome 1.

#### 2.4.2. *ecfB*: involvement in temperature-dependent processes

This ECF lineage has a broad distribution (Fig. 4). Related genes could be found in the  $\alpha$  and  $\gamma$ -proteobacteria. *ecfB* DNA probings of Bcc genomic DNA blots confirmed its large distribution in the Bcc (Table 2). The deduced EcfB protein is phylogenetically closely related to PrtI which is involved in a temperature dependent regulation of certain proteases [31]. However, no gene at the proximity of *ecfB* was found to encode a protease in *Bcen*-J2315 and other *Burkholderia* genomes. Such a situation was also observed for the PrtI-like locus of the *P. putida* genome [6]. *ecfB* could thus be involved in temperature dependent regulatory processes but seems to have been recruited for the regulation of other protein families than proteases.

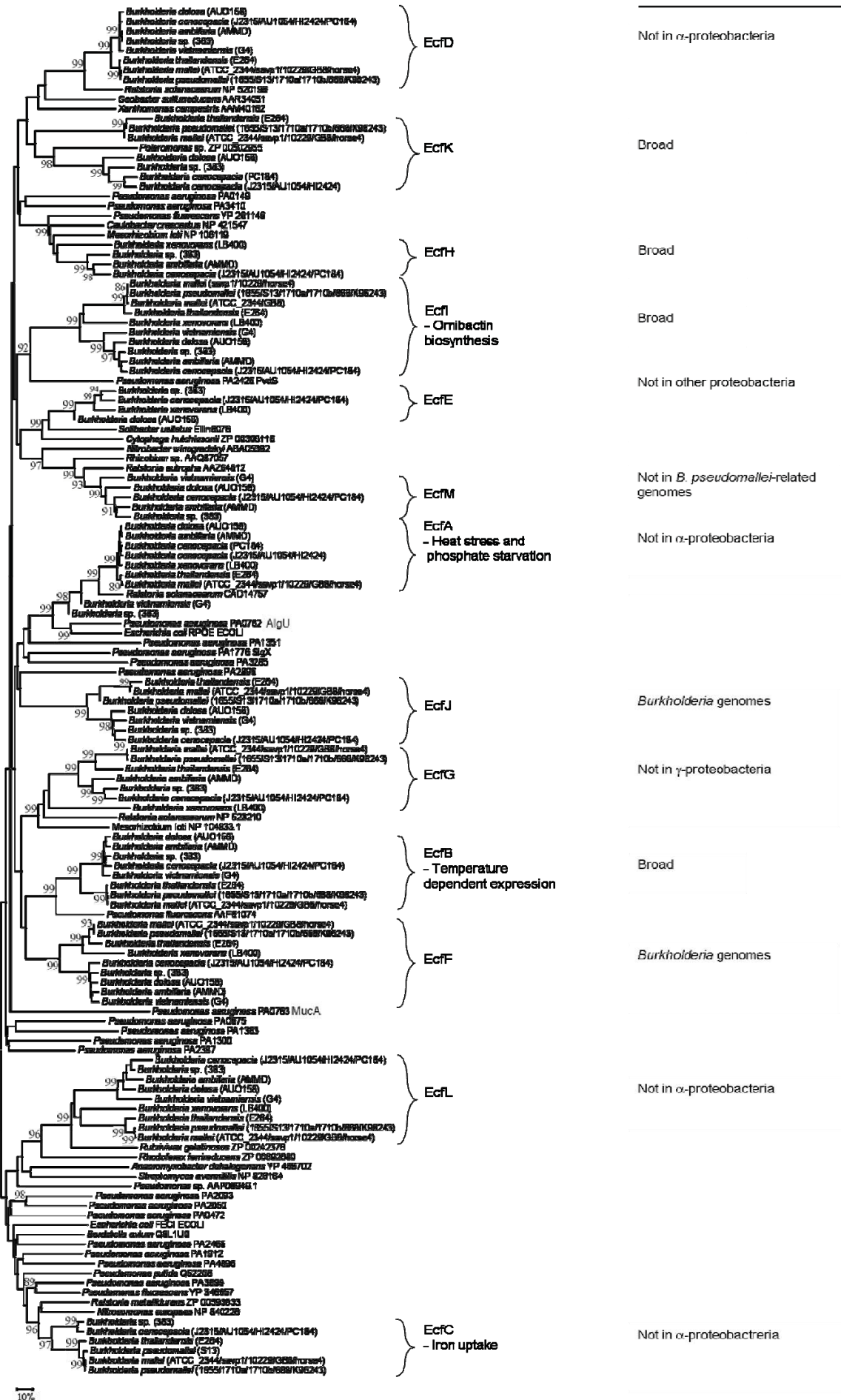
#### 2.4.3. *ecfC* and *ecfI*: ECF gene sequences involved in iron-uptake

The ECF NJ tree (Fig. 4) shows *Bcen*-J2315 deduced EcfI sigma factor to group with the PvdS factor from *Pseudomonas aeruginosa*, and the EcfC deduced factor to group with FecI-like factors. EcfI was previously shown to be involved in siderophore (ornibactin) biosynthesis [32] (EcfI being named OrbS in this latter article). Here, it is found to have a broad distribution (Fig. 4). FecI was shown to be involved in the regulation of iron dicitrate uptake in *E. coli* [33]. In other bacterial species, the *fecIRA* system can regulate iron uptake by other siderophores than citrate [34]. Phylogenetic analysis of EcfC/FecI homologs suggests an emergence of this particular sub-group after the diversification of the  $\alpha$ -proteobacteria (no *fecI* homologs) from the other proteobacteria. Interestingly, FecI-like factors represented 26% of *P. aeruginosa* deduced ECF factors, and 68% of those of *P. putida* [6]. Among *P. aeruginosa* and *P. putida*, these systems were all physically linked to genes involved in the synthesis of outer membrane siderophore receptor proteins. The high numbers of FecI paralogs in these two pseudomonads were considered indicative of a shared ability at colonizing certain niches that require distinct iron uptake systems or particular interactions with siderophores [6]. In the



case of the Bcc, a single FecI-like factor was detected. The *ecfC* gene, encoding this factor, was found physically linked to *fecR*- and *fecA*-like gene sequences. FecI can interact with FecR to activate the *fecABCDE* operon involved in the transport of extracellular Fe(III)-citrate complexes to the cytoplasm. FecR interaction is modulated by FecA which will activate the FecIR interactions after perception of Fe(III)-citrate complexes. TBLASTN searches using FecR were performed on selected Bcc genomes to ensure that no FecIRA-like system was missed during our analyses. These searches did not detect any other FecIRA system in *Bcen-J2315*, *B. vietnamiensis*, *B. dolosa*, and *B. ambifaria* genomes. Other TBLASTN searches using HemO, CirA, Has, HxuC, and *P. aeruginosa* SigX, FiuI, PigD, and PupI-homologs were performed to complete this analysis. None of these BLAST searches revealed novel coding sequences involved in haem- or iron-uptake [6] in *Bcen-J2315*.

Interestingly, *B. cenocepacia* was recently shown able to get its iron from ferritin by proteolytic degradation of this iron-binding protein. *B. cenocepacia* is apparently the only known pathogenic bacteria able to use ferritin as an iron source [35]. The iron uptake strategies selected by *P. aeruginosa* and the Bcc are thus quite different; with *P. aeruginosa* showing a greater versatility in ECF regulated processes.



**Figure 4:** Neighbor-Joining phylogenetic tree of ECF  $\sigma^{70}$  factors.

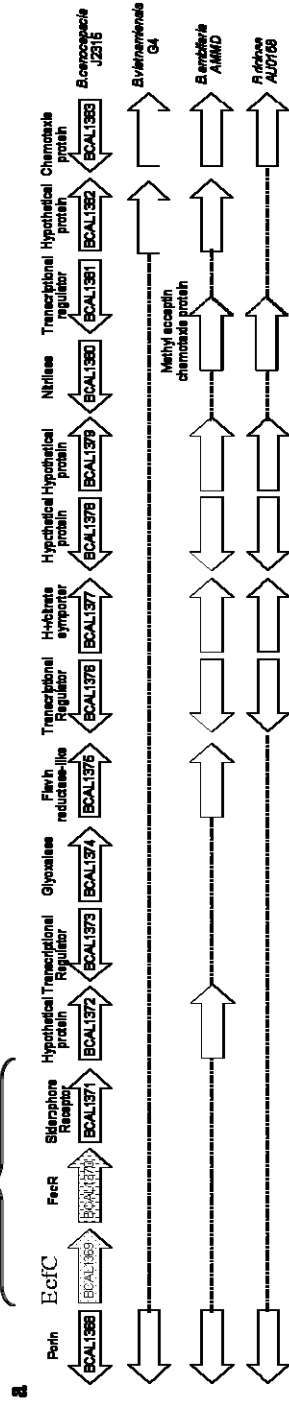
The sequences were retrieved from sequenced *Burkholderia* genomes or the GenBank database. Within one species, when sequences were orthologs, only one sequence per sigma factor was kept. A total of 746 sites were analyzed (with pairwise deletion of gap-containing sites). Distances are proportional to evolutionary divergences expressed in substitutions per 100 sites. Bootstrap values higher than 85% are given. Distribution among other proteobacterial groups is indicated. Likely role of the identified ECF factors is indicated.

#### 2.4.4. *ecfD*, *ecfE*, *ecfG*, *ecfH*, *ecfL* and *ecfM* loci: unknown physiological implications

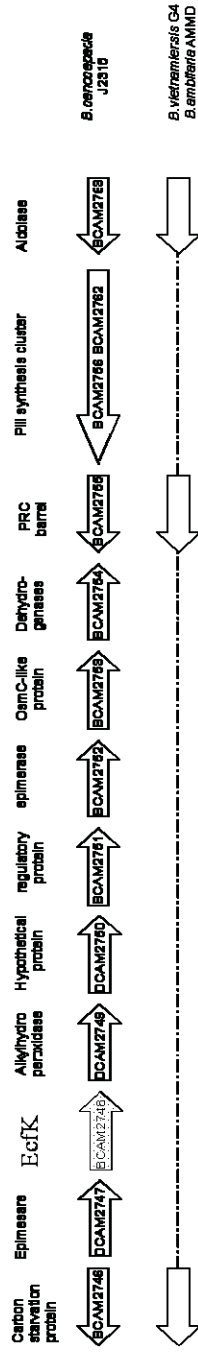
The physiological processes involving these regulatory factors remain unknown. However, their distribution and evolutionary history reveal several observations suggesting that they play significant roles.

EcfD is grouping with ECF factors deduced from genomes of other proteobacterial groups ( $\gamma$  and  $\Delta$ ) but not  $\alpha$ -proteobacterial ones. *ecfD* DNA probings of Bcc genomic DNA blots showed a large distribution in the Bcc (Table 2). EcfE is not grouping with ECF factors from other proteobacteria but is closely related to  $\sigma$  factors found among *Solibacter usitatus* and *Cytophaga hutchinsonii*. *ecfE* DNA probings gave positive hybridization signals on *B. cepacia*, *B. cenocepacia*, *B. stabilis*, *B. dolosa* and *B. pyrrocinia* genomic DNA blots (Table 2). However, restricted genomic DNA from all strains of *B. multivorans* and *B. vietnamiensis*, tested in this study, did not yield positive hybridization signals with the *ecfE* probe. Phage DNA might have played a role in the loss of *ecfE* in these bacterial species (see section on genomic rearrangements). Surprisingly, the *ecfE* deduced AA sequence of *B. dolosa* AU0158 was found distantly related to other homologs of the Bcc. EcfG is significantly grouping with  $\sigma$  factors deduced from other  $\beta$ - and  $\alpha$ -proteobacterial genomes but not with  $\gamma$ -proteobacterial ones. In this case, a deletion of *ecfG*-like sequences probably occurred early in the evolution of the  $\gamma$ -proteobacteria. Concerning the deduced AA sequences of *ecfK* and *ecfH* coding regions, homologs could be detected in genomes of the three main proteobacterial groups ( $\alpha$ ,  $\beta$ , and  $\gamma$ ). *Bcen*-J2315 EcfH deduced homologs were mainly found in genomes of  $\alpha$ -proteobacteria but a sequence was also detected among one  $\gamma$ -proteobacterial genome, the one of *Pseudomonas fluorescens*. *ecfH* was not detected in the *B. vietnamiensis* G4 and *B. ambifaria* AMMD genomes (see section on genomic rearrangements). EcfK homologs were found in other *Burkholderia* genomes and in the one of a species of *Polaromonas*, which is also part of the  $\beta$ -proteobacteria. *ecfK* was not detected in the *B. vietnamiensis* G4 and *B. ambifaria* AMMD genomes (see section on genomic rearrangements). EcfL is clustered with ECF without any known function but encoding a TPR (Tetratricopeptide repeat) motif involved in protein-protein interactions. This motif is not found in EcfL and its related *Burkholderia* ECF factors. In fact, deduced EcfL-like sequences retrieved from *B. ambifaria* AMMD, *B. cenocepacia* PC184 and *B. pseudomallei* 668 genomes are misleadingly annotated with a TPR repeat motif in the databases. EcfL coding gene sequences are found in the genomes of the actinobacteria and proteobacteria with the exception of the  $\alpha$ -proteobacterial ones. EcfM deduced orthologs were found among most proteobacterial sub-groups with the exception of *B. pseudomallei* and closely related species.

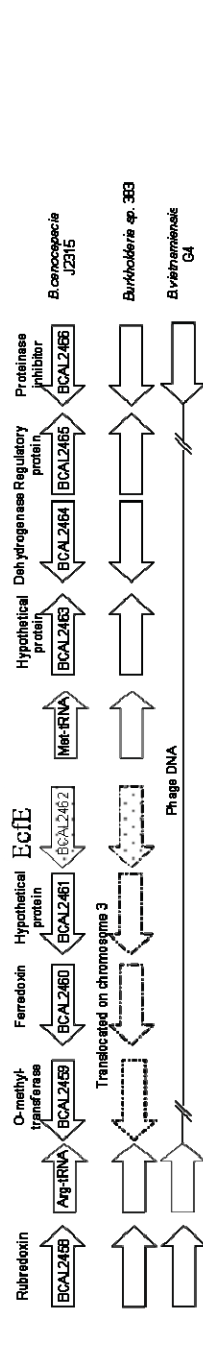
### FecIRA cluster



**b**

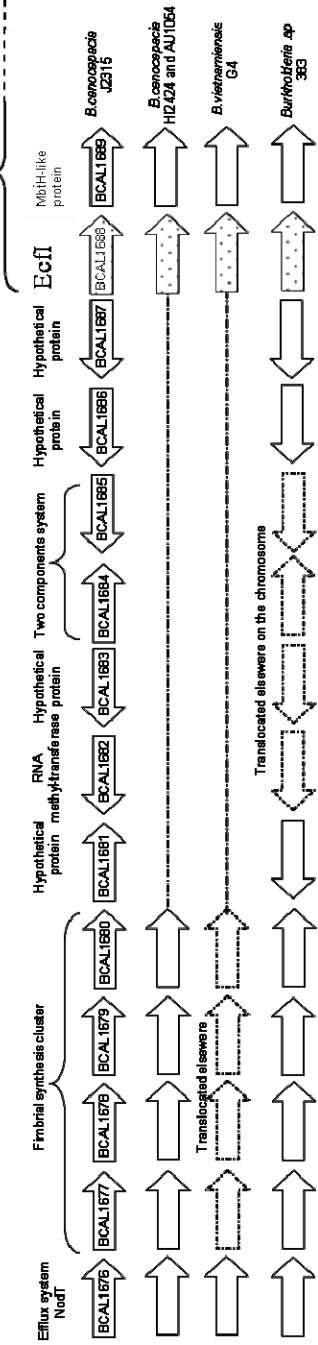


**c**



**d**

### Oribactin biosynthesis cluster



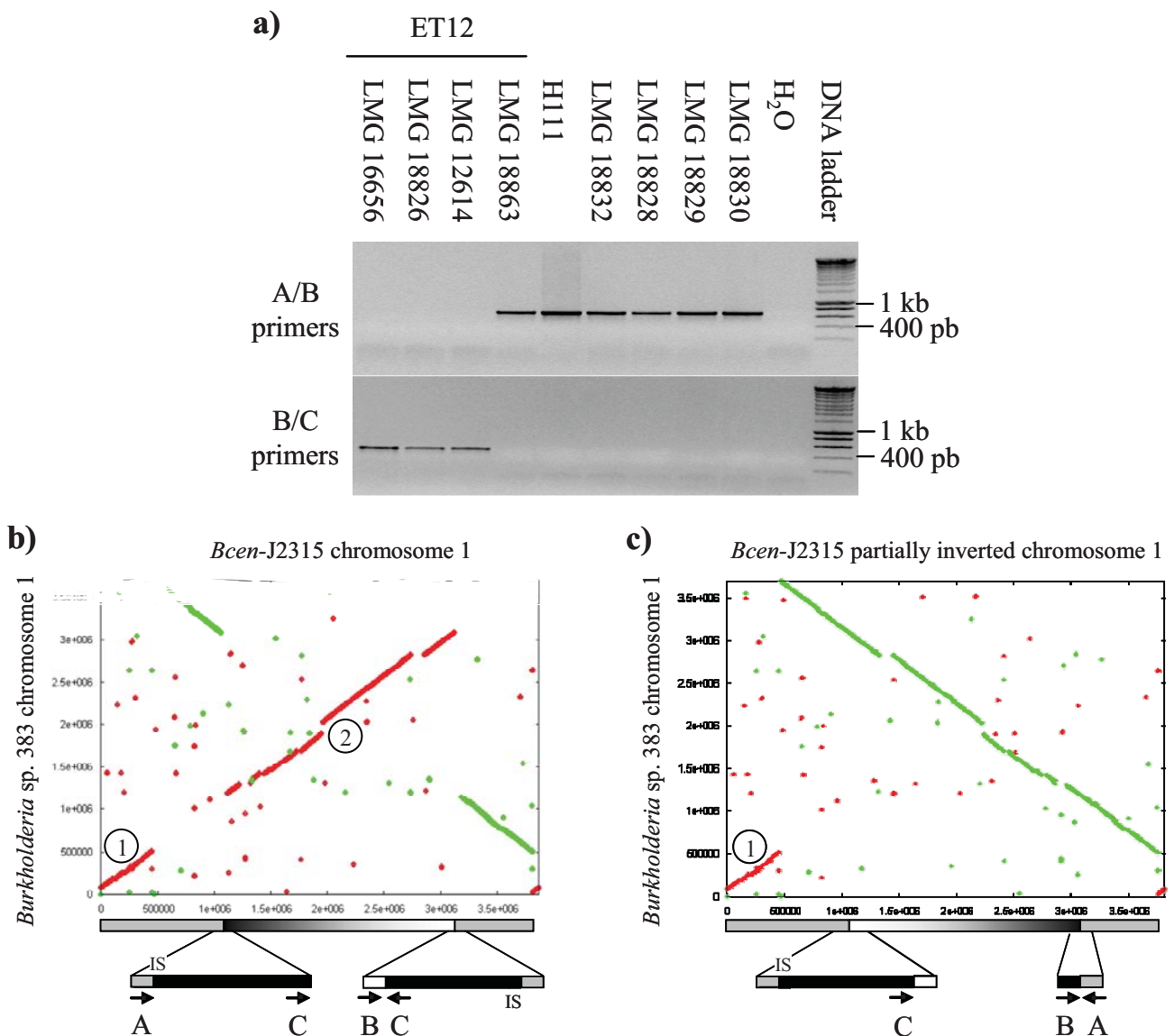
**Figure 5: Synteny analysis of *B. cereus* strain J2315 *ecf* genomic regions showing rearrangements in other *B. cereus* sequenced genomes.** (a) *ecfC*, (b) *ecfK*, (c) *ecfE*, and (d) *ecfI* concatenated genomes were compared to other *B. cereus* concatenated genomes using ACT [36]. *B. cereus* J2315 ORF annotations were assigned using other annotated *Burkholderia* genomes. ORF numbers assigned by the Sanger Institute are indicated. Dashes indicate missing regions. Phage DNA in (c) indicates a potential phage insertion. Arrows indicating ORF of sigma factors are filled with dots, and of anti-sigma factors are filled with vertical bars.

#### 2.4.5. *ecfF* and *ecfJ*: evidence of *Burkholderia*-specific lineages

*EcfF* and *EcfJ* are distinct deduced  $\sigma^{70}$  factors which do not group on the NJ tree (Fig. 4) with any of the previously described ECF sigma factors available in the databases. They can thus be considered *Burkholderia*-specific. However, *ecfJ* was not detected in the genome of *B. xenovorans* LB400. The processes making use of these ECF factors remain unknown.

#### 2.5. Genomic rearrangements of $\sigma^{70}$ genetic loci among Bcc sequenced genomes

Inside the Bcc, synteny analyses of the  $\sigma^{70}$  genetic loci were performed using ACT [36]. Some of these results are shown in Fig. 5. *sigE*, *sigL*, *sigM*, *sigK*, *ecfA*, *ecfB*, *ecfD*, *ecfF*, *ecfJ*, *ecfL* and *ecfM* genetic loci were conserved over at least 10 kb (upstream and downstream) among the tested Bcc genomes. *ecfB* is located 5,5 kb upstream *ecfM*. *sigI*, *sigJ*, *ecfC*, *ecfF*, *ecfG*, *ecfH*, *ecfI*, and *ecfK* loci were affected by various genomic rearrangements. *ecfA*, *ecfB* and *ecfF* are physically linked to anti-sigma gene sequences. Three major categories of genomic rearrangements were defined, one leading to a loss of a sigma gene through unknown processes, one leading to a loss of a sigma gene through integration of phage DNA, and one leading to a reorganization of the genetic environment surrounding the sigma gene.



**Figure 6: The *ecfA* (*rpoE*) gene duplication.** (a) PCR analysis of the duplicated 40 kb *ecfA* region among *Bcc* genomes. Strains used are described in Table 2. PCR primers are presented in Table 3. The DNA size marker is shown. Orientation of the PCR primers is indicated in the b and c sections. B/C primers detect the *Bcen*-J2315 configuration (duplication event)-*ecfA2* allele, and A/B primers detect the *Burkholderia* sp. 383 configuration (no inversion). (b) and (c) Promer plot comparison of *Burkholderia* sp. 383 chromosome 1 against either (b) *Bcen*-J2315 chromosome 1 or (c) the partially inverted *Bcen*-J2315 chromosome 1. ① inverted region #1 and ② inverted region #2 of *Bcen*-J2315. ISBcen8 position is indicated. Promer analyses were performed according to [25].

### 2.5.1. Loss of $\sigma^{70}$ gene sequences in the Bcc: cases of *ecfC*, *ecfG*, *ecfH*, *ecfK*, *sigI*, and *sigJ* loci

In this category, in one instance, a very precise deletion of a sigma gene and its respective anti-sigma was observed. This situation was observed for *ecfG* and its anti-sigma in the *B. vietnamiensis* G4 genome. However, in most instances, the loss of a sigma gene was concomitant to the loss of important pieces of DNA. *sigI* and *sigJ* are, in fact, found deleted in respectively *Burkholderia* sp. 383, and *B. cenocepacia* PC184, and the extension of the deleted regions were respectively of 20 and 150 kb. Loss of *ecfH* in *B. vietnamiensis* G4 and *B. dolosa* AU0158 genomes were also associated with the deletion of a DNA region (9 and 12 kb, respectively). In the case of the *ecfK* locus, which is found at the proximity of a pilus synthesis cluster and genes involved in the synthesis of an alkylhydroperoxidase (Fig. 5b), its deletion involved a loss of about 7 kb in the genomes of *B. vietnamiensis* G4 and *B. ambifaria* AMMD. Interestingly, the configuration of these latter deletions were identical in these two genomes, suggesting that it had either occurred prior to the separation of these bacteria into two distinct species or involved similar processes. *ecfC* deletion event was observed among several genomes but the extent of the deletion events was variable. Nevertheless, one border was always conserved. *ecfC* (*fecI*-like) is deleted in *B. vietnamiensis* G4, *B. ambifaria* AMMD and *B. dolosa* AU0158 genomes (Fig. 5a). Part of the deletion concerned genes involved in iron-uptake. These genomic rearrangements of the *ecfC* genetic cluster could explain the differences observed between *B. cepacia* and *B. vietnamiensis* in response to iron starvation. *B. vietnamiensis* was only found to produce ornibactin under iron starvation while *B. cepacia* clinical isolates could also produce pyochelin and cepabactin [37]. Among the *B. ambifaria* and *B. dolosa* *ecfC* loci, 4 to 6 ORFs, over the eleven likely involved in iron-uptake in *Bcen*-J2315, were found conserved. One of these was found encoding a citrate symporter involved in the transport of Fe(III)-citrate complexes. To our knowledge, it is not yet known if *B. ambifaria* and *B. dolosa* *ecfC* loci are functional.

### 2.5.2. Loss of *ecfE* in *B. vietnamiensis*: involvement of phage DNA

Concerning this second category, synteny analysis showed the *Bcen*-J2315 *ecfE* region to have been modified by the integration of phage DNA in *B. vietnamiensis* strain G4. In this genome, *ecfE* and an overall 5 kb region are replaced by phage DNA. All genes upstream the tRNA gene of the *ecfE* region are conserved. The loss of *ecfE* in *B. vietnamiensis* was confirmed by DNA blot analyses (Table 2). All restricted genomic DNA of the *B. vietnamiensis* strains tested did not show any DNA hybridization with an *ecfE* probe. Similarly, no positive hybridization signal could be obtained on *B. multivorans* DNA blots (Table 2), while positive signals were obtained for the other Bcc species. It is not known if the integration of phage DNA was involved in the deletion of *ecfE* among the *B. vietnamiensis* and *B. multivorans* strains analysed by DNA probings. However, we consider that this is highly probable.

### 2.5.3. Reorganized $\sigma^{70}$ genetic loci in the Bcc: cases of *ecfI* and *ecfE* loci

Concerning this third category, two cases were observed. One concerned the *ecfI* (*pvdS* homolog) region, and the other one the *ecfE* region. Regarding *ecfI* and its related genetic cluster, the ornibactin biosynthesis cluster, genes upstream *ecfI* are found deleted in *B. vietnamiensis* G4, *B. cenocepacia* HI2424 and AU1054 genomes, and translocated elsewhere in the *Burkholderia* sp. 383 genome. Genes involved in fimbrial synthesis, and the synthesis of a two component system were partly affected by these rearrangements (see Fig. 5d). Similarly, the *ecfE* region and three of its neighboring genes proximal to a tRNA gene sequence in *B. cenocepacia* J2315 chromosome 1 were translocated into a distinct region of the genome of *Burkholderia* sp. 383, while all other genes were kept in place (Fig. 5c).

### 3. Conclusion

This paper was dedicated to the analysis of the  $\sigma^{70}$  gene family, its architecture and plasticity, through an analysis of about 20 genome sequences of the eubacterial *Burkholderia* genus. Emphasis was made on a bacterial group of this genus which is named the "*Burkholderia cepacia* complex" or Bcc. Several epidemic clones have been described in the Bcc, including bacteria responsible of the "*cepacia* syndrome", a dramatic necrotizing pneumonia. The great versatility of these Bcc species could be linked to an important diversity in regulatory processes including those of the  $\sigma^{70}$  gene family. The  $\sigma^{70}$  gene family is widely distributed among the eubacteria, and probably emerged early in bacterial evolution. Duplication events led to the emergence of several Bcc  $\sigma^{70}$  paralogs which were found limited to certain bacterial groups like the proteobacteria, *Burkholderia*, the *cepacia* complex, and particular Bcc species or strains. This gene family was found to be under strong selective pressures that could lead to acquisition/deletion (*ecfE*, *ecfG*, *sigI*, *sigJ*, *ecfC*, *ecfH*, and *ecfK*), and duplication events (*ecfA* and *rpoD*) modifying its architecture. These changes are likely to be key events in Bcc evolution, generating novel gene expression profiles that could be more suited for the colonization of particular biotopes including the respiratory tract of CF patients. A key conclusion was that the comparative analysis of the  $\sigma^{70}$  gene architecture of Bcc species and of *Pseudomonas aeruginosa*, a species found in similar habitats including CF lungs, revealed distinct evolutionary strategy, with the Bcc having selected several alternative primary factors, something not recorded among *P. aeruginosa* and only reported to occur among the actinobacteria. Only three ECF paralogs (FecI-like, PvdS and AlgU) were found in common between the Bcc and *P. aeruginosa*. The next step in these studies will be to investigate the impact of gains and losses of  $\sigma^{70}$  determinants in the ecology of the Bcc and the colonization of the CF lungs.



## 4. Methods

### 4.1. *B. cenocepacia* genome sequence, and $\sigma^{70}$ BLAST and ORF searches

TBLASTN searches of *B. cenocepacia* J2315 genome were performed using the Wellcome Trust Sanger Institute web site [26]. The following ten  $\sigma^{70}$  factors were used for these searches: RpoD (accession #AAA24601), RpoS (#CAA34435), RpoH (#AAA24587), SigF (#AAA22788), FliA (#BAA00389), AlgU (#AAC43714), CnrH (#AAA21967), HrpL (#AAD00805), PpuI (#CAA54870), and RfaY (#AAA92044). All loci encoding putative proteins showing good similarities with the target proteins were analysed by FramePlot [38], ORF finder [39] and testcode [40]. BLASTP analyses of the deduced ORF proteins were performed [41]. Conserved domains among sigma factors were detected using CDD 2.02 [42]. To achieve saturation in the detection of  $\sigma^{70}$  factors, identified putative factors were concatenated and the resulting sequence compared to *Bcen*-J2315 chromosomes using ACT. This graphical tool allows visualisation of a BLAST comparison of two sequences. A TBLASTN was performed using *Bcen*-J2315 chromosomes as the database and the concatenated sigma sequences as the query. Visual identification of DNA regions showing similarity with different sigma factors was performed and candidate ORF were analysed as stated above.

### 4.2. Molecular phylogenetic analyses

Multiple alignments were computed using CLUSTALW [43]. Distances between sequence pairs, the inferred phylogenetic trees, and bootstrap values were all computed through the Mega3 graphic tool [44]. Phylogenetic trees were built using the Neighbor-Joining (NJ) method [45]. Bootstrap replicates were performed using 1000 replicates. Gap containing sites were deleted globally or by pair of sequences. Alignments can be downloaded using the UMR5557 http server [46,47].

### 4.3. Bacterial strains, DNA extractions, PCR amplification and DNA blot hybridizations

Strains used in this study are listed in Table 2. Bacteria were grown in LB (Luria-Bertani) [48] broth at 30°C. DNA extractions were performed as previously described [49]. PCR amplifications (25  $\mu$ l) were done according to the *Taq* polymerase manufacturer (Invitrogen, Cergy Pontoise France) without 1% w-1 (detergent) but 1,25  $\mu$ l of DMSO. PCR primers are indicated in Table 3. PCR cycles using total DNA were as follow: (1) 95°C for 5 min, (2) 95°C for 30 sec, selected annealing temperature for 30 sec (see Table 3), 72°C for 30 sec (35 cycles), (3) 72°C for 5 min.

Genomic DNA blots were performed on Genescreen (USA) nylon membranes following [50]. The DNA (5  $\mu$ g) was digested with *Eco*RI (Fermentas) at 1 U/ $\mu$ g DNA, for 16 h at 37°C. DNA probes were labelled using the random priming labelling kit (Amersham-Pharmacia, Orsay-France). DNA hybridizations were carried out according to the nylon membrane manufacturer (Genescreen, USA). Membranes were hybridized and washed at 65°C. Autoradiography was performed according to [50].

## 5. Acknowledgements

We thank the CNRS, Université Lyon1, and Rhône-Alpes Region (France) for their financial supports. AM was allocated a Ph. D. scholarship by the Ministère de la Recherche et de l'Enseignement Supérieur (France). AG was allocated a Ph. D scholarship from the «Cluster Environnement» of Rhône-Alpes Region (France). We thank the Sanger and Joint Genome Institutes for having made available *Burkholderia* genome sequences to the scientific community.

Table 3. PCR primers used in this study.

DNA targets	PCR primer code	DNA sequence	position on <i>Bcen-12315</i> genome	annealing temperature
<i>ecfA1</i> region	A	CACGGTCATCACGGCTCGC	Chr1 3124945-3124962	55°C
<i>ecfA1</i> and <i>ecfA2</i> region	B	TTCGGTGCATAACCAGC	Chr1 1056358-1056374	55°C
	C	CGGGCGCAAGAACATACA	Chr1 1056771-1056789	55°C
<i>ecfB</i>	ecfB-F	CGCTTCGGCGCTCTGGCTG	Chr1 3817718- 3817735	58°C
<i>ecfB</i>	ecfB-R	TCAGCCGCAAGGAAAGGAGC	Chr1 3818171- 3818153	58°C
<i>ecfD</i>	ecfD-F	CGCAACTGCGGGCGCCCG	Chr1 2115044-2115060	55°C
<i>ecfD</i>	ecfD-R	CGAGCGTTTCGAGCGGGGTC	Chr1 2615449-2615467	55°C
<i>ecfE</i>	ecfE-F	GTGAACGAAACGAGGCTAGG	Chr1 2728191-2728210	55°C
<i>ecfE</i>	ecfE-R	CTCCTTGAAAGGACTGTCC	Chr1 2727807-2727824	55°C
<i>sigI</i>	sigI-F	ATCCCGCTCGATGCCGCGC	Chr2 2667864-2667882	55°C
<i>sigI</i>	sigI-R	GAGAAATTTGAAAGCCCGGCCG	Chr2 2668242-2668261	55°C
<i>sigI</i>	sigIv-F	ATCATGCCCGCTGGCTC	Chr3* 1181202- 1181218	55°C
<i>sigI</i>	sigIv-R	AACAGCCCGATGTTGCCT	Chr3* 1181508- 1181491	55°C
<i>sigJ</i>	sigJ-F	CATGCCGATGGCACCCGTTTC	Chr2 1383013-1383032	55°C
<i>sigJ</i>	sigJ-R	GAGAAATTTGAAAGCCCGGCCG	Chr2 2668242-2668261	55°C

## References

1. Mahenthiralingam, E; Urban, TA; Goldberg, JB. The multifarious, multireplicon *Burkholderia cepacia* complex. *Nat Rev Microbiol.* 2005;3:144–156.
2. Isles, A; Maclusky, I; Corey, M; Gold, R; Prober, C; Fleming, P; Levison, H. *Pseudomonas cepacia* infection in cystic fibrosis: an emerging problem. *J Pediatr.* 1984;104:206–210.
3. LiPuma, JJ; Mortensen, JE; Dasen, SE; Edlind, TD; Schidlow, DV; Burns, JL; Stull, TL. Ribotype analysis of *Pseudomonas cepacia* from cystic fibrosis treatment centers. *J Pediatr.* 1988;113:859–862.
4. Coenye, T; Spilker, T; Van Schoor, A; LiPuma, JJ; Vandamme, P. Recovery of *Burkholderia cenocepacia* strain PHDC from cystic fibrosis patients in Europe. *Thorax.* 2004;59:952–954.
5. GOLD: Genomes OnLine database V.2 . <http://www.genomesonline.org>
6. Martinez-Bueno, MA; Tobes, R; Rey, M; Ramos, JL. Detection of multiple extracytoplasmic function (ECF) sigma factors in the genome of *Pseudomonas putida* KT2440 and their counterparts in *Pseudomonas aeruginosa* PA01. *Environ Microbiol.* 2002;4:842–855.
7. Rojo, F. Mechanisms of transcriptional repression. *Curr Opin Microbiol.* 2001;4:145–151.
8. Lonetto, MA; Rhodius, V; Lamberg, K; Kiley, P; Busby, S; Gross, C. Identification of a contact site for different transcription activators in region 4 of the *Escherichia coli* RNA polymerase sigma70 subunit. *J Mol Biol.* 1998;284:1353–1365.
9. Bentley, SD; Chater, KF; Cerdeno-Tarraga, AM; Challis, GL; Thomson, NR; James, KD; Harris, DE; Quail, MA; Kieser, H; Harper, D; Bateman, A; Brown, S; Chandra, G; Chen, CW; Collins, M; Cronin, A; Fraser, A; Goble, A; Hidalgo, J; Hornsby, T; Howarth, S; Huang, CH; Kieser, T; Larke, L; Murphy, L; Oliver, K; O'Neil, S; Rabbinowitsch, E; Rajandream, MA; Rutherford, K; Rutter, S; Seeger, K; Saunders, D; Sharp, S; Squares, R; Squares, S; Taylor, K; Warren, T; Wietzorrek, A; Woodward, J; Barrell, BG; Parkhill, J; Hopwood, DA. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature.* 2002;417:141–147.
10. Himmelreich, R; Hilbert, H; Plagens, H; Pirkl, E; Li, BC; Herrmann, R. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* 1996;24:4420–4449.
11. Dandekar, T; Huynen, M; Regula, JT; Ueberle, B; Zimmermann, CU; Andrade, MA; Doerks, T; Sanchez-Pulido, L; Snel, B; Suyama, M; Yuan, YP; Herrmann, R; Bork, P. Re-annotating the *Mycoplasma pneumoniae* genome sequence: adding value, function and reading frames. *Nucleic Acids Res.* 2000;28:3278–3288.
12. Lonetto, M; Gribskov, M; Gross, CA. The sigma 70 family: sequence conservation and evolutionary relationships. *J Bacteriol.* 1992;174:3843–3849.
13. Lavire, C; Blaha, D; Cournoyer, B. Selection of unusual actinomycetal primary sigma 70 factors by plant-colonizing *Frankia* strains. *Appl Environ Microbiol.* 2004;70:991–998.
14. Maeda, H; Fujita, N; Ishihama, A. Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res.* 2000;28:3497–3503.
15. Lavenir, R; Jocktane, D; Laurent, F; Nazaret, S; Cournoyer, B. Improved reliability of *Pseudomonas aeruginosa* PCR detection by the use of the species-specific ecfX gene target. *J Microbiol Methods.* 2007;70:20–29. doi: 10.1016/j.mimet.2007.03.008.
16. Aguilar, C; Bertani, I; Venturi, V. Quorum-sensing system and stationary-phase sigma factor (rpoS) of the onion pathogen *Burkholderia cepacia* genomovar I type strain, ATCC 25416. *Appl Environ Microbiol.* 2003;69:1739–1747.

17. Subsin, B; Thomas, MS; Katzenmeier, G; Shaw, JG; Tungpradabkul, S; Kunakorn, M. Role of the stationary growth phase sigma factor RpoS of *Burkholderia pseudomallei* in response to physiological stress conditions. *J Bacteriol.* 2003;185:7008–7014.
18. DeMaio, J; Zhang, Y; Ko, C; Young, DB; Bishai, WR. A stationary-phase stress-response sigma factor from *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A.* 1996;93:2790–2794.
19. Potuckova, L; Kelemen, GH; Findlay, KC; Lonetto, MA; Buttner, MJ; Kormanec, J. A new RNA polymerase sigma factor, sigma F, is required for the late stages of morphological differentiation in *Streptomyces* spp. *Mol Microbiol.* 1995;17:37–48.
20. Brochier, C; Bapteste, E; Moreira, D; Philippe, H. Eubacterial phylogeny based on translational apparatus proteins. *Trends Genet.* 2002;18:1–5.
21. Kim, J; Kang, Y; Choi, O; Jeong, Y; Jeong, JE; Lim, JY; Kim, M; Moon, JS; Suga, H; Hwang, I. Regulation of polar flagellum genes is mediated by quorum sensing and FlhDC in *Burkholderia glumae*. *Mol Microbiol.* 2007;64:165–179.
22. Martin, DW; Holloway, BW; Deretic, V. Characterization of a locus determining the mucoid status of *Pseudomonas aeruginosa*: AlgU shows sequence similarities with a *Bacillus* sigma factor. *J Bacteriol.* 1993;175:1153–1164.
23. Devescovi, G; Venturi, V. The *Burkholderia cepacia* rpoE gene is not involved in exopolysaccharide production and onion pathogenicity. *Can J Microbiol.* 2006;52:260–265.
24. Mahenthiralingam, E; Simpson, DA; Speert, DP. Identification and characterization of a novel DNA marker associated with epidemic *Burkholderia cepacia* strains recovered from patients with cystic fibrosis. *J Clin Microbiol.* 1997;35:808–816.
25. Kurtz, S; Phillippy, A; Delcher, AL; Smoot, M; Shumway, M; Antonescu, C; Salzberg, SL. Versatile and open software for comparing large genomes. *Genome Biol.* 2004;5:R12.
26. Wellcome Trust Sanger Institute web site. [www.sanger.ac.uk/Projects/B\\_cenocepacia/](http://www.sanger.ac.uk/Projects/B_cenocepacia/)
27. Echeita, MA; Usera, MA. Chromosomal rearrangements in *Salmonella enterica* serotype typhi affecting molecular typing in outbreak investigations. *J Clin Microbiol.* 1998;36:2123–2126.
28. Stibitz, S; Yang, MS. Genomic plasticity in natural populations of *Bordetella pertussis*. *J Bacteriol.* 1999;181:5512–5515.
29. Romling, U; Schmidt, KD; Tummeler, B. Large genome rearrangements discovered by the detailed analysis of 21 *Pseudomonas aeruginosa* clone C isolates found in environment and disease habitats. *J Mol Biol.* 1997;271:386–404.
30. Hughes, D. Evaluating genome dynamics: the constraints on rearrangements within bacterial genomes. *Genome Biol.* 2000;1
31. Burger, M; Woods, RG; McCarthy, C; Beacham, IR. Temperature regulation of protease in *Pseudomonas fluorescens* LS107d2 by an ECF sigma factor and a transmembrane activator. *Microbiology.* 2000;146 Pt 12:3149–3155.
32. Agnoli, K; Lowe, CA; Farmer, KL; Husnain, SI; Thomas, MS. The ornibactin biosynthesis and transport genes of *Burkholderia cenocepacia* are regulated by an extracytoplasmic function sigma factor which is a part of the Fur regulon. *J Bacteriol.* 2006;188:3631–3644.
33. Enz, S; Mahren, S; Stroehrer, UH; Braun, V. Surface signaling in ferric citrate transport gene induction: interaction of the FecA, FecR, and FecI regulatory proteins. *J Bacteriol.* 2000;182:637–646.
34. Mahren, S; Schnell, H; Braun, V. Occurrence and regulation of the ferric citrate transport system in *Escherichia coli* B, *Klebsiella pneumoniae*, *Enterobacter aerogenes*, and *Photobacterium luminescens*. *Arch Microbiol.* 2005;184:175–186.
35. Whitby, PW; Vanwagoner, TM; Springer, JM; Morton, DJ; Seale, TW; Stull, TL. *Burkholderia cenocepacia* utilizes ferritin as an iron source. *J Med Microbiol.* 2006;55:661–668.

36. Carver TJ, RKM., Berriman M, Rajandream MA, Barrell BG, Parkhill J ACT: the Artemis Comparison Tool. *Bioinformatics*. 2005;21:3422–3423.
37. Meyer, JM; Van, VT; Stintzi, A; Berge, O; Winkelmann, G. Ornibactin production and transport properties in strains of *Burkholderia vietnamiensis* and *Burkholderia cepacia* (formerly *Pseudomonas cepacia*). *Biometals*. 1995;8:309–317.
38. FramePlot 2.3.2. <http://www.nih.gov/~jun/cgi-bin/frameplot.pl>
39. ORF Finder. <http://www.ncbi.nlm.nih.gov/gorf/gorf.html>
40. Fickett, JW. Recognition of protein coding regions in DNA sequences. *Nucleic Acids Res*. 1982;10:5303–5318.
41. Basic Local Alignment Search Tool. <http://www.ncbi.nlm.nih.gov/BLAST>
42. Marchler-Bauer, A; Anderson, JB; Cherukuri, PF; DeWeese-Scott, C; Geer, LY; Gwadz, M; He, S; Hurwitz, DI; Jackson, JD; Ke, Z; Lanczycki, CJ; Liebert, CA; Liu, C; Lu, F; Marchler, GH; Mullokandov, M; Shoemaker, BA; Simonyan, V; Song, JS; Thiessen, PA; Yamashita, RA; Yin, JJ; Zhang, D; Bryant, SH. CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res*. 2005;33:D192–6.
43. Thompson, JD; Higgins, DG; Gibson, TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994;22:4673–4680.
44. Kumar, S; Tamura, K; Nei, M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform*. 2004;5:150–163.
45. Saitou, N; Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987;4:406–425.
46. HTTP server UMR5557 ECF. [http://ecomicro.univ-lyon1.fr/poe/ECF\\_alignment.fas](http://ecomicro.univ-lyon1.fr/poe/ECF_alignment.fas)
47. HTTP server UMR5557 NON ECF. [http://ecomicro.univ-lyon1.fr/poe/NON-ECF\\_alignment.fas](http://ecomicro.univ-lyon1.fr/poe/NON-ECF_alignment.fas)
48. Bertani, G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol*. 1951;62:293–300.
49. Pitcher, DG., Saunders, N.A., Owen, R.J. Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Lett Appl Microbiol*. 1989;8:151–156.
50. Sambrook, J., Fritsch E. F., and Maniatis T. *Molecular cloning: a laboratory manual*. 2nd . , editor. Cold Spring Harbor, N.Y. , Cold Spring Harbor Laboratory Press; 1989.
51. J. Craig Venter Institute. <http://www.tigr.org>
52. The Joint Genome Institute. [http://genome.jgi-psf.org/mic\\_home.html](http://genome.jgi-psf.org/mic_home.html)
53. Wellcome Trust Sanger Institute Bacterial Genomes . <http://www.sanger.ac.uk/Projects/Microbes/>
54. Broad Institute. <http://www.broad.mit.edu>
55. Palleroni, NJ; Holmes, B. *Pseudomonas-cepacia* Sp-Nov, Nom-Rev. *Int J Syst Bacteriol*. 1981;31:479–481.
56. LiPuma, JJ; Spilker, T; Gill, LH; Campbell, PW, 3rd; Liu, L; Mahenthiralingam, E. Disproportionate distribution of *Burkholderia cepacia* complex species and transmissibility markers in cystic fibrosis. *Am J Respir Crit Care Med*. 2001;164:92–96.
57. Vandamme, P; Mahenthiralingam, E; Holmes, B; Coenye, T; Hoste, B; De Vos, P; Henry, D; Speert, DP. Identification and population structure of *Burkholderia stabilis* sp. nov. (formerly *Burkholderia cepacia* genomovar IV). *J Clin Microbiol*. 2000;38:1042–1047.
58. Ramette, A; LiPuma, JJ; Tiedje, JM. Species abundance and diversity of *Burkholderia cepacia* complex in the environment. *Appl Environ Microbiol*. 2005;71:1193–1201.
59. Vandamme, P; Holmes, B; Coenye, T; Goris, J; Mahenthiralingam, E; LiPuma, JJ; Govan, JR. *Burkholderia cenocepacia* sp. nov.--a new twist to an old story. *Res Microbiol*. 2003;154:91–96.
60. Romling, U; Fiedler, B; Bosshammer, J; Grothues, D; Greipel, J; von der Hardt, H; Tummler, B. Epidemiology of chronic *Pseudomonas aeruginosa* infections in cystic fibrosis. *J Infect Dis*. 1994;170:1616–1621.

61. Richau, JA; Leitaó, JH; Correia, M; Lito, L; Salgado, MJ; Barreto, C; Cescutti, P; Sa-Correia, I. Molecular typing and exopolysaccharide biosynthesis of *Burkholderia cepacia* isolates from a Portuguese cystic fibrosis center. *J Clin Microbiol.* 2000;38:1651–1655.
62. Vandamme, P; Holmes, B; Vancanneyt, M; Coenye, T; Hoste, B; Coopman, R; Revets, H; Lauwers, S; Gillis, M; Kersters, K; Govan, JR. Occurrence of multiple genomovars of *Burkholderia cepacia* in cystic fibrosis patients and proposal of *Burkholderia multivorans* sp. nov. *Int J Syst Bacteriol.* 1997;47:1188–1200.
63. Gillis, M; Van Van, T; Bardin, R; Goor, M; Hebbbar, P; Willems, A; Segers, P; Kersters, K; Heulin, T; Fernandez, M. Polyphasic taxonomy in the genus *Burkholderia* leading to an emended description of the genus and proposition of *Burkholderia vietnamiensis* sp. nov. for N<sub>2</sub>-fixing isolates from rice in Vietnam. *Int J Syst Bacteriol.* 1995;45:274–289.
64. Menard, A; Monnez, C; Estrada de Los Santos, P; Segonds, C; Caballero-Mellado, J; Lipuma, JJ; Chabanon, G; Cournoyer, B. Selection of nitrogen-fixing deficient *Burkholderia vietnamiensis* strains by cystic fibrosis patients: involvement of nif gene deletions and auxotrophic mutations. *Environ Microbiol.* 2007;9:1176–1185.
65. Vermis, K; Coenye, T; LiPuma, JJ; Mahenthalingam, E; Nelis, HJ; Vandamme, P. Proposal to accommodate *Burkholderia cepacia* genomovar VI as *Burkholderia dolosa* sp. nov. *Int J Syst Evol Microbiol.* 2004;54:689–691.
66. Vandamme, P; Henry, D; Coenye, T; Nzula, S; Vancanneyt, M; LiPuma, JJ; Speert, DP; Govan, JR; Mahenthalingam, E. *Burkholderia anthina* sp. nov. and *Burkholderia pyrrocinia*, two additional *Burkholderia cepacia* complex bacteria, may confound results of new molecular diagnostic tools. *FEMS Immunol Med Microbiol.* 2002;33:143–149.

# Analyses complémentaires de la structure génétique de la famille des facteurs $\sigma 70$ chez les Bcc: mise à jour et cas du clone B&B de *B. cenocepacia*

## 1. Introduction

Les résultats obtenus précédemment nous ont permis de mettre en évidence une grande instabilité au sein de la famille des facteurs transcriptionnels  $\sigma 70$  des *Burkholderia* du complexe *cepacia* (Bcc). Nous avons ainsi détecté des phénomènes d'insertion/délétion, et de duplication de gènes codant ces facteurs, au sein des génomes de ces bactéries. De plus, ces génomes ont présenté l'originalité d'abriter des facteurs primaires alternatifs, des formes n'ayant jamais été décrites auparavant chez les bactéries hors du groupe des actinobactéries [2]. Nous avons pu également déduire des stratégies évolutives différentes entre les Bcc et *P. aeruginosa* concernant la sélection de ces facteurs transcriptionnels; les *P. aeruginosa* ne présentant aucun sigma primaire alternatif mais une plus forte densité de gènes  $\sigma 70$  (3,67/Mb pour PAO1 et 2,86/Mb pour J2315) avec une grande partie de ceux-ci montrant une probable implication dans l'acquisition du fer. Les niches écologiques colonisées par ces deux microorganismes sont pourtant similaires voire identiques dans le cas des patients CF [9, 10]. Les Bcc ont la capacité de coloniser une multitude de niches écologiques (sol, plantes, eaux, rhizosphère, solutions antiseptiques, etc...), et la grande diversité de facteurs de la famille  $\sigma 70$  pourrait avoir contribué à l'émergence de ce caractère versatile. Les principaux travaux réalisés chez les Bcc décrivent l'implication des  $\sigma 70$  dans les processus d'expression des propriétés de virulence [11-13].

Dans cette seconde partie, nous avons réalisé une réactualisation des données concernant la famille  $\sigma 70$  chez les Bcc depuis notre publication de 2007. De plus, nous avons effectué des analyses complémentaires permettant d'intégrer les nouvelles données concernant le génome de *B. multivorans* ATCC17616 (nouvellement assemblé). Nous avons également développé une stratégie PCR/séquençage afin de vérifier la présence ou l'absence de certains gènes de cette famille  $\sigma 70$  au sein du clone B&B de *B. cenocepacia*. Nous avons cherché à vérifier certaines hypothèses émises dans la publication de 2007. Ces travaux nous ont permis d'identifier neuf nouveaux gènes susceptibles de coder des facteurs sigma70 dont des facteurs primaires alternatifs et des facteurs de type ECF (extracytoplasmic function). Ils nous ont également conduits à observer des insertions/délétions (indels) de gènes proximaux à certains gènes *ecf* tel qu'*ecfI*. Ces indels ont permis de confirmer la divergence entre le clone ET12 et les autres clones épidémiques de *B. cenocepacia*.

## 2. Résultats et discussion

### 2.1. Détection de nouveaux facteurs $\sigma 70$

Les nouvelles analyses réalisées par génomique comparative ont été effectuées selon la méthode décrite précédemment par Menard et al. [14]. Dans le but de détecter de nouveaux facteurs sigma, nous avons utilisé un concaténât de l'ensemble des séquences d'acides aminés déduites des gènes identifiés précédemment et additionné des séquences d'acides aminés déduites des bases de données (cf. matériels et méthodes). Ces analyses ont permis de détecter neuf nouveaux facteurs  $\sigma 70$  (Tableau 4, en gras). Nous avons ensuite étudié la distribution de ces facteurs/gènes en utilisant les génomes séquencés et assemblés. Nous avons ainsi pu déterminer que ces facteurs, absents chez J2315, étaient présents chez d'autres souches de *Burkholderia* : SigN, SigO, EcfU, EcfV, EcfW, EcfX chez *B. xenovorans* LB400 ; SigP chez *B. dolosa* AU0158 et EcfN, EcfO chez *B. sp383* (récemment renommé *B. lata* [15]). Certains de ces  $\sigma 70$  semblent être spécifiques à une espèce, c'est le cas des facteurs SigN, EcfV, EcfW et EcfX de *B. xenovorans* LB400, ceci pourrait indiquer une sélection spécifique suite à des phénomènes de duplication et dérive génétique ou pourrait indiquer une acquisition d'îlot génomique en provenance d'espèces éloignées. Les analyses par BLASTP/TBLASTN réalisées sur NCBI avec ces nouvelles séquences confirment la faible distribution de ces gènes chez les *Burkholderia* (e.g. section suivante). Cependant, des séquences correspondant à EcfN et EcfO ont respectivement été détectées de ces banques de données chez trois et deux souches appartenant au Bcc ; et, des séquences proches de EcfU ont été identifiées chez toutes les souches de *B. pseudomallei* testées. De plus, une séquence homologue à SigP a été détectée chez *B. thailandensis* E264 (Tableau 4, Fig. 7, Fig. 8). La densité des  $\sigma 70$  chez les souches analysées montre une relative conservation aux alentours de 2,2 $\sigma$ /Mb (de 1,65/Mb pour *B. multivorans* à 2,8/Mb pour *B. dolosa*). Ces densités restent en dessous de celle calculée pour *P. aeruginosa* (3,67/Mb). Les blasts réalisées en utilisant les génomes de PA14 et PAO1 ne nous ont pas permis de mettre en évidence de séquence homologue de ces neuf facteurs  $\sigma 70$ . Cependant les résultats des TBLASTN/BLASTP réalisés sur la banque de données GenBank ont permis de détecter un homologue d'EcfU chez *P. aeruginosa* 19660.



Tableau 4. Distribution des facteurs  $\sigma 70$  au sein des génomes de *Burkholderia* séquencés et chez *P. aeruginosa*

<i>Burkholderia</i> $\sigma^{70}$ factor <sup>♦</sup>	nom associé	<i>B. cenocepacia</i>	<i>B. vietnamiensis</i>	<i>B. multivorans</i>	<i>B. sp. 383</i>	<i>B. ambifaria</i>	<i>B. dolosa</i>	<i>B. mallei</i>	<i>B. pseudomallei</i>	<i>B. thailandensis</i>	<i>B. xenovorans</i>	<i>P. aeruginosa</i>
		G4	ATCC17616	AMMD	AU0158	E264	LB400					
<i>Bce-SigE</i>	RpoD	4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	2/2
<i>Bce-SigI</i>		4/4	1/1	1/1	0/1	1/1	1/1	0/4	0/6	0/1	0/1	0/2
<i>Bce-SigJ</i>		3/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	0/1	0/2
<b><i>Bxe-SigN</i></b>	<b>RpoD-like</b>	<b>0/4</b>	<b>0/1</b>	<b>0/1</b>	<b>0/1</b>	<b>0/1</b>	<b>0/1</b>	<b>0/4</b>	<b>0/6</b>	<b>0/1</b>	<b>1/1</b>	<b>0/2</b>
<b><i>Bxe-SigO</i></b>	<b>RpoD-like</b>	<b>0/4</b>	<b>0/1</b>	<b>0/1</b>	<b>1/1</b>	<b>0/1</b>	<b>0/1</b>	<b>0/4</b>	<b>0/6</b>	<b>0/1</b>	<b>1/1</b>	<b>0/2</b>
<i>Bce-SigL</i>	RpoH-like	4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	2/2
<i>Bce-SigM</i>	Flia-like	4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	2/2
<b><i>Bdo-SigP</i></b>	<b>Flia-like</b>	<b>0/4</b>	<b>0/1</b>	<b>0/1</b>	<b>0/1</b>	<b>0/1</b>	<b>1/1</b>	<b>0/4</b>	<b>0/6</b>	<b>1/1</b>	<b>0/1</b>	<b>0/2</b>
<i>Bce-SigK</i>	RpoS	4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	2/2
<i>Bce-EcfA1</i>	RpoE	4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	2/2
<i>Bce-EcfB</i>	PrtI-like	4/4	1/1	1/1	1/1	1/1	1/1	0/4	6/6	1/1	0/1	0/2
<i>Bce-EcfC</i>	FecI	4/4	0/1	1/1	1/1	0/1	0/1	4/4	6/6	1/1	0/1	2/2
<b><i>Bxe-EcfV</i></b>		<b>0/4</b>	<b>0/1</b>	<b>0/1</b>	<b>0/1</b>	<b>0/1</b>	<b>0/1</b>	<b>0/4</b>	<b>0/6</b>	<b>0/1</b>	<b>1/1</b>	<b>0/2</b>
<b><i>Bxe-EcfW</i></b>	<b>FecI</b>	<b>0/4</b>	<b>0/1</b>	<b>0/1</b>	<b>0/1</b>	<b>0/1</b>	<b>0/1</b>	<b>0/4</b>	<b>0/6</b>	<b>0/1</b>	<b>1/1</b>	<b>0/2</b>
<b><i>Bxe-EcfX</i></b>		<b>0/4</b>	<b>0/1</b>	<b>0/1</b>	<b>0/1</b>	<b>0/1</b>	<b>0/1</b>	<b>0/4</b>	<b>0/6</b>	<b>0/1</b>	<b>1/1</b>	<b>0/2</b>
<i>Bce-EcfD</i>		4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	0/2
<i>Bce-EcfE</i>		4/4	0/1	1/1	1/1	0/1	1/1	0/4	0/6	0/1	1/1	0/2
<i>Bce-EcfF</i>		4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	0/2
<i>Bce-EcfG</i>		4/4	0/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	0/2
<i>Bce-EcfH</i>		4/4	0/1	0/1	1/1	1/1	0/1	0/4	0/6	0/1	1/1	0/2
<i>Bce-EcfI</i>	OrbS	4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	2/2
<i>Bce-EcfJ</i>		4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	0/1	0/2
<i>Bce-EcfK</i>		4/4	0/1	0/1	1/1	0/1	1/1	4/4	6/6	1/1	0/1	0/2
<i>Bce-EcfL</i>		4/4	1/1	1/1	1/1	1/1	1/1	4/4	6/6	1/1	1/1	0/2
<i>Bce-EcfM</i>		4/4	1/1	1/1	1/1	1/1	1/1	0/4	0/6	0/1	0/1	0/2
<b><i>B383-EcfN</i></b>		<b>0/4</b>	<b>1/1</b>	<b>1/1</b>	<b>1/1</b>	<b>0/1</b>	<b>0/1</b>	<b>0/4</b>	<b>0/6</b>	<b>0/1</b>	<b>0/1</b>	<b>0/2</b>
<b><i>B383-EcfO</i></b>		<b>0/4</b>	<b>0/1</b>	<b>0/1</b>	<b>1/1</b>	<b>1/1</b>	<b>0/1</b>	<b>0/4</b>	<b>0/6</b>	<b>0/1</b>	<b>1/1</b>	<b>0/2</b>
<b><i>Bxe-EcfU</i></b>		<b>0/4</b>	<b>0/1</b>	<b>0/1</b>	<b>0/1</b>	<b>0/1</b>	<b>0/1</b>	<b>0/4</b>	<b>6/6*</b>	<b>0/1</b>	<b>1/1</b>	<b>0/2*</b>

La détection de nouveaux facteurs sigma70 (indiqués en gras) a été réalisée en TBLASTN entre le concaténat de tous les  $\sigma 70$  déjà décrits par Menard *et al.* (2007 Table1) et les génomes séquencés. Les génomes non référencés des souches analysés correspondent à: *B. cenocepacia*, souches I2315, AU1054, HI2424, PC184 (négative pour *Bcen-SigJ*); *B. mallei*, ATCC23344, 10229, GB8 horse 4, SAVP1; *B. pseudomallei*, 1710a, 1710b, S13, 1655, 668, K96243; *P. aeruginosa*, PAO1, PA14. \* détecté dans la banque de données GenBank. ♦ : Numéros d'accès des nouveaux  $\sigma 70$  : SigN = YP\_554902 ; SigO = YP\_552774 ; SigP = ZP\_04948391; EcfV = YP\_554415 ; EcfW = YP\_556268 ; EcfX = YP\_553645 ; EcfN = YP\_366282 ; EcfO = YP\_366672 ; EcfU = YP\_558841.

## 2.2. Classification fonctionnelle des nouveaux $\sigma 70$

Afin de connaître les catégories fonctionnelles de ces nouveaux facteurs nous avons utilisé l'arbre décisionnel Neighbor-Joining (NJ) contenant des séquences représentatives des différentes classes de  $\sigma 70$  [2]. Ces analyses ont permis de classer six des neuf nouveaux  $\sigma 70$  au sein de la famille des ECFs (EcfN, EcfO, EcfU, EcfV, EcfW et EcfX). Les trois autres  $\sigma 70$  ont été classés au sein des  $\sigma 70$  primaires alternatifs pour SigN et SigO et au niveau de FliA pour SigP (Tableau 4). La détection des différents domaines protéiques au sein des nouvelles séquences a été réalisée par analyse CDD (Conserved Domain Database [16]). SigO et SigN présentent les quatre domaines décrits au sein de la famille des  $\sigma 70$  [17], SigP, EcfU, EcfN, EcfW et EcfX présentent les domaines 2 et 4; EcfN ne présente que le domaine 2 et EcfO ne présente aucun de ces domaines bien que l'analyse CDD le classe dans la famille des  $\sigma 70$  de type ECF.

Ces séquences ont par la suite été intégrées aux jeux de données NON-ECF et ECF de Menard et al (2007) et analysées par les outils de la phylogénie moléculaire, les versions non condensées de ces deux arbres phylogénétiques sont disponibles sur le serveur du laboratoire d'écologie microbienne [18, 19].

## 2.3. Phylogénie moléculaire des facteurs $\sigma 70$ non-ECF

Les analyses phylogénétiques ont permis de déterminer que les séquences d'acides aminés (AA) déduites des loci *sigN* et *sigO* étaient significativement reliées au groupe des facteurs  $\sigma 70$  primaires et primaires alternatifs. La séquence d'AA déduite de *sigP* a été reliée au groupe des  $\sigma 70$  de type FliA (Figure 7). Ces séquences forment, cependant, de nouvelles lignées au sein de ces groupes. La séquence SigN déduite ne groupe avec aucune séquence des banques de données. SigO partage 50% d'identité avec SigN. SigO a pu être retrouvée chez *B. phytofirmans* et *Ralstonia eutropha* (*Cupriavidus necator*, [20]).

Les analyses réalisées par Menard et al. 2007 avaient montré l'absence des facteurs  $\sigma 70$  primaires alternatifs SigI et SigJ chez la souche *B. xenovorans* LB400. Ces nouvelles analyses ont permis de montrer la présence de facteurs primaires alternatifs chez *B. xenovorans* et autres *Burkholderia*. Il est fort probable que ces facteurs présentent une diversité plus large que celle observée à ce jour, et que leur niveau de diversité pourrait atteindre celui des facteurs de type ECF. Il est à noter qu'un locus codant une séquence similaire à SigI n'avait pu être détecté dans le génome de la souche *B. sp.383* (*B. lata*, [15]). Ces nouvelles analyses ont permis de détecter une séquence déduite similaire à SigO chez cette dernière souche (Tableau 4, Fig. 7). Ces données présagent d'incompatibilité probable entre ces facteurs primaires alternatifs e. g. SigO ne pourrait être présent dans une cellule exprimant SigI. Cette hypothèse est confortée par le fait que le génome de *B. phytofirmans* abrite une copie de SigO et ne contient pas de copie de SigI, comme observé précédemment pour *B. xenovorans* et *B. sp383*. La fonction de ces facteurs

primaires alternatifs n'est cependant pas connue. Ces analyses ont également permis d'observer pour la première fois la présence de deux lignées distinctes de facteurs de type FliA chez les *Burkholderia*, le type classique appelé SigM et le type SigP. Ces paralogues sont fortement différenciés et le locus sigP aurait été acquis probablement par transferts horizontaux. Cependant, la présence du type SigP chez plusieurs espèces de *Burkholderia* suggère un maintien de ce locus au sein des populations de ces espèces, avec une probable incompatibilité avec la forme SigM.

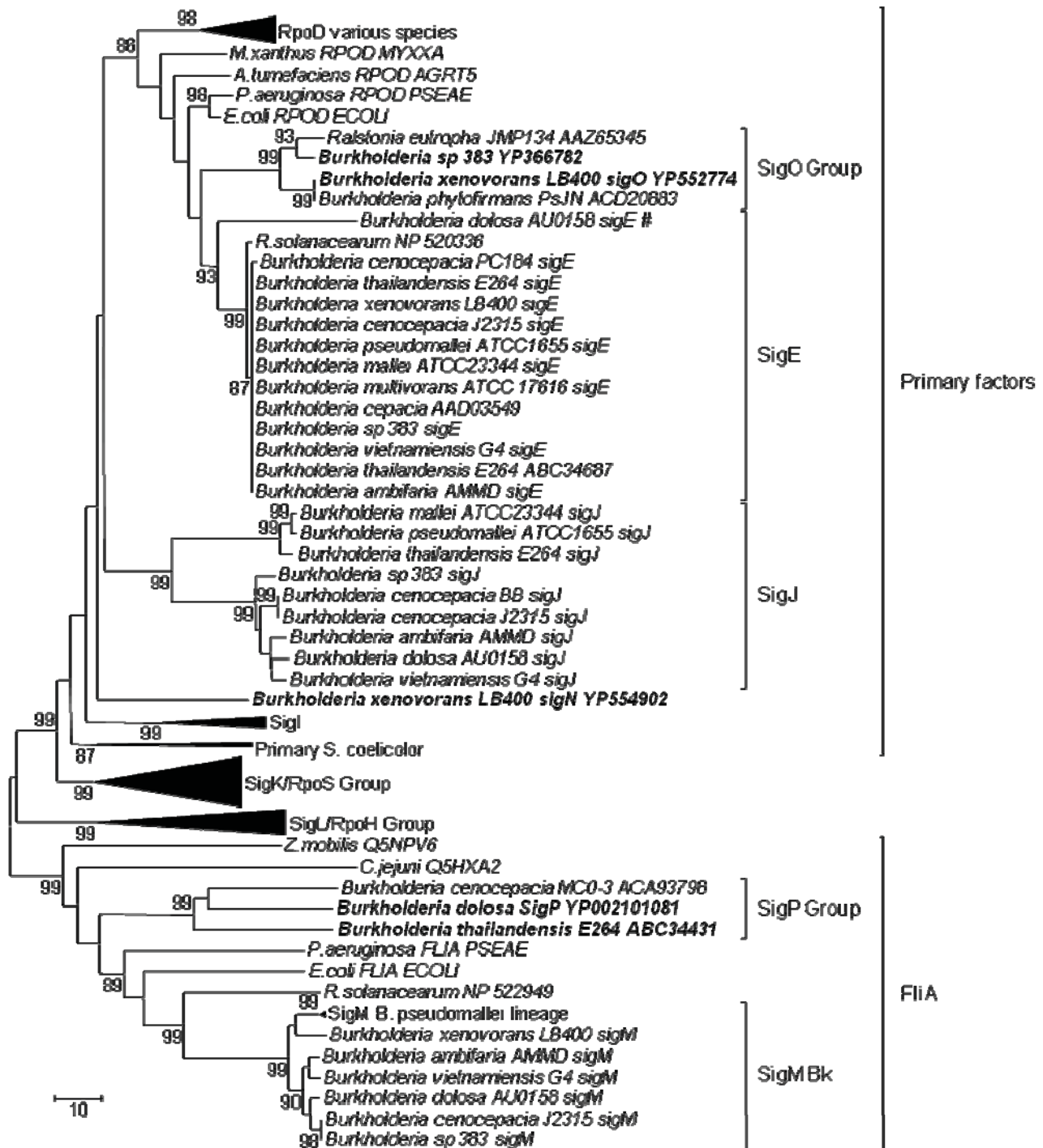
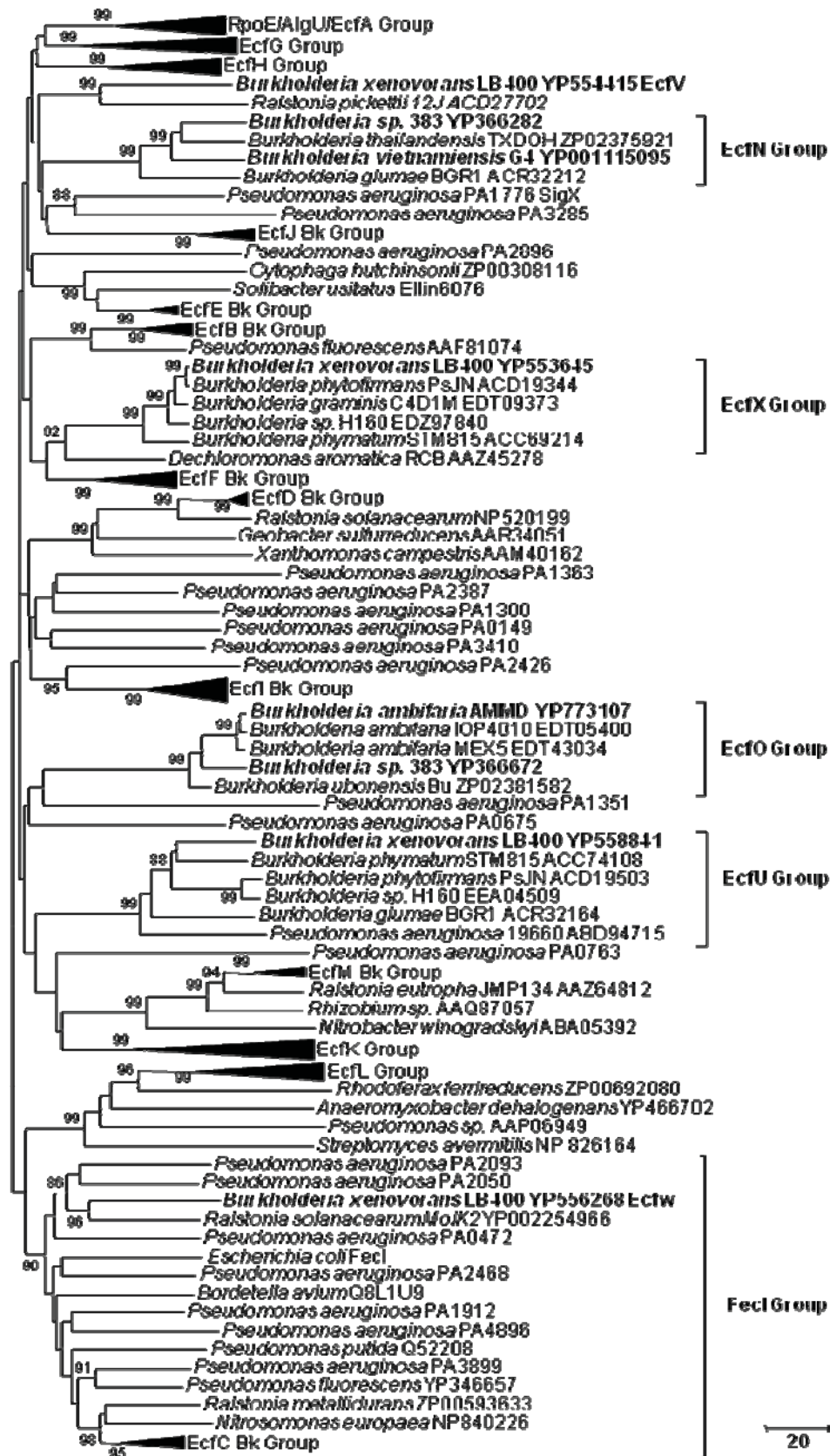


Figure 7. Arbre phylogénétique en Neighbor-Joining des facteurs  $\sigma_{70}$  non-ECF

Les séquences acides aminées  $\sigma_{70}$  ont été obtenues dans les bases de données Genbank et à partir des travaux de Lavire *et al.* [2]. Un total de 150 sites ont été analysés (avec une délétion globale des sites contenant des gaps). La distance est proportionnelle à la divergence évolutive et exprimée en substitution pour 100 sites. Seules les valeurs de bootstrap supérieure à 85% sont indiquées. Les sous groupes condensés sont visibles dans la Figure 3. Bk : correspond à des séquences appartenant au genre *Burkholderia* uniquement et condensées pour plus de clarté. # : correspond à la séquence SigE de *B. dolosa* présentant un décalage de cadre sûrement du à une erreur de séquençage et toujours pas corrigé à ce jour. Les alignements sont disponibles sur le serveur de l'umr5557, [http://ecomicro.univ-lyon1.fr/poe/Non-ECF\\_align\\_these.txt](http://ecomicro.univ-lyon1.fr/poe/Non-ECF_align_these.txt).

#### 2.4. Phylogénie moléculaire des facteurs $\sigma 70$ ECF

Les nouvelles séquences ECF, ainsi que les séquences plus proches identifiées dans les bases de données ont été combinées aux jeux de données ECF de Menard et al. (2007). Ce nouveau jeu de données a été analysé par phylogénie moléculaire (Figure 8). Quatre nouveaux groupes supportés par des valeurs élevées de bootstrap ont été obtenus (EcfN, EcfX, EcfO et EcfU) et deux séquences (EcfV et EcfW) se retrouvent dans des groupes non-résolus. La séquence EcfV déduite groupe avec un ECF retrouvé chez *Ralstonia pickettii* dont la fonction n'a pas été décrite pour l'instant. Le facteur putatif EcfW de *B. xenovorans* LB400 se retrouve dans le cluster des facteurs de type FecI et pourrait donc être impliqué dans les processus d'acquisition du Fer, comme décrit précédemment. Cependant la séquence la plus proche retrouvée chez *R. solanacearum* MolK2 correspond au facteur ECF PhrI (YP\_002254966) qui est impliqué dans l'activation du système de sécrétion de type III lors du contact de *R. solanacearum* avec la plante [21]. Bien que *B. xenovorans* LB400 ne soit pas connue comme une bactérie phytopathogène stricte, cette souche a été isolée de la rhizosphère de certaines plantes, de plus l'analyse de son génome démontre la présence de gènes pouvant être impliqués dans l'infection de plantes (hemagglutinines, adhesines, système de sécrétion de type III et IV) [22]. Le cluster EcfN regroupe des séquences putatives uniquement déduites des génomes de *Burkholderia*. Le cluster EcfX regroupe des séquences déduites des génomes de quatre souches de *Burkholderia* et d'une souche de *Dechloromonas aromatica* (genre *Rhodocyclaceae*,  $\beta$ -protéobactérie). La fonction potentielle de ces ECF n'a pas pu être déduite. Cependant, cet ECF putatif semble restreint aux  $\beta$ -protéobactéries. Des séquences d'orthologues putatifs d'EcfO ont été retrouvées strictement chez des espèces du Bcc. EcfU n'a pas été détecté chez les Bcc mais des homologues ont été détectés chez *P. aeruginosa* et plusieurs espèces de *Burkholderia*.



**Figure 8. Arbre phylogénétique en Neighbor-Joining des facteurs  $\sigma_{70}$  de type ECF**

Les séquences acides aminées  $\sigma_{70}$  ont été obtenues dans les bases de données Genbank et à partir des travaux de Lavire *et al.* [2]. Un total de 540 sites ont été analysés (avec une déletion par paire des sites contenant des gaps). La distance est proportionnelle à la divergence évolutive et exprimée en substitution par 100 sites. Seules les valeurs de bootstrap supérieure à 85% sont indiquées. Les sous-groupes condensés sont visibles dans la Figure 4. Bk : correspond à des séquences appartenant au genre *Burkholderia* uniquement et condensées pour plus de clarté. Les alignements sont disponibles sur le serveur de l'umr5557, [http://ecomicro.univ-lyon1.fr/poe/ECF\\_align\\_these.txt](http://ecomicro.univ-lyon1.fr/poe/ECF_align_these.txt).

## 2.5. Distribution et environnement génétique de gènes *ecf* chez les Bcc

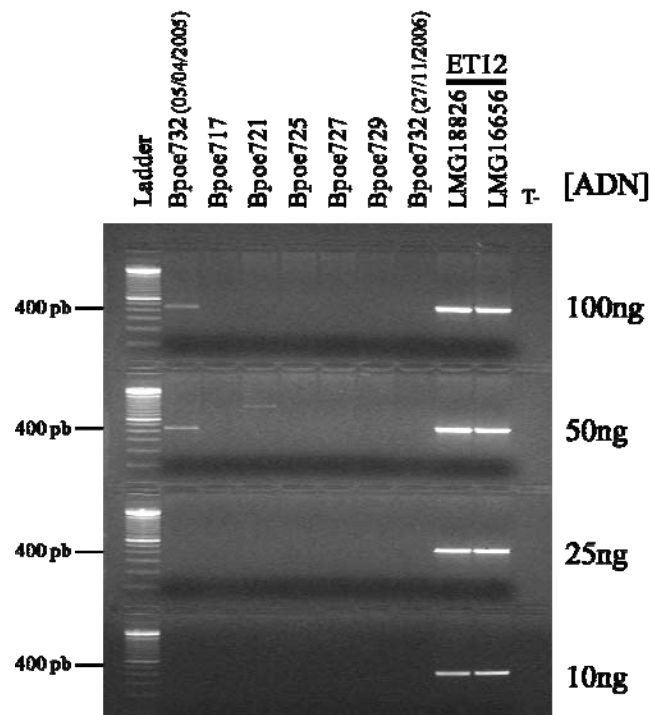
Des cribles PCR ont été élaborés pour préciser la distribution des gènes *ecf* détectés chez *B. cenocepacia* J2315 et ayant montré une certaine instabilité génétique ou une répartition espèce ou souche spécifique par génomique comparative. Ces cribles ont été utilisés pour étudier la distribution des gènes *ecfA*, *ecfB*, *ecfC*, *ecfD*, *ecfE*, *ecfG*, *ecfH*, *ecfI*, et *ecfK* chez les 3 espèces du Bcc retrouvées les plus souvent au sein des poumons de patients atteints de mucoviscidose (*B. cenocepacia*, *B. multivorans* et *B. vietnamiensis*) [23]. 10 souches de *B. cenocepacia* représentant les quatre groupes RFLP *recA* de cette espèce en incluant le clone B&B ; 5 souches de *B. multivorans* et 6 souches de *B. vietnamiensis* ont été étudiées. Les résultats de ces cribles sont présentés dans le Tableau 5.

### 2.5.1. *ecfA*

Comme décrit précédemment par Menard et al. 2007, un réarrangement génétique majeur (inversion d'une partie du chr1 de J2315) a conduit à la duplication d'une séquence de 40 kb contenant le gène *ecfA* (Figure 6). Ce phénomène semblait restreint à certaines souches de la lignée ET12 chez *B. cenocepacia*. Les résultats obtenus par PCR et séquençage en utilisant les amorces A, B, et C (Tableau 6) sur un ensemble de souches et d'espèces de Bcc plus important ont confirmé l'absence de ce réarrangement chez toutes les souches de *B. multivorans* et *B. vietnamiensis* testées. Les analyses complémentaires sur les souches de *B. cenocepacia* appartenant ou non à la lignée ET12 ont également confirmé les données précédentes mais une certaine forme d'instabilité du phénomène d'inversion/duplication a été observée (Tableau 5). Les deux souches J2315 et LMG18826 appartenant à la lignée ET12 ont été détectées avec une configuration inversion/duplication du locus *ecfA* (*ecfA2* positif en utilisant les amorces B/C de Menard et al., 2007). La majorité des souches n'appartenant pas à la lignée ET12 a présenté une configuration sans inversion/duplication (*ecfA2* négatif ; produits PCR avec les amorces A/B de Menard et al., 2007). Cependant, les quatre dernières souches, LMG12614 (III appartenant à la lignée ET12), LMG21461 (IIID), LMG19240 (IIIC) et le clone B&B (IIIA) ont présenté des résultats de crible PCR surprenant et suggérant la présence des deux configurations génétiques, avec ou sans inversion, au sein des cultures testées (*ecfA2* positif - amorces B/C, et négatif - amorces A/B). Le séquençage des produits PCR a confirmé la présence des deux configurations chez ces quatre souches (Annexe B1), cependant l'intensité des produits PCR de la région *ecfA2* était beaucoup plus faible que celle des produits PCR indiquant son absence (résultats non montrés). Ces analyses de séquence ont donc confirmé l'hypothèse de la présence de deux populations au sein des cultures testées. Les ADN totaux utilisés pour les criblages PCR sont obtenus à partir de cultures pures réalisées en milieu liquide. Pendant cette phase de croissance des réarrangements génétiques

peuvent avoir lieu et pourraient donc aboutir à la présence des deux populations, l'une d'elles présentant l'inversion (*ecfA2* positive) et l'autre ne présentant pas l'inversion (*ecfA2* négative).

Dans le but de connaître la quantité relative des deux populations, nous avons réalisé des PCR *ecfA2* en utilisant différentes quantités d'ADN des souches du clone B&B dont les extractions d'ADN avaient été réalisées à différentes dates. Nous avons ainsi pu détecter *ecfA2* uniquement chez une des souches du clone B&B, la souche bpoe732, et seulement à partir de l'ADN extrait en 2005 (Fig. 9). L'utilisation de différentes quantités d'ADN nous a permis d'avoir une estimation relative des deux populations au niveau de cet ADN ; les PCR avec les amorces A/B (*ecfA2* absent) ont été toujours positives jusqu'à la plus petite dilution testée (10 ng d'ADN) pour toutes les souches du clone B&B (résultats non montrés) ; les PCR *ecfA2* (amorces B/C) ont été négatives à partir de 25 ng d'ADN. Un facteur >2,5 en faveur d'une absence d'*ecfA2* semble donc s'appliquer aux concentrations observées. Cependant, l'extraction réalisée en 2006 de cette même souche ainsi que toutes les autres souches appartenant au clone B&B n'ont présenté qu'une population *ecfA2* négative (Fig. 9). L'inversion pourrait donc uniquement avoir eu lieu pendant la phase de croissance de la souche bpoe732, avant extraction d'ADN, en 2005. Cet événement serait donc rare. Une autre explication pourrait être que toutes les souches du clone B&B présenteraient cette inversion au départ et l'aurait "perdu" pendant leur croissance en milieu liquide, mais ceci semble moins probable. Ce type de réarrangement génétique pourrait avoir un effet sur la virulence comme cela a déjà été montré chez un clone de *P. aeruginosa* [24]. Des phénomènes similaires impliquant des séquences d'insertion ont déjà été décrits chez les *Burkholderia* et *P. aeruginosa* [14, 25, 26].



**Fig. 9** Crible PCR de la région *ecfA2* à partir d'extrait d'ADN dilué de *B. cenocepacia*. Voir tableau 5 pour l'origine des souches. Clone B&B= Bpoe717, 721, 725, 727, 729 et 732. Les dates indiquées pour Bpoe732 correspondent aux dates d'extraction d'ADN.



Tableau 5. Souches de *Burkholderia* utilisées dans cette étude. Criblages par analyse des génomes\*, PCR/séquençage<sup>#</sup> et/ou Southern-blot<sup>o</sup>

Souches	Refs	<i>ecfA2</i>	<i>ecfB</i>	<i>ecfC</i>	indel- <i>ecfC</i> -viet	<i>ecfD</i>	<i>ecfE</i>	région <i>ecfE</i> -Been	phage- <i>ecfE</i> -viet	<i>ecfG</i>	<i>ecfH</i>	<i>ecfI</i>	indel- <i>ecfI</i> -PHDC	indel- <i>ecfI</i> -viet	<i>ecfK</i>	indel- <i>ecfK</i> -viet	<i>sigI</i>	<i>sigJ</i>	
<i>B. cenocepacia</i>																			
J2315 <sup>#*</sup>	[27]	+	+	+	-	+	+	+	-	+	+	+	-	-	+	-	+	+	
HI2424 <sup>*</sup>	[28]	-	+	+	-	+	+	+	-	+	+	+	+	-	+	-	+	+	
AU1054 <sup>*</sup>	[29]	-	+	+	-	+	+	+	-	+	+	+	+	-	+	-	+	+	
PC184 <sup>*</sup>	[30]	-	+	+	-	+	+	+	-	+	+	+	+	-	+	-	+	+	
HI11 <sup>#</sup>	[31]	-	+	+	-	+	+	+	-	+	+	+	-	-	+	-	+	+	
LMG 18828 <sup>#</sup>	[32]	-	+	+	-	+	+	+	-	+	+	+	-	-	+	-	+	+	
LMG 18863 <sup>#</sup>	[33]	-	+	+	-	+	+	+	-	+	+	+	-	-	+	-	+	+	
LMG 18826 <sup>#</sup>	[34]	+	+	+	-	+	+	+	-	+	+	+	-	-	+	-	+	+	
LMG 12614 <sup>#</sup>	[35]	+	+	+	-	+	+	+	-	+	+	+	-	-	+	-	+	+	
LMG 21461 <sup>#</sup>	[36]	+	+	+	-	+	+	+	-	+	+	+	+	-	+	-	+	+	
LMG 19240 <sup>#</sup>	[37]	+	+	+	-	+	+	+	-	+	+	+	-	-	+	-	+	+	
LMG 16654 <sup>#</sup>	[37]	-	+	+	-	+	+	+	-	+	+	+	+	-	+	-	+	+	
LMG 18832 <sup>#</sup>	[34]	-	+	+	-	+	+	+	-	+	+	+	+	-	+	-	+	+	
B&B clone <sup>#o</sup>	[8]	+	+	+	-	+	+	+	-	+	+	+	-	-	+	-	+	+	
<i>B. multivorans</i>																			
LMG 13010 <sup>#o</sup>	[38]	-	+	+	-	+	+	-	-	+	-	+	-	-	-	-	+	+	
LMG 18822 <sup>#o</sup>	[38]	-	+	+	-	+	+	-	-	nd	-	+	-	-	-	-	+	+	
LMG 16665 <sup>#</sup>	[38]	-	+	+	-	+	+	-	-	nd	-	+	-	-	-	-	+	+	
Bpoe 734 <sup>#</sup>	[8]	-	+	+	-	+	+	-	-	nd	-	+	-	-	-	-	+	+	
ATCC 17616 <sup>*</sup>	[39]	-	+	+	-	+	+	-	-	+	-	+	-	-	-	-	+	+	
<i>B. vietnamiensis</i>																			
LMG 10929 <sup>#</sup>	[40]	-	+	+	+	+	-	-	+	-	-	+	-	+	-	+	+	+	
G4 <sup>*</sup>	[40]	-	+	+	+	+	-	-	+	-	-	+	-	+	-	+	+	+	
LMG 16232 <sup>#</sup>	[40]	-	+	+	+	+	-	-	+	-	-	+	-	+	-	+	+	+	
LMG 16228 <sup>#</sup>	[40]	-	+	+	+	+	-	-	+	-	-	+	-	+	-	+	+	+	
LMG 10823 <sup>#</sup>	[40]	-	+	+	+	+	-	-	+	-	-	+	-	+	-	+	+	+	
Bpoe 736 <sup>#</sup>	[8]	-	+	+	+	+	-	-	+	-	-	+	-	+	-	+	+	+	
Bpoe 737 <sup>#</sup>	[8]	-	+	+	+	+	-	-	+	-	-	+	-	+	-	+	+	+	

Les Southern Blot ont été réalisés suivant le protocole décrit dans la section matériels et méthodes. Les amorces utilisées pour les cribles PCR sont décrites dans le Tableau 6. nd: non déterminé. (+) : signal positif, (-) : signal négatif.

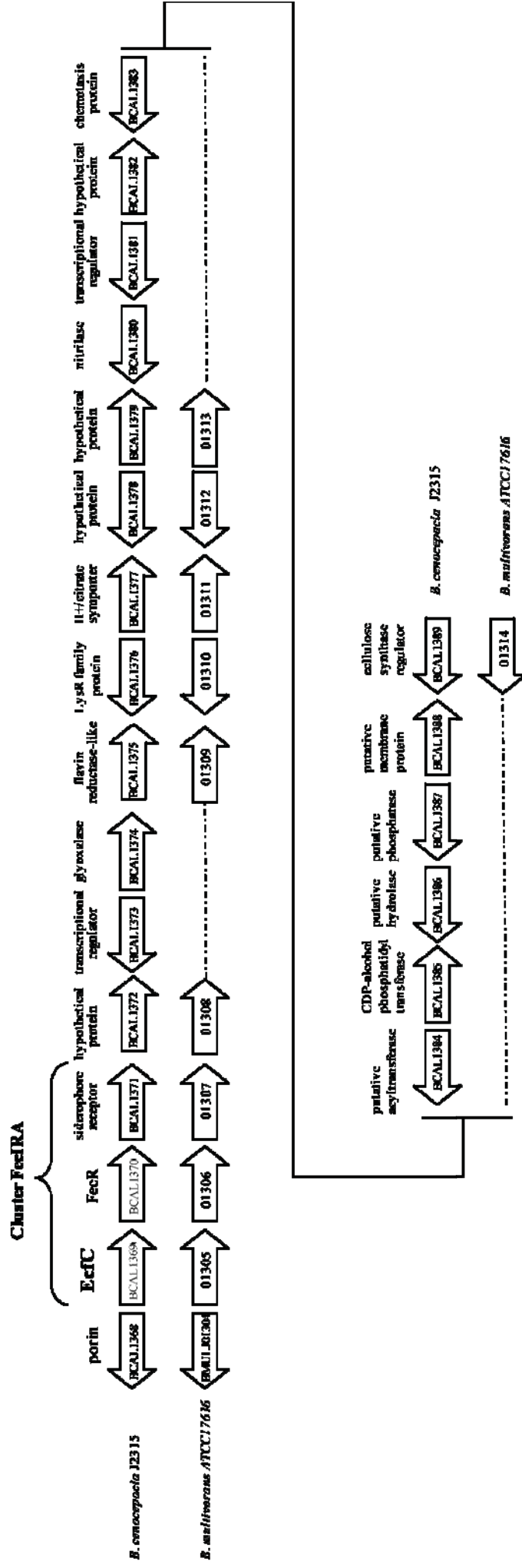
### 2.5.2. *ecfC*

Le gène *ecfC* (FecI-like), potentiellement impliqué dans l'acquisition du Fer chez *B. cenocepacia*, montrait une distribution hétérogène au niveau de ce genre et a été relié à une région d'instabilité génétique (Tableau 4). En utilisant les différentes copies d'*ecfC* détectées au sein des génomes séquencés de Bcc, des amorces PCR permettant l'amplification de ce gène (*ecfC*-F, *ecfC*-R, Tableau 6) ont été définies. Une large région de 14 kb contenant ce gène étant absente chez *B. vietnamiensis*, des amorces PCR permettant de déduire l'absence/présence de cette région ont, également, été définies (région *ecfC*, Tableau 6). Les résultats obtenus sont présentés dans le Tableau 5 et l'annexe B9. Ils confirment les données obtenues par Menard et al. 2007. Toutes les souches de *B. vietnamiensis* n'ont pas montré de copie d'*ecfC* dans leur génome. De plus, le criblage par PCR du locus *ecfC* en utilisant les régions flanquantes a confirmé cette absence d'*ecfC* chez toutes les souches testées de cette espèce (*indel-ecfC-viet* Tableau 5). Ces résultats ont été validés par analyse des séquences d'ADN des fragments PCR amplifiés. La délétion de ce locus semble s'être produite très tôt dans la radiation évolutive des *B. vietnamiensis*. Le génome de *B. multivorans* étant maintenant assemblé, la synténie de cette région entre *B. cenocepacia* et *B. multivorans* a également pu être comparée. Les résultats obtenus sont présentés dans la Figure 10. Une organisation génétique différente de celles décrites à ce jour pour cette région chez les génomes de Bcc a pu être observée. Une délétion de 5 gènes additionnels en amont du gène BCAL1383 a été détectée dans le génome de *B. multivorans*. Cependant, chez cette dernière bactérie, en opposition aux situations observées pour les génomes de *B. ambifaria* et *B. dolosa*, un cluster de type *fecIRA*-like complet a pu être observé. Les protéines FecI et FecR pourraient donc s'associer et permettre l'activation de l'opéron débutant classiquement par le gène *fecA*. FecA est un activateur des gènes *fecIR* lorsqu'il y a une présence du complexe Fe(III)-citrate. Les analyses de synténie de cette région chez les Bcc (Fig. 5 et Menard et al. 2007) suggère que l'opéron régulé par le système EcfC-FecR pourrait s'étendre jusqu'au locus BCAL1372 ou BCAL1379 du génome de *B. cenocepacia* J2315. Une démonstration expérimentale est cependant nécessaire pour préciser ce point. Il est important de noter que le gène *ecfC* a été détecté pour l'ensemble des souches de *B. multivorans* testées (tableau 5, annexe B11).

### 2.5.3. *ecfE*

Les analyses réalisées précédemment suggéraient une instabilité de la région contenant le gène *ecfE* chez les souches *Burkholderia lata* sp. 383 et chez *B. vietnamiensis* G4 (Figure 5). Afin de confirmer ces résultats, des stratégies PCR ont été mises en place pour les espèces *B. cenocepacia* et *B. vietnamiensis*. La première a consisté à amplifier la région comprise entre la CDS BCAL2459 et le gène *ecfE* chez les souches de *B. cenocepacia* (amorces BCAL2459 et *ecfE*-R,

Tableau 6 et Figure 11a). La deuxième stratégie a consisté à amplifier la bordure droite du site d'intégration du phage chez *B. vietnamiensis* (amorces BCAL2466vietF et BCAL2466vietR, Tableau 6 et Figure 11a). Les résultats obtenus ont démontré une conservation de la région contenant les CDS BCAL2459 à BCAL2462 chez toutes les souches de *B. cenocepacia* testées (Tableau 5, Annexe B2). Concernant l'intégration d'un phage chez *B. vietnamiensis* au niveau de cette région, les résultats obtenus par PCR et séquençage démontrent une conservation de cette intégration chez toutes les souches testées et pourrait donc représenter être un événement spécifique de l'espèce *B. vietnamiensis* (Tableau 5, annexe B3 et annexe B10). L'intégration de ce phage semble donc avoir eu lieu tôt dans l'évolution des *B. vietnamiensis*. Cependant l'hypothèse émise par Menard et al. 2007 selon laquelle l'absence d'*ecfE* chez l'espèce *B. multivorans* aurait été relié à l'intégration de ce même phage a été infirmée. En effet, les nouvelles analyses réalisées sur le génome de la souche *B. multivorans* ATCC17616 ont montré l'absence de la région de *B. cenocepacia* J2315 codant *ecfE* chez *B. multivorans* (Figure 11a), cependant une copie d'*ecfE* a été détectée ailleurs dans le génome, au niveau d'une région présente sur le chromosome 1 de la souche séquencée de *B. multivorans* (Figure 11b). Ce résultat a été confirmé par PCR, toutes les souches de *B. multivorans* testées présentent une copie de l'*ecfE* par un crible PCR redéfini en fonction des séquences de *B. multivorans* (Tableau 5). La région contenant *ecfE* chez *B. multivorans* ATCC17616 contient divers gènes impliqués dans la régulation, le métabolisme général ainsi que de nombreuses protéines putatives de fonction encore inconnue. Cette région ne partage aucun locus avec la région *ecfE* de *B. cenocepacia*. L'analyse par BLASTN montre une absence de cette région de 21 kb chez les autres espèces de *Burkholderia*. Il est intéressant de noter que cette région spécifique de *B. multivorans* ATCC17616 est remplacée par un îlot génomique, IG6, chez *B. cenocepacia* J2315 (chapitre 4 Figure 11b). Cette région pourrait être instable en raison de la présence d'un gène codant un ARNt en bordure ; ces gènes peuvent jouer un rôle dans l'intégration de phages, plasmides et îlot génomiques chez les bactéries [41-45]. Un crible PCR pour suivre la région *ecfE*-IG6 a permis de déduire sa présence chez toutes les souches (n=14) de *B. cenocepacia* testées (Tableau 5, annexe B12 et annexe B13) mais n'a pas permis sa détection chez les autres espèces testées. Cependant les analyses par BLASTN démontrent la présence de cette région au sein des génomes de *B. phytofirmans* PsJN et *B. lata* sp.383. Les sigmas EcfE *B. multivorans* et *B. cenocepacia* jouent fort probablement des rôles différents chez ces bactéries malgré une faible divergence au niveau de leurs séquences.



**Figure 10: Analyse de la synténie de la région *ecfC* entre *B. cenocepacia* strain J2315 et *B. multivorans* ATCC17616**

Le génome de *B. cenocepacia* J2315 a été comparé au génome de *B. multivorans* ATCC17616 nouvellement assemblé en utilisant le logiciel ACT. Les ORF de J2315 ont été annotées d'après la publication de ce génome par Holden et al. 2009 [27]; l'annotation du génome de *B. multivorans* a été réalisée selon en suivant les code des protéines sur le site NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Les régions non retrouvées chez *B. multivorans* sont indiquées par des pointillés. Le facteur  $\sigma70$  *ecfA* est indiqué en gras.

#### 2.5.4. *ecfI*

Les analyses précédentes avaient démontré une délétion de la région située en amont du gène *ecfI* chez les souches *B. cenocepacia* AU1054 et HI2424 et une délétion plus importante de cette région chez la souche *B. vietnamiensis* G4 (Figure 5). Afin de déterminer la conservation de ces deux phénomènes chez d'autres souches de *B. cenocepacia* et *B. vietnamiensis* nous avons dessiné des amorces PCR de part et d'autres de ces délétions. Les amorces *ecfI*-R couplée à BCAL1680F ont été utilisées pour déterminer si la région absente chez les deux souches PHDC pouvait l'être chez d'autres souches de *B. cenocepacia*, et les amorces *indel-ecfI-vietF* et *indel-ecfI-vietR* ont été utilisées pour déterminer si la région absente chez la souche G4 l'était aussi chez les autres souches de *B. vietnamiensis* (Tableau 6). Les séquençages et analyses ACT obtenus pour les souches de *B. vietnamiensis* démontrent une conservation de cette délétion chez toutes les souches testées (Tableau 5, annexe B4 et annexe B8), ce qui confirme les résultats obtenus précédemment. Les analyses ACT réalisées à l'aide des produits PCR séquencés et des séquences extraites des génomes de PHDC ont montré des situations différentes selon les souches (Tableau 5, Annexe B5) et une différence avec la synténie décrite en Figure 5. En effet, les CDS BCAL1686 et BCAL1687 non détectés lors des travaux précédents ont été retrouvés chez ces deux souches à une position en accord avec l'annotation de *B. cenocepacia* J2315 (Figure 12). L'analyse du génome de *B. multivorans* ATCC17616 pour cette région a permis d'observer cette même organisation (Fig. 12). Cependant, ces loci ont été confirmés comme absent de la région correspondante du génome de *B. vietnamiensis* G4. Concernant la délétion en amont de ces gènes chez les souches PHDC, les analyses ACT des séquences obtenues à partir des fragments PCR démontrent une hétérogénéité pour ce phénomène de délétion (Tableau 5, colonne *indel-ecfI*-PHDC); les souches *B. cenocepacia* PC184, LMG21461, LMG16654 et LMG18832 sont les seules souches à présenter cette délétion. Les autres souches testées sont négatives pour cette délétion, cependant nous ne pouvons pas savoir si ces souches présentent la même synténie que J2315. Il apparait tout de même que cette délétion est partagée par des souches de complexes clonaux distincts, ce qui suggère une instabilité de cette région. Cependant, aucun gène associé à de l'instabilité génomique comme les gènes codant les ARNt, les IS et aucune divergence dans le biais en %GC n'ont pu être observé pour cette région.

#### 2.5.5. *ecfK*

Les analyses précédentes démontraient une délétion de la région contenant le gène *ecfK* chez les souches *B. vietnamiensis* G4 et *B. ambifaria* AMMD (Figure 5). Dans le but de

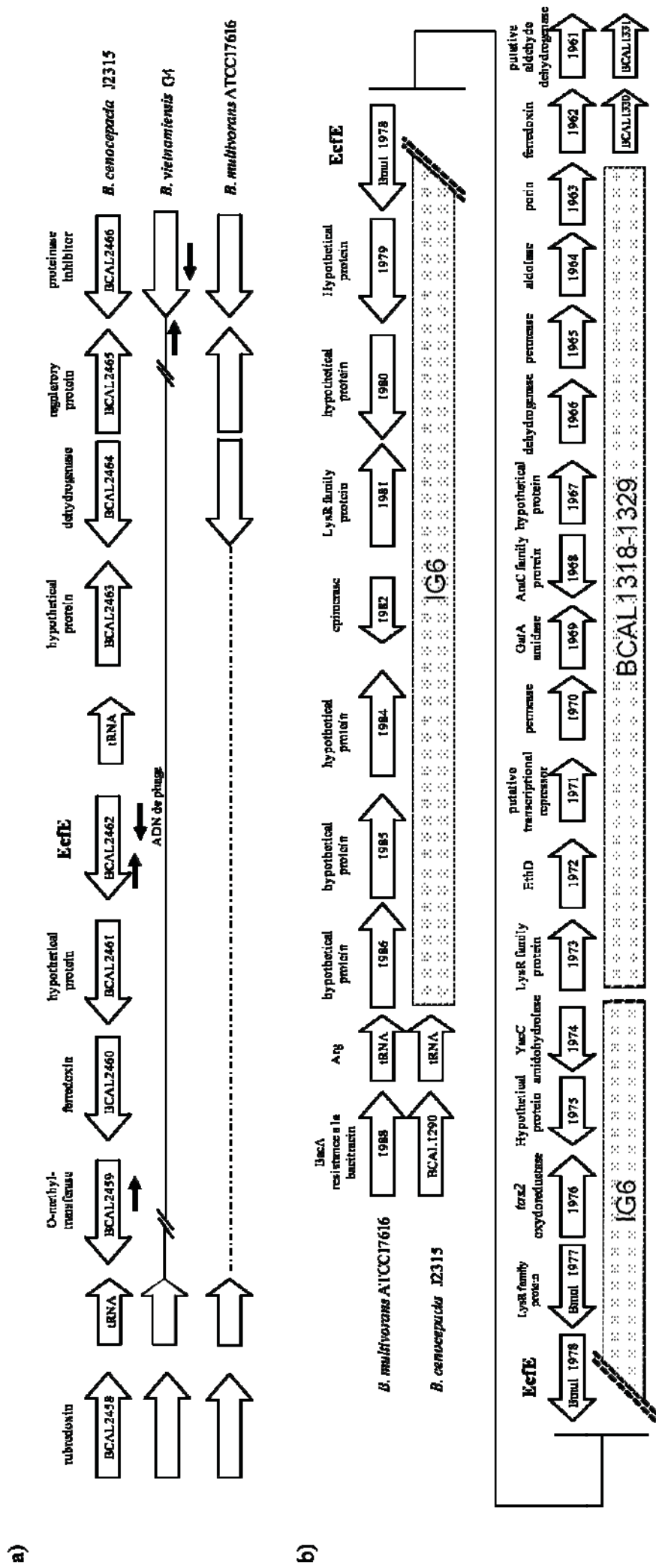
connaître la distribution de ce phénomène chez l'espèce *B. vietnamiensis*, nous avons mis au point un crible PCR utilisant des amorces de part et d'autre de cette séquence délétée (amorces BCAM2746 et BCAM2755, Tableau 6). Les résultats obtenus par séquençage et analyses ACT confirment la présence de cette délétion chez toutes les souches de *B. vietnamiensis* (Tableau 5 et annexe B6 et annexe B7). La conservation de cette même délétion mise en évidence précédemment chez *B. ambifaria* AMMD supporte l'hypothèse d'une perte de cette région chez ces deux espèces avant leur divergence. Les analyses de synténie de la région contenant le gène *ecfK* ont par ailleurs été réalisées chez les deux souches *B. cenocepacia* AU1054 et HI2424 ainsi que chez la souche *B. multivorans* ATCC17616. Concernant les deux génomes de *B. cenocepacia*, les analyses ont révélé une insertion d'un fragment d'ADN de 40 kb en amont des loci BCAM2747 et *ecfK* (Figure 13). Cette région ou IG contient des CDS dont certains ont des homologies avec des CDS de transposases, d'intégrase de phage, et d'antigènes de surface. Un biais dans le %GC a pu être mis en évidence, avec un %GC de 62% au lieu des 67% attendus pour les deux génomes de *B. cenocepacia* analysés. Deux des trois CDS de transposases ont été positionnés aux extrémités de cette région, et aurait pu contribuer à l'intégration de cette séquence de 40 kb. L'une des ces transposase est l'IS1356-c, et correspond à une cible décrite pour le suivi du clone épidémique ET12 [46]. Une analyse plus poussée de cette séquence d'insertion est présentée dans le chapitre suivant. Les deux CDS codant des antigènes pourraient jouer un rôle dans l'attachement de cette bactérie aux cellules eucaryotes. Les analyses ACT ont confirmé l'absence de cette région dans le génome de *B. cenocepacia* J2315 ce qui suppose une intégration de ce fragment d'ADN après la divergence de ces deux complexes clonaux. Concernant les analyses de synténie réalisées entre les génomes de *B. cenocepacia* J2315 et *B. multivorans* ATCC17616, une séquence de 8,5 kb insérée chez la souche ATCC17616 a été mise en évidence à la place de la région *ecfK* de J2315 (Figure 13). Les gènes présents au niveau de cette région pourraient être impliqués dans la dégradation du benzoate (Figure 14). Un cluster de gènes similaire a été mis en évidence au niveau du chromosome 3 de la souche *B. cenocepacia* J2315 (CDS BCAS0492 à BCAS0498). Cependant, le gène *benD* (BCAS0493) de J2315 présente un décalage de cadre de lecture après le codon 180, conduisant à une protéine non fonctionnelle. Bien qu'il existe d'autres voies métaboliques permettant la dégradation du benzoate, notamment via la CoA ligation ou via le métabolisme de la Tyrosine, seule la voie de dégradation via l'hydroxylation de la souche de *B. multivorans* ATCC17616 apparaît fonctionnelle.

### 2.5.6. *ecfB*, *ecfD*, *ecfG*, *ecfH*, *sigI* et *sigJ*

Des tests PCR pour le suivi des gènes *ecfB*, *ecfD*, *ecfG*, *ecfH*, *sigI* et *sigJ* ont été développées pour tester leur présence chez le clone B&B de *B. cenocepacia*, et vérifier les tendances observées par Menard et al. (2007) concernant leur distribution au sein du Bcc (Tableau 6). Ces cribles PCR ont permis de détecter ces gènes chez le clone B&B. Ces détections sont en accord avec les tendances observées à ce jour chez les *B. cenocepacia*. Cependant, *sigJ* n'a pu être détecté chez la souche *B. cenocepacia* LMG21461. Cette souche apparaissait proche de la lignée ET12 lors de l'étude de l'allèle *ecfA2*, mais l'absence de *sigJ* donne à cette souche une configuration uniquement observée à ce jour chez la souche PC184 du clone épidémique Midwest. Ceci nécessitera une clarification par MLST. En ce qui concerne les cribles effectués sur les souches de *B. vietnamiensis*, ils ont validé la présence d'*ecfB*, *ecfD*, *sigI* et *sigJ*, et confirmé l'absence d'*ecfG* et *ecfH* chez cette espèce. En ce qui concerne *B. multivorans*, la présence d'*ecfB*, *ecfD*, d'*ecfG*, *sigI* et *sigJ* a pu être démontrée. Cependant, *ecfH* n'a pu être détecté chez cette espèce (Tableau 5).

## 3. Conclusions

Dans ce second volet sur la structure génétique de la famille  $\sigma 70$ , neuf nouveaux gènes susceptibles de coder des facteurs primaires alternatifs et des facteurs de type ECF ont pu être identifiés. La détection de ces gènes chez d'autres espèces que *B. cenocepacia* démontre une diversité encore plus importante de ces facteurs au sein du genre *Burkholderia*. En parallèle, différentes stratégies de PCR/séquençage ont été développées afin de vérifier la présence ou l'absence de certains gènes de la famille  $\sigma 70$  au sein de souches non séquencées, incluant le clone B&B. Ces travaux nous ont conduit à observer des insertions/délétions (indels) de gènes proximaux à certains gènes *ecf*, permettant ainsi de confirmer la divergence entre les clones ET12 et certains autres clones épidémiques de l'espèce *B. cenocepacia*. Cependant, quatre souches, LMG12614, LMG21461, LMG19240 et le clone B&B ont présenté des résultats de crible PCR surprenant au niveau du locus *ecfA*, suggérant la présence des deux configurations génétiques, avec ou sans inversion de cette région. Ces résultats ont été confirmés par séquençage suggérant la présence de deux populations au sein des cultures testées. Cet événement, bien que rare, pourrait avoir un effet sur la virulence de ces souches, comme cela a déjà été montré chez un clone de *P. aeruginosa*. Enfin, des tests PCR ont permis de détecter les gènes *ecfB*, *ecfD*, *ecfG*, *ecfH*, *sigI* et *sigJ* chez le clone B&B confirmant une structure génétique similaire entre les clones B&B et ET12.



**Figure 11: Analyse de la syténie de la région *ecfE* entre *B. cenocepacia* J2315 et *B. multivorans* ATCC17616**  
 Le génome de *B. cenocepacia* J2315 a été comparé au génome de *B. multivorans* ATCC17616 nouvellement assemblé en utilisant le logiciel ACT. **a)** région *ecfE* de *B. cenocepacia* J2315 comparée à la même région chez *B. multivorans* ATCC17616. **b)** région *ecfE* de *B. multivorans* ATCC17616 comparée à la même région chez *B. cenocepacia* J2315. Les ORF de J2315 ainsi que l'îlot génomique 6 (IG6) ont été annotés d'après la publication de ce génome par Holden et al. 2009 [27]. La nomenclature des gènes de *B. multivorans* ATCC17616 a été réalisée selon les codes des protéines sur le site NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Les régions non retrouvées chez *B. multivorans* sont indiquées par des pointillés. La région contenant un phage chez *B. vietnamiensis* est indiquée par un trait continu noir. Les régions contenant le phage chez *B. vietnamiensis* G4 ainsi que l'IG6 et la région contenant les CDS BCAL1318 à 1329 chez J2315 ne sont pas représentées en taille réelle. Le facteur  $\sigma^{70}$  *ecfE* est indiqué en gras. Les flèches noires correspondent aux amorces utilisées dans les cribles PCR (e. g. Tableau 5).



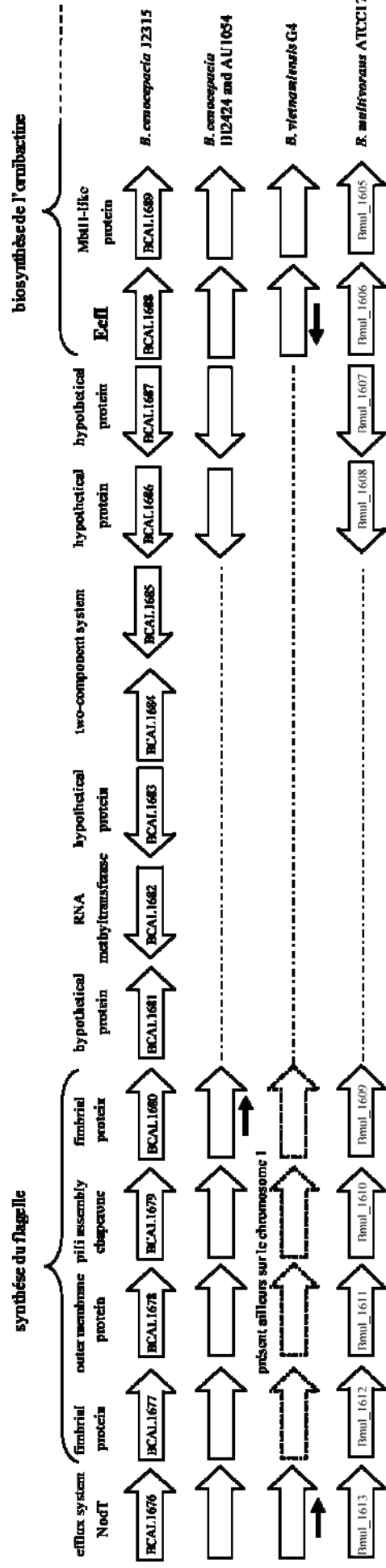
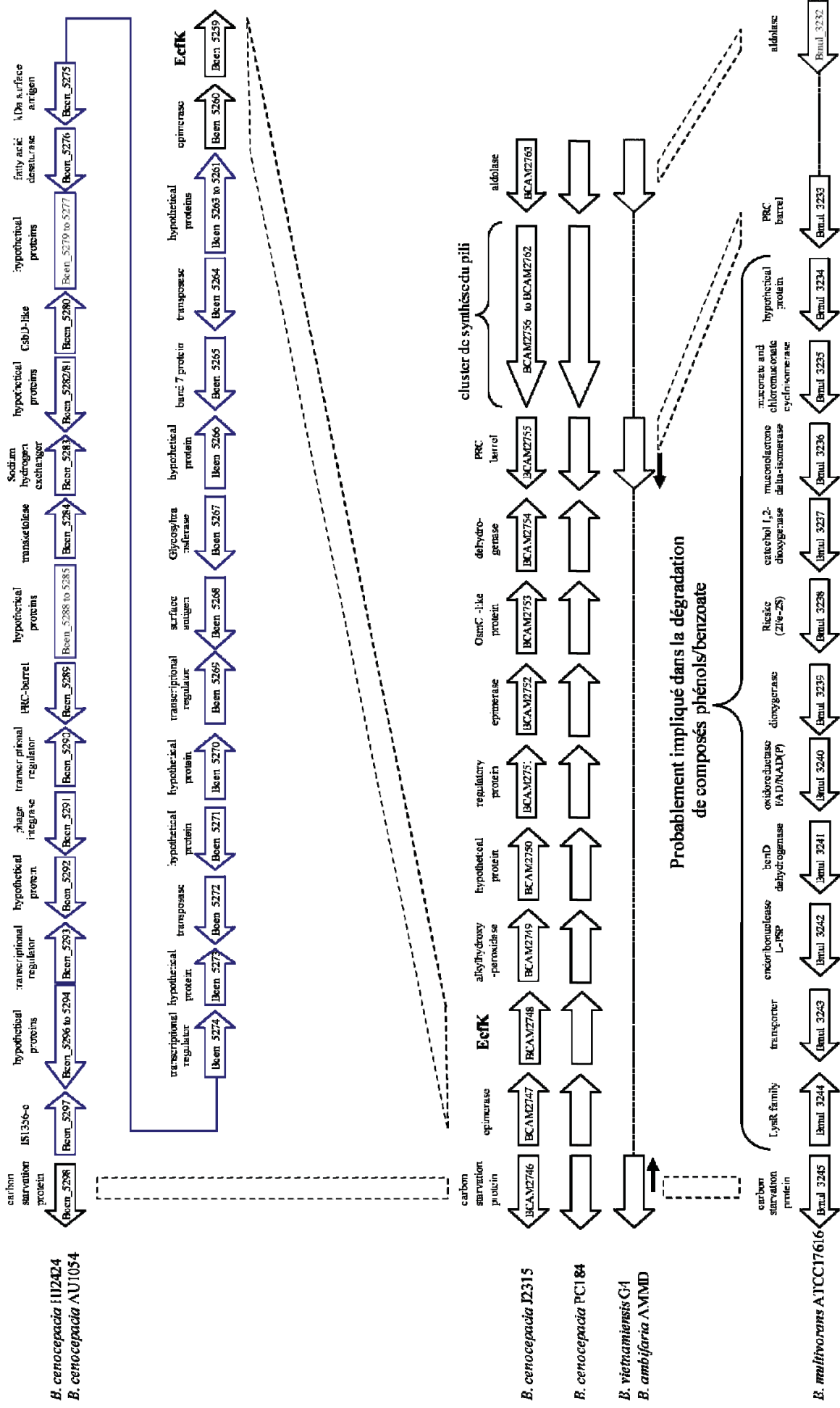
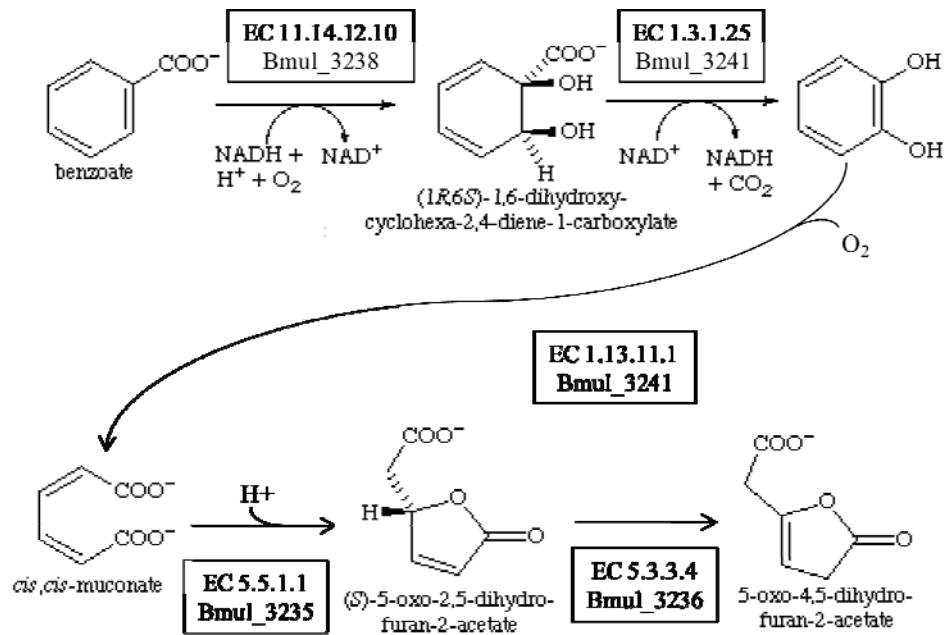


Figure 12: Analyse de la syténie de la région de *ecfH* entre *B. cenocepacia* strain J2315 et différentes souches de *Burkholderia*

Le génome de *B. cenocepacia* J2315 a été comparé aux autres génomes en utilisant le logiciel ACT. Les ORF de J2315 ont été annotés d'après la publication de ce génome par Holden et al.2009 [27]; La nomenclature des gènes de *B. multivorans* ATCC17616 a été réalisée selon le code des protéines sur le site NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Les régions non retrouvées chez les autres *Burkholderia* sont indiquées par des pointillés. Le facteur  $\sigma 70$  *ecfH* est indiqué en gras. Les flèches noires correspondent aux amorces utilisées dans les cribles PCR (c. g. Tableau 5).



**Figure 13: Analyse de la synténie de la région de *ecfK* entre *B. cenosepacia* J2315 et différents génomes de *Burkholderia***  
 Le génome de *B. cenosepacia* J2315 a été comparé au génome d'autres *Burkholderia* en utilisant le logiciel ACT (cf. Matériel et Méthode). Les ORF de J2315 ont été annotés selon la publication du génome par Holden et al. 2009; les annotations des génomes de *B. multivorans* ATCC17616, *B. cenosepacia* H12424 et AU1054 ont été réalisées selon le code des protéines sur le site NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Les régions présentées au sein des autres génomes sont indiquées par des traits pointillés. Les parallélogrammes pointillés indiquent les gènes conservés entre les différents génomes. En bleu sont décrits les CDS présentes chez les deux souches PHDC (AU154, HI2424) et non retrouvées chez J2315. Le facteur  $\sigma 70$  *ecfK* est indiqué en gras. Les flèches noires correspondent aux amorces utilisées dans les cribles PCR (e. g. Tableau 5)



**Figure 14: Voie de dégradation du benzoate chez *B. multivorans* ATCC17616**

Les nombres EC des enzymes ainsi que la correspondance avec les gènes de *B. multivorans* ATCC17616 sont indiqués. La nomenclature des gènes de *B. multivorans* ATCC17616 a été réalisée selon les codes de protéine sur le site NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)).

## 4. Matériels et méthodes

### 4.1. Génome de *Burkholderia*, $\sigma^{70}$ BLAST et recherches d'ORF.

Les analyses par BLAST au sein du génome de *B. cenocepacia* J2315 ont été réalisées en utilisant le génome publié par Holden *et al.* [27]. La détection de neufs nouveaux facteurs  $\sigma^{70}$  a été réalisée par concaténation des séquences obtenues par Menard *et al.* [14] et comparaison de celles-ci par BLAST avec les génomes du Bcc disponibles [47]. Les BLAST ont été réalisés en utilisant les génomes de Bcc comme "database" et le concaténât des séquences acides aminés ou acides nucléiques des  $\sigma^{70}$  comme fichier "query". La visualisation des résultats a été réalisée en utilisant le logiciel ACT [48]. Ce logiciel permet une visualisation graphique des résultats obtenus par BLAST. Les nouveaux facteurs obtenus ont été analysés par BLASTP, BLASTX et par phylogénie moléculaire afin de déterminer leurs groupes d'appartenance (cf section suivante). La synténie de certains  $\sigma^{70}$  a également été étudiée à l'aide des génomes assemblés via le site <http://www.ncbi.nlm.nih.gov>.

### 4.2. Analyses par phylogénie moléculaire

Les alignements multiples ont été calculés en utilisant CLUSTALW [49]. La distance entre paires de séquences, les arbres phylogénétiques obtenus et les valeurs de bootstrap ont été calculés à l'aide du logiciel Mega4 [50]. Les arbres phylogénétique ont été construits en utilisant la méthode Neighbor-Joining (NJ) [51]. Les valeurs de bootstrap ont été obtenues en utilisant 1000 répliques. Les sites contenant des Gap ont été supprimés globalement ou par paires de séquences selon les analyses réalisées.

### 4.3. Souches bactériennes, extraction d'ADN, amplifications PCR et Southern blot

Les souches utilisées dans cette étude sont décrites dans le Tableau 5. Ces souches ont été cultivées sur milieu LB [52] à 28°C. L'extraction des ADN totaux a été réalisée selon le protocole de Pitcher *et al.* [53]. Les réactions PCR ont été effectuées selon les recommandations du fournisseur (Invitrogen, Cergy-Pontoise, France) avec 5% de DMSO. Les amorces listées dans le Tableau 6 ont été utilisées pour l'amplification et/ou le séquençage des séquences d'intérêt. Les amplifications ont été effectuées comme suit : une dénaturation de 5min à 95°C puis 35 cycles de 45s à 95°C, 45s à la température d'hybridation souhaitée (Tableau 6) et 1min à 72°C ; pour finir une élongation de 5 min à 72°C a été réalisée. Les séquences d'ADN sont présentés en annexe.

Les Southern blots ont été réalisés comme suit : les ADN totaux (5 $\mu$ g) ont été digérés avec 1U/ $\mu$ g d'enzyme de restriction *EcoRI* (Fermentas) pendant 16h à 37°C. Une électrophorèse en gel

d'agarose 1% a ensuite été réalisée afin de séparer les différents fragments d'ADN. Un transfert de l'ADN séparé a ensuite été effectué sur membrane de nitrocellulose selon les recommandations du fabricant Genescreen (USA) et suivant le protocole de Sambrook *et al.* [54]. Les hybridations ADN ont été effectuées en accord avec les recommandations du fabricant de membrane (Genescreen, USA). Les sondes ADN utilisées pour réaliser les blots (e. g. tableau 5) ont été marquées à l'aide du kit Amersham "random priming labelling kit" (Amersham-Pharmacia, Orsay-France). Les membranes ont été hybridées et lavées à 65°C. Les autoradiographies ont été révélées selon le protocole de Sambrook *et al.* [54].

**Tableau 6. Amorces PCR utilisées dans ce travail.**

Gène cible	nom de l'amorce PCR	séquence de l'amorce	position chez <i>Bcen-J2315/Bviet-G4*</i>	température d'hybridation
région <i>ecfA1</i>	A	CAGGGTCATCACGGCTCGC	Chr1 3124945-3124962	55°C
région <i>ecfA1</i> et <i>ecfA2</i>	B	TTCCGGTGCATAACCAGC	Chr1 1056358-1056375	55°C
région <i>ecfA2</i>	C	CGGGCGCAAGAACAATACA	Chr1 1056772-1056789	55°C
<i>ecfB</i>	ecfB-F	CGTTCGCGCTCTGGCTG	Chr1 3817718-3817735	58°C
<i>ecfB</i>	ecfB-R	TCAGCCGCAAGGAAGGAGC	Chr1 3818171-3818153	58°C
<i>ecfC</i>	ecfC-F	CGACGCCCTCTATGCCGACCACCAC	Chr1 1506365-1506389	63°C
<i>ecfC</i>	ecfC-R	AGGCCGCCATCGCGAAAAGCACTG	Chr1 1506683-1506687	63°C
région <i>ecfC</i>	*BCAL1382viet	CGCGTCGAACTTCACCG	Chr1 1431976 -1431960	58°C
région <i>ecfC</i>	*BCAL1368viet	GTTTGACGATCGTCTGCG	Chr1 1429775-1429792	58°C
<i>ecfD</i>	ecfD-F	CGCAACTGGCGGCCCG	Chr1 2115044-2115060	55°C
<i>ecfD</i>	ecfD-R	CGAGCGTTTCGAGCGGGTC	Chr1 2615449-2615467	55°C
<i>ecfE</i>	ecfE-F •	CGTTTGGCGCGGAGCAG	Chr1 2727757-2727773	57°C
<i>ecfE</i>	ecfE-R •	GACCGGACATCACTG	Chr1 2728223-2728238	57°C
région <i>ecfE</i>	BCAL2459 •	CAGATAGGGTCCGGCCC	Chr1 2726433-2726449	58°C
région <i>ecfE</i>	*BCAL2466vietF •	CGTTCTGTGCAGGTGCG	Chr1 2745220-2745204	58°C
région <i>ecfE</i>	*BCAL2466vietR •	CTCGCAGCTACATCTGCG	Chr1 2744378-2744395	58°C
<i>ecfG</i>	ecfG-F	GCGCGCGGAGCGGGTTC	Chr1 2146406-2146390	64°C
<i>ecfG</i>	ecfG-R	TCGATAGCCTCCGATCGGGTTCC	Chr1 2145828-2145850	64°C
<i>ecfH</i>	ecfH-F	GGCCCGACCGATGCAAGGAGAGCAC	Chr2 1846236-1846212	63°C
<i>ecfH</i>	ecfH-R	GCTAGGTGGTCCCGGCCAGTTGC	Chr2 1845576-1845600	63°C
<i>ecfI</i>	ecfI-F	ATGGCGGAAAGTGCTCGA	Chr1 1846460-1846476	56°C
<i>ecfI</i>	ecfI-R	GCGTCTGCGACACGTTT	Chr1 1847009-1846993	56°C
<i>ecfI</i>	*ecfI-F-viet	ATGGCGCAAAGTGCTCGA	Chr1 1702224-1702240	56°C
région <i>ecfI</i>	BCAL1680F •	CGAGGACAGCCAGGGC	Chr1 1840202-1840217	56°C
région <i>ecfI</i>	*indel-ecfI-vietF •	CGGCCGGCTCCGACCTACCTC	Chr1 1701889-1701911	64°C
région <i>ecfI</i>	*indel-ecfI-vietR •	GCCGTCGTCCTCCGTCGGTGATG	Chr1 1702589-1702566	64°C
<i>ecfK</i>	ecfK-F	GCACTACACGGGTATCTG	Chr2 3109321-3109338	56°C
<i>ecfK</i>	ecfK-R	TTGGCTGCACGAGAAATC	Chr2 3109935-3109918	56°C
région <i>ecfK</i>	*BCAM2746 •	GTACTGGAAGCCCGGATTC	Chr2 3108000-3108017	56°C
région <i>ecfK</i>	*BCAM2755 •	GCGCTGAAAGGCCACGC	Chr2 3114684-3114669	56°C
<i>sigI</i>	sigI-F	ATCCCGCTCGATGCCCGCC	Chr2 2667864-2667882	55°C
<i>sigI</i>	sigI-R	GAGAAATTTGAAGCCCGGCCG	Chr2 2668242-2668261	55°C
<i>sigJ</i>	sigJ-F	CATGCGCATGGCACCCCGTTC	Chr2 1383013-1383032	55°C
<i>sigJ</i>	sigJ-R	GAGAAATTTGAAGCCCGGCCG	Chr2 2668242-2668261	55°C

\* : amorces dessinées en utilisant le génome de *B. vietnamiensis* G4. • : correspond aux amorces indiquées dans les figures 11, 12 et 13. Le couple d'amorces BCAL2459/ecfE-R a été utilisé pour tester la présence de la région *ecfE* chez les souches de *B. cereopecacia* testées (e. g. Tableau 5).

## Bibliographie (partie B du chapitre 2)

1. Lonetto MA, Rhodius V, Lamberg K, Kiley P, Busby S, Gross C: **Identification of a contact site for different transcription activators in region 4 of the Escherichia coli RNA polymerase sigma70 subunit.** *J Mol Biol* 1998, **284**(5):1353-1365.
2. Lavire C, Blaha D, Cournoyer B: **Selection of unusual actinomycetal primary sigma70 factors by plant-colonizing Frankia strains.** *Appl Environ Microbiol* 2004, **70**(2):991-998.
3. Llamas MA, van der Sar A, Chu BC, Sparrius M, Vogel HJ, Bitter W: **A Novel extracytoplasmic function (ECF) sigma factor regulates virulence in Pseudomonas aeruginosa.** *PLoS Pathog* 2009, **5**(9):e1000572.
4. Shaw LN, Lindholm C, Prajsnar TK, Miller HK, Brown MC, Golonka E, Stewart GC, Tarkowski A, Potempa J: **Identification and characterization of sigma, a novel component of the Staphylococcus aureus stress and virulence responses.** *PLoS One* 2008, **3**(12):e3844.
5. Mittenhuber G: **An inventory of genes encoding RNA polymerase sigma factors in 31 completely sequenced eubacterial genomes.** *J Mol Microbiol Biotechnol* 2002, **4**(1):77-91.
6. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D *et al*: **Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).** *Nature* 2002, **417**(6885):141-147.
7. Jishage M, Ishihama A: **Variation in RNA polymerase sigma subunit composition within different stocks of Escherichia coli W3110.** *J Bacteriol* 1997, **179**(3):959-963.
8. Graindorge A, Menard A, Neto M, Bouvet C, Miollan R, Gaillard S, de Montclos H, Laurent F, Cournoyer B: **Epidemiology and molecular characterization of a clone of Burkholderia cenocepacia responsible for nosocomial pulmonary tract infections in a French intensive care unit.** *Diagn Microbiol Infect Dis* 2009.
9. Riedel K, Hentzer M, Geisenberger O, Huber B, Steidle A, Wu H, Hoiby N, Givskov M, Molin S, Eberl L: **N-acylhomoserine-lactone-mediated communication between Pseudomonas aeruginosa and Burkholderia cepacia in mixed biofilms.** *Microbiology* 2001, **147**(Pt 12):3249-3262.
10. Saiman L, Cacalano G, Prince A: **Pseudomonas cepacia adherence to respiratory epithelial cells is enhanced by Pseudomonas aeruginosa.** *Infect Immun* 1990, **58**(8):2578-2584.
11. Keith KE, Hynes DW, Sholdice JE, Valvano MA: **Delayed association of the NADPH oxidase complex with macrophage vacuoles containing the opportunistic pathogen Burkholderia cenocepacia.** *Microbiology* 2009, **155**(Pt 4):1004-1015.
12. Lengwehasatit I, Nuchtas A, Tungpradabkul S, Sirisinha S, Utaisinchaoen P: **Involvement of B. pseudomallei RpoS in apoptotic cell death in mouse macrophages.** *Microb Pathog* 2008, **44**(3):238-245.
13. Solis R, Bertani I, Degrassi G, Devescovi G, Venturi V: **Involvement of quorum sensing and RpoS in rice seedling blight caused by Burkholderia plantarii.** *FEMS Microbiol Lett* 2006, **259**(1):106-112.
14. Menard A, de Los Santos PE, Graindorge A, Cournoyer B: **Architecture of Burkholderia cepacia complex sigma70 gene family: evidence of alternative primary and clade-specific factors, and genomic instability.** *BMC Genomics* 2007, **8**:308.
15. Vanlaere E, Baldwin A, Gevers D, Henry D, De Brandt E, LiPuma JJ, Mahenthiralingam E, Speert DP, Dowson C, Vandamme P: **Taxon K, a complex within the Burkholderia cepacia complex, comprises at least two novel species, Burkholderia contaminans sp. nov. and Burkholderia lata sp. nov.** *Int J Syst Evol Microbiol* 2009, **59**(Pt 1):102-111.

16. Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z *et al*: **CDD: a Conserved Domain Database for protein classification**. *Nucleic Acids Res* 2005, **33**(Database issue):D192-196.
17. Lonetto M, Gribskov M, Gross CA: **The sigma 70 family: sequence conservation and evolutionary relationships**. *J Bacteriol* 1992, **174**(12):3843-3849.
18. **HTTP server UMR5557 ECF** [ <http://ecomicro.univ-lyon1.fr/poe/ECF.txt>]
19. **HTTP server UMR5557 NON-ECF** [ <http://ecomicro.univ-lyon1.fr/poe/NON-ECF.txt>]
20. Vandamme P, Coenye T: **Taxonomy of the genus Cupriavidus: a tale of lost and found**. *Int J Syst Evol Microbiol* 2004, **54**(Pt 6):2285-2289.
21. Brito B, Aldon D, Barberis P, Boucher C, Genin S: **A signal transfer system through three compartments transduces the plant cell contact-dependent signal controlling *Ralstonia solanacearum* hrp genes**. *Mol Plant Microbe Interact* 2002, **15**(2):109-119.
22. Chain PS, Denef VJ, Konstantinidis KT, Vergez LM, Agullo L, Reyes VL, Hauser L, Cordova M, Gomez L, Gonzalez M *et al*: **Burkholderia xenovorans LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility**. *Proc Natl Acad Sci U S A* 2006, **103**(42):15280-15287.
23. Coenye T, Vandamme P: **Diversity and significance of Burkholderia species occupying diverse ecological niches**. *Environ Microbiol* 2003, **5**(9):719-729.
24. Kresse AU, Dinesh SD, Larbig K, Romling U: **Impact of large chromosomal inversions on the adaptation and evolution of *Pseudomonas aeruginosa* chronically colonizing cystic fibrosis lungs**. *Mol Microbiol* 2003, **47**(1):145-158.
25. Kresse AU, Blocker H, Romling U: **ISP<sub>a</sub>20 advances the individual evolution of *Pseudomonas aeruginosa* clone C subclone C13 strains isolated from cystic fibrosis patients by insertional mutagenesis and genomic rearrangements**. *Arch Microbiol* 2006, **185**(4):245-254.
26. Lessie TG, Hendrickson W, Manning BD, Devereux R: **Genomic complexity and plasticity of *Burkholderia cepacia***. *FEMS Microbiol Lett* 1996, **144**(2-3):117-128.
27. Holden MT, Seth-Smith HM, Crossman LC, Sebahia M, Bentley SD, Cerdeno-Tarraga AM, Thomson NR, Bason N, Quail MA, Sharp S *et al*: **The genome of *Burkholderia cenocepacia* J2315, an epidemic pathogen of cystic fibrosis patients**. *J Bacteriol* 2009, **191**(1):261-277.
28. LiPuma JJ, Spilker T, Coenye T, Gonzalez CF: **An epidemic *Burkholderia cepacia* complex strain identified in soil**. *Lancet* 2002, **359**(9322):2002-2003.
29. Chen JS, Witzmann KA, Spilker T, Fink RJ, LiPuma JJ: **Endemicity and inter-city spread of *Burkholderia cepacia* genomovar III in cystic fibrosis**. *J Pediatr* 2001, **139**(5):643-649.
30. LiPuma JJ, Mortensen JE, Dasen SE, Edlind TD, Schidlow DV, Burns JL, Stull TL: **Ribotype analysis of *Pseudomonas cepacia* from cystic fibrosis treatment centers**. *J Pediatr* 1988, **113**(5):859-862.
31. Vandamme P, Holmes B, Coenye T, Goris J, Mahenthiralingam E, LiPuma JJ, Govan JR: ***Burkholderia cenocepacia* sp. nov.-a new twist to an old story**. *Res Microbiol* 2003, **154**(2):91-96.
32. Ramette A, LiPuma JJ, Tiedje JM: **Species abundance and diversity of *Burkholderia cepacia* complex in the environment**. *Appl Environ Microbiol* 2005, **71**(3):1193-1201.
33. Darling P, Chan M, Cox AD, Sokol PA: **Siderophore production by cystic fibrosis isolates of *Burkholderia cepacia***. *Infect Immun* 1998, **66**(2):874-877.
34. Henry DA, Campbell ME, LiPuma JJ, Speert DP: **Identification of *Burkholderia cepacia* isolates from patients with cystic fibrosis and use of a simple new selective medium**. *J Clin Microbiol* 1997, **35**(3):614-619.



35. LiPuma JJ, Dulaney BJ, McMenamin JD, Whitby PW, Stull TL, Coenye T, Vandamme P: **Development of rRNA-based PCR assays for identification of Burkholderia cepacia complex isolates recovered from cystic fibrosis patients.** *J Clin Microbiol* 1999, **37**(10):3167-3170.
36. Mahenthiralingam E, Coenye T, Chung JW, Speert DP, Govan JR, Taylor P, Vandamme P: **Diagnostically and experimentally useful panel of strains from the Burkholderia cepacia complex.** *J Clin Microbiol* 2000, **38**(2):910-913.
37. Balandreau J, Viallard V, Cournoyer B, Coenye T, Laevens S, Vandamme P: **Burkholderia cepacia genomovar III Is a common plant-associated bacterium.** *Appl Environ Microbiol* 2001, **67**(2):982-985.
38. Richau JA, Leitao JH, Correia M, Lito L, Salgado MJ, Barreto C, Cescutti P, Sa-Correia I: **Molecular typing and exopolysaccharide biosynthesis of Burkholderia cepacia isolates from a Portuguese cystic fibrosis center.** *J Clin Microbiol* 2000, **38**(4):1651-1655.
39. Stanier RY, Palleroni NJ, Doudoroff M: **The aerobic pseudomonads: a taxonomic study.** *J Gen Microbiol* 1966, **43**(2):159-271.
40. Vandamme P, Holmes B, Vancanneyt M, Coenye T, Hoste B, Coopman R, Revets H, Lauwers S, Gillis M, Kersters K *et al*: **Occurrence of multiple genomovars of Burkholderia cepacia in cystic fibrosis patients and proposal of Burkholderia multivorans sp. nov.** *Int J Syst Bacteriol* 1997, **47**(4):1188-1200.
41. Buchrieser C, Brosch R, Bach S, Guiyoule A, Carniel E: **The high-pathogenicity island of Yersinia pseudotuberculosis can be inserted into any of the three chromosomal asn tRNA genes.** *Mol Microbiol* 1998, **30**(5):965-978.
42. Hentschel U, Hacker J: **Pathogenicity islands: the tip of the iceberg.** *Microbes Infect* 2001, **3**(7):545-548.
43. Lesic B, Carniel E: **Horizontal transfer of the high-pathogenicity island of Yersinia pseudotuberculosis.** *J Bacteriol* 2005, **187**(10):3352-3358.
44. Sentchilo V, Czechowska K, Pradervand N, Minoia M, Miyazaki R, van der Meer JR: **Intracellular excision and reintegration dynamics of the ICEclc genomic island of Pseudomonas knackmussii sp. strain B13.** *Mol Microbiol* 2009, **72**(5):1293-1306.
45. Boyd EF, Almagro-Moreno S, Parent MA: **Genomic islands are dynamic, ancient integrative elements in bacterial evolution.** *Trends Microbiol* 2009, **17**(2):47-53.
46. Tyler SD, Rozee KR, Johnson WM: **Identification of IS1356, a new insertion sequence, and its association with IS402 in epidemic strains of Burkholderia cepacia infecting cystic fibrosis patients.** *J Clin Microbiol* 1996, **34**(7):1610-1616.
47. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**(17):3389-3402.
48. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J: **ACT: the Artemis Comparison Tool.** *Bioinformatics* 2005, **21**(16):3422-3423.
49. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**(22):4673-4680.
50. Kumar S, Tamura K, Nei M: **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Brief Bioinform* 2004, **5**(2):150-163.
51. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4**(4):406-425.
52. Bertani G: **Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli.** *J Bacteriol* 1951, **62**(3):293-300.
53. Pitcher DG, Saunders, N.A., Owen, R.J.: **Rapid extraction of bacterial genomic DNA with guanidium thiocyanate.** *Letters in Applied Microbiology* 1989, **8**:151-156.

54. Sambrook J, E. F. Fritsch, and T. Maniatis.: **Molecular cloning: a laboratory manual.** Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press; 1989.

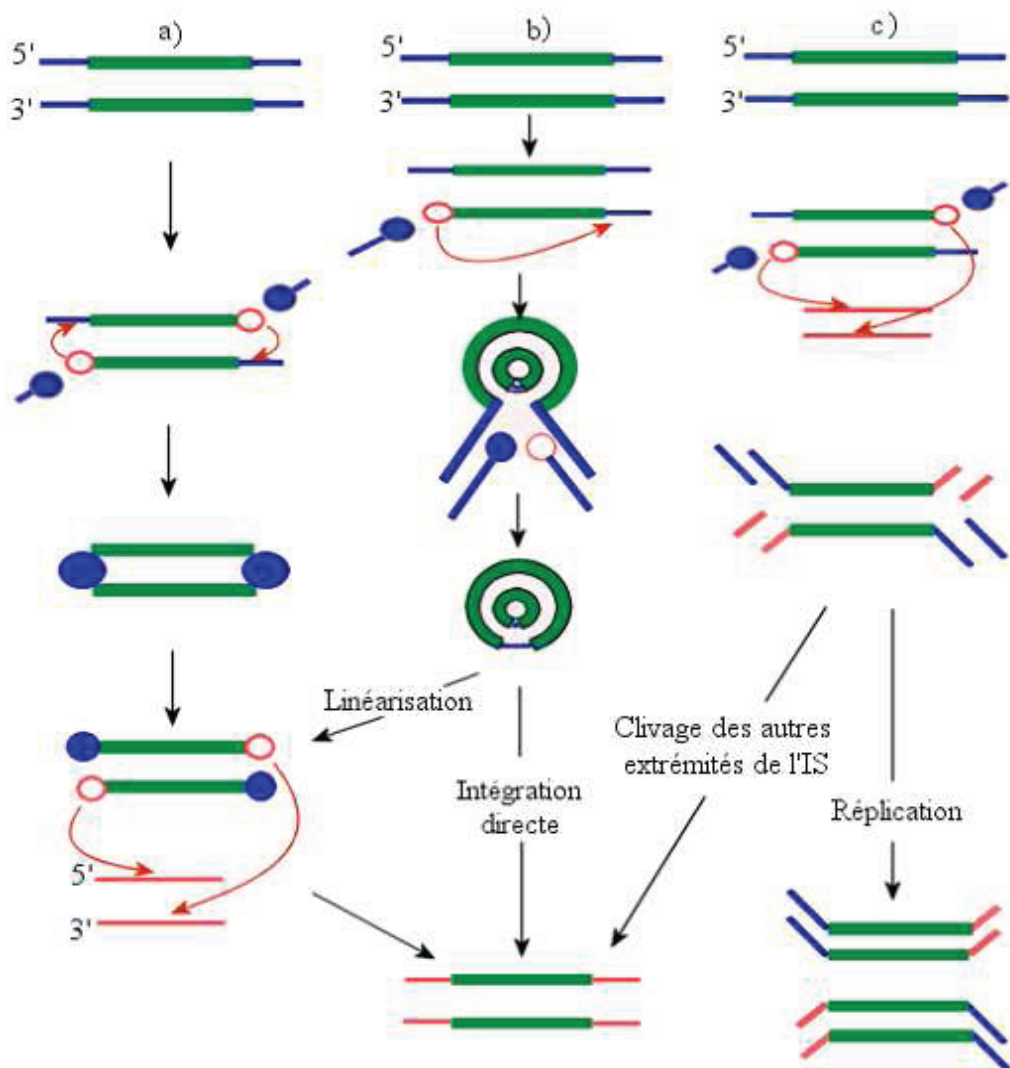
## 6. Conclusion de chapitre

Ce chapitre avait pour objectif d'analyser l'architecture, la distribution et la plasticité de la famille des facteurs  $\sigma 70$  chez les *Burkholderia* du complexe *cepacia* tout en précisant les divergences et similarités entre les principaux clones épidémiques isolés de patients CF et le clone B&B. Les résultats nous ont permis de mettre en évidence une distribution hétérogène des facteurs  $\sigma 70$  au niveau des différentes espèces de *Burkholderia*. Bien que les facteurs de la famille des  $\sigma 70$  soient communs à toutes les protéobactéries, nous avons mis en évidence des stratégies évolutives différentes entre les *Burkholderia* du complexe *cepacia* et l'espèce *P. aeruginosa*. Ces bactéries sont capables de coloniser les mêmes biotopes cependant la souche *P. aeruginosa* PAO1 contient de nombreux facteurs  $\sigma 70$  de type ECF impliqués dans l'acquisition du Fer alors que chez les *Burkholderia* seulement trois gènes homologues à ces ECF ont été détectés. L'identification de facteurs primaires alternatifs chez les *Burkholderia* indique également une stratégie différente quant à l'adaptation aux biotopes colonisés. Ces facteurs primaires alternatifs n'avaient jamais été identifiés ailleurs que chez les actinobactéries. Cette étude a par ailleurs permis de détecter de nombreux réarrangements génétiques telles que des duplications, des délétions ou des inversions de séquences d'ADN, pouvant expliquer l'hétérogénéité de la distribution des  $\sigma 70$  au sein des espèces de *Burkholderia* et parfois même au niveau infra-spécifique. L'hétérogénéité de cette distribution peut conduire à une adaptabilité différente des souches aux divers biotopes ; ces facteurs étant impliqués dans la régulation de nombreux gènes, la délétion ou l'introduction de l'un d'entre eux pourrait conduire à des profils d'expressions différents et ainsi à l'expression de phénotypes différents en fonction des souches. On peut alors se demander si la duplication de l'*ecfA* permettrait aux souches ayant les deux copies de répondre plus rapidement à un stress thermique. Les réarrangements génétiques menant à cette distribution hétérogène des  $\sigma 70$  pourraient avoir impliqués des gènes codant des ARNt, des séquences d'insertion (IS) ou encore des phages. Dans les cas de la duplication d'*ecfA*, une séquence d'insertion (*ISBcen8*) semble avoir été impliquée dans la duplication et l'inversion d'une partie d'un chromosome chez certaines souches de *B. cenocepacia*. Les IS semblent donc jouer un rôle majeur dans la plasticité des génomes notamment chez les *Burkholderia*. Ces IS peuvent être impliquées dans la réorganisation globale des génomes, dans la fusion de deux chromosomes, et dans l'acquisition d'ilots génomiques pouvant permettre une augmentation de la virulence, une capacité à dégrader certains composés, etc. Le chapitre suivant présentera nos études relatives à la distribution des IS au sein des génomes de *Burkholderia* disponibles et à leur caractérisation, et leur implication possible dans l'émergence du clone B&B.

# Relation entre émergence de clones épidémiques de *Burkholderia cenocepacia* et amplification d'éléments répétés de type IS

Les analyses épidémiologiques du clone B&B de *Burkholderia cenocepacia* ont permis d'observer sa large distribution et proximité avec des clones responsables d'infections pulmonaires chez les patients atteints de mucoviscidose (chapitre 1). Ces clones ayant leur génome séquencé, nous avons orienté la suite des travaux vers des analyses de génomique comparative visant à identifier des processus ou gènes (îlots génomiques) ayant pu contribuer à son émergence et forte transmissibilité. Dans le chapitre 2, nous avons étudié la famille de gènes codant des facteurs sigma 70 et ainsi évaluer l'importance des phénomènes de réarrangements génétiques entre ce clone et les souches de *Burkholderia cenocepacia* séquencées à ce jour. Ces travaux ont confirmé la forte proximité de ce clone avec le clone ET12 fortement virulent et responsable de pneumonies nécrotiques foudroyantes chez les individus atteints de mucoviscidose (CF). Ils ont permis d'identifier de nombreux facteurs sigma70 susceptibles d'avoir contribué à l'adaptation ou l'acclimatation de ces souches à l'homme. Ils ont également permis d'observer des phénomènes de délétion, duplication, et inversion de segments de chromosomes suggérant une contribution importante des réarrangements génomiques dans l'émergence de ces clones.

Les mécanismes à la base de cette plasticité génomique sont nombreux mais impliquent de façon récurrente, directe ou indirecte, des éléments répétés au sein des génomes bactériens, appelés IS (séquence d'insertion). Dès 1996, Lessie et al., avaient constaté une grande plasticité au sein des génomes de *Burkholderia* et soulignaient le rôle probable des IS dans cette plasticité. Les IS sont de petits éléments génétiques mobiles codant uniquement l'enzyme (transposase) essentielle à leur mobilité (e. g. voir Mahillon and Chandler 1998). Elles sont délimitées par des séquences inversées répétées, les IR. Ces IR et les similarités de séquences entre transposases sont utilisées pour la classification de ces IS en famille. Les familles d'IS n'utilisent pas nécessairement des modes de transposition identique, et dans certaines situations, ces transpositions peuvent maintenir, réduire ou augmenter le nombre de copies d'une IS au sein d'un génome (Fig. 1).



**Fig. 1 : Différents modes de transposition d'IS.**

a et c : la première étape est un clivage aux deux extrémités de l'IS. a : un intermédiaire linéaire de l'IS est libéré. c : l'IS est toujours liée en une structure décrite par Shapiro (1979). b : le clivage ne se fait que sur un brin d'ADN, pour former un intermédiaire en huit résolue en une IS circulaire où les deux IR sont séparées par quelques paires de bases. Le devenir de la séquence de 3 pb n'est pas détaillé car mal compris. Ce cercle peut être soit linéarisé soit directement intégré dans la molécule cible. En vert les deux brins de l'IS, l'ADN donneur est représenté par des lignes bleu alors que l'ADN cible est en rouge. Les groupements phosphates sont représentés par des disques bleus alors que l'extrémité 3'-OH (réactive) est sous la forme d'un cercle rouge. Les flèches en rouge matérialisent l'attaque du 3'-OH sur l'ADN cible. Schéma modifié de Mahillon et Chandler (1998) et Turlan et Chandler (2000).

Les IS au sein d'un génome peuvent être le support de recombinaisons homologues ou homéologues conduisant à des réarrangements au sein des génomes bactériens. Ces réarrangements peuvent être intra-génomique (Schneider et al., 2000) ou conduire à l'intégration d'ADN exogène (Schneider et al., 2002). L'implication des IS dans un réarrangement est caractérisée par la présence de copies de l'IS en bordure. Nous avons mis en évidence dans le chapitre 2 une configuration de ce type, impliquant une IS nommée *ISBcen8* chez *B. cenocepacia* J2315. Cette IS a pu être reliée à un large réarrangement chromosomique. L'impact fonctionnel de tels réarrangements est variable mais une relation avec des changements dans la virulence de certaines bactéries pathogènes a été observé (e. g. Romling et al., 1997). Plus localement, les IS peuvent également modifier l'environnement des gènes situés à proximité de leur site d'insertion. Par exemple, l'*IS407* code une séquence « promotrice » permettant l'attachement de facteurs de type sigma70 et pouvant ainsi modifier l'expression des gènes en aval (Wood et al., 1990). Par contre, l'insertion d'une IS dans un gène peut conduire à son inactivation. DeShazer et al., (2001) ont montré que l'insertion d'une copie de l'*IS407* bloquait la synthèse de la capsule chez *B. mallei*, et provoquait une baisse du pouvoir pathogène. Il est à noter que deux IS situées à proximité l'une de l'autre peuvent être co-mobilisées ainsi que la portion d'ADN qu'elles encadrent conduisant à la formation d'un transposon composite avec un module phénotypique (Toussaint and Merlin 2002). Ces IS composites peuvent être impliquées dans certains transferts horizontaux de gènes fonctionnels entre espèces ou souches bactériennes.

Deux hypothèses sont avancées pour expliquer le maintien des IS au sein des génomes bactériens. La première (Charlesworth et al., 1994) considère que les IS sont des parasites des génomes bactériens provoquant des mutations parfois délétères. Malgré leur coût, leur maintien dans ces génomes est assuré par la colonisation de nouveaux sites d'insertion notamment sur des plasmides ; ce qui permet leur propagation par transferts horizontaux. La seconde hypothèse (Blot 1994) retient le fait que les IS peuvent également provoquer des mutations adaptatives bénéfiques pour l'hôte. Leur maintien est alors assuré par les quelques mutations bénéfiques apportées. Cette thèse est aussi soutenue par Holden et al., 2009 qui expliquent que la présence de nombreux pseudogènes (inactivés par des mutations, délétion, insertion d'IS) pourrait jouer un rôle dans l'adaptation d'une souche à un biotope particulier. Afin d'évaluer l'impact des IS sur l'évolution bactérienne, des études ont été menées sur la bactérie modèle *E. coli*. Papadopoulos et al., (1999), après un suivi sur 10 000 générations de cultures d'*E. coli*, ont mis en évidence une baisse du taux de modifications phénotypiques au cours du temps, avec des IS présentes à une proportion relativement stable au cours du temps.

Ces auteurs suggèrent donc une contribution négligeable des IS dans l'évolution des phénotypes, au détriment d'autres processus tels que la sélection de mutations ponctuelles. Cependant, certaines mutations impliquant des IS ont été étudiées en détail et leur bénéfice démontré (Schneider and Lenski 2004).

Nos études du chapitre 2, et des analyses préliminaires permettant d'estimer l'importance des IS en termes de nombre, ont suggéré un rôle de ces éléments dans l'évolution des génomes de *B. cenocepacia*. Un bilan exhaustif des IS présentes dans les génomes de cette espèce est donc présenté dans ce chapitre, et complété d'une analyse de ces éléments chez le clone B&B. Les implications fonctionnelles de ces insertions d'IS sont discutées.

# *Genomic organization of Burkholderia cenocepacia IS elements and its relation with the emergence of epidemic clones*

## **1. Introduction**

*B. cenocepacia* is a member of the *Burkholderia cepacia* complex (Bcc), a group of seventeen closely related species. Members of this group are able to colonize several niches (maize rhizosphere, soils, freshwaters, rice field, polluted soils etc...) and most of them have been isolated from lungs of Cystic Fibrosis (CF) patients. In fact, the Bcc species are considered to be opportunistic pathogens infecting immuno-compromised or CF individuals. Prevalence of the Bcc species among CF patients is lower than the one of *Pseudomonas aeruginosa* (Lambiase et al., 2006). However, life expectancy of these patients is reduced when Bcc infection is occurring (Hutchison and Govan 1999). *B. multivorans* and *B. cenocepacia* are the main Bcc species recovered from CF patients. *B. cenocepacia* can lead to a dramatic necrotic pneumonia named the “*cepacia* syndrome” (e. g. Saiman and Siegel 2004). Several epidemic clones of *B. cenocepacia* have been described from the mid-80’s. One of them, the ET12 clone is widely distributed in North America and the United Kingdom, and was shown to be highly transmissible (Isles et al., 1984, Johnson et al., 1994). Two other epidemic clones, named PDHC and Midwest, have been described and mainly founded in North America (LiPuma et al., 1988, Coenye et al., 2004). The B&B clone represent another epidemic clone recently described (Graindorge et al., 2009). This clone and its related clonal complex (ST32), were found involved in nosocomial and CF pulmonary tract infections, and have disseminated among several occidental countries. Strains of this clonal complex were recently reported as the most prevalent among French CF cohorts (Abalain et al., 2009). Emergence and differentiation of these epidemic clones could have involved several molecular processes including DNA acquisition/deletion, transcriptional changes of gene expression, and genetic disruptions of genes by insertion sequences or small *indels* (insertion or deletion of a few nucleotides).

In this work, an in depth investigation of the insertion sequences (IS) elements genomic organization among *Burkholderia* genomes was undertaken. IS are small bacterial



mobile genetic elements. They can be found in high numbers within a genome and, can apparently move autonomously inside the bacterial cell. Insertion of an IS within a gene can lead to its inactivation (DeShazer et al., 2001). Insertion in the promoter region can also modify the expression of a gene as shown for the IS407 and IS406. These IS were shown to enhance expression of a flanking gene by providing an exogenous  $\sigma 70$  promoter (Wood et al., 1990, Wood et al., 1991). IS elements can also play a role in homologous recombination within a replicon or between replicons of a genome like it has been described by Byrne et al. (1994) and Menard et al. (2007). All these events can thus modify dramatically the phenotype of these bacteria. In some cases, a concordance with the emergence of clinical strains was inferred. For example, Tyler et al. (1996) identified a hybrid element containing a copy of IS1356 inserted into IS402 (iso-IS*Bcen20*), and this IS organization appeared specific of the ET12 lineage. Moreover, Liu et al. (2003) identified an IS element (IS1363, renamed IS*Bcen2*) in high copy numbers within PHDC strains but in a single copy in ET12. Similarly, Kenna et al. (2006) investigated the distribution of three IS within the Bcc and found the IS407 in high copy numbers among ET12 strains but not among the PHDC clone. A potential “linkage” between IS, virulence and transmissibility could thus be hypothesized.

Here, all IS elements found among *B. cenocepacia* genome sequences were allocated to their respective family, and sub-groups, and their origin investigated by molecular phylogeny. Processes involved in their occurrence at a particular site in the sequenced genomes were inferred. Inferences about the impact of these IS on the properties of the sequenced strains were also made. These data were then used to investigate the IS elements genomic organization of the B&B clone. Genomic stability of IS407 insertions was experimentally investigated using a panel of *B. cenocepacia* strains from the main *recA* PCR RFLP sub-groups. These data further support the hypothesis of a close proximity between the ET12 and B&B clones; both clones showing several similarities in their IS elements genomic organizations.

**Table 1: Bacterial strains used in this study**

<i>B. cenocepacia</i>	Other code	Origin	Reference
LMG 16656	J2315	ET12, Cystic Fibrosis	Vandamme et al., 2003
LMG 12614		ET12, Cystic Fibrosis	Vandamme et al., 1997
LMG 18826	BC7	ET12, Cystic Fibrosis	Sajjan et al., 1995
LMG 18829	PC184	Cystic Fibrosis	Vandamme et al., 2003
LMG 16654		Cystic Fibrosis	Vandamme et al., 2003
LMG 18828		Cystic Fibrosis	Mahenthiralingam et al., 2000
LMG 18832	ATCC 17765	Urinary tract infection	Stanier <i>et al.</i> 1966
H111		Cystic Fibrosis	Romling et al., 1994
LMG 19240		Continuous wheat plot	Viallard <i>et al.</i> 1998
LMG 21461		Cystic Fibrosis	Vandamme et al., 2002
B&B clone	Bpoe732	Pulmonary infection	Graindorge et al., 2009

## 2. Materials and Methods.

### 2.1. Strain used, DNA extractions, PCR amplification and DNA blot hybridization.

*Burkholderia* strains used in this study are listed in Table 1. Bacteria were grown in Luria-Bertani (LB) broth (Bertani 1951) at 30°C between 24h 48h. DNA extractions were performed according to Pitcher 1989 and resuspended in ultra-pure water.

PCR (polymerase chain reaction) amplifications (25-50 µl) were done according to the *Taq* polymerase manufacturer (Invitrogen, Cergy Pontoise, France) without 1% w-1 detergent but with 5% DMSO. PCR primers used in this study are described in Table 2. PCR cycles using 50-100ng of total DNA were as follow: (1) 95°C for 5 min, (2) 95°C for 30 sec, selected annealing temperature (e.g. Table 2) for 45 sec, 72°C for 60 sec (35 cycles), (3) 72°C for 5 min.

DNA sequences of several B&B clone IS elements used for phylogenetic analysis (*IS407*, *ISBcen20*, *ISBcen4*, *ISBcen6* and *ISBcen9*) were obtained after cloning PCR products using the pGEMT-easy cloning kit (Promega, Charbonnières, France) within *Escherichia coli* DH5α strains. These sequences are given in the annex C1 section, at the end of this manuscript.

Plasmids (pGEM-T) purification was performed using QIAGEN® Plasmid Plus Midi Kit (Qiagen, Courtaboeuf, France).

DNA sequencing of the PCR products and pGEM-T plasmids was performed by Cogenics® (Grenoble, France).

Southern blots were performed as follow: each strain genomic DNAs (5µg) were digested by 1U/µg of *EcoRI* (Fermentas), for 16h at 37°C. Digested DNAs were separated by electrophoresis using a 1% agarose gel, for 6h at 100Volts (60A). DNA blots were done using Genescreen (USA) nylon membranes and using the procedure described by the manufacturer. DNAs probes were labeled using the random priming labeling kit (Amersham-Pharmacia, Orsay, France) and blots were performed following Sambrook (1989). DNA hybridizations were carried out according to the nylon membrane manufacturer (Genescreen, USA), hybridized and washed at 65°C. Autoradiography was performed according to Sambrook (1989).

## ***2.2. Identification of IS and insertion sites among sequenced genomes.***

The full genome sequence of *B. cenocepacia* J2315 was retrieved from the Sanger Institute web site ([http://www.sanger.ac.uk/Projects/B\\_cenocepacia/](http://www.sanger.ac.uk/Projects/B_cenocepacia/)). *B. cenocepacia* AU1054, *B. cenocepacia* HI2424, *B. ambifaria* AMMD, *Burkholderia* sp. 383, *B. vietnamiensis* G4, *B. multivorans* ATCC17616 and *B. xenovorans* LB400 genome sequences were retrieved from the Joint Genome Institute web site ([http://genome.jgi-psf.org/mic\\_home.html](http://genome.jgi-psf.org/mic_home.html)). *B. mallei* ATCC 23344, *B. pseudomallei* K96243 and *B. thailandensis* E264 genome sequences were retrieved from Genbank. *Burkholderia cenocepacia* PC184 genome was retrieved from the Broad Institute ([http://www.broadinstitute.org/annotation/genome/burkholderia\\_cenocepacia/](http://www.broadinstitute.org/annotation/genome/burkholderia_cenocepacia/)).

BLAST analyses were performed locally using the BLAST-2.2.20 package available from the NCBI web page (<ftp://ftp.ncbi.nih.gov/blast/executables/LATEST>).

Genomic analyses of insertion sites were performed using a local BLAST approach allowing comparison of IS sequences retrieved from Genbank, the IS databases (<http://www-is.biotoul.fr/>) and the above genome sequences. The Artemis Comparison Tool software (Carver et al., 2005) was used to visualize graphically the results of these different BLAST.

**Table 2: PCR primer used in this study**

name	sequence (5'-3')	first nucleotide position on <i>Bcen</i> -J2315 genome	annealing temperature
IS407aF	GCTCAACCTCGGGATACTGT	(Chr1 450247)	55°C
IS407aR	GGCGTTGTCGGATTGTT	(Chr1 451630)	55°C
IS407cF	GACTGATGATGGCATTCC	(Chr1 1276718)	59°C
IS407cR	GAGCTGAGACAGTGATACCG	(Chr1 1278500)	58°C
IS407eF	GGGCCTCGTCGCAAACCAT	(Chr1 1426953)	61°C
IS407eR	GAGCCCCGCATCGATCCATTCTT	(Chr1 1425422)	61°C
IS407fF	AACCTGGTTTCGAACACG	(Chr1 2560114)	58°C
IS407fR	GAGTCTGCCTTGTAGATCGG	(Chr1 2561809)	58°C
IS407iF	CCTTCAGGTGTTGGCTTACA	(Chr1 2848475)	58°C
IS407iR	CGAGACGAGTAGCGACTTGT	(Chr1 2849466)	58°C
IS407jF	GATCTGCTTGGCAATATG	(Chr1 2849773)	57°C
IS407jR	AACGCAAGGAAGAATAAGAC	(Chr1 2851308)	57°C
SI11	CAAATTGCCACGCGACTAC	(Chr1 454987)	58°C
SI10	TGACCGAAGATCGAGACG	(Chr1 456554)	56°C
SI8	AGTCGTTCTGTTCCCGAT	(Chr1 1416963)	58°C
SI17	GAGATCGCAGGTAGCTTAG	(Chr1 1418385)	58°C
SI21	TAGTACTGCGGAAGGCATG	(Chr1 2800281)	62°C
SI22	CATGAGATCAGGCGACAAC	(Chr1 2801819)	58°C
SI23	GAACAGAGCGACGTGTAAG	(Chr1 2840657)	58°C
SI24	TCCTTGCCATGACGGATG	(Chr1 2842068)	58°C
SI5	GCAACGAGCGACCATTGAC	(Chr1 3525140)	60°C
SI7	CGGCTTGGAAATTATTGTGAC	(Chr1 3526591)	58°C
SI27	TTCACGTGGTTAGTTGG	(Chr2 2128444)	50°C
SI20	GTGGTACTCGCATTATTCCG	(Chr2 2130034)	56°C
SI3	CGGTAAACGGGAAACTACG	(Chr3 789840)	58°C
SI2	GTGTCGTGATGGACTGCAT	(Chr3 788422)	58°C
SI14	AACCGGAACAACATGCATTC	(Plasmid 80958)	58°C
SI15	TTCGGTCGATCTACGGTC	(Plasmid 82552)	56°C
ISBcen4F	GTAGACATCGCCAAGCGG	(Chr3 713400)	58°C
ISBcen4R	GCTGCAGTAACTGGAGGC	(Chr3 712823)	58°C
ISBcen6F	GCAGGAAGTGATGGAAGTG	(Chr3 713044)	58°C
ISBcen6R	CGATAGCCATAGCTGGGC	(Chr3 712292)	58°C
ISBcen8F	CCTGAAGCCCGAACGGTTAC	(Chr1 1102132)	62°C
ISBcen8R	CTGGATGGCGCAATGATAG	(Chr1 1102704)	62°C
ISBcen9F	CCGAACGAGAAGTGGGTTAC	(Chr1 54671)	60°C
ISBcen9R	GATGCGCTCGTGATTGTAG	(Chr1 55078)	60°C
ISBcen20F	CTCAAGACCGGACTACCG	(Chr1 626666)	58°C
ISBcen20R	CGGAGACGACGGAAGTG	(Chr1 626073)	58°C
IS1363F	GCTTAATAGGATGGTCAG	(Chr3 749398)	56°C
IS1363R	TCTATGGTCAGCCCGATTTTTG	(Chr3 750739)	56°C

### *2.3. Molecular phylogenetic analyses.*

Multiple alignments of amino acids (AA) and DNA sequences of transposases were computed using CLUSTALW. Distance between sequence pairs, the inferred phylogenetic trees, and bootstrap values were all computed through the Mega4 graphic tool (Kumar et al., 2004). Phylogenetic trees were built using the Neighbor-Joining (NJ) method (Saitou and Nei 1987). Bootstrap replicates were performed using 1000 replicates. Alignments realized in this study are available on the microbial ecology laboratory FTP server: <http://ecomicro.univ-lyon1.fr/poe/IS-phylogeny-AN-alignments.rar> for nucleic acids alignments and <http://ecomicro.univ-lyon1.fr/poe/IS-phylogeny-AA-alignments.rar> for amino acids alignments.

**Table 3: IS elements among *Burkholderia* genomes**

Insertion sequences	<i>B. cenocepacia</i> clones/strains					<i>B. multivorans</i>	<i>B. ambifaria</i>	<i>B. vietnamiensis</i>	<i>B. mallei</i>	<i>B. pseudomallei</i>	<i>B. xenovorans</i>	<i>B. thailandensis</i>	
	ET12	B&B	PHDC		Mid-west								
Family	Sub group	J2315	Bpoe717*	AU1054	HI2424	PC184	ATCC 17616	AMMD	G4	ATCC 23344	K96243	LB400	E264
IS3	ISBcen6	(2)	≥1										
	ISBcen7	2											
	ISBcen9	9	≥5										
	ISBcen15	3				1							
	ISBcen16	1						5					
	ISBcen17	1											
	ISBcen21					1 (+2)							
	ISBcen22			1	1	1							
	ISBcen23			2	2							2	
	ISBp1												
	IS1416									(2)			
	IS401						1						
	IS407	13 (+1)	≥6				3		1 (+1)	91	3 (+1)		
	ISBcen10	4					9						
	ISButh1			1	2				1 (+1)				1
	ISButh2												6
	ISBxe1											7	
	ISBxe2											1	
	ISBph1											1	
	ISBmu6						4						
	ISBmu11					2	2			2			
	ISBvi3									1			
	ISBvi4									5			
IS4	ISBvi1									5			
IS5	ISBcen20	9	≥8					(1)					
	IS402						3			11			
	ISBmu2						10 (+1)			3			
IS21	ISBcen28				1	3				(1)			
	IS408	1					3 (+1)			4			
	ISBmu1						6						
	ISBmu3						2				1 (+1)	(1)	
	ISBcen13	5								1			
IS66	ISBcen19	2											
	ISBcen14	4											
IS110	ISBcen1	1				(1)							
	ISBcen2	1	0	5	5								
	ISBcen3	2											
	ISBcen4	1 (+1)	≥3									(1)	
	ISBcen5	3											
	ISBfun1											1	
	ISBcen8	12	≥7	2	1	3	4						
	ISBma3			8	3		2			3	1 (+1)	3 (+1)	5
	ISBfun2											1	
IS256	IS1356	2		4	2		7	1	3			(1)	6 (+1)
	IS1413											2	
	ISBcen18	1				3					(1)		
	IS931									7			
	IS1490									1			
	IS406						6						
	ISRso7			2	2	1							
IS481	ISBcen25			2	2								
	ISBcen26			1	1								
	ISBxe3											5	
	ISBxe4											3	
ISL3	ISBma1					2			1	17 (+3)	6		7
others	ISBma2									46	5	(2)	
	ISBfun4											1	
	IS1071						(1)					2	
	unknown	1		6	4				14				
<b>Total</b>		<b>82</b>	<b>≥31</b>	<b>34</b>	<b>26</b>	<b>20</b>	<b>65</b>	<b>7</b>	<b>74</b>	<b>157</b>	<b>22</b>	<b>35</b>	<b>26</b>

\*: IS elements among the B&amp;B clone were identified by DNA blot analyses. Numbers in brackets represent partial IS elements.

### 3. Results and Discussion.

*Burkholderia* genomes are known to harbour several IS elements. Holden et al., (2009) recently reported several IS-inactivated gene sequences among *B. cenocepacia* strain J2315 genome, a strain isolated from a CF patient. These IS-driven inactivations could have been selected by the strong selective pressures taking place among *B. cenocepacia* colonized habitats including the CF lungs or could have occurred by chance, without much selection except the one related to lethal inactivation. Here, we present a study that investigated the genomic organization of IS elements among *B. cenocepacia* sequenced genomes, and attempted at establishing the benefits of the related insertion events. Conservation of the insertion events among *B. cenocepacia*, the Bcc, and other *Burkholderia* species was investigated using full genome sequences. These analyses were completed by experimental investigations of the *B. cenocepacia* B&B clone IS genomic organization. These analyses suggested similarities in the IS genomic organization of the B&B and ET12 clones, and a relation with their emergence.

#### 3.1. IS numbers among *Burkholderia cenocepacia* genomes

Fully assembled *B. cenocepacia* genomes of strains J2315, AU1054, HI2424 and PC184 were used to estimate IS numbers and build a set of reference IS to investigate their spread among the *Burkholderia*. Sequenced genomes of the following *Burkholderia* strains were used to complete the analyses: *B. multivorans* ATCC17616, *B. vietnamiensis* strain G4, *B. ambifaria* strain AMMD, *B. mallei* strain ATCC23344, *B. pseudomallei* K96243, *B. xenovorans* LB400, *Burkholderia* sp. 383 and *B. thailandensis* E264. Contigs of *Burkholderia* genomes under sequencing (in January 2009) were not included in these investigations. Shotgun DNA sequencings do not cover entire repeated regions making hard quick assembly of regions with repeated sequences. IS were detected through BLASTN searches and ACT analyses. These investigations were completed by DNA blot analyses of IS numbers of certain families/sub-groups among the *B. cenocepacia* B&B genome. DNA probes used in these analyses were obtained by PCR from strain J2315 DNA, using primers indicated in Table 2.

*Burkholderia* genomes were found to harbour between 7 to 157 IS elements. *B. cenocepacia* J2315 was found to harbour 6 families (IS3, IS256, IS110, IS5, IS21, IS66) of IS for a total number of 82 (including 4 partial sequences) elements. The *B. cenocepacia* genomes of the PHDC strains AU1054 and HI2424 were found, respectively, to harbour 34 and 26 IS elements, being part of, respectively, 4 (IS3, IS256, IS110, and IS481) and 5 (IS3, IS256, IS110, IS21, and IS481) established families, and families not yet defined. *B. cenocepacia* genome of strain PC184 of the Midwest clone showed 20 IS elements being part of five distinct families (IS3, IS256, IS110, IS21, ISL3). IS numbers can thus vary significantly within *B. cenocepacia*, with some genomes showing as

much as 4 times more IS elements than others. The families detected in all sequenced *B. cenocepacia* strains were IS3, IS110, and IS256. The IS66 and IS5 families were only recorded among strain J2315, the IS481 among the sequenced PHDC genomes, and ISL3 among the PC184 genome (Table 3). However, *Burkholderia* other than *B. cenocepacia* were found to harbour these IS.

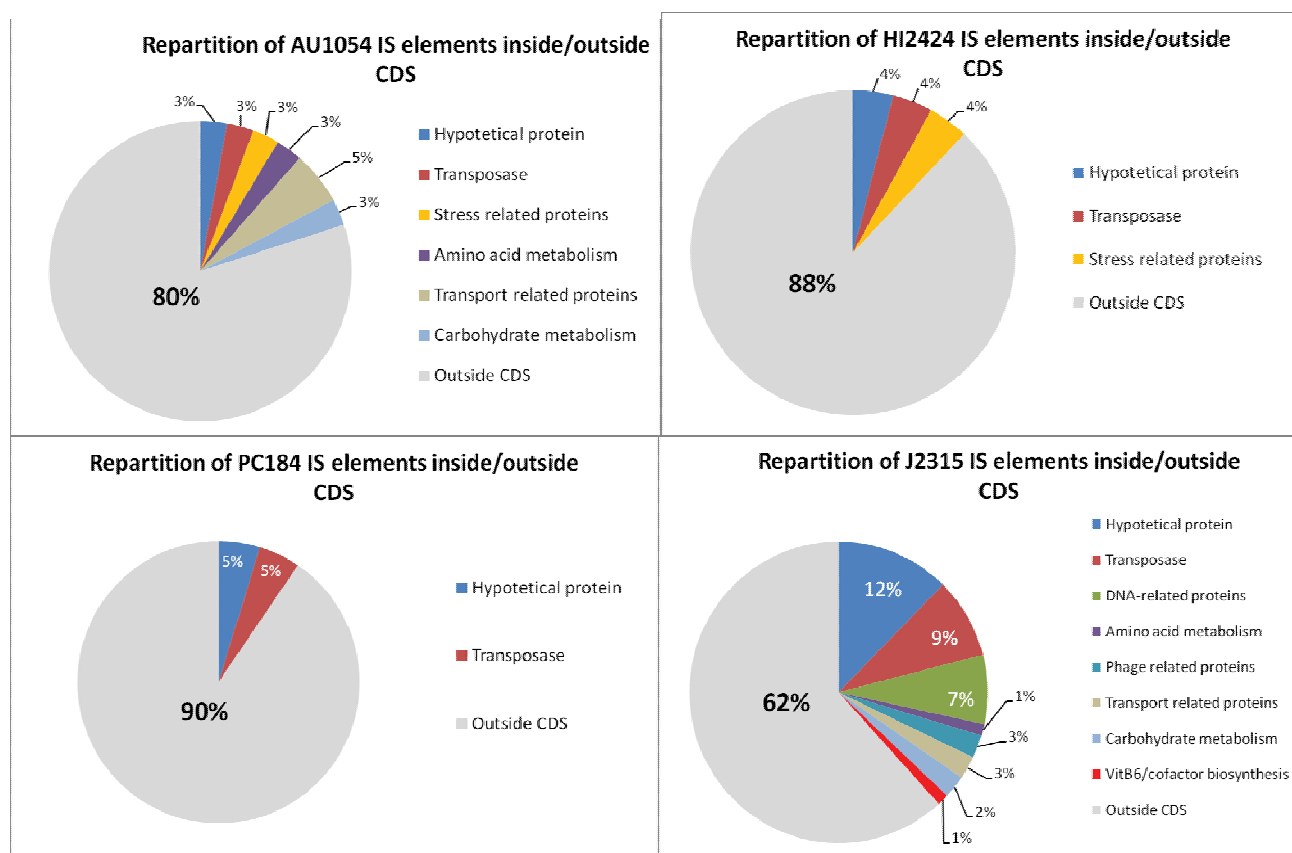
IS DNA probings of genomic DNA blots of strain bpoe717 of the B&B clone were performed to estimate the copy numbers of three IS families, IS3, IS5 and IS110. DNA probings with selected IS elements of the IS3 family of the J2315 genome of *B. cenocepacia* gave hybridizing signals indicative of more than twelve IS elements in bpoe717. DNA probings with selected IS elements of the IS5 family of the J2315 genome gave hybridizing signals indicative of more than eight IS elements, and DNA probings with selected J2315 elements of the IS110 family revealed at least 10 elements. These hybridizing data showed that IS5 elements restricted, so far, to the ET12 clone of *B. cenocepacia* could also be detected in the B&B clone, supporting the idea of a high genetic proximity between these two clones. The hybridization results with elements of the IS3 and IS110 families showed high copy numbers of these elements in bpoe717. Such high numbers were not surprising for the IS110 family but were considered a trademark of ET12 for the IS3 family. This result further supports the idea of a high genetic proximity between the ET12 and B&B clones.

The *B. cenocepacia* IS composition could be compared with the ones of four other species of the Bcc, and of species outside the Bcc like *B. pseudomallei* and *B. xenovorans* (Table 3). *B. lata* strain 383 genome, a close relative of *B. cenocepacia*, did not show any IS element in its genome. This absence of IS element among *B. lata* strain 383 might be indicative of cellular processes that can prevent IS genomic colonization. The other Bcc species showed 7 to 74 IS elements. IS families recorded among *B. multivorans* were IS3, IS256, IS110, IS5, IS21, were IS3, IS256, and IS5 for *B. ambifaria*, and were IS3, IS4, IS256, IS110, IS5, IS21 and ISL3 for *B. vietnamiensis*. The sequenced genome of this latter Bcc species was found to harbour more IS families than the one of *B. cenocepacia* strain J2315. The other *Burkholderia* genomes showed lower numbers of IS families, with *B. pseudomallei* (22 IS elements) and *B. xenovorans* (35 elements) harbouring five families. *B. mallei* ATCC 23344 genome which showed the highest number of IS elements (157 IS elements) showed only two known families. *B. thailandensis* E264 genome showed 26 elements and four families.

These variations in IS numbers and families indicate a relationship with the evolutionary history of the various strains that were investigated. These relationships might be related to recent acquisitions of DNA through horizontal transfers or amplification of IS elements of an existing family. In the next section, we present analyses of these phenomena. The localization of these IS inside and outside coding sequences (CDS) was investigated for the available *B. cenocepacia* genomes (Fig. 1). These analyses showed a high repartition of more than 80% of the IS elements



outside CDS among the PC184 and PHDC (HI2424 and AU1054) genomes. However, 38% of J2315 IS elements were found inside CDS. IS found among CDS showed a strong bias towards a colonization of other IS elements (Fig. 1). Other trends in the colonization of particular gene classes were not clear suggesting a lack of correlation with such classes. Other constraints like DNA sites targeted by the IS transposases, and methylation of DNA, could have a more significant effect on the distribution of these elements. These issues were partly investigated in the section below by analyzing the direct repeats detected at the proximity of the IS elements (Table 4). These analyses showed biases among IS DNA targets (or direct repeats) often matching the separation observed by phylogenetic analyses of the IS elements of a same family into sub-groups.



**Fig. 1: Gene classes of disrupted ORF among *B. cenocepacia* strains investigated.**

AU1054 and HI2424 (PHDC clones), and PC184 (Midwest clone) genomes. For PHDC and Midwest genomes, disrupted genes are potentially under-estimated because of an annotation still in progress.

### ***3.2. IS family sub-groups broadly distributed among B. cenocepacia***

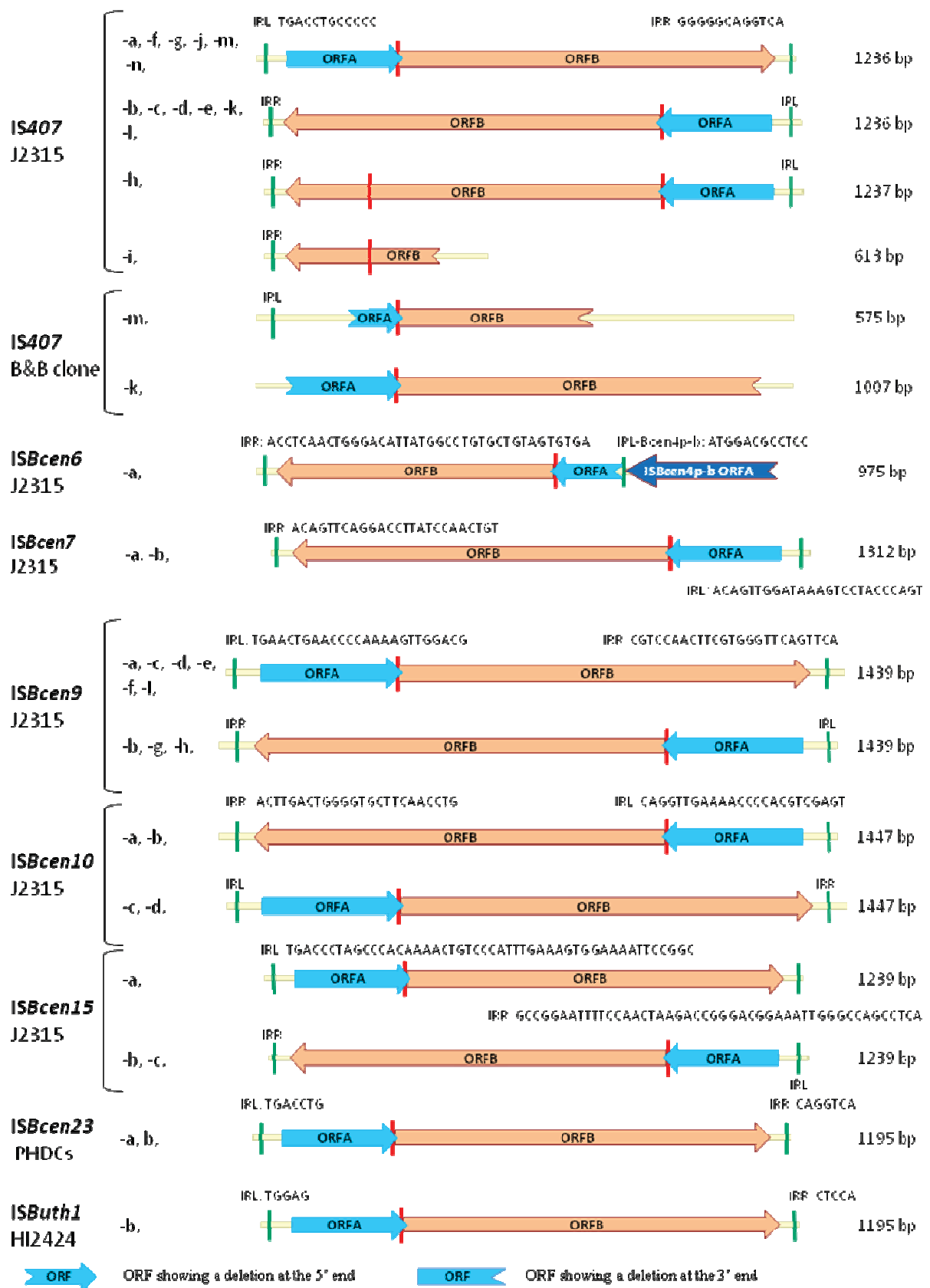
Significant DNA sequence variations were observed among each IS family, and were used to investigate phylogenetic relationships. Phylogenetic analyses of the inferred amino acids (AA) sequences of the encoded IS transposases allowed the definition of sub-groups among most families. Distribution of these sub-groups was investigated among *B. cenocepacia* and other *Burkholderia*. These analyses allowed identification of IS founder elements (most likely ancestral sequence) among *B. cenocepacia*, and to investigate their likely involvement in the emergence of clonal lineages. DNA targets (or direct repeats) of these IS elements are indicated in Table 4.

#### **3.2.1. IS3**

The IS3 family was found the most diverse among the *B. cenocepacia* J2315 genome, with at least four large clusters (named I, II, III, and IV) of related transposase AA sequences which could then be sub-divided into 14 significant sub-groups, by molecular phylogeny (Fig. 3). Relationships within these sub-groups could not be resolved using the AA data set of the IS-encoded transposases but could be partly resolved using the DNA sequences. This allowed identification of highly active elements which appeared to have undergone recent amplification events. The genetic and genomic organizations of these IS are presented respectively in Fig. 2 and Fig. 5.

Transposases of cluster I were divided into an IS407 sub-group, an IS*Bcen22* sub-group, an IS*Bcen15* one, and an IS*Bcen23* one (Fig. 2; Fig. 3). Most sub-groups of this cluster showed a narrow distribution and were detected in a few strains of *B. cenocepacia*. AA sequence similarities were observed with encoded transposases of IS elements of cluster I from other genera like *Ralstonia solanacearum* for IS*Bcen15* and IS407, and *Nitrosospora multiformis* for IS*Bcen22*. Among the IS*Bcen23* sub-group, two distinct founder elements, a and b, could be observed among the PHDC strain HI2424, while only the “a” founder element appeared responsible for the two copies found among the other strain of PHDC investigated i.e. AU1054 (Fig. 3). The IS407 sub-group was found the most widespread and showed the highest number of elements among the *Burkholderia* but was not recorded among all *B. cenocepacia* clones. This is indicative of genomic instability among the DNA regions colonized by these elements. This led us to investigate more deeply the behaviour of this sub-group. These analyses are presented in the sub-section below.

Transposases of cluster II were divided into an IS*Bcen9* sub-group, an IS*Bcen10* one, IS*Bmu11* one, and IS*Bcen16* one (Fig. 2; Fig. 3). Interestingly, a common IS element apparently led to the emergence of the IS*Bcen9*, IS*Bcen10* and IS*Bcen16* sub-groups, with these last two sub-groups being closely related and also detected in other Bcc species.



**Fig. 2: Main features of IS3 elements detected among *B. cenocepacia* strains.**

Inverted DNA repeats are indicated by a green vertical line; red vertical lines are used to indicate a frameshift (FS); nonsense mutations (amber type) are indicated by a brown vertical line (Ab). ORFs are indicated by arrows and ORFs with a deletion are represented by arrows without a vertical extremity. ORFA: total transposase sequence (IS5, IS110, IS256 families) or amino-terminal end (IS3, IS21 families), ORFB: carboxy-terminal end (IS3, IS21 families), ORFC: carboxy-terminal end (IS66 family). ORF orientation of IS elements are according to the main orientation of the genomes, beginning from the origin of replication. PHDC indicates genomes of the HI2424 and/or AU1054 strains, and Midwest the one of strain PC184.

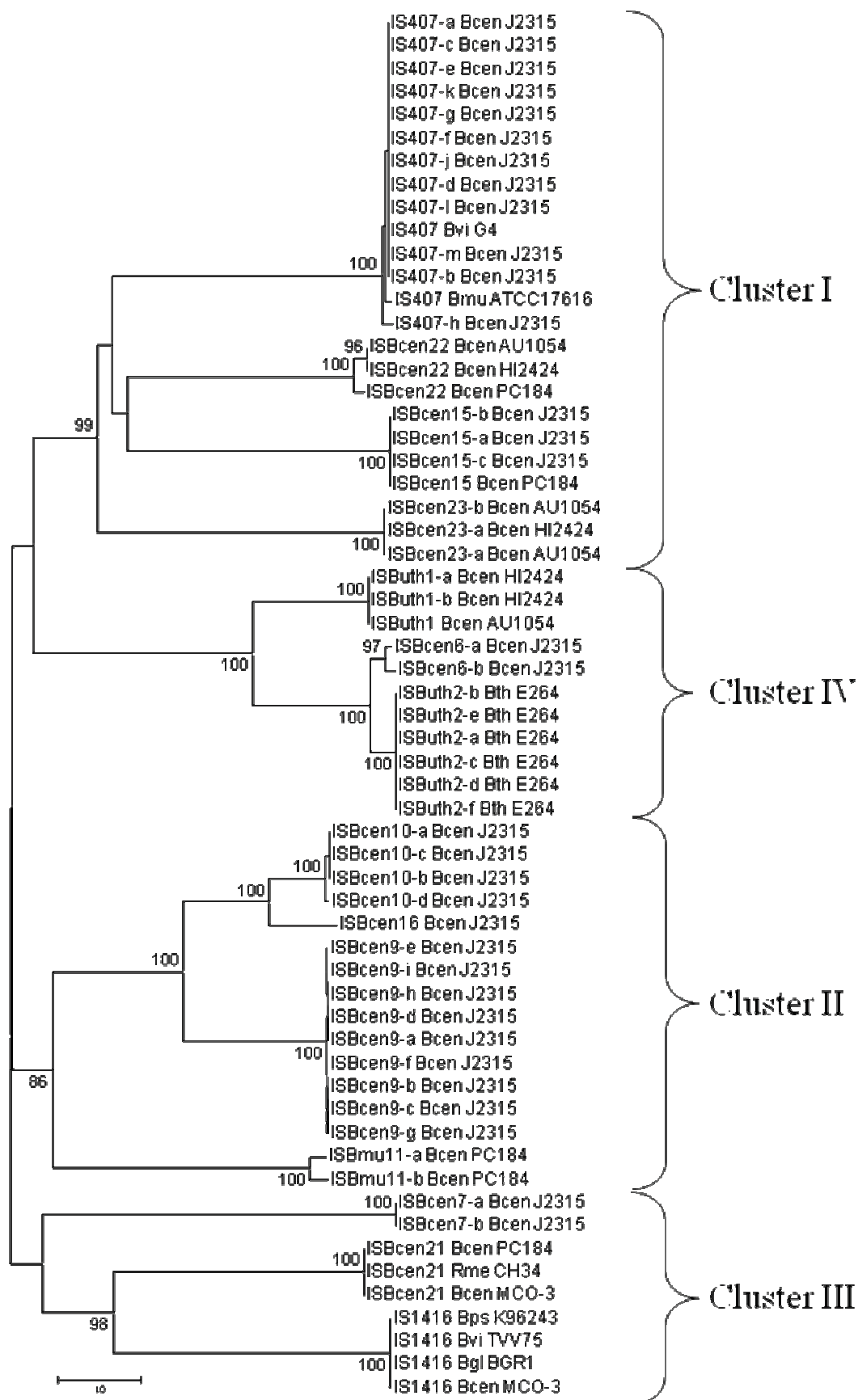
Analysis of the DNA sites targeted by these IS elements i. e. the direct repeats observed after insertion, revealed similarities for the *ISBcen10* and *ISBcen16* sub-groups (Table 4). However, the *ISBcen9* DR were found more AT-rich than those of these two sub-groups. Interestingly, relatively high IS numbers were observed for *ISBcen9* and *ISBcen10*, with respectively, nine and four elements detected in the genome of J2315. The *ISBcen16* was found in a single copy in the J2315 genome but five elements were detected in the *B. ambifaria* AMMD genome. This sub-group was the most important, in number, among this last genome. Each of these elements apparently emerged from a single but distinct founder element. Similar IS elements were observed among other bacterial genera like the one of *Ralstonia metallidurans* CH34 (renamed *Cupriavidus metallidurans* by Vandamme and Coenye 2004).

Transposases of cluster III were divided into an *ISBcen7* sub-group, an *ISBcen21* one, and an *IS1416* one (Fig. 2; Fig. 3), with the last two sub-groups being part of a common significant sub-lineage. The position of *ISBcen7* in this sub-group was not supported by high bootstrap values. The *ISBcen7* sub-group could not be detected in other Bcc strains but similar elements could be detected in species like *B. phytofirmans* strain PsJN and *R. pickettii* strain 12j. *ISBcen21* was observed among strain PC184 and MCO-3 (two elements) of *B. cenocepacia*, and could also be detected among *C. metallidurans* CH34 (1 element). The *IS1416* sub-group was recorded among *B. pseudomallei*, *B. vietnamiensis* (strain TVV75), *B. phymatum* STM815, *B. glumae* BGR1 (5 copies) and *B. cenocepacia* MCO-3 (5 copies).

Transposases of cluster IV were divided into three sub-groups, *ISBcen6*, *ISButh1* and *ISButh2* (Fig. 2; Fig. 3). IS of this cluster could be detected in several species of *Burkholderia* like *B. thailandensis*, *B. ambifaria* (strain MEX-5), *B. oklahomensis*, and, in other genera like *Ralstonia*. *ISBcen6* was found restricted to the J2315 genome of *B. cenocepacia* and to *B. ambifaria* (strain MEX-5) but to be a close relative of the *ISButh2* sub-group observed among *B. thailandensis* (n=6) (Fig. 3). *ISButh1* was detected among the PHDC genomes (Fig. 4a) and those of *B. vietnamiensis*, *B. thailandensis*, and of strain 1106b of *B. pseudomallei*. Similarities were also observed with elements found among *C. metallidurans* strain CH34. The two elements of the *ISBcen6* sub-group of the J2315 genome were found physically linked, with the *ISBcen6-a* harbouring a partial *ISBcen4* element responsible of the disruption of its orf-b (Fig. 2 and Fig 4c) and the *ISBcen6-b* harbouring a disruption of its orf-a by the *ISBcen4-a* (Fig 4c). *ISBcen4* is part of the *IS110* family which is described in the next section.

IS3 elements of clusters I, II, and IV were screened among the B&B genome. Elements of all three clusters could be detected: *IS407* (cluster I), *ISBcen9* (cluster II), and *ISBcen6*, (cluster IV). At

least six elements of *IS407*, five elements of *ISBcen9*, and one element of *ISBcen6* could be detected among the genome of strain bpoe717 of the B&B clone (Table 3 and annexe C1). DNA sequence analyses of some bpoe717 DNA sequences of *ISBcen9* showed 100% identity with the ones of J2315 suggesting a recent acquisition or a acquisition prior their divergences. Data concerning the *IS407* are presented in the following sub-section.



**Fig. 3: Neighbor-Joining phylogenetic tree of IS3 family insertion sequences.**

The amino acid sequences were retrieved from sequenced *Burkholderia* genomes or the GenBank database. A total of 110 sites were analyzed (with global removal of gap-containing sites). Distances are proportional to evolutionary divergences expressed in substitutions per 100 sites. Bootstrap values higher than 85% are given. Four clusters were defined, and are indicated. Bcen = *B. cenocepacia*; Bth = *B. thailandensis*; Bgl = *B. glumae*; Bvi = *B. vietnamiensis*; Rme = *R. metallidurans*; Bps = *B. pseudomallei*; Bmu = *B. multivorans*

**Table 4: Characterization of IS direct DNA repeats (DR) among *B. cenocepacia* J2315 genome.**

Family	Sub-Group	DR	Chr.	ORF disruption	
IS3	IS <i>Bcen7-a</i>	TGT	1	no	
	IS <i>Bcen7-b</i>	TGT	2	no	
	cons	TGT			
	IS <i>Bcen9-a</i>	.TTG	1	no	
	IS <i>Bcen9-b</i>	.TTT	1	yes	
	IS <i>Bcen9-c</i>	GAAA	1	no	
	IS <i>Bcen9-d</i>	GCAA	1	no	
	IS <i>Bcen9-e</i>	.CTT	1	no	
	IS <i>Bcen9-f</i>	TTTC	1	no	
	IS <i>Bcen9-g</i>	GTAG	1	no	
	IS <i>Bcen9-h</i>	.GAA	2	no	
	IS <i>Bcen9-i</i>	GCAA	3	no	
	cons	NTT/AA			
	IS <i>Bcen15-a</i>	none	1	no	
	IS <i>Bcen15-b</i>	none	1	yes	
	IS <i>Bcen15-c</i>	none	2	no	
	IS <i>Bcen16</i>	GTG	1	no	
	IS <i>Bcen17</i>	AAC	1	yes	
	IS407-a	G	1	no	
	IS407-b	CGCA	1	yes	
	IS407-c	CGTA	1	yes	
	IS407-d	CGAG	1	yes	
	IS407-e	G	1	yes	
	IS407-f	GATC	1	no	
	IS407-g	ACCG	1	yes	
	IS407-h	none	1	no	
	IS407-i	none	1	no	
	IS407-j	GCAA	1	yes	
	IS407-k	GGAC	1	yes	
	IS407-l	none	2	yes	
	IS407-m	none	3	no	
	IS407-n	C	Plasmid	no	
	cons	CGNN			
	IS <i>Bcen10-a</i>	GCTT	1	no	
	IS <i>Bcen10-b</i>	.CTA	1	no	
	IS <i>Bcen10-c</i>	ATCA	2	no	
	IS <i>Bcen10-d</i>	GTC.	3	no	
	cons	GC/TC/TA			
	IS256	IS1356-a	ACGGCGC.	1	yes
		IS1356-b	CGAACC CA	1	yes
	cons	NNNNCNCN			
	IS110	IS <i>Bcen18</i>	GAAAGGCAGC	1	yes
		IS <i>Bcen1</i>	GGACTTGT	2	no
IS <i>Bcen2</i>		none	3	yes	
IS <i>Bcen4-a</i>		none	3	yes	
IS <i>Bcen4p-b</i>		none	3	yes	
IS <i>Bcen5-a</i>		G	1	no	
IS <i>Bcen5-b</i>		G	1	no	
IS <i>Bcen5-c</i>		G	2	no	
IS <i>Bcen8-a1</i>		none	1	yes	
IS <i>Bcen8-b</i>		none	1	no	
IS <i>Bcen8-c</i>		none	1	no	
IS <i>Bcen8-d</i>		none	1	no	
IS <i>Bcen8-e</i>		none	1	no	
IS <i>Bcen8-f</i>		none	1	no	
IS <i>Bcen8-a2</i>		none	1	yes	
IS <i>Bcen8-g</i>		none	2	yes	
IS <i>Bcen8-h</i>		none	2	no	
IS <i>Bcen8-i</i>		none	3	no	
IS <i>Bcen8-j</i>		none	3	no	
IS <i>Bcen8-k</i>		none	3	yes	
IS5	IS <i>Bcen20-a</i>	AAT	1	no	
	IS <i>Bcen20-b</i>	AAA	1	yes	
	IS <i>Bcen20-c</i>	AAT	1	yes	
	IS <i>Bcen20-d</i>	TTA	1	no	
	IS <i>Bcen20-e</i>	ATT	1	no	
	IS <i>Bcen20-f</i>	TTA	1	yes	
	IS <i>Bcen20-g</i>	AAT	1	no	
	IS <i>Bcen20-h</i>	TTA	2	no	
	IS <i>Bcen20-i</i>	TAA	2	yes	
cons	A/TA/TA/T				
IS21	IS408	none	1	no	
	IS <i>Bcen13-a</i>	CCAA	1	no	
	IS <i>Bcen13-b</i>	GCTT	1	yes	
	IS <i>Bcen13-c</i>	TCTA	2	yes	
	IS <i>Bcen13-d</i>	AGCC	2	no	
	IS <i>Bcen13-e</i>	ACAT	3	yes	
cons	ACA/TA/T				
IS66	IS <i>Bcen14-a</i>	TAAGTCCT	1	no	
	IS <i>Bcen14-b</i>	ACGGCTGG	1	yes	
	IS <i>Bcen14-c</i>	GGTCATGC	1	no	
	IS <i>Bcen14-d</i>	GGACTTGT	2	yes	
cons	GGACTTGT				

Seven IS elements outside ORF and harboring no DR have been excluded of this table (IS*Bcen6*, IS*Bcen3*, IS*Bcen19*). Cons: consensus direct repeats DNA sequences. Only nucleotides being found in more than 40% for a single site of a DR were considered in the definition of these consensus.

### *The particular case of the IS407 sub-group*

In the IS407 sub-group, four clusters of these elements could be inferred from the available genome sequences (Fig. 6). Three of these were recorded among Bcc genomes, and one was specific of the *B. pseudomallei*/*B. mallei* genomes. These IS elements showed a strong attraction for “CG” DNA targets as observed in Table 4. One cluster, named “a” (active), was found highly active and led to several copies in the genomes of *B. cenocepacia* J2315 (Fig. 2) and *B. vietnamiensis* G4. This highly active group matches the one previously observed by Miché et al. (2001), in *B. vietnamiensis*, using an IS-trapping system. The other cluster named “b” grouped elements found among the *B. cenocepacia* J2315 genome and the one of *B. multivorans* ATCC17616. However, the elements found in J2315 appeared to be inactive.

The third cluster was found among the fully sequenced genome of strain ATCC17616 of *B. multivorans*. So far, it appears restricted to this species. It is noteworthy that Kenna et al., (2006) investigated the distribution of the IS407 sub-group among a panel of *Burkholderia* strains, by DNA probings. This sub-group was not found in the strains of *B. cepacia*, *B. dolosa*, *B. ambifaria* and *B. pyrrocinia* that were tested. However, all *B. cenocepacia* ET12 strains that were tested (i. e. J2315, LMG18826, K56-2 and LMG18827) were found to harbour this sub-group, and strains allocated to other clones or groups were confirmed to rarely harbour this sub-group (Kenna et al., 2006). Higher IS407 copy numbers were also reported among ET12 strains than among any other Bcc.

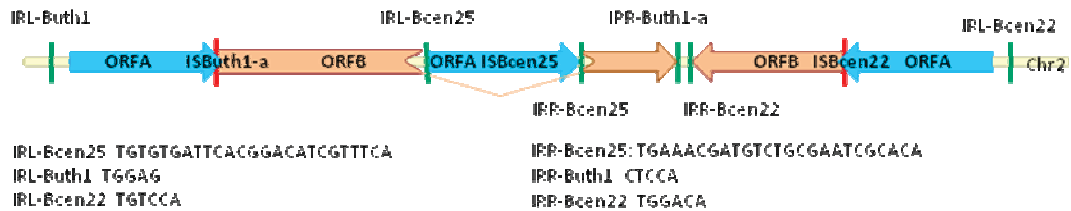
#### - *the B. cenocepacia J2315 IS407 inactive elements*

The two inactive elements, IS407-h and IS407-i, found in J2315 are physically linked (Fig. 5; Table 5), and show major sequence alterations (Fig. 2). IS407-h shows nucleotide insertions at positions 497 and 498 of ORF-b (BCAL2581) encoding the transposase, creating a frameshift which is inactivating this CDS. IS407-i shows the same nucleotide insertions but also a size reduction including a complete deletion of orf-a and a partial deletion of ORF-b (BCAL2581) encoding the transposase. One of these two IS is likely to be the result of a duplication of the more ancestral element. This hypothesis is supported by the conservation of the two nucleotide insertions in orf-b in both, the IS407-h and IS407-i elements.

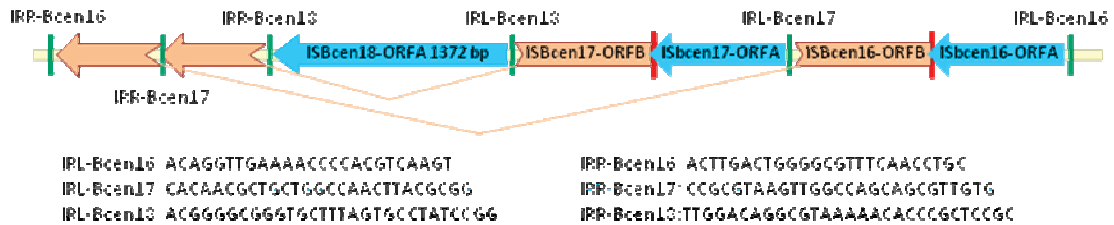
These inactive IS could be genetic markers of *B. cenocepacia* diversification into clonal complexes. A PCR method, targeting the flanking regions of these IS, was thus developed to follow these insertions among other strains. These elements were found outside coding sequences. The PCR approach showed a conservation of these inactive IS among other strains of ET12 and among the B&B clone, which was previously suggested as being closely related to ET12 (Table 6, annex C1 and C2). Both elements were also confirmed to be present among the



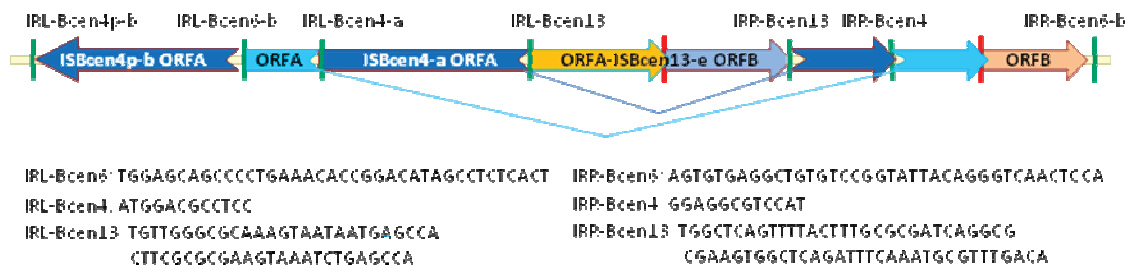
**a) Case of ISButh1-a, ISBcen25 and ISBcen22 among PHDC strains**



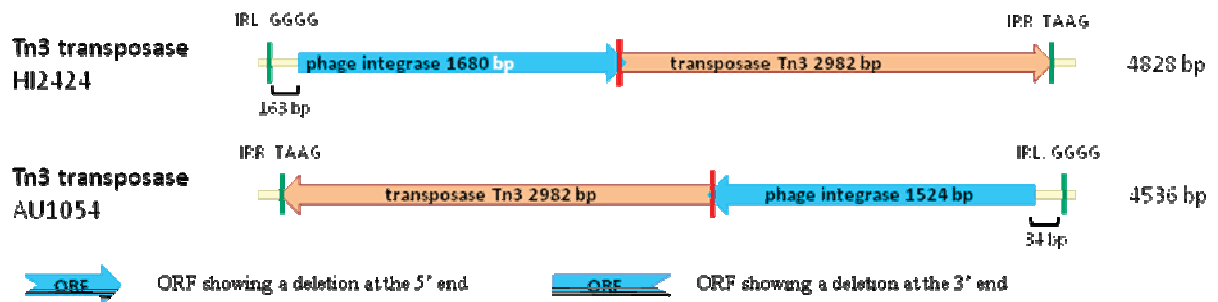
**b) Case of ISBcen18/ISBcen17/ISBcen16 among J2315**



**c) Case of ISBcen6-b/ISBcen4-a, -b/ISBcen13-e among J2315**



**Transposase Tn3 among PHDC strains**



**Fig. 4: Main features of DNA regions showing particular transposase CDS organization among *B. cenocepacia* strains.**

Inverted DNA repeats are indicated by a green vertical line; red vertical lines are used to indicate a frameshift. ORFs are indicated by arrows and ORFs with a deletion are represented by arrows without a vertical extremity. ORFA: full transposase sequence (IS5, IS110, IS256 families) or amino-terminal end (IS3, IS21 families) with ORFB being the carboxy-terminal end (IS3, IS21 families). ORF orientation of IS elements are according to orientation from the origin of replication. PHDC indicates genomes of the HI2424 and AU1054 strains, and Midwest the one of strain PC184.

bpoe717 genome of the B&B clone by DNA sequence analyses (Fig.6, annex C1 and C2). The IS407-h element was also found among strain LMG21461 which was isolated from a CF patient, and found part of *B. cenocepacia* IIID *recA* PCR RFLP group. These data suggest selective forces maintaining the IS407-h DNA region among the *B. cenocepacia* ET12 and B&B ones, but its counter-selection among strains distantly related from these epidemic clones. Distantly related strains did not show the presence of this IS and of its targeted DNA region (Table 6), suggesting their deletion. Regarding IS407-i, all strains of *B. cenocepacia* tested showed its presence (Table 6), except LMG18828, PC184, HI2424 and AU1054. These data thus suggest selective forces maintaining this insertion and its DNA region. DNA sequence analyses of IS407-i DNA regions among *B. cenocepacia* suggest an origin that can be traced back to the root of its emergence. This element is thus likely to be the ancestral one, and the one at the origin of the emergence of IS407-h (duplication event).

- *the B. cenocepacia J2315 IS407 active elements*

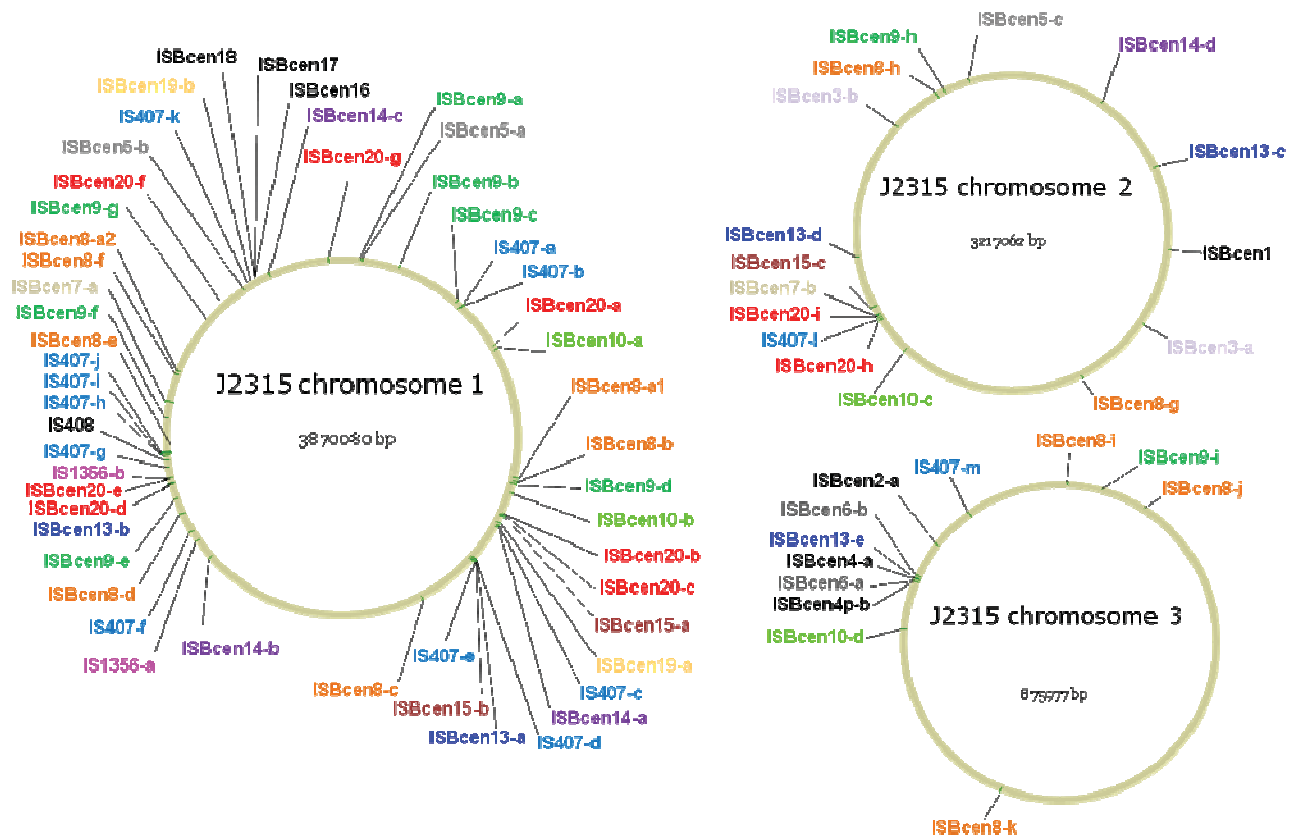
The associations observed between the emergence of the ET12 and B&B clones, and the inactive IS407 elements, led us to complete these investigations by studying the genomic localization of the active “a” elements. The genomic localization of these elements is indicated in Table 5 and Fig. 5. The IS407-j element was found physically close to the inactive IS407 elements. The IS407-a and IS407-b elements were found to be physically linked as well as the IS407-d and IS407-e elements. 8 IS407 “a” elements were found outside CDS, and 6 inside CDS (Table 5). IS407 elements outside coding regions might have a strong polar effect on the downstream CDS; this IS carries an outwardly directed promoter that can activate closely related genes (Wood et al., 1990). For example, IS407-f may have affected the expression of an rRNA 5S gene and modified the expression pattern of other genes. Regarding the IS insertions among J2315 CDS, IS407-e is inserted within a helicase gene. IS407-k is found in a gene involved in capsule biosynthesis, confirming data obtained by DeShazer et al., (2001). IS407-b is inserted in a gene encoding a methylase (*hsdM*) of a type I restriction-modification (RM) system, and is located upstream the *hsdS* and *hsdR* genes which play an essential role in this RM system. All three genes are required for the restriction activities, while only *hsdM* and *hsdS* are required for the methylation activity (for review see Wilson and Murray 1991, Bickle and Kruger 1993). In consequence, the IS407 insertion in this *hsd* operon inactivated both methylation and restriction activity. These RM systems protect bacteria against foreign DNA including phage DNA (Bickle and Kruger 1993). *B. cenocepacia* J2315 may thus be more sensitive to bacteriophage infection. An ACT comparison was performed between IS407 DNA sequences, the J2315 genome, and the HI2424 genome (Fig. 7), in order to determine the presence or not of the DNA regions colonized by IS407 in J2315 among the PHDC genome. This analysis showed that most IS407 elements are located in regions not detected in the PHDC genome. This analysis showed that orthologs of the RM system, described above, are not

found in the PHDC *B. cenocepacia* genomes. Only IS407-f (Table 6) was found inserted in a DNA region also observed in HI2424, AU1054 and PC184 genomes (but without an IS407 insertion). These observations led us to further investigate the distribution of J2315 DNA regions targeted by active IS407 elements among a panel of *B. cenocepacia* strains, including additional *recA* PCR RFLP IIIA strains such as the B&B clone strains, and strains from the IIIB, IIIC, and IIID *recA* groups. These IS407 insertion sites were screened among these bacterial genomes using a PCR approach targeting the IS flanking regions or using a tripartite ACT analysis (Table 6; Fig. 7, annex C1 and C2). PCR products ranging from 1000 to 1802 bp were obtained when a site was colonized by the IS407, and products from 193 to 700 bp when the site was not colonized by an IS. On one hand, these analyses showed IS407-f, IS407-j and IS407-c loci to be conserved among most strains: 11/13, 8/13 and 8/13, respectively. This suggests a presence of these loci quite early in the evolution of *B. cenocepacia*. On the other hand, these analyses also showed the presence of more recent loci showing a more restricted distribution like the IS407-b and IS407-g loci only found among the ET12 clone, and the IS407-l locus found among the ET12 and B&B strains. A dendrogram of the distribution of all loci colonized by the IS407 was built to better visualize the similarities in the distribution profiles. Interestingly, this cladogram shows a good proximity in the distribution profiles of the loci among the *B. cenocepacia* ET12 and B&B clones, and IIID strains (Fig. 8). The IIIA strain named H111 did not show a profile in line with those of the ET12 and B&B clones, suggesting an early divergence from these last two clones. This proximity is in line with phylogenetic trees built to investigate *B. cenocepacia* infra-specific diversity e. g. Graindorge et al., (2009). This cladogram also shows a high level of conservation of the IS407 targeted DNA loci among strains of the ET12 clone (Table 4; Fig. 8), confirming their close proximity. These analyses were completed by a further investigation of these loci by visualizing the number of IS407 found among these loci (Fig. 9). This analysis shows one strain of ET12, LMG18826, to have the same IS407 genomic organization as the one of strain J2315 but another strain, LMG12614, to have a profile much different from these strains. DNA sequence analyses confirmed this conservation and showed high sequence identities between the LMG18826 IS407 insertion sites and those of J2315 (annexe C2). Interestingly, the bpoe717 IS407 profile was found close to the one of J2315. Regarding strain LMG12614, while all the J2315 loci showing an IS407 in J2315 could be detected (Fig. 8), only one site, the IS407-n locus, out of 12 was colonized by an “active” IS407 element (Table 6).

**Table 5: Characterization of IS407 flanking ORFs in the J2315 genome.**

Strain	ORF upstream	IS407 ORFs	ORF downstream
J2315			
IS407-a	BCAL0409 hypothetical protein	BCAL0410-0411	BCAL0412 hypothetical protein
IS407-b	0414 HsdM DNA methylase:Type I restriction-modification system	BCAL0415-0416	0414 HsdM DNA methylase:Type I restriction-modification system ZP_00427263
IS407-c	1169 conserved hypothetical protein (pseudogene)	BCAL1171-1170	1169 conserved hypothetical protein (pseudogene)
IS407-d	1302 hypothetical protein	BCAL1304-1303	1302 hypothetical protein
IS407-e	1308 putative SNF2-related helicase (pseudogene)	BCAL1313-1312	1308 putative SNF2-related helicase (pseudogene)
IS407-f	2309 transcriptional regulator, MerR family	BCAL2310-2311	r2231a rRNA 5S
IS407-g	2533 Peptidase M24A, methionine aminopeptidase, subfamily 1 YP_555027	BCAL2534-2535	2536 alpha/beta hydrolase with Thioesterase domain
IS407-h	2580 hypothetical protein	BCAL2581-2583 (frameshift)	2584 Metallo-beta-lactamase
IS407-i	2588 putative transposase (fragment)	BCAL2589 (partial)	2591 conserved hypothetical protein (pseudogene)
IS407-j	2591 conserved hypothetical protein (pseudogene)	BCAL2592-2593	2591 conserved hypothetical protein (pseudogene)
IS407-k	3223 Capsule polysaccharide biosynthesis ABB07469	BCAL3225-3224	3223 Capsule polysaccharide biosynthesis ABB07469
IS407-l	BCAM1914 hypothetical phage protein (pseudogene)	BCAM1916-1915	BCAM1914 hypothetical phage protein (pseudogene)
IS407-m	BCAS0717 hypothetical protein	BCAS0718-0719	BCAS0720 putative lipoprotein (fragment)
IS407-n	pBCA082 conserved hypothetical protein	pBCA083-084	pBCA085 putative carbamoyltransferase

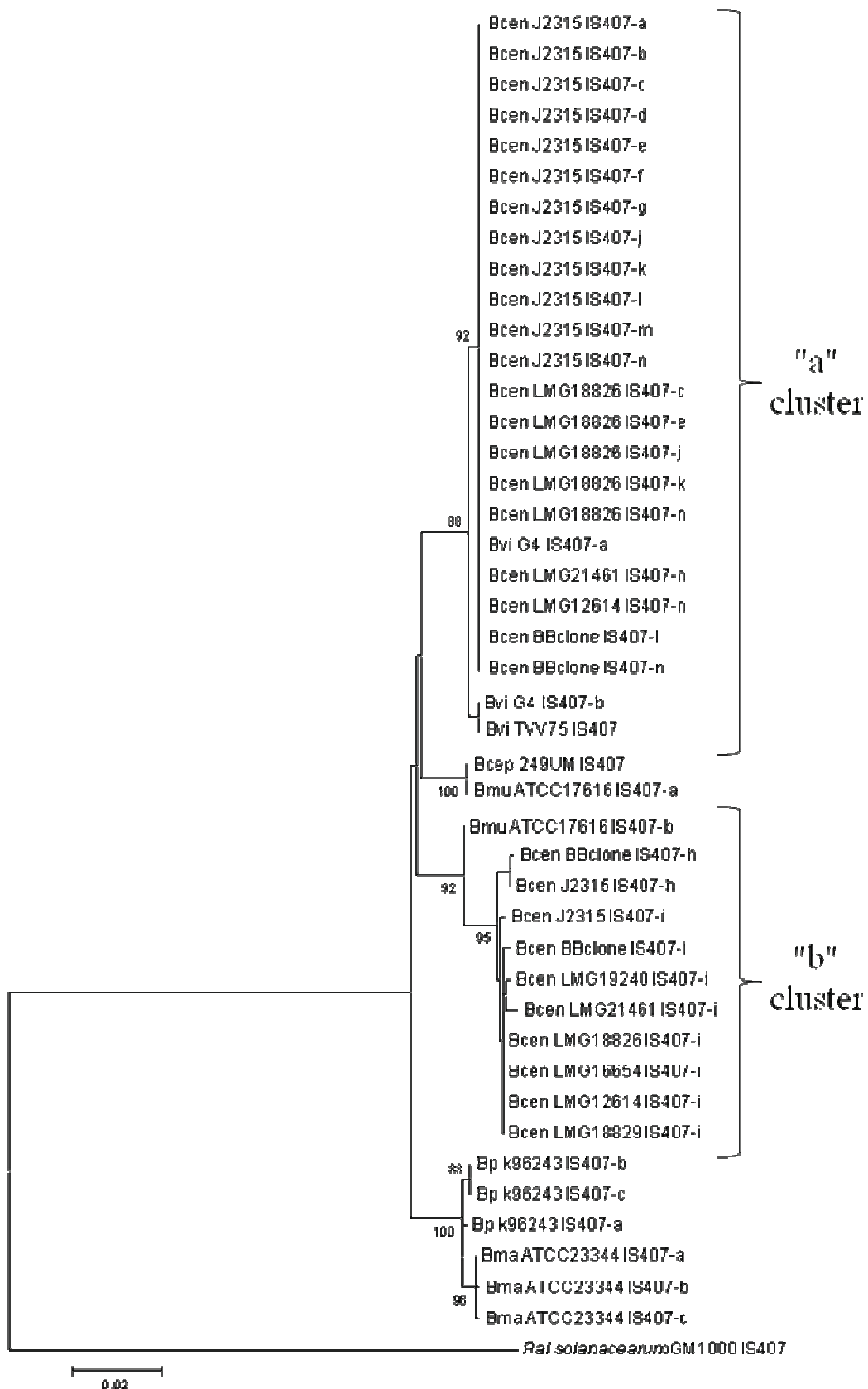
Flanking ORFs were characterised by BLASTP against Genbank and genome annotations Holden *et al.* 2009



**Fig. 5: Genomic distribution of IS elements among *B. cenocepacia* J2315 genome.**

IS407-n copy is not shown because of its localization on a plasmid. Copies of IS elements belonging to a same group are indicated with a same color code.

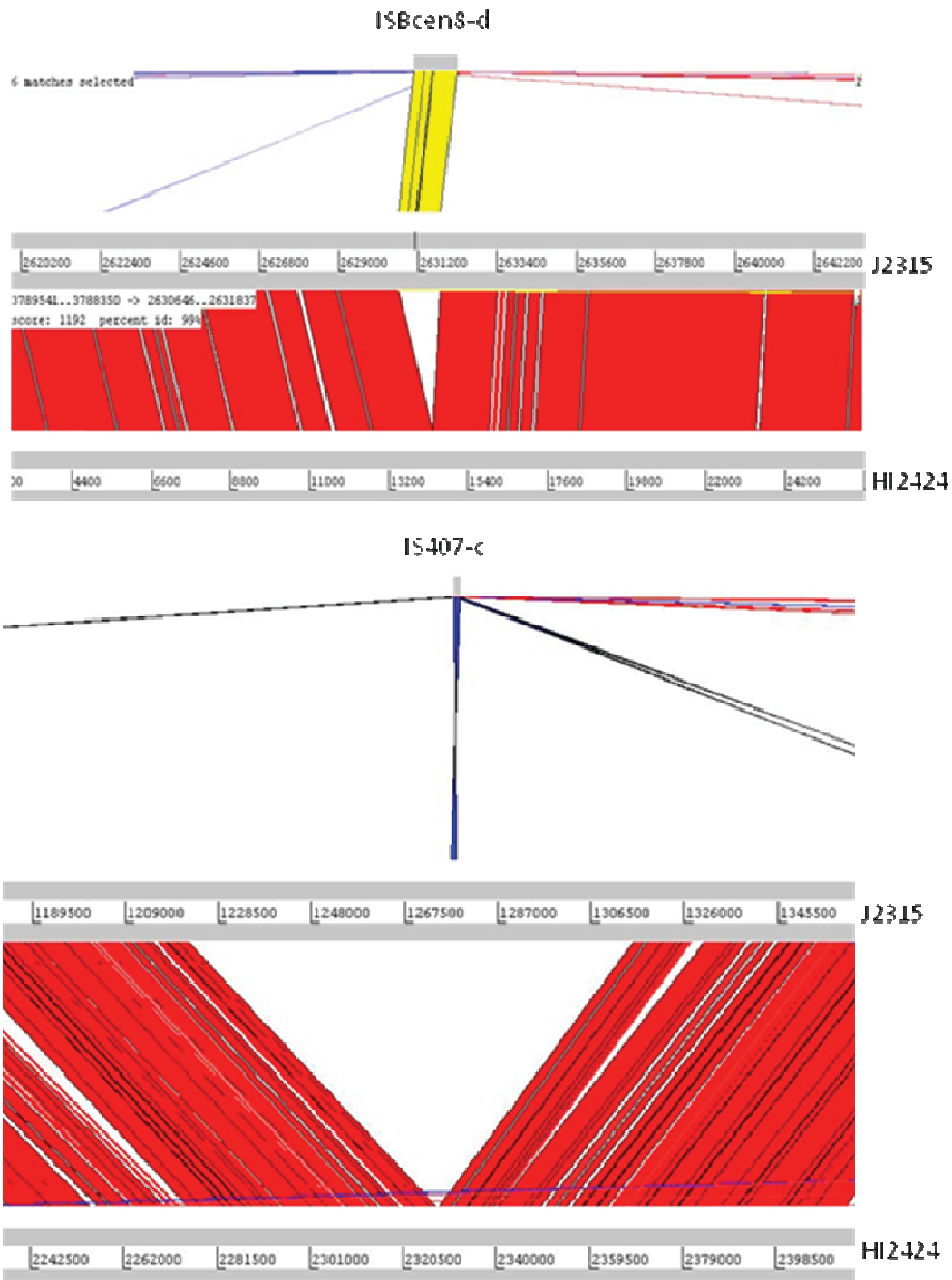
However, two of these loci (the IS407-a and IS407-1 DNA loci, see Table 6) were found colonized by a different DNA insert, not related to an IS element (Table 7). These inserts led to PCR products of other sizes than the ones expected (Table 6, annex C2). All other loci of LMG12614 were found “empty”. The IS407 genomic organization of strain LMG12614 could thus be indicative of the state of ET12 strains prior to an IS407 burst as the ones inferred for J2315-like genomes. This would suggest the IS element of the IS407-n locus to be the founder element of this active group. Interestingly, this latter element was found to be plasmid-borne. Acquisition of this plasmid could thus explain the IS407 amplification events observed among these strains. This plasmid has a size of 92 661 bp. It has a G+C content of 62.6%, and encodes 99 CDSs (see Holden et al., 2009). Looking at the IS407-a loci of strain LMG12614, which was not found colonized by an IS, a DNA segment encoding part of the C subunit of an excinuclease described in *Verminephrobacter eiseniae* EF01-2 was detected (Table 7, annex C2). This segment is inserted at the position 450373 of the J2315 genome, corresponding to the DR sequence of the IS407-a insertion site. Surprisingly, in the same locus but in the genome of strains H111, LMG19240, and LMG21461, a similar situation was observed. The locus of each strain was colonized, at the same site, by an insert matching the same DNA segment encoding a part of the subunit C of an excinuclease. So far, the reason for having a conservation of this insert among such distantly related strains is not clear. If this is the ancestral state, the IS407 insertion at this locus would have led to a deletion of this DNA segment. One argument in favor of this hypothesis is the observation of apparently no strain without any insert (empty site) at this locus. It is noteworthy that the matching locus in the *V. eiseniae* genome shows the presence of a partial IS element, IS *Ve*2 of the IS481 family, upstream this excinuclease gene, which is indicative of genomic instability. The situation of the IS407-1 loci, also shown among strain LMG12614 to have an insert not matching an IS, was found quite different. This locus was found deleted in most of the strains tested except those of the ET12 and B&B clones. In strain LMG12614, this locus was colonized by an insert encoding a putative protein showing DNA sequence similarities with a phage tail protein (Table 7, annex C2). All other strains showed an IS407 element at this locus. Interestingly, these data about the distribution of the IS407 colonized DNA loci confirmed again the close proximity of the ET12



**Fig. 6: Neighbor-Joining phylogenetic tree of IS407 transposase DNA sequences.**

DNA sequences were retrieved from sequenced *Burkholderia* genomes, or obtained by sequencing PCR products or through the GenBank database. A total of 1122 sites were analyzed (with removal of gap-containing sites by pair) using Kimura's 2-parameter model. Distances are proportional to evolutionary divergences expressed in substitutions per 100 sites. Bootstrap values higher than 85% are given. *Bcen* = *B. cenocepacia*; *Bvi* = *B. vietnamiensis*; *Bmu* = *B. multivorans*; *Bp* = *B. pseudomallei*; *Bcep* = *B. cepacia*;

(annex C1). In the other two cases, signs of convergence towards an inactivation by the IS407 of the loci were observed. For the IS407-k locus of bpoe717, the IS407 was found in the opposite direction and nucleotide deletions inactivating orf-a and orf-b of the transposase were observed (Fig. 2). This IS element disrupted genes involved in capsule synthesis (Table 5). In the case of the other locus, IS407-m, genomic rearrangements significantly modified the IS407 element; orf-a and orf-b were found to be partial, and located upstream its left inverted repeat (IR-L) (Fig. 2).



**Fig. 7 : Example of tripartite comparisons of *B. cenocepacia* conserved DNA regions showing presence or absence of an IS element, using the ACT graphic tool.**

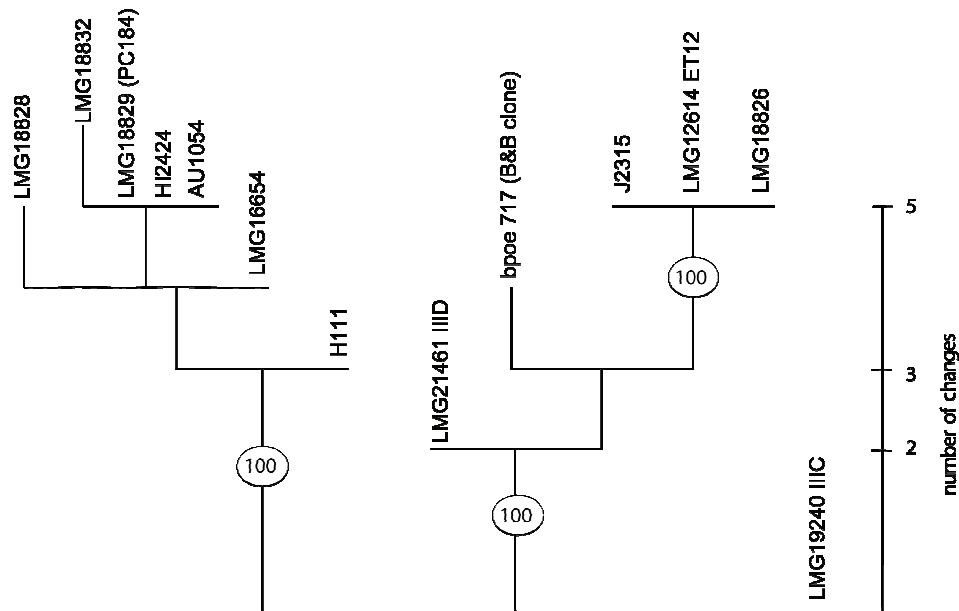
*B. cenocepacia* J2315 genome sequence was compared to IS407 or ISBcen8, and then the J2315 DNA region showing an IS element was compared with the one of *B. cenocepacia* strain HI2424. First panel: a simple insertion event of ISBcen8 into the J2315 genome showing a lack of significant identities over a segment of the size of the IS in the matching DNA region of HI2424 genome. This analysis shows a conservation of the synteny between both genomes for the targeted DNA region. Second panel: large scale 90 kb DNA insertion detected in the J2315 genome and showing an IS407-c element. The lack of synteny with the HI2424 genome indicates an absence of this 90 kb region in HI2424.



**Table 6: PCR screenings and DNA sequences analysis of J2315 IS407 insertion sites among a collection of *B. cenocepacia* strains.**

IS407 copies	Primers	IIIA				IIIB				IIIC	IIID	PCR screenings					
		ET12		B&B clone		LMG		AU1054#				LMG		expected PCR products (pb)	IS empty		
IS407-a	407aF/407aR	J2315	LMG 12614	full	18826	full	H111	18828	16654	III2424#	AU1054#	18829#	18832	LMG 19240	LMG 21461	IS full	700
IS407-b	SI11/SI10	full	insert*	full	full	empty	insert*	-	-	-	-	-	insert*	-	insert*	1400	1400
IS407-c	407cF/407cR	full	empty	full*	full*	empty	empty	empty	empty	-	-	-	-	-	empty*	1802	600
IS407-d	SI8/SI17	full	empty	full	full	empty*	-	-	-	-	-	-	-	empty*	empty*	1439	383
IS407-e	407eF/407eR	full	empty*	full*	full*	-	-	-	-	-	-	-	-	-	empty*	1411	385
IS407-f	407fF/407fR	full	empty	full	full	empty*	empty	empty	empty	empty	empty	empty	-	-	empty*	1715	475
IS407-g	SI21/SI22	full	empty*	full*	full*	-	-	-	-	-	-	-	-	-	-	1557	321
IS407-h	SI23/SI24	full	full	full	full	full*	-	-	-	-	-	-	-	-	full	1430	193
IS407-i	407iF/407iR	full	full*	full	full	full*	full	-	full*	-	-	full	full*	full*	full*	1000	470
IS407-j	407jF/407jR	full	empty*	full*	full*	empty	empty	-	empty*	-	-	-	empty*	empty*	empty*	1555	380
IS407-k	SI5/SI7	full	empty*	full	full	full*	-	-	-	-	-	-	-	-	empty*	1470	444
IS407-l	SI27/SI20	full	insert*	full	full	full*	-	-	-	-	-	-	-	-	-	1408	382
IS407-m	SI3/SI2	full	empty	full	full	full*	-	-	-	-	-	-	full	full	insert*	1437	201
IS407-n	SI14/SI15	full	full*	full*	full*	full*	-	-	-	-	-	-	full*	full*	full*	1612	376

SI and 407 primers were designed to amplify insertion sites. These sites can be absent (no amplification, -) or present and containing a copy of IS407 (large PCR fragment, full) or present but not containing an IS (small PCR fragment, empty) or present and containing sequence different from IS407 (insert), for more details see Table X. \*: represent PCR fragments which have been sequenced and in brackets are given % identity with J2315 matching region #: represent strains which have been analyzed using sequenced genomes and AC1 software.



**Fig. 8: Cladogram illustrating the conservation of *B. cenocepacia* DNA loci colonized by the *IS407* among the *B. cenocepacia* J2315 genome.**

14 loci of J2315 were analysed. The cladogram is unrooted and shows gains, losses or conservation of IS-colonized DNA loci. 27 MP (maximum parsimony) trees involving 19 steps were obtained. Ovals indicate the % of occurrence of the main clusters among all MP trees. These analyses were performed using the “mix” program of the PHYLIP package.



**Fig. 9: Cladogram illustrating the relation between the number of *IS407* and the strains of *B. cenocepacia*.**

14 loci of J2315 were analysed. The cladogram is unrooted and shows gains, losses or conservation of *IS407* elements. 100 MP (maximum parsimony) trees involving 15 steps were obtained. Ovals indicate the % of occurrence of the main clusters among all MP trees. These analyses were performed using the “mix” program of the PHYLIP package.

**Table 7: Characterization of IS407 insertion sites containing no IS element.**

IS407 site Strain	PCR product size Sequenced size accession number	% identity with flanking regions of insertion site	inserted DNA region	
			zone of no homology with the database	similarity with Genbank access, BlastX bests hits
<b>IS407a</b>				
LMG12614	850bp 777bp	124bp, 100% id (1-107 and 760-777)	108 to 402bp	excinuclease ABC subunit C, <i>Verminephrobacter eiseniae</i> EF01-2, Bases 403 to 756, Id: 107/118 (90%), Pos= 114/118 (96%), Gaps = 0/118
H111	850bp 806bp	152bp, 98% id (1-125 and 775-806)	126 to 420bp	excinuclease ABC subunit C, <i>Verminephrobacter eiseniae</i> EF01-2, Bases 421 to 774, Id: 107/118 (90%), Pos= 114/118 (96%), Gaps = 0/118
LMG19240	850bp 795bp	142bp, 100% id (1-117 and 770-795)	118 to 412bp	excinuclease ABC subunit C, <i>Verminephrobacter eiseniae</i> EF01-2, Bases 413 to 766, Id: 107/118 (90%), Pos= 114/118 (96%), Gaps = 0/118
LMG21461	850bp 791bp	138bp, 100% id (1-116 and 770-791)	117 to 411bp	excinuclease ABC subunit C, <i>Verminephrobacter eiseniae</i> EF01-2, Bases 412 to 765, Id: 107/118 (90%), Pos= 114/118 (96%), Gaps = 0/118
<b>IS407I</b>				
LMG12614	1800bp 1696bp	69bp, 100% id (1-30 and 1558-1696)	31 to 927bp 1504 to 1557	phage tail protein I, <i>Burkholderia pseudomallei</i> 112, Bases 928 to 1503, Id: 100/194 (51%), Pos= 131/194 (67%), Gaps = 11/194 (5%)
<b>IS407m</b>				
LMG21461	550bp 452bp	201bp, 100% id (1-155 and 407-452)	156 to 406bp	No hit founded using BlastN and BlastX

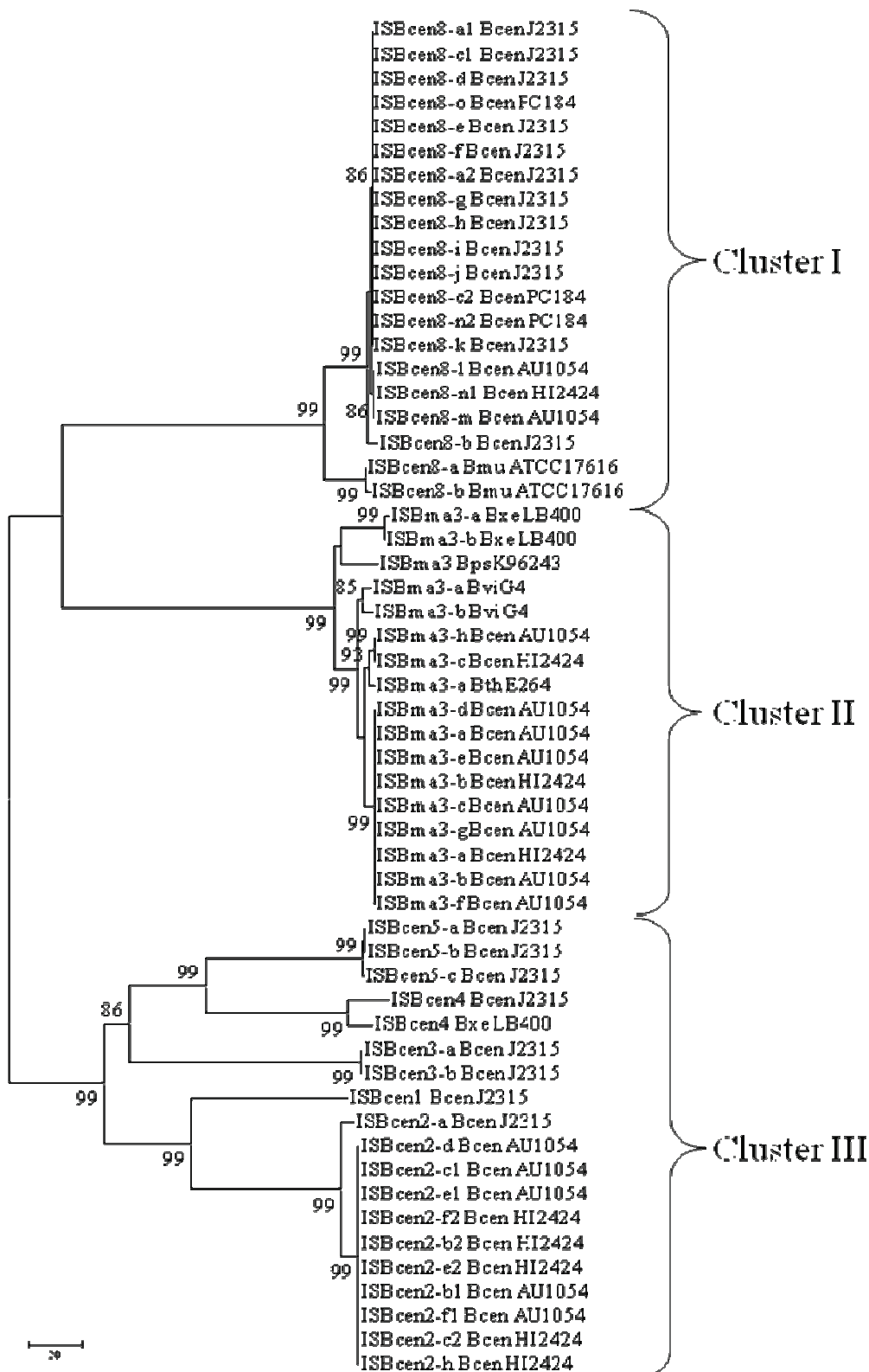
BlastX analyses were done using the NCBI web site. Characterization of left and right borders has been realized using a local BlastN alignment between sequences obtained by PCR and the J2315 genome. Visualization and % identity have been obtained using the ACT software. Sequences obtained are presented in the annex C2.

### **3.2.2. IS110**

The transposases of the IS110 family were found quite diverse, with three main clusters, of which two were highly conserved clusters; the cluster I being made of *ISBcen8* and cluster II of the *ISBma3* one. Cluster III grouped transposases that could be sub-divided into five significant sub-groups, with *ISBcen4* and *ISBcen5* being part of a same sub-lineage grouping with the *ISBcen3* sub-group, and with *ISBcen1* and *ISBcen2* being apart from this sub-group but found to form a significant sub-lineage (Fig. 10 and Table 3). DNA targets of these IS110 elements could only be inferred from the direct repeats observed for the *ISBcen1* and *ISBcen5* sub-groups (Table 4). Relationships within these sub-groups could only be partly resolved using the AA data set of the IS-encoded transposases but could be improved using the DNA data set of the CDS. This allowed identification of highly active elements which appeared to have undergone recent amplification events (Fig. 11). The genetic and genomic organizations of these IS are presented in Fig. 5 and Fig. 12.

#### *- cluster I*

Transposases of cluster I were divided into a main *ISBcen8* lineage grouping sequences from *B. cenocepacia* strains J2315 (n=12), the PHDC strains (n=2 for AU1054, and n=1 for HI2424), the Midwest strain PC184 (n=3) and the *B. multivorans* strain ATCC17616 (n=4). However, DNA sequence analyses of bpoe717 (B&B clone) PCR products of *ISBcen8* revealed two lineages distant from the ones of the main lineage (data not shown, but sequences available in the annexe C1). These are strong indications of DNA acquisitions likely specific of the B&B clone. All these *ISBcen8* elements suggest the presence of at least three distinct founder elements among this sub-group (named A, B, C in Fig. 11). Elements of the “A” lineage appear to have emerged recently, few nucleotide substitutions being observed between pair of sequences. Five sub-lineages of *B. cenocepacia* *ISBcen8* could be inferred from the NJ tree derived from the DNA data set (Fig. 11): one grouping the *ISBcen8*-j and -k of strain J2315, one being made of a single element i. e. the *ISBcen8*-b of J2315, one grouping all other copies of *ISBcen8* of strain J2315, one grouping all *ISBcen8* sequences of the PHDC strains and one grouping all copies of the Midwest lineage. This suggests that three waves of amplification of *ISBcen8* have likely occurred in strain J2315. *ISBcen8*-b of J2315 was found the most divergent element (Fig. 10; Fig. 11). DNA sequence analysis of this particular element showed a substitution leading to a stop codon (amber mutation) and a small variable region of 7 nucleotides with a nucleotide insertion leading to an inactivation of the transposase CDS (Fig. 12). Genetic divergence was also observed between the J2315 elements, the PHDC ones, and those of strain PC184, suggesting an acquisition of *ISBcen8* prior their diversification. *ISBcen8*

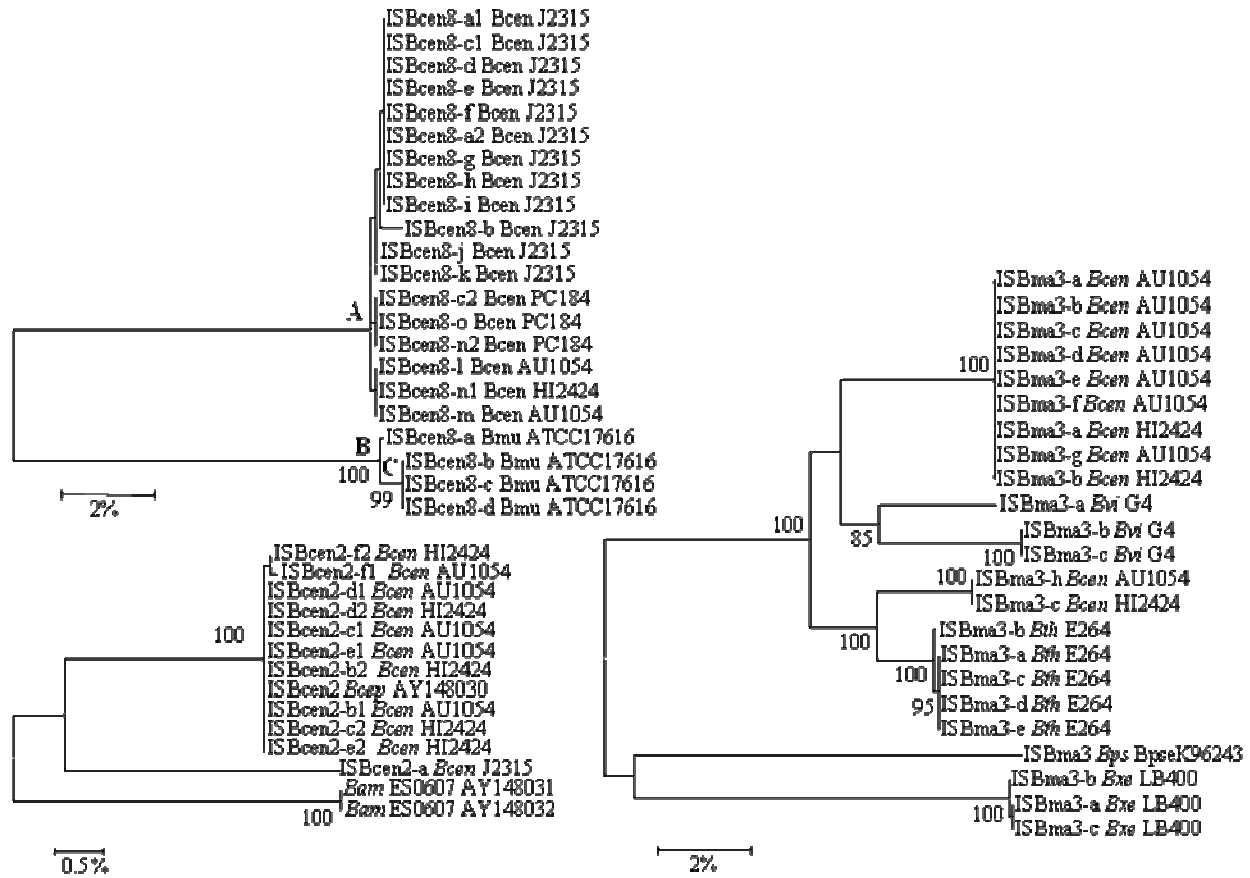


**Fig. 10: Neighbor-Joining phylogenetic tree of transposases of IS110 family.**

The amino acids sequences were retrieved from sequenced *Burkholderia* genomes or the GenBank database. A total of 469 sites were analyzed (with pairwise deletion of gap-containing sites). Distances are proportional to evolutionary divergences expressed in number of difference. Bootstrap values higher than 85% are given. The three main clusters of IS110 are indicated. Bcen = *B. cenocepacia*; Bth = *B. thailandensis*; Bvi = *B. vietnamiensis*; Bps = *B. pseudomallei*; Bmu = *B. multivorans*; Bxe = *B. xenovorans*.

elements could be detected among *B. multivorans*, and other genera like *Acidovorax* (Fig. 10; Fig. 11). DNA blot analyses showed the presence of more than 7 elements of *ISBcen8* among the B&B bpoe717 strain (Table 3).

In the J2315 genome, the *ISBcen8*-a element was found to have been duplicated, as a part of a 40 kb region, leading to two copies, *ISBcen8*-a1 and *ISBcen8*-a2. This duplication was found involved in a large scale inversion of a segment of J2315 chromosome 1, as first reported by Menard et al., (2007). Interestingly, flanking each 40 kb duplicated regions, *ISBcen8* elements were detected i. e. *ISBcen8*-b (an inactive IS, see above) and *ISBcen8*-f (Fig. 5). The *ISBcen8*-a1 / *ISBcen8*-a2 elements were found inactivating a *uvrC* gene (Table 8). No other copy of *uvrC* could be detected in this genome. In *B. cenocepacia* AU1054, HI2424 and PC184, *uvrC* is intact. *uvrC* encodes an endonuclease. It is the third component of the *uvrABC* DNA repair system. *E. coli uvrC* mutants are more sensitive to UV light than wild type strains. In *E. coli*, a second endonuclease, *cho* was found to be able to partially complement *uvrC* mutations (Moolenaar et al., 2002). In the J2315 genome, no *cho* ortholog could be found. Thus, UV sensitivity should be different among *B. cenocepacia* strains, with the ET12 strains like J2315 likely to be more sensitive. The localization of other *ISBcen8* elements in the *B. cenocepacia* genomes was also investigated. Eight *ISBcen8* elements were found outside CDS, and four inside CDS among strain J2315 genome (Table 8). One element, *ISBcen8*-c1, was found positioned in the same DNA site among the PC184 genome (Table 8). This was the only element having a position conserved between J2315 and other sequenced *B. cenocepacia* genomes. This investigation was completed by using a tripartite ACT approach developed in the *IS407* sub-section (above), to identify DNA sites colonized by an *ISBcen8* element in J2315 that are conserved among the other sequenced *B. cenocepacia* genomes (Fig. 7). Eight out of the 12 *ISBcen8* colonized DNA regions of strain J2315 were found, by this approach, in the genomes of strain HI2424 (*ISBcen8*-a1, -b, -c1, -d, -f, -g, -h and -k DNA regions) and 7 out of the 12 *ISBcen8* within the PC184 genome (*ISBcen8*-a1, -b, -c1, -d, -f, -h and -k DNA regions). Thus, contrary to the *IS407* situation, 58 to 66% of the *ISBcen8* insertions of the J2315 genome occurred in chromosomal regions which are respectively conserved in the PHDC and Midwest genomes. However, IS insertions were not detected in the matching DNA regions with one exception for *ISBcen8*-c1 among PC184 strain. It is noteworthy that one element of the *ISBcen8* sub-group of strain HI2424 was found in a conserved location among the AU1054 sequenced strain (Table 8).

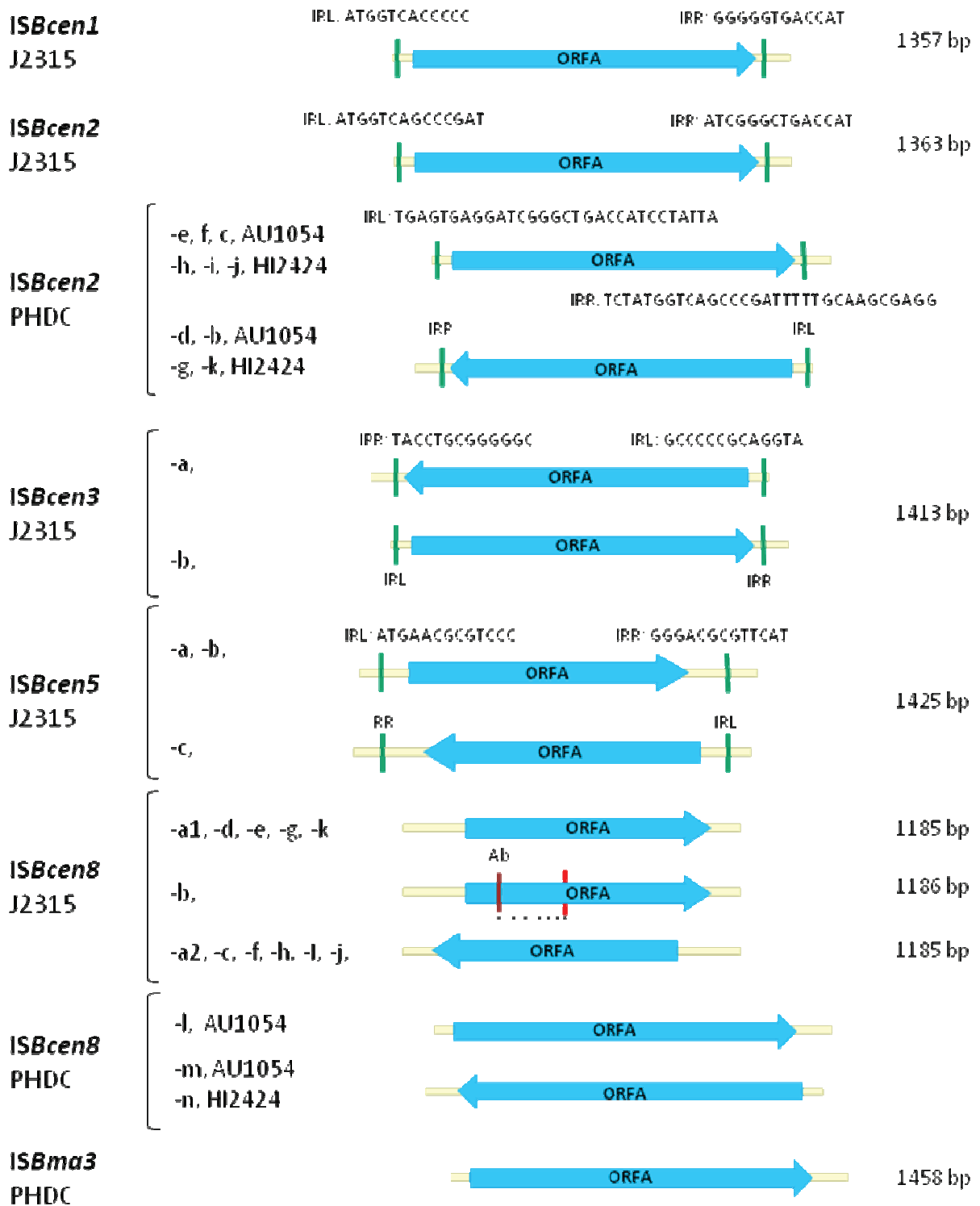


**Fig. 11: Neighbor-Joining phylogenetic trees of ISBcen8, ISBma3, and ISBcen2 DNA sequences from *B. cenocepacia* and closely related species.**

DNA sequences were retrieved from Genbank and *Burkholderia* sequenced genomes. Alignments were performed using CLUSTALW and NJ trees were inferred using Mega3 graphic tool (gap containing sites were deleted by pairs). Distances are proportional to evolutionary divergences expressed in substitutions per 100 sites. Bootstrap values higher than 85% are given. *Bcen* = *B. cenocepacia*; *Bps* = *B. pseudomallei*; *Bam* = *B. ambifaria*; *Bth* = *B. thailandensis*; *Bmu* = *B. multivorans*; *Bcep* = *B. cepacia*; *Bvi* = *B. vietnamiensis*.

Regarding gene inactivations by *ISBcen8*, the *ISBcen8-k* element of strain J2315 was found inserted in a gene encoding a component of an ABC transporter. This inactivation could thus have modified uptake and export of nutrients in this strain. *ISBcen8-g* was found to have inactivated a gene involved in drug transport (Table 8). However, *B. cenocepacia* strain J2315 is known to be highly resistant to a large panel of antibiotics. It is thus probably not related to antibiotic resistance mechanisms. Other *ISBcen8* elements of strain J2315 were found located next to CDS encoding hypothetical proteins, regulators or metabolic enzymes. No evidence for a contribution of these insertions in any modification of the virulence properties could be inferred. The *ISBcen8-e* element of the J2315 genome was found located on the left border of a large scale insertion of 120 kb. This DNA region was found sharing identities with phage related DNA. *ISBcen8-j* was found located on a 15 kb insertion. These two DNA insertions were not detected in the HI2424, AU1054 and PC184 genomes.





**Fig. 12: Main features of IS110 elements found among *B. cenocepacia* strains.**

Inverted DNA repeats are indicated by a green vertical line; red vertical lines are used to indicate a frameshift; nonsense mutations (amber type) are indicated by a brown vertical line (Ab). \* correspond to the seven nucleotides of the *ISBcen8-b* which were different compared with other *ISBcen8* elements of the J2315 genome. ORFs are indicated by arrows and ORFs with a deletion are represented by arrows without a vertical extremity. ORFA: total transposase sequence. ORF orientation of IS elements are according to the main orientation of the genomes, beginning from the origin of replication. PHDC indicates the HI2424 and AU1054 genomes, and Midwest the one of strain PC184. The two *ISBcen4* copies are described within Fig. 4.

**Table 8: Characterization of IS*Bcen8* flanking ORFs.**

Strain	ORF upstream	IS <i>Bcen8</i> ORF	ORF downstream
J2315			
IS <i>Bcen8</i> -a1	BCAL1014, UvrABC system protein C	BCAL1017	BCAL1014, UvrABC system protein C 2
IS <i>Bcen8</i> -b	BCAL1026, putative GAF sensor protein	BCAL1028	BCAL1029, Transcriptional regulator, LysR family
IS <i>Bcen8</i> -c1	BCAL1488, MerR family regulatory protein	BCAL1489	BCAL1490, putative exported protein
IS <i>Bcen8</i> -d	BCAL2376, quaternary ammonium compound-resistance protein SugE	BCAL2377a	BCAL2378, putative exported protein
IS <i>Bcen8</i> -e	BCAL2591, conserved hypothetical protein	BCAL2595	BCAL2596, conserved hypothetical protein
IS <i>Bcen8</i> -f	BCAL2844, putative branched-chain amino acid aminotransferase IlvE	BCAL2845	BCAL2846, putative GAF sensor protein
IS <i>Bcen8</i> -a2	BCAL2854, UvrABC system protein C	BCAL2855	BCAL2854, UvrABC system protein C
IS <i>Bcen8</i> -g	BCAM1251, Drug resistance transporter EmrB/QacA family,	BCAM1252	BCAM1251, Drug resistance transporter EmrB/QacA family
IS <i>Bcen8</i> -h	BCAM2605, putative 2,5-diketo-d-gluconic acid reductase	BCAM2606	BCAM2607, putative NADH:flavin oxidoreductase/NADH oxidase
IS <i>Bcen8</i> -i	BCAS0003, ParA protein	BCAS0004	BCAS0005, AraC family regulatory protein
IS <i>Bcen8</i> -j	BCAS0067, AraC family regulatory protein	BCAS0068	BCAS0069, two-component regulatory system, response regulator protein
IS <i>Bcen8</i> -k	BCAS0421, ABC transporter ATP-binding protein	BCAS0422a	BCAS0421, ABC transporter ATP-binding protein
AU1054			
IS <i>Bcen8</i> -l	Bcen2207, Phosphotransferase system, N-acetylglucosamine-specific IIBC component, YP_622081	Bcen_2208	Bcen2209, Phosphoenolpyruvate-protein phosphotransferase, YP_622083
IS <i>Bcen8</i> -m	Bcen_1816, transcriptional regulator, XRE family, YP_621692	Bcen_1817	Bcen_1818, argininosuccinate lyase, YP_621694
HI2424			
IS <i>Bcen8</i> -n1	Bcen2424_3473, CsbD family protein, YP_837105	Bcen2424_3472	Bcen2424_3471, LuxR family transcriptional regulator, YP_837103
PC184			
IS <i>Bcen8</i> -o	BCPG_00538, hypothetical protein, YP_002091830	BCPG_00539	BCPG_00540, hypothetical protein, YP_002091832
IS <i>Bcen8</i> -c2	BCPG_00313, regulatory protein, MerR, YP_002091614	BCPG_00312	BCPG_00311, hypothetical protein, YP_002091612
IS <i>Bcen8</i> -n2	BCPG_04002, hypothetical protein, YP_002095152	BCPG_04003	BCPG_04004, PAS/PAC domain, YP_002095154

Flanking ORFs were characterised by BLASTP against Genbank and described by Holden et al., 2009.

- cluster II

Transposases of cluster II were divided into a main IS*Bma3* lineage grouping sequences from *B. cenocepacia* PHDC strains (n=8 for AU1054 and n=3 for HI2424), *B. multivorans* (n=2), *B. vietnamiensis* (n=3), *B. pseudomallei* (n=1), *B. xenovorans* (n=3), and *B. thailandensis* (n=5). Among PHDC genomes (HI2424 and AU1054), IS*Bma3* was the most abundant IS element. About 66% of the *Burkholderia* genomes showed at least one copy of the IS*Bma3* element. However, no copy could be detected among *B. cenocepacia* J2315 and PC184 genomes (Table 3). Similarities among AA sequences translated from the *C. solanacearum* genome of strain IPO1609 (90% id), of *Escherichia coli* strain CFT073, and of *Shigella dysenteriae* strain 1012 (80% id) were detected. Amino acid

sequences analyses suggested that at least three significant lineages led to the elements observed (Fig. 10), with some divided into sub-lineages matching their species of origin. Phylogenetic analysis of the DNA data set confirmed these divisions, and improved the overall significance of the inferred phylogenetic tree (Fig. 11). Most *ISBma3* elements of the PHDC genomes (7/8 for AU1054 and 2/3 for HI2424) have identical transposase DNA sequences, suggesting a common recent origin or founder element. The other PHDC copies, *ISBma3-c* for HI2424 and *ISBma3-h* for AU1054, are identical but belong to a distinct lineage showing 92% id with the other lineage. These two elements group with *B. thailandensis* *ISBma3* elements (Fig. 11). They suggest the presence of a second group of active *ISBma3* elements among the PHDC genomes. Another sub-lineage was found restricted to *B. vietnamiensis*, with two elements, *ISBma3-b* and *-c*, being identical, and a third one, *ISBma3-a*, harboring 96% id with the other sequences. The last two sub-lineages are composed of three *ISBma3*-like sequences from *B. xenovorans* and one from *B. pseudomallei*. These elements appear to have diverged early from the other sub-lineages. The large distribution of *ISBma3* suggests a former colonization by this IS element early in the evolution of the *Burkholderia* species.

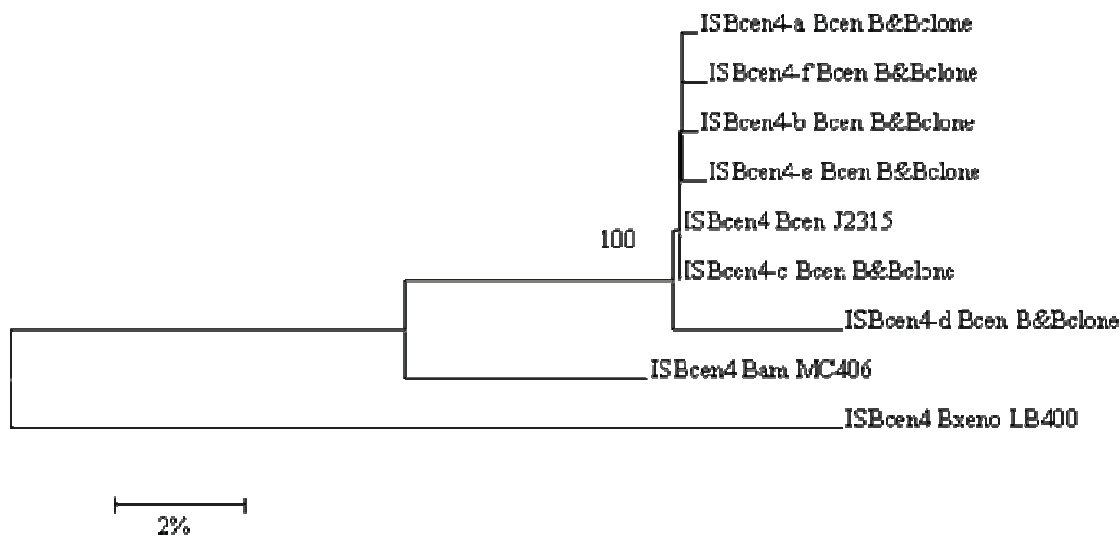
The localization of *ISBma3* elements among *B. cenocepacia* PHDC genomes (AU1054 and HI2424) was investigated. All IS elements were found outside CDS, except *ISBma3-f* of AU1054 which was found inserted in a gene encoding a glutamine amidotransferase [Bcen\_4743]. Enzymes of this family can remove the ammonia group from a glutamate molecule and transfer it to a specific substrate, apparently tryptophan, in this particular case. The selective forces that could have led to this inactivation are not known. However, strain AU1054 was isolated from blood, and strain HI2424 (not disrupted in this gene) from soil. Tryptophan being available in blood, strain AU1054 may not need this enzyme for growth. Comparison of the DNA regions colonized by *ISBma3* among PHDC strains was performed using a tripartite ACT approach. All DNA regions colonized by *ISBma3* in AU1054 genome were detected in HI2424 genome. Moreover, three of these regions were found colonized by elements localized at the same site in both genomes (AU1054 *ISBma3-e* [Bcen\_0716]; *ISBma3-g* [Bcen\_6212]; *ISBma3-h* [Bcen\_4714] = HI2424 *ISBma3-a* [Bcen2424\_1195]; *ISBma3-b* [Bcen2424\_1867]; *ISBma3-c* [Bcen2424\_3649] elements). ACT comparisons were also performed between AU1054 and J2315 genomes. Most DNA regions (7/8) were found in both genomes despite the absence of *ISBma3* in J2315 genome. The exception was the DNA region harboring AU1054 *ISBma3-a* which was not found in the J2315 genome. This could indicate that this IS element was introduced via this particular genomic region which could be a genomic island. Interestingly, the *ISBma3-h* DNA locus of AU1054 showed similarities with J2315 genome over a 5.6 kb DNA sequence encoding an AraC regulator, a hypothetical protein and an isoquinolone oxidoreductase genes. Conservation of this locus between J2315 and PHDC genomes supports the hypothesis that

IS*Bma3*-h of AU1054 (and IS*Bma3*-c from HI2424 strain) could be the first IS*Bma3* element responsible of the global colonization of these strains by these elements.

- cluster III

AA of the transposases of cluster III were divided into an IS*Bcen5* sub-group, an IS*Bcen4* one, IS*Bcen3* one, IS*Bcen1* one, and IS*Bcen2* one (Fig. 10 and Fig. 11), with IS*Bcen3*, IS*Bcen4*, IS*Bcen5* being found among a common phylogenetic group, and IS*Bcen2* and IS*Bcen1* into another one. Among the *Burkholderia*, most IS elements of the cluster III appeared restricted to *B. cenocepacia* with the exception of IS*Bcen4* which was detected in *B. xenovorans* LB400 genome. BLASTP analysis of IS*Bcen1* transposase sequence showed similarities with sequences translated from *Burkholderia* sp. H160 (77% id), *Acidovorax delafieldii* strain 2AN (77% id) and *Herminiimonas arsenicoxydans* (76% id) DNA sequences but most *Burkholderia* genomes did not show significant similarities with this IS (Table 3). Surprisingly, IS*Bcen1* and IS*Bcen14* of the IS66 family (described below) showed similar DNA target according to their DR (Table 4). IS*Bcen2* AA sequences showed similarities with proteins of *Herminiimonas arsenicoxydans* (66% id) and *Azotobacter vinelandii* strain DJ (66% id). IS*Bcen3* AA sequences showed similarities with putative proteins of *Polaromonas naphthalenivorans* strain CJ2 (72% id) and *Verminephrobacter eiseniae* strain EF01-2 (72% id). These species belong to the  $\beta$ -proteobacteria, and our previous analyses of IS407 insertion sites showed likely DNA transfers between *B. cenocepacia* strains and *Verminephrobacter eiseniae*. IS*Bcen4* was detected in *B. cenocepacia* J2315 and B&B genomes (annexe C1), and among *B. xenovorans* LB400 genome. AA sequence analysis allowed detection of a copy of this IS element among the *Burkholderia ambifaria* strain MC40-6 (98% id) but no copy was detected among the *B. ambifaria* AMMD genome. Interestingly, the two IS*Bcen4* copies among J2315 genome were inactivated, with one element being disrupted by IS*Bcen13*-e (Fig. 4c) and the second one, being close to the other one and showing a partial DNA sequence (Fig. 4c). IS*Bcen4* DNA sequences were obtained for the B&B clone and compared with other sequences. These analyses showed one B&B IS*Bcen4* element to group with the J2315 IS*Bcen4* lineage but the second sequence was outside this group (Fig. 13) and apart the lineages of IS*Bcen4* elements from *B. xenovorans* LB400 and *B. ambifaria* MC40-6. These data could indicate an early colonization of the *Burkholderia* by IS*Bcen4* elements but no major wave of amplification of these elements apparently occurred. The B&B clone showed the highest number of these elements (Table 3), indicating that a recent acquisition of some elements might have occurred, as supported by the DNA data set. However, we were unable to determine if the B&B clone IS*Bcen4* elements were disrupted or partial (as observed in J2315, Fig. 4c). IS*Bcen5* was found quite conserved among J2315 genome but Blast analyses allowed to detect two copies among the genome of *B. thailandensis* strain MSMB43. IS*Bcen5* was also found to harbor 91% identity with the IS884 and IS*Bfun2* from,

respectively, the *Ralstonia eutropha* and *B. xenovorans* genomes. This suggests a common origin between these elements.



**Fig. 13: NJ phylogenetic trees of ISBcen4, from *B. cenocepacia* J2315, B&B clone and closely related species.**

DNA sequences were retrieved from Genbank, PCR sequencing and *Burkholderia* sequenced genomes. Alignments were performed using CLUSTALW and NJ trees were inferred using Mega3 graphic tool (gap containing sites were deleted globally). Distances are proportional to evolutionary divergences expressed in substitutions per 100 sites. Bootstrap values higher than 85% are given. *Bcen* = *B. cenocepacia*; *Bam* = *B. ambifaria*; *Bxe* = *B. xenovorans*.

ISBcen2 is the last IS element of this cluster. It was recorded among J2315 (n=1) and the PHDC genomes (n=5 per genome). No copy was detected in the B&B genome. The position of these IS was investigated using the ACT approach. All *B. cenocepacia* ISBcen2 copies of the PHDC strains were found in the same insertion sites but the ISBcen2 copy of J2315 was found in a site not detected in the PHDC strains (Table 9). Phylogenetic analysis of ISBcen2 DNA sequences divided copies according to the strains or clone analyzed (Fig. 11). Most PHDC copies are grouped together suggesting a recent wave of amplification. However, ISBcen2-f1 and ISBcen2-f2 represent a distinct lineage found conserved (orthologous insertion sites) in both strains. The J2315 ISBcen2 element grouped apart from these sequences and was found to be inserted inside a novel CDS showing similarities (98% id AA sequences) with a phage-like Mu integrase (BCAS0681, Table 9). This element could have been acquired recently through phage transduction.

**Table 9 : Characterisation of IS*Bcen2* flanking ORFs**

Strain	ORF upstream	IS <i>Bcen2</i> ORF	ORF downstream
J2315			
IS <i>Bcen2</i> -a	BAS0681, Integrase, catalytic region, ZP_00424558	BCAS0682	BCAS0681, Integrase, catalytic region, ZP_00424558
AU1054			
IS <i>Bcen2</i> -b1	Bcen_5871, polyferredoxin-like, YP_625709	Bcen_5872	Bcen_5873, hypothetical protein, YP_625711
IS <i>Bcen2</i> -c1	Bcen_0282, DedA family transmembrane protein, YP_620169	Bcen_0283	Bcen_0284, MscS mechano-sensitive ion channel YP_620170
IS <i>Bcen2</i> -d1	Bcen_2599, extradiol ring-cleavage dioxygenase, class III enzyme, subunit B, YP_622471	Bcen_2600	Bcen_2601, amino acid permease-associated region YP_622472
IS <i>Bcen2</i> -e1	<i>lpxC</i> , UDP-3-0-acyl N-acetylglucosamine deacetylase, YP_619972	Bcen_0085	Bcen_0086, hypothetical protein, YP_619973
IS <i>Bcen2</i> -f1	Bcen_2351, 2-nitropropane dioxygenase, YP_622225	Bcen_2352	Bcen_2353, transcriptional regulator, LysR family, YP_622226
HI2424			
IS <i>Bcen2</i> -f2	Bcen2424_2965, 2-nitropropane dioxygenase, NPD, YP_836606	Bcen2424_2966	Bcen2424_2967, LysR family transcriptional regulator, YP_836608
IS <i>Bcen2</i> -d2	Bcen2424_0506, extradiol ring-cleavage dioxygenase III subunit B, YP_834153	Bcen2424_0505	Bcen2424_0504, amino acid permease-associated region, YP_834151
IS <i>Bcen2</i> -b2	Bcen2424_2206, polyferredoxin-like protein, YP_835850	Bcen2424_2205	Bcen2424_2204, hypothetical protein, YP_835848
IS <i>Bcen2</i> -e2	<i>lpxC</i> , UDP-3-0-acyl N-acetylglucosamine deacetylase, YP_834213	Bcen2424_0567	Bcen2424_0568, hypothetical protein, YP_834215
IS <i>Bcen2</i> -c2	Bcen2424_0766, DedA family transmembrane protein, YP_834412	Bcen2424_0767	Bcen2424_0768, MscS mechanosensitive ion channel, YP_834414

Flanking ORFs were characterised by BLASTP against Genbank and described by Holden et al., 2009. IS element founded in the same insertion site are linked by a black bar.

### **3.2.3. IS256**

Phylogenetic analysis of the IS256 encoded transposases divided this family into at least five significant clusters (Fig. 14, Table 3). Cluster I grouped highly conserved elements of the IS1356 sub-group. Cluster II was divided into at least four distinct sub-groups, two corresponding to IS*Rso7*-like elements, and two to IS*Bcen18*-like elements. Cluster III was found grouping identical sequences of the IS406 sub-group. Cluster IV grouped transposases of the IS1490 and IS1413 sub-groups. Cluster V grouped transposases of the IS931 sub-group. DNA targets of these IS256 elements could only be inferred from the direct repeats observed for IS1356-a, IS1356-b and IS*Bcen18* of the J2315 genome (Table 4). Relationships within IS1356 and IS*Bcen18* sub-groups could only be partly resolved using the AA data set of the IS-encoded transposases but could be refined using the DNA data set of the transposase CDS. This allowed identification of highly active elements which appeared to have undergone recent amplification events (Fig. 15). The genetic and genomic organizations of some these IS is presented in Fig. 16 and Fig. 5.

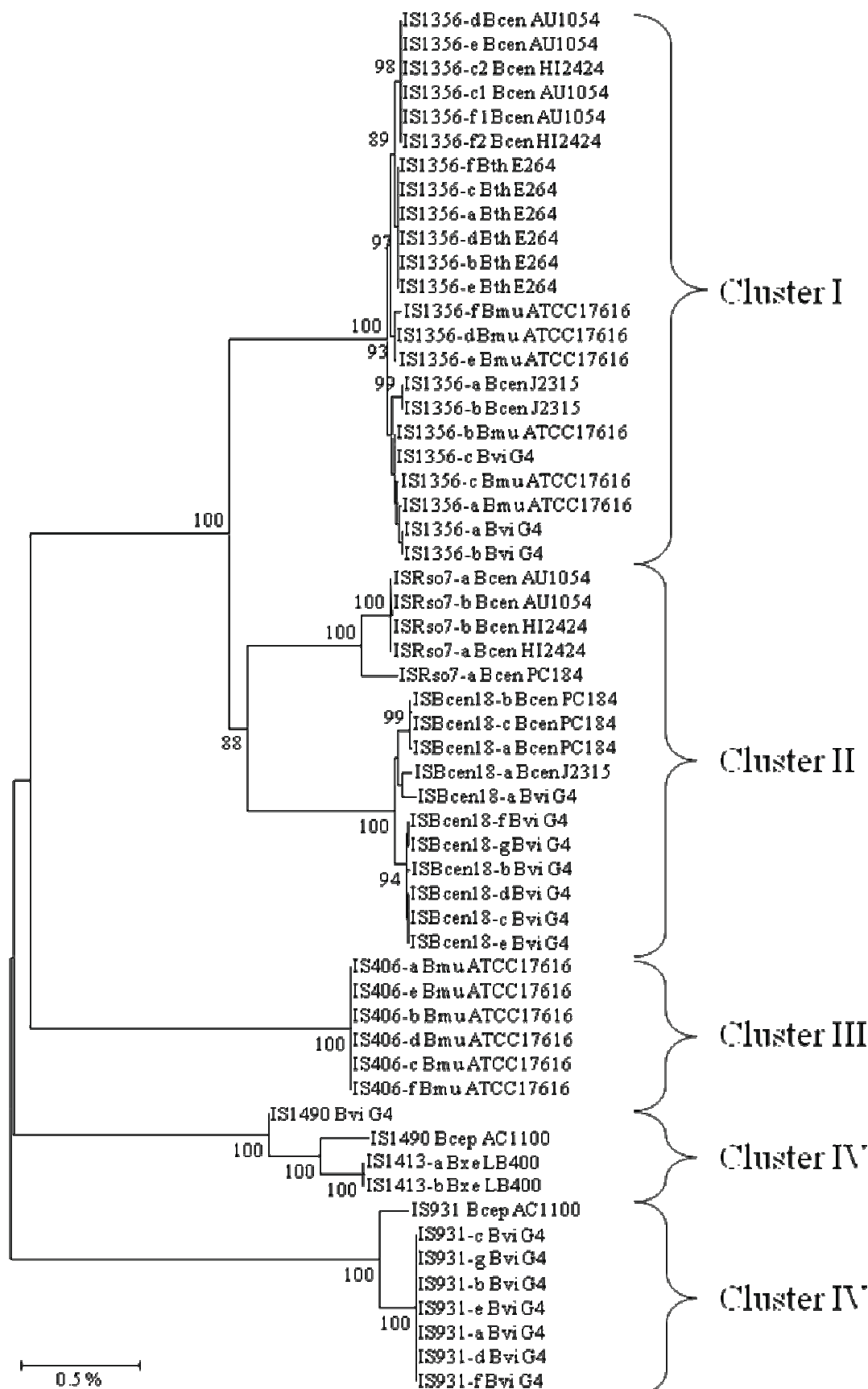
#### *- cluster I*

Phylogenetic analysis of the deduced AA sequences of the transposases of cluster I (IS1356) or their respective DNA sequences gave similar NJ tree topologies (Fig. 14 et 15). At least five sub-groups, supported by high bootstrap values, could be observed. Most were species-specific. In one instance, two sub-groups were found to form a significant cluster not matching the actual systematic of the *Burkholderia*, with the PHDC IS1356 sequences being grouped with those extracted from the *B. thailandensis* genome rather than being grouped with those extracted from the other genomes of the Bcc (Fig. 15). In the other instances, the IS1356 elements showed either identical sequences, even though the sequences were from distinct species of the Bcc, or important divergences within a same genome like the one of *B. multivorans* ATCC17616. This suggests important movement of these IS among the Bcc, and multiple independent acquisitions. However, recent waves of amplification have also apparently occurred as observed for the PHDC genomes (e. g. IS1356-c1, -d, -e, -f1 of AU1054). The position of these IS was investigated within the three *B. cenocepacia* genomes (Table 10; Fig. 5). The two IS1356 copies of the HI2424 genome (IS1356-c2 and -f2) were found inserted in the same sites as the ones of the AU1054 genome (IS1356-c1, -f1). Moreover, IS1356-f1/IS1356-f2 and IS1356-d were found inserted respectively among genes encoding a glutathione S-transferase-like and a Mannose-6-phosphate isomerase. IS1356 was involved in 33% of the gene disruptions by IS recorded among the PHDC genomes (Fig.1). The two IS1356 elements of the J2315 genome formed a distinct group, and were localized in clone-specific insertion sites. IS1356-a disrupted an ORF encoding a Vgr-like protein, and IS1356-b was found inserted within an IS*Bcen20* (Fig 18, Table 10).

- cluster II

Phylogenetic analysis of the deduced AA sequences of the transposases of cluster II led to a separation of the sequences into two sub-groups, supported by high bootstrap values, one grouping sequences of *ISRso7* from *B. cenocepacia* strains AU1054 (n=2), HI2424 (n=2) and PC184 (n=1), and one grouping sequences of *ISBcen18* from *B. cenocepacia* strains J2315 (n=1), PC184 (n=3) and *B. vietnamiensis* G4 (n=7) (Fig. 14). One common element probably led to the emergence of these two sub-groups. *ISRso7* was absent of *B. cenocepacia* strain J2315 genome but homologs were detected in *Ralstonia solanacearum* GMI1000 (91% id), *C. metallidurans* CH34 (92% id), *B. xenovorans* MC40-6 (92% id) and *Pseudomonas aeruginosa* C3719 (87% id) genomes. This suggests a large distribution of this element in  $\beta$  and  $\gamma$ -proteobacteria. Investigation of *ISRso7* insertion sites among PHDC strains showed conserved locations (data not shown). However, the matching PC184 DNA site, colonized by *ISRso7*-a in the PHDC genomes, was not colonized by this IS but instead by the *ISBcen18*-a, at exactly the same site. In fact, the PC184 *ISRso7*-a copy was instead localized among a 60 kb DNA region absent from PHDC and J2315 genomes. This region was found to have a 60% of GC and to harbor three other IS elements (*ISBcen28*, *ISBcen21* and *ISBmu11*), representing 25% of the IS content of the PC184 genome. This DNA region can thus be considered a hotspot of IS elements. This PC184 60 kb DNA region encodes LysR-like, and AraC-like regulatory proteins, but most translated CDS showed no significant similarity with known sequences. BLASTN analysis detected this 60 kb DNA region among *B. cenocepacia* MCO-3 genome. This region was not detected in J2315. However, the matching genomic region of J2315 harbored a 14 kb DNA sequence which was only detected among the J2315 genome, so far. This J2315 sequence showed 57% of GC content and no IS element but, on its left hand, a partial IS element showing homologies with *ISBcen20* (IS5 family) was detected. This element could be a vestige of an old colonization of this site. *ISBcen20* could have been partially deleted during the integration of the 14 kb region. This particular genomic region seems quite unstable among all genomes investigated. The *ISBcen18* sub-group of transposases of cluster II was divided into two sub-lineages (Fig. 14). These sub-groups were confirmed using the DNA data set (Fig. 15), and were found to match their origin. One sub-group is corresponding to *ISBcen18* sequences from *B. vietnamiensis* G4 genome, and the second one sequences from PC184 genome. BLASTX analysis detected *ISBcen18* homologs among *Burkholderia dolosa* AUO158, *Burkholderia phytofirmans* PsJN, *Burkholderia cenocepacia* MC0-3, *B. pseudomallei* MSHR346, *Burkholderia thailandensis* TXDOH and *Burkholderia oklahomensis* EO147 genomes. *ISBcen18* showed 70% identity with the deduced transposase of *ISRta2* from *Cupriavidus eutropha* JMP134. Two elements seem to have recently been active among the *B. vietnamiensis* genome, one grouping *ISBcen18*-c, -d, -e and -h; and one grouping *ISBcen18*-g and -f. These data show identical IS sequences indicative of recent amplification events. *ISBcen18*-a of *B. vietnamiensis* was the most divergent element, and may result of an ancient DNA acquisition.





**Fig 14. Neighbor-Joining phylogenetic tree of IS256 family transposases.**

The AA sequences were retrieved from sequenced *Burkholderia* genomes or the GenBank database. A total of 459 sites were analyzed (with pairwise deletion of gap-containing sites). Distances are proportional to evolutionary divergences expressed in substitutions per 100 sites. Bootstrap values higher than 85% are given. The two different clusters of IS256 family are indicated. Bcen = *B. cenocepacia*; Bth = *B. thailandensis*; Bmu = *B. multivorans*; Bvi = *B. vietnamiensis*; Bcep = *B. cepacia*; Bxe = *B. xenovorans*;

*ISBcen18* sequences of PC184 strain formed a distinct group among the NJ trees (Fig.14 and Fig. 15). The *ISBcen18-a* element was found inserted in a conserved site between the PHDC and the PC184 genomes (section above). The *ISBcen18-a* sequences of J2315 genome were found inserted within another IS element (*ISBcen17*), which was itself inserted in the *ISBcen16* (Fig. 4c), suggesting an important attraction of these IS element towards transposase DNA sequences.

Concerning cluster III, only transposase AA sequences from *B. multivorans* ATCC17616 were found to be part of this cluster, and to match a single sub-group of IS elements, the one of *IS406*. All sequences of this cluster were found identical, suggesting a recent amplification which led to six elements among this genome (Fig. 14). No homolog was detected in the databases. However, *IS406* was previously demonstrated to increase expression of neighboring genes, even in heterologous hosts (Wood et al., 1990, Wood et al., 1991). Interestingly, a DNA sequence encoding a putative type VI secretion system (Vgr family protein: Bmul\_5730) was identified on the 3' region of *ISBcen18-c* from ATCC17616 strain. This suggests a potential role of this IS element in the regulation of genes involved in virulence.

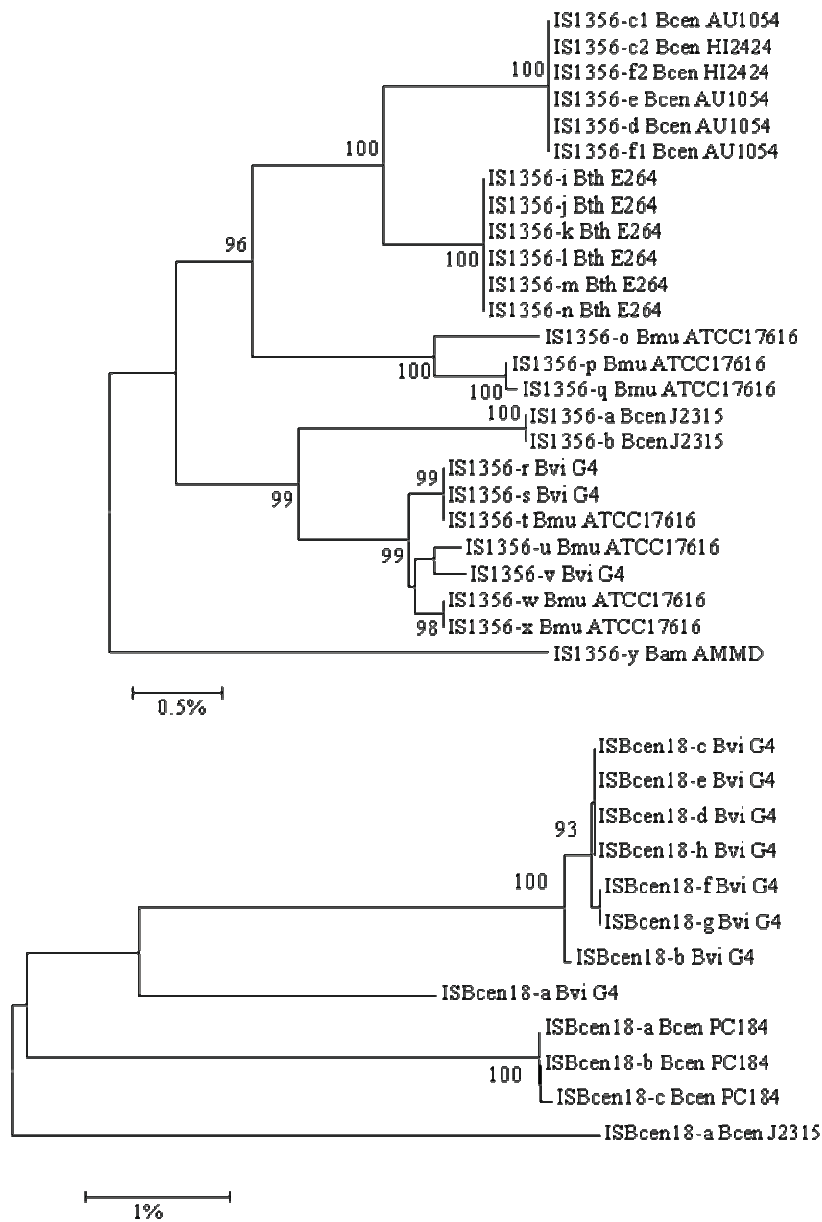
Regarding cluster IV, two sub-groups were detected, *IS1490* and *IS1413*. These elements, detected in both the *B. vietnamiensis* G4 and *B. xenovorans* LB400 genomes, were physically linked to the *IS931* (below). These sub-groups have been shown to have a high-level of promoter activity that can lead to transcription of genes as far as 500 bp downstream from their insertion site (Hubner and Hendrickson 1997).

Phylogenetic analysis of IS elements of cluster V revealed a situation similar to the one of cluster III. All elements were part of the same sub-group, *IS931*. However, this sub-group could be detected among *B. vietnamiensis* G4 and *Burkholderia cepacia* AC1100 genomes. A recent amplification event probably led to the seven copies of this IS observed among *B. vietnamiensis* G4. Interestingly, this IS element was associated with the emergence of a cluster of genes involved in the metabolism 2,4,5-trichlorophenoxyacetic acid among *Burkholderia cepacia* AC1100. Two of these IS elements were identified on both sides of the genes cluster involved in degradation of 2,4,5-trichlorophenoxyacetic acid and seemed to be involved in the integration of it among the AC1100 strain (Haugland et al., 1991). However, using BLASTP analysis non genes of this cluster were identified among the G4 genome suggesting another mechanism involved in the integration of these IS elements.

**Tables 10 : Characterisation of IS1356 flanking ORFs among Bcc genomes**

Strain	ORF upstream	IS1356 ORF	ORF downstream
J2315			
IS1356-a	BCAL2279 Rhs element Vgr protein	BCAL2278	BCAL2279 Rhs element Vgr protein
IS1356-b	BCAL2497 ISBcen20	BCAL2498	BCAL2497 ISBcen20
AU1054			
IS1356-c1	Bcen_5296 Hypothetical protein, YP_625143	Bcen_5297	Bcen_5298 carbon starvation protein CstA, YP_625145
IS1356-d	Bcen_0397 Mannose-1-phosphate guanylyltransferase/mannose-6-phosphate isomerase, YP_620282	Bcen_0398	Bcen_0399 mannose-6-phosphate isomerase, type II YP_620284
IS1356-e	Bcen_1200 MarR family transcriptional regulator, YP_621080	Bcen_1201	Bcen_1202 hypothetical protein, YP_621082
IS1356-f1	Bcen_1203 glutathione S-transferase, YP_621083	Bcen_1204	Bcen_1205 extracellular solute-binding protein, YP_621085
HI2424			
IS1356-c2	Bcen2424_5563 hypothetical protein, YP_839189	Bcen2424_5562	Bcen2424_5561 carbon starvation protein CstA, YP_839187
IS1356-f2	Bcen2424_1682 glutathione S-transferase, YP_835327	Bcen2424_1683	Bcen2424_1684 extracellular solute-binding protein, YP_835329

Flanking ORFs were characterised by BLASTP against Genbank and described by Holden et al., 2009. IS found in identical insertion sites are linked.



**Fig. 15: Neighbor-Joining phylogenetic trees of IS1356 (cluster I) and ISBcen18 (cluster II) DNA sequences encoding transposases of the IS256 family.**

DNA sequences were retrieved from *Burkholderia* sequenced genomes and Genbank. A total of 1351 and 1283 sites were analyzed respectively for IS1356 and ISBcen18 (with global deletion of gap-containing sites). Distances are proportional to evolutionary divergences expressed in substitutions per 100 sites. Bootstrap values higher than 85% are given. *Bcen* = *B. cenocepacia*; *Bam* = *B. ambifaria*; *Bth* = *B. thailandensis*; *Bmu* = *B. multivorans*; *Bvi* = *B. vietnamiensis*.

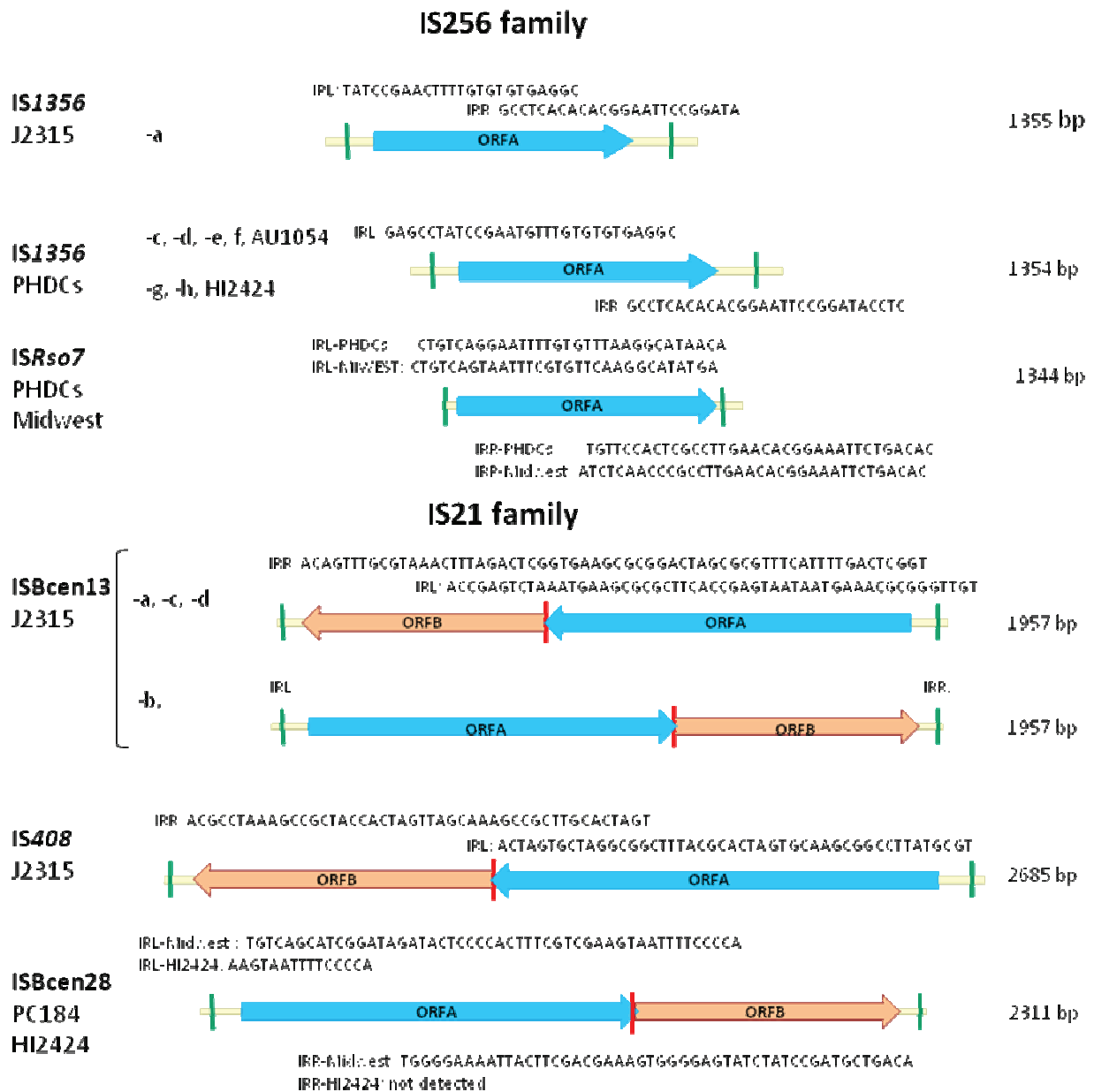
### **3.2.4. IS21**

Molecular phylogeny of IS21family transposases was analyzed. The inferred phylogenetic tree showed four main sub-groups: an IS408 sub-group, an ISBcen13 one, an ISBmu1 one and an ISBcen28 one (Fig. 17). The genetic organization of these IS is presented in Fig. 16, and their localization on Fig. 5. The IS408 sub-group was detected among *B. vietnamiensis* G4 (n=4), *B. multivorans* ATCC17616 (n=3) and *B. cenocepacia* J2315 (n=1) (Fig. 16; Table 3). DNA sequences of this sub-group were identical between all genomes, suggesting a recent acquisition. BLASTN analysis did not detect other copies of this element among the databases. IS408 was originally detected in *Burkholderia cepacia* as an insertion into the *lacZ* gene of Tn951, located on the broad host-range plasmid pGC91.14 (Wood et al., 1990). IS408 could have been involved in the rearrangement pTGL1 which led the pTGL6 (Gaffney and Lessie 1987). Recently, a study demonstrated an increase of IS408 transposition at high-temperature (Ohtsubo et al., 2005), confirming the mobility of this element among *B. multivorans* ATCC17616 strain. No DR was observed for this sub-group (Table 4).

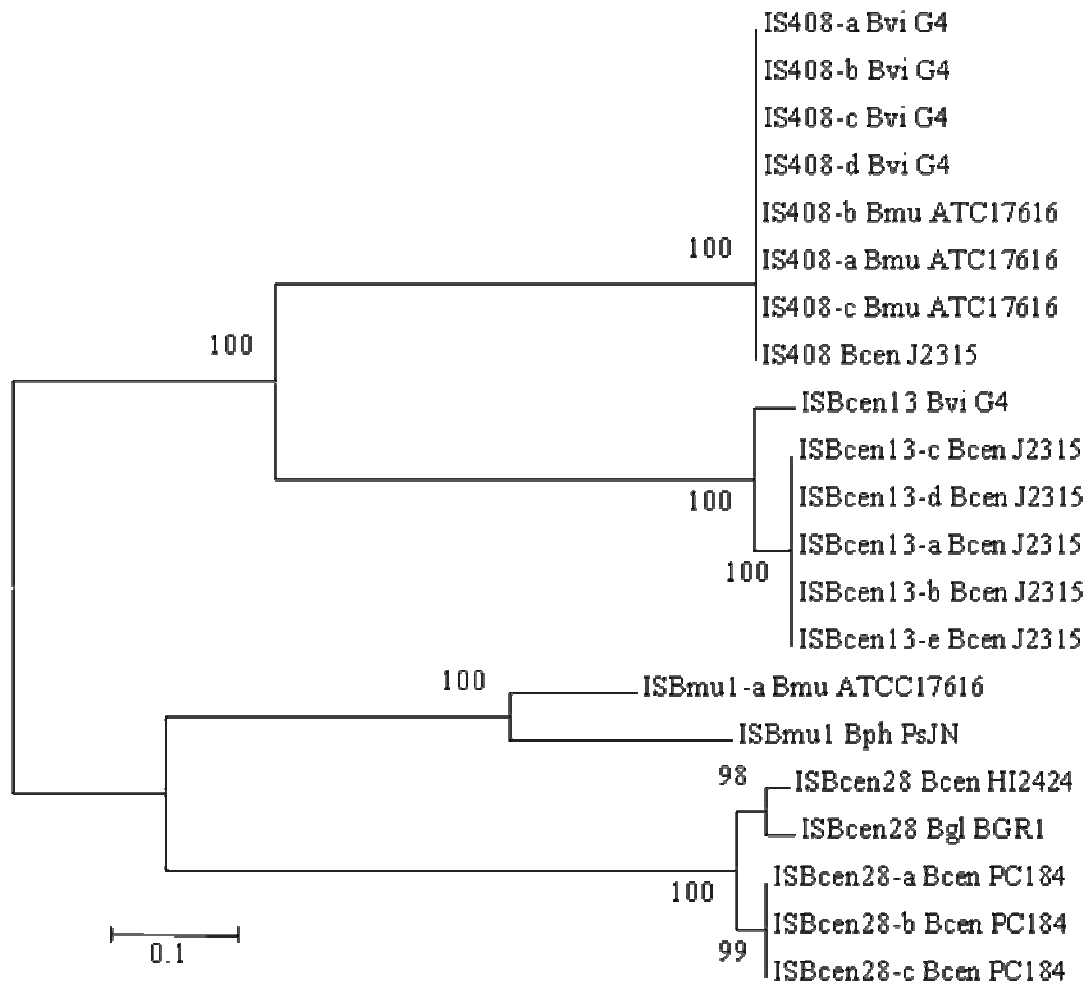
The ISBcen13 sub-group was found among *B. cenocepacia* J2315 (n=5) and *B. vietnamiensis* (n=1) (Table 3). No other copies of these elements have been detected in the databases. Conservation of DNA sequences among J2315 suggests a recent amplification of these elements among this genome. Three of these, ISBcen13-b, -c and -e, have been shown inserted within CDS coding for a protein of unknown function or an IS element (ISBcen13-e) (Fig. 4). DRs of ISBcen13 among J2315 showed variable sequences of 4 nucleotides suggesting no specificity for the insertion sites (Table 4).

The ISBcen28 sub-group was found among *B. cenocepacia* PC184 (n=3), HI2424 (n=1) and *B. glumae* BGR1 (n=1) (Table 3). ISBcen28 elements of the PC184 genome have identical sequences, suggesting a recent amplification of this element among this strain. No other significant homolog was detected in the databases. Analysis of ISBcen28 DR among HI2424 and PC184 showed no conservation of these sequences.

Regarding the ISBmu1 sub-group, six elements were detected among *B. multivorans* ATCC17616 genome. These elements have identical DNA sequences but no DR could be found. This sub-group was described as being active in strain ATCC17616, during starvation (Ohtsubo et al., 2005). Two elements of this sub-group were detected among *Burkholderia phytofirmans* PsJN genome, but no other copy could be detected in the databases.



**Fig. 16: Main features of IS21 and IS256 families elements found among *B. cenocepacia* strains.** Inverted DNA repeats are indicated by a green vertical line; red vertical lines are used to indicate a frameshift; ORFs are indicated by arrows and ORFs with a deletion are represented by arrows without a vertical extremity. ORFA: total transposase sequence (IS256 family) or amino-terminal end (IS21 family), ORFB: carboxy-terminal end (IS21 family). ORF orientation of IS elements are according to the main orientation of the genomes, beginning from the origin of replication. PHDCs indicate genomes of the HI2424 and AU1054 strains, and Midwest the one of strain PC184. The J2315 IS1356-b is described within Fig. 18 and ISBcen18 and ISBcen13-e copies are described within Fig. 4.



**Fig 17: IS21 family Neighbor-Joining phylogenetic trees.**

DNA sequences were retrieved from *Burkholderia* sequenced genomes and Genbank. Alignments were performed using CLUSTALW and NJ trees were inferred using Mega4 graphic tool. A total of 1769 sites were analyzed (with global deletion of gap-containing sites). Distances are proportional to evolutionary divergences expressed in substitutions per 100 sites. Bootstrap values higher than 85% are given. *Bcen* = *B. cenocepacia*; *Bph* = *B. phenazinium*; *Bgl* = *B. glumae*; *Bmu* = *B. multivorans*; *Bvi* = *B. vietnamiensis*.

### 3.3. IS family sub-groups with a narrow distribution in *B. cenocepacia*

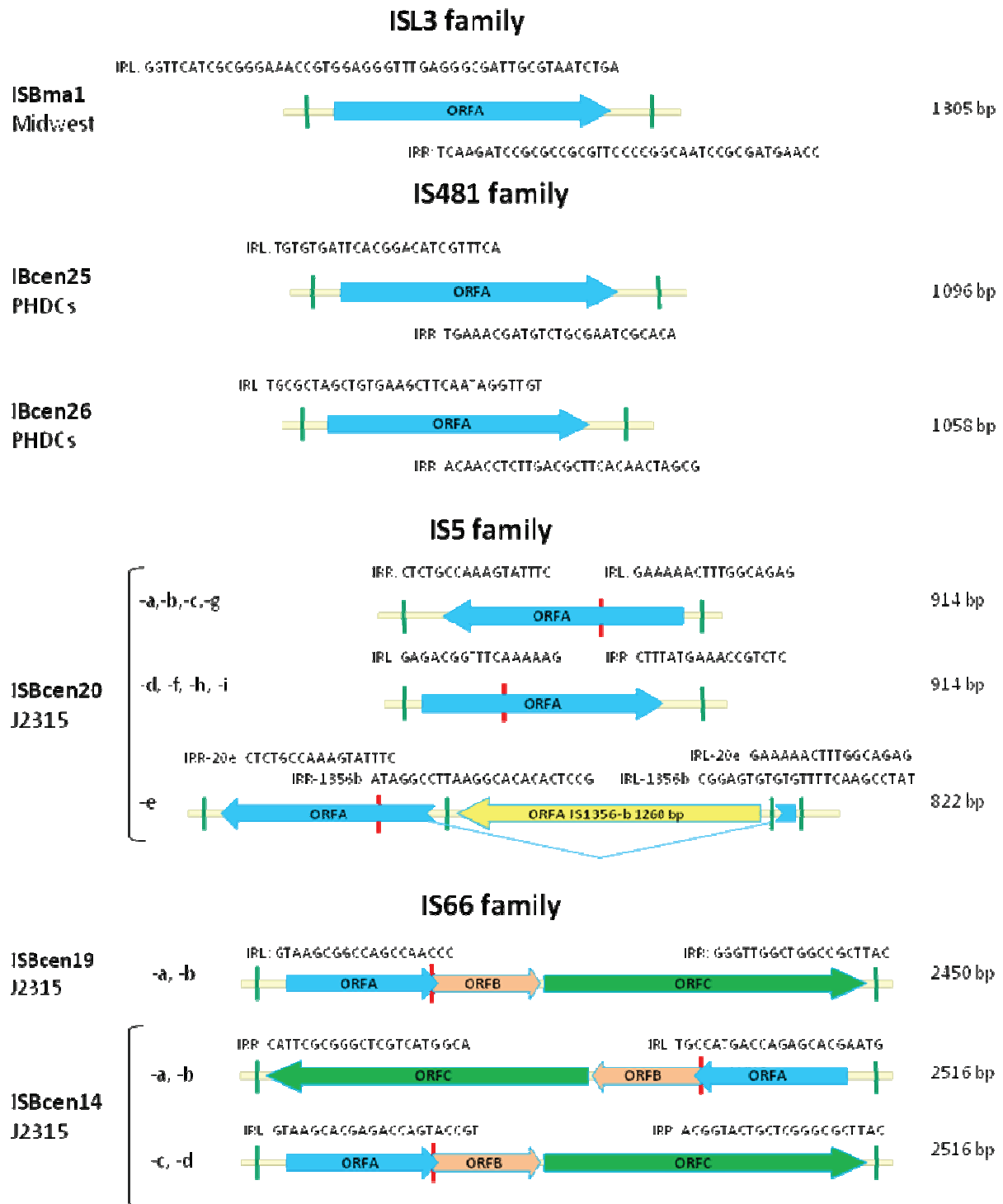
IS elements of the IS5, IS66, ISL3 and IS481 families showed a narrow distribution in the *B. cenocepacia* species. In this study, phylogenetic analyses were performed to infer sub-groups among these families and investigate the origin of these elements. The genetic organization of these elements is described in Fig. 18, and their respective DR in Table 4. The position of these IS is shown on the genome of *B. cenocepacia* J2315 (Fig. 5).

#### 3.3.1. IS5

Phylogenetic analysis of the IS5 family among the *Burkholderia* was performed using their respective DNA sequences encoding putative transposases (Fig. 19). Two main clusters were obtained, one grouping sequences named IS*Bcen20* and IS402 according to IS biotoul, and one sequences of the IS*Bmu2* sub-group. IS402 and IS*Bcen20* are found to be closely related elements, they share 92% identity and DR sequences are often the same (TTA). These two IS elements are considered to be two isoforms of the same element. Nine copies of IS*Bcen20* were detected in the J2315 genome, with one found to be disrupted by the IS1356-e (Fig. 18). One partial IS*Bcen20* was detected among *B. ambifaria* AMMD but no other copy was found in the databases. All J2315 sequences of IS*Bcen20* were identical, suggesting a recent wave of amplification among this genome. Four of these IS*Bcen20* disrupted CDS: IS*Bcen20*-b, -c were inserted in sequences coding for hypothetical proteins; and IS*Bcen20*-f and -i were respectively inserted among a CDS encoding a putative glycosyl-transferase (BCAL3125) and a CDS encoding a putative phage integrase (BCAM1923). A phage integrase can mediate unidirectional site-specific recombination between two DNA recognition sequences (*attB* and *attP*) (Groth and Calos 2004), thus inactivation of its CDS could fix a phage in a genome. Other phage CDS in this region were found colonized by the IS407-1 and IS*Bcen20*-h elements.

IS402 elements have been detected among *B. vietnamiensis* G4 (n=11) and *B. multivorans* ATCC17616 (n=3) (Table 3). No copy of IS402 was detected in the J2315 genome. IS402 AA sequence share similarities with sequences from *Burkholderia glumae* BGR1 (76% id) and with IS*Bcen20* e. g. *Burkholderia phytofirmans* PsJN (60% id). IS402 transposase seemed to be a "fusion" between two ORFs, as suggested by comparison with other IS5 family members (Ferrante and Lessie 1991). In fact, IS402 was described as promoting replicon fusion (Barsomian and Lessie 1986) and to increase foreign gene expression (in both directions) (Scordilis et al., 1987). These elements were detected in high copy number in *B. vietnamiensis* G4 and *B. multivorans* ATCC17616 genomes, and seemed to be highly active. The DNA sequences of these IS among these two genomes were found identical or very closely related (Fig. 19).





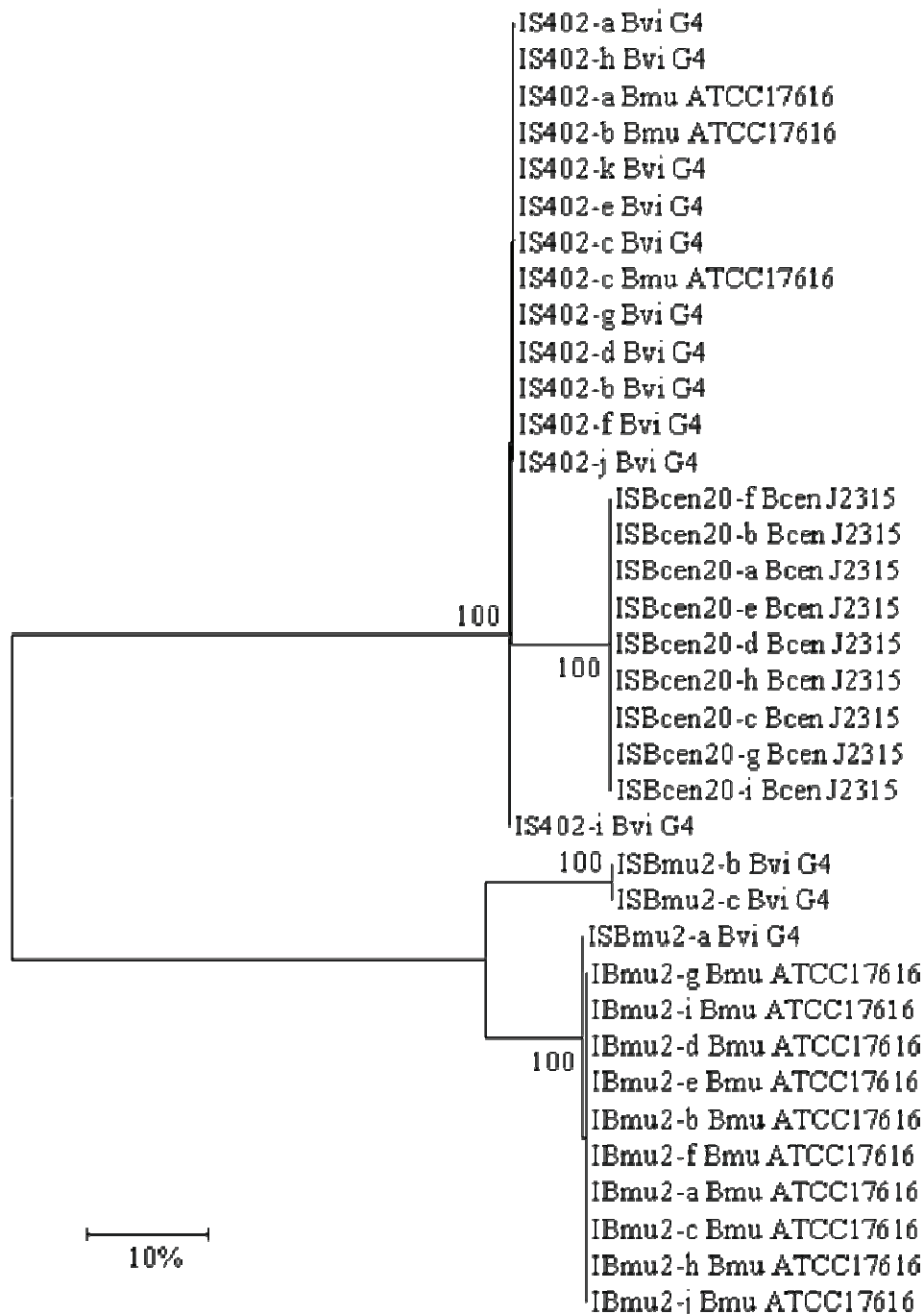
**Fig. 18: Main features of IS elements with a narrow distribution among *Burkholderia cenocepacia* strains.**

Inverted DNA repeats (IR) are indicated by a green vertical line; red vertical lines are used to indicate a frameshift. ORFs/CDS are indicated by arrows and CDS with a deletion are represented by arrows without a vertical extremity. ORFA: total transposase sequence (ISL3, IS5, IS481 families) or amino-terminal end (IS66 family), ORFC: carboxy-terminal end (IS66 family). ORF orientation of IS elements are according to the main orientation of the genomes, beginning from the origin of replication. PHDCs indicate genomes of the HI2424 and AU1054 strains, and Midwest the one of strain PC184.

IS*Bmu2* elements of the other main cluster of IS5 elements among the *Burkholderia* were detected among *B. multivorans* ATCC17616 (n=11) and *B. vietnamiensis* G4 (n=3) (Table 3). No copy of this sub-group was found among *B. cenocepacia* strains, except *Burkholderia glumae* BGR1 which showed a single element (85% id). The closest putative transposase among other bacterial genomes was the one of IS1021 of *R. solanacearum* GMI1000 which show 80% identity with IS*Bmu2* transposases. Phylogenetic analysis of IS*Bmu2* transposase CDS extracted from *B. vietnamiensis* G4 and *B. multivorans* ATCC17616 showed two sub-lineages, one grouping sequences of ATCC17616 genome and one sequence from the G4 genome, and the other lineage grouping the other two elements from the G4 genome (Fig. 19). The eleven copies detected among ATCC17616 presented identical sequences, suggesting a recent wave of amplification. IS*Bmu2* have been described as the most active sequence of ATCC17616. This sub-group represents 75% of the IS elements transposed under starvation, oxidative stress, high temperature and normal conditions (Ohtsubo et al., 2005). This IS element seemed to play a role in the adaptation of *B. multivorans* ATCC17616 to changing environmental conditions.

### **3.3.2. IS66**

Only two IS elements of the IS66 family have been detected among strains investigated in this study. These elements, named IS*Bcen19* and IS*Bcen14*, were detected respectively in two and four copies among J2315 genome (Table 3; Fig. 5). DR of IS*Bcen14* could be identified (Table 4). A partial copy of IS*Bcen14* and a full copy of IS*Bcen19* were detected among *B. cenocepacia* MC0-3, but no other copy of these elements was detected in the database. This suggests a restriction of IS*Bcen14* and IS*Bcen19* to the *B. cenocepacia* species. Nevertheless, no copy of these two IS elements could be detected among PHDC and Midwest genomes. The four IS*Bcen14* DNA sequences of J2315 are identical (data not shown) suggesting a recent amplification wave of this element. Two of these IS elements were found integrated inside CDS, with one encoding a hypothetical protein (BCAL2219) and one a putative pyridoxal-dependent decarboxylase (BCAM0245). This last copy of IS*Bcen14* is part of the *Burkholderia cepacia* Epidemic Strain Marker (BCESM), and was detected only among a few *B. cenocepacia* strains (Baldwin et al., 2004). The two IS*Bcen19* copies of J2315 showed identical DNA sequences and share 98% identity with the copy detected among *B. cenocepacia* MC0-3 genome.



**Fig. 19: IS5 family Neighbor-Joining phylogenetic tree using transposase CDS DNA sequences from the *Burkholderia*.**

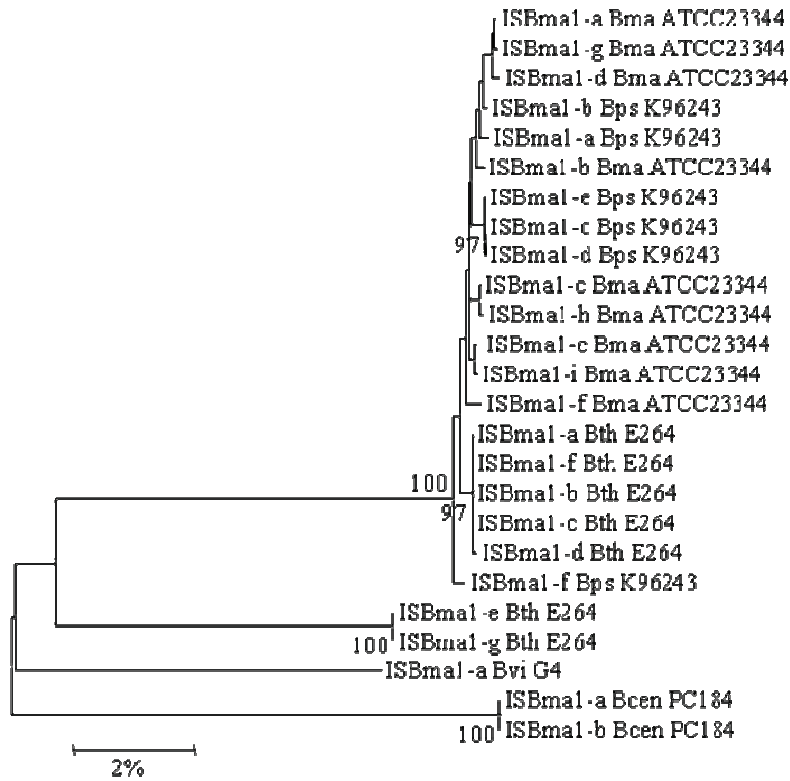
DNA sequences were retrieved from *Burkholderia* sequenced genomes and Genbank. Alignments were performed using CLUSTALW and NJ trees were inferred using Mega4 graphic tool. A total of 795 sites were analyzed (with global deletion of gap-containing sites). Distances are proportional to evolutionary divergences expressed in substitutions per 100 sites. Bootstrap values higher than 85% are given. *Bcen* = *B. cenocepacia*; *Bmu* = *B. multivorans*; *Bvi* = *B. vietnamiensis*.

### **3.3.3. ISL3**

*ISBma1* was the only IS sub-group of the *ISL3* family detected among the *Burkholderia* genomes. This sub-group was detected among *B. cenocepacia* PC184 (n=2), *B. vietnamiensis* G4 (n=1), *B. mallei* ATCC23343 (n=17), *B. pseudomallei* K96243 (n=5) and *B. thailandensis* E264 (n=7) (Table 3). Phylogenetic analysis of the *ISBma1* DNA sequences encoding a transposase revealed four major sub-lineages; one corresponding to the *B. pseudomallei*, *B. mallei* and most of *B. thailandensis* sequences, one corresponding to the other *B. thailandensis* sequences, one to the *B. vietnamiensis* element, and one to the two *B. cenocepacia* PC184 elements (Fig. 20). All *B. mallei*, *B. pseudomallei* and *B. thailandensis* sequenced strains were found to harbour these elements. These three species are closely related, *B. mallei* being considered a clone of *B. pseudomallei*, as well as *B. thailandensis*, but a non-virulent one (Redfearn et al., 1966, Brett et al., 1998). The genomic organization of these IS elements confirmed this proximity between these three species. Nevertheless, two *ISBma1* sequences of strain E264 grouped outside this group suggesting an independent acquisition. *ISBma1* was suggested to have played part in the genomic rearrangements leading to the genome size reduction of *B. mallei* (Holden et al., 2004).

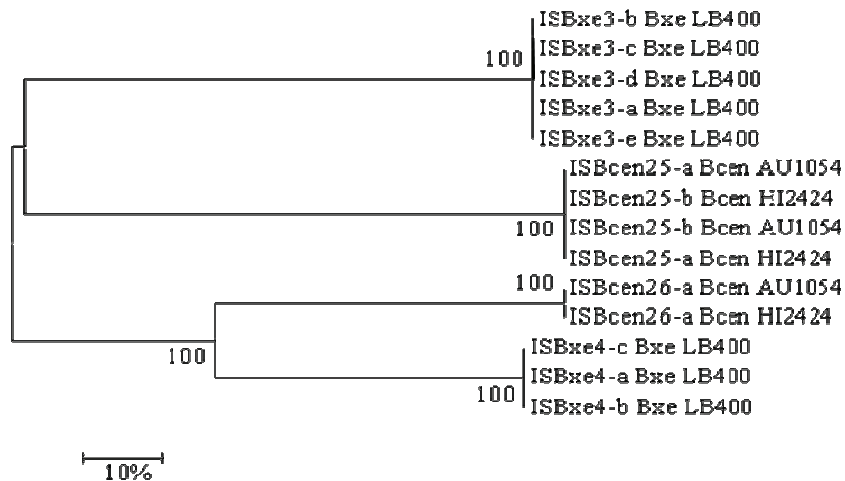
### **3.3.4. IS481**

Transposases of the *IS481* family were found only among three species of *Burkholderia*. Phylogenetic analysis of their CDS encoding these transposases led to the observation of four main sub-groups: *ISBxe3*, *ISBxe4*, *ISBcen25*, and *ISBcen26* (Fig. 21). Two of these sub-groups, *ISBxe3*, *ISBxe4*, were only detected among the *B. xenovorans* LB400 genome. *ISBcen25* elements were detected among *B. cenocepacia* AU1054 (n=2) and HI2424 (n=2) (Table 3). These elements have identical DNA sequences, and suggest a colonization of these strains prior their divergence. Analysis of their insertion regions confirmed this hypothesis; the two *ISBcen25* of the PHDC genomes were found inserted in the same site. Interestingly, these two sites were absent from J2315. Regarding the *ISBcen26* sub-group, similar observations were made. The same DNA region was colonized by these elements among the PHDC genomes, and this DNA region was absent from J2315. These observations might be indicative of an acquisition of these elements after the divergence of the PHDC and ET12 lineages. No copy of *ISBcen26* disrupted a CDS, but one copy of *ISBcen25* was inserted among a transposase CDS (*ISBtuh1*, Fig. 4). BLASTN analysis detected a copy of *ISBcen26* among *B. cenocepacia* MCO-3 strain (99% id). The potential role of these two IS elements are still unknown but *ISBcen25* seemed restricted to the PHDC lineage, and might thus be used as a marker of this clone.



**Fig. 20: *Burkholderia* ISBma1 Neighbor-Joining phylogenetic tree using CDS DNA sequences of their respective transposase.**

DNA sequences were retrieved from *Burkholderia* sequenced genomes and Genbank. Alignments were performed using CLUSTALW and NJ trees were inferred using Mega4 graphic tool. A total of 1305 sites were analyzed (with global deletion of gap-containing sites). Distances are proportional to evolutionary divergences expressed in substitutions per 100 sites. Bootstrap values higher than 85% are given. Bcen = *B. cenocepacia*; Bps = *B. pseudomallei*; Bma= *B. mallei*; Bth = *B. thailandensis*.



**Fig. 21: *Burkholderia* IS481 family Neighbor-Joining phylogenetic tree using CDS DNA sequences of their respective transposase..**

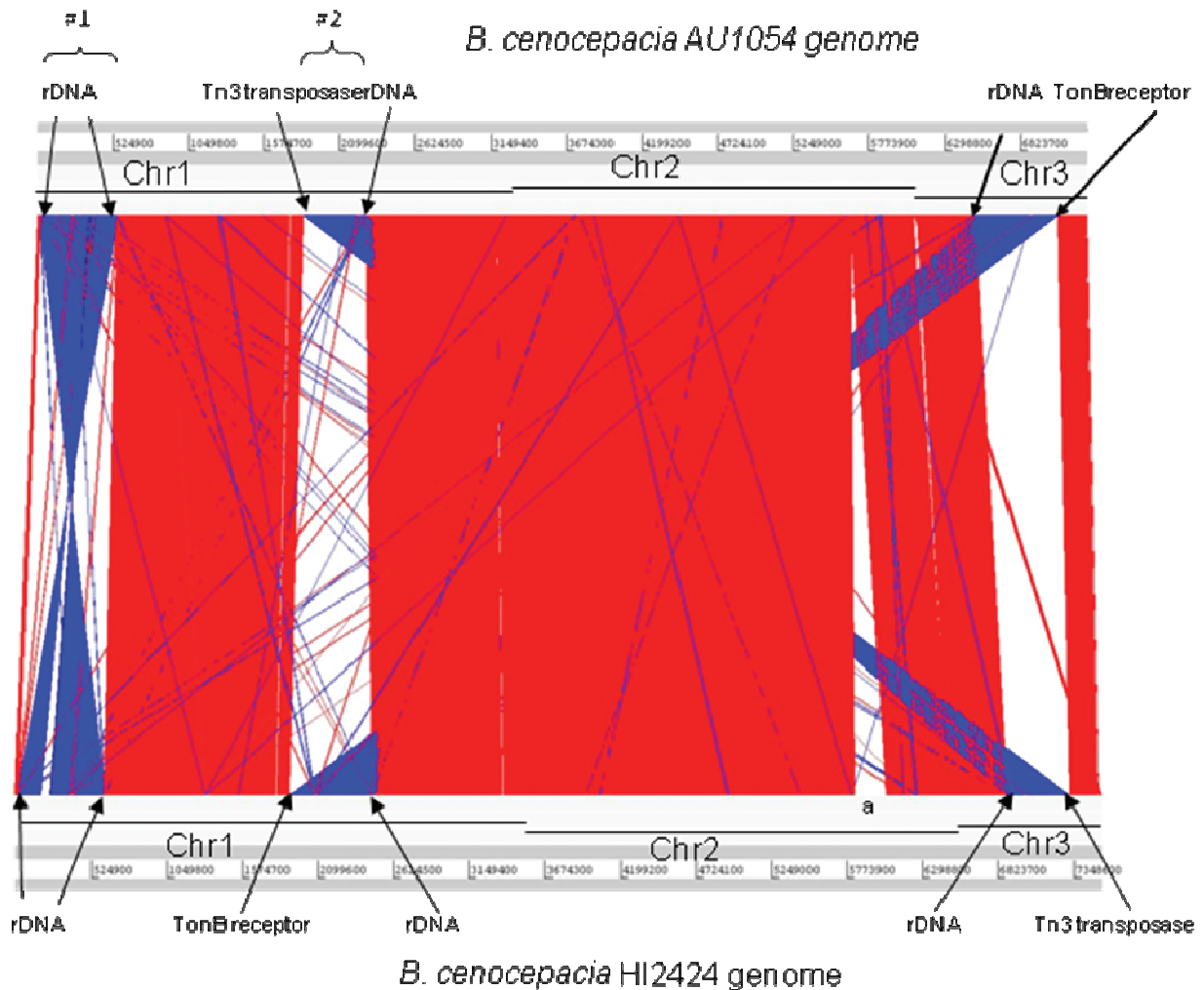
DNA sequences were retrieved from *Burkholderia* sequenced genomes and Genbank. Alignments were performed using CLUSTALW and NJ trees were inferred using Mega4 graphic tool. A total of 978 sites were analyzed (with global deletion of gap-containing sites). Distances are proportional to evolutionary divergences expressed in substitutions per 100 sites. Bootstrap values higher than 85% are given. Bcen = *B. cenocepacia*; Bxe= *B. xenovorans*.

### 3.4. IS and large scale genomic rearrangements.

To detect genomic rearrangements involving IS elements, the two *B. cenocepacia* strains of the PHDC epidemic clone were compared using the ACT approach. This comparison is presented in Fig. 22. In order to trace these genomic rearrangements, genome sequences of these two strains were aligned using the same starting nucleotide position. For chromosome 1, the starting position was set at the beginning of *dnaA*, a gene at the proximity of the replication origin. For chromosome 2, the starting position was set at the beginning of *parA*, a gene involved in replication and close to the replication origin (Dubarry et al., 2006).

The ACT analysis of these chromosomes showed an inversion of 500 kb close to the replication origin (at about 15 kb from the origin) of chromosome 1, and named rearrangement #1 on Fig. 22. Inside this region, a 60 kb DNA fragment was only found in the HI2424 genome (and not in the one of AU1054). Anyhow, this segment was detected among the J2315 genome. Phage related DNA was found in this fragment and suggests that it could have been lost from the AU1054 genome through an excision process. A second rearrangement was observed in chromosome 1, and named rearrangement #2 on Fig. 22. A segment of chromosome 1 was exchanged with a segment of chromosome 3 of about the same size (420 kb). The HI2424 chromosome 1 organization of this region was found conserved with the one of J2315 chromosome 1. It thus appears that the rearrangement occurred recently in the AU1054 genome. The organization of chromosome 2 was found more stable than the one of chromosome 1, with the exception of a 200 kb region only found in the HI2424 genome, and named "a" on Fig. 22. The genetic environment of the above two rearrangements, 500 kb inversion and 420 kb exchanged segment, observed among chromosome 1, were investigated to infer the processes involved. The 500kb area close to the origin of chr1 was found bordered by two *rrn* (rRNA) operons suggesting a direct involvement in the inversion process. The exchanged region between chr1 and chr3 is bordered on the left hand by an IS-like element of the Tn3 family (Fig. 4) and by a rDNA (*rrn*) operon on the right hand. This IS was classified into the Tn3 group because of high AA sequence similarities with transposases of this group. This IS-like Tn3 element was found to encode a CDS showing identities with a phage-like integrase (Fig. 4). Among other bacterial species like *Francisella tularensis*, a duplicated pathogenicity island was found bordered by a rDNA operon and an IS element (Larsson et al., 2005). Interestingly, in these analyses of the PHDC clone, the *rrn* operons appeared major elements in genomic rearrangements. These operons are longer than IS elements, and found in multiple copies in *Burkholderia* genomes, making them good candidates to trigger recombination events. In fact, frequent recombinations between the *rrn* operons are assumed to be one of the mechanisms explaining their concerted evolution (Liao 1999). It is noteworthy that replication of chromosome 2 and 3 of *B. cenocepacia* are based on plasmid-like

mechanisms and thus their copy numbers might be different from those of the more “classic” chromosome 1 replication system (*oriC* and *dna* operon). In consequence, translocation from chromosome 1 to chromosome 3 may have an impact on gene expression.



**Fig. 22: Global synteny analysis of the *B. cenocepacia* AU1054 and HI2424 genomes.**

*B. cenocepacia* AU1054 and HI2424 concatenated genomes were compared using ACT. Chromosome 1 sequences in the databases were modified to start with the *dnaA* gene sequence. Chromosome 2 sequences were modified to start with *parA*. Genes found on synteny breaking points are indicated. #1 and #2 correspond to rearrangements described within the text. A: DNA fragments of 200 kb that was only detected in the HI2424 genome.

Lessie et al., (1996) proposed that IS may contribute significantly to the genomic plasticity of *Burkholderia* but they only had examples of a local action of IS with a recruitment of catabolic genes such as those involved in the ability to degrade the 2,4,5-T herbicide. Holden et al., 2004 in their paper describing the *B. mallei* ATCC 23344 genome emphasized the key role of IS in the genome reduction processes. Holden et al., (2009) presented IS invasion as an adaptation process which could select or counter-select genes depending on the biotope. Menard et al., (2007) presented an example of IS implication in a global genomic rearrangement. Here, we have confirmed the likely involvement of IS elements in *B.*

*cenoepectia* genomic rearrangements, and showed these events to be linked not only to the IS110 family but also other families like those showing similarities with Tn3 transposons.

## 4. Conclusions.

*Burkholderia* genomes are characterized by a high genomic plasticity. The richness of IS elements among their genomes may explain parts of this plasticity. Indeed, IS elements are able to activate or disrupt genes, leading to phenotypic changes but also to favor large rearrangements by recombination between elements. Virulent species and epidemic clones like *B. mallei* and *B. cenoepectia* have been shown to harbor high copy number of IS elements. A potential “linkage” between IS elements, virulence and emergence of epidemic clones was thus hypothesized. In order to verify this hypothesis, an in depth investigation of insertion sequences (IS) elements genomic organization among *Burkholderia* of the *cepacia* complex was undertaken.

These investigations were performed using full genome sequences and completed by experimental investigations. The results showed a heterogeneous distribution of IS elements among the twelve genomes investigated. IS copy numbers ranged between 7 to 157 among the *Burkholderia*. IS numbers varied significantly within *B. cenoepectia*; J2315 genomes harboring as much as 3 times more IS elements than PHDC or Midwest genomes. Nine families were identified, some of them broadly distributed (IS3, IS21, IS110 and IS256) and some other showing a narrow distribution (IS4, IS5, IS66, IS481 and ISL3). Several IS families were restricted to a few strains, IS66 and IS5 for J2315 and B&B clones, IS481 for PHDC genomes, and ISL3 among the PC184 genome. Variations observed for IS numbers and families indicate a strong relationship with the evolutionary history of the strains. These relationships might be related to recent acquisitions or amplification of IS elements of an existing family.

Interestingly, IS distribution and organization showed similarities between the ET12 and B&B clones. IS*Bcen4*, IS*Bcen6*, IS*Bcen8*, IS*Bcen9*, IS*Bcen20*, and IS407 which are present in the J2315 genome, were also detected among the B&B clone. Investigation of IS407 distribution confirmed this genetic proximity between the two clones but also showed the absence of some loci/alleles of J2315 in the B&B clone. This heterogeneity apparently results from recent DNA acquisitions and distinct amplification waves of the IS407. In addition, IS*Bcen9* and IS*Bcen6* were only detected among the J2315 and B&B genomes. Phylogenetic analysis of IS*Bcen9* showed 100% identity with the ones of J2315. These results confirmed the genetic proximity of these two clones, confirming inferences made in the previous chapters. Nevertheless, several divergences were detected between J2315 and the B&B clones, notably concerning the IS*Bcen4*. This element, found in two copies in J2315 genome, was found in at



least 5 copies within the B&B genome. Moreover, phylogenetic analysis revealed only one sequence identical with J2315 *ISBcen4*. These data suggest a recent acquisition by the B&B clone of a novel and active *ISBcen4*

Functional implications of IS insertions were also investigated. Thereby, CDS disrupted by IS elements were examined for the four sequenced *B. cenocepacia* genomes. Results showed differences between ET12 and other lineages, 10% to 20 % of the IS elements being detected inside CDS for PC184 and PHDC genomes while 38% were found inside CDS for J2315. Interestingly, a large part of IS disruptions occurred among other transposases CDS. Looking at the direct repeats of the IS elements, some DNA targets could be identified but without any relation with a particular gene family. However, several gene disruptions could have played a role in the emergence of genetic variants. For example *IS407* was found involved in UV resistance making the strain less likely to survive in outdoor environments.

The role of insertion sequences in genetic rearrangements is just beginning to be documented and the ecological role of these events is still not clear. Some IS element are likely to activate transcription of genes near their integration site but the benefits of such events remain unknown.

This study gave us an overview of IS elements distribution and their potential role in the emergence of epidemic clones among *B. cenocepacia*. Dynamic of IS amplification events under stress conditions would now need to be investigated.

# Ilots génomiques et éclosions épidémiques chez les *Burkholderia cenocepacia* : une association étroite

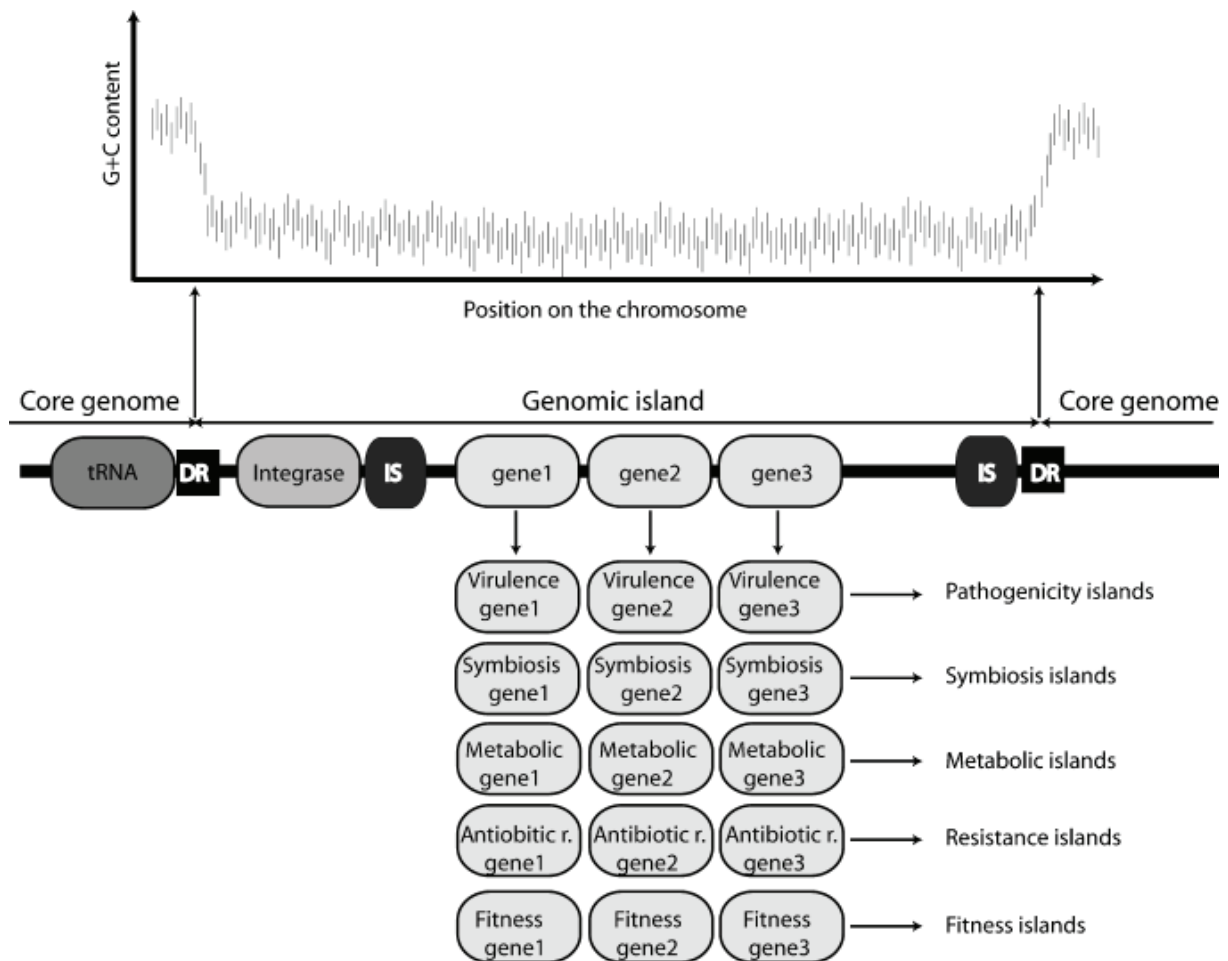
Au cours des chapitres précédents, la répartition mondiale du clone B&B, sa prévalence en France et sa proximité avec la lignée ET12 ont été étudiées. La structure génétique des facteurs transcriptionnels de la famille sigma70 ( $\sigma70$ ) suggère deux stratégies évolutives différentes entre *P. aeruginosa* et *B. cenocepacia*, et a permis d'identifier une structure génétique différente de celles des autres espèces du Bcc. Ceci a mis en évidence l'importance des réarrangements génétiques dans la distribution des  $\sigma70$  et l'évolution de l'espèce *B. cenocepacia*. De plus, l'étude des éléments génétiques répétés de la famille des séquences d'insertion (IS) a permis d'observer une organisation génomique du clone B&B distincte de celle de la lignée ET12. Ces différences ont été reliées à des phénomènes d'instabilité au niveau des loci colonisés par ces éléments, pouvant être expliquées par des phénomènes d'acquisition d'éléments génétiques mobile de type îlot génomique (IG). De plus, une colonisation préférentielle de certaine IS chez les clones ET12 et B&B a été observée. Les réarrangements génétiques, notamment l'acquisition/délétion d'ilots génomiques, semblent jouer un rôle important dans l'évolution de ces bactéries ; il apparaît donc fondamental de déterminer l'importance de ces phénomènes au niveau de l'espèce *B. cenocepacia* et du clone B&B, et de mieux définir leurs implications dans l'émergence de clones épidémiques.

Chez les procaryotes, différents facteurs sont impliqués dans l'évolution des génomes dont les mutations ponctuelles, la recombinaison, l'inactivation de gènes par des *indels*, et le transfert de matériel génétique entre différentes espèces bactériennes. Ce dernier processus, dénommé transfert horizontal de gènes (HGT), représente une pièce maitresse de l'évolution des bactéries et peut conduire à des changements drastiques dans la composition des génomes microbiens sur des périodes relativement courtes (Ochman et al., 2000). L'augmentation exponentielle récente des séquençages de génomes a permis le développement d'outils de génomique comparative contribuant à mieux apprécier ces phénomènes de HGT. Différents travaux de génétique et de comparaison de génomes appartenant à des souches plus ou moins éloignées ont mis en évidence des régions d'ADN non conservées entre ces souches, donnant ainsi naissance à la notion "d'îlot génomique" (IG) (Hacker and Carniel 2001). Les ilots génomiques sont des fragments d'ADN : 1. ayant une taille allant d'environ 5kb à plusieurs centaines de kb ; 2. portant généralement des gènes impliqués dans la mobilité, des

transposases et des intégrases ; 3. présentant la plupart du temps un G+C% différent du reste du génome ; 4. souvent intégrés au niveau d'ARNt, et 5. couramment flanqués de régions répétées (Figure 1). Les IG peuvent être classés dans différentes catégories en fonction des gènes qu'ils contiennent ; un des premiers îlot génomique décrits abrite des gènes impliqués dans la virulence d'une souche d'*Escherichia coli*, il a ainsi été nommé "îlot de pathogénie" (PAI) par Hacker et al., (1990). Il y a quelques années, la distribution de ces PAI était considérée comme faible. Cependant, le développement des techniques de séquençage de génomes ainsi que les études de génomique comparative ont permis de montrer une répartition plus large de ces IG. Les termes "îlot génomique" ou "îlots de fitness" ont alors été utilisés pour décrire l'ensemble des types d'IG (Figure 1) (Strauss and Falkow 1997; Hacker and Kaper 2000; Hacker and Carniel 2001). De nombreux IG ont été identifiés chez les *Burkholderia* notamment grâce au séquençage de nombreuses souches au sein de ce genre, qui compte désormais plus de 50 génomes séquencés ou en cours de séquençage ([www.genomesonline.org](http://www.genomesonline.org)). Une analyse réalisée sur 5 génomes de l'espèce *B. pseudomallei* a démontré une grande plasticité génétique de ces souches, liée notamment à la distribution hétérogène de leurs IG (Tuanyok et al., 2008). Ces phénomènes de transfert d'ADN semblent, de plus, jouer un rôle important dans l'émergence de souches hypervirulentes, comme cela a été récemment démontré chez *Streptococcus pneumoniae*, où des loci récemment acquis au sein du chromosome joueraient un rôle dans la résistance à des composés antimicrobiens (Croucher et al., 2009). De nombreux autres travaux, notamment au sein des espèces *Yersinia pestis* et *Yersinia pseudotuberculosis*, ont démontré l'importance des IG et des IS dans les processus évolutifs de ces espèces et leur spéciation (Chain et al., 2006; Eppinger et al., 2007).

Afin d'identifier de potentiels IG impliqués dans l'émergence de clones épidémiques tels que les souches J2315 et le clone B&B, une comparaison des génomes de la souche J2315 et des deux souches PHDC a été entreprise durant cette thèse. Ces travaux ont permis de détecter de nombreuses séquences d'ADN spécifiques de la souche J2315, avec plus d'une vingtaine d'entre elles présentant les caractéristiques d'îlots génomiques. Ces travaux ont également pu mettre en évidence une relation particulière entre les IG et certaines IS, permettant ainsi la détection d'IG chez le clone B&B. Ces différents résultats sont présentés dans la publication ci-dessous. Les analyses présentées dans ce chapitre ont été initiées en 2007. Depuis, le génome de la souche J2315 a été publié par Holden et al., (2009). Cet article a permis de mieux comprendre le "succès" évolutif de la souche *B. cenocepacia* J2315 et la façon dont la lignée ET12 semble s'être récemment adaptée à la niche "patient CF". Dans

cette publication, les auteurs ont comparé différents génomes leur permettant ainsi d'identifier 13 IG chez J2315. Ces 13 IG avaient également été identifiés au cours de mon doctorat.



**Fig. A: Caractéristiques générales des îlots génomiques.**

Les IG sont des segments relativement importants d'ADN dont les caractéristiques diffèrent souvent du reste du chromosome. Les IG sont souvent insérés au niveau de gènes codant des ARNt et flanqués par des régions répétées (DR). Les gènes présents au niveau des IG codent généralement pour des facteurs impliqués dans la mobilité de l'îlot, comme les intégrases, transposases et des séquences d'insertion (IS). En fonction de leur contenu en gènes, les IG peuvent être décrits comme des îlots de pathogénie, symbiotiques, métaboliques, de fitness ou de résistance. (schema tirée de Juhas et al., 2009).

# Genomic islands and IS elements: a significant association among the *Burkholderia cenocepacia* ET12 and B&B genomes

## 1. Introduction

*B. cenocepacia* belongs to the *Burkholderia cepacia* complex (Bcc), a group of seventeen closely related species. These species are able to colonize a large panel of environmental niches such as soils, freshwaters, plants, amoebae. However, some are also opportunistic pathogens infecting immuno-compromised individuals or individuals with cystic fibrosis (CF). *B. cenocepacia* is the main Bcc species recovered from the respiratory tract of CF patients in France, and can lead to a dramatic necrotic pneumonia named the “*cepacia* syndrome” (e. g. Saiman and Siegel 2004). Several epidemic clones of *B. cenocepacia* have been described among CF cohorts. The ET12 clone was found to have affected North American, Canadian, and United Kingdom CF cohorts, and was shown to be highly transmissible (Isles et al., 1984; Johnson et al., 1994). The PDHC and Midwest clones have been mainly detected in North American CF patients (LiPuma et al., 1988; Coenye et al., 2004). The B&B clone was first described as the etiological agent of a nosocomial outbreak (Graindorge et al., 2009) but recent MLST analyses showed this clone to belong to the most prevalent Bcc clonal complex found among French CF patients (Abalain et al., 2009). This clonal complex can thus be considered highly transmissible among CF cohorts but also among intensive care unit patients. Differentiation of these epidemic clones might have involved several genomic changes such as spontaneous mutations, disruption of genes by insertion sequences and DNA acquisition/deletion. This last phenomenon includes gains and losses of DNA such as “Genomic Islands” (GI).

GI acquisition/excision is broadly occurring among bacteria and can lead to major benefits, and an improved fitness (Hacker and Carniel 2001; Dobrindt et al., 2004). The first GI was characterized among a pathogenic strain of *Escherichia coli* and termed a “Pathogenic Island or PAI” because of its involvement in virulence (Blum et al., 1994). A number of other pathogenic islands (PAI) have been characterized, notably among *Salmonella typhimurium* and *Vibrio cholerae* (Mills et al., 1995; Prager et al., 1995). The analysis of fully sequenced

genomes helped to identify different kinds of GI like symbiosis islands, saprophytic islands and ecological islands. All these types of GI could be grouped as “fitness island” because of their implications in bacterial adaptation to new environments (Preston et al., 1998). Different characteristics are shared by these GI like a distinct bias in G+C %, presence of integrases, transposases and direct repeated sequences on the two borders of their integration (Dobrindt et al., 2004). Integration sites are often located in regions coding for tRNA (Boyd et al., 2009). GI are often unstable and their excision could occur with varying frequency. They have been described in many bacterial species, Gram negative, but in some species of Gram-positive bacteria (Hacker and Carniel 2001). Little is known about GI origin but it has been hypothesized that they may have been derived from integrative plasmids or phages that have lost the genes implicated in their replication and self transfer (Dobrindt et al., 2004). Some genomic islands code for unknown phenotype, but most of them encode biochemical pathways or factors of pathogenicity in the case of PAI. Interestingly, insertions sequences (IS) are often detected among GI but relations between GI and IS are still unclear. IS elements could be involved in genetic recombination within GI leading to phenotypic changes like it has been described for *Salmonella enterica* multidrug (MDR) genes cluster which was colonized by diverse IS elements (Doublet et al., 2008). IS elements could also disrupt genes leading to phenotypic variation including virulence patterns (DeShazer et al., 2001).

In this study, an identification of GI into the B&B clone using the different available genomes of *B. cenocepacia* as a tool has been performed. J2315 genome (ET12 lineage) has been compared with the two HI2424 and AU1054 genomes (PHDC lineage) and thus 58 different DNA sequences specific of the J2315 strain could be detected. These DNA sequences were briefly described and several of them, related to genomic island, were deeply characterised. Six of these GI presenting high density of IS elements were selected and their presence among the B&B clone using a PCR/sequencing approach was investigated. Results demonstrated the presence of only two of these GI within the B&B clone. Finally, using IS407 as a DNA probe to screen a genomic DNA library of the B&B clone, two GI specific of the B&B clone were detected. These two GI showed homologies with *P. aeruginosa* and *R. solanacearum* genomes and harboured genes potentially involved in virulence.

## 2. Material and methods

### 2.1. Strains, DNA extractions, PCR amplification and sequencing.

*Burkholderia* strains used in this study are listed in Table 1. Bacteria were grown in Luria-Bertani (LB) broth (Bertani 1951) at 28°C for 24 to 48h. DNA extractions were performed according to Pitcher 1989 and resuspended in distilled water.

PCR (polymerase chain reaction) amplifications (25-50 µl) were done according to the *Taq* polymerase manufacturer (Invitrogen, Cergy Pontoise, France) without 1% w-1 detergent but with 5% DMSO. PCR primers used in this study are described in Table 2. PCR cycles using total DNA were as follow: (1) 95°C for 5 min, (2) 95°C for 30 sec, selected annealing temperature (e.g. Table 2) for 45 sec, 72°C for 60 sec (35 cycles), (3) 72°C for 5 min.

DNA sequences of B&B and LMG12614 strains were obtained after cloning PCR products in pGEMT-easy (Promega, Charbonnieres, France) into *Escherichia coli* DH5α strains. Plasmids (pGEM-T) extraction was performed using QIAGEN® Plasmid Plus Midi Kit (Qiagen, Courtaboeuf, France). DNA sequencing of PCR products and pGEM-T plasmids was performed by Cogenics® (Grenoble, France).

**Table 1: Bacterial strains used in this study**

<i>B. cenocepacia</i>	Other code	Origin	Reference
LMG 16656	J2315	ET12, Cystic Fibrosis	Vandamme et al., 2003
LMG 12614		ET12, Cystic Fibrosis	Vandamme et al., 1997
HI2424		PHDC, soil	LiPuma et al., 2002
AU1054		PHDC, Cystic Fibrosis	Chen et al., 2001
B&B clone	Bpoe732	Pulmonary infection	Graindorge et al., 2009

## 2.2. B&B clone genomic library

Bpoe732 genomic DNA was extracted as described by Pitcher (1989) and partially digested with *Sau3AI* (Fermentas, St. Rémy Les Chevreuse, France). The digested DNA fragments were separated by sucrose density gradient centrifugation. Fragments in the range of 8 to 10 kb were ligated into pCPP47, which contains tetracycline and ampiciline resistance genes. These vectors were previously digested with *BamHI* (Fermentas, St. Rémy Les Chevreuse, France) and dephosphorylated with calf intestinal alkaline phosphatase (CIAP) (Stratagen, Massy, France). After enzyme inactivation, the dephosphorylated plasmid was purified using a QIAGEN PCR purification kit (QIAGEN, Courtaboeuf, France). Ligation reactions were performed using a 1:1 ratio of vector and insert. Ligated DNA has been encapsided using Gigapack III Gold Packaging Extract (Stratagen GmbH, Zurich, Switzerland) following the manufacturer recommendations. Transduction was performed using 200µl of XL1-Blue *E. coli* cells (Stratagen, Massy, France) (DO=0.7), 100µl of phage added to 700µl of LB liquid medium and incubation at 37°C for one hour. After incubation, cells were plated on LB agar medium supplemented with tetracycline (10µg/ml) and ampicillin (100µg/ml) and incubated overnight at 37°C. quality of the gene library was determined on 5 randomly selected transformants. Extraction of plasmids were performed using Qiagen Plasmid purification kit (QIAGEN, Courtaboeuf, France), and 100ng of extracted plasmids were digested with 1U of *EcoRI* (Fermentas, St. Rémy Les Chevreuse, France). Restricted plasmids were separated using a 0.6% agarose gel electrophoresis. Transformants were transferred into 96-well plates containing LB media supplemented with tetracycline (12.5 µg/ml), and were also transferred into plates containing LB supplemented with tetracycline (12.5 µg/ml) and ampiciline (30 µg/ml). These last plates allowed to determine the proportion of empty pCPP47 vectors. Using this method, 8.3% of plasmids were detected as not harboring B&B DNA inserts. Validation of the genomic library was performed using DNA blot screening of 1248 clones with an *ecfB* DNA probe known to be in a single copy in the B&B clone genome (Graindorge et al., 2009). Five transformants gave a positive hybridization confirming the representation of at least three copies of the B&B clone genome in this library.



**Table 2: PCR Primers used in this study**

Name	sequence (5'-3')	base position on <i>Bcen</i> -J2315	annealing T
GI10-LBo	GGCAACGCCGAGCCAAGCC	(Chr1 1222684..1222702)	63°C
GI10-LBi	GGCCTTTCGGATAGTGACGTC	(Chr1 1223111..1223131)	63°C
GI10-RBo	GCCCGACAACAGGCGCTCG	(Chr1 1315844..1315862)	60°C
BCAL1130F	CCCTCGTCCGTATTCATCAC	(Chr1 1234352..1234371)	60°C
BCAL1130R	ATCTGGCAATGACGGAAGAC	(Chr1 1234789..1234808)	60°C
GI11-LBo	GCTGCATCCGCAGCCGG	(Chr1 1402624..1402640)	60°C
GI11-LBi	CAGACCCAAGACGGGACAC	(Chr1 1403183..1403201)	62°C
GI11-RBo	CCAGCGCGATGATGCGCC	(Chr1 1442644..1442661)	62°C
BCAL1298F	GCGCCCCTGCGTCTTCAGC	(Chr1 1412064..1412082)	60°C
BCAL1298R	GCCGACCGCCGTGGGTGCC	(Chr1 1412617..1412635)	60°C
GI20-LBo	CGACGCACCATTTCGTTCCC	(Chr1 2737257..2737275)	62°C
GI20-LBi	GCCGACGCCACTTCCAG	(Chr1 2737929..2737946)	62°C
GI20-RBo	GACGCTCGCGATGGCGC	(Chr1 2860299..2860315)	60°C
BCAL2520F	CGCGGCACCCGCGCCTACC	(Chr1 2785949..2785967)	60°C
BCAL2520R	CGCGCCCTGGACGAGCGGT	(Chr1 2786332..2786350)	60°C
GI26-LBo	CGCTTCTCGGATGCGTGC	(Chr1 3521997..3522015)	62°C
GI26-LBi	CGAGCAATTCCTCGATATCG	(Chr1 3522515..3522534)	60°C
GI26-RBo	CACGGCGGCAAGACGATC	(Chr1 3562365..3562382)	60°C
BCAL3245F	CGGTGGCGCCTCCCTGCGC	(Chr1 3553490..3553508)	60°C
BCAL3245R	GCGACCTGGCCGGGCACC	(Chr1 3553874..3553892)	60°C
GI41-LBo	GTCACCGGTCAGGCATTCCG	(Chr2 2091499..2091517)	62°C
GI41-LBi	GTGAACCGGTACATGCGGC	(Chr2 2091953..2091971)	60°C
GI41-RBo	CGCTTCTCCGTGTGCAGG	(Chr2 2138103..2138121)	62°C
BCAM1890F	CAGCGAGCCGCTAGCGCGGG	(Chr2 2105938..2105957)	60°C
BCAM1890R	GGCGTCGGCCGTGTTCCGCA	(Chr2 2106318..2106337)	60°C
GI57-LBo	GCGCGCAAGCCATTTCGCG	(Chr3 702520..702537)	62°C
GI57-LBi	CCGCGGCTGTCCGGCTGC	(Chr3 703102..703118)	62°C
GI57-RBo	GTTTCATGCAGTAGCGCAGG	(Chr3 795432..795450)	62°C
BCAS0698F	CACGGCCTGCCCCGCCGACC	(Chr3 761733..761751)	60°C
BCAS0698R	GCCGGCCGCGCCGACGATC	(Chr3 762164..762182)	60°C
GIBCESM-LBo	CGTCGTGATCGCGCTGG	(Chr2 236706..236722)	58°C
GIBCESM-LBi	CCGATCGTGCATTGTTGC	(Chr2 237229..237246)	58°C
GIBCESM-RBo	GCGTAGGGATTGAATCCCC	(Chr2 379739..379757)	58°C

B&B clone GI sequencing primers\*:

IS407sqF	CTAGACTCCGGCTGGCC	IS407sqR	CCGCTTTCAGGAGTTGG
5H3_F1	GTGTTGAATCAGCGGCTTG	5H3_F2	CTCGTCGAGATACCGGCC
5H3_F3	CAGCGAGCAACGGGCGCTC	5H5_R1	GTCGAACGCAAACGGCTC
5H3_R2	GAGCCGTTTGGTTTCGAC	5H3_R3	CATCACGCGTGAGCATCTC
1C7_F1	GGTAGCGTGGGCGCTG	1C7_F2	GCCAACGGCCGCGAATC
1C7_F3	GGAGATCAATGCCATGAGCAG	1C7_R1	CTACCTCGGATTCCGCC
1C7_R2	CAAACAGGTTGTGGAAGC	BLACT-F	GCGCATGGCCATACTCG
BLACT-R	GATCGTGCTTCTGGTCCG	NRBE-F	AAGATGCTGGCCTATGTCCGAGTC
NRBE-R	CGCCAAGGGATAGCAAATGAGC		

GI: correspond to the six genomic islands investigated in this study. Primers position is indicated following the J2315 genome numbering system. LBo: primers binding sites are positioned in the Left Border and outside the genomic island. LBi: primers binding sites are positioned in the Left Border and inside the genomic island. RBo: primers binding sites are positioned in the Right Border and outside the genomic island (e.g. Fig. 8).\*: DNA sequencing primers have been used at 55°C by Cogenics™ to sequence the two clones (5H3, 1C7) of the B&B clone genomic library (e.g. Fig. 9). BLACT-F/BLACT-R and NRBE-F/NRBE-R have been used on the B&B genomic DNA to detect the Beta-lactamase and the *nrBE* genes respectively.

### ***2.3. DNA blot analyses***

Genomic DNAs (5 µg) were digested by 1 U/µg *EcoRI* (Fermentas, St. Leon-Rot, Germany) for 16h at 37°C. DNA blots were performed using Genescreen Plus nylon membranes (PerkinElmer Life Science, Boston, USA) and the procedure described by Sambrooke et al. (1989). IS407 PCR products were labelled using the random priming labelling kit (Amersham-Pharmacia, Orsay, France). DNA hybridizations and nylon membrane washes were carried out according to the membrane manufacturer. Membranes were hybridized and washed at 65°C. Autoradiography was performed according to Sambrooke et al. (1989).

### ***2.4. Comparative genomic analyses.***

BLASTN analyses of genomes were performed locally using BLAST-2.2.20 package available on the NCBI web page (<ftp://ftp.ncbi.nih.gov/blast/executables/LATEST>). The Artemis Comparison Tool (Carver et al., 2005), was used to visualize BLASTN and TBLASTX results (Altschul et al., 1990). The full Genome of *B. cenocepacia* J2315 was retrieved from the Sanger Institute web site ([http://www.sanger.ac.uk/Projects/B\\_cenocepacia](http://www.sanger.ac.uk/Projects/B_cenocepacia)). Full genome sequences of *B. cenocepacia* strains AU1054 (CP000378, CP000379, and CP000380) ([http://genome.jgi-psf.org/finished\\_microbes/burca/burca.home.html](http://genome.jgi-psf.org/finished_microbes/burca/burca.home.html)) and HI2424 (CP000458, CP000459, and CP000460) (LiPuma et al., 2002) were retrieved from the Joint Genome Institute web site ([http://genome.jgi-psf.org/mic\\_home.html](http://genome.jgi-psf.org/mic_home.html)). Only DNA sequences of more than 4 kb, absent from the HI2424 and AU1054 genomes, were considered in this analysis, and further investigated in terms of content.

### 3. Results and discussion

Among the clonal complexes described so far among *B. cenocepacia* species, ET12 was found the most closely related to the B&B ST32 complex. One strain of this complex, named J2315, had its genome fully sequenced and could thus be used as a reference to better understand the events which led to the emergence of the B&B clone. Here, investigations were carried out in order to identify large DNA regions, over 4 kb, specific of both complexes, ET12 and B&B, and to identify particular trends or genomic markers that could allow identification of B&B specific DNA regions among a genomic DNA library cloned in *E. coli*.

#### 3.1. DNA regions specific of the J2315 genome

Identification of ET12/J2315 specific DNA regions was performed by comparing its genome with those of the PHDC complex (HI2424 and AU1054), and investigating by Blast analyses and PCR the distribution of these DNA regions among other bacteria, the Bcc and ET12 complex. Using a tripartite ACT comparison, 1.35 Mb of J2315 DNA was found absent from the PHDC genomes. These segments of the J2315 genome were divided into 58 DNA regions of a size going from 4.1 to 142.2 kb, and a G+C content between 48.1 and 71.4 % (Table 3). These DNA regions were named genomic islands (GI) in the text below for simplicity but not all were found to match all the criteria used to define such islands. BLASTN analysis revealed that 33/58 of these DNA regions showed partial or complete similarities with genomes of other organisms: *B. lata* 383 for a large part of them (23/33), *B. multivorans* ATCC17616 (9/33), *B. ambifaria* AMMD (6/33), MC40-6 (8/33) and few of them showed similarities with *B. pseudomallei* genomes (2/33), *B. cenocepacia* MC0-3 (4/33) and *Ralstonia eutropha* H16 (renamed *Cupriavidus necator* by Vandamme and Coenye 2004) (2/33) (Table 3). These 33 GI, even though they are not observed among the PHDC genomes, thus appear to have been selected by *Burkholderia* species long before the emergence of the ET12 complex. Interestingly, *B. lata* was the species sharing the highest number of J2315 GI (not found in the PHDC genomes). *B. lata* is known to be a close phylogenetic neighbor of *B. cenocepacia* (Vanlaere et al., 2008), and this proximity thus appears to have resulted in similarities in the organization of their respective genomic islands architecture. However, most *B. lata* strains have been isolated from soil or water and only three were recovered from human samples (Vanlaere et al., 2009), suggesting that the GI conserved between J2315 and *B. lata* 383 genomes were mainly selected by environmental forces. Identification of 6 genes encoding membrane proteins (BCAL0654, BCAL0657, BCAL1683, BCAM0408, BCAM0420 and BCAM1604), 4 genes encoding permeases (BCAL0850, BCAM1173,

Table 3: DNA regions only detected in J2315 genome when performing comparisons with the PHDC genomes.

#GI	coordinates	size kb	BLASTN results	%GC	IS detected	left border	right border
Chromosome 1							
1	100361..125265	25,00	none	68.5	no	tRNA-Arg	partial tRNA-Arg
2	188011..204429	16,40	none	55.7	IS <i>Bcen9</i> -b	tRNA-Ala	partial tRNA-Ala
3	449978..463082	13,10	none	56.5	IS <i>407</i> -a, -b	none	none
4	492222..497525	5,30	<i>Bam</i> MCO-3 <i>Bla sp383</i>	68	no	none	none
5	712071..721620	9,50	<i>Bla sp383</i> <i>Bcen</i> MCO-3 partial	69.7	no	none	none
6	835703..846355	10,00	<i>Bla sp383</i> partial	65.16	no	none	none
7	920522..926765	6,24	<i>Bla sp383</i> <i>Bam</i> AMMD <i>Bam</i> MC40-6 <i>Bvi</i> G4 <i>Rme</i> CH34 partial <i>Ctai</i> LMG19424 partial <i>Reu</i> JMP134 partial	70.69	no	none	none
8	1032360..1036537	4,18	none	42.82	no	none	tRNA-Met
9	1162576..1170560	7,99	<i>Bam</i> MC40-6 <i>Bmu</i> ATCC17616 <i>Bam</i> AMMD	70.73	no	none	none
10	1222566..1315385	92,80	none	60.3	IS <i>Bcen20</i> -b,-c IS <i>Bcen15</i> -a IS <i>Bcen19</i> -a IS <i>407</i> -c IS <i>Bcen14</i> -a	Group I intron	none
11	1402882..1437088	34,20	none	58.8	IS <i>407</i> -d IS <i>Bcen13</i> -a IS <i>Bcen15</i> -b IS <i>407</i> -e	tRNA-Arg	partial tRNA-Arg
12	1466013..1473381	7,37	<i>Bla sp383</i> <i>Bcen</i> MCO-3 partial	68.49	no	none	none
13	1481512..1503701	22,19	none	66.55	no	none	none
14	1728919..1766584	37,70	<i>Burkholderia</i> phage KS10	62.9	no	BCAL1558p	BCAL1558p
15	1839588..1846574	6,99	<i>Bla sp383</i>	68.11	no	none	none
16	1911719..1923252	11,53	<i>Bam</i> MC40-6 <i>Bam</i> AMMD <i>Bla sp383</i> <i>Bma/Bps</i> partial	71.23	no	none	none
17	2497660..2507470	9,81	<i>Bla sp383</i> <i>Reu</i> H16	68.67	no	none	none
18	2513164..2530552	17,39	<i>Bla sp383</i> partial	62.7	IS <i>1356</i> -a	none	tRNA-Ser at 3kb
19	2656516..2666220	9,71	none	67.76	no	none	none
20	2738065..2859923	121,90	none	61.4	IS <i>Bcen13</i> -b IS <i>Bcen20</i> -d -e IS <i>1356</i> -b IS <i>407</i> -g IS <i>408</i> IS <i>407</i> -h -i -j IS <i>Bcen8</i> -e	none	BCAL2601p
21	2898047..2903025	4,98	<i>Bla sp383</i> <i>Bam</i> AMMD <i>Bam</i> MC40-6 <i>Bps/Bma</i> partial	69.8	no	none	none
22	3241588..3257909	16,30	none	62.8	no	tmRNA	none
23	3286423..3291776	5,35	<i>Bla sp383</i> <i>Bam</i> AMMD <i>Bam</i> MC40-6 partial	68.66	no	tRNA-Thr	none
24	3366796..3381064	14,27	none	59.4	no	tRNA-Gly	partial tRNA-Gly
25	3411938..3424737	12,80	none	59.01	IS <i>Bcen20</i> -f	none	none
26	3522290..3562993	40,70	none	55.54	IS <i>407</i> -k IS <i>Bcen19</i> -b IS <i>Bcen18</i> IS <i>Bcen17</i> IS <i>Bcen16</i>	none	none
Chromosome 2							
27	70049..84707	14,66	none	55.41	no	partial IS <i>Bcen20</i>	none
28	168377..182238	13,86	none	59.49	no	none	none

Table 3: continued.

29*	236925..379167	142,24	<i>Bcen</i> MCO-3 partial	65	<i>ISBcen14</i>	none	none
30	456383..481899	25,52	<i>Bla sp383</i>	67.8	no	none	none
31	842504..854406	11,90	<i>Bla sp383</i>	67.0	no	none	IS element
32	862760..866992	4,20	<i>Bmu</i> ATCC17616	48.1	no	none	none
33	1140183..1186975	46,80	none	63.2	no	tRNA-SeC	partial tRNA-SeC
34	1271236..1298260	27,00	<i>Bla sp383</i> <i>Bmu</i> ATCC17616	68.58	no	none	none
35	1444014..1450309	6,30	<i>Bla sp383</i>	68.63	no	none	none
36	1471324..1488524	15,00	<i>Bmu</i> ATCC17616 partial	60.93	no	none	none
37	1713161..1720512	7,30	<i>Bla sp383</i> partial	67	no	none	none
38	1787650..1836115	48,50	<i>Bla sp383</i>	68.79	no	none	none
39	1890941..1895098	4,10	none	65.78	no	none	none
40	2011726..2019173	7,50	none	66.1	no	none	none
41	2091701..2137922	46,20	none (prophage)	54.3	<i>ISBcen20-h</i> <i>IS407-l</i> <i>ISBcen20-i</i>	BCAM1874p	BCAM1874p
42	2249572..2256842	7,20	<i>Bla sp383</i> <i>Bmu</i> ATCC17616	68.46	no	none	none
43	2260643..2275008	14,30	none (TTSS)	71.37	no	none	none
44	2332339..2395714	63,30	none	63.04	<i>ISBcen13-e</i>	none	none
45	2540911..2556578	15,60	<i>Bla sp383</i> <i>Bmu</i> ATCC17616 partial	66.36	no	none	none
46	2877034..2891133	20,20	none	65.5	no	none	none
47	3114602..3126850	12,20	<i>Bcen</i> BC7 <i>Bam</i> MC40-6 partial	64.97	no	none	none
48	3184563..3191106	6,50	<i>Bmu</i> ATCC17616 <i>Bam</i> MC40-6 <i>Bla sp383</i> <i>Bxe</i> LB400 <i>Bphy</i> STM815	70.39	no	none	none
Chromosome 3							
49	39761..57020	17,20	<i>Bla sp383</i> partial	70.5	no	none	none
50	75872..89859	14,00	<i>Bmu</i> ATCC17616 partial	66.3	<i>ISBcen8-j</i>	none	none
51	101812..107517	5,70	<i>Bla sp383</i> <i>Bmu</i> ATCC17616	67.89	no	none	none
52	114943..122014	7,00	<i>Bla sp383</i> partial	67.41	no	none	none
53	370817..385704	14,80	<i>Bcen</i> AU154 <i>Bam</i> AMMD <i>Bam</i> MC40-6 <i>Bglu</i> BGR1 partial	69.06	no	none	none
54	472087..478502	6,40	<i>Bam</i> MC40-6 <i>Bam</i> AMMD partial	63.47	no	none	none
55	524936..535707	10,70	none	69.68	no	none	none
56	573207..609954	36,70	none	62.8	no	none	none
57	703084..794700	91,60	none	61.33	<i>ISBcen6-a</i> <i>ISBcen4-a</i> <i>ISBcen6-b</i> <i>ISBcen4-e</i> <i>ISBcen13-b</i> <i>ISBcen2-a</i> <i>IS407-m</i>	none	none
58	839553..856005	16,40	none	69.9	no	none	none

Coordinates indicated positions of GI among the J2315 genome. BLASTN results indicate strains showing DNA similarities with the J2315 genome for the full length of the GI tested, “*partial*” results indicated strains harboring similarities with the J2315 genome for a partial length of the GI tested. *Bcen* = *B. cenocepacia*; *Bphy* = *B. phymatum*; *Bglu* = *B. glumae*; *Bmu* = *B. multivorans*; *Bvi* = *B. vietnamiensis*; *Bla* = *B. lata*; *Bam* = *B. ambifaria*; *Bps* = *B. pseudomallei*; *Bma* = *B. mallei*; *Bxe* = *B. xenovorans*; *Rme* = *Ralstonia metallidurans*; *Reu* = *Ralstonia eutropha*; *Ctai* = *Cupriavidus taiwanensis*. \*GI29 contained the *cenocepacia* island (BCESM) and corresponds to the GIBCESM in the text. The 14 Holden et al., 2009 BcenGI correspond respectively to GI1, GI2, GI3, GI8, GI10, GI11, GI14, GI20, GI22, GI24, GI29, GI33, GI41 and GI56.

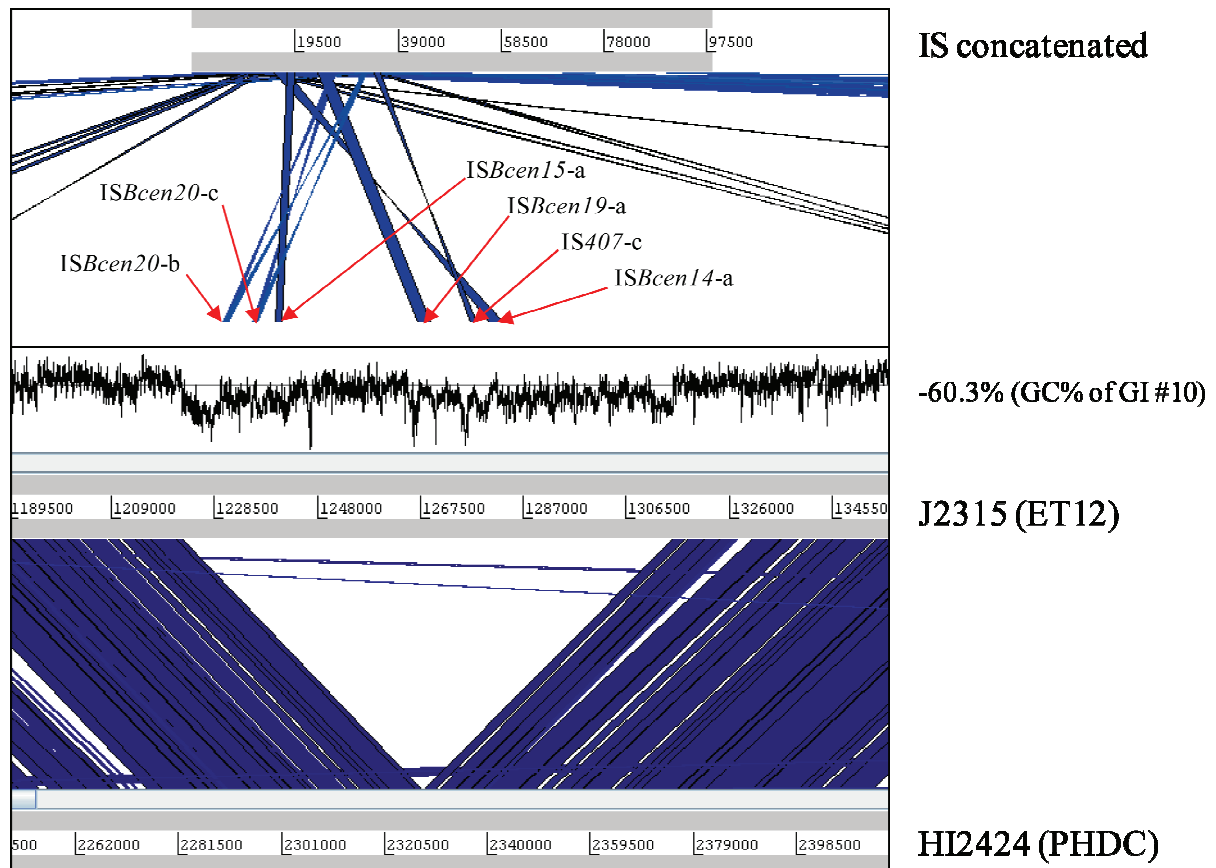
BCAM1180 and BCAM2271) and 2 genes encoding porin proteins (BCAL1734 and BCAM0407) within these conserved GI support an implication in the interaction with the extracellular environment. These GI could thus be metabolism islands rather than PAI. A similar observation was made from the comparative analyses of GI conserved with *B. ambifaria* strains AMMD and MC40-6 genomes. These GI harbored genes encoding permeases, porins and membrane proteins (Table 3). *B. ambifaria* was generally found in outdoor environments where it is often associated with plant roots (Dalmastrri et al., 2007; Jacobs et al., 2008). The selective forces which led to the deletion of these 33 GI among the PHDC lineage are unclear since strains of this clone have been detected in the environment. GI are known to condense during evolution of organisms (Hacker and Kaper 2000); thus, a part of these GI could have been lost or rearranged among these strains. In fact, some sequences of these GI could be detected in the PHDC genomes like GI53.

The remaining 25 GI of the J2315 genome showed no similarities with the databases suggesting a recent acquisition (Table 3). Several of these GI showed a different bias in G+C content than the one of the core genome, showed insertion sequences (IS) and a tRNA gene or integrases CDS (Hacker and Carniel 2001). Interestingly, 51% of the IS elements detected among the J2315 genome was found inserted within these 25 GI (42/82). No IS element was found among the previously described conserved GI (section above and Table 3). IS elements seemed to have preferentially colonized newly acquired DNA sequences, and 6 of these GI (GI10, GI11, GI20, GI26, GI41 and GI57) harbored 83% of the total number of IS located on GI (35/42). Only four of these GI have been described by Holden et al. (2009) (GI10=BcenGI5, GI11=BcenGI6, GI20=BcenGI8 and GI41=BcenGI13). In the next section, GI showing high number of IS elements are presented.

### ***3.2. B. cenocepacia J2315 GI harboring high number of IS elements***

GI10 has a size of 92.8 kb, a G+C content of 60%, and was described as a miscellaneous island by Holden et al., 2009, sharing some genes involved in metabolism and resistance to fusaric acid. ACT visualization of this GI and position of IS elements are described in Fig. 1 and Table 3. This GI harbored 6 IS elements (*ISBcen20-b,-c* from *IS5* family, *ISBcen15-a* and *IS407-c* from *IS3* family, *ISBcen14-a* and *ISBcen19-a* from *IS66* family). Three of these disrupted genes (see Chapter 3). No particular function could be attributed to these disrupted genes. In fact, a large part of the genes found on this GI appeared to encode proteins of unknown function, and only the genes implicated in fusaric acid

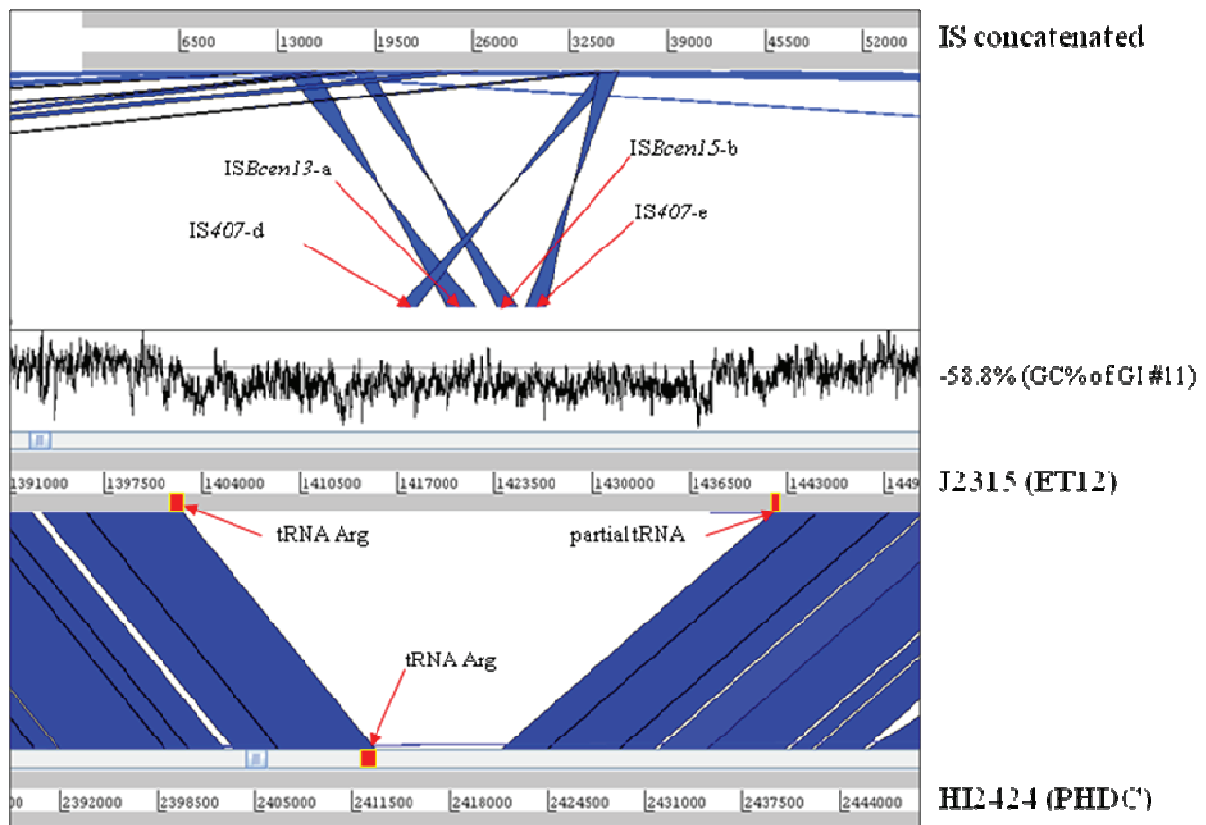
resistance could be linked to virulence. Mechanisms involved in integration of this GI are unknown but an integrase gene was detected on the right border of this GI.



**Fig. 1: The 92.8 kb Genomic Island 10 of *B. cenocepacia* J2315 and its high number of IS elements.** Tripartite comparison of *B. cenocepacia* J2315, HI2424 genomes and IS concatenated sequences using the ACT graphic tool. *B. cenocepacia* J2315 genome sequence was compared to a concatenated set of IS elements and then the J2315 DNA region was compared with the one of *B. cenocepacia* strain HI2424. In blue are conserved DNA sequences showing 93% to 100% identity. IS elements are indicated on the J2315 genome following the nomenclature of chapter 3. G+C content of the GI is given (knowing that J2315 core genome G+C content is 68%). Six IS elements are located in this GI. The lack of synteny with the HI2424 genome indicates an absence of this GI in HI2424. Coordinates of GI among J2315 chromosome 1 is 1222566..1315385.

GI11 was identified on chromosome 1, showed a length of 34.2 kb and a G+C content of 58.8%. It was classified as a miscellaneous island by Holden et al. (2009). ACT visualization of this GI and position of IS elements in this GI are shown in Fig. 2 and Table 3. No particular function could be inferred from GI11 DNA sequences, beside a phage integrase gene (BCAL1317) on its right border, suggesting a possible phage origin. A tRNA-Arg gene was detected on the left border of GI11 suggesting a potential implication in the integration event. This integration event appeared to have led to the duplication of the 3' end of this tRNA-Arg gene on the right hand of the GI. Interestingly, this tRNA-Arg gene was also detected in the matching DNA region of the PHDC genomes (Fig. 2). However, the GI found in the PHDC genomes was different and of a shorter size (8.5 kb). This PHDC GI has 58%

G+C. Analysis of this PHDC DNA sequence also revealed genes encoding putative phage



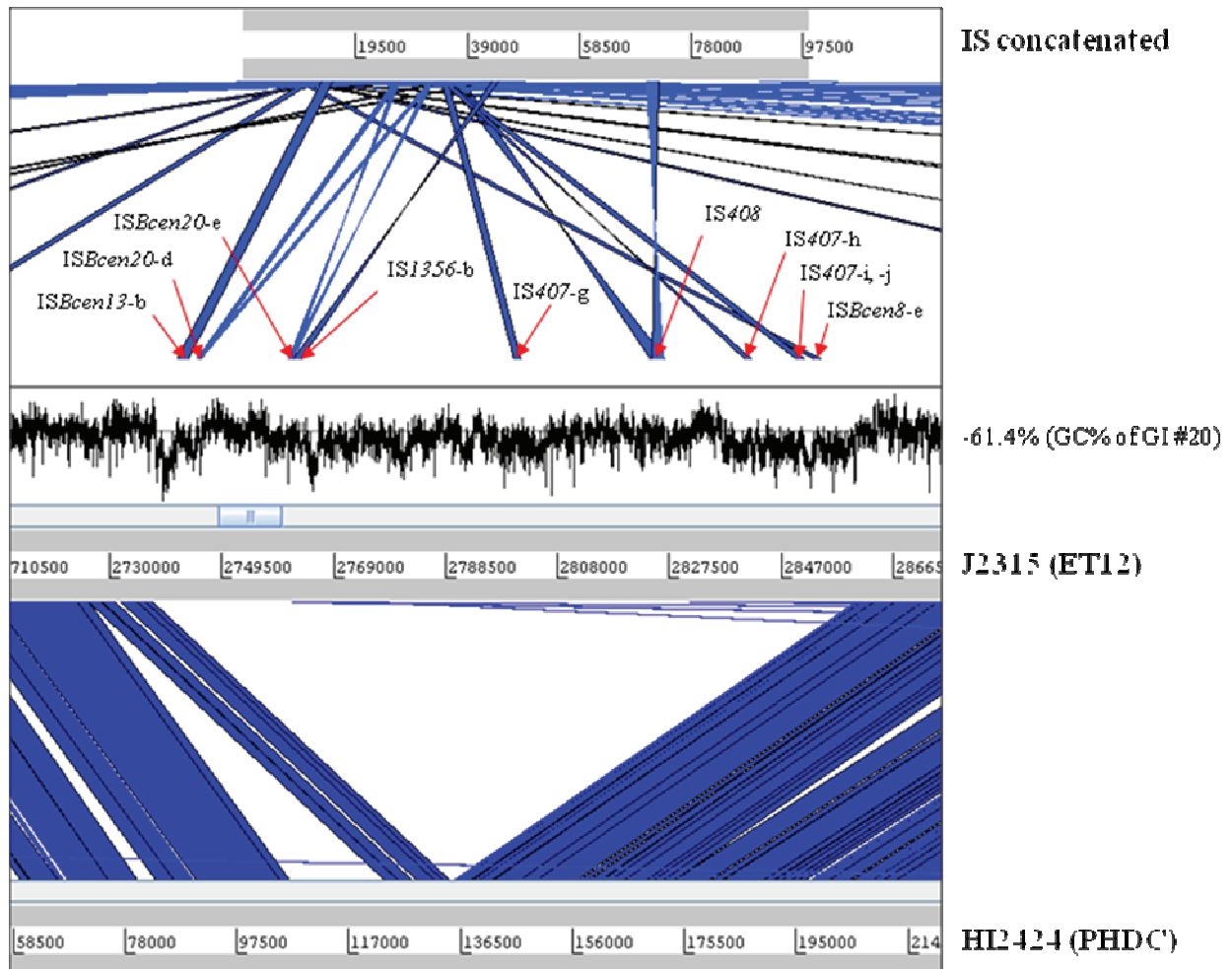
**Fig. 2: The 34.2 kb Genomic Island 11 of *B. cenocepacia* J2315 and its high number of IS elements.**

Tripartite comparison of *B. cenocepacia* J2315, HI2424 genomes and IS concatenated sequences using the ACT graphic tool. *B. cenocepacia* J2315 genome sequence was compared to a concatenated set of IS elements and then the J2315 DNA region was compared with the one of *B. cenocepacia* strain HI2424. In blue are conserved DNA sequences showing 93% to 100% identity. IS elements are indicated on the J2315 genome following the nomenclature of chapter 3. G+C content of the GI is given (knowing that J2315 core genome G+C content is 68%). Six IS elements are located in this GI. The lack of synteny with the HI2424 genome indicates an absence of this GI in HI2424. Coordinates of hotspot among J2315 chromosome 1 is 1402882..1437088. tRNA sequence is indicated by red boxes

proteins (Bcen\_0845: phage integrase; Bcen\_0842: Phage-plasmid primase P4-like; Bcen\_0839: phage terminase small subunit). Detection of these genes strongly suggests a phage integration mediated by the tRNA-Arg gene, among the analysed PHDC genomes. No other copies of this phage were observed among the *Burkholderia* strains sequenced. Interestingly, the deduced PHDC integrase showed 68% similarities with the AA sequence of the J2315 integrase (BCAL1317). The location of two distinct GI in a same tRNA-Arg gene suggests this gene to play a major role in GI or phage integration events. Four IS elements were detected within this GI (IS407-d, -e and ISBcen15-b from IS3 family, ISBcen13-a from IS21 family) and three of them disrupted genes (Table 3, and see chapter 3). Interestingly IS407-d and ISBcen15-b disrupted the same CDS encoding a putative helicase (BCAL1308). This suggests a strong counter-selection of helicase genes. This hypothesis is supported by the observation of another helicase gene (BCAL2480), found on the J2315 GI20 (described



below), and also disrupted by an IS. No IS element was detected into the phage DNA remnant found among PHDC genomes, but its shorter size might have been due to major rearrangements and deletions. These results suggest a preferential colonization of newly integrated DNA by IS elements. These IS would benefit of these new DNA segments by increasing their number and the bacterial cell would benefit of these IS inactivations by preventing expression of unnecessary or deleterious genes.

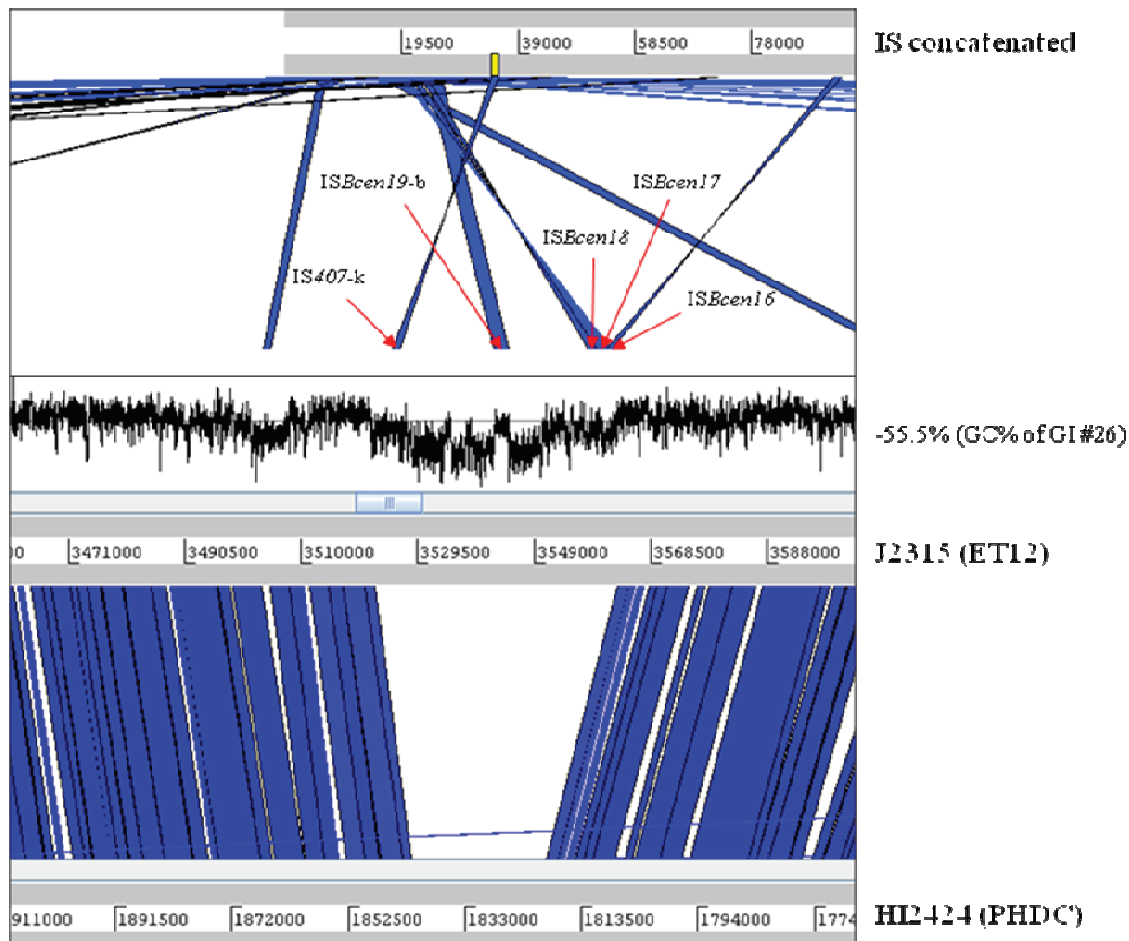


**Fig. 3: The 121.9 kb Genomic Island 20 of *B. cenocepacia* J2315 and its high number of IS elements.**

Tripartite comparison of *B. cenocepacia* J2315, HI2424 genomes and IS concatenated sequences using the ACT graphic tool. *B. cenocepacia* J2315 genome sequence was compared to a concatenated set of IS elements and then the J2315 DNA region was compared with the one of *B. cenocepacia* strain HI2424. In blue are conserved DNA sequences showing 93% to 100% identity. IS elements are indicated on the J2315 genome following the nomenclature of chapter 3. G+C content of the GI is given (knowing that J2315 core genome G+C content is 68%). Six IS elements are located in this GI. The lack of synteny with the HI2424 genome indicates an absence of this GI in HI2424. Coordinates of hotspot among J2315 chromosome 1 is 2738065..2859923.

GI20 was detected on chromosome 1 and harbors a size of 121.9 kb, and a G+C content of 61.4% (Fig. 3, Table 3). GI20 integration occurred in the BCAL2601 gene encoding a putative ABC transporter. Other CDS of this GI appeared involved in iron transport, O-antigen acetylation and other miscellaneous metabolisms. This GI showed the

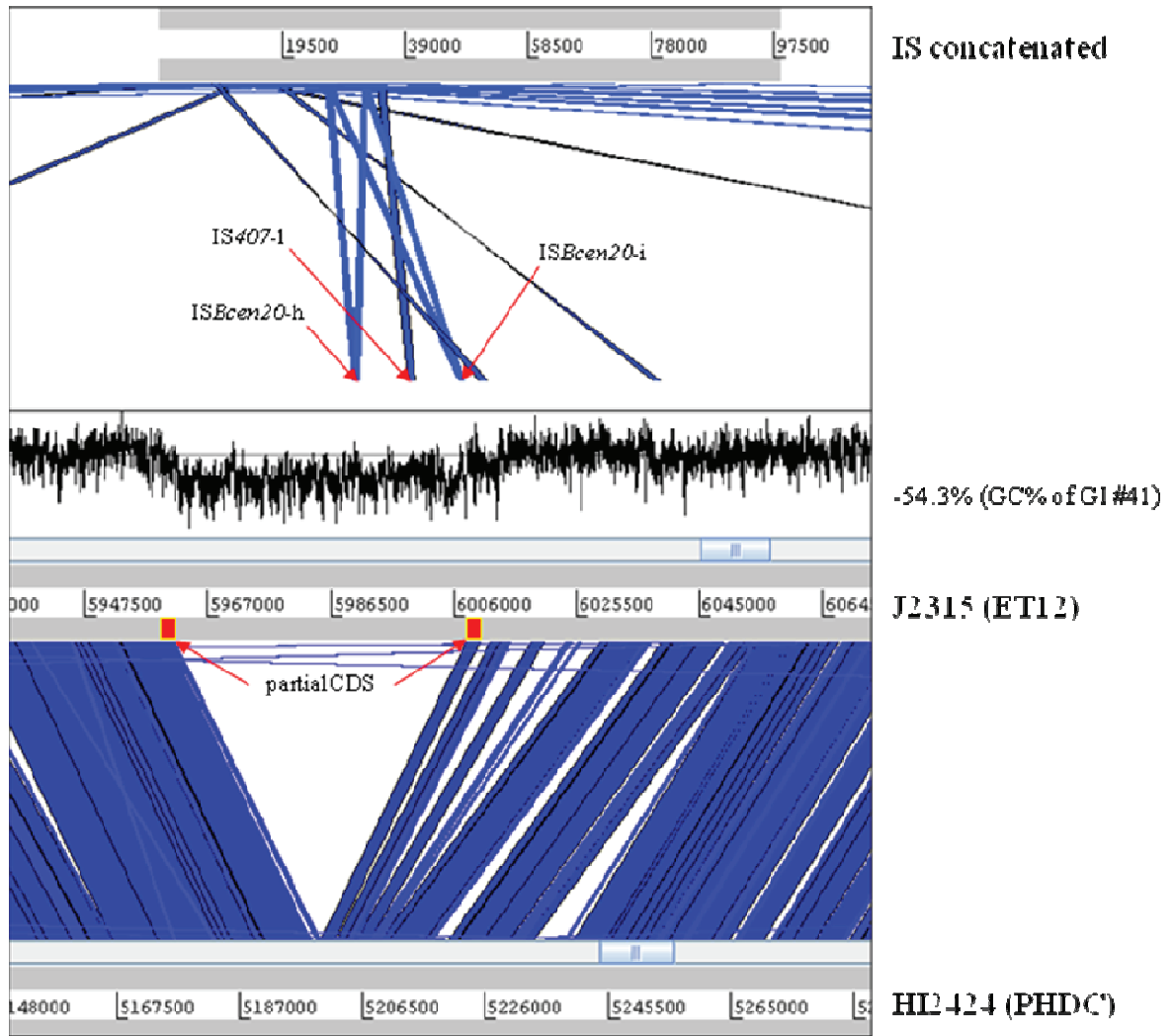
highest number of IS insertions with ten different IS elements identified belonging to the IS3, IS5, IS21, IS110 and IS256 families. As indicated above, a gene encoding a putative helicase (BCAL2080) was disrupted by the *ISBcen13-b*. Three other IS elements were found disrupting genes encoding respectively for another IS element (*ISBcen20-e*), a methionin aminopeptidase and a hypothetical protein. These data confirmed the high concentration of IS elements among GI, and also confirmed data obtained in the previous chapter which suggested an "attraction" of IS for the DNA of other IS elements.



**Fig. 4: The 40.7 kb Genomic Island 26 of *B. cenocepacia* J2315 and its high number of IS elements.** Tripartite comparison of *B. cenocepacia* J2315, HI2424 genomes and IS concatenated sequences using the ACT graphic tool. *B. cenocepacia* J2315 genome sequence was compared to a concatenated set of IS elements and then the J2315 DNA region was compared with the one of *B. cenocepacia* strain HI2424. In blue are conserved DNA sequences showing 93% to 100% identity. IS elements are indicated on the J2315 genome following the nomenclature of chapter 3. G+C content of the GI is given (knowing that J2315 core genome G+C content is 68%). Six IS elements are located in this GI. The lack of synteny with the HI2424 genome indicates an absence of this GI in HI2424. Coordinates of hotspot among J2315 chromosome 1 are 3522290..3562993.

GI26 was not previously described by Holden et al. (2009). This GI harbors a length of 40.7 kb with a G+C content of 55.54%. Five IS elements were detected in this GI (Fig. 4, Table 3). No particular features have been detected on its borders, hampering the understanding of the mechanisms involved in its integration. BLASTX analysis revealed a

cluster of genes related to capsule biosynthesis (BCAL3221 to BCAL3255) with homologies to the *B. mallei* and *B. pseudomallei* sequences. Two GI sharing homologies with GI26 were previously described among *B. mallei* and also several strains of the Bcc using subtractive hybridization (DeShazer et al., 2001; Parsons et al., 2003). Interestingly, its distribution among the IIIA *recA* RFLP group of *B. cenocepacia* was found heterogeneous; ET12 lineage strains showed a conservation of the entire cluster but only a partial cluster was observed among strains outside the ET12 lineage (Parsons et al., 2003). Moreover, all strains belonging to the IIIB RFLP group were negative for this cluster confirming absence of this sequence among PHDC strains which belongs to the IIIB RFLP subgroup. Deeper investigations revealed an insertion of IS407-k within the *wcbO* gene (BCAL3223) of this cluster among the J2315 genome. This insertion seemed to be restricted to the J2315 genome (Parsons et al., 2003) but analysis performed in the previous chapter demonstrated the presence of this insertion among the B&B clone strain, suggesting a wider distribution of this phenomenon. This IS inactivation apparently led to a loss of the ability to synthesize a capsule in J2315. However, no capsule or mucoid colonies were detected among the other ET12 strains tested (not having the IS407 insertion) by Parsons et al. (2003), although RT-PCR analysis showed expression of these genes. Thus, the real implication of this GI26 in virulence has not been demonstrated, but genes present could be involved in different processes related or not to virulence (EPS, capsule or biofilms formation).

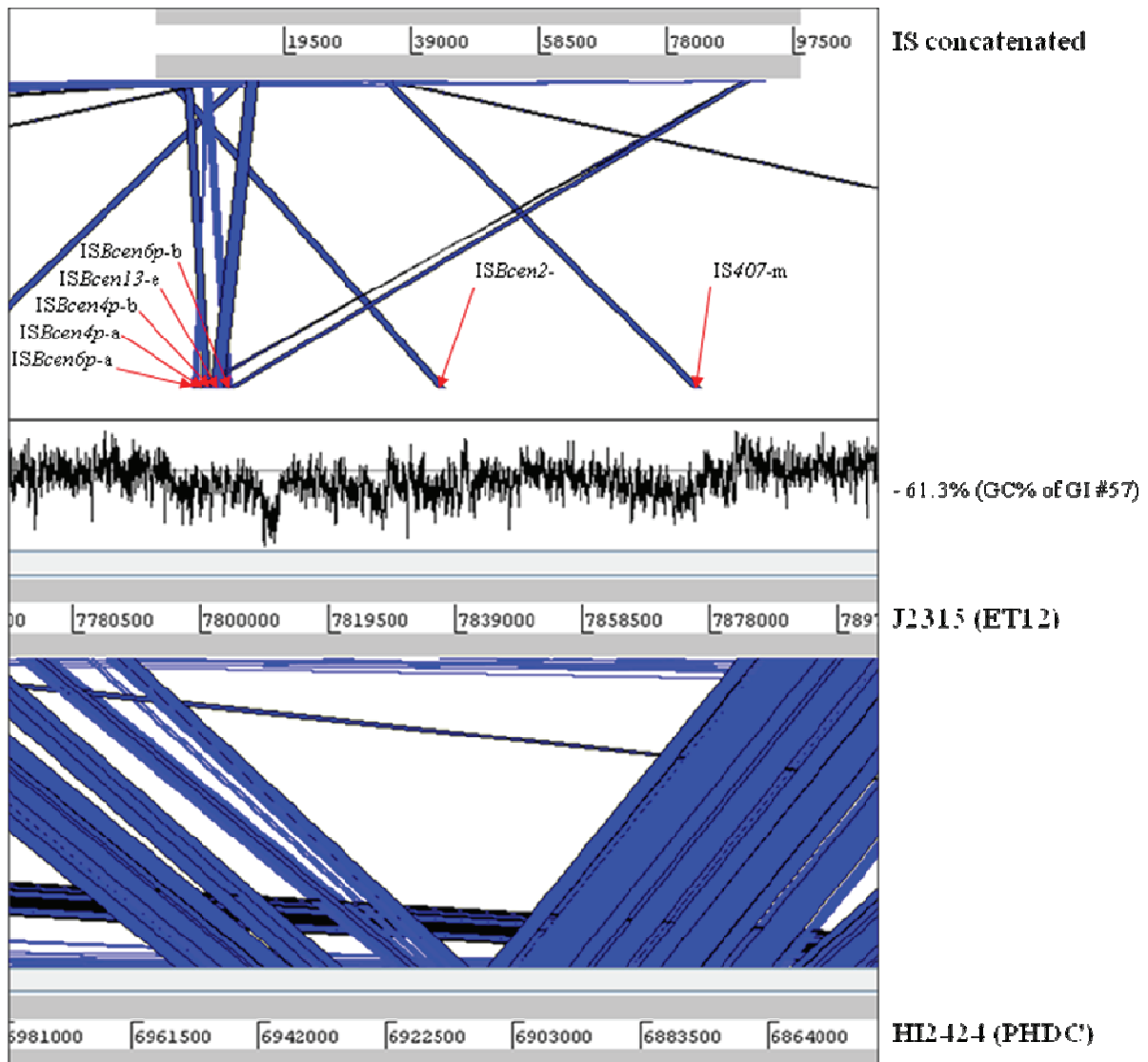


**Fig. 5: The 46.2 kb Genomic Island 41 of *B. cenocepacia* J2315 and its high number of IS elements.**

Tripartite comparison of *B. cenocepacia* J2315, HI2424 genomes and IS concatenated sequences using the ACT graphic tool. *B. cenocepacia* J2315 genome sequence was compared to a concatenated set of IS elements and then the J2315 DNA region was compared with the one of *B. cenocepacia* strain HI2424. In blue are conserved DNA sequences showing 93% to 100% identity. IS elements are indicated on the J2315 genome following the nomenclature of chapter 3. G+C content of the GI is given (knowing that J2315 core genome G+C content is 68%). Six IS elements are located in this GI. The lack of synteny with the HI2424 genome indicates an absence of this GI in HI2424. Coordinates of GI among J2315 chromosome 2 is 2091701..2137922. GI integration occurred within a CDS sequence indicated by red boxes.

GI41 was detected on chromosome 2; it harbors a G+C content of 54.3% and a length of 46.2 kb (Table 3, Fig. 5). This GI was also detected by Holden et al. (2009) and was defined as being a prophage. Integration of this GI occurred among a CDS which encodes a putative GTP cyclohydrolase of type II (BCAM1874). Most CDS identified on this GI showed no similarities with genes already described in the database. However, three CDS were identified as IS elements and one CDS was identified as encoding a putative phage integrase similar to a phage integrase (BCAM1923) encoded by *Rhodopseudomonas palustris*. Interestingly, the *ISBcen20-i* insertion occurred within this phage integrase gene

leading to its inactivation, and IS407-1 was found inserted in a CDS encoding an hypothetical phage protein (BCAM1914). These two disruptions suggest an inability for this phage to undertake its lytic cycle. It can thus be considered as a GI undergoing rearrangement.



**Fig. 6: The 91.6 kb Genomic Island 57 of *B. cenocepacia* J2315 and its high number of IS elements.** Tripartite comparison of *B. cenocepacia* J2315, HI2424 genomes and IS concatenated sequences using the ACT graphic tool. *B. cenocepacia* J2315 genome sequence was compared to a concatenated set of IS elements and then the J2315 DNA region was compared with the one of *B. cenocepacia* strain HI2424. In blue are conserved DNA sequences showing 93% to 100% identity. IS elements are indicated on the J2315 genome following the nomenclature of chapter 3. G+C content of the GI is given (knowing that J2315 core genome G+C content is 68%). Six IS elements are located in this GI. The lack of synteny with the HI2424 genome indicates an absence of this GI in HI2424. Coordinates of GI among J2315 chromosome 3 is 703084..794700.

The last GI investigated, GI57, shows a length of 91.60 kb, a G+C content of 61.33%. Seven IS elements were detected with most being on the left hand of this GI (Table 3, Fig. 6, chapter 3). Two additional CDS were identified to be potential IS elements (BCAS0658-0660; BCAS0680) but homology searches could not identify their family of origin. No other copies of these IS were detected in the J2315 genome. Over the 82 CDS encoded by this GI, 26

encode hypothetical proteins, 6 were identified as pseudo-genes presenting frameshifts within their CDS, 4 encoded putative membrane proteins, and the other CDS were related to miscellaneous metabolism. However, one putative protein encoded by these CDS showed similarities with a metallo-beta-lactamase super family protein (BCAS0689) potentially implicated in the degradation of antibiotics with a  $\beta$ -lactam core (Payne 1993). Furthermore, a putative partial Group II intron belonging to Class E (BCAS0657) was identified close to the zone showing multiple IS elements. This intron showed 68% identity with one detected on the plasmid of J2315 suggesting a possible dissemination by this plasmid as it has been described for a large part of bacterial type II intron (Toro et al., 2002; Tourasse and Kolsto 2008). The role of this intron is unclear but relationship of group II introns in eubacteria with mobile elements suggests a possible role in events such as resistance “cassette” formation and/or GI evolution (Centron and Roy 2002). 3 CDS related to a two-component system (BCAS0707 to BCAS0709) were also identified. AraC and two components-system regulators have been described as being the two major classes of regulators encoded by PAI (Hacker and Kaper 2000). All these results led us to define the GI57 as a potential PAI.

### ***3.3. B&B clone screening of selected J2315 GI***

The above six GI are likely to have different origins. However, features like IS elements insertions, bias in G+C content, presence of an integrase gene, the proximity of a tRNA gene, phage protein genes, are consistent with the definition of them as genomic islands. The real impact of these GI is still unknown but their acquisition could have affected traits such as antibiotic resistance, fitness, and others. In chapter I, we detected a GI among the B&B clone: the one named BCESM (Graindorge et al., 2009) and which has been reported to play a role in the virulence of *B. cenocepacia* strains (Baldwin et al., 2004). Tripartite analysis using the PHDC genomes showed that the BCESM island was inserted among a longer GI, GI29, in the J2315 genome (Fig.7, Table 3). Furthermore, a large DNA sequence absent in J2315 was found to have replaced the BCESM GI in the HI2424 genome. This suggests a hot-spot of recombination within this region.

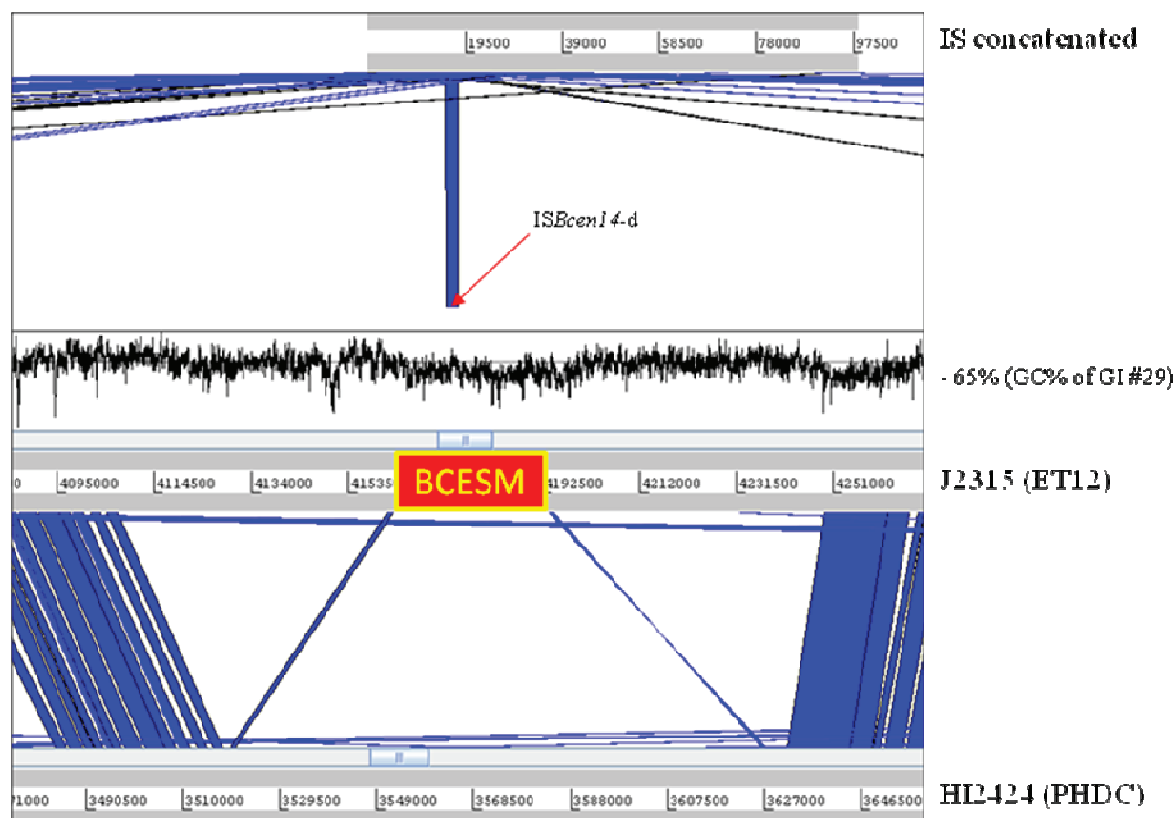
Table 4: PCR screenings of GI identified in *B. cenocepacia* J2315 genome.

strains:		J2315		LMG12614		Bpoe732	
		PCR	sequencing	PCR	sequencing	PCR	sequencing
GI10	abs	0	nd	0	nd	+	ok
	pst	+	ok	+	ok	0	nd
	BCAL1130	+	ok	+	nd	0	nd
GI 11	abs	0	nd	0	nd	+	ok
	pst	+	ok	+	ok	0	nd
	BCAL1298	+	ok	+	nd	0	nd
GI 20	abs	0	nd	0	nd	0	nd
	pst	+	ok	+	ok	0	nd
	BCAL2520	+	ok	+	nd	0	nd
GI 26	abs	0	nd	0	nd	0	nd
	pst	+	ok	+	ok	+	ok
	BCAL3245	+	ok	+	nd	+	nd
GI 41	abs	0	nd	0	nd	0	nd
	pst	+	ok	+	ok	0	nd
	BCAM1890	+	ok	+	nd	0	nd
GI 57	abs	0	nd	0	nd	0	nd
	pst	+	ok	+	ok	+	ok
	BCAS0698	+	ok	+	ok	+	ok
GI 29*	abs	0	nd	0	nd	0	nd
	pst	+	ok	+	ok	+	ok

+: PCR amplification with the size expected. ok: PCR results confirmed by sequencing and ACT analysis (e.g. Fig. 8). 0: no amplification. nd: not determined. \*: GI 29 correspond to the GIBCESM described in the text.

In the previous chapters, a close proximity between J2315 genome and the ones of the B&B clone was observed. The *B. cenocepacia* sequenced genomes, notably the one of J2315, were used to infer some of the events which might have led to the emergence of this clone. Here, these investigations were completed by investigating the distribution of the six GI described above. PCR screenings described in Fig. 8 were developed. Three different primers were designed per GI to test for their presence/absence (Fig. 8a, Table 2). Size of the PCR products were estimated and used to establish the presence of a GI (Fig. 8b). A cloning/sequencing of these PCR fragments was also performed, and the DNA sequences were analysed by BLAST and compared with those of the J2315 genome. Results have been visualized using the ACT® software (Fig. 8c). Using this approach, out of the 7 GI (including the BCESM GI) tested, all were detected in the LMG12614 strain (ET12 lineage) but only 3 were detected the B&B clone strain bpoe717: GI26, GI57 and GI29 (Table 4). GI10 and GI11 were absent from the B&B genome, and no PCR products could be obtained for GI20 and GI41 (Table 4). These last failures in getting the expected PCR products could be explained by an absence of the integration sites through a large deletion going over the region of these GI. Presence of GI29 in the B&B clone confirms the results obtained in chapter I concerning its detection in this clone. This GI was previously described as playing part in virulence (Clode et al., 2000; Baldwin et al., 2004; Golini et al., 2006). GI26 and GI57 were shared by the B&B clone and the J2315 strain and were respectively inferred as playing a role in capsule synthesis and in virulence. The entire GI26 was previously described to be restricted to ET12 (Parsons et al., 2003). In this study, presence of the left border of this GI was demonstrated among the B&B clone using a PCR-screening, and the BCAL3245 gene encoding the *wcbA* gene was also amplified and shown to harbour IS407-k in the B&B clone (as observed in chapter 3). These data strongly suggest a conservation of the entire cluster in the B&B clone. GI57 was first described as a miscellaneous GI; nevertheless, several CDS could be playing part in antibiotic resistance, and a two components system was inferred, suggesting a possible role in B&B pathogenicity. All these results demonstrated an heterogeneous distribution of these seven GI within the B&B clone genome. Three of them (GI16, GI29 and GI57) seemed to have colonized the J2315 and B&B genomes prior their divergence; the last four GI have either colonized J2315 specifically or been lost by the B&B clone. This suggests many genomic rearrangements during evolution of these two clones, confirming the important plasticity and versatility of *Burkholderia* strains.





**Fig. 7: The 142.2 kb Genomic Island 29 harbouring the *Cenocepacia* island (BCESM).**

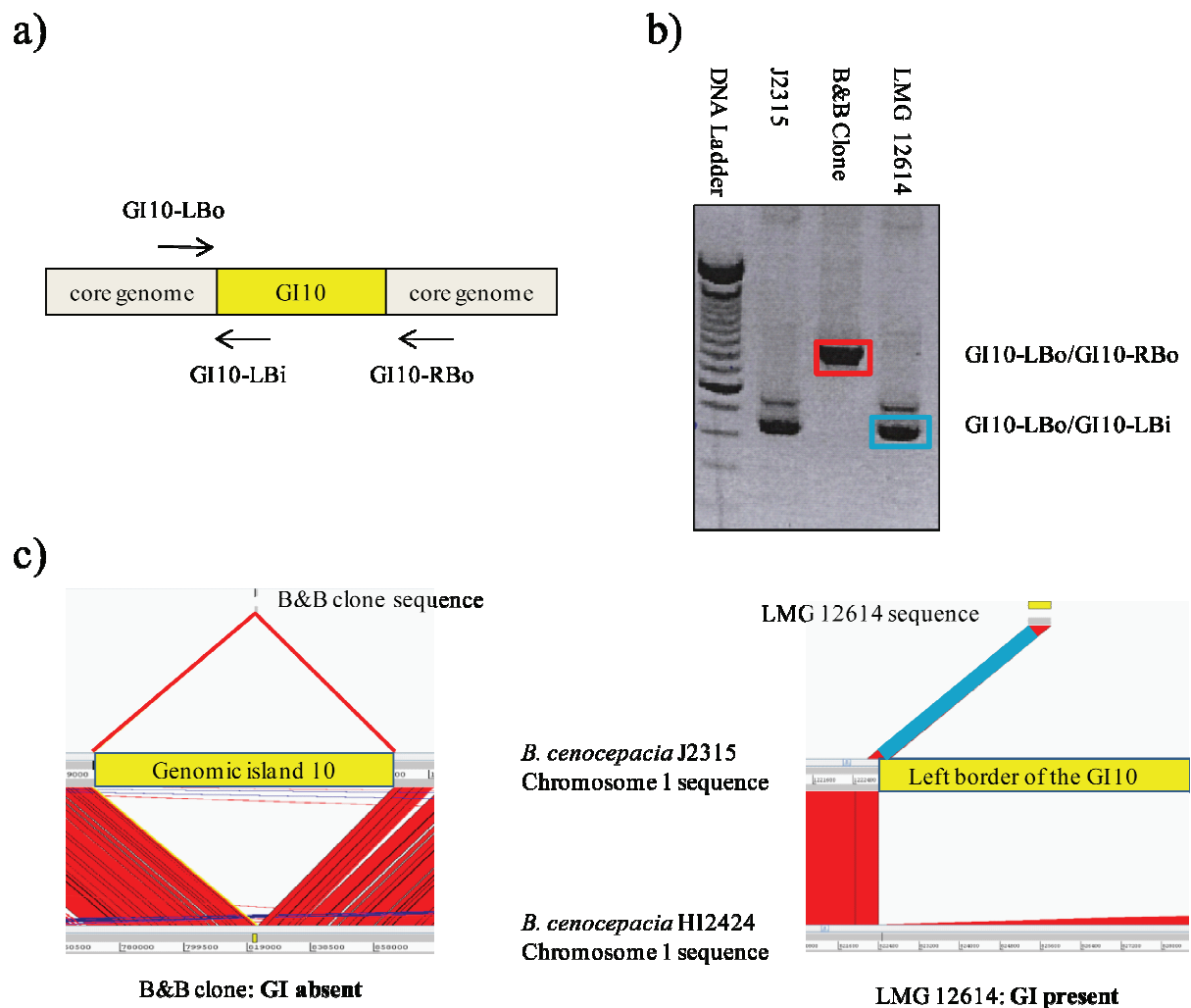
Tripartite comparison of *B. cenocepacia* J2315, HI2424 genomes and IS concatenated sequences using the ACT graphic tool. *B. cenocepacia* J2315 genome sequence was compared to a concatenated set of IS elements and then the J2315 DNA region was compared with the one of *B. cenocepacia* strain HI2424. In blue are conserved DNA sequences showing 93% to 100% identity. IS elements are indicated on the J2315 genome following the nomenclature of chapter 3. G+C content of the GI is given (knowing that J2315 core genome G+C content is 68%). Six IS elements are located in this GI. The lack of synteny with the HI2424 genome indicates an absence of this GI in HI2424. Coordinates of this region among J2315 chromosome 2 is 236925..379167. The BCESM (44.1 kb, 62%GC, coordinates is 290291..334378) present in this region is indicated.

### 3.4. Detection of B&B specific GI

The analyses performed of the J2315 genome allowed detection of 58 DNA fragments absent in the PHDC genomes. 25 of these DNA sequences showed no homologies with other bacteria in the databases and shared several features leading us to consider these sequences as GI. Interestingly, these GIs harbored more than half of the IS listed in J2315, suggesting a tropism of IS elements for GI. Although relations between IS and GI are unclear, IS elements seemed to have preferentially colonize newly acquired DNA such as GI. In fact, 80% of GI detected among five *B. pseudomallei* strains contain IS elements, with the IS407 and ISBp1 of the IS3 family being the most numerous (Tuanyok et al., 2008). It is noteworthy that IS407 was identified to be an active IS element in diverse Bcc species (Wood et al., 1991; Miche et al., 2001; Ohtsubo et al., 2005 and chapter 3). In this study, this IS was detected among all the

six GI investigated and among GI3 (Table 3). This particular distribution of IS407 was used to detect GI specific of the B&B clone.

A screening of a B&B clone genomic DNA library was performed using IS407 as a DNA probe. Two positive clones within the library were selected and partially sequenced (clone 5H3=GI-a and 1C7=GI-b). The two sequences obtained were annotated and results are described in Fig. 9 (sequences of these GI are available on the Microbial Ecology laboratory website: <http://ecomicro.univ-lyon1.fr/poe/BBcloneGI.zip>). These sequences were analysed by BLAST and compared with J2315 genome. However, no significant hits were detected, indicating their absence in the J2315 genome. The G+C content of GI-a and GI-b were 63.57 and 61.74% respectively, supporting the definition of these two sequences as being part of a GI. BLASTX analysis of GI-a and GI-b sequences showed similarities respectively with DNA regions of *R. metallidurans* CH34 (recently renamed *Cupriavidus metallidurans* by Vandamme and Coenye 2004) and *P. aeruginosa* PA14 (Fig. 9) genomes. The B&B DNA sequences encoded putative proteins showing strong similarities with *R. metallidurans* CH34; a  $\beta$ -lactamase-like (Rmet\_1266) protein and an outer membrane protein involved in an efflux system (Rmet\_1267). This result was confirmed using PCR screening of the Rmet\_1266 gene among B&B clone total genomic DNA. Amplification was obtained for the B&B clone but not for J2315. Furthermore, the B&B PCR fragment was found longer than the one expected from the genome of *R. metallidurans* CH34: 2.8 kb instead of 1.41 kb (Annexe D1). This change in size was found to match the one expected from the Rmet\_1266 CDS with the insertion of an IS407. Moreover, the Rmet\_1267 homologous CDS was found disrupted by an ISBmu1 which was itself disrupted by another copy of the same IS element (Fig. 9a). These multiple insertions in GI-a are consistent with our previous results suggesting a preferential colonization of newly acquired DNA by IS elements (chapter 3). Interestingly, *R. metallidurans* and the B&B clone shared the same synteny for these two genes suggesting an horizontal gene transfer between these two strains/species. These genes are part of a *R. metallidurans* gene cluster involved in resistance to antibiotics. Using the Genoscope website, we looked at the region containing Rmet\_1267 and Rmet\_1266 genes among *R. metallidurans* CH34 genome. This region harbors two transposases from IS21 (Rmet\_1263-64) and IS1090 (Rmet\_1262) families, a type four secretion system (Rmet\_1259), and *traF* genes (Rmet\_1258) and other features typical of an Integrative Conjugative Element (ICE) (annexeD1). This ICE could have played part in the emergence of the B&B clone. Several



**Fig. 8. Methodology used for GI screening among non-sequenced genomes; case of GI10.**

a) Position of primers used to screen the GI10 region. b) Electrophoresis of PCR fragments obtained after screening of the three strains. c) Tripartite comparison of PCR fragments sequenced, *B. cenocepacia* J2315, HI2424 genomes using ACT graphic tool. Primers used and annealing temperature are described in the Table 2.

genes seem to have been counter-selected by IS insertions in this DNA region. This could be related to the metabolic burden represented by expression of these genes.

The second DNA region identified in the B&B clone genome using the IS407 as a DNA probe harbors sequences showing identities with *P. aeruginosa* PA14 genome. Two of these were identified as genes encoding hypothetical proteins (PA14\_31230 and PA14\_31220), and one to encode a partial CDS showing 80% similarities with a deduced NrbE-like protein (BCAL31240) (Fig. 9b). This protein has a domain close to the major facilitator superfamily (MFS) protein. MFS is the largest family of transporter proteins which are highly represented in bacteria and implicated in different ways in multidrug transport (Fluman and Bibi 2009). Interestingly, PA14 and the B&B clone showed the same synteny for these three genes, excepted for IS407 which was not detected into *P. aeruginosa* PA14.

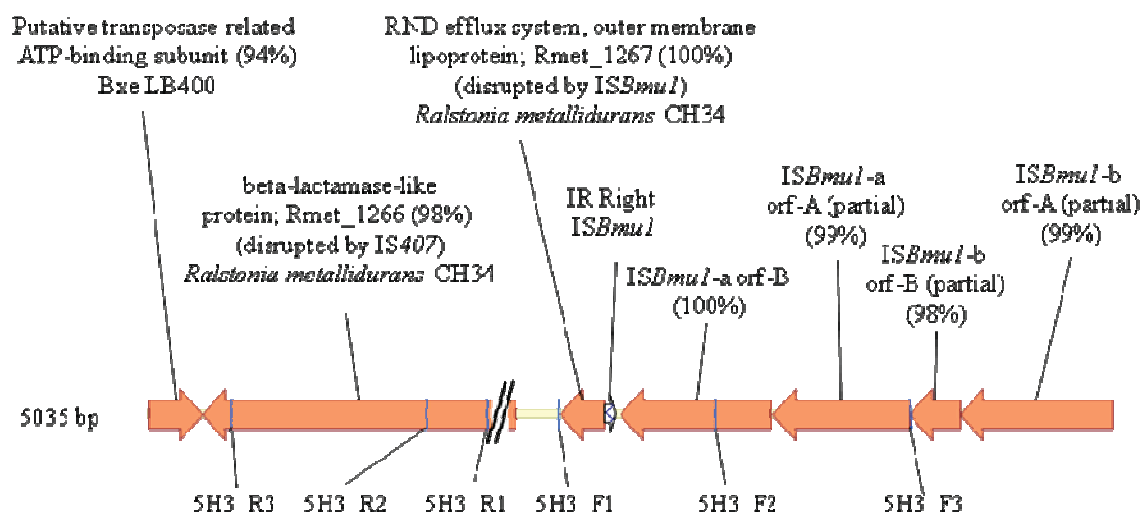
Presence of these genes with the same synteny between PA14 and the B&B clone strongly suggest a horizontal transfer between these two species. Moreover, using the *Pseudomonas* Genome Database, we looked at the region containing PA14\_31230 and PA14\_31220 genes among *P. aeruginosa* PA14 genome (Annexe D2). This region seemed typical of a genomic island; with no homologies with other *Pseudomonas* genomes and a CDS encoding an integrase (PA14\_31280). This GI acquisition may have occurred during the “pathogenic” life of these bacteria which are able to form mixed biofilms among CF patients. Moreover, a communication *via* Quorum Sensing has been demonstrated between these two species, suggesting a close collaboration in CF lungs (Saiman et al., 1990; Riedel et al., 2001; Chambers et al., 2005). This supports a potential coevolution of these two species leading to gene transfers such as acquisition of GI.

These data highlight the role of GI and IS elements in the “fitness of bacteria”. Detection of GI26, GI29 and GI57 in the B&B clone confirms the closer proximity of this clone with ET12 strains than PHDC ones. However, absence of GI10, GI11, GI20 and GI41 in the B&B clone suggests a significant divergence with the ET12 lineage. This diversification was further supported by an identification of two DNA regions (probable GI) in the clone B&B which were not found in J2315 (and the PHDC strains).

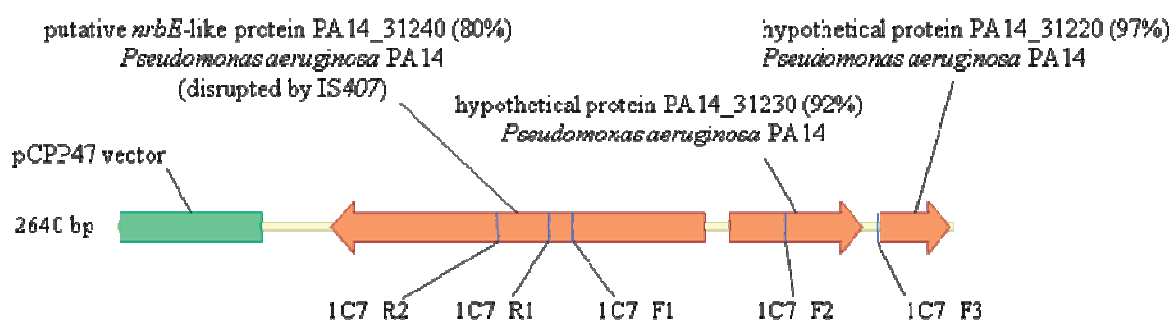
## 4. Conclusion

The data presented in this study suggested a heterogeneous distribution of GI among the B&B clone and J2315 genomes. Different horizontal gene transfers seem to have contributed to this heterogeneity, with recent acquisitions likely to have occurred among each clone. Molecular events involved in GI integration are not well understood but tRNA gene mediated site specific recombination appears to play an important part. Roles of GI are still unclear but they represent a major source of genomic diversity probably leading to improve fitness. IS elements, notably IS407, appear to preferentially colonize GI. The use of IS407 as a DNA probe allowed detection of DNA regions sharing similarities with GI among the B&B clone. Indeed, this approach IS DNA probing, could be used for rapid and efficient detection of GI among the *Burkholderia* genus, complementary to classical genome sequencing projects. Additional studies, such as functional genomics and dynamics of the insertion/excision of these GI into the B&B clone, will be required to further improve our understanding of this clone.

a) GI-a B&B clone 5H3



b) GI-b B&B clone 1C7



**Fig. 9: Genomic islands detected among the B&B clone.**

Genomic islands have been detected using a B&B clone genomic DNA library. *IS407* have been used as probe to screen the library (e. g. material and methods). a) correspond to the first genomic island identified, b) correspond to the second genomic island identified. The two parallel bars are located at the potential insertion site of *IS407* element. Sequencing has been performed using primers indicated at the bottom of each GI. Sequences of primers are described within table 2. Annotations have been performed using BLASTN/BLASTX analysis and % identities are indicated between brackets. Sequences of these GI are available on the Microbial Ecology laboratory website: <http://ecomicro.univ-lyon1.fr/poe/BBcloneGI.zip>.

## Conclusion et perspectives

Les travaux réalisés ces dernières années sur les Bcc portent principalement sur leur taxonomie, les techniques d'identification, de classification ainsi que sur la caractérisation de gènes ou de processus impliqués dans la virulence. Peu d'études ont tenté de déterminer les processus liés au phénomène d'éclosion épidémique. Les pathogènes opportunistes sont, par définition, retrouvés accidentellement chez l'Homme, au niveau d'infection de personnes immunodéprimées ou fragiles tels que les individus CF. La plupart des espèces du Bcc devrait donc être retrouvée dans des écosystèmes naturels. Cependant bien que certaines études démontrent que les souches environnementales sont comparables à celles retrouvées chez l'Homme, aucun réservoir significatif de *B. cenocepacia* et *B. multivorans* n'a été décrit.

Un fait intéressant dans la versatilité de colonisation de ces organismes est leur capacité à occuper des milieux très différents. Cette adaptation et/ou acclimatation à ces environnements implique des processus évolutifs mis en place d'une manière temporelle en réponse aux différentes pressions de sélection rencontrées. Les génomes des *Burkholderia* et leur plasticité facilitent cette évolution. A long terme cette évolution peut conduire à une existence totalement dépendante d'un hôte comme c'est le cas pour les symbiotes et les parasites obligatoires, ou alors à une aptitude à coloniser des milieux extrêmement variés comme cela semble être le cas pour les espèces du Bcc. Cependant, les publications relatives à l'épidémiologie et l'étude de la prévalence des espèces du Bcc retrouvées chez les patients CF suggèrent une hétérogénéité dans la capacité de certaines espèces à coloniser certains milieux ; en effet, ces études montrent une prévalence élevée de *B. multivorans* et *B. cenocepacia* chez les patients CF mais faible dans l'environnement. De plus, l'espèce *B. vietnamiensis* bien que présente au niveau de ces patients, est plus fréquemment retrouvée dans des milieux pollués par des xénobiotiques ou dans la rhizosphère de diverses plantes. Il se dégage donc que certaines espèces du Bcc semblent plus "adaptées" à la niche Homme alors que d'autres seraient plus aptes à coloniser les écosystèmes naturels. Cette constatation semble aussi applicable au sein même de l'espèce *B. cenocepacia*. Cette dernière est, effectivement, subdivisée en sous-groupes correspondant aux différents profils de restriction du gène *recA* (sous-groupe IIIA, IIIB, IIIC, IIID et IIIE). Le sous groupe IIIC abrite des souches provenant uniquement de l'environnement, tandis que le sous groupe IIIA abrite la plupart des isolats cliniques de cette espèce.

Notre travail s'est focalisé sur l'étude d'une éclosion épidémique et sur la compréhension des processus évolutifs liés à ce phénomène. Le clone épidémique B&B de

l'espèce *B. cenocepacia*, responsable d'une infection nosocomiale en 2004, a été utilisé comme modèle.

Dans la première partie de ce travail, nous avons caractérisé le clone responsable d'une épidémie survenue dans le service de réanimation d'un hôpital de la région de Bourg-en-Bresse. La virulence et la forte transmissibilité de ce clone nous ont amené à étudier son origine phylogénétique et sa proximité génétique avec les clones virulents du Bcc. L'utilisation de divers outils de biologie moléculaire nous ont permis de démontrer son appartenance à l'espèce *B. cenocepacia* ainsi que la présence de nombreux gènes de virulence. Les analyses par phylogénie et MLST ont permis de mettre en évidence la distribution mondiale de ce clone et sa prévalence au niveau des infections de patients CF en France. Nous avons également pu établir sa proximité avec les différents clones épidémiques connus, notamment la lignée ET12 de l'espèce *B. cenocepacia* qui représente la lignée la plus dommageable de cette espèce. Une cinquantaine de génomes étant désormais disponibles au sein du genre *Burkholderia*, nous avons employé la génomique comparative afin d'identifier les configurations génétiques originales pouvant être impliquées dans l'émergence de clones épidémiques tel que le clone B&B. Ainsi, nous avons considéré ce processus d'émergence de manière globale en appréciant (i) les phénomènes d'acclimatation, par l'étude de régulateurs de l'expression des gènes (ii) les mécanismes d'adaptation, en développant une approche relative à la présence d'éléments génétiques mobiles de type séquences d'insertion et îlot génomique.

Dans ce contexte, nous nous sommes, tout d'abord, focalisés sur l'étude d'une famille de régulateurs transcriptionnels, la famille des facteurs  $\sigma 70$ . L'étude de certains facteurs sigma (e.g. ECF) peut permettre d'accéder aux sigmulons impliqués dans la régulation de l'expression de certains facteurs de virulence bactériens. Nous avons donc analysé l'architecture, la distribution et la plasticité de la famille des facteurs  $\sigma 70$  chez les Bcc tout en précisant les divergences et similarités entre les principaux clones épidémiques isolés de patients CF et le clone B&B. Un de ces facteurs  $\sigma 70$ , le gène *ecfB*, a été utilisé pour développer un nouveau marqueur, exploité dans la première partie de ce travail lors de l'étude des relations phylogénétiques du clone B&B avec les autres clones épidémiques de *B. cenocepacia*. L'analyse de la distribution et le positionnement phylétique des facteurs déduits nous a également permis de mettre en évidence quelques spécificités liées aux Bcc. Plusieurs gènes codant des facteurs primaires alternatifs ont été détectés pour la première fois hors du groupe des actinobactéries. Nous avons également identifié une distribution restreinte de certains facteurs ECF au niveau du Bcc. Par ailleurs, l'étude de la distribution de ces facteurs

a permis de confirmer la proximité du clone B&B avec la lignée ET12. En effet, une distribution similaire des  $\sigma 70$  a pu être mise en évidence entre ces clones épidémiques. Ces travaux ont mis en évidence deux stratégies évolutives différentes chez *B. cenocepacia* et *P. aeruginosa* au niveau des ECF liés à l'acquisition du fer. Chez les *Pseudomonas*, près d'une vingtaine de sigmulons liés à l'acquisition du fer (de type *fecI*) a été sélectionnée au cours de l'évolution. Un seul paralogue a pu être décelé chez les Bcc et la fonction de celui-ci semble différente de celle assurée par son homologue chez *P. aeruginosa*. La distribution inégale des  $\sigma 70$  au sein du Bcc a été expliquée, entre autre, par des phénomènes de réarrangements génétiques. Certains d'entre eux ont pu être reliés à la présence d'éléments génétiques répétés de la famille des séquences d'insertion et à certains phénomènes d'acquisition d'îlot génomiques.

Nous avons poursuivi notre étude par l'évaluation de l'implication d'IS dans l'émergence de clones épidémiques dont le clone B&B. En nous basant sur l'analyse des nombreux génomes de *Burkholderia* disponibles, nous avons mis en évidence l'abondance et la diversité des IS présentes chez ces bactéries. Ceci nous a permis d'observer une organisation génomique du clone B&B distincte, bien que très proche, de celle de la lignée ET12 suggérant une divergence récente de ces deux souches. Des phénomènes d'acquisition/délétion, d'amplification et de réarrangements de ces IS ont pu être identifiés. De plus, l'implication des IS dans l'adaptation des clones épidémiques de l'espèce *B. cenocepacia* a été évaluée, par l'analyse des sites d'insertion et de leur distribution au sein de l'espèce *B. cenocepacia* ; une distribution très hétérogène des loci colonisés par les IS au sein des différents génomes testés et au sein d'une même espèce a été mise en évidence. Il semble donc que les différents types d'IS n'aient pas joué le même rôle dans l'émergence des différents clones épidémiques de *B. cenocepacia* : certains ont permis l'inactivation de gènes, et d'autres, des réarrangements génétiques locaux et globaux ayant favorisé l'acquisition ou la perte de fragments d'ADN. Ces phénomènes d'acquisition et perte d'ADN ont pu être reliés à la présence d'îlots génomiques.

Dans la dernière partie de ce travail, nous avons développé différentes stratégies pour identifier de potentiels îlots génomiques chez les clones B&B et ET12. Ces travaux ont permis de détecter de nombreuses séquences d'ADN spécifiques de la souche J2315, avec plus d'une vingtaine d'entre elles présentant les caractéristiques d'îlots génomiques. Six de ces îlots potentiels ont été caractérisés et leur distribution a été testée chez une souche de la lignée ET12 et le clone B&B. Les résultats obtenus ont démontré la présence de l'ensemble de ces îlots chez les souches ET12 et seulement trois ont pu être détectés chez le clone B&B.



Différents transferts horizontaux semblent avoir contribué à cette hétérogénéité, des acquisitions/délétions sont soupçonnées chez les différents clones testés. Certains de ces IG ont pu être reliés à des propriétés de virulence. Ces travaux ont également mis en évidence une relation particulière entre les IG et certaines IS. En effet, l'IS407, semble coloniser préférentiellement ces six IG. Nous avons donc utilisé cette observation comme outil de détection d'IG chez le clone B&B. Deux IG potentiels portant des gènes homologues aux espèces *P. aeruginosa* et *R. metallidurans* ont ainsi été identifiés. Ces travaux ont ainsi validé l'utilisation d'IS comme "marqueur" d'IG. Cette approche pourrait être utilisée pour la détection rapide et efficace d'IG chez les Bcc, en utilisant le bilan IS du chapitre 3.

Cette thèse s'est inscrite dans la problématique des éclosions épidémiques et des processus impliqués. Les différents thèmes abordés nous ont permis d'approfondir les connaissances relatives à la plasticité des génomes de *Burkholderia* et particulièrement de l'espèce *B. cenocepacia*. Les travaux futurs découlant de ces études devraient conduire à tester certaines des hypothèses formulées au cours de cette thèse, comme par exemple le rôle potentiel des *ISBcen4* dans l'émergence du clone B&B. A cette fin, deux approches peuvent être utilisées. Tout d'abord, il est possible d'analyser l'impact de l'inactivation d'un gène par l'*ISBcen4* en effectuant sa complémentation génétique. L'utilisation des différents modèles d'étude présentés en introduction permettrait ensuite d'évaluer l'impact de ces modifications sur la virulence bactérienne. La seconde approche pourrait consister à introduire l'*ISBcen4* et suivre l'évolution de cette souche. Ces bactéries « enrichies d'IS » pourraient être inoculées à un hôte modèle. Si conformément à notre hypothèse, certaines IS peuvent générer de la diversité, après passage dans l'hôte, les bactéries isolées pourraient présenter différents degrés de virulence. L'analyse de la distribution et de l'impact des IS au sein de ces bactéries permettra de confirmer leurs implications dans la modulation de la virulence bactérienne. Ces analyses pourraient être complétées par le suivi de certains ilots génomiques selon divers stress mimant des conditions d'infection *versus* conditions environnementales. Ceci permettrait de déterminer l'instabilité de telles régions génétiques et pourrait participer à la compréhension de leur rôle dans la fitness de ce clone.

Dans le second chapitre de ce travail nous avons émis des hypothèses fortes concernant l'implication de facteurs ECF dans la régulation de sigmulons associés à la virulence des espèces du Bcc. De plus, nous avons identifié des CDS codant des facteurs primaires alternatifs dont le rôle est indéterminé. Il serait donc opportun d'inactiver certains de ces gènes afin de déterminer leur rôle. D'autre part, les récentes techniques développées pourraient être utilisées pour étudier les réponses de souches à l'ajout ou la délétion d'un

ECF. En effet, le séquençage à haut débit, allié aux nombreux génomes séquencés désormais disponibles, permet l'obtention d'une multitude de données en peu de temps. La caractérisation de l'ensemble des sigmulons pourrait potentiellement être réalisée dans le temps imparti à un travail de thèse.

Les analyses réalisées dans le chapitre 1 ont démontré que des souches provenant de l'environnement et de patients CF appartenaient au même complexe clonal que le clone B&B. Il apparaîtrait intéressant désormais de connaître la distribution des IS et des IG détectées chez le clone B&B au sein des souches retrouvées dans ce complexe.

# Bibliographie

- Abalain M. L. et al., (2009). Clonally identical *Burkholderia cepacia* complex (Bcc) sequence type strain ST32 isolated from cystic fibrosis (CF) patients in French western Brittany compared to epidemiology of Bcc strains in France using Multilocus sequence typing (MLST). *Journal of Cystic Fibrosis* 8(Supplement 2): S34.
- Aballay A. and Ausubel F. M. (2002). *Caenorhabditis elegans* as a host for the study of host-pathogen interactions. *Curr Opin Microbiol* 5(1): 97-101.
- Adamo R. et al., (2004). *Pseudomonas aeruginosa* flagella activate airway epithelial cells through asialoGM1 and toll-like receptor 2 as well as toll-like receptor 5. *Am J Respir Cell Mol Biol* 30(5): 627-34.
- Agodi A. et al., (2001). *Burkholderia cepacia* complex infection in Italian patients with cystic fibrosis: prevalence, epidemiology, and genomovar status. *J Clin Microbiol* 39(8): 2891-6.
- Altschul S. F. et al., (1990). Basic local alignment search tool. *J Mol Biol* 215(3): 403-10.
- Alvarez-Lerma F. et al., (2008). Moisturizing body milk as a reservoir of *Burkholderia cepacia*: outbreak of nosocomial infection in a multidisciplinary intensive care unit. *Crit Care* 12(1): R10.
- Aubert D. F. et al., (2008). A novel sensor kinase-response regulator hybrid controls biofilm formation and type VI secretion system activity in *Burkholderia cenocepacia*. *Infect Immun* 76(5): 1979-91.
- Balandreau J. et al., (2001). *Burkholderia cepacia* genomovar III Is a common plant-associated bacterium. *Appl Environ Microbiol* 67(2): 982-5.
- Baldwin A. et al., (2008). Elucidating global epidemiology of *Burkholderia multivorans* in cases of cystic fibrosis by multilocus sequence typing. *J Clin Microbiol* 46(1): 290-5.
- Baldwin A. et al., (2007). Environmental *Burkholderia cepacia* complex isolates in human infections. *Emerg Infect Dis* 13(3): 458-61.
- Baldwin A. et al., (2005). Multilocus sequence typing scheme that provides both species and strain differentiation for the *Burkholderia cepacia* complex. *J Clin Microbiol* 43(9): 4665-73.
- Baldwin A. et al., (2004). The *Burkholderia cepacia* epidemic strain marker is part of a novel genomic island encoding both virulence and metabolism-associated genes in *Burkholderia cenocepacia*. *Infect Immun* 72(3): 1537-47.
- Bamford S. et al., (2007). Highly purified lipopolysaccharides from *Burkholderia cepacia* complex clinical isolates induce inflammatory cytokine responses via TLR4-mediated MAPK signalling pathways and activation of NFkappaB. *Cell Microbiol* 9(2): 532-43.
- Barsomian G. and Lessie T. G. (1986). Replicon fusions promoted by insertion sequences on *Pseudomonas cepacia* plasmid pTGL6. *Mol Gen Genet* 204(2): 273-80.
- Bertani G. (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol* 62(3): 293-300.
- Bevivino A. et al., (2002). *Burkholderia cepacia* complex bacteria from clinical and environmental sources in Italy: genomovar status and distribution of traits related to virulence and transmissibility. *J Clin Microbiol* 40(3): 846-51.
- Bickle T. A. and Kruger D. H. (1993). Biology of DNA restriction. *Microbiol Rev* 57(2): 434-50.
- Blackburn L. et al., (2004). 'Cepacia syndrome' with *Burkholderia multivorans*, 9 years after initial colonization. *J Cyst Fibros* 3(2): 133-4.
- Blot M. (1994). Transposable elements and adaptation of host bacteria. *Genetica* 93(1-3): 5-12.
- Blum G. et al., (1994). Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infect Immun* 62(2): 606-14.
- Boas S. R. et al., (2000). Postexercise immune correlates in children with and without cystic fibrosis. *Med Sci Sports Exerc* 32(12): 1997-2004.
- Boyd E. F. et al., (2009). Genomic islands are dynamic, ancient integrative elements in bacterial evolution. *Trends Microbiol* 17(2): 47-53.
- Brett P. J. et al., (1998). *Burkholderia thailandensis* sp. nov., a *Burkholderia pseudomallei*-like species. *Int J Syst Bacteriol* 48 Pt 1: 317-20.
- Burkholder W. H. (1950). Sour skin, a bacterial rot of onion bulbs. *Phytopathology* 40: 115-117.
- Byrne A. M. and Lessie T. G. (1994). Characteristics of IS401, a new member of the IS3 family implicated in plasmid rearrangements in *Pseudomonas cepacia*. *Plasmid* 31(2): 138-47.
- Campana S. et al., (2005). Transmission of *Burkholderia cepacia* complex: evidence for new epidemic clones infecting cystic fibrosis patients in Italy. *J Clin Microbiol* 43(10): 5136-42.
- Cardona S. T. et al., (2005). Diverse pathogenicity of *Burkholderia cepacia* complex strains in the *Caenorhabditis elegans* host model. *FEMS Microbiol Lett* 250(1): 97-104.

- Carta L. K. (2000). Bacterial-feeding nematode growth and preference for biocontrol isolates of the bacterium *Burkholderia cepacia*. *Journal of Nematology* 32(4): 362-369.
- Carver T. J. et al., (2005). ACT: the Artemis Comparison Tool. *Bioinformatics* 21(16): 3422-3.
- Cascales E. and Christie P. J. (2003). The versatile bacterial type IV secretion systems. *Nat Rev Microbiol* 1(2): 137-49.
- Centron D. and Roy P. H. (2002). Presence of a group II intron in a multiresistant *Serratia marcescens* strain that harbors three integrons and a novel gene fusion. *Antimicrob Agents Chemother* 46(5): 1402-9.
- Cesarini S. et al., (2009). *recA* gene sequence and Multilocus Sequence Typing for species-level resolution of *Burkholderia cepacia* complex isolates. *Lett Appl Microbiol*.
- Chain P. S. et al., (2006). Complete genome sequence of *Yersinia pestis* strains Antiqua and Nepal516: evidence of gene reduction in an emerging pathogen. *J Bacteriol* 188(12): 4453-63.
- Chambers C. E. et al., (2005). Identification of N-acylhomoserine lactones in mucopurulent respiratory secretions from cystic fibrosis patients. *FEMS Microbiol Lett* 244(2): 297-304.
- Charlesworth B. et al., (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371(6494): 215-20.
- Chen J. S. et al., (2001). Endemicity and inter-city spread of *Burkholderia cepacia* genomovar III in cystic fibrosis. *J Pediatr* 139(5): 643-9.
- Chen W. M. et al., (2005). Proof that *Burkholderia* strains form effective symbioses with legumes: a study of novel Mimosa-nodulating strains from South America. *Appl Environ Microbiol* 71(11): 7461-71.
- Christie P. J. et al., (2005). Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annu Rev Microbiol* 59: 451-85.
- Cieri M. V. et al., (2002). Correlation between an in vitro invasion assay and a murine model of *Burkholderia cepacia* lung infection. *Infect Immun* 70(3): 1081-6.
- Clode F. E. et al., (2000). Distribution of genes encoding putative transmissibility factors among epidemic and nonepidemic strains of *Burkholderia cepacia* from cystic fibrosis patients in the United Kingdom. *J Clin Microbiol* 38(5): 1763-6.
- Coenye T. and LiPuma J. J. (2002). Multilocus restriction typing: a novel tool for studying global epidemiology of *Burkholderia cepacia* complex infection in cystic fibrosis. *J Infect Dis* 185(10): 1454-62.
- Coenye T. and LiPuma J. J. (2003). Molecular epidemiology of *Burkholderia* species. *Front Biosci* 8: e55-67.
- Coenye T. et al., (2001). *Burkholderia ambifaria* sp. nov., a novel member of the *Burkholderia cepacia* complex including biocontrol and cystic fibrosis-related isolates. *Int J Syst Evol Microbiol* 51(Pt 4): 1481-90.
- Coenye T. et al., (2004). Recovery of *Burkholderia cenocepacia* strain PHDC from cystic fibrosis patients in Europe. *Thorax* 59(11): 952-4.
- Coenye T. and Vandamme P. (2003). Diversity and significance of *Burkholderia* species occupying diverse ecological niches. *Environ Microbiol* 5(9): 719-29.
- Compant S. et al., (2008). Diversity and occurrence of *Burkholderia* spp. in the natural environment. *FEMS Microbiol Rev* 32(4): 607-26.
- Corbett C. R. et al., (2003). An extracellular zinc metalloprotease gene of *Burkholderia cepacia*. *Microbiology* 149(Pt 8): 2263-71.
- Croucher N. J. et al., (2009). Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone *Streptococcus pneumoniae* Spain 23F ST81. *J Bacteriol* 191(5): 1480-9.
- Cunha M. V. et al., (2007). Exceptionally high representation of *Burkholderia cepacia* among *B. cepacia* complex isolates recovered from the major Portuguese cystic fibrosis center. *J Clin Microbiol* 45(5): 1628-33.
- Dalmastri C. et al., (2007). Investigating *Burkholderia cepacia* complex populations recovered from Italian maize rhizosphere by multilocus sequence typing. *Environ Microbiol* 9(7): 1632-9.
- Dalmastri C. et al., (2005). Efficacy of species-specific *recA* PCR tests in the identification of *Burkholderia cepacia* complex environmental isolates. *FEMS Microbiol Lett* 246(1): 39-45.
- Darling P. et al., (1998). Siderophore production by cystic fibrosis isolates of *Burkholderia cepacia*. *Infect Immun* 66(2): 874-7.
- DeShazer D. et al., (2001). Identification of a *Burkholderia mallei* polysaccharide gene cluster by subtractive hybridization and demonstration that the encoded capsule is an essential virulence determinant. *Microb Pathog* 30(5): 253-69.
- Dobrindt U. et al., (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* 2(5): 414-24.
- Doublet B. et al., (2008). Novel insertion sequence- and transposon-mediated genetic rearrangements in genomic island SG11 of *Salmonella enterica* serovar Kentucky. *Antimicrob Agents Chemother* 52(10): 3745-54.
- Drevinek P. et al., (2003). Genomovar distribution of the *Burkholderia cepacia* complex differs significantly between Czech and Slovak patients with cystic fibrosis. *J Med Microbiol* 52(Pt 7): 603-4.

- Drevinek P. et al., (2005). Widespread clone of *Burkholderia cenocepacia* in cystic fibrosis patients in the Czech Republic. *J Med Microbiol* 54(Pt 7): 655-9.
- Dubarry N. et al., (2006). ParABS systems of the four replicons of *Burkholderia cenocepacia*: new chromosome centromeres confer partition specificity. *J Bacteriol* 188(4): 1489-96.
- Duff C. et al., (2006). Differences in invasion and translocation of *Burkholderia cepacia* complex species in polarised lung epithelial cells in vitro. *Microb Pathog* 41(4-5): 183-92.
- Eberl L. (2006). Quorum sensing in the genus *Burkholderia*. *Int J Med Microbiol* 296(2-3): 103-10.
- Engledow A. S. et al., (2004). Involvement of a plasmid-encoded type IV secretion system in the plant tissue watersoaking phenotype of *Burkholderia cenocepacia*. *J Bacteriol* 186(18): 6015-24.
- Enright M. C. and Spratt B. G. (1999). Multilocus sequence typing. *Trends Microbiol* 7(12): 482-7.
- Eppinger M. et al., (2007). The complete genome sequence of *Yersinia pseudotuberculosis* IP31758, the causative agent of Far East scarlet-like fever. *PLoS Genet* 3(8): e142.
- Estrada-De Los Santos P. et al., (2001). *Burkholderia*, a genus rich in plant-associated nitrogen fixers with wide environmental and geographic distribution. *Appl Environ Microbiol* 67(6): 2790-8.
- Ferrante A. A. and Lessie T. G. (1991). Nucleotide sequence of IS402 from *Pseudomonas cepacia*. *Gene* 102(1): 143-4.
- Finlay B. B. and Falkow S. (1997). Common themes in microbial pathogenicity revisited. *Microbiol Mol Biol Rev* 61(2): 136-69.
- Fluman N. and Bibi E. (2009). Bacterial multidrug transport through the lens of the major facilitator superfamily. *Biochim Biophys Acta* 1794(5): 738-47.
- France M. W. et al., (2008). The changing epidemiology of *Burkholderia* species infection at an adult cystic fibrosis centre. *J Cyst Fibros* 7(5): 368-72.
- Fredrickson J. K. et al., (2004). Geomicrobiology of high-level nuclear waste-contaminated vadose sediments at the Hanford site, Washington state. *Appl Environ Microbiol* 70(7): 4230-41.
- Gaffney T. D. and Lessie T. G. (1987). Insertion-sequence-dependent rearrangements of *Pseudomonas cepacia* plasmid pTGL1. *J Bacteriol* 169(1): 224-30.
- Gautam V. et al., (2008). Two cases of *Burkholderia cenocepacia* in septicemic patients. *Jpn J Infect Dis* 61(2): 133-4.
- Ghazal S. S. et al., (2006). Outbreak of *Burkholderia cepacia* bacteremia in immunocompetent children caused by contaminated nebulized sulbutamol in Saudi Arabia. *Am J Infect Control* 34(6): 394-8.
- Gillis M. et al., (1995). Polyphasic taxonomy in the genus *Burkholderia* leading to an emended description of the genus and proposition of *Burkholderia vietnamiensis* sp. nov. for N<sub>2</sub>-fixing isolates from rice in Vietnam. *Int J Syst Bacteriol* 45: 274-289.
- Gingues S. et al., (2005). Distribution and expression of the ZmpA metalloprotease in the *Burkholderia cepacia* complex. *J Bacteriol* 187(24): 8247-55.
- Glendinning K. J. et al., (2004). Sequence divergence in type III secretion gene clusters of the *Burkholderia cepacia* complex. *FEMS Microbiol Lett* 235(2): 229-35.
- Golini G. et al., (2006). Molecular epidemiology and antibiotic susceptibility of *Burkholderia cepacia*-complex isolates from an Italian cystic fibrosis centre. *Eur J Clin Microbiol Infect Dis* 25(3): 175-80.
- Gonzalez C. F. et al., (1997). Mobilization, cloning, and sequence determination of a plasmid-encoded polygalacturonase from a phytopathogenic *Burkholderia* (*Pseudomonas*) *cepacia*. *Mol Plant Microbe Interact* 10(7): 840-51.
- Govan J. R. et al., (2007). Evolving epidemiology of *Pseudomonas aeruginosa* and the *Burkholderia cepacia* complex in cystic fibrosis lung infection. *Future Microbiol* 2: 153-64.
- Govan J. R. et al., (1993). Evidence for transmission of *Pseudomonas cepacia* by social contact in cystic fibrosis. *Lancet* 342(8862): 15-9.
- Govindarajan M. et al., (2008). Effects of the inoculation of *Burkholderia vietnamiensis* and related endophytic diazotrophic bacteria on grain yield of rice. *Microb Ecol* 55(1): 21-37.
- Graindorge A. et al., (2009). Epidemiology and molecular characterization of a clone of *Burkholderia cenocepacia* responsible for nosocomial pulmonary tract infections in a French intensive care unit. *Diagn Microbiol Infect Dis*.
- Groth A. C. and Calos M. P. (2004). Phage integrases: biology and applications. *J Mol Biol* 335(3): 667-78.
- Haas B. et al., (1991). Siderophore presence in sputa of cystic fibrosis patients. *Infect Immun* 59(11): 3997-4000.
- Hacker J. et al., (1990). Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extraintestinal *Escherichia coli* isolates. *Microb Pathog* 8(3): 213-25.
- Hacker J. and Carniel E. (2001). Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. *EMBO Rep* 2(5): 376-81.
- Hacker J. and Kaper J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* 54: 641-79.

- Han J. et al., (2009). Diversity of culturable bacteria isolated from root domains of moso bamboo (*Phyllostachys edulis*). *Microb Ecol* 58(2): 363-73.
- Harrison F. (2007). Microbial ecology of the cystic fibrosis lung. *Microbiology* 153(Pt 4): 917-23.
- Haugland R. A. et al., (1991). Cloning and characterization of a chromosomal DNA region required for growth on 2,4,5-T by *Pseudomonas cepacia* AC1100. *Gene* 100: 65-73.
- Henry D. A. et al., (2001). Phenotypic methods for determining genomovar status of the *Burkholderia cepacia* complex. *J Clin Microbiol* 39(3): 1073-8.
- Hens D. K. et al., (2006). New temperate DNA phage BcP15 acts as a drug resistance vector. *Arch Virol* 151(7): 1345-53.
- Hobson R. et al., (1995). *Burkholderia (Pseudomonas) cepacia* as a cause of brain abscesses secondary to chronic suppurative otitis media. *Eur J Clin Microbiol Infect Dis* 14(10): 908-11.
- Holden M. T. et al., (2009). The genome of *Burkholderia cenocepacia* J2315, an epidemic pathogen of cystic fibrosis patients. *J Bacteriol* 191(1): 261-77.
- Holden M. T. et al., (2004). Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proc Natl Acad Sci U S A*.
- Holmes A. et al., (1999). An epidemic of *Burkholderia cepacia* transmitted between patients with and without cystic fibrosis. *J Infect Dis* 179(5): 1197-205.
- Huang C. H. et al., (2001). Characteristics of patients with *Burkholderia cepacia* bacteremia. *J Microbiol Immunol Infect* 34(3): 215-9.
- Hubner A. and Hendrickson W. (1997). A fusion promoter created by a new insertion sequence, IS1490, activates transcription of 2,4,5-trichlorophenoxyacetic acid catabolic genes in *Burkholderia cepacia* AC1100. *J Bacteriol* 179(8): 2717-23.
- Hutchison M. L. and Govan J. R. (1999). Pathogenicity of microbes associated with cystic fibrosis. *Microbes Infect* 1(12): 1005-14.
- Ierano T. et al., (2008). The structure and proinflammatory activity of the lipopolysaccharide from *Burkholderia multivorans* and the differences between clonal strains colonizing pre and posttransplanted lungs. *Glycobiology* 18(11): 871-81.
- Isles A. et al., (1984). *Pseudomonas cepacia* infection in cystic fibrosis: an emerging problem. *J Pediatr* 104(2): 206-10.
- Jacobs J. L. et al., (2008). Identification and onion pathogenicity of *Burkholderia cepacia* complex isolates from the onion rhizosphere and onion field soil. *Appl Environ Microbiol* 74(10): 3121-9.
- Jepson M. A. et al., (2000). Localization of dysfunctional tight junctions in *Salmonella enterica* serovar typhimurium-infected epithelial layers. *Infect Immun* 68(12): 7202-8.
- Jimenez L. et al., (2000). Molecular detection of *Burkholderia cepacia* in toiletry, cosmetic, and pharmaceutical raw materials and finished products. *J AOAC Int* 83(4): 963-6.
- Johnson W. M. et al., (1994). Linkage analysis of geographic and clinical clusters in *Pseudomonas cepacia* infections by multilocus enzyme electrophoresis and ribotyping. *J Clin Microbiol* 32(4): 924-30.
- Jones A. M. et al., (2001). *Burkholderia gladioli*: recurrent abscesses in a patient with cystic fibrosis. *J Infect* 42(1): 69-71.
- Juhas M. et al., (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* 33(2): 376-93.
- Kalish L. A. et al., (2006). Impact of *Burkholderia dolosa* on Lung Function and Survival in Cystic Fibrosis. *Am J Respir Crit Care Med* 173(4): 421-5.
- Kenna D. T. et al., (2006). Distribution and genomic location of active insertion sequences in the *Burkholderia cepacia* complex. *J Med Microbiol* 55(Pt 1): 1-10.
- Kim J. Y. et al., (2005). Disruption of tight junctions during traversal of the respiratory epithelium by *Burkholderia cenocepacia*. *Infect Immun* 73(11): 7107-12.
- Kooi C. et al., (2006). *Burkholderia cenocepacia* ZmpB is a broad-specificity zinc metalloprotease involved in virulence. *Infect Immun* 74(7): 4083-93.
- Kostman J. R. et al., (1992). Molecular epidemiology of *Pseudomonas cepacia* determined by polymerase chain reaction ribotyping. *J Clin Microbiol* 30(8): 2084-7.
- Kothe M. et al., (2003). Killing of *Caenorhabditis elegans* by *Burkholderia cepacia* is controlled by the *cep* quorum-sensing system. *Cell Microbiol* 5(5): 343-51.
- Kumar A. et al., (1997). Genetic relatedness of *Burkholderia (Pseudomonas) cepacia* isolates from five cystic fibrosis centers in Michigan. *Respir Med* 91(8): 485-92.
- Kumar S. et al., (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 5(2): 150-63.
- Kyselkova M. et al., (2009). Comparison of rhizobacterial community composition in soil suppressive or conducive to tobacco black root rot disease. *Isme J* 3(10): 1127-38.

- Lambiase A. et al., (2006). Microbiology of airway disease in a cohort of patients with Cystic Fibrosis. *BMC Infect Dis* 6: 4.
- Lambiase A. et al., (2007). *Burkholderia cepacia* complex infection in a cohort of Italian patients with cystic fibrosis. *J Microbiol* 45(3): 275-9.
- Lamothe J. et al., (2007). Intracellular survival of *Burkholderia cenocepacia* in macrophages is associated with a delay in the maturation of bacteria-containing vacuoles. *Cell Microbiol* 9(1): 40-53.
- Lamothe J. et al., (2004). *Burkholderia cepacia* complex isolates survive intracellularly without replication within acidic vacuoles of *Acanthamoeba polyphaga*. *Cell Microbiol* 6(12): 1127-38.
- Lamothe J. and Valvano M. A. (2008). *Burkholderia cenocepacia*-induced delay of acidification and phagolysosomal fusion in cystic fibrosis transmembrane conductance regulator (CFTR)-defective macrophages. *Microbiology* 154(Pt 12): 3825-34.
- Langley R. et al., (2003). Lysogeny and bacteriophage host range within the *Burkholderia cepacia* complex. *J Med Microbiol* 52(Pt 6): 483-90.
- Laraya-Cuasay L. R. et al., (1977). Chronic pulmonary complications of early influenza virus infection in children. *Am Rev Respir Dis* 116(4): 617-25.
- Larsson P. et al., (2005). The complete genome sequence of *Francisella tularensis*, the causative agent of tularemia. *Nat Genet* 37(2): 153-9.
- Leitao J. H. et al., (2008). Variation of the antimicrobial susceptibility profiles of *Burkholderia cepacia* complex clonal isolates obtained from chronically infected cystic fibrosis patients: a five-year survey in the major Portuguese treatment center. *Eur J Clin Microbiol Infect Dis* 27(11): 1101-11.
- Lemke M. J. and Leff L. G. (1999). Bacterial Populations in an Anthropogenically Disturbed Stream: Comparison of Different Seasons. *Microb Ecol* 38(3): 234-243.
- Lennon E. and DeCicco B. T. (1991). Plasmids of *Pseudomonas cepacia* strains of diverse origins. *Appl Environ Microbiol* 57(8): 2345-50.
- Lessie T. G. et al., (1996). Genomic complexity and plasticity of *Burkholderia cepacia*. *FEMS Microbiol Lett* 144(2-3): 117-28.
- Leveau J. H. and Gerards S. (2008). Discovery of a bacterial gene cluster for catabolism of the plant hormone indole 3-acetic acid. *FEMS Microbiol Ecol* 65(2): 238-50.
- Liao D. (1999). Concerted evolution: molecular mechanism and biological implications. *Am J Hum Genet* 64(1): 24-30.
- LiPuma J. J. (1998). *Burkholderia cepacia*. Management issues and new insights. *Clin Chest Med* 19(3): 473-86, vi.
- Lipuma J. J. (2005). Update on the *Burkholderia cepacia* complex. *Curr Opin Pulm Med* 11(6): 528-33.
- LiPuma J. J. et al., (1990). Person-to-person transmission of *Pseudomonas cepacia* between patients with cystic fibrosis. *Lancet* 336(8723): 1094-6.
- LiPuma J. J. et al., (1988). Ribotype analysis of *Pseudomonas cepacia* from cystic fibrosis treatment centers. *J Pediatr* 113(5): 859-62.
- LiPuma J. J. et al., (2002). An epidemic *Burkholderia cepacia* complex strain identified in soil. *Lancet* 359(9322): 2002-3.
- LiPuma J. J. et al., (2001). Disproportionate distribution of *Burkholderia cepacia* complex species and transmissibility markers in cystic fibrosis. *Am J Respir Crit Care Med* 164(1): 92-6.
- Liu L. et al., (2003). Identification by subtractive hybridization of a novel insertion element specific for two widespread *Burkholderia cepacia* genomovar III strains. *J Clin Microbiol* 41(6): 2471-6.
- Lo Cascio G. et al., (2006). A napkin-associated outbreak of *Burkholderia cenocepacia* bacteraemia in haemodialysis patients. *J Hosp Infect* 64(1): 56-62.
- Loukil C. et al., (2003). Epidemiologic investigation of *Burkholderia cepacia* acquisition in two pediatric intensive care units. *Infect Control Hosp Epidemiol* 24(9): 707-10.
- Loutet S. A. et al., (2006). A complete lipopolysaccharide inner core oligosaccharide is required for resistance of *Burkholderia cenocepacia* to antimicrobial peptides and bacterial survival in vivo. *J Bacteriol* 188(6): 2073-80.
- Lynch J. P., 3rd (2009). *Burkholderia cepacia* complex: impact on the cystic fibrosis lung lesion. *Semin Respir Crit Care Med* 30(5): 596-610.
- Mahenthalingam E. et al., (2008). *Burkholderia cepacia* complex bacteria: opportunistic pathogens with important natural biology. *J Appl Microbiol* 104(6): 1539-51.
- Mahenthalingam E. et al., (2006). Multilocus sequence typing breathes life into a microbial metagenome. *PLoS One* 1: e17.
- Mahenthalingam E. et al., (2000a). DNA-Based diagnostic approaches for identification of *Burkholderia cepacia* complex, *Burkholderia vietnamiensis*, *Burkholderia multivorans*, *Burkholderia stabilis*, and *Burkholderia cepacia* genomovars I and III. *J Clin Microbiol* 38(9): 3165-73.

- Mahenthiralingam E. et al., (1996). Epidemiology of *Burkholderia cepacia* infection in patients with cystic fibrosis: analysis by randomly amplified polymorphic DNA fingerprinting. *J Clin Microbiol* 34(12): 2914-20.
- Mahenthiralingam E. et al., (2000b). Diagnostically and experimentally useful panel of strains from the *Burkholderia cepacia* complex. *J Clin Microbiol* 38(2): 910-3.
- Mahenthiralingam E. et al., (1997). Identification and characterization of a novel DNA marker associated with epidemic *Burkholderia cepacia* strains recovered from patients with cystic fibrosis. *J Clin Microbiol* 35(4): 808-16.
- Mahenthiralingam E. et al., (2005). The multifarious, multireplicon *Burkholderia cepacia* complex. *Nat Rev Microbiol* 3(2): 144-56.
- Mahenthiralingam E. et al., (2001). Infection with *Burkholderia cepacia* complex genomovars in patients with cystic fibrosis: virulent transmissible strains of genomovar III can replace *Burkholderia multivorans*. *Clin Infect Dis* 33(9): 1469-75.
- Mahillon J. and Chandler M. (1998). Insertion sequences. *Microbiol Mol Biol Rev* 62(3): 725-74.
- Manno G. et al., (2004). Epidemiology and clinical course of *Burkholderia cepacia* complex infections, particularly those caused by different *Burkholderia cenocepacia* strains, among patients attending an Italian Cystic Fibrosis Center. *J Clin Microbiol* 42(4): 1491-7.
- Marolda C. L. et al., (1999). Intracellular survival and saprophytic growth of isolates from the *Burkholderia cepacia* complex in free-living amoebae. *Microbiology* 145 (Pt 7): 1509-17.
- Martin D. W. and Mohr C. D. (2000). Invasion and intracellular survival of *Burkholderia cepacia*. *Infect Immun* 68(1): 24-9.
- Matoba A. Y. (2003). Polymicrobial keratitis secondary to *Burkholderia ambifaria*, *enterococcus*, and *staphylococcus aureus* in a patient with herpetic stromal keratitis. *Am J Ophthalmol* 136(4): 748-9.
- McClellan S. and Callaghan M. (2009). *Burkholderia cepacia* complex: epithelial cell-pathogen confrontations and potential for therapeutic intervention. *J Med Microbiol* 58(Pt 1): 1-12.
- Mena A. et al., (2008). Genetic adaptation of *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients is catalyzed by hypermutation. *J Bacteriol* 190(24): 7910-7.
- Menard A. et al., (2007). Architecture of *Burkholderia cepacia* complex sigma70 gene family: evidence of alternative primary and clade-specific factors, and genomic instability. *BMC Genomics* 8: 308.
- Meyer J. M. et al., (1995). Ornibactin production and transport properties in strains of *Burkholderia vietnamiensis* and *Burkholderia cepacia* (formerly *Pseudomonas cepacia*). *Biometals* 8(4): 309-17.
- Miche L. et al., (2001). Detection and activity of insertion sequences in environmental strains of *Burkholderia*. *Environ Microbiol* 3(12): 766-73.
- Miller S. C. et al., (2002). Culture-based and non-growth-dependent detection of the *Burkholderia cepacia* complex in soil environments. *Appl Environ Microbiol* 68(8): 3750-8.
- Mills D. M. et al., (1995). A 40 kb chromosomal fragment encoding *Salmonella typhimurium* invasion genes is absent from the corresponding region of the *Escherichia coli* K-12 chromosome. *Mol Microbiol* 15(4): 749-59.
- Moolenaar G. F. et al., (2002). Cho, a second endonuclease involved in *Escherichia coli* nucleotide excision repair. *Proc Natl Acad Sci U S A* 99(3): 1467-72.
- Moore J. E. et al., (2007). Environmental persistence of *Pseudomonas aeruginosa* and *Burkholderia multivorans* in sea water: preliminary evidence of a viable but non-culturable state. *Br J Biomed Sci* 64(3): 129-31.
- Moulin L. et al., (2001). Nodulation of legumes by members of the beta-subclass of Proteobacteria. *Nature* 411(6840): 948-50.
- Mullen T. et al., (2007). Role of lipase in *Burkholderia cepacia* complex (Bcc) invasion of lung epithelial cells. *Eur J Clin Microbiol Infect Dis* 26(12): 869-77.
- Nasser R. M. et al., (2004). Outbreak of *Burkholderia cepacia* bacteremia traced to contaminated hospital water used for dilution of an alcohol skin antiseptic. *Infect Control Hosp Epidemiol* 25(3): 231-9.
- Nolan G. et al., (1982). Antibiotic prophylaxis in cystic fibrosis: inhaled cephaloridine as an adjunct to oral cloxacillin. *J Pediatr* 101(4): 626-30.
- Nzula S. et al., (2000). Sensitivity of the *Burkholderia cepacia* complex and *Pseudomonas aeruginosa* to transducing bacteriophages. *FEMS Immunol Med Microbiol* 28(4): 307-12.
- Nzula S. et al., (2002). Influence of taxonomic status on the in vitro antimicrobial susceptibility of the *Burkholderia cepacia* complex. *J Antimicrob Chemother* 50(2): 265-9.
- O'Carroll M. R. et al., (2003). *Burkholderia pseudomallei*: another emerging pathogen in cystic fibrosis. *Thorax* 58(12): 1087-91.
- O'Quinn A. L. et al., (2001). *Burkholderia pseudomallei* kills the nematode *Caenorhabditis elegans* using an endotoxin-mediated paralysis. *Cell Microbiol* 3(6): 381-93.



- O'Sullivan L. A. et al., (2007). Identifying the genetic basis of ecologically and biotechnologically useful functions of the bacterium *Burkholderia vietnamiensis*. *Environ Microbiol* 9(4): 1017-34.
- Ochman H. et al., (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405(6784): 299-304.
- Ohtsubo Y. et al., (2005). High-temperature-induced transposition of insertion elements in *Burkholderia multivorans* ATCC 17616. *Appl Environ Microbiol* 71(4): 1822-8.
- Olapade O. A. et al., (2005). Abundance of three bacterial populations in selected streams. *Microb Ecol* 49(3): 461-7.
- Olapade O. A. and Leff L. G. (2005). Seasonal response of stream biofilm communities to dissolved organic matter and nutrient enrichments. *Appl Environ Microbiol* 71(5): 2278-87.
- Onofre-Lemus J. et al., (2009). ACC (1-aminocyclopropane-1-carboxylate) deaminase activity, a widespread trait in *Burkholderia* and its growth promoting effect on tomato. *Appl Environ Microbiol*.
- Ortega X. et al., (2005). Reconstitution of O-specific lipopolysaccharide expression in *Burkholderia cenocepacia* strain J2315, which is associated with transmissible infections in patients with cystic fibrosis. *J Bacteriol* 187(4): 1324-33.
- Panlilio A. L. et al., (1992). Infections and pseudoinfections due to povidone-iodine solution contaminated with *Pseudomonas cepacia*. *Clin Infect Dis* 14(5): 1078-83.
- Papadopoulos D. et al., (1999). Genomic evolution during a 10,000-generation experiment with bacteria. *Proc Natl Acad Sci U S A* 96(7): 3807-12.
- Parke J. L. and Gurian-Sherman D. (2001). Diversity of the *Burkholderia cepacia* complex and implications for risk assessment of biological control strains. *Annu Rev Phytopathol* 39: 225-58.
- Parsons Y. N. et al., (2003). Suppression-subtractive hybridisation reveals variations in gene distribution amongst the *Burkholderia cepacia* complex, including the presence in some strains of a genomic island containing putative polysaccharide production genes. *Arch Microbiol* 179(3): 214-23.
- Paul M. L. et al., (1998). Molecular epidemiology of *Burkholderia cepacia* in two Australian cystic fibrosis centres. *J Hosp Infect* 38(1): 19-26.
- Payne D. J. (1993). Metallo-beta-lactamases--a new therapeutic challenge. *J Med Microbiol* 39(2): 93-9.
- Payne G. W. et al., (2005). Development of a *recA* gene-based identification approach for the entire *Burkholderia* genus. *Appl Environ Microbiol* 71(7): 3917-27.
- Peeters E. et al., (2008). Resistance of planktonic and biofilm-grown *Burkholderia cepacia* complex isolates to the transition metal gallium. *J Antimicrob Chemother* 61(5): 1062-5.
- Pitcher D. G., Saunders, N.A., Owen, R.J. (1989). Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Letters in Applied Microbiology* 8: 151-156.
- Pitt T. L. et al., (1996). Type characterisation and antibiotic susceptibility of *Burkholderia* (*Pseudomonas*) *cepacia* isolates from patients with cystic fibrosis in the United Kingdom and the Republic of Ireland. *J Med Microbiol* 44(3): 203-10.
- Prager R. et al., (1995). Genomic and biochemical relatedness between *Vibrio cholerae* serovar O139 and serovar O1 eltor strains. *Zentralbl Bakteriol* 283(1): 14-28.
- Preston G. M. et al., (1998). Bacterial genomics and adaptation to life on plants: implications for the evolution of pathogenicity and symbiosis. *Curr Opin Microbiol* 1(5): 589-97.
- Pritt B. S. and Clark C. G. (2008). Amebiasis. *Mayo Clin Proc* 83(10): 1154-9; quiz 1159-60.
- Rabkin C. S. et al., (1989). *Pseudomonas cepacia* typing systems: collaborative study to assess their potential in epidemiologic investigations. *Rev Infect Dis* 11(4): 600-7.
- Ramette A. et al., (2005). Species abundance and diversity of *Burkholderia cepacia* complex in the environment. *Appl Environ Microbiol* 71(3): 1193-201.
- Rasolomampianina R. et al., (2005). Nitrogen-fixing nodules from rose wood legume trees (*Dalbergia* spp.) endemic to Madagascar host seven different genera belonging to alpha- and beta-Proteobacteria. *Mol Ecol* 14(13): 4135-46.
- Redfearn M. S. et al., (1966). A comparative study of *Pseudomonas pseudomallei* and *Bacillus mallei*. *J Gen Microbiol* 43(2): 293-313.
- Reik R. et al., (2005). Distribution of *Burkholderia cepacia* complex species among isolates recovered from persons with or without cystic fibrosis. *J Clin Microbiol* 43(6): 2926-8.
- Richardson J. et al., (2002). Diversity of *Burkholderia* isolates from woodland rhizosphere environments. *J Appl Microbiol* 93(4): 616-30.
- Riedel K. et al., (2001). N-acylhomoserine-lactone-mediated communication between *Pseudomonas aeruginosa* and *Burkholderia cepacia* in mixed biofilms. *Microbiology* 147(Pt 12): 3249-62.
- Riethmuller J. et al., (2006). Membrane rafts in host-pathogen interactions. *Biochim Biophys Acta* 1758(12): 2139-47.

- Romling U. et al., (1994). Epidemiology of chronic *Pseudomonas aeruginosa* infections in cystic fibrosis. *J Infect Dis* 170(6): 1616-21.
- Romling U. et al., (1997). Large chromosomal inversions occur in *Pseudomonas aeruginosa* clone C strains isolated from cystic fibrosis patients. *FEMS Microbiol Lett* 150(1): 149-56.
- Rose H. et al., (2009). Biocide susceptibility of the *Burkholderia cepacia* complex. *J Antimicrob Chemother* 63(3): 502-10.
- Saiman L. et al., (1990). *Pseudomonas cepacia* adherence to respiratory epithelial cells is enhanced by *Pseudomonas aeruginosa*. *Infect Immun* 58(8): 2578-84.
- Saiman L. and Prince A. (1993). *Pseudomonas aeruginosa* pili bind to asialoGM1 which is increased on the surface of cystic fibrosis epithelial cells. *J Clin Invest* 92(4): 1875-80.
- Saiman L. and Siegel J. (2004). Infection control in cystic fibrosis. *Clin Microbiol Rev* 17(1): 57-71.
- Saitou N. and Nei M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4(4): 406-25.
- Sajjan S. U. et al., (2008a). A type IV secretion system contributes to intracellular survival and replication of *Burkholderia cenocepacia*. *Infect Immun* 76(12): 5447-55.
- Sajjan U. et al., (2002). Interaction of *cblA*/adhesin-positive *Burkholderia cepacia* with squamous epithelium. *Cell Microbiol* 4(2): 73-86.
- Sajjan U. S. et al., (1992). Binding of *Pseudomonas cepacia* to normal human intestinal mucin and respiratory mucin from patients with cystic fibrosis. *J Clin Invest* 89(2): 648-56.
- Sajjan U. S. et al., (2008b). *Burkholderia cenocepacia* ET12 strain activates TNFR1 signalling in cystic fibrosis airway epithelial cells. *Cell Microbiol* 10(1): 188-201.
- Sajjan U. S. et al., (1995). Cable (*cbl*) type II pili of cystic fibrosis-associated *Burkholderia (Pseudomonas) cepacia*: nucleotide sequence of the *cblA* major subunit pilin gene and novel morphology of the assembled appendage fibers. *J Bacteriol* 177(4): 1030-8.
- Sajjan U. S. et al., (2006). Intracellular trafficking and replication of *Burkholderia cenocepacia* in human cystic fibrosis airway epithelial cells. *Cell Microbiol* 8(9): 1456-66.
- Saldias M. S. et al., (2008). *Burkholderia cenocepacia* requires the RpoN sigma factor for biofilm formation and intracellular trafficking within macrophages. *Infect Immun* 76(3): 1059-67.
- Saldias M. S. and Valvano M. A. (2009). Interactions of *Burkholderia cenocepacia* and other *Burkholderia cepacia* complex bacteria with epithelial and phagocytic cells. *Microbiology* 155(Pt 9): 2809-17.
- Salles J. F. et al., (2002). Molecular method to assess the diversity of *Burkholderia* species in environmental samples. *Appl Environ Microbiol* 68(4): 1595-603.
- Sambrook J., E. F. Fritsch, and T. Maniatis. (1989). Molecular cloning: a laboratory manual. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.
- Samuel J. E. et al., (1990). Comparison of the glycolipid receptor specificities of Shiga-like toxin type II and Shiga-like toxin type II variants. *Infect Immun* 58(3): 611-8.
- Santos A. V. et al., (2004). Occurrence of the antibiotic producing bacterium *Burkholderia sp.* in colonies of the leaf-cutting ant *Atta sexdens rubropilosa*. *FEMS Microbiol Lett* 239(2): 319-23.
- Schneider D. et al., (2000). Long-term experimental evolution in *Escherichia coli*. IX. Characterization of insertion sequence-mediated mutations and rearrangements. *Genetics* 156(2): 477-88.
- Schneider D. et al., (2002). Genomic comparisons among *Escherichia coli* strains B, K-12, and O157:H7 using IS elements as molecular markers. *BMC Microbiol* 2: 18.
- Schneider D. and Lenski R. E. (2004). Dynamics of insertion sequence elements during experimental evolution of bacteria. *Res Microbiol* 155(5): 319-27.
- Schwab U. et al., (2002). Patterns of epithelial cell invasion by different species of the *Burkholderia cepacia* complex in well-differentiated human airway epithelia. *Infect Immun* 70(8): 4547-55.
- Scordilis G. E. et al., (1987). Identification of transposable elements which activate gene expression in *Pseudomonas cepacia*. *J Bacteriol* 169(1): 8-13.
- Seed K. D. and Dennis J. J. (2009). Experimental bacteriophage therapy increases survival of *Galleria mellonella* larvae infected with clinically relevant strains of the *Burkholderia cepacia* complex. *Antimicrob Agents Chemother* 53(5): 2205-8.
- Segonds C. et al., (1997). Genotypic analysis of *Burkholderia cepacia* isolates from 13 French cystic fibrosis centers. *J Clin Microbiol* 35(8): 2055-60.
- Sfalanga A. et al., (1999). Isolation and characterisation of a new antagonistic *Burkholderia* strain from the rhizosphere of healthy tomato plants. *Res Microbiol* 150(1): 45-59.
- Silipo A. et al., (2007). The complete structure and pro-inflammatory activity of the lipooligosaccharide of the highly epidemic and virulent gram-negative bacterium *Burkholderia cenocepacia* ET-12 (strain J2315). *Chemistry* 13(12): 3501-11.

- Simonovic I. et al., (2001). Enteropathogenic *Escherichia coli* activates ezrin, which participates in disruption of tight junction barrier function. *Infect Immun* 69(9): 5679-88.
- Sokol P. A. (1986). Production and utilization of pyochelin by clinical isolates of *Pseudomonas cepacia*. *J Clin Microbiol* 23(3): 560-2.
- Spilker T. et al., (2009). Expanded multilocus sequence typing for *Burkholderia* species. *J Clin Microbiol* 47(8): 2607-10.
- Springman A. C. et al., (2009). Genetic diversity and multihost pathogenicity of clinical and environmental strains of *Burkholderia cenocepacia*. *Appl Environ Microbiol* 75(16): 5250-60.
- Stapleton A. E. et al., (1998). The globoseries glycosphingolipid sialosyl galactosyl globoside is found in urinary tract tissues and is a preferred binding receptor *In vitro* for uropathogenic *Escherichia coli* expressing pap-encoded adhesins. *Infect Immun* 66(8): 3856-61.
- Strauss E. J. and Falkow S. (1997). Microbial pathogenesis: genomics and beyond. *Science* 276(5313): 707-12.
- Summer E. J. et al., (2004). *Burkholderia cenocepacia* phage *BcepMu* and a family of Mu-like phages encoding potential pathogenesis factors. *J Mol Biol* 340(1): 49-65.
- Sylvester F. A. et al., (1996). *Burkholderia* (basonym *Pseudomonas*) *cepacia* binding to lipid receptors. *Infect Immun* 64(4): 1420-5.
- Tablan O. C. et al., (1985). *Pseudomonas cepacia* colonization in patients with cystic fibrosis: risk factors and clinical outcome. *J Pediatr* 107(3): 382-7.
- Thomas M. S. (2007). Iron acquisition mechanisms of the *Burkholderia cepacia* complex. *Biomaterials* 20(3-4): 431-52.
- Thomassen M. J. et al., (1986). *Pseudomonas cepacia*: decrease in colonization in patients with cystic fibrosis. *Am Rev Respir Dis* 134(4): 669-71.
- Thomassen M. J. et al., (1985). *Pseudomonas cepacia* colonization among patients with cystic fibrosis. A new opportunist. *Am Rev Respir Dis* 131(5): 791-6.
- Tomich M. et al., (2003). Attenuated virulence of a *Burkholderia cepacia* type III secretion mutant in a murine model of infection. *Infect Immun* 71(3): 1405-15.
- Tomich M. et al., (2002). Role of flagella in host cell invasion by *Burkholderia cepacia*. *Infect Immun* 70(4): 1799-806.
- Tomich M. and Mohr C. D. (2004). Genetic characterization of a multicomponent signal transduction system controlling the expression of cable pili in *Burkholderia cenocepacia*. *J Bacteriol* 186(12): 3826-36.
- Toro N. et al., (2002). Identification and characterization of bacterial class E group II introns. *Gene* 299(1-2): 245-50.
- Tourasse N. J. and Kolsto A. B. (2008). Survey of group I and group II introns in 29 sequenced genomes of the *Bacillus cereus* group: insights into their spread and evolution. *Nucleic Acids Res* 36(14): 4529-48.
- Toussaint A. and Merlin C. (2002). Mobile elements as a combination of functional modules. *Plasmid* 47(1): 26-35.
- Tran Van V. et al., (2000). Repeated beneficial effects of rice inoculation with a strain of *Burkholderia vietnamiensis* on early and late yield components in low fertility sulphate acid of Vietnam. *Plant Soil* 218: 273-284.
- Tuanyok A. et al., (2008). Genomic islands from five strains of *Burkholderia pseudomallei*. *BMC Genomics* 9: 566.
- Tumapa S. et al., (2008). *Burkholderia pseudomallei* genome plasticity associated with genomic island variation. *BMC Genomics* 9: 190.
- Tyler S. D. et al., (1996). Identification of IS1356, a new insertion sequence, and its association with IS402 in epidemic strains of *Burkholderia cepacia* infecting cystic fibrosis patients. *J Clin Microbiol* 34(7): 1610-6.
- Uehlinger S. et al., (2009). Identification of specific and universal virulence factors in *Burkholderia cenocepacia* strains by using multiple infection hosts. *Infect Immun* 77(9): 4102-10.
- Urban T. A. et al., (2005). Cable pili and the 22-kilodalton adhesin are required for *Burkholderia cenocepacia* binding to and transmigration across the squamous epithelium. *Infect Immun* 73(9): 5426-37.
- Ursing J. and Aleksic S. (1995). *Yersinia frederiksenii*, a genotypically heterogeneous species with few differential characteristics. *Contrib Microbiol Immunol* 13: 112-6.
- van den Berg J. M. et al., (2009). Chronic granulomatous disease: the European experience. *PLoS One* 4(4): e5234.
- van Laer F. et al., (1998). An outbreak of *Burkholderia cepacia* with septicemia on a cardiology ward. *Infect Control Hosp Epidemiol* 19(2): 112-3.
- Van Nostrand J. D. et al., (2007). Reduction of nickel and uranium toxicity and enhanced trichloroethylene degradation to *Burkholderia vietnamiensis* PR1301 with hydroxyapatite amendment. *Environ Sci Technol* 41(6): 1877-82.

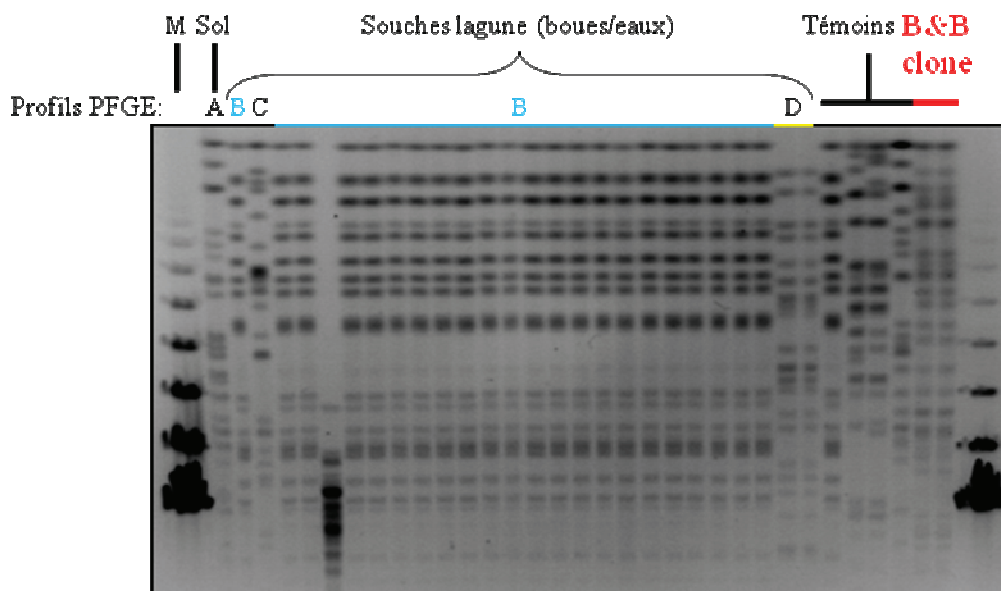
- Vandamme P. and Coenye T. (2004). Taxonomy of the genus *Cupriavidus*: a tale of lost and found. *Int J Syst Evol Microbiol* 54(Pt 6): 2285-9.
- Vandamme P. et al., (2002). *Burkholderia anthina* sp. nov. and *Burkholderia pyrrocinia*, two additional *Burkholderia cepacia* complex bacteria, may confound results of new molecular diagnostic tools. *FEMS Immunol Med Microbiol* 33(2): 143-9.
- Vandamme P. et al., (2003). *Burkholderia cenocepacia* sp. nov. - a new twist to an old story. *Res Microbiol* 154(2): 91-6.
- Vandamme P. et al., (1997). Occurrence of multiple genomovars of *Burkholderia cepacia* in cystic fibrosis patients and proposal of *Burkholderia multivorans* sp. nov. *Int J Syst Bacteriol* 47(4): 1188-200.
- Vandamme P. et al., (2000). Identification and population structure of *Burkholderia stabilis* sp. nov. (formerly *Burkholderia cepacia* genomovar IV). *J Clin Microbiol* 38(3): 1042-7.
- Vanlaere E. et al., (2009). Taxon K, a complex within the *Burkholderia cepacia* complex, comprises at least two novel species, *Burkholderia contaminans* sp. nov. and *Burkholderia lata* sp. nov. *Int J Syst Evol Microbiol* 59(Pt 1): 102-11.
- Vanlaere E. et al., (2005). A novel strategy for the isolation and identification of environmental *Burkholderia cepacia* complex bacteria. *FEMS Microbiol Lett* 249(2): 303-7.
- Vanlaere E. et al., (2008a). *Burkholderia latens* sp. nov., *Burkholderia diffusa* sp. nov., *Burkholderia arboris* sp. nov., *Burkholderia seminalis* sp. nov. and *Burkholderia metallica* sp. nov., novel species within the *Burkholderia cepacia* complex. *Int J Syst Evol Microbiol* 58(Pt 7): 1580-1590.
- Vecerek B. et al., (2007). Control of Fur synthesis by the non-coding RNA RyhB and iron-responsive decoding. *Embo J* 26(4): 965-75.
- Venter J. C. et al., (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304(5667): 66-74.
- Vermis K. et al., (2003). Isolation of *Burkholderia cepacia* complex genomovars from waters. *Syst Appl Microbiol* 26(4): 595-600.
- Vermis K. et al., (2004). Proposal to accommodate *Burkholderia cepacia* genomovar VI as *Burkholderia dolosa* sp. nov. *Int J Syst Evol Microbiol* 54(Pt 3): 689-91.
- Vermis K. et al., (2002). Evaluation of species-specific *recA*-based PCR tests for genomovar level identification within the *Burkholderia cepacia* complex. *J Med Microbiol* 51(11): 937-40.
- Vinion-Dubiel A. D. et al., (2004). Correlation of *wbiI* genotype, serotype, and isolate source within species of the *Burkholderia cepacia* complex. *J Clin Microbiol* 42(9): 4121-6.
- Vlot A. C. et al., (2002). Role of the alfalfa mosaic virus methyltransferase-like domain in negative-strand RNA synthesis. *J Virol* 76(22): 11321-8.
- Wang Y. et al., (2008). Degradation of dimethyl carboxylic phthalate ester by *Burkholderia cepacia* DA2 isolated from marine sediment of South China Sea. *Ecotoxicology* 17(8): 845-52.
- Waters L. S. and Storz G. (2009). Regulatory RNAs in bacteria. *Cell* 136(4): 615-28.
- Whitby P. W. et al., (2006). *Burkholderia cenocepacia* utilizes ferritin as an iron source. *J Med Microbiol* 55(Pt 6): 661-8.
- Whitman W. B. et al., (1998). Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* 95(12): 6578-83.
- Wilson G. G. and Murray N. E. (1991). Restriction and modification systems. *Annu Rev Genet* 25: 585-627.
- Winiecka-Krusnell J. and Linder E. (2001). Bacterial infections of free-living amoebae. *Res Microbiol* 152(7): 613-9.
- Wise M. G. et al., (1995). Genetic structure of a lotic population of *Burkholderia* (*Pseudomonas*) *cepacia*. *Appl Environ Microbiol* 61(5): 1791-8.
- Wood M. S. et al., (1991). IS406 and IS407, two gene-activating insertion sequences for *Pseudomonas cepacia*. *Gene* 105(1): 101-5.
- Wood M. S. et al., (1990). Activation of the lac genes of Tn951 by insertion sequences from *Pseudomonas cepacia*. *J Bacteriol* 172(4): 1719-24.
- Woods C. W. et al., (2004). Virulence associated with outbreak-related strains of *Burkholderia cepacia* complex among a cohort of patients with bacteremia. *Clin Infect Dis* 38(9): 1243-50.
- Worgall S. et al., (2005). Similarity of gene expression patterns in human alveolar macrophages in response to *Pseudomonas aeruginosa* and *Burkholderia cepacia*. *Infect Immun* 73(8): 5262-8.
- Yabuuchi E. et al., (1992). Proposal of *Burkholderia* gen. nov. and transfer of seven species of the genus *Pseudomonas* homology group II to the new genus, with the type species *Burkholderia cepacia* (Palleroni and Holmes 1981) comb. nov. *Microbiol Immunol* 36(12): 1251-75.
- Yabuuchi E. et al., (1995). Transfer of two *Burkholderia* and an *Alcaligenes* species to *Ralstonia* gen. Nov.: Proposal of *Ralstonia pickettii* (Ralston, Palleroni and Doudoroff 1973) comb. Nov., *Ralstonia solanacearum* (Smith 1896) comb. Nov. and *Ralstonia eutropha* (Davis 1969) comb. Nov. *Microbiol Immunol* 39(11): 897-904.

- Yoder-Himes D. R. et al., (2009). Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. *Proc Natl Acad Sci U S A* 106(10): 3976-81.
- Zahariadis G. et al., (2003). Cepacia-like syndrome caused by *Burkholderia multivorans*. *Can J Infect Dis* 14(2): 123-5.
- Zhang L. and Xie G. (2007). Diversity and distribution of *Burkholderia cepacia* complex in the rhizosphere of rice and maize. *FEMS Microbiol Lett* 266(2): 231-5.
- Zhang R. et al., (2009). Two type IV secretion systems with different functions in *Burkholderia cenocepacia* K56-2. *Microbiology*.
- Zulianello L. et al., (2006). Rhamnolipids are virulence factors that promote early infiltration of primary human airway epithelia by *Pseudomonas aeruginosa*. *Infect Immun* 74(6): 3134-47.

# Annexes

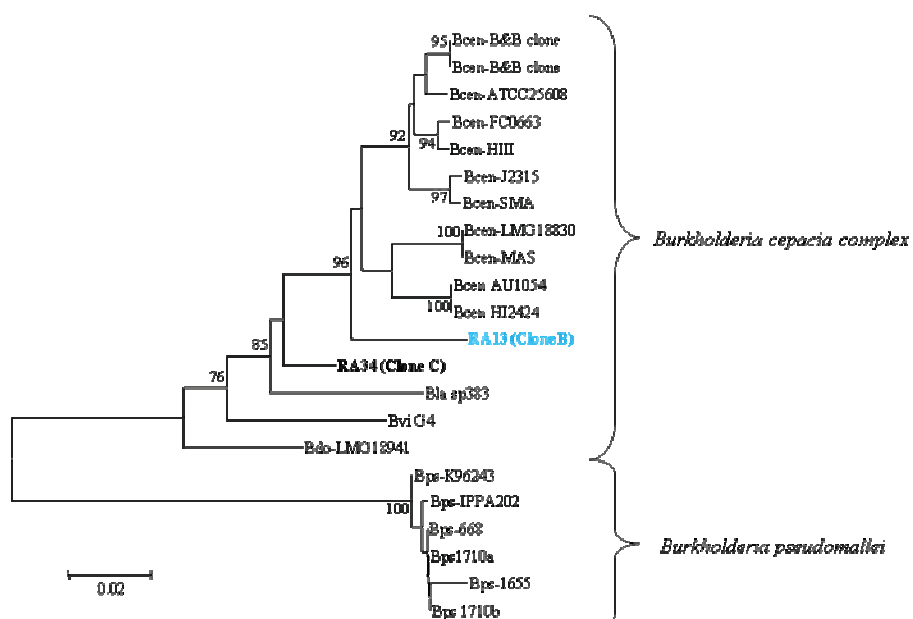
## A. Chapitre 1

### 1. PFGE des isolats obtenus près de l'hôpital Fleyriat



PFGE des souches obtenues au cours de la première campagne d'échantillonnage.

### 2. Phylogénie *ecfB* des clones B et C obtenus lors du premier échantillonnage



Neighbor-joining phylogenetic tree of *ecfB* gene sequences from the first environmental sampling and representative strains of the *Burkholderia*. DNA sequences were either obtained in this work or retrieved from the GenBank database. Distances are proportional to evolutionary divergences expressed in substitutions per 100 sites. Bootstrap values above 90% are indicated on the tree branches. Bcen = *B. cenocepacia*; Bvi = *B. vietnamiensis*; Bla = *B. lata*; Bam = *B. ambifaria*; Bdo = *B. dolosa*; Bpe = *B. pseudomallei*

### 3. Analyse MLST des clones obtenus lors du premier échantillonnage

#### Clone B :

RA 13	allèle	remarques	st
atpD	2	99.77% (mismatches: 1; gaps: 0; alignment: -441/443)	<b>4</b> <i>Burkholderia cepacia</i>
gltB	55	99.75% (mismatches: 1; gaps: 0; alignment: 400/400)	
gyrB	2	98.90% (mismatches: 5; gaps: 0; alignment: 454/454)	
lepA	1	99.75% (mismatches: 1; gaps: 0; alignment: 397/397)	
phaC	93	99.74% (mismatches: 0; gaps: 1; alignment: 385/385)	
recA	37	Identical match	
trpB	48	100.00% (mismatches: 0; gaps: 0; alignment: 281/301)	

#### Clone C :

RA 34	allèle	remarques	st
atpD	93	99.10% (mismatches: 4; gaps: 0; alignment: -441/443)	?
gltB	96	99.50% (mismatches: 2; gaps: 0; alignment: 400/400)	
gyrB	142	Identical match	
lepA	129	99.48% (mismatches: 2; gaps: 0; alignment: 383/397)	
phaC	65	99.48% (mismatches: 2; gaps: 0; alignment: 385/385)	
recA	82	Identical match	
trpB	58	100.00% (mismatches: 0; gaps: 0; alignment: 297/301)	

Résultats obtenus par analyse MLST des souches obtenues au cours de la première campagne d'échantillonnage.

## B. Chapitre 2

### 1. Séquences de la région *ecfA2*

>*ecfA2\_12614*

```
CGGGCGCAAGAACATACAAGCGCGCCGCGCGACCCACCGCCCCACACAAGAGTGATAAGAGAAAAGTCTCATG
AGTGAATCAAGGAAATGCCGCTGGTTCGAGCTGACGGCCAAGGATCTTCCCGCCTACTGCCCCAACCCGGCCA
TGACCCGCTGGAACGCGCACCCGCGCGTCTTCATCGACGTGTGCGACGGCGAGGCGCGCTGCCCGTACTGCGG
CACGCGCTACAAGTGTGCGGACGGCGAGGTCTCAAGGGCCACTGAGCCCCGGCGGCCGGCCTGAAGGCCCA
CCGACCCGCAAGGCGGCGCAGCCGGAAGTCCGGCGCGCCGCCGTTCCCCATTCTGGCAACCCGAACGCGGCGC
CCGAGCGCCGTGCTTCATTACGACATCGGAAGTCGACCTGATGCGTCGAGCGCTGGTTATCGCACCCGAAAG
```

>*ecfA2\_21461*

```
CGGGCGCAAGAACATACAAGCGCGCCGCGCGACCCACCGCCCCACACAAGAGTGATAAGAGAAAAGTCTCATG
AGTGAATCAAGGAAATGCCGCTGGTTCGAGCTGACGGCCAAGGATCTTCCCGCCTACTGCCCCAACCCGGCCA
TGACCCGCTGGAACGCGCACCCGCGCGTCTTCATCGACGTGTGCGACGGCGAGGCGCGCTGCCCGTACTGCGG
CACGCGCTACAAGTGTGCGGACGGCGAGGTCTCAAGGGCCACTGAGCCCCGGCGGCCGGCCTGAAGGCCCA
CCGACCCGCAAGGCGGCGCAGCCGGAAGTCCGGCGCGCCGCCGTTCCCCATTCTGGCAACCCGAACGCGGCGC
CCGAGCGCCGTGCTTCATTACGACATCGGAAGTCGACCTGATGCGTCGAGCGCTGGTTATCGCACCCGAAAG
```

>*ecfA2\_19240*

```
GAGTGATAAGAGAAAAGTCTCATGAGTGAAATCAAGGAAATGCCGCTGGTTCGAGCTGATGGCCAAGGATCTTCC
CGCTACTGCCGAACCCGGCCATGACCCGCTGGAACGCGCACCCGCGCGTCTTCATCGACGTGTGCGACGGC
GAGGCGCGCTGCCCGTACTGCGGCACGCGCTACAAGTGTGCGGACGGCGAGGTCTCAAGGGCCACTGAGCCC
GGCGGCCGGCCTGAAGGCCACCGACCCGCAAGGCGGCGCAGCCGGAAGTCCGGCGCGCCGCCGTTCCCCA
TTCTGGCAACCCGAACGCGGCGCCCGAGCGCCGTGCTTCATTACGACATCGGAAGTCGACCTGATGCGTCGAG
CGCTGGTTATCGCACCCGAAAG
```

>*ecfA2\_B&B clone*

```
CGGGCGCAAGAACATACAAGCGCGCCGCGCGACCCACCGCCCCACACAAGAGTGATAAGAGAAAAGTCTCATG
AGTGAATCAAGGAAATGCCGCTGGTTCGAGCTGACGGCCAAGGATCTTCCCGCCTACTGCCCCAACCCGGCCA
TGACCCGCTGGAACGCGCACCCGCGCGTCTTCATCGACGTGTGCGACGGCGAGGCGCGCTGCCCGTACTGCGG
CACGCGCTACAAGTGTGCGGACGGCGAGGTCTCAAGGGCCACTGAGCCCCGGCGGCCGGCCTGAAGGCCCA
CCGACCCGCAAGGCGGCGCAGCCGGAAGTCCGGCGCGCCGCCGTTCCCCATTCTGGCAACCCGAACGCGGCGC
CCGAGCGCCGTGCTTCATTACGACATCGGAAGTCGACCTGATGCGTCGAGCGCTGGTTATCGCACCCGAAAG
```

### 2. Séquences de la région *ecfE* des souches de *B. cenocepacia*

>*région ecfE-Bcen\_18830*

```
TCATTCCGGCGCCGATCCGCGACCCGGACGATGCCGAGGACGTTCTGCAGGATGTGCTTCACGAATTCGTACA
GGCTTATCGACTCCCTGCGCCGATCGAACAGGCGAGCGGTGGCTGTCCGTACCGCGCGCAATCGCATCATC
GACCGTTTTTCGAAGAAGAAGGAGCAGCCGCTGACCGATCTGCTCGATGCCGACGACGAAGCGGACAGTGAG
TATCGCCTGGATCTTGCACTACCGTCGACGATGCCGGCCCCGAGGCGCTCTACGCTCGCACGCTGTTGCTCAA
GGCGTTGCAGGATGCGCTCGACGAATTGCCGGCGAATCAGCGCGAGTTTTTCATCGCGCACGAGCTGGAGGGG
CAGTCGTTCAAGGAGATGGCCGCGCAGAGCGGCGTACGCTCAATACGCTGCTCGCGCGCAAACGCTACGCGG
TCCTGCATCTGCGTGTCTGACTGCAACCCATTTATGACGAACCTGGATAACTAGAGGAGTTCGACGGATGAATTTT
CGAATCAGATGTGCAGGCAAAGCGCTGCTCGTGGTGTATCGCCATCGGTGTGCTCGGATGGGTGCTCATGATGC
TGTGGAATGGGTGATTCGGGGTTATTCGTCGTTGCCCGACGATCGACTTCGCCCATCGCTGGGCTGCTC
GTGTTGAGCCGATCCTGTTCCGGCGATTCCGCGGACATCCGCGGATCCGCGCTGCGACGGCCGACGCTGGC
GGGAAGCCATGACGCTGAAGAACGCGAGCAGTTTTGGGGCCGATGGCGATCGCGCGGCAAGCGGGCCACGG
AGGAGTGACATGCGCGATAAATCAACGGCCGACT
```

>*région ecfE-Bcen\_18863*

```
CATTCGGCGCCGGATCCGCGACCCGGACGATGCCGAGGACGTTCTGCAGGATGTGCTTCACGAATTCGTACAG
GCCTATCGACTTCCCGCGCAATCGAACAGGCGAGCGGTGGCTGTCCGTACCGCGCGCAACCCGCATCATG
ATCGTTTTTCGAAGAAGAAGGAGCAACCCGCTGACCGATCTGCTCGACGCCGACGACGAAGCGGACAGCGAT
ATCGCTTGATCTGGCTGCTGCCGGCGCACGATGCCGGCCCCGAGGCGCTCTACGCTCGCACGCTGTTGCTCAA
GGCGTTGCAGGATGCGCTCGACGAGTTGCCGGCGAATCAGCGCGAGTTTTTCATCGCGCACGAGCTGGAAGGA
CAGTCCTTCAAGGAGATGGCCGCGCAAAGCGGTGTACGCTCAATACGCTGCTCGCGCGCAAACGCTACGCGG
TCCTGCATCTGCGTGTCTGACTGCAACCCATTTACGACGAACCTGGATATCTAGAGGAGTTCGACGGATGAATTTT
CGAATCAGATGTGCGGGCAAGGCGCTGCTCGTGGTGGTGGCCATCGGCGTGTCTCGGTGGGGTCTGTGATGATG
TGTGGAATGGGTGATGCCGGCGCTATTCGTTGGCGCCCGGACGATCGACTTCGCTCATGCGCTGGGCTGCTC
GTGTTGAGCCGATCCTGTTCCGGCGATTCCGCGGACACCCGGCTGGCACGGCCGGCCACTGGCGCAAGT
GGGAAGCCATGACGCGAAGAGCGGAGCAGTTCCGGGCCGATGGCGGTGCGCGGCAAGCGGGCCACCGG
AGGAGTGACCATGCGCGAGAAATCAGCGGCCGACTGCAAGCAGAACGCGGGCGTGATTGCGCCGGTCTCGA
TCTGAAGCGTTGCGAAGGCAAGGGCGATTGCGTACCGTTTTGCC
```

>*région ecfE-Bcen\_12614*

```
GCTAGGGAATTTCAATTCGGCGCCGGATCCGCGACCCGGACGATGCCGAGGACGTTCTGCAGGATGTGCTTAC
GAATTCGTACAGGCTATCGACTTCCCGCGCAATCGAACAGGCGAGCGCGTGGCTGTTCCGTACCGCGCGCA
ACCGCATCGATCGTCTTTTCGAAGAAGAAGGAGCAACCCGCTGACCGATCTGCTCGACGCCGACGACGACG
GGACAGCGGATATCGGCTGATCTGGCGCTGCCGGCGCACGATGCCGGCCCCGAGGCGCTCTACGCTCGCACG
CTGTTGCTCAAGGCGTTGCAAGGATGCGCTCGACGAGTTGCCGGCGAATCAGCGCGAGGTTTTTCATCGCGCACG
AGCTGGAAGGACAGTCTTCAAGGAGATGGCCGCGCAAAGCGGTGTACGCTCAATACGCTGCTCGCGCGCAA
ACGCTACGCGGTCTGCATCTGCGTGTCTGACTGCAACCCATTTACGACGAACCTGGATATCTAGAGGAGTCCA
CGGATGAATTTTCAATCAGATGTGCGGGCAAGGCGCTGCTCGTGGTGGCCATCGGCGTGTCTCGGTGGG
```



TCGTGATGATGTTGTGGAAGTGGGTGATGCCGGCGCTATTTCGTTGGCGCCCGGACGATCGACTTCGCTCATGCC  
CTGGGCTGCTCGTGTGAGCCGCATCCTGTTCCGGCGGATCCCGCGGACACCGCGGCTGGCACGGCCGGCGCC  
ACTGGCGCAAGTGGGAAGCCATGACGCCGGAAGAGCGCGAGCAGTTCCGGGCCGCATGGCGGTTCGCGCGGA  
AGCGGCCACGGAGGATGACCATGCGCGAGAAATCAGCGGCCGACTGCAAGCAGAACCGCGGGCGTGATT  
>région ecfE-Bcen\_21461  
CGAACGCGGATCACGCTTCCCCGAAGGGCATAGCGTCTCCCGTGCACGCCGATCATGGATCGGCAGCGCGG  
CAGGCTGACGGCAGGTAGCAGTCGNNGGATATCGCCGTCGACGTGATCAACTTCCTCAAGGAACGCACTTGG  
CTTCCACCACCCACGCTCTTTTGAACGGCTCCATGAAGAGAGCCAGCGGAAATCATCGCCCGCGCACGCGC  
GCAGTTCAGGGGCCGAACGAGGCGTTTCCGACGCTCGTCGACGAATACGAAGCACTCGCCCCGAGGCTGCGC  
GGGCGGTTCTCGAAACGTTCCGGCTTGTTCACCGAAGATTTTCGAGTCTTCGGCAGGCAGTGCTCTATCGTTGGC  
CGTGTCCGGGAAACGGGACGATTCTTGCACAACATGGTGTGCACAACGGCCGGCCAGAATCCTCGAGCTC  
GGCAGTTCGTGCGCGGTGTCGACGCTGATTTTGGCGATGCGTTGCGGACGCTGGGCGGTGGCGTGGTTCATCG  
GACCGAATGCGATGCGTTCAAGTGCAGCGGTTGCGGGCTCATGTGAAACAGCGGGCCTGGGCGCCAGTGC  
ACGTGCGGGAGGGCGATGTATTTCGAGACCGT  
>région ecfE-Bcen\_21461  
TTCATTCGGCGCCGGATCCGTGACCCGGACGACGCCGAGGACGTTCTGCAGGATGTCCTTCACGAATTCGTACA  
GGCTTATCGACTTCCTGCGCCGATCGAGCAGGCGAGCGCGTGGCTGTTCCGTGCCGCGCGTAACCGCATCGTGC  
ATCGTTTTTCGAAGAAGAAGGAGCAGCCGCTGACCGACCTGCTCGATGCCGACGACGAAGCCGACAGCGAGT  
ATCGCCTGGATCTCGCTACCGGCGCACGATGCGGGGCGTGAAGCGCTCTACGCTCGCACGCTGTTGCTCAA  
GGCGTTGCAGGACGCGCTCGACGAATTGCCGGCGAATCAGCGCGAGGTTTTTCATCGCGCACGAAGTGGACGGG  
CTGTCCTTCAGGGAGATGGCCGCGCAAAGCGGTGTACGCTCAATACGCTGCTCGCGCGCAAACGCTACGCGG  
TTTTGCATCTGCGTGTCTGACTGCAACCCATTTATGACGAAGTGGATATCTAGAGGAGTCCGACGGATGAATTTT  
CGAATCAGATGTGCGGGCAAGGCGCTGCTCGTGGTGGTGGCGATCGGCGTGTCTGGGCTGGGTCGTCATGATGC  
TGTGGAAGTGGGTGATGCCGGCGCTATTTCGTGCGCGCGCGGACGATCGACTTCGCTCATGCGCTGGGCTGCTC  
GTGTTGAGCCGCATCCTGTTCCGGCGATTTCGCGGACACCGCGGCTGGCACGGCCGGCGCCACTGGCGCAAGT  
GGGAAGCCATGACGCCCGAGGCGGAGGTTCCGCGCCGATGGCGATCGCGCGGCAAGCGGCCCGCGC  
AGGAGTGAACATGCGCGAGAAAGCAACGGCCGACT  
>région ecfE-Bcen\_19240  
GAGTGAACATGCGCGAAAAATCATCGGACGACTGCAAACAGAAGCGGGTGTGATTGCGCCGGTTCGTCGATCT  
GAAACGTTGCGAAGGCAAGGGGATTCGCTCACCGTTTGCCCGGAAAACGTTTCGAGGTGCGGCGCATCGAT  
CAGGCCGATTACGTCGACCTCGGTCTGATGCATCGCTTGAAGTGCAGTGGAAATGAAAGTGCCTACA  
CGCCGAATGCGCAAGCCTGCCGATCATGCGGGCTTTGCGTGACGGCGTGTCCCGAACGCGCGATACGCTCGC  
CAGAAGGCGTAGCCTGCGCTGCTCGCAGATGATGAGCGGTATCGCGGAGTTCGCGCGGATCGGCGGATCGCAGCCG  
CTGGATACGCCGTTTCATGTGATCAATCTCGTCAAGGAATGCACTTGGCTTCAACCATCGACGTTCTTCTGCAAC  
GGCTTCATGAAGAAAGCCGAGCGAAATCATCGCCCGGGCCCGCGCGCAGTTCCAGGGGCCGAATGCGGCGTT  
TCCGACGCTCGTCGACGAATACGAAGCGCTCGCACCCCGGCTGCGAGGGCGCTTTCTCGAAACGTTCCGGCCTCT  
TCAGCGACGATTCGAATCGTCGAAAGCGGTGCGCTATCGTTGGCCGTGTCGGCGGAAACGGGACGATTCTT  
GCGCAACACCGTGTGTCGCAACGGTTCGGGCGAAGTCTTGAAGTGGCAGTTCCTGTGGCGTGTGACTCTGT  
ATTTCCGCGACGCGCTGCGTATGTTGGGCGGTGGCGTGGTTCATCGCGACCGAACGTTGATGCGCTCAAATGCGC  
GCGTTTTGCGGCTCACGTCGAAACAGCAGGGCTGGACGCCTATGTGACTTGGCGGAGGGCGATGTGTTTCGAG  
ACGGTTGCCGAATCGACGGAC  
>région ecfE-Bcen\_19240  
GGACTTCATTCCGGCGCCGATACCGGACCCGGACGATGCCGAGGACGTTCTGCAGGATGTGCTCCACGAATTC  
GTCCAGGCTTATCGACTTCCCGCGCCGATCGAACAGGCGAGTGCCTGGTGTCCGTGCCGCGCGCAACCCGCA  
TCATCGATCGTTTTTCGAAAAGGAAGGAGCAGCCGCTGACCGATCTGGTTCGATGCCGACGACGAAGCGGACAG  
CGAATATCGGCTGGATCTCGGTTACCGGCGCACGATGCCGGCCCTGAAGCGCTCTACGCTCGCACGCTGTTGC  
TCAAGGCGTTGACGATGCGCTCGACGAGTTGCCGGCGAATCAGCGCGAGGTTTTTATCGCACACGAAGTGA  
GGGCTTCTCTTCAAGGAGATGGCCGCGCAAAGCGGCGTACGCTCAATACGCTGCTTGGCGCAAACGCTAC  
GCGGTCTTGCATCTGCGTGTGCGGCTGCAACCCATTTATGACGAAGTGGATAGCTAGAGGAGTCCGACGGATGA  
ATTTTCGAATCAGATGTGCGGGCAAGGCGCTGCTCGTGGTGGTCCGATAGGCGTGTGCGGATGGGTGCTCAT  
GATGCTGTGGAAGTGGTGGTTCGGGCGCTATTCGTCGGCGCCCGGACGATCGACTTCGCCATGCGCTGGGCG  
TCCTCGTGTGAGCCGCATCCTGTTCCGTGGATTTCCGCGACACCGCGGCTGGCACGGCCGGCGCCACTGGGCG  
AAGTGGGAAGCCATGACGCCGGAAGAGCGCGAGCAGTTCCGGGCCGATGGCGATCGCGCGCACCCCGCGCA  
CGGAGGAGTGAACATGCGCGAAAAATCATCGGACGACT  
>région ecfE-Bcen\_B&Bclone  
CGGCGCCGGATCCGCGACCCGGACGATGCCGAGGACGTTCTGCAGGATGTGCTTACGAATTCGTACAGGCCT  
ATCGACTTCCCGCGCCAATCGAACAGGCGAGCGGTGGTGTTCGTTACCGCGCGCAACCGCATCATCGATCG  
TTTTCGCAAGAAGAAGGAGCAACCGCTGACCGATCTGCTCGACGCCGACGACGAAGCGGACAGCGAGTATCG  
CCTGGATCTGGCGCTGCCGCGCACGATGCCGGCCCCGAGGCGCTCTACGCTCGCACGCTGTTGCTCAAGGCG  
TTGCAGGATGCGCTCGACGAGTTGCCGGCGAATCAGCGCGAGGTTTTTCATCGCGCACGAGCTGGAAGGACAGT  
CCTTCAAGGAGATGGCCGCGCAAAGCGGTGTACGCTCAATACGCTGCTCGCGCGCAAACGCTACGCGGCTCT  
GCATCTGCGTCTCGACTGCAACCCATTTACGACGAAGTGGATATCTAGAGGAGTCCGACGGATGAATTTTCGA  
ATCAGATGTGCGGGCAAGGCGCTGCTCATGGTGGTGGCCATCGGCGTGTGCTGGGTTGGGCTGATGATGTTGT  
GGAAGTGGGTGATGCCGGCGCTATTCGTTGGCGCCCGGACGATCGACTTCGCTCATGCGCTGGGCTGCTCGT  
TTGAGCCGCATCCTGTTCCGGCGATTCCGCGGACACCGCGGCTGGCACGGCCGGCGCCACTGGCGCAAGTGGG  
AAGCCATGACGCCGGAAGAGCGCGAGCAGTTCCGGGCGCATGGCGGTCCGCGCGGAAGCG

### 3. Séquences de la région *ecfE* contenant un phage chez *B. vietnamiensis*

> phage-ecfE-Bvi\_LMG16232

```
TGGTCAGCTGCTGCGGTACAACAGCCGGTTGCCGTTGGTCGTGTATGTGCCGGAAGGGTTTGAGGTGCGGTATC
GCGTCTGGCATGCGGCGAAGGATGTGGGGCAGGCGGGGATGCAGTGATGCTCGCGACGGCTCCCGCCGTCGCG
CGCCTGAATCGACATTCGCGGGTTTGCCGTCACGTGGCAGCGCGCAATCGCATCGAGGACCGCGTGGGCTGAG
AACTATGCCGTGCGGTGACGCTCCCCAAGTTTCCGGACGCGTGGTTGACCTACCGAAGAAAGCCAAACTTCGCC
GATCGACAACGGTAGGTCGATCATCAACGATCATGAATCGCTGTCCACGCCATGAAGCGCCAGTGTTCGAATG
AGAAAAATGACGGAGTAAGTTTCAATGAGTTTTCACACTCAGGTTTGGGTTTCATCCGTCTGACGGCGAGCCGT
TGAATGTTGCAGGTGTCCGCGAGGAGCTCTTGAAGTTTCTTGATGAATATTCGCTCAGCCACGACGTATTGACC
GACCTCACCGAAGCATGTGGTTCCGGCCCTCCCGACTTCAACTTTGTTCTATTTGGACAGCATGGCGATCGTCAT
GCTGTTGACAAAATTGGCTTCATGCCTACCGAATGTGTCGTTCCGAGTAAAGGGTTCTGGCGAGGATCACCGCG
ATATATGGGTGCGAGAGTATGCCGATGGAGTGGCCACGTTTGAAGCGAGGCCGTTTCGAGGAGTGACGTGGGG
GCAGTCATCTGATTACGTTTACGATGTGCCATCGGGCGCTCTACGACGTTCCGGCCGAAC
```

>phage-ecfE-Bvi\_LMG16228

```
TGGTCAGCTGCTGCGGTACAACAGCCGGTTGCCGTTGGTCGTGTATGTGCCGGAAGGGTTTGAGGTGCGGTATC
GCGTCTGGCATGCGGCGAAGGATGTGGGGCAGGCGGGGATGCAGTGATGCTCGCGACGGCTCCCGCCGTCGCG
CGCCTGAATCGACATTCGCGGGTTTGCCGTCACGTGGCAGCGCGCAATCGCATCGAGGACCGCGTGGGCTGAG
AACTATGCCGTGCGGTGACGCTCCCCAAGTTTCCGGACGCGTGGTTGACCTACCGAAGAAAGCCAAACTTCGCC
GATCGACAACGGTAGGTCGATCATCAACGATCATGAATCGCTGTCCACGCCATGAAGCGCCAGTGTTCGAATG
AGAAAAATGACGGAGTAAGTTTCAATGAGTTTTCACACTCAGGTTTGGGTTTCATCCGTCTGACGGCGAGCCGT
TGAATGTTGCAGGTGTCCGCGAGGAGCTCTTGAAGTTTCTTGATGAATATTCGCTCAGCCACGACGTATTGACC
GACCTCACCGAAGCATGTGGTTCCGGCCCTCCCGACTTCAACTTTGTTCTATTTGGACAGCATGGCGATCGTCAT
GCTGTTGACAAAATTGGCTTCATGCCTACCGAATGTGTCGTTCCGAGTAAAGGGTTCTGGCGAGGATCACCGCG
ATATATGGGTGCGAGAGTATGCCGATGGAGTGGCCACGTTTGAAGCGAGGCCGTTTCGAGGAGTGACGTGGGG
GCAGTCATCTGATTACGTTTACGATGTGCCATCGGGCGCTCTACGACGTTCCGGCCGAAC
```

>phage-ecfE-Bvi\_bpoe736

```
TGGTCAGCTGCTGCGGTACAACAGCCGGTTGCCGTTGGTCGTGTATGTGCCGGAAGGGTTTGAGGTGCGGTATC
GCGTCTGGCATGCGGCGAAGGATGTGGGGCAGGCGGGGATGCAGTGATGCTCGCGACGGCTCCCGCCGTCGCG
CGCCTGAATCGACATTCGCGGGTTTGCCGTCACGTGGCAGCGCGCAATCGCATCGAGGACCGCGTGGGCTGAG
AACTATGCCGTGCGGTGACGCTCCCCAAGTTTCCGGACGCGTGGTTGACCTACCGAAGAAAGCCAAACTTCGCC
GATCGACAACGGTAGGTCGATCATCAACGATCATGAATCGCTGTCCACGCCATGAAGCGCCAGTGTTCGAATG
AGAAAAATGACGGAGTAAGTTTCAATGAGTTTTCACACTCAGGTTTGGGTTTCATCCGTCTGACGGCGAGCCGT
TGAATGTTGCAGGTGTCCGCGAGGAGCTCTTGAAGTTTCTTGATGAATATTCGCTCAGCCACGACGTATTGACC
GACCTCACCGAAGCATGTGGTTCCGGCCCTCCCGACTTCAACTTTGTTCTATTTGGACAGCATGGCGATCGTCAT
GCTGTTGACAAAATTGGCTTCATGCCTACCGAATGTGTCGTTCCGAGTAAAGGGTTCTGGCGAGGATCACCGCG
ATATATGGGTGCGAGAGTATGCCGATGGAGTGGCCACGTTTGAAGCGAGGCCGTTTCGAGGAGTGACGTGGGG
GCAGTCATCTGATTACGTTTACGATGTGCCATCGGGCGCTCTACGACGTTCCGGCCGAAGCCGAGTTTA
```

>phage-ecfE-Bvi\_bpoe737

```
TGGTCAGCTGCTGCGGTACAACAGCCGGTTGCCGTTGGTCGTGTATGTGCCGGAAGGGTTTGAGGTGCGGTATC
GCGTCTGGCATGCGGCGAAGGATGTGGGGCAGGCGGGGATGCAGTGATGCTCGCGACGGCTCCCGCCGTCGCG
CGCCTGAATCGACATTCGCGGGTTTGCCGTCACGTGGCAGCGCGCAATCGCATCGAGGACCGCGTGGGCTGAG
AACTATGCCGTGCGGTGACGCTCCCCAAGTTTCCGGACGCGTGGTTGACCTACCGAAGAAAGCCAAACTTCGCC
GATCGACAACGGTAGGTCGATCATCAACGATCATGAATCGCTGTCCACGCCATGAAGCGCCAGTGTTCGAATG
AGAAAAATGACGGAGTAAGTTTCAATGAGTTTTCACACTCAGGTTTGGGTTTCATCCGTCTGACGGCGAGCCGT
TGAATGTTGCAGGTGTCCGCGAGGAGCTCTTGAAGTTTCTTGATGAATATTCGCTCAGCCACGACGTATTGACC
GACCTCACCGAAGCATGTGGTTCCGGCCCTCCCGACTTCAACTTTGTTCTATTTGGACAGCATGGCGATCGTCAT
GCTGTTGACAAAATTGGCTTCATGCCTACCGAATGTGTCGTTCCGAGTAAAGGGTTCTGGCGAGGATCACCGCG
ATATATGGGTGCGAGAGTATGCCGATGGAGTGGCCACGTTTGAAGCGAGGCCGTTTCGAGGAGTGACGTGGGG
GCAGTCATCTGATTACGTTTACGATGTGCCATCGGGCGCTCTACGACGTTCCGGCCGAAGCCGA
```

>phage-ecfE-Bvi\_ATCC53617

```
AACCGAATACCAGAGCTGGGTCTGCCTGATTTGCGGGTGGGTCTACAACGAAGCGGAAGGCCCTGCCGACGAC
GGCATCCCCGCCGACGCGTTTCGCCGATATCCCCGCCGACTGGCGATGCCCGCTGTGCGATGTGGGCAAGG
AAGATTTTCGTGGTTGTGATTTCTGAGGCGGGTTTCTGAGGCGCAGCTTCTGAGGCGATTGCGCGGAGCGAAT
CGTATTCGCTCCGCCGGCGCGCCGCGTTTGTATACTCTGCGCGCTGTCCACCTTGCCGTCCCGGTGCGGTCT
TGAGCACACGATATCGCGGACATGGCCCTCCCCGATGTTCAATGGATAGAACAAGCGCCTCCTAAGCGCTAGA
TACAGGTTGATTCCTGTGCGGGGGACCAAGATCCTCCAAAGTTTACCAGATTCGCCACAAGAACCCCGTC
AGGCTTGGCTTTACGGGTTTTTTGTTGCAAACCTCCACCAGCGACCACAGTGGAAACAACCCCGGATTTGGG
GTTCTTTTGGGGCTTCGCAAAACACCCCTTCAGACAACCCCGAGGCACCAGCATATGGCGCTACCGACCT
CAAGGTCCGCACGGCCAAGCCAGCGGATAAG
```

### 4. Séquences de la région *ecfI* délétée chez *B. vietnamiensis*

>indel-ecfI-viet\_LMG10923

```
GCGTCTCCTCGCGCAGCCGCACCATCTCGAACGCCGCGCGGCTGCGCGGGGAGATCGTCGAGCGCGGCGCA
CACGCGGGCAGCGCGTCGCGCGTCATCAGCGCGGCCCTCGGGCGTGGGCTCCGGCGACGGCACGTGCAAGCCG
TCGTCTCGTCCGTGTGATGGATGCTCTCGAACTTTGCCGGCGCAGCGCGTCGATCGACGCATTGCGACCAT
```

CCGCGTCACGTACGCGACCGGCTGGCGCACCGCGTCTGGTTTCGGAACTCGACGAGCTTGACGAACACGTCG  
TGTACGACGTCTCGGCGCGGCTCGCGCAGCCGACGAAGCCGCGCGACGTTACCAGCATCGCGGATGCC  
CGACCAGCAGCTCAGCAGCGCGCCGTGCGCGTGCGGCGCGCCGCGGCCCGCGCGCACGGCCGCGCGCGGCG  
CAAGATCGGGACAACCTGCCGAGGAACGCATTGCGCCGTGCCGCCGCCGCCGTTTCGAGCA

>indel-ecfI-viet\_LMG16232

CGTCTGTGTGATGGATGCTCTCGAAACTCTGCCGGCGCAGCGCGTTCGATCGACGCATTGCGCACCATCCGCGTC  
ACGTACGCGACCGGCTGGCGCACCGCGTCTGGTTTCGGAAACTCGACGAGCTTGACGAACACGTCGTGCACGA  
CGTCTCGGCGCGGCTCGCGCAGCCGACGAAGCCGCGCGACGTTACCAGCATCGCGGATGCGCGACCAG  
CACGTCGAGCAGCGCGCCGTGCGCGTGCGGCGCGCCGCGGCCCGCGCGCACGGCCGCGCGCGGCGCAAGATC  
GGGACAACCTGCCGAGGAACGCATTGCGCCGTGCCGCCGCCGCCGTTTCGAGCACTTGCGCCATTGCCATCGAT  
CCGCTCCATCAGTTTCGAGATGGGGCAATTTAACGTAACGCAAAATCATTCTCAATCCGATTAACGGAAGCC  
CTTCGATTTTGTTCGGGATTTGTCTGATCAAAGTGCGGCGGGGGCCGCCGTTGCTGCGTGTGCCGCCGCCGCG  
CGGGCGGATCACGCGCGCGGCTCCGCGCCTTCGGCGGGCGTCCAGCCGCCCGGAGCGCGGATAACAGCGCG  
ATCGCGTTCGCGAGGCGCAGCTCCTTCAGCCGATCAGCGCCTGGCCGACTCGTAGGTGCTGCGCTGCGCGTC  
GAGCAGCTCGAGGTAGGTC

>indel-ecfI-viet\_bpoe737

TCCGTGTGATGGATGCTCTCGAAACTCTGCCGGCGCAGCGCGTTCGATCGACGCATTGCGCACCATCCGCGTCAC  
GTATGCGACCGGCTGGCGCACAGCGTCTGGTTTCGGAAACTCGACGAGCTTGACGAACACGTCGTGCACGACG  
TCCTCGGCGCGGCTCGCGCAGCCGACGAAGCCGCGCGACGTTACCAGCATCGCGGATGCGCGACCAGCA  
CGTCGAGCAGCGCGCCGTGCGCGTGCGGCGCGCCGCGGCCCGCGCGCACGGCCGCGCGCGGCGCAAGATCGG  
GACAACCTGCCGAGGAACGCATTGCGCCGTGCCGCCGCCGCCGTTTCGAGCACTTGCGCCATTGCCATCGATCC  
GCTCCATCAGTTTCGAGATGGGGCAATTTAACGTAACGCAAAATCATTCTCAATCCGATTAACGGAAGCCCTT  
CGATTTTGTTCGGGATTTGTCTGATCAAAGTGCAGGGGGCCGCCGTTGCTGCGTGTGCCGCCGCCGCGCG  
GGCGGATCACGCGCGCGGCTGGGGCGCCTTCGGCGGGCGTCCAGCCGCCCGGAGCGCACGATAACAGCGCGAT  
CGCGTTCGCGAGGCGCAGCTCCTTCAGCCGATCAGCGCCTGGCCGACTCGTAGGTGCTGCGCTGCGCGTCC  
AGCAGCTCGAGGTAGGAC

>indel-ecfI-viet\_bpoe736

CGTCCGTGTGATGGATGCTCTCGAAACTCTGCCGGCGCAGCGCATCGATCGACGCATTGCGCACCATCCGCGTC  
ACGTACGCGACCGGCTGGCGCACCGCGTCTGGTTTCGGAAACTCGACGAGCTTGACGAACACGTCGTGCACGA  
CGTCTCGGCGCGGCTCGCGCAGCCGACGAAGCCGCGCGACGTTACCAGCATCGCGGATGCGCGACCAG  
CACGTCGAGCAGCGCGCCATGCGCGTGCGGCGCGCCGCGGCCCGCGCGCACGGCCGCGCGCGGCGCAAGATCGG  
GGGACAACCTGCCGAGGAACGCATTGCGCCGTGCCGCCGCCGCCGTTTCGAGCACTTGCGCCATTGCCATCGAT  
CCGCTCCATCAGTTTCGAGATGGGGCAATTTAACGTAACGCAAAATCATTCTCAATCCGATTAACGGAAGCC  
CTTCGATTTTGTTCGGGATTTGTCTGATCAAAGTGCAGGGGGCCGCCGTTGCTGCGTGTGCCGCCGCCGCG  
CGGGCGGATCACGCGCGCGGCTGGGGCGCCTTCGGCGGGCGTCCAGCCGCCCGGAGCGCACGATAACAGCGCG  
ATCGCGTTCGCGAGGCGCAGCTCCTTCAGCCGATCAGCGCCTGGCCGACTCGTAGGTGCTGCGCTGCGCGTCC  
CGAGCAGCTCGAGGTAGG

## 5. Séquences de la région *ecfI* délétée chez les *B. cenocepacia* PHDC

>indel-ecfI\_PHDC\_LMG21461

CGTGATGCTCGACGCGGTGCTGAAGTCGTTCCGACCGCCGAAGATCGGGCCGTTCAACGCGTCGGCGACGATC  
CGGATTTTCTGTACTGAGCGCGGTGCGGCATGAGCGCGTGCGCCGCGCAACATGCTGCCCGCACGGGATTT  
TCATTGGTGTACCGACGCAGGAGAAGGCGGCGATGCGGCGAAGCGTCTGCCGCATGCCCGTCATCCCGGCACG  
GCTCGCGTCTGCGCGGCTGTTGTTGCCGGCCATGCGCTAGCGAGCATGGTCGCGTGCAGGTGAAGCGATGAAA  
CGAGCGCGCGGTTTGCGGCGCCTCGTTTCCGTTTACCGCACCCAGTGACACACCCGGTGTGTTGATGCTCGAAG  
TGACACACGCGGTGCGGATGCGCTTCGGCGATGGCCGGCACGAGCACGGCAGCGCGCCACGGCGCGCGCGG  
AGGGTTGTCAGTAGCATCTTCATGTCGAGTCTGTCGTTGAGTGATGGATGACGTGAGCGCAGCGGCATCAC  
CACTGCGGCCGTCGTTGTCGCGGCGGTGCTCCCGGTAGTCTGCCCGGATCGTACGCGGGCCGGCATGCC  
GGACTTCCAGTCGCGGCGTCCCAGTAGCGATTTCCGTTCCAGTCACGATCGCCATGCCAGCCGATCGTGACT  
TCGGCCGGGCCAGGGTAGGCGACGCAGCCGTTGAGCATCACGCTCGACCGCGCGCCACCATGACGGCGGAT  
AACACGATGGACGTCATGTCGTTCTCCGTTGGGAACGGCATGTCGATCGCACGAGCCCTCAAATTTACGCG  
GTAAGAAGTTGCTACCGCCGATGACACGTGCAGGCATTCGTCATGAATGTGAAAGCATGCGTCCGAGAAATGA  
AGGGAGCCCGTGCGGCGCAGACGTGTACCGGTCCGACCCGCGACGATTCCGGCGATTGTTTCAGACAAATCCA  
GACAACCGGAAGGGTCATCTCGTAATCGGATTGAGAATGATTTGCGTTTACGTTAAATTGCGCTATCTCGGAAT  
TTGACGGAGCAGATCGATGGCCATGGCGGAAGTGTCTGACCGACCGGCGGCGACAGCGGCAAAACCCGTTTCTC  
GGCAGTTATCCGGACCTGCCGCCGCGCGCCGCGCGCGTGCAGCGCGGGCGTCCGAGCCCCGTGCGCAGGGTG  
CATTGCTCGACGTGCTGATCGCCACCGCGCATGCTCGTCAACGTCGCGCGCGGCTTCGTCGGTTGCGCGAGC  
CGCGCGGAAGACGTCGTGCACGACGTGTTGTCGAAGCTCGTCAATTCGGAACAGGACGCGGTGCGCCAGC  
CGGTGCGGTACGTGACGCGGATGGTGCCTAATGCGTCGATCGATGCCGTGCCG

>indel-ecfI\_PHDC\_LMG16654

GTCATGAACCAGAGCGTATTGATGGACGCGGTGCTGAAGTCGTTCCGGGCCGCCGAAGATCGGCCGTTTCGACG  
CGACGGCGACGATCCGGTTGTTTCATCTATTGATATCGAGGGCGATGCCCGGCACATCCATCCGCGTGGCCCG  
CCATGCGGGGAACGGATGGCGCGGCGGTTTCGCGAAAGCAGGATCGGCTTCCACCGGCACCGGCCGTGACGCTC  
GCGACCGCGTGCACGTGAATACCAATACGGCGGCCCGCGCAATTCGAGCCGCCAAAACGAAAAAGCGCC  
GGCGTCCGGCGCTTTCCCGCTTCCGCGCGATGCTCAACGCCACCCAGTGGCACACCCCGGTGATGGTGATG  
CTCGAAGTGACACACGCGGTGCGGATGCGCTTCGGCGGACGCGGACGAGCACGGCACGGCCACGGCGCGG  
GGCGGTGAGGTTTTTCAGTCTTCATGTCGAGGTCCTGTCGTTCAATGATGGATCAAGGGTTCGCGAGCG  
GCATCACCACTGCGGCCGGGTTTCGTCGCGGCGGTCGTCACGCGGGCCGGCATGCCGGGCTTCCAGTCGCGG  
CGTCCAGTAGCGATTTCCGTTCCAGTAACGATGCCATGCCAGCCGATCGTGACTTCGGCCGGGCCAGGGT

AGGCGACGCAGCCGGTGAGCATCACGCTCGACAGCGCGCCACCATGACGGCGGATAACACGATGGATTTTCAT  
GTTGTTCTCCCTTGAAACGGCAAGTCGATCGTACGAGCCCCCAAATTTACGTTGGTAAGAAGTTGCTACCGC  
CGATGACACGTGCAAGCCCCTCGTCAGAAACGTGAAAGCCTGCGTCGGCGAAGTGAAGGCAGTCCGCGCGGC  
CCGGACGCGTAC

>indel-ecfI\_PHDC\_LMG18832

ACCTCGCGGANNTCCCGNCCAACGTCATGAACCAGAGNGTATTGATGGACGCGGAGCTGAAGTCGTTCCGGGCCG  
CCGAAGATCGGCCGGTTCGACGCGACGGCGACGATCCGGTTGTTTCATCTATTGATATCGAGGGCGATGCCCGG  
GCCACATCCATCCGCGTGGCCCGCCTGCGGGAAACGGATGGCGCGGCGCGTTCTCGAAAGCAGGATCGGCTTC  
CCCCGGCACCGGCTGACGCTCGCGACGCCGTGCGCGTGAATCCCGATACGGCGGCACCCGCGCAATTCGAGC  
CGCCAAAACGAAAAAAGCGCCGGCGTGCCGGCGCTTTCCCGCATCTGCCGCGCGATGCTCAACGCACCCAGTG  
GCACACCCGGTGATGGTGATGCTCGAAGTGACACACGCGGTGCGGATGCGCTTCGGCGATGGCCGGCACNNGC  
ACGCNCCGCGCCACGGCGGGCGGTGAGGGTTTTAGTAGCGTCTTCATGTCGAGGTCCTGTCGTTCAATGA  
TGGATCAAGGGTTCGCGACGGCATCACCACTGCGGCCGGGGTTCGTCGCGGGCGTTCGTCACGGTAGTCGTCC  
GATGATCCTCGCGGCGATCGTCACGCGGGCCGGCATGCCGGGCTTCCAGTCGCGGCGCTCCAGTAGCGATT  
TCCGTCCCAGTAACGATCGCCATGCCAGCCGATCGTGACTTCGGCCGGGCCAGGGTAGGCGACGCAGCCGGT  
AGCATCACGCTCGACAGCGCGCCCGCCATGACGGCGGATAACACGATGGATTTTCATGTTGTTCTCCCTTGAA  
ACGGCAAGTCGATCGTACGAGCCCCGAAATTTACGTTGGTAAGAAGTTGCTACCGCCGATGACACGTGCAAG  
CCCGTCGTAGAAACGTGAAAGCCTGCGTCGGCGAAGTGAAGGCAGTCCGCGCGGCCCGGACGCGTACGCGTC  
CACTCCGGCAACGATTCCGG

## 6. Séquences de la région *ecfK* déléetée chez les *B. vietnamiensis*

>indel-ecfK-viet\_LMG10929

GCGCTGAAGGCCACGCCGGCATTCCAGTACACGAAGGTGCCGTGCGCCCGCAAGCCCCGCAAGGTCGACGAG  
AAGCACTGACCCTCCGCGCCCGGTACGCAACGTGCCGGGCCGGCCGAGCGCGCGCGATGTCGCCGCGCGCGT  
CGGCCAGCGCCCATCCGCTGGCATCTACGTAATTCTACGTAGCGCTGCATACGTAGTTGTGCGCTTGTACGCC  
TCCCCCTTGACGGAATACTACCGACCACCGCCGACAAACAGTGCGCGGCGACGCATCGACGTCACGCGGGTT  
CGGTACACCTCTCGCCGAATCCACTTAAAACCACATAGGAGACGAGGAACCGCCATGAATCGGGCTTCCAGT  
AC

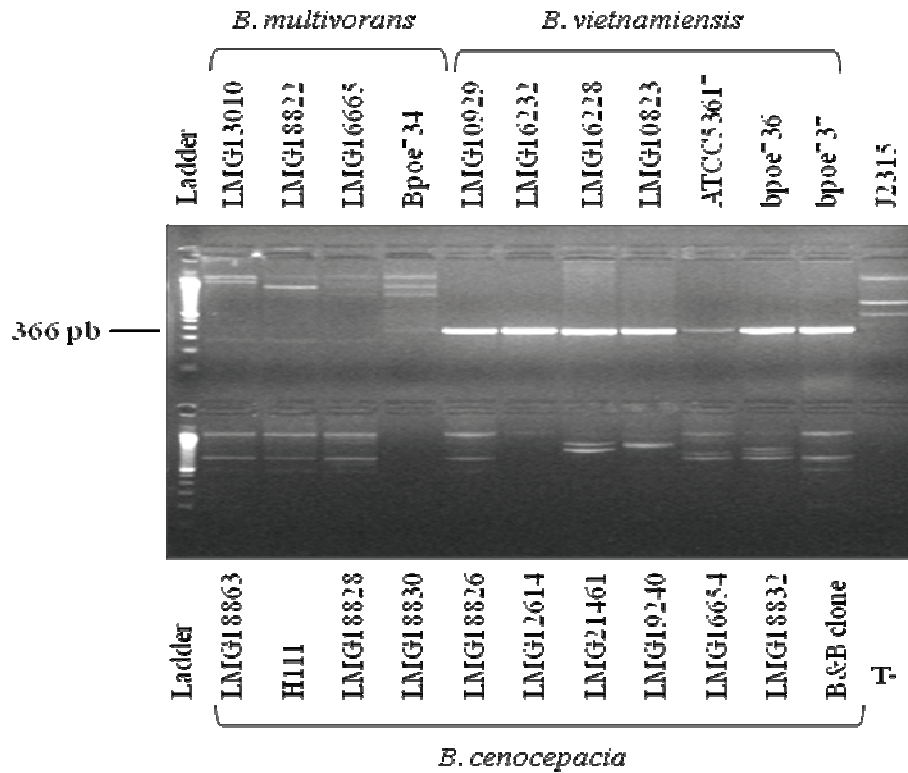
>indel-ecfK-viet\_bpoe737

GCGCTGAAGGCCACGCCGGCATTCCAGTACACGAAGGTGCCGTGCGAGCCGAAGCCGCGCAAGGTCGACGAG  
AAGCACTAACCCTCCGCGCCCGGTACGCAACGTGCCGGGCCGGCCGAGCGCGCGCGATGTCGCCGCGCGCGT  
CGGCCAGCGCCCATCCGCTGGCATCTACGTAATTCTACGTAGCGCTGCATACGTAGTTGTGCGCTTGTACGCC  
TCCCCCTTGACGGAATACTACCGACCACCGCCGACAAACAGTGCGCGGCGACGCATCGACGTCACGCGGGTT  
CGGTACACCTCTCGCCGAATCCACTTAAAACCACATAGGAGACGAGGAACCGCCATGAATCGGGCTTCCAGT  
AC

>indel-ecfK-viet\_bpoe736

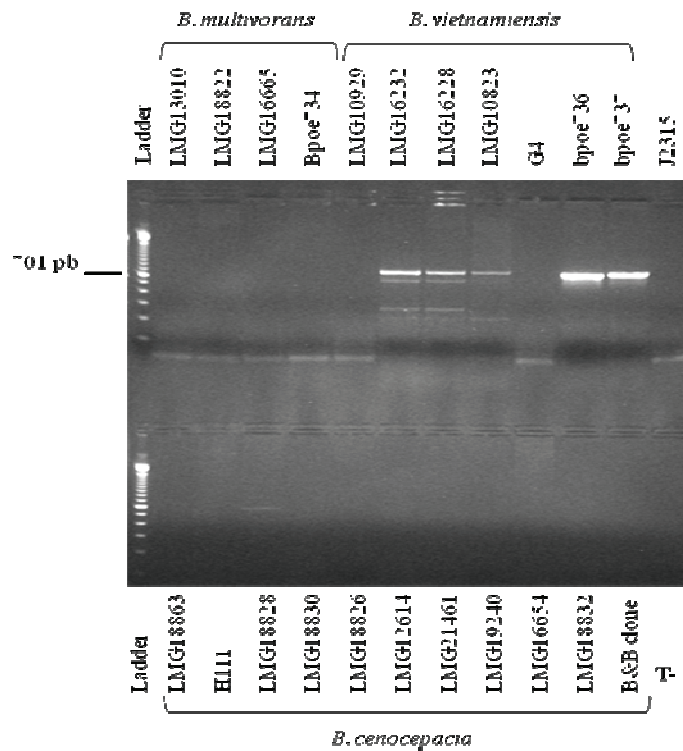
GCGCTGAAGGCCACGCCGGCATTCCAGTACACGAAGGTGCCGTGCGCCCGCAAGCCCCGCAAGGTCGACGAG  
AAGCACTGACCCTCCGCGCCCGGTACGCAACGTGCCGGGCCGGCCGAGCGCGCGCGATGTCGCCGCGCGCGT  
CGGCCAGCGCCCATCCGCTGGCATCTACGTAATTCTACGTAGCGCTGCATACGTAGTTGTGCGCTTGTACGCC  
TCCCCCTTGACGGAATACTACCGACCACCGCCGACAAACAGTGCGCGGCGACGCATCGACGTCACGCGGGTT  
CGGTACACCTCTCGCCGAATCCACTTAAAACCACATAGGAGACGAGGAACCGCCATGAATCGGGCTTCCAGT  
ACA

7. Crible PCR de la région *ecfK* délétée chez les *B. vietnamiensis*



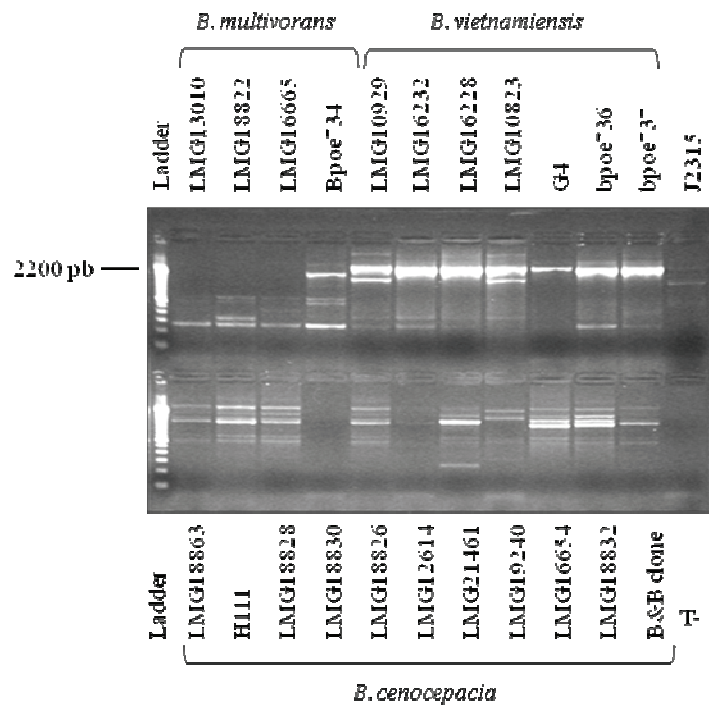
PCR indel-*ecfK*-viet réalisée sur différentes souches du Bcc. Cette PCR a été réalisée avec le couple d'amorces BCAM2746/BCAM2755 (e. g. Tableau 6). La taille attendue est de 366 pb.

8. Crible PCR de la région *ecfI* délétée chez les *B. vietnamiensis*



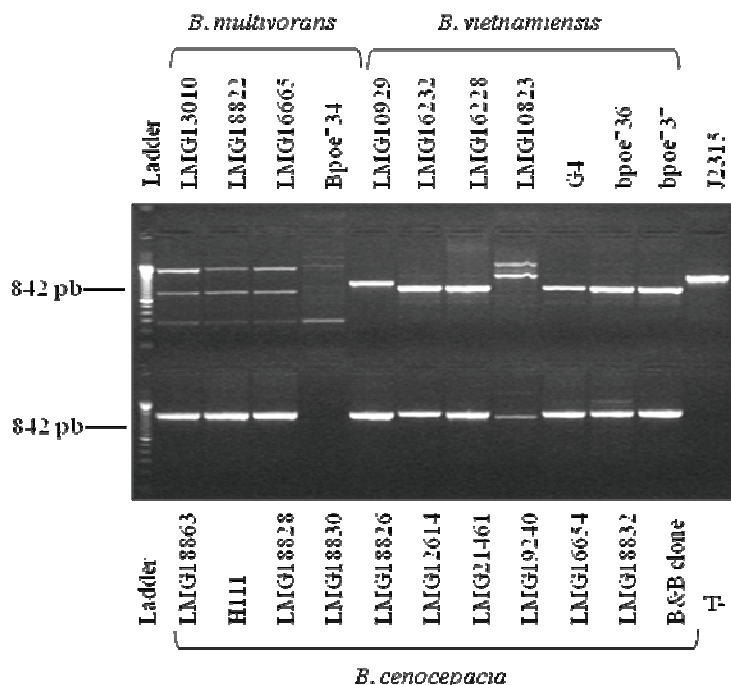
PCR indel-*ecfI*-viet réalisée sur différentes souches du Bcc. Cette PCR a été réalisée avec le couple d'amorces indel-*ecfI*-vietF/indel-*ecfI*-vietR (e. g. Tableau 6). Un manque d'ADN pour la souche G4 est responsable d'une absence d'amplification (e. g. génome de cette souche). La taille attendue est de 701 pb

9. Crible PCR de la région *ecfC* déléetée chez les *B. vietnamiensis*



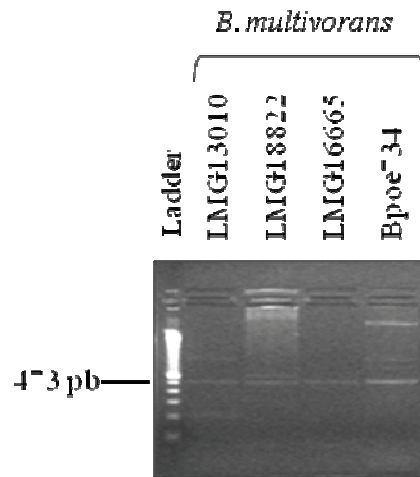
PCR indel-*ecfC*-viet réalisée sur différentes souches du Bcc. Cette PCR a été réalisée avec le couple d'amorces BCAL1382viet/BCAL1368viet (e. g. Tableau 6). La taille attendue est de 2200 pb

10. Crible PCR de la région *ecfE* déléetée chez les *B. vietnamiensis*



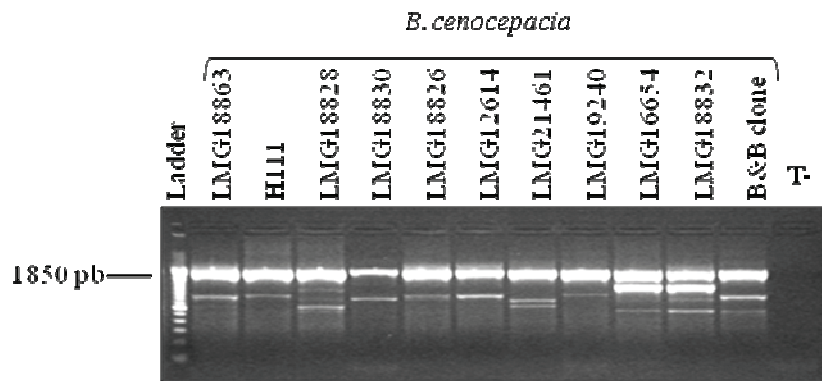
PCR indel-*ecfE*-viet réalisée sur différentes souches du Bcc. Cette PCR a été réalisée avec le couple d'amorces BCAL2466vietF/BCAL2466vietR (e. g. Tableau 6). La taille attendue est de 842 pb. Les séquençages des amplifias obtenus à partir des souches de *B. cenocepacia* et *B. multivorans* ne corresponde pas a la région désirée et sont donc le fruit d'une hybridation non spécifique du couple d'amorces.

### 11. Crible PCR du gène *ecfC* chez les *B. multivorans*



PCR *ecfC* réalisée sur différentes souches de *B. multivorans*. Cette PCR a été réalisée avec le couple d'amorces *ecfC-F/ecfC-R* (e. g. Tableau 6).

### 12. Crible PCR de la région *ecfE* chez les *B. cenocepacia*



PCR de la région *ecfE* réalisée sur différentes souches de *B. cenocepacia*. Cette PCR a été réalisée avec le couple d'amorces BCAL2459/*ecfE-R* (e. g. Tableau 6).

### 13. Séquences de la région *ecfE* amplifiée chez les *B. cenocepacia*

>région-*ecfE*\_LMG18830

```
TCATTCCGGCGCCGGATCCGCGACCCGGACGATGCCGAGGACGTTCTGCAGGATGTGCTTCACGAATTCGTACA
GGCTTATCGACTGCCTGCGCCGATCGAACAGGCGAGCGCGTGGCTGTCCGTACCGCGCGCAATCGCATCATC
GACCGTTTTCGCAAGAAGAAGGAGCAGCCGCTGACCGATCTGCTCGATGCCGACGACGAAGCGGACAGTGAG
TATCGCCTGGATCTTGCACTACCGTCGCACGATGCCGGCCCCGAGGCGCTCTACGCTCGCACGCTGTTGCTCAA
GGCGTTGCAGGATGCGCTCGACGAATTGCCGGCGAATCAGCGCGAGGTTTTTCATCGCGCACGAGCTGGAGGGG
CAGTCGTTCAAGGAGATGGCCGCGCAGAGCGGCGTCACGCTCAATACGCTGCTCGCGCGCAAACGCTACGCGG
TCCTGCATCTGCGTGCTCGACTGCAACCCATTTATGACGAACTGGATAACTAGAGGAGTCGACGGATGAATTTT
CGAATCAGATGTGCAGGCAAAGCGCTGCTCGTGGTGATCGCCATCGGTGTGCTCGGATGGGTGCTCATGATGC
TGTGGAAGTGGGTGATTCCGGGGTTATTCGTCGGTGCCCGGACGATCGACTTCGCCCATGCGCTGGGCTGCTC
GTGTTGAGCCGCATCCTGTTCCGGCGATTCCGCGGACATCGCGGCTGGCACGGCCGACGTCACCTGGCGCAAGT
GGGAAGCCATGACGCTGAAGAACGCGAGCAGTTTCGGGCCGATGGCGATCGCGCGGCAAGCGGGCCACGG
AGGAGTGAGCATGCGGATAAATCAACGGCCGACT
```

>région-*ecfE*\_LMG18863

```
CATTCGGCGCCGGATCCGCGACCCGGACGATGCCGAGGACGTTCTGCAGGATGTGCTTCACGAATTCGTACAG
GCCTATCGACTTCCCGCGCAATCGAACAGGCGAGCGCGTGGCTGTCCGTACCGCGCGCAACCGCATCATCG
ATCGTTTTCGCAAGAAGAAGGAGCAACCGCTGACCGATCTGCTCGACGCCGACGACGAAGCGGACAGCGAGT
ATCGCCTGGATCTGGCGCTGCCGGCGCACGATGCCGGCCCCGAGGCGCTCTACGCTCGCACGCTGTTGCTCAA
GGCGTTGCAGGATGCGCTCGACGAGTTGCCGGCGAATCAGCGCGAGGTTTTTCATCGCGCACGAGCTGGAAGGA
```

CAGTCCTTCAAGGAGATGGCCGCGCAAAGCGGTGTCACGCTCAATACGCTGCTCGCGCGCAAACGCTACGCGG  
TCCTGCATCTGCGTGCTCGACTGCAACCCATTTACGACGAACTGGATATCTAGAGGAGTCGACGGATGAATTTT  
CGAATCAGATGGTGCAGGCAAGGCGCTGCTCGTGGTGGTGGCCATCGGCGTGCTCGGTGGGGTCGTGATGATG  
TGTGAAGATGGTGCAGGCTATCCGGCGCTATTCGTTGGCGCCCGACGATCGACTTCGCTCATGCGTGGGCCTGCTC  
GTGTTGAGCCGCATCTGTTCCGGCGATTCCGCGGACACCGCGCTGGCACGGCCGGCCACTGGCGCAAGT  
GGGAAGCCATGACGCCGAAGAGCGCGAGCAGTTCCGGGCCGATGGCGTCCGCGCGCAAAGCGGCCACGG  
AGGAGTGACCATGCGCGAGAAATCAGCGGCCGACTGCAAGCAGAACGCGGGCGTGATTGCGCCGGTCTCGA  
TCTGAAGCGTTGCGAAGGCAAGGGCGATTGCGTCAACGTTTGCCC

>région-ecfE\_LMG12614

GCTAGGGAATTTCAATTCGGCGCCGATCCGCGACCCGGACGATGCCGAGGACGTTCTGCAGGATGTGCTTAC  
GAATTCGTACAGGCCTATCGACTTCCCAGCCAATCGAACAGGCGAGCGGTGGCTGTTCCGTACCGCGCGCA  
ACCGCATCATCGATCGTTTTTCGCAAGAAGAAGGAGCAACCGCTGACCGATCTGCTCGACGCCGACGACGAAGC  
GGACAGCGAGTATCGCCTGGATCTGGCGCTGCCGGCGACGATGCCGGCCCCGAGGCGCTCTACGCTCGCACG  
CTGTTGCTCAAGGCGTTGCAGGATGCGCTCGACGAGTTGCCGGCGAATCAGCGCGAGGTTTTTCATCGCGCACG  
AGCTGGAAGGACAGTCCCTCAAGGAGATGGCCGCGCAAAGCGGTGTCACGCTCAATACGCTGCTCGCGCGCAA  
ACGCTACGCGGTCTGCATCTGCGTGCTCGACTGCAACCCATTTACGACGAACTGGATATCTAGAGGAGTCGA  
CGGATGAATTTTCAATCAGATGTGCGGGCAAGGCGCTGCTCGTGGTGGTGGCCATCGGCGTGCTCGGTGGGG  
TCGTGATGATGTTGTGGAAGTGGGTGATGCCGGCGCTATTCGTTGGCGCCCGACGATCGACTTCGCTCATGCG  
CTGAGCCTGCTCGTGTGAGCCGCATCCTGTTCCGGCGGATTCGCGGACACCGCGCTGGCCGCGGCC  
ACTGGCGCAAGTGGGAAGCCATGACGCCGAAGAGCGCGAGCAGTTCCGGGCCGATGGCGGTCCGCGCGCA  
AGCGGCCACGGAGGAGTGACCATGCGCGAGAAATCAGCGGCCGACTGCAAGCAGAACGCGGGCGTGATT

>région-ecfE\_LMG21461

TTCATTCGGCGCCGGATCCGTGACCCGGACGACGCCGAGGACGTTCTGCAGGATGTCCTTCACGAATTCGTACA  
GGCTTATCGACTTCTGCGCCGATCGAGCAGGCGAGCGGTGGCTGTTCCGTGCCGCGCGTAACCGCATCGTCCG  
ATCGTTTTTCGAAGAAGAAGGAGCAGCCGCTGACCGACCTGCTCGATGCCGACGACGAAGCCGACAGCGAGT  
ATCGCCTGGATCTCGCCTACCCGGCGCACGATGCGGGGCTGAAGCGCTTACGCTCGCACGCTGTTGCTCAA  
GGCGTTGCAGGACGCGCTCGACGAATTGCCGGCGAATCAGCGCGAGGTTTTTCATCGCGCACGAACTGGACGGG  
CTGTCCTTCAGGGAGATGGCCGCGCAAAGCGGTGTCACGCTCAATACGCTGCTCGCGCGCAAACGCTACGCGG  
TTTTGCATCTGCGTGCTCGACTGCAACCCATTTATGACGAACTGGATATCTAGAGGAGTCGACGGATGAATTTT  
CGAATCAGATGTGCGGGCAAGGCGCTGCTCGTGGTGGTGGCGATCGGCGTGCTCGGCTGGGTCTCATGATGC  
TGTGGAAGTGGGTGATGCCGGCGCTATTCGTCGGCGCGCGGACGATCGACTTCGCTCATGCGCTGGGCCTGCTC  
GTGTTGAGCCGCATCCTGTTCCGGCGATTTCGCGGACACCGCGGTGGCACGGCCGGCCACTGGCGCAAGT  
GGGAAGCCATGACGCCGAGGAGCGCGAGTTCCGCGCCGATGGCGATCGCGCGGCAAGCGGCCCGCGC  
AGGAGTGAACATGCGCGAGAAAGCAACGGCCGACT

>région-ecfE\_LMG19240

GAGTGAACATGCGCGAAAAATCATCGGACGACTGCAAACAGAAGCGGGTGTGATTGCGCCGGTCTCGATCT  
GAAACGTTGCGAAGGCAAGGGGATGCGTCAACGTTTGCCCCGAAAACGTTTCGAGGTGCGGCGCATCGAT  
CAGGCCGATTACGTCGACCTCGGTCTGATGCATCGCTTAAACTGCGCGTGCATGGAATGAAAGTCGCCTACA  
CGCGAATGCGCAAGCCATGCCGATCATGCGGGTTCGCGTGCAGGCGGTGCCGAACCGCGGATCAGCTCGC  
CAGAAGCCATGACGTCGGCGTCTCGCAGATGCGGAGCGGTATCGCGGAGTCCGCGGATGCGGAGCCG  
CTGGATACGCCGTTTCATGTGATCAATCTCGTCAAGGAATGCACTTGGCTTCAACCATCGACGTTCTTCTGCAAC  
GGCTTTCATGAAGAAAGCCGGAGCGAAATCATCGCCCCGGCCCGCGCGCAGTTCAGGGGCCGAATGCGGCGTT  
TCCGACGCTCGTCGACGAATACGAAGCGCTCGCACCCCGGCTGCGAGGGCGCTTTCGAAACGTTCCGGCCTCT  
TCAGCGACGATTTCAATCGTCCGAAAGCGGTGCGCTATCGTTGGCCGTGTCGGCGGAAACGGGACGATTCCT  
GCGCAACACCGTGTGTCGCAACGGTCCGGCAGAAATCTTGAAGTTGGCAGTTCCTGTGGCGTGTGACTCTGT  
ATTTCCCGGACGCGCTGCGTATGTTGGGCCGTGGCGTGGTTCATCGCGACCGAACGTTGATGCGCTCAAATGCGC  
CGTTTTGCGGGCTCAGCTCGAAACAGCAGGGCTGACGCCTATGTCGACTTCCGGGAGGGCGATGTGTTCCGAG  
ACGGTTGCCGAACTGGACGGAC

>région-ecfE\_B&Bclone

CGGCGCCGGATCCGCGACCCGGACGATGCCGAGGACGTTCTGCAGGATGTGCTTACGAATTCGTACAGGCCT  
ATCGACTTCCCAGCGCAATCGAACAGGCGAGCGGTGGCTGTTCCGTACCGCGCGCAACCGCATCATCGATCG  
TTTTCGCAAGAAGAAGGAGCAACCGCTGACCGATCTGCTCGACGCCGACGACGAAGCGGACAGCGAGTATCG  
CCTGGATCTGGCGCTGCCGGCGCAGATGCCGGCCCCGAGGCGCTTACGCTCGCACGCTGTTGCTCAAGGCG  
TTGAGGATGCGCTCGACGAGTTGCCGGCGAATCAGCGCAGGTTTTTCATCGCGCACGAGCTGGAAGGACAGT  
CCTTCAAGGAGATGCCCGCGCAAAGCGGTGTCACGCTCAATACGCTGCTCGCGCGCAAACGCTACCGGTCCT  
GCATCTGCGTGCTCGACTGCAACCCATTTACGACGAACTGGATATCTAGAGGAGTCGACGGATGAATTTTCGA  
ATCAGATGTGCGGGCAAGGCGCTGCTCATGTTGGTGGCCATCGGCGTGCTCGGTGGGGTCTGTGATGATGTTGT  
GGAAGTGGGTGATGCCGGCGCTATTCGTTGGCGCCCGACGATCGACTTCGCTCATGCGTGGGCCTGCTCGTG  
TTGAGCCGCATCCTGTTCCGGCGATTCCGCGGACACCGCGGCTGGCACGGCCGGCGCCACTGGCGCAAGTGGG  
AAGCCATGACGCCGGAAGAGCGCGAGCAGTTCCGGGCGCATGGCGGTGCGCGGCAAGCG



## C. Chapitre 3

### 1. Séquences ADN des IS amplifiées chez le clone B&B clone

>IS407-h\_B&Bclone

GTATTCCGCATCGCCGGCGACATTGACGGCAGTATTGGAAGAAAGTGACGCGTCTGACCTGCCCCCTGCAAAC  
AGGGCCAGCTGGAGTCTAGTAAAGTTTCGTTTTTCGGAGAGAAGACGAACATGAAGAAGCGCTTTACGGAACAG  
CAAATCATCGGGTTTCTGAAGGAAGCCGAGGCCGGTATGCCGGTCAAGGAAGTGTGCAGGAAGCACGGGTTCA  
GTGACGCGTCTGTTACACCTGGCGCGCAAGTTTGGCGGCATGGAGGTCTCGGAAGCCCGCCGGCTCAAGGA  
CCTCGAGGTGGAGAATGCCCGGCTGAAGAACTGCTGGCCGAAGCAATGCTCGATATGGAAGCGTTGAAGGTC  
GTCGTCAAGGGAAAGCCCTGAGCCCGCAAGCCAAACGCGAGGCAGTGTGCGGATTTCGGGAGAAGGTCAACC  
TCTCCGAGCGCCGCGCTGCCGGCTTGTGCGGCTTCTCGCAGCGTGTGCATTACGACGCGAAGCCGGACCAT  
GAGAATGATGTGCTAGCGGCGCTGTTGGAGTTGGCGCACGAACGTCGTCGATTTCGGTTACCGCCGACTGC  
ATGCTCTGGTGGAGCGGAAGGCACGCACGCCAATCAACAAGCGCATCTATCGCCTGTACCGTGACGCAGGGCT  
GGCTGTGCGGCGCCGTCGCAAGCGCCACGGCGTAATGATTGAGCGGAGCAACTGGCATTGCCAGGGCACCC  
AACGAGGTGTGGTCAATCGATTTCGTGATGGATGCGCTTTCCAACGGCCGCGTGAAGTGCCTGACCGCTC  
TCGACGATTTACGAAAGAGGCTGTGACATCGTCTGACCATGGTATCTCAGGTTTATATGTCGCCCCGGGAG  
CATTGGACCGTGCAGCTCGCTTCCGTGGCTATCCCAAGGCGGTGCGAACAGACCAGGGACCCGAATTTACGAG  
CCGCGCGCTTGACCAGTGGGCATATGCCAACGGCGTCACTCTGAAGTTGATTTCAGGCGGGCAAGCCAACGCAG  
AATGCGTACATCGAATCGTTCAACGGCAAATTCGCGACGAATGCCCTAACGAGCACTGGTTCACGACGCTCG  
CGCACGCTCGGGCTGTCATCGCGCATGGCGTCAAGGACTACAACGAGCAACGGCCGCACAGTGCCTGAAC  
CCTTGCCTGTCAGAGTTTGCGGCGAAACATCGGGCAACCGCGGATGCTCCTGCCGCTTCCAGGAGCTGGTTT  
AAAGGACTTTGCTAGAAGCCATTGGCCCTATTGAAGGGGGCAGGTCAACTATGATGGGTAGCATAGTGCAT  
GCTACTATGACGACCATCA

>IS407i\_B&Bclone

CAATGGGCTTCTAGCAAAGTCCCTTTAAACCAGCTCCTGGAAAGCGGCAGGAGCATCCGCGGTTGCCCGATGT  
TTCGCCGCAAACCTCTGACGGCGCAAGGTAGTTTCAGTGCAGTGTGCGGCCGTTGCTCGTTGTAGTCCGACGCCA  
TGCCGCGATGACAGCCCGAGCGTGCAGGAGCGTCTGTAACAGTGTCTGTTAAGGCATTTCGTCGCGGAATTTG  
CCGTTGAACGATTTCGATGACGCATTCTGCGTTGGCTTGCCTGCCCCGCTGAATCAACTTCAGAGTGACCGGTTCCG  
ATATGCCACTGGTCAAGCGCGGCTCGTAAATTCGGTCCCTGGTCTGTTCCGACCGCTTGGGATAGCCAC  
GGAAGCGAGCTGCACGGTCCAATGCTCCCGAGCGACATATAAACCTGAGATAACCATGGTTCGACGACGATGTCG  
ACAGCCTCTTTCGTGAAATCGTCCACGACGGTCAAGCACTTCACGCGCCGCGCTTGAAAGCGCATCCATCA  
CGAAATCGATTGACCACACCTCGTTGGGTGCGCCTGGCAATGCCAGTTGCTCGCGCTCAATCATTACGCCGTGC  
CGACGTGCAACGGCCGACACCGATGCTCCCGCTCCAATGTTTCTGACGATGGCGATTTTTTCTTGCGCCGT  
GCGCCGCGGACGGCGCTCCGGCTCGGTCACAATTTGATGCTTTCCACGTAATGACCAGGCTTACTGATAGGCA  
CAAGACTATCCCTATTTTAAAGAGAGTCCCTCGTGTCTCAGATAACGTGGGGCCGCTCCAATAGCAGATTA  
CCACGGCGGAAATAATCTATATTTCCCGTCTGCGCCGATGCTTCATGTC

>IS407-k\_B&Bclone

CGTCACATCGTCGAAGCCAACGACTTTACCGATGTCTTCTGTTTGGCGATTGCCGCGAAATCCATCGACCGAT  
ACATGCAGTCTGTTACGATCTCGGCTGCGCGTACACGTTTTTCGAGGAGGGCTATGTTCTGTCCTCCACTGGATCA  
CGATGGAGCATCACGGCGTCAATGGTCTGCTGTTGCAATCATCGGGTTTCTGAAGGAAGCCGAGGCCGGTATG  
CCGGTCAAGGAAGTGTGCAGGAAGCACGGTTCAGTGTGCGTCTTACACTTGGCGTGCAGGTTTCGGCG  
GCATGGAGGTCTCGGAAGCCCGCCGGCTCAAGAACCTCGAGGTGGAGAATGCCCGGCTGAAGAAGCTGTGG  
CCGAAGCAATGCTCGATATGGAAGCGTTGAAGTCTGCTCAAGGAAAGCCCTGAGCCCGCAAGCCAAACG  
CGAAGCAATGCTCGGATTCGGGGAAAGTCAACATCTCCGAGCGCCGCGCTGCGGCTTGTGCGGGCTTCT  
CGCAGCGTGTGATTACGACGCGAAGCCAGACCAGAGAATGAGGTGCTAGCGGCGGCTGTTGGAGTTGG  
CACACGAACGTCGTCGATTTCGGCTACCGCCGAATGCACGCCCTAGTGAACCGGAAGGCACGCACGCCAATCA  
CAAGCGCATCTATCGCCTGTACCGTGAAGCAGGGCTGGCTGTGCGGCGCCGTCGCAAGCGCCACGGCGTAATG  
ATTGAGCGGAGCAACTGGCATTGCCGGGCGCCCCAACGAGGTGTGGTCAATCGATTTCTGATGGATGCGC  
TTTCCAACGGCCGGCGGTGAAGTGCCTGACCGTCTGACGACTTCACGAAAGAGGCTGTGACATCGTTGT  
CGACCATGGCATCTCAGGTTTGTATGCTGCTCGGCAATTGACCGCGAGCTCGCTTCCGTGGCTATCCCAAGG  
CGGTGCGAACAGACCAGGGCCCCGAATTTACGAGCCGCGCTTGACCAGTGGGCGTATGCCAAGCGCGTCTG  
TCTGAAGTTGATTACGCGGGCAAGCCAACGCAGAATGCGTACATCGAATCGTTCAACGGCAAGTTCCGCGAC  
GAATGCCTAACGAGCACTGGTTCCCGCCGCTCGCGCATGCTCGGGCAGTCATCGCGCATGGCGTCAAGGACT  
ACAACGAGCAAAGGCCGACAGTCCGCTTGAATATTGTGACAAGGCGGAAGGCTTGTGCGCCTGGTACCGT  
CACATCGTCGAAGCCAACGACTTTACCGATGTCTTCTGTTTGGCGATTGCCGCGAAATCCATCGACCGATA  
TGCAGTCTGTTACGATCTCGGCTGCGCGTACACGTTTTTCGAGGAGGGCTATGTTCTGTCCTCC

>IS407-l\_B&Bclone

TTTGGCATCAACAGGCGCGCTGACCTGCCCCATCAATAGGGCCAATGGGCTTCTAGCAAAGTCCCTTT  
AAACCAACTCCTGGAAAGCGGCAGGAGTATCCGCGGTTGCCGATGTTTCGCCGCAAACCTCTGACGGCGCAAG  
GTAGTTCAGTGCCTGTGCGGCTTTGCTCGTTGATGCTGACGCCATGCCGCGATGACTGCCCGAGCATGCG  
CGAGCGTGTGAACAGTGTCTGTTAAGGCATTCGTCGCGAACTTGCCGTTGAACGATTTCGATGTACGCATTC  
TGCGTTGGCTTCCCGCTGAATCAACTTCAGAGCGACCGGTTGTCATACGCCCACTGGTCAAGCGCGCGGCT  
CGTAAATTCGGGGCCCTGGTCTGTTGACCGCCTGGGATAGCCACGGAAGCGAGCTGCGCGGTTCAATGCC

CGAGCGACATACAAACCTGAGATGCCATGGTTCGACAACGATGTCGACAGCCTCTTTTCGTGAAGTCGTGCGACGA  
CGGTCAGGCACCTTCACGCGCCGGCCGTTGGAAAGCGCATCCATCACGAAATCGATTGACCACACCTCGTTGGG  
GCGCGCCGGCAATGCCAGTTGCTCGCGCTCAATCATTACGCCGTGGCGCTTTCGACGCGCCGACGAGCCAGC  
GCTCCCTACGGTACAGCGATAGATGCGCTTGTGATTGGCGTGGCTGCTTCGCTTCCACTGACCGCGCAA  
TCGGCGGTAGCCGAATCGACGACGTTTCGTGTGCCAACTCCACCAGACGCGCCGCTAGCACCTCATTCTCGTGGT  
CTGGCTTCGCGTCGTAATGCAGCACGCTGCGAGAAAGCCCGACAAGCCGGCAGGGCGGGCGCTCGGAGATGTT  
GACCTTCCCCGAATCGCCGATACTGCTTCGCGTTTGGCTTTCGCGGCTCAGGGCTTTCCTTGACGACGACCTT  
CAACGCTTCCATATCGAGCATTGCTTCGGCCAGCAGCTTCTCAGCCGGCATTCTCCACCTCGAGGTTCTTGA  
GCCGGCGGGCTTCCGAGACCTCCATGCCGCGCAACTTCGCACGCCAAGTGTAGAACGACGCATCACTGAACCC  
GTGCTTCTGCACAGTTTCTTGACCGGCATACCGGCTTCGGCTTCTTCAGAAACCCGATGATTTGCTGTTCCGT  
AAAGCGCTTCTTCAAGTTTCGTTCTTCTCCGAAAACGAACCTTACTAGACTCCGGCTGGCCCTGTTTGCAGGG  
GGCAGGTCAGGGAGTACTGATTGGAGAACGGCACAC

>IS407-m\_B&Bclone

CCGGCTGAAGAAGCTGCTGGCCGAAGCAATGCTCGATATGGAAGCGTTGAAGGTCGTGCTCAAGGGAAAGCC  
CTGAGCCCCGAAGCAAACGCGAAGCAGTATCGGCGATTCCGGGGGAAGGTCAACATCTCCGAGCGCCGCGCCT  
GCCGGCTTGTTCGGGCTTTCTCGCAGCGTCTGCATTACGACGCGAAGCCAGACCAGAGAATGAGGTGCTAGC  
GGCGCGTCTGGTGGAGTTGGCACACGAACGTCGTCGATTTCGGCTACCGCCGATTGCACGCCCTAGTGAACCG  
GAAGGCACGCACGCCAATCACAAGCGCATCTATCGCTGTACCCTGAGGCAGGGCTGGCTGTGCGGCGCCGTC  
GCAAGCGCCACCGGTAATGATTGAGCGCAGGCAACTGGCATTGCGGGGCGCCCCAACGAGGTGTGGTCAAT  
CGATTTTCGTGATGGATGCGCTTTCCAACGGCCGGCGCGTGAAGTGCTGACCGTCGTCGACGACTTCACGAAA  
GAGGCTGTCGACATCGTTGTCGACCATGGCATCTCAGGTTTGTATGTCGCTCGGGCATTGGACCGNGTGTCTGT  
NTGGANTGCATCGTTCGCTCAAGAATCGATCAAGAAATCAGGGGATTACCGCGGAAGTTCGAGGTCTGCTACGA  
GAGTAGACGACATTATTTTGGTGAGATTTCAATAAAGCCGTCAACTCGTGTCCCAGTGCCTTGGTCTACGGG  
GATTCTGGAGTGACGGAATTTTATGATGTCTACCGCGGACGTGGAGTACGGCGTCAACCCGATATCCCACCC  
GGAGATTGCAATAGCCCCGCCATCCGCCCCCGCGTTGCGGAGTTCGGACATCTGTTCCGGCCGTGGGCAGAT  
AGACCGCAATCTCTCTGCCCCAGCGCCCCAGCCCTGAATCTCAGTTTCGAGCCGAATCGGCAGCCGCTGATTC  
GTCGGCGCTTCCGCGAGCAATTCAAGCTGACGTACCCGCGTACCGACGCGCAGCGGGCTAGTTTC

>IS407-n\_B&Bclone

TGAATCGGACCCGTAGCGCCGGCCACAGTAGATGTCGCGCAGACCGCCTTCCGGAGATACGTCGAGCTTTGAC  
GAAAGTGCCACGCGGCCCCAGGCTGTTGCTTTCGTCGCCACACGATGGAATGACGTAGAAGTCTGCGATTG  
ACGGATGTTGGCTCAGGTGTTGATTTAGCTTGACGTTTCATGAAAGCACCTCCCGCCACCGCAACCGAGTCAGTT  
CCGAGTGACCTGCCCCATCAATAGGGCCAAATGGGCTTCTAGCAAAGTCCCTTTAAACCACTCTGGAAAGC  
GGCAGGATACCGCGTTGCCGATGTTTCGCGCAAACCTGACGCGCGAAGGTACTGAGTTCAGTGCCTGTGCG  
GGCCTTTGCTCGTTGTAGTCTGACGCCATGCCGCGATGACTGCCCGAGCATGCGCGAGCGTCTGTAACCAAGTG  
CTCGTTAAGGCATTCGTCGCGAACTTGCCGTTGAACGATTTCGATGTACGCATTCTGCGTTGGCTTGCCCGCT  
GAATCAACTTCAGAGCGACGCCGTTTCGCATACGCCACTGGTCAAGCGCGCGGCTCGTAAATTCGGGGCCCTG  
GTCTGTTTCGACCGCCTTGGGATAGCCACGGAAGCGAGCTGCGCGTCCAATGCCCGAGCGACATACAAACCT  
GAGATGCCATGGTCGACAACGATGTCGACAGCCTCTTTCGTGAAGTTCGTCGACGACGGTCAGGCACTTCACGC  
GCCGGCCGTTGGAAGCGCATCCATCACGAAATCGATTGACCACACCTCGTTGGGGGCGCCCGCAATGCCAG  
TTGCTCGCGCTCAATCATTACGCGTGGCGTTGCGACGGCCGACAGCCAGCCCTGCCTACGCTACAGGC  
GATAGATGCGCTTGTGATTGGCGTGCCTTCGCGTTCCACTAGGGCGTGAATCGGCGGTAGCCGAATCG  
ACGACGTTTCGTGTGCCAACTCCACCAGACGCGCCGCTAGCACCTCATTCTCGTGGTCTGGCTTTCGCGTCTGTAAT  
GCAGCAGCTGCGAGAAAGCCCGACAAGCCGGCAGGCGCGCGCTCGGAGATGTTGACCTTCCCCGAATCGC  
CGATACTGCTTCGCGTTTGGCTTTCGCGGCTCAGGGCTTCCCTTGACGACGACCTTCAACGCTTCCATATCGAG  
CATTGCTTCGGCCAGCAGCTTCTTCAGCCGGGCACTTCCACCTCGAGGTTCTTGAGCCGGCGGGCTTCCGAGA  
CCTCCATCCGCGCAACTTCGCACGCCAAGTGTAGAACGACGCATCACTGAACCCGTGCTTCTGCACAGTTCC  
TTGACCGGCATCCCGAATCGGCGAATCGGCGAATCGCGTTCCTTCAGAAACCCGATGATTTGCTGTTCCGTAAGCGCTTCTTCATGTT  
GTCTTCTTCCGAAAACGAACCTTACTAGACTCCGCGTGGCCCTGTTTGCAGGGGGCAGGTCAGTACAACCG  
TGCTCGGCCGACGCTCCGAGCGGATTGGGCATC

>IS402-a\_B&Bclone

TCGTAACGGCCAGTTCGACACGATGCCAGACTTCATGGGAAGCTGATCTTATGTCCGTTCCGTGGCACCTGGA  
CTTGAGCGTCGATTTTTGTGGATGTTGTCAGGGGCGAGCCTATGAAACGCCAGCACGCGGCTTTTACGTCTGCT  
TTGCTGCTTTGCTCACCATGTTCTTTCGCGTATCCCTGATTCTGTGGATACGTATAACGCTTTGAGGGAGCTGAT  
ACCTCGCGCAGCCGAACGACCGAGCGCAGGAGTCAGTGAGCGAGGAAGCGGAAGAGCCCAATACGAAA  
CCGCTTCCCCGCGCTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCAGCTGGAAAGCGGGCA  
GTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTTACACTTTATGCTTCCGGCT  
CGTATGTTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAACAGCTATGACCATGATTACGCCAAGC  
TATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGGGAGCTTCCCATATGGTCGACCTGC  
AGGCGGCCCGCAATTCAGTATGATTCCGAGACGACGGAAGTGTGCAACCAGGCATGCGTGCCTTCAACGAC  
CCAGCGATGTTTCCAAGGCCGCTGCCGTGTTCCGTACGGCGCCTGGCAATCACTGGCTCGATACCGCGATTGC  
GCAATGCTCTTCGATGTCGCTCGGAGTCGTAACCCGATCGGCGTAGACAACGCGGCTCTCTGCAGTGGGTG  
GCCACGCAATCCGCAATCGGCGAATCGCGTTCGATCGAGCGGACGAAATTCGCTGACGCTCGTGGACGTTGGC  
CCAGTCAGGATCGCGCGAGCGGCGTGGCGTTGGCGTGGTGACGATATGGTGTGTTGAAACCGGGTCTGTGCGC  
GATCAGTGGGGTTTGGCCAGTTTTTGGCCGCCCCAACAGCACGAATCGATGAAGAGTCGACAGCGGCTCGC  
GAGAAATCGATCTGGTTCGCGCACGAGCTTCGCGAGCAGCAACACGTGCAACCGGTCACACGCGCAGCAG  
CTTGCCAGTCGGAAGCCGTCGCCAACATGTCACGCCGAGCCGAACCCATCTCGCGGGCAAGTCGCGCCA  
CGGTAGTCCGGTCTTGAATCGAATTTCCGCGGCCGCCATGGCGCCGGGAGCATGCGACGTCGGGCCAAT

TCGCCCTATAGTGAGTCGTATTACAATTCACCTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCTGGCGT  
TACCCAACCTAATCGCCTTGACGACATCCCCCTTTCCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCCGATC  
GCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAAAGCGCCGGGT  
GTGGTGGTTACCGGACCGTACCCTACACTTGGCAGCGCCCTAGCGCCCTAGCGCCCGCTCTTCGCTTTCCGCTTCC  
TTTCTCGCACGTTCCGCCGCTTTCCCGTCAAGCTCTAAATCGGGGCTCCCTTAGGGTTCCGATTTAGTGCTTT  
AGGCACTCGACCCCAAACTGATTAGGGTGATGTTACGTATGGCATCGCCTGATAGACGTTTTCCGCTTGACG  
TGGAGTCCACGTTCTTTAATAGGTGACCTCT

>IS402-b\_B&Bclone

TAACCGGTTCTGGCCTTTGGCTGGCCTTTTGCTCAACATGTTCTTTCTGCGTATCCCCTGATCTGTGGATAACC  
GTATTACCGCCTTTGAGTGAGCTGAATCCGCTCGCCGACGCCAAGACCGAGCGCAGCGAGTCAGTGAGCGA  
GGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCTGGCCGATTCAATTAATGCAGCTGGCAGC  
ACAGGTTTCCCGACTGGAAAGCGGGCAGTGACGCGAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACC  
CCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAA  
ACAGCTATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTT  
GGGAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGGAATTCAGTAGTGATTGACGGAAGTGATGCAACCAG  
GCATGCGTGGTTCAACGACCCAGCGATGTTTGCCAAGGCCGCTGCCGTGTTCCGTACGGCGCCTGGCAATCA  
CTGGCTCGATAACCGGATTGCGCAATGCTCTTCGATGTCGCTCGGAGTCGTAACCGCGATCGGCGTAGACAAC  
GCGCGGTCTCTGCAGTGGGTGGCCACGCAATCCGCGAATCGGCGGAATCGCGTCGATCAGCGGCAGCAATTGC  
GTGACGTCGTGGACGTTGGCGCCAGTCAGGATCGCGGCGAGCGGCGTCCGTTGGCGTCCGCTGGGATAGGT  
GCTTGAACCGGGTCTGTGCGGATCAGTGGGGTTTTGGCCAGTTTTTGGTCCGCCCAACAGCACGAATCGAT  
GAAGAGTCGACAGCGGCTCGCGAGAAATCGATCTGGTCTGCCGCACGCAGCTTCGCGAGCAGCAACACGTGCA  
ACCGGTCCACACGCCAGCAGCTTGCCAGTCGCGAAGCCGTCGCCAACATGTCACGCCCGAGCCGCAACCCAT  
CTCGGCGGGCAAGTCGCGCCAGCGTAGTCCGGTCTTGAGACGTAGTCCGGTCTTGGGCGTAGTCCGGTCTTGA  
GGAATCGAATTTCCCGCGCCGCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGT  
CGTATTACAATTCACCTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCTAATCGC  
CTTGACGACATCCCCCTTTCCGAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTT  
GCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAAAGCGCGGGGTGTGGTGGTTACGCGCA  
GCGTGACCGCTACACTTGCAGCGCCCTAGCGCCCGCTCCTTTCCGCTTTCTCCCTTCTTTCTCGCACGTTCCG  
CGGCTTTCCCGTCACTCTAAATCGGGGCTCCCTTAGGGTCCGATTAGTGCTTTACGGCACCTCGACCCCAA  
AAACTTGATTAGGTGATGTTACCGTAATGGCATCGCCTGATAGACGATTTTCGTTGGACGTGAAGTCACGTT  
CTTATAGTGGACTCTTGTTCAAACTGGATCA

>IS402-c\_B&Bclone

TAACCGGTTCTGGCCTTTGGCTGGCCTTTTGCTCAACATGTTCTTTCTGCGTATCCCCTGATCTGTGGATAACC  
GTATTACCGCCTTTGAGTGAGCTGAATCCGCTCGCCGACGCCAAGACCGAGCGCAGCGAGTCAGTGAGCGA  
GGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCTGGCCGATTCAATTAATGCAGCTGGCAGC  
ACAGGTTTCCCGACTGGAAAGCGGGCAGTGACGCGAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACC  
CCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAA  
ACAGCTATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTT  
GGGAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGGAATTCAGTAGTGATTGACGGAAGTGATGCAACCAG  
GCGATCGTGGTTCAACGACCCAGCGATGTTTGCCAAGGCCGCTGCGGTGTTCCGTACCGGCTGGCAATCA  
CTGGCTCGATAACCGGATTGCGCAATGCTCTTCGATGTCGCTCGGAGTCGTAACCGCGATCGGCGTAGACAAC  
GCGCGGTCTCTGCAGTGGGTGGCCACGCAATCCGCGAATCGGCGGAATCGCGTCGATCAGCGGCAGCAATTGC  
GTGACGTCGTGGACGTTGGCGCCAGTCAGGATCGCGGCGAGCGGCGTCCGTTGGCGTCCGCTGGGATAGGT  
GCTTGAACCGGGTCTGTGCGGATCAGTGGGGTTTTGGCCAGTTTTTGGTCCGCCCAACAGCACGAATCGAT  
GAAGAGTCGACAGCGGCTCGCGAGAAATCGATCTGGTCTGCCGCACGCAGCTTCGCGAGCAGCAACACGTGCA  
ACCGGTCCACACGCCAGCAGCTTGCCAGTCGCGAAGCCGTCGCCAACATGTCACGCCCGAGCCGCAACCCAT  
CTCGGCGGGCAAGTCGCGCCAGCGTAGTCCGGTCTTGAGACGTAGTCCGGTCTTGGGCGTAGTCCGGTCTTGA  
GGAATCGAATTTCCCGCGCCGCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGT  
CGTATTACAATTCACCTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCTAATCGC  
CTTGACGACATCCCCCTTTCCGAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTT  
GCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAAAGCGCGGGGTGTGGTGGTTACGCGCA  
GCGTGACCGCTACACTTGCAGCGCCCTAGCGCCCGCTCCTTTCCGCTTTCTCCCTTCTTTCTCGCACGTTCCG  
CGGCTTTCCCGTCACTCTAAATCGGGGCTCCCTTAGGGTCCGATTAGTGCTTTACGGCACCTCGACCCCAA  
AAACTTGATTAGGTGATGTTACCGTAATGGCATCGCCTGATAGACGATTTTCGTTGGACGTGAAGTCACGTT  
CTTATAGTGGACTCTTGTTCAAACTGGATCA

>ISBcen4-a\_B&Bclone

GGAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGGAATTCAGTAGTGATTGCTGCAGTAACCTGGAGGGGA  
TAGCTCTCGCAGTATGAGCGGCCCTGAGCTAGCAGAGCCAGATCGTCCGCGGATCTTGTGGCCAGTGCAA  
CGACTACGACATTGAGCGGACGCCGATGCTCAACTGCACTGCCAGGAACCCGGGTCTTTTGGCTGCTTCATG  
ACGGACCGTGCCTATGAATCATCAGCGCGCGAGGTATCGATCGCCTCTTGTGATGCCGAGCAGCCGGA  
CCTTGCCACCAGTTCCAAACTGCGCAGGAGCAGACCCAGCCATGCTGCAAATTCGGGCGCCGACCGAAACGC  
TTTGGGATCGCCATAGCTGCTACGGCAGCAGTGGCGGTGAGGAAACCCAAACCCGGGATTTCCGCAATCGTT  
CTGCAGGCCGGTTTTTCATGCAGCCACGCGCGGAGTCGCTGCTCGATGAGGGTTATCTGCTTGTGATGTCGAC  
AAGGCGTTGCCACTGTTCCCGCAGGGAATCGATCAAGACTCGCGGACGGCGCAATTCGAGGCGTAGCAGCAGG  
TCGGGCATGGCGGATTCATTGCTGCGCTCCACGCCATGACCTCTCCGCACTCGCCGAGCAAGCCTCTCAG  
GCAGTTAGATTGAGAGCGCCGAAATATAACCAACTGCTGTGCAATACGATGCAGTGCCAGAACCCTGTGG  
GCCTCGGTCTTACCGGGACGGCTTTACGCCCCGGTGTCTGTCAGCTGCCAGATTGCGCGCGCATCGGCGAC

ATCGTTCCTGTTTCCTGTACGAACGCTTTGACCAGTTTGCCTGGCATCAATTTGACCTGATGGCCGAGCTCGGT  
GAGACGGCGTGCCAGTGTGTGAGCCTCCGAGGCTCCATGCCGATGAGACAGGGCTCACGATTGGCGAAG  
TGCTCGAGAAAATCACTGCGGCGGAGCTGCCGGTGAACACCTCGCCGGTGTCCGCAGCGATCCAGTGCAGTT  
GGAAGACCCGCTGCGGATGTCTACAATCGAATTCCCAGCGCCGCCATGGCGGCCGAGC

>ISBcen4-b\_B&Bclone

CGCGTTGGGAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGGAATTCAGTGTGATTGCTGCAGTAACTGG  
AGGGCGATAGCTCTCGCAGTATGAGCGGCCATGAGCTAGCAGAGCCCAGATCGTCCGCGCATCTTGTGGCC  
AGTGCAACGACTACGACATTGAGCGGACGCCGATGCTCAACTGCACTGCCAGGAACCCGGGTCTTTGCGT  
GTTTCATGACGGACCGTGCGCCATGAATCATCAGCGTGCAGGTATCGATCGCCTCTCTGCTGATGCCGAGC  
AGCCGGACCTTGCCACCAGTTCCAAACTGCGCAGGGGCGAGACCCAGCCATGCTGCAAATTCGCGGCCGACC  
GAAAACGCTTTGGGATCGCCCATAGCTGTACGGCAGCAGTGGCGGTGAGGAACCCAAACCCGGGATTTCCGC  
AATCGTTCTGCAGGCGGGTTCATGCAGCCACGCGCGGAGTGCCTGCTCGATGAGGGTTATCTGCTTGTGCA  
TGTCGACAAGGCGTTGCCACTGTTCCACAGGAATCGATCAAGACTCGCGGCAGGCGCAATTCGAGGCGTAG  
CAGCAGGTGGGCATGGCGGATTATTGCTGCGCGTCCACGCCATGACCTCTCCGTACTCGCCGAGCAAG  
CCTCTCAGGCAGTTAGATTGAGAGCGCCGAAATATAACCAACTGCTGTGCAATACGATGCAGTGCCAGAACCG  
CTTGCTGGGCTCGGTCTTACCCGGGACGGCTTTCACGCCGGGTGTCTGTCAGCTGCCAGATTGCGCGCGCA  
TCGGCGACATCGCTCTTGTTCCTGTACGAACGCTTTGACCAGTTTGCCTGGCATCAATTTGACCTGATGGCCG  
AGCTCGGTGAGACGGCGTCCCAGTGTGTGAGCTCCGAGGCCATGCCGATGAGACAGGGCTCACGAT  
TGGCGAAGTCTCGAGAAAATCACTGCGGCGGAGCTGCCGGTGAACACCTCGCCGGTGTCCGCAGCGATCCA  
GTGCAGTTGGAAGACCCGCTTGGCGATGTCTACAATCGAATTCCCAGCGCCGCCATGGCGGCCGAGC

>ISBcen4-c\_B&Bclone

GTTGGGAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGGAATTCAGTGTGATTGCTGCAGTAACTGGAGG  
CGGATAGCTCTCGCAGTATGAGCGGCCATGAGCTAGCAGAGCCCAGATCGTCCGCGCATCTTGTGGCCAGT  
GCAACGACTACGACATTGAGCGGACGCCGATGCTCAACTGCACTGCCAGGAACCCGGGTCTTTTGGCGTGT  
TCATGACGGACCGTGCGCCATGAATCATCAGCGTGCAGGTATCGATCGCCTCTCTGCTGATGCCGAGCAGC  
CGGACCTTGCCACCAGTTCCAAACTGCGCAGGGACGAGACCCAGCCATGCTGCAAATTCGCGGCCGACC  
ACGCTTTGGGATCGCCCATAGCTGTACGGCAGCAGTGGCGGTGAGGAACCCAAACCCGGGATTTCCGCAAT  
CGTTCTGCAGGCGGGTTCATGCAGCCACGCGCGGAGTGCCTGCTCGATGAGGGTTATCTGCTTGTGATGT  
CGACAAGGCGTTGCCACTGTTCCCGCAGGGAATCGATCAAGACTCGCGGCAGGCGCAATTCGAGGCGTAGCAG  
CAGGTCCGGCATGGCGGATTATTGCTGCGCGTCCCACGCCATGACCTCTCCGTACTCGCCGAGCAAGCCTC  
TCAGGCAGTTAGATTGAGAGCGCCGAAATATAACCAACTGCTGTGCAATACGATGCAGTGCCAGAACCGCTT  
CTGGGCTCGGTCTTACCGGACGGCTTTCACGCCGGGTGTCTGTCAGCTGCCAGATTGCGCGCGCATCCGG  
CGACCTGTTCTTGTTCCTGTACGAACGCTTTGACCAGTTTGCCTGGCATCAATTTGACCTGATGGCCGAGCT  
CGGTGAGACGGCGTCCCAGTGTGTGAGCCTCCGAGGCTCCATGCCGATGAGACAGGGCTCACGATTGGC  
GAAGTGTCTGAGAAAATCACTGCGGCGGAGTGCCTGTCGAACACCTCGCCGGTGTCCGCAGCGATCCAGTGC  
AGTTGGAAGACCCGCTTGGCGATGTCTACAATCGAATTCCCAGCGCCGCCATGGCGGCCGAGCA

>ISBcen4-d\_B&Bclone

GGAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGGAATTCAGTGTGATTGCTGCAGTAACTGGAGGCGGA  
TAGCTCTCGCAGTATGAGCGGCCATGAGCTAGCAGAGCCCAGATCGTCCGCGCATCTTGTGGCCAGTGCAA  
CGACTACGACATTGAGCGGACGCCGATGCTCAACTGCACTGCCAGGAACCCGGGTCTTTTGGCGTCTTCATG  
ACGGACCGTGCGCCATGAATCATCAGCGTGCAGGTATCGATCGCCTCTCTGCTGATGCCGAGCAGCCGGA  
CCTTGCCACCAGTTCCAAACTGCGCAGGGACGAGACCCAGCCATGCTGCAAATTCGCGGCCGAGCCGAAACCG  
TTTGGGATCGCCCATAGCTGTACGGCAGCAGTGGCGGTGAGGAACCCAAACCCGGGATTTCCGCAATCGTT  
CTGCAGGCGGGTTCATGCAGCCACGCGCGGAGTGCCTGCTCGATGAGGGTTATCTGCTTGTGATGTGCGAC  
AAGGCGTTGCCACTGTTCCCGCAGGGAATCGATCAAGACTCGCGGCAGGCGCAATTCGAGGCGTAGCAGCAGG  
TCGGGATGGCGGATTATTGCTGCGCGTCCCACGCCATGACCTCTCCGTACTCGCCGAGCAAGCCTCTCAG  
GCAGTTAGATTGAGAGCGCCGAAATATAACCAACTGCTGTGCAATACGATGCAGTGCCAGAACCGCTTGG  
GCCTCGGTCTTACCCGGGACGGCTTTCACGCCGGGTGTCTGTCAGCTGCCAGATTGCGCGCGCATCGGCGAC  
ATCGTTCTTGTTCCTGTACGAACGCTTTGACCAGTTTGCCTGGCATCAATTTGACCTGATGGCCGAGCTCGGT  
GAGACGGCGTCCCAGTGTGTGAGCCTCCGAGGCTCCATGCCGATGAGACAGGGCTCACGATTGGCGAAG  
TGCTCGAGAAAATCACTGCGGCGGAGTGCCTGTCGAACACCTCGCCGGTGTCCGCAGCGATCCAGTGCAGTT  
GGAAGACCCGCTTGGCGATGTCTACAATCGAATTCCCAGCGCCGCCATGGCGGCCGAGC

>ISBcen4-e\_B&Bclone

GAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGGAATTCAGTGTGATTGCTGCAGTAACTGGAGGCGGAT  
AGCTCTCGCAGTATGAGCGGCCATGAGCTAGCAGAGCCCAGATCGTCCGCGCATCTTGTGGCCAGTGCAAC  
GACTACGACATTGAGCGGACGCCGATGCTCAACTGCACTGCCAGGAACCCGGGTCTTTTGGCGTCTTCATGA  
CGGACCGTGCGCCATGAATCATCAGCGTGCAGGTATCGATCGCCTCTCTGCTGATGCCGAGCAGCCGGAC  
CTTGCCACCAGTTCCAAACTGCGCAGGGACGAGACCCAGCCATGCTGCAAATTCGCGGCCGAGCCGAAACGCT  
TTGGGATCGCCCATAGCTGTACGGCAGCAGTGGCGGTGAGGAACCCAAACCCGGGATTTCCGCAATCGTT  
TGCAGGCGGGTTTTTCATGCAGCCACGCGCGGAGTGCCTGCTCGATGAGGGTTATCTGCTTGTGATGTGCA  
AGGGCTTGCACCTGTTCCCGCAGGGAATCGATCAAGACTCGCGGCAGGCGCAATTCGAGGCGTAGCAGCAGG  
CGGGATGGCAGTGTCCCGCAGGGAATCGATCAAGACTCGCGGCAGGCGCAATTCGAGGCGTAGCAGCAGG  
CGGGATGGCAGTGTCCCGCAGGGAATCGATCAAGACTCGCGGCAGGCGCAATTCGAGGCGTAGCAGCAGG  
CAGTTAGATTGAGAGCGCCGAAATATAACCAACTGCTGTGCAATACGATGCAGTGCCAGAACCGCTTGTGGG  
CCTCGGTCTTACCCGGGACGGCTTTCACGCCGGGTGTCTGTCAGCTGCCAGATTGCGCGCGCATCGGCGACA  
TCGTTCTTGTTCCTGTACGAACGCTTTGACCAGTTTGCCTGGCATCAATTTGACCTGATGGCCGAGCTCGGT  
AGACGGCGTGCCAGTGTGTGAGCCTCCGAGGCTCCATGCCGATGAGACAGGGCTCACGATTGGCGAAGT

GCTCGAGAAAATCACTGCGGCGGAGCTGCCGGTTCGAACACCTCGCCGGTGTCCGCGAGCGATCCAGTGCAGTTG  
GAAGACCCGCTTGCGGATGTCTACAATCGAATTCGCCGCGGCCCATGGCGGCCGGAGC

>ISBcen4-e\_B&Bclone

GGAGCTCTCCCATATGGTTCGACCTGCAGGCGGCCGGAATTCAGTACTGATTGCTGCAGTAACTGGAGGCGGA  
TAGCTCTCGCAGTATGAGCGGCCATGAGCTAGCAGAGCCAGATCGTCCGCGCGATCTTGTGGCCAGTGCAA  
CGACTACGACATTGAGCGGACGCCGATGCTCAACTGCACTGCCAGGAACCCGGGTCTTTTGGCTGCTTCATG  
ACGGACCGTGCGCCATGAATCATCAGCGTGGCAGGTATCGATCGCCTCTTGTGATGCCGAGCAGCCGGA  
CCTTGCCACCAGTTCCAACTGCGCAGGACGAGACCCAGCCATGCTGCAAATTCGCGGCCGGACCGAAACGC  
TTTGGGATCGCCATAGCTGCTACGGCAGCAGTGGCGGTTCAGGAACCCAACACCCGGGATTTTCGGCAATCGTT  
CTGCAGGCCGGTTCATGCAGCCACGCGCGGAGTCTGCTCGATGAGGGTTATCTGCTTGTGATGTCGAC  
AAGCGTTGCCACTGTTCCCGCAGGGAATCGATCAAGACTCGCGGACGGCGCAATTCGAGGCGTAGCAGCAGG  
TCGGGCATGGCGGATTCATTGCTGCGCGTCCCACGCCCATGACCTCTCCGTACTCGCCGAGCAAGCCTCTCAG  
GCAGTTAGATTGAGAGCGCCGAAATATAACCAACTGCTGTGCAATACGATGCAGTGCCAGAACCCTGTGCTGG  
GCCTCGGTCTTACCGGGACGGCTTTCACGCCGGGTGCTGTCAGCTGCCAGATTGCGCGCGCATCGGCGAC  
ATCGTTCTGTTTCCGTGCACGAACGCTTTGACCAGTTTGCCTGGCATCAATTTGACCTGATGGCCGAGCTCGGT  
GAGACGGCGTGGCCAGTGTGTGAGCCTCCGACGGCTCCATGCCGATGAGACAGGGCTCACGATTGGCGAAG  
TGCTCGAGAAAATCACTGCGGCGGAGCTGCCGGTTCGAACACCTCGCCGGTGTCCGCGAGCGATCCAGTGCAGTT  
GGAAGACCCGCTTCCGATGTCTACAATCGAATTCGCCGCGGCCCATGGCGGCCGGAGCA

>ISBcen6-a\_B&Bclone

GCAGGAAGTGATGGAAGTGTGACGGGCCCCGAGCGTCCGCGCGCTGGACGACGGAGCAGAAGCTGGCGAT  
GGTTCGCGAGAGTTTCGAGCCGGGAAATCGGTTTCGATGGTTCGCGCGCCAGCACGGCGTGAATCCGAACCAA  
CTGTTCCACTGGCGCAAGCTGTACCAGGTGGGAGCCTGTCAGCGGTCAAGGCTGGCGAGGAAGTGGTCCCGG  
CCTCGGAACTGGCCGACGCGCTCAAGCAGATTTCGCGAGTTGCAACGGATGCTCGGCAAAAAAAGACCATGGA  
GAATGAGATTCTTCGGAAGCAGTCCAGTACGGCCGGCAAAAAATGGATAGCGCACTCGCCCTCGTGGCG  
GAGGACGACCAGTAAACTGGTTTGTGAAGTTCTCGGCGTGTGCGCTCGAACGTATCGGCACGACTGTGCGG  
TCCGGCGACGTGGCGGATGGCCGTCAATCGAGGCAGAGTGACGACGCGGCCGCGGTTCGAGGAGATCCGCCG  
GGTCAATTGGCGATTTGCCAGCTATGGCTATCG

>ISBcen6-b\_B&Bclone

GCAGGAAGTGATGGAAGTGTGACGGGCCCCGAGCGTCCGCGCGCTGGACGCGGAGCAGAAGCTGGCGAT  
GGTTCGCGAGAGTTTCGAGCCGGGAAATCGGTTTCGATGGTTCGCGCGCCAGCACGGCGTGAATCCGAACCAA  
CTGTTCCACTGGCGCAAGCTGTACCAGGTGGGAGCCTGTCAGCGGTCAAGGCTGGCGAGGAAGTGGTCCCGG  
CCTCGGAACTGGCCGACGCGCTCAAGCAGATTTCGCGAGTTGCAACGGATGCTCGGCAAAAAAAGACCATGGA  
GAATGAGATTCTTCGGAAGCAGTCCAGTACGGCCGGCAAAAAATGGATAGCGCACTCGCCCTCGTGGCG  
GAGGACGACCAGTAAACTGGTTTGTGAAGTTCTCGGCGTGTGCGCTCGAACGTATCGGCACGACTGTGCGG  
TCCGGCGACGTGGCGGATGGCCGTCAATCGAGGCAGAGTGACGACGCGGCCGCGGTTCGAGGAGATCCGCCG  
GGTCAATTGGCGATTTGCCAGCTATGGCTATCG

>ISBcen8-a\_B&Bclone

TTCCTGAAGCCCGAACGGTTACGCGCGCCATTTACATCGAGTGGCGATGTACAAGCATGGGATCCGAAGATCA  
TGCTGCTCCTTCTGCCGTAACGGTCCGCGATTGACGTCGCAAAAGCTCACGTCGATGTTTCCGTAAGGGTGGC  
AATCTGGACATCCCGGATTCACCAACGATGCCGAGGGCCACTCGGCCCTGGCTGTTGCCCTCAGCCGCTGAA  
CGTGTAGTTGGTCTGATGGAGGCGACCGGTGGCTACGAGACCGAGCTGGCGTGGCCTTGAAGGTGCGGGT  
TTGCCGGTCTGCTGCTCAACCCTCGGCAAGCACGCGATTTTCCCGCTCGATGGGACGCTCTGGCCAAAACCGA  
CGCCATCGACGCACGTATGCTTGGCAACTGGCTTCCGTGCTACTGCGTCCGGACGATCTGTCCCGTTGCTAC  
GTCCCGTTCCCGACGAGCGGCAGCAGTGGCTGGCTGCACTGGTGACCCGACGGCGCAACTACTCACCATGCT  
CCTTTCGGAGCGACAGCGGCTCCAGATCACGCCGGTGGTCTGCATTCAAGTCTGCTGGCTATCATTGCCGCCA  
TCCAGAATCGAATTCGCCGCGGCCATGGCGCCGGGAGCATGGCAGCTCGGGCCCAATTCGCCCTATAGTG  
AGTCGTATTACAATTCAGTGGCCGTGTTTACAACGCTGCTGACTGGGAAAACCTGGCGTTACCCAATTAAT  
CGCCTTGCAGCACATCCCCCTTTCGCCAGTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACA  
GTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAAAGCGCGGGGGTGTGGTGGTTACGC  
GCAGCGTGACCGCTACACTTGGCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTTCTCCCTTCTTTCGCCACGT  
TCGCCGGCTTTCGCCGTAAGCTCTAAATCGGGGGTCCCTTTCAGGTTCCGATTTAGTGCTTACGCACTCGAC  
CCCAAAAATGATTAGGTGATGATCACGTATGGCATCGCCTGAATAGACGATTTTCGCCCTTACGTTGGAGTCC  
ACGTTCTTTAAATAGGTGGAATC

>ISBcen8-b\_B&Bclone

CTGAAGCCCGAACGGTTACGCGCGCCATTTACATCGAGTGGCGATGTACAAGCATGGGATCCGAAGATCATGT  
CTGCTCCTTCTGCCGTAACGGTCCGCGATTGACGTCGCAAAAGCTCACGTCGATGTTTCCGTAAGGGTGGC  
CTGGACATCCGTCGATTCAACAACGATGCCGAGGGCCACTCGGCCCTGGCTGTTGCCCTCAGCCGCTGAACGT  
GCAGTTGGTCTGATGGAGGCGACCGGTGGCTACGAGACCGAGCTGGCGTGGCCTTGAAGGTGCGGGTTG  
CCGGTGGCTGCTGCTCAACCCTCGGCAAGCACGCGATTTTCCCGCTCGATGGGACGCTCTGGCCAAAACCGACG  
CCATCGACGCACGTATGCTTGGCGAATGGCTTCCGTGCTACTGCGTCCGGACGATCTGTCCCGTTGCTACGT  
CCCCTTCCCGACGAGCGGACGAGTGGCTGGCTGCACTGGTGACCCGACGGCGCAACTACTACCATGCTCC  
TTTCGGAGCGACAGCGGCTCCAGATCACGCCGGTGGTCTGCAATTCAGTCTGCTGGTATCATGCGGCCATC  
CAGAATCGAATTCGCCGCGGCCCATGGCGGCCGGGAGCATGGCAGCTCGGGCCCAATTCGCCCTATAGTGAG  
TCGTATTACAATTCAGTGGCCGTGCTTTTACAACGCTGCTGACTGGGAAAACCTGGCGTTACCCAATTAATCG  
CCTTGCAGCACATCCCCCTTTCGCCAGTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGT  
TGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAAAGCGCGGGGGTGTGGTGGTTACGCGC  
AGCGTGACCGCTACACTTGGCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTTCTCCCTTCTTTCGCCACGTTCC

CCGGCTTTCCCGTCAAGCTCTAAATCGGGGGCTCCCTTAGGGTCCGATTTAGTGCTTACCACCTCGACCCGA  
AAACTTGTATTAGGTGATGTTACGTATGGGGCATCGCCTGAATGACGTTTTCCGCATGACGTGGAGTCACGT  
CTTAATATGGATCTTGTCAACTGGACAACCTCATCCTATTTCCGGCTATTTCTTTGAGATATTA

>ISBcen8-c\_B&Bclone

CTGAAGCCCGAACGGTTACGCGCGCCATTTACATCGAGTGCATGTACAAGCATGGGATCCGAAGATCATGT  
CTGCTCCTTCTGCCGTAACGGTCCGCATTGACGTCGCAAAAGCTCACGTCGATGTTTCCGTAAGTGGTGGCGAT  
CTGGACATCCGTCGATTCAACAACGATGCCGAGGGCCACTCGGCCCTGGCTGTTGCCCTCAGCCGCTGAGCGT  
GCAGTTGGTTCGTGATGGAGGCGACCGGTGGCTACGAGACCGAGCTGGCGTGCCTTGAAGGTGCGGGTTG  
CCGGTCCGCTGTCGTAACCTCGGCAAGCACGCGATTTTGGCCGCTCGATGGGACGCTCTGGCCAAAACCGACG  
CCATCGACGCACGTATGCTTGCCGAACCTGGCTTCCGTCGTAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT  
CCGTTCCCGACGAGCGGCGAGCAGTGGCTGGCTGCATGGTGACCCGACGGCGCCAATACTACTCACCATGCTCC  
TTTCGGAGCGACAGCGGCTCCAGATCACGCCGGTTGGTCTGCATTCAAGTCTGCTGGCTATCATTGCCGCCATC  
CAGAATCGAATTCCCGCGGCCGCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCCAATTGCCCTATAGTGAG  
TCGTATTACAATTCAGTGGCCGTCGTTTACAACGTCGTGACTGGGAAAACCTGCGTTACCCAATAATCGCC  
TTGCAGCACATCCCCCTTTTCGAGCTGGCGTATAGCGAGAGCCACGATCGCTTCCCAACAGTTGCGCAGCTG  
ATGGCGATGACGCGCCCTGTAGCGGCGCATTAGCGCGCGGGTGTGGTGTACGCGCAGCGTGACGCTAACACT  
TGCAGCGCCTTAGCGCCCGCTTCTTTTCGCTCCT

>ISBcen8-d\_B&Bclone

TTCCTGAAGCCCGAACGGTTACGCGCGCCATTTACATCGAGTGCATGTACAAGCATGGGATCCGAAGATCA  
TGTCTGCTCCTTCTGCCGTAACGGTCCGCATTGACGTCGCAAAAGCTCACGTCGATGTTTCCGTAAGTGGTGGC  
AATCTGGACATCCGTCGATTCAACAACAATGCCGAGGGCCACTCGGCCCTGGCTGTTGCCCTCAGCCGCTGAA  
CGTGCAGTTGGTTCGTGATGGAGGCGACCGGTGGCTACGAGACCGAGCTGGCGTGCCTTGAAGGTGCGGGT  
TTGCCGGTCCGCTGTCGTAACCTCGGCAAGCACGCGATTTTGGCCGCTCGATGGGACGCTCTGGCCAAAACCGA  
CGCCATCGACGCACGTATGCTTGCCGAACCTGGCTTCCGTCGTAAGTGGTGGTGGTGGTGGTGGTGGTGGT  
GTCCTTCCCGACGAGCGGCGAGCAGTGGCTGGCTGCATGGTGACCCGACGGCGCCAATACTACTCACCATGCTC  
CCTTTCGGAGCGACAGCGGCTCCAGATCACGCCGGTTGGTCTGCATTCAAGTCTGCTGGCTATCATTGCCGCCA  
TCCAGAATCACTAGTGAATTCGCGGCCCGCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCA  
TAGCTTGAGTATTCTATAGTGTACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCCTGTGTGAAATTG  
TTATCCGCTCACAATTCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTG  
AGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTCCAGTCGGGAAACCTGTGCTGCCAGCTGCATTA  
ATGAATCGGCCAACGCGCGGGGAGAGGGCGGGTTTGGCTATTGGGCGCTCTTCCGCTTCCCTCGCTCACTGACTCG  
CTGCGCTCGGTGCTTCCGCTGCGCGAGGGTATCAGCTCCCTCAAAGGCGGTAATACGGTTATCCACAGATCAG  
GGATACGCAGAAAGAATGTGAGCAAAGCCAGCAAGCCAGGACCCGTTAAAAGGCCGCGTTGCTGCGCGA  
CTCA

>ISBcen9-a\_B&Bclone

GTACGCCACATACCAGGCTCATGGTAAGGACGGCGTGCCTAGGAAGTACAACCTCATAACGATGTGGCGTTCAGG  
CTCAAAGTTGTTTCAGCACATGCGCGAGCATGGCGTGTGAGCAAGGAAGCGGCGCGCGGTTCAATATTCGCA  
ATCCGAGCGCGGTGCTCGAATGGGCTCGCCGCTATGATGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT  
AGGCGCGCGCCAAACCGCAATGCCAAAAACACCGCTGCTCAGCCTCTGAACCCTACCGATGGCAGCGGATACC  
CGTACCCGCGAGCTGCTCAGGAACTGAACCTACCTGCGCATGGAGAACGCGTATCTAAAAAAGCTGAGGG  
CCTTGGTTCAGGCGCAGGCTGTACCGCGCAAAAAGCGCAAAATAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT  
GCTGGCTTGTCCGGTGGCCGACTGGCTCGCAGCACGTTTTATTACCAGTGCAAGGCGCTGGCCGCTCCCGA  
TCGACATGCGAGTGTGAAAGCGAAAATTTCGAGCACTGTTTCGAGCAGCACAAGGGGCGCTATGGCTATCGACGC  
ATCACGCTGGCGCTGCGCGCTCTGGGCGAGATGATTAACCACAAGACGGTTCACGCGCTGATGCGCAAGATGC  
AGTTGAAGTCTGCGTGGCGGTCAAGAAGTACCGTGGCTATAGGGGCAACACGTCGAAGACGGCCCCGCATCT  
GCTCCAGCGCCAGTTCCATGCCGCCGTCGGAACGAGAAGTGGTTACCGACATCACCGAGTTACGCGTAGGT  
GGCCAAAAGCTGTATCTGTACCAAGTGTGGATCTGTACAACGGGAGATCATCGCGTACCAGACGCATACTC  
GGCCAGCCTTCCAGATGGTACTGACATGTGCGCAAGGCTTTGCG

## 2. Séquences ADN obtenues après amplifications des régions flanquantes des IS407

>locus\_IS407-a\_H111

TCAACCTCGGGATACTGTTTCAGCTTCTTAGGACGTGACTTACTATAGGCGTTAGCCATCTTCCGACTCCTTGTG  
CGAGTTGGTGGTGGTTCAGCGGGTTCGATGGGTGGCAGCCATGCGGCCCGCGCTTACTGCTTTGCTCGCCTC  
AGCCTTCTTCTGATTTCGAGCTTTGAAGCGAGACGCCACTCCTTATCGGTGGCGTAGCAAGTAGTTTAGTGAT  
ATCGCATCTCATAAATCCGTTGCAATTGGACTAGCTAAACGGCAGGTGCGACGATCAACGTGAGCCGAATG  
GCATAAGGAAGTTCTCACGCACGGCCAGTCTAGATTGGCGACTACAATTGAGTGTGGCGAACTAAATCCGCT  
GCAGAATGCGAGCGAGACACTGCCTGAGTGAAGACAGAGGGCCCTAGTGCAACCATTTTCCATCACAC  
TATTTGCAACACCGGGGATCCGGAGGGAATTGCTCACCTCGACAAATCGAACTGGTGGGTTACGGGGTTGT  
CTTACGAGGGAACCTTTCATCTGCTGAAGCAGGAGCCCGCATCTCTCAGGCAGGTATTTACATCTTAGTTG  
GTAACGACGAGGAGACGATTTACATCGGAGAGGACAGCCAGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT  
CAAACAAGGAGGATGGGTCTGGGGCGTGTATTTCTTTGACAGAAATCACAAAATCGGAAAGACAGAGGTCC  
AATTTCTAGAATCTGAACTGGTGCATTGGCGAAGAAGCACGGTCTGCAATCTTACTGAACAAAACAATCG

>locus\_IS407-a\_LMG21461

GGATACTGTTTGGCTTCTTGGGACGTGACTTACTATAGGCGTTAGCCATCTTCCGACTCCTTGTCCGAGTTGGT  
GGTGGTTCAGCGGGTTCGATGGGTGGCAGCCATGCGGCCCGCGCTTACTGCTTTGCTCGCCTCAGCCTTCT

CCTGATTTCGAGCTTTGAAGCGAGACGCCACTCCTTATCGGTGGCGTAGCAAGTAGTTTAGTGCATATCGCATCT  
CATAAATCCGTGTTGCAATTGGACTAGCTAAACGGCAGGTGCGACGATCAACGTGAGCCGAATGGCATAAGGA  
AGTTCTCACGCACGGCCAGTCTAGATTGGCGACTACAATTGAGTGTTGGCGAACTAAATCCGCTGCAGAATGC  
GAGCGAGACACTGCCTGAGTGACTTGAAGACAGAGGGGCTAGTGCAACCATTTTCCATCACACTATTTGCAA  
CCACCGGGGATCCGGAGGGAATTCGTACAGTCGACAAATCGAACTGGTTCGGGTTACGGGGTTGTCTTCACGAG  
GGAACTCTTTTCATCTGCTGAAGCAGGAGCCCGGCATCTCTCAGGCAGGTATTTACATCTTAGTTGGTAACGCAG  
CAGAGGAGACGATTTACATCGGAGAGGCAGACCCAGTTGGTGATCGACTAAAAGAAATCACGTGCGAAACAAGG  
AGGGATGGGTCTGGGGCGTGTATTTCTTTGACAGAAATCACAAAATCGGAAAGACAGAGGTCCAATTTCTAGA  
ATCTGAACTGGTTCGATTGGCGAAGAAGCACGGTTCGTGCAATCTTACTGAACAAAAC

>locus\_IS407-a\_LMG19240

GGGATACTGTTTGGAGCTTCTTGGGACGTGACTTACTATAGGCGTTAGCCATCTTCCGACTCCTTGTCCGAGTTG  
GTGGTGGTCAGCGGGTTCGTATGGGTGGCAGCCCATGCGGCCCGCGCTTACTGCTTTGCTCGCCTCAGCCTTC  
TTCCTGATTTCGAGCTTTGAAGCGAGACGCCACTCCTTATCGGTGGCGTAGCAAGTAGTTTAGTGCATATCGCAT  
CTCATAAATCCGTGTTGCAATTGGACTAGCTAAACGGCAGGTGCGACGATCAACGTGAGCCGAATGGCATAAG  
GAAGTTCTCACGCACGGCCAGTCTAGATTGGCGACTACAATTGAGTGTTGGCGAACTAAATCCGCTGCAGAAT  
GCGAGCGAGACACTGCCTGAGTGACTTGAAGACAGAGGGGCTAGTGCAACCATTTTCCATCACACTATTTGC  
AACCACCGGGGATCCGGAGGGAATTCGTACAGTCGACAAATCGAACTGGTTCGGGTTACGGGGTTGTCTTCACG  
AGGGAACTCTTTTCATCTGCTGAAGCAGGAGCCCGGCATCTCTCAGGCAGGTATTTACATCTTAGTTGGTAACGC  
AGCAGAGGAGACGATTTACATCGGAGAGGCAGACCCAGTTGGTGATCGACTAAAAGAAATCACGTGCGAAACA  
GGAGGGATGGGTCTGGGGCGTGTATTTCTTTGACAGAAATCACAAAATCGGAAAGACAGAGGTCCAATTTCTA  
GAATCTGAACTGGTTCGATTGGCGAAGAAGCACGGTTCGTGCAATCTTACTGAACAAAACAT

>locus\_IS407-a\_LMG12614

TTGAGCTTCTTGGGACGTGACTTACTATAGGCGTTAGCCATCTTCCGACTCCTTGTCCGAGTTGGTGGTGGTCA  
GCGGGTTCGTATGGGTGGCAGCCCATGCGGCCCGCGCTTACTGCTTTGCTCGCCTCAGCCTTCTTCTGATTCCG  
AGCTTTGAAGCGAGACGCCACTCCTTATCGGTGGCGTAGCAAGTAGTTTAGTGCATATCGCATCTCATAAATCC  
GTGTTGCAATTGGACTAGCTAAACGGCAGGTGCGACGATCAACGTGAGCCGAATGGCATAAGGAAGTTCTCAC  
GCACGGCCAGTCTAGATTGGCGACTACAATTGAGTGTTGGCGAACTAAATCCGCTGCAGAATGCGAGCGAGAC  
ACTGCCTGAGTGACTTGAAGACAGAGGGGCTAGTGCAACCATTTTCCATCACACTATTTGCAACCACCGGGG  
ATCCGGAGGGAATTCGTACAGTCGACAAATCGAACTGGTTCGGGTTACGGGGTTGTCTTCACGAGGGAATCTT  
TCATCTGCTGAAGCAGGAGCCCGGCATCTCTCAGGCAGGTATTTACATCTTAGTTGGTAACGCAGCAGAGGAG  
ACGATTTACATCGGAGAGGCAGACCCAGTTGGTGATCGACTAAAAGAAATCACGTGCGAAACAAGGAGGGATGG  
GTCTGGGGCGTGTATTTCTTTGACAGAAATCACAAAATCGGAAAGACAGAGGTCCAATTTCTAGAATCTGAAC  
TGGTTCGATTGGCGAAGAAGCACGGTTCGTGCAATCTTACTGAAC

>locus\_IS407-c\_LMG18826

CCCTCAAGCGCGGATCAAAGTGAAGTAGCTGTGCCAGGCAGCTGCAAGCCGCTCCACTGAACCTCACGGC  
GGTCTGTTACCTGCCTGATCGTTTGTCCGATCAGCTTGCTTCCCTTCGGAATGGCAACGTTGCCAAAACGGTCA  
TAGACGTTGAGCGACACGATGCCGATCATCATGCTGACCTGCCCTGCAAACAGGGCCAGCCGGAGTCTAGT  
AAAGTTCGTTTTCCGAGAAGAAGACGAACATGAAGAAGCGCTTTACGGAACAGCAAATCATCGGGTTTCTGAA  
GGAAGCCGAGGCCGATGCCGCTCAAGGAAGTGTGCAGGAAGCAGCGGTTCAAGTATGCGTCTGAGGTTGAGT  
TGGCTGCGAAGTTCGGCGCATGGAGTCTCGGAAGCCCGGCTCAAGAACCTCGAGGTGGAGAATGCCC  
GGCTGAAGAAGCTGCTGGCCGAAGCAATGCTCGATATGGAAGCGTTGAAGGTCGTGCGTCAAGGAAAGCCCT  
GAGCCCGCAAGCCAAACGCGAAGCAGTATCGGCGATTTCGGGGGAAGGTCAACATCTCCGAGCGCCGCGCTG  
CCGGCTTGTCCGGCTTTCTCGCAGCGTGTGCTGATTACGACGCGAAGCCAGACCACGAGAATGAGGTGCTAGCG  
GCCGCTGCTGGAGTTGGCACACGAACGTCGTCGATTCCGCTACCGCCGATTGCACGCCCCTAGTGAACCGG  
AAGGCACGCACGCCAATACAAGCGCATCTATCGCTGTACCGTGAGGCAGGGCTGGCTGTGCGGCGCCGCTCG  
CAAGCGCCACGGCGTAATGATTGAGCGCGAGCAACTGGCATTGCCGGGCGCCCCAACGAGGTGTGGTCAATC  
GATTTCTGATGGATGCGCTTTCCAACGGCCGGCGCTGAAGTGCCTGACCGTCTGACGACTTCACGAAAG  
AGGCTGTGACATCGTTGTGACCATGGCATCTCAGGTTTGTATGTCGCTCGGGCATTGGACCGCGCAGCTCGC  
TTCCGTGGCTATCCCAAGCGGTGCGAACAGACCAGGGCCCCGAATTTACGAGCCGCGCGCTTGACCAGTGGG  
CGTATGCGAACGGCGTCTGCTCTGAAGTTGATTACGGCGGGCAAGCCAACGCAGAATGCGTACATCGAATCGTT  
CAACGGCAAGTCCGCGACGAATGCCTAACGAGCACTGGTTCACGACGCTCGCGCATGCTCGGGCAGTCATC  
GCGGCATGGCGTCAGGACTACAACGAGCAAAGGCCGACAGCGCACTGAACTACCTTGCGCCGTCAGAGTTTG  
CGGGCAAACATCGGGCAACCGCGGATACTCCTGCCGCTTTCCAGGAGTTGGTTTAAAGGGACTTTGCTAGAAG  
CCCATTGGCCCTATTGATGGGGCAGGTCAATGCCGCAATTCGGGATCACCCGTTTCCAAGCCAACAGCGA  
TGGGAGTGTCTCCGTGATTTCTGACGGCCGATGTGATGGAATTTCCGCCAGATGGGCGCATCCGGATCCGTG  
CCATCGGTGGCTTGGGCCGCGCTCCATCCGAGAGCATCAACATCACACGCACGACAGGATGATCAGGGGCA  
ACTTCACGGCGAACTTTTCCGTTGCTGAAGGGTCGGTGCCAGCCACAGGCGAGCAAATACAGTGTCTTGATCC  
GATCATAACGAGACTCGATATCAGGTGACAGCCACATGTGGCGCGA

>locus\_IS407-c\_B&Bclone

ATTCCGATCGCACAGCACGGCCATCGCGCCACATGTGGCTGTACCTGATATCGAGTCTCGTATGATCGGATCA  
AGACACTGTATTTGCTCGCCTGTGGCTGGCACCGACCCTTCAGCAACGGAAAAGTTCGCCGTGAAGTTGCCCT  
GATCATCTGTCTGTGCGTGGTGATGTTGATGCTCTCGGATGGAGCGCGGGCCCAAGCCACCGATGGCACGGAT  
CCGGATGCGCCCATCGTGCGCAAATTCCATCACATCCGGCCGTACGAAATCACGGAGGACACTCCCATCGCTG  
TTGGCTTGAAACGGGTGATCCCGGAATTGGCGGCATGATGATCGGCATCGTGTGCTCAACGTCTATGACCGT  
TTTGCAACGTTGCCATTCCGAAGGGAAGCAAGCTGATCGGACAAACGATCAGGCAGGTGAACGACCGCCGTG  
AGGTTCACTGGAGCGGCTTGCAGCTGCCTGGCACAGCTACTTTGCAGTTTGATCCGCCGCTTGAGGGTTGTATG  
CCGGATGGTTCGGCCGGTATCACTG

>locus\_IS407-c\_LMG21461

GGCATTCCGATCGCACAGCACGGCCATCGCGCCACATGTGGCTGTACCTGATATCGAGTCTCGTATGATCGGA  
TCAAGACACTGTATTTGCTCGCCTGTGGCTGGCACCGACCCTTCAGCAACGGAAAAGTTCGCCGTGAAGTTGCC  
CCTGATCATCTGTCTGTGCGTGGTGATGTTGATGCTCTCGGATGGAGCGCGGGCCCAAGCCACCGATGGCACG  
GATCCGGATGCGCCATCGTGCGCAAATTCCATCACATCCGGCCGTACGAAATCACGGAGGACACTCCCATCG  
CTGTTGGCTTGAAACGGGTGATCCCGGAATTGGCGGCATGATGATCGGCATCGTGTGCTCAACGTCTATGAC  
CGTTTTGGCAACGTTGCCATTCCGAAGGGAAGCAAGCTGATCGGACAAACGATCAGGCAGGTGAACGACCGCC  
GTGAGGTTCACTGGAGCGGCTTGCAGCTGCCTGGCACAGCTACTTTGCAGTTTGATCCGCCGCTTGAGGGTTGT  
ATGCCGGATGGTTCGGCCGGTATCACTG

>locus\_IS407-d\_B&Bclone

GAGATCGCAGGTAGCTTAGATTAGATTGATAGATAGTTACCGGTACGGCTGGCGAAGCCTTCGACTACAAC  
ACGCTTAAGCCACGTTTCGAGCTCTTTGCGCGAGCCACCTTCAACACCTCGGGGACGAAGCTGCGAATAAAG  
CGCTCGTCGGCCGACCTCTCAAACGAGTGGTGCTGATCGGCGAACAGAACGACT

>locus\_IS407-d\_LMG19240

CGGTACGGCTGGCGAAGCCTTCGACTACAACACGCTTAAGCCATGTTTCGAGCTCTTTGCGCGAGCCACCTTC  
AACACCTCGGGGACGAAGCTGCGAATAAAGCGCTCGTCGGCCGACCTCTCGAACGAGTGGTGCTGATCGGCGA  
ACAGAACGACT

>locus\_IS407-d\_LMG21461

GAGATCGCAGGTAGCTTAGATTAGATTGATAGATAGTTACCGGTACGGCTGGCGAAGCCTTCGACTACAAC  
ACGCTTAAGCCACGTTTCGAGCTCTTTGCGCGAGCCACCTTCAACACCTCGGGGACGAAGCTGCGAATAAAG  
CGCTCGTCGGCCGACCTCTCAAACGAGTGGTGCTGATCGGCGAACAGAACGACT

>locus\_IS407-e\_LMG12614

GGGCCTCGTCGCAAACCATGACCGACCACTTCTGCGCAGCTAAGGAGAATTCCAGGTGCGCGCAATGTCTCATA  
GGTCGTGACACCACGTTTGCCGAACCGAGCCAGTCCGATGCTAAGAGCTTGGGCGAACCCTGACTCAATCAAC  
CCGTCATCAAGCGCGGACCTCGGCGCTCGCTTTGCCGCGACGTTCCGGCCATACAGCGTGAGGACATCCATCG  
CTCCCTCAACGAAAAATTTCTCGATTTCCGCTTCCAATTGTCCAGCAACGAAACGGGCGCGACGACCAAGAA  
TGGATCGATGCGGGGCTC

>locus\_IS407-e\_LMG18826

ATCGAGAAATTTTCGTTGAGGGAGCGATGGATGTCTCACGCTGTATGGCCCGAACGTGCGGGCAAAGCGAG  
CGCCGAGGTCCGCGCTTGATGACGGTTGATTGAGTCGGTTTCGTCTGACCTGCCCCATCAATAGGGCCAATG  
GGCTTCTAGCAAAGTCCCTTTAAACCAACTCCTGAAAGCGGCAGGAGTATCCGCGGTTGCCGATGTTTCGCC  
GCAAACCTGACGGCGCAAGGTAGTTCACTGCGCTGTGCGGCTTTGCTCGTTGTAGTCTGACGCCATGCCGC  
GATGACTGCCCCGAGCATGCGCGAGCGTCTGAACCACTGCTCGTTAAGGCATTCTGCGGGAACCTTCCGTTG  
AACGATTGATGATGACGATTCTGCGTTGGCTTGGCCGCTGAATCAACTTCAGAGCGACGCCGTTTCGCATACGC  
CCACTGGTCAAGCGCGCGGCTCGTAAATTCGGGGCCCTGGTCTGTTTCGCACCGCCTTGGGATAGCCACGGAAG  
CGAGCTGCGCGTCCAATGCCCGAGCGACATACAAACCTGAGATGCCATGGTCGACAACGATGTCGACAGCCT  
CTTTTCGTGAAGTCGTGACGACGGTCAGGCACTTACGCGCCGGCCGTTGAAAGCGCATCCATCACGAAAT  
CGATTGACCACACCTCGTTGGGGGCGCCCGCAATGCCAGTTGCTCGCGCTCAATCATTACGCCGTGGCGCTTG  
CGACGGCGCCGACAGCCAGCCCTGCCTCACGGTACAGGCGATAGATGCGCTTGTGATTGGCGTGCCTGCTT  
CGCGTTCCACTAGGGCGTGCAATCGCGGTAGCCGAATCGACGACGTTCTGTGCCAACTCCACCAGACGCGC  
CGCTAGCACCTCATTCTCGTGGTCTGGCTTTCGCTCGTAATGCAGCACGCTGCGAGAAAGCCCGACAAGCCGG  
CAGGCGCGGCGCTCGGAGATGTTGACCTTCCCCGAATCGCCGATACTGCTTCCGCTTTGGCTTTCGGGGCTCAG  
GGCTTTCCCTGACGACGACCTTCAACGCTTCCATATCGAGCATTGCTTCGGCCAGCAGCTTCTTCAGCCGGGC  
ATTCTCCACCCTCGAGTTCTTGAGCCGGCGGGCTTCCGAGACCTCCATGCCGCCGAACTTCGCACGCCAAGTG  
GTAGAACGACGCATCACTGAACCCGTGCTTCTGACAGTTCCTTGACCGGCATACCGGCCTCGGCTTCTTCA  
GAAACCCGATGATTGCTGTTCGTAAAGCGCTTCTCATGTTCTGCTTCTTCTCCGAAAACGAACTTTACTAGA  
CTCCGGCTGGCCCTGTTTGCAGGGGGCAGGTCACGCCAAGCTTTAGCATCGGACTGGCTCGGTTCCGGCAA  
CGTGGTGCTGACGACCTATGAGACATTGCGCGACCTGGA



>locus\_IS407-e\_LMG21461

GGGCCTCGTGCCAAACCATGACCGACCACTTCTGCGCAGCTAAGGAGAATTCCAGGTTCGCGCAATGTCTCATA  
GGTCTGTCAGCACCACGTTTGCCGAACCGAGCCAGTCCGATGCTAAGAGCTTGGGCGAACCCGACTCAATCAAC  
CCGTCATCAAGCGCGGACCTCGGTGCTCGCTTTGCCGCGACGTTTCGGGCCATACAGCGTGAGGACATCCATCG  
CTCCCTCAACGAAAAATTTCTCGATTTTCGGCCTTCCAATTGTCCAGCAACGAAACGGGCGCGACCAAGAA  
TGGATCG

>locus\_IS407-f\_B&Bclone

CCTGTAGATCGGCGCCGTGTGAAAGCACTGGCCGCGATCCAGAGCGACATGTTGGATTCTCGTGTGTGATACA  
CACAACCTAAATTAAGTCTTCTTCCAAGATTGGTTGCGCTGCGAAGCAACGCAACAACCCCTTTGCTGATGA  
CCATAGCGAGTCCGTTCCACCCCTTCCCATCCCGAACAGGACCGTGAACGACTCTACGCCGATGATAGTGCG  
GATTTCCCGTGTGAAAGTAGTAATCGTCAGGCTCCCTAAGCCAGAAACCCCGCCCCGACAAGGCGGGGTTTT  
TGCATTGGCGTGCCAGAAAAGCGTGCGGCGAAGCGCGAAACGGTATGCGCGCCAGATCAGATCCTCCGGCTCC  
CGCCACCAAGCCGTCCCCCGAACAAGTCCCAGGATTTGATACGATCGAACTGTCGTGACAATGGCATCCGC  
CGACGTTTCGTGTTT

>locus\_IS407-f\_LMG21461

CTGTAGATCGGCGCCGTGTGAAAGCACTGGCCGCGATCCAAAGCGACAAGTTGGATTCTCGTGTGTGATACAC  
ACAATCAAATTAAGTCTTCTTCCAAGATTGGTTGCGCTGCGAAGCAACGCAACAACCCCTTTGCTGATGA  
CCATAGCGAGTCCGTTCCACCCCTTCCCATCCCGAACAGGACCGTGAACGACTCTACGCCGATGATAGTGCG  
GATTTCCCGTGTGAAAGTAGTAATCGTCAGGCTCCCTAAGCCAGAAACCCCGCCCCGACAAGGCGGGGTTT  
TTTGCATTTGAGCGGCGCAAAAATAGCATGCAGGAAGGCGCGCCCTTCTCGTACCGGCGACGTGCGCGCCAGC  
TCAGCTCGGCTGCGTTCCGCCACCAACCCCGCCCCGAAACGAACTCCCCATTTGATACGATCTACGTGTCT  
CAACGTCGGCATCCGTCGGCGTTCGTGTTT

>locus\_IS407-g\_LMG12614

TTAGTACTGCGCGAAGGCATGTTTTACCATAGAGCCGATGGTGAACCAAGGACGACGTGCCATTCGAACTG  
AGGACGACGGCTGGACGGTCTCACGCGCGACGGCAACTGTGCGCTCAGTTCGAGCATACCGTTGCCGTCAC  
CGGGAGCGCGTGCCTGTTGACTGCGCCCCGGGCGAAAAAGCGTTGCACTGATCCAGTGTGGTAGCGATA  
CGGACTTTGCTCAGGACCTGACTTTGCCTTCGGATAAGAACACGCCGTTGCCGGCTGCTCTCGGTGCGTGC  
AAGGTAGTTGTCGCTGATCTCATGA

>locus\_IS407-g\_LMG18826

GACGACGTGCCATTCGAACTGAGGACGACGGCTGGACGGTCTCACGCGCGACGGGCAACTGTGCGCTCAGTT  
CGAGCATACCGTGACCTGCCCCCTGCAAACAGGGCCAGCCGGAGTCTAGTAAAGTTCGTTTTCGGAGAAGAAG  
ACGAACATGAAGAAGCGCTTTACGGAACAGCAAATCATCGGGTTTTCTGAAGGAAGCCGAGCCGGTATGCCG  
GTCAAGGAAGTGGCAGGAAGCACGGGTTGAGTGTGCGTCTACACTTGGCGTGCGAAGTTTCGGCGGCA  
TGGAGGTCTCGGAAGCCCGCGGCTCAAGAACCTCGAGGTGGAGAATGCCCGGCTGAAGAAGCTGCTGGCCG  
AAGCAATGCTCGATATGGAAGCGTTGAAGGTCGTGTCAGGGAAAGCCCTGAGCCCGCAAGCCAAACCGCA  
AGCAGTATCGGCGATTTCGGGGGAAGGTCAACATCTCCGAGCGCCGCGCCTGCCGGCTTGTCCGGCTTTCTCGC  
AGCGTGCTGCATTACGACGCGAAGCCAGACCACGAGAATGAGGTGCTAGCGGCGCGCTCTGGTGGAGTTGGCAC  
ATGAACGTCGTCGATTCGGCTACCGCCGATCGCCGCAACGGCGTGTCTTATCCGAAGGCAAAGTCAAGTCC  
TGAGCGAAAGTCCGTATCGTACCACACTGGATCAGTGAACGCTTTTTCGCCCGGGCGCATGTCAAGACAC  
GCACGCGACTCCCGGTGACGGCAACGGTTGAGTGTCCCCCAATAGGGCCAATGGGCTTCTAGCAAAGTCC  
CTTTAAACCAACTCCTGGAAGCGGCAGGAGTATCCGCGGTTGCCGATGTTTCGCCGCAAACCTCTGACGGCG  
CAAGGTAGTTCAGTGCCTGTGCGGCCCTTTGCTCGTTGTAGTCTGACGCCATGCCGCGATGACTGCCCGAGCA  
TGCGCGAGCGTCTGTAACCAGTGTCTGTTAAGGCATTCGTCGCGGAACTTGCCGTTGAACGATTCGATGTACGC  
ATTCTGCGTTGGCTTGCCCGCTGAATCAACTTCAGAGCGACGCGGTTGCGCATAACGCCACTGGTCAAGCGCGC  
GGCTCGTAAATTCGGGGCCCTGGTCTGTTGCAACCGCCTTGGGATAGCCACGGAAGCGAGCTGCGCGGTCAA  
TGCCCGAGCGACATACAAACCTGAGATGCCATGGTGCACAACGATGTCGACAGCCTTTTCGTGAAGTCGTG  
ACGACGGTCAGGCACTTACGCGCCCGGCTTGGAAAGCGCATCCATCACGAAATCGATTGACCACACCTCGT  
TGGGGGCGCCCGCAATGCCAGTTGCTCGCGCTCAATCATTACGCCGT

>locus\_IS407-i\_LMG12614

CTAAGTTCGATATTGGTCAGAGTACTAGATCGAGCGAGAACGACATGAAGCATCGGCCACGACGGGGAATAT  
AGATGTATTTCCGCCGTGGCTTTAATCTGCTATTGGAGCGGCCCCACGTATCTGAGGACACGAGGACTCTCT  
TAAAATAAGGGATAGTCTTGTGCCTATCAGTAAGCCTGGTCAATTACGTGGAAAGCATCGAAATTTGACCGAG  
CCGGAGCGCCGTCGGCGGCGCACGGCGCAAGAAAAAATGCCATCGTGCAGGAAACATTGGAGCCGGGAGCA  
TCGGTGTCCGGCTTGCACGTCGCGCACGGCGTAATGATTGAGCGGAGCAACTGGCATTGCCAGGCGCACCCA  
ACGAGGTGTGGTCAATCGATTTCTGTGATGGATGCGCTTTCCAACGGCCGGCGCGTGAAGTGCCTGACCGTCTG  
GACGATTTACGAAAGAGGCTGTGACATCGTCTGACCATGGTATCTCAGTTTATATGTCGCTCGGGAGCA  
TTGGACCGTGCAGCTCGCTTCCGTGGCTATCCAAAGGCGGTGCGAACAGACCAGGGACCCGAATTTACGAGCC  
GCGCGCTTGACAGTGGCATATGCGAACGGCGTCACTCTGAAGTTGATTCAGGCGGGCAAGCCAACGCAGAA  
TGCGTACATCGAATCGTTCAACGGCAAATTCGCGACGAATGCCTTAACGAGCACTGGTTCACGACGCTCGCG  
CACGCTCGGGCTGTATCGCGCATGGCGTCAGGACTACAACGAGCAACGGCCGACAGTGCATGAACTACC  
TTGCGCCCTCAGAGTTTGGCGGAAACATCGGGCAACCGCGGATGCTCTGCCGCTTCCAGGAGCTGGTTTAA  
AGGGACTTTGCTAGAAGCCCA

>locus\_IS407-i\_LMG18826

CTAAGTTCGATATTGGTCAGAGTGACTAGATCGAGCGAGAACGACATGAAGCATCGGCCACGACGGGGAATAT  
AGATGATATTTCCGCGTGGCTTTTAATCTGCTATTGGAGCGGCCACGTATCTGAGGACACGAGGACTCTCT  
TAAAATAAGGGATAGTCTTGTGCCTATCAGTAAGCCTGGTCATTACGTGGAAGCATCGAAATTCTGACCGAG  
CCGGAGCGCCGTGCGCGGCGCACGGCGCAAGAAAAAATCGCCATCGTGCAGGAAACATTGGAGCCGGGAGCA  
TCGGTGTGCGCCGTTGCACGTGCGCACGGCGTAATGATTGAGCGCGAGCAACTGGCATTGCCAGGGCACCCA  
ACGAGGTGTGGTCAATCGATTTTCGTGATGGATGCGCTTTCCAACGGCCGGCGCGTGAAGTGCCTGACCGTCGTC  
GACGATTTACGAAAGAGGCTGTGACATCGTGTGACCATGGTATCTCAGGTTTATATGTCGCTCGGGAGCA  
TTGGACCGTGCAGCTCGCTTCCGTGGCTATCCCAAGGCGGTGCGAACAGACCAGGGACCCGAATTTACGAGCC  
GCGCGCTTGACCAGTGGGCATATGCGAACGGCGTCACTCTGAAGTTGATTACAGGCGGGCAAGCCAACGCAGAA  
TGGTACATCGAATCGTTCAACGGCAAATTCGCGACGAATGCCTTAACGAGCACTGGTTCACGACGCTCGCG  
CACGCTCGGGCTGTCATCGCGGCATGGCGTCAGGACTACAACGAGCAACGGCCGCACAGTGCCTGAACACTACC  
TTGCGCCGTGAGAGTTTTCGCGCGAAACATCGGGCAACCGCGGATGCTCCTGCCGCTTTCCAGGAGCTGGTTTAA  
AGGGACTTTGCTAGAAGCCCAT

>locus\_IS407-i\_LMG16654

CTAAGTTCGATATTGGTCAGAGTGACTAGATCGAGCGAGAACGACATGAAGCATCGGCCACGACGGGGAATAT  
AGATGATATTTCCGCGTGGCTTTTAATCTGCTATTGGAGCGGCCACGTATCTGAGGACACGAGGACTCTCT  
TAAAATAAGGGATAGTCTTGTGCCTATCAGTAAGCCTGGTCATTACGTGGAAGCATCGAAATTCTGACCGAG  
CCGGAGCGCCGTGCGCGGCGCACGGCGCAAGAAAAAATCGCCATCGTGCAGGAAACATTGGAGCCGGGAGCA  
TCGGTGTGCGCCGTTGCACGTGCGCACGGCGTAATGATTGAGCGCGAGCAACTGGCATTGCCAGGGCACCCA  
ACGAGGTGTGGTCAATCGATTTTCGTGATGGATGCGCTTTCCAACGGCCGGCGCGTGAAGTGCCTGACCGTCGTC  
GACGATTTACGAAAGAGGCTGTGACATCGTGTGACCATGGTATCTCAGGTTTATATGTCGCTCGGGAGCA  
TTGGACCGTGCAGCTCGCTTCCGTGGCTATCCCAAGGCGGTGCGAACAGACCAGGGACCCGAATTTACGAGCC  
GCGCGCTTGACCAGTGGGCATATGCGAACGGCGTCACTCTGAAGTTGATTACAGGCGGGCAAGCCAACGCAGAA  
TGGTACATCGAATCGTTCAACGGCAAATTCGCGACGAATGCCTTAACGAGCACTGGTTCACGACGCTCGCG  
CACGCTCGGGCTGTCATCGCGGCATGGCGTCAGGACTACAACGAGCAACGGCCGCACAGTGCCTGAACACTACC  
TTGCGCCGTGAGAGTTTTCGCGCGAAACATCGGGCAACCGCGGATGCTCCTGCCGCTTTCCAGGAGCTGGTTTAA  
AGGGACTTTGCTAGAAGCCCA

>locus\_IS407-i\_LMG19240

GGTGGATGTAATGCAACTAAGTTCGATATTGGTCAGAGTGACTAGATCGAGCGAGAACGACATGAAGCATCG  
GCCACGACGGGGAATATAGATGATATTTCCGCGTGGCTTTTAATCTGCTATTGGAGCGGCCACGTATCTGA  
GGACACGAGGACTCTCTTAAAATAAGGGATAGTCTTGTGCCTATCAGTAAGCCTGGTCATTACGTGGAAGCA  
TGCAAATTCTGACCGAGCCGGAGCGCGCTCGGGCGCGCACGGCGCAAGAAAAAATCGCCATCGTGCAGGAAA  
CATTGGAGCCGGGAGCATCGGTGTGCGCCGTTGCACGTGCGCACGGCGTAATGATTGAGCGCGAGCAACTGGC  
ATTGCCAGGGCACCCAACGAGGTGTGGTCAATCGATTTTCGTGATGGATGCGCTTTCCAACGGCCGGCGCGTG  
AAGTGCCTGACCGTCGTCGACGATTTACGAAAGAGGCTGTGACATCGTGTGACCATGGTATCTCAGGTTT  
ATATGTCGCTCGCGAGCATTGGACCGTGCAGCTCGCTTCCGTGGCTATCCCAAGGCGGTGCGAACAGACCAGG  
GACCCGAATTTACGAGCCGCGCGCTTGACCAGTGGGCATATGCGAACGGCGTCACTCTGAAGTTGATTACAGG  
GGGCAAGCCAACGCAGAATGCGTACATCGAATCGTTCAACGGCAAATTCGCGACGAATGCCTTAACGAGCAC  
TGGTTACGACGCTCGCGAGCCTCGCGGCTGTCATCGCGGCTGTCAGGACTACAACGAGCAACGGCCGCACG  
ACAGTGCCTGAACACTTTCGCGCGTCAAGTTTTCGCGCGAAACATCGGGCAACCGCGGATGCTCCTGCCG  
TTTCCAGGAGCTGGTTTAAAGGACTTTGCTAGAAGCCCAT

>locus\_IS407-i\_LMG21461

TGCAACTAAGTTCGATATTGGTCAGAGTGACTAGATCGAGCGAGAACGACATGAAGCATCGGCCACGACGGGC  
AATATAGATGATATTTCCGCGTGGCTTTTAATCTGCTATTGGAGCGGCCACGTATCTGAGGACACGAGGAC  
TCTCTTAAAATAAGGGATAGTCTTGTGCCTATCAGTAAGCCTGGTCATTACGTGGAAGCATCGAAATTCTGAC  
CGAGCCGGGCGCGTGGCGCGCACGGCGCAAGAAAAAATCGCCATCGTGCAGGAAACATTGGAGCCGGG  
AGCATCGGTGTGCGCCGTTGCACGTGCGCACGGCGTAATGATTGAGCGCGAGCAACTGGCATTGCCAGCGCA  
CCCAACGAGGTGTGGTCAATCGATTTTCGTGATGGATGCGCTTTCCAACGGCCGGCGCGTGAAGTGCCTGACCG  
TCGTGACGATTTACGAAAGAGGCTGTGACATCGTGTGACCATGGTATCTCAGGTTTATATGTCGCTCGA  
GAGCATTGGACCGTGCAGCTCGCTTCCGTGGCTATCCCAAGGCGGTGCGAACAGACCAGGCACCCGAATTTAC  
GAGCCGCGCGCTTGACCAGTGGGCATATGCGAACGGCGTCACTCTGAAGTTGATTACAGGCGGGCAAGCCAACG  
CAGAATGCGTACATCGAATCGTTCAACGGCAAAGTTCCGCGACGAATGCCTTAACGAGCACTGGTTCACGACG  
CTCGCGACGCTCGGGCTGTCATCGCGGCATGGCGTCAGGACTACAACGAGCAACGGCCGCACAGTGCCTGA  
ACTACCTTGCCTGAGAGTTTTCGCGCGAAACATCGGGCAACCGCGGATGCTCCTGCCGCTTTCCAGGAGCT  
GGTTTAAAGGACTTTGCTAGAAGCCCAT

>locus\_IS407-J\_LMG12614

GCAATATGCTCGTGGTGTGCTCTGCCGAGTAGAACAAGATCACATGCAAATGCAAGCCCGTGTCTCGCC  
TTGTTCCGATAGCCCATACGTAGTTTCGCAATACCTGACATCAACGTATTAACCGACGAGCGGCGAAGAAGCGG  
TCACGATGCTGCTGGATCACTTCCGGCGTAATCCAGCGACGAAACTGAGGTTTGTAGCGCAACGTCAATGACA  
GCACCAACCAACGCGATCTACCCGTCTCTTCGAGAACATGGCCTCCTCCAACGCGAGTACATTTGTCTTATC  
T

>locus\_IS407-J\_LMG16654

GATCTGCTTGGCAATATGCTCGTGGTGTGCTCTGCCGAGTAGAACAAGATCACATGCAAATGCAAGCCC  
GTGCTCTCGCCTTGTTCGATAGCCCATACGTAGTTTCGCAATACCTGACATCAACGTATTAACCGACGAGCGGC  
GAAGAAGCGGTACGATGCTGCTGGATCACTTCCGGCGTAATCCAGCGACGAAACTGAGGTTTGTAGCGCAAC

GTC AATGACAGCACCAACCAACGCGATCTACCCGCTCTCCTTCGAGAACATGGCCTCCTCCAACGCGAGTACATT  
TGTCTTATTCTTCCTTGCGTTA

>locus\_IS407-J\_LMG18826

AGGAGGCCATGTCTCGAAGGAGACGGGTAGATCGCGTTGGTTGGTGTGTCATTGACGTTGCGCTACAAACC  
TCAGTTTCGTCGCTGGATTACGCCGGAAGTATCCAGCAGCATCGTGACCGCTTCTTCGCCGCTCGTCGGTTTA  
ATACGTTGATGTCAGGTATTGCTGACCTGCCCCATCAATAGGGCCAATGGGCTTCTAGCAAAGTCCCTTTAAA  
CCAACCTCTGGAAAGCGGCAGGAGTATCCGCGGTTGCCCGATGTTTCGCCGCAAACCTCTGACGGCGCAAGGTA  
GTTACAGTGCCTGTGCGGCCCTTTGCTCGTTGTAGTCTGACGCCATGCCGCGATGACTGCCCGAGCATGCGCGA  
GCGTCGTGAACCAAGTGTCTGTTAAGGCATTGCTCGCGGAACTTGGCGTTGAACGATTGATGTACGCATTCTGC  
GTTGGCTTGGCCGCTGAATCAACTTCAGAGCGACGCCGTTGCGATACGCCACTGGTCAAGCGCGCGGCTCGT  
AAATTCGGGGCCCTGGTCTGTTGCAACCGCCTTGGGATAGCCACGGAAGCGAGCTGCGCGGTCCAATGCCCGA  
GCGACATACAAACCTGAGATGCCATGGTGCACAACGATGTCGACAGCCTCTTTCGTGAAGTGTGTCGACGACGG  
TCAGGCACTTCACGCGCCGCGCTTGGAAAGCGCATCCATCACGAAATCGATTGACCACACCTCGTTGGGGGC  
GCCCGCAATGCCAGTTGCTCGCGCTCAATCATTACGCCGTTGGCGCTTGGACGCGCGCCGACAGCCAGCCCT  
GCCTCACGGTACAGGCGATAGATGCGCTTGTGATTGGCGTGCCTTCGCGTTCCTACTAGGGCGTGCAATCG  
GCGGTAGCCGAATCGACGACGTTGCTGTGCCAACTCCACCAGACGCGCCGCTAGCACCTCATTCTCGTGGTCTG  
GCTTCGCGTGTGAATGCAGCAGCTGCGGAGAAAGCCCGACAAGCCGGCAGGCGCGGCGCTCGGAGATGTTGAC  
CTTCCCCGAATCGCCGATACTGCTTCGCGTTTGGCTTGGCGGCTCAGGGCTTTCCTTGACGACGACCTTCAA  
CGCTTCCATATCGAGCATTGCTTCGGCCAGCAGCTTCTTCAGCCGGGCTTCTCCACCTCGAGGTTCTTGAGCC  
GGCGGGCTTCGAGACCTCCATGCCGCCGAACCTTCGCACGCCAAGTGTAGAACGACGCATCACTGAACCCGTG  
CTTCTGCACAGTTCTTGACCGGCATACCGGCCCTCGGCTTCTTCAGAAACCCGATGATTTGCTGTTCCGTAA  
AGCGCTTCTCATGTTCTGTTCTTCTCCGAAAACGAACTTTACTAGACTCCGGCTGGCCCTGTTTGCAGGGGGC  
AGGTCATTGCGAACTACGTATGGGCTATCGAACAAGGCGAGGACACGGGCTTGCATTTGCATGTGATC

>locus\_IS407-J\_LMG21461

GATCTGCTTGGCAATATGCTCGTGTGGTTGTGCTCTGCCGAGTAGAACAAGATCACATGCAAATGCAAGCCC  
GTGTCCTCGCCTTGTTCGATAGCCCATACGTAGTTTCGCAATACCTGACATCAACGTATTAACCGACGAGCGGC  
GAAGAAGCGTACGATGCTGCTGGATCACTTCCGGCGTAATCCAGCGACGAAACTGAGGTTTGTAGCGCAAC  
GTCAATGACAGCACCAACCAACGCGATCTACCCGCTCTCCTTCGAGAACATGGCCTCCTCCAACGCGAGTACATT  
TGTCTTATTCTTCCTTGCGTTA

>locus\_IS407-J\_LMG19240

GATCTGCTTGGCAATATGCTCGTGTGGTTGTGCTCTGCCGAGTAGAACAAGATCACATGCAAATGCAAGCCC  
GTGTCCTCGCCTTGTTCGATAGCCCATACGTAGTTTCGCAATACCTGACATCAACGTATTAACCGACGAGCGGC  
GAAGAAGCGTACGATGCTGCTGGATCACTTCCGGCGTAATCCAGCGACGAAACTGAGGTTTGTAGCGCAAC  
GTCAATGACAGCACCAACCAACGCGATCTACCCGCTCTCCTTCGAGAACATGGCCTCCTCCAACGCGAGTACATT  
TGTCTTATTCTTCCTTGCGTTA

>locus-IS407-k\_LMG12614

TCGGCTTGAATTATTGTGACAAGGCGGAAGGCTTGTGCGCCTGGTACCGTCACATCGTGAAGCCAACGACT  
TTACCGATGTCTTCTGTTTGGCGATTGCCGCGAAATCCATCGACCGATAACATGCAGTGTTCAGGATCTCGGC  
CTGCGCGTACACGTTTTTCAGAGGGGCTATGTTGCTCCCCACTGGATCACGATGGAGCATCACGGCGTCAATG  
G

>locus\_IS407-k\_LMG21461

GCAACGAGCGACCATTGACGCCGTGATGCTCCATCGTGATCCAGTGGGGACGAACATAGCCCTCCTCGAAAAC  
GTGTACGCGCAGGCCGAGATCCTGAACGACTGCATGTATCGGTGATGGATTTCCGGGCAATCGCCAAACAGG  
AAGACATCGGTAAAGTGTGCTTGGCTTCGACGATGTGACGGTACCAGGCCGACAAGCCTTCCGCCTTGTACAAT  
AATTCCAAGCCG

>locus\_IS407-k\_LMG18826

CTGGTACCGTCACATCGTGAAGCCAACGACTTTACCGATGTCTTCTGTTTGGCGATTGCCGCGAAATCCATC  
GACCGATACATGCAGTCTGTTTCAGGTGACCTGCCCTGCAAACAGGGCCAGCCGAGTCTAGTAAAGTTCGTT  
TTCGGAGAAGAAGACGAACATGAAGAAGCGCTTACGGAACAGCAAATCATCGGTTTCTGAAGGAAGCCGA  
GGCCGGTATGCCGTCAAGGAACTGTGCAGGAAGCACGGGTTCAAGTATGCGTCTGTTTACACTTGGCGTGGC  
AAGTTCCGGCGCATGGAGGTCTCGGAAGCCCGCGGCTCAAGAACCTCGAGGTGGAGAATGCCCGGCTGAAG  
AAGCTGCTGGCCGAAGCAATGCTCGATATGGAAGCGTTGAAGTCTGTCGTAAGGGAAAGCCCTGAGCCCGCA  
AGCCAAACGCGAAGCAGTATCGGCGATTCCGGGGAAGGTCAACATCTCCGAGCGCCGCGCTGCCGGCTTGTG  
GGGCTTCTCGCAGCGTGTGCTTACGACGCGAAGCCAGACCACGAGAATGAGGTGCTAGCGGCGCGTCTGG  
TGGAGTTGGCACACGAACGTCGATTCCGCTACCGCCGATTGCACGCCCTAGTGAACGCGAAGGCACGCA  
CGCCAAATCACAAGCGCATCTATCGCCTGACCCTGAGGCAGGGCTGGCTGTGCGGCGCCGTCGCAAGCGCCAC  
GGCGTAATGATTGAGCGCGAGCAACTGGCATTGCCGGCGCCCGCAACGAGGTGTGGTCAATCGATTTCTGTA  
TGGATGCGCTTTCCAACGGCCGGCGGCTGAAGTGCCTGACCGTCTGCGACGACTTCACGAAAGAGGCTGTGCA  
CATCGTTGTGACCATGGCATCTCAGGTTTGTATGTCGCTCGGGCATTGGACCGCGCAGCTCGCTTCCGTGGCT  
ATCCAAGGGGTGCGAACAGACCAGGGCCCCGAATTTACGAGCCGCGCGCTTGACCAGTGGGCGTATGCGAA  
CGGCGTCTGCTTGAAGTTGATTCAGGCGGGCAAGCCAACGCAGAATGCGTACATCGAATCGTTCAACGGCAAAG

TTCCGCGACGAATGCCTTAACGAGCACTGGTTCACGACGCTCGCGCATGCTCGGGCAGTCATCGCGGCATGGC  
GTCAGGACTACAACGAGCAAAGGCCGCACAGCGCACTGAACTACCTTGGCGCCGTGAGAGTTTGGCGGCAAACA  
TCGGGAACCCGCGGACTACTCTGCCGCTTCCAGGAGTTGGTTAAAGGGACTTTGCTAGAAGCCCATTGGCCC  
TATTGATGGGGCAGTACAGGATCTCGGCCTGCGCGTACACGTTTTTCGAGGAGGGCTATGTTTCGTCGCCAC

>locus\_IS407-l\_LMG12614

GTCCTTTTGGCGATCAACGAGGCGCGCTCGTCCTGCCGTTGCGGCATACGATTTTCAGTTGCGCAGTGACTGCG  
CGATTAGGCGGGGAGGGTGAATAATGAGGGTGAAGACTATTCCGACCGGAACCTGACAGACGCCGCT  
TGAATTGCGGCCCTTCAAGACTTTCCACGCGTACGCCGTACCGTCAGTGGTATCAACCGATACGGTGACGTTGCG  
ACGGCTCGGGAACGTGACGGATATGGCGTTGGTTTATGACGACCCCGCGCCAAGCTGAGTTTCGTTGGCACCG  
TCAGCACCGTTGGTCCCAGTCAGGAAGCGTTGGATTGCGCGCTTGAGCCACGGCACGGAAAAACAGTAACCAT  
CGCCCTTGTAGAGTTCCACGTGGCGATGCGCTTATAGATATCGTCCGTCGCTACGTACGTTTCCGTTACTTCGC  
GCTTCGTGTACCGTTGTAAGTCTGATGGAGTTGATACGGAGTGCCTGTTGATGGCCCCCTTGTCCGCGATCATTGCG  
GCCGGGAGTGGGGGCGTTGGATGCCGTATAGGTTGTTACAGCAGTAATCCAGACCGGCCCTTGTAGTCACGG  
TGTTCCCGACATAGATCGGAAATCAATCGAGTTGAACAGTCTAGATAGTCTTGAGCGACCCCGTTGTAGGCA  
TCCGTAAACGCGCGGATGTTGGGGTCGTTCTGATATTGGTAGAACAGATATGCATCCACGGCGCGCCATAA  
CCTGAATGTTCTGATTCGGGTACATGATTGCCTTGTATTCCGCAATGTTGTACGTCATCGCAATCCAGTAGCCG  
GACGAGTCGATCAACTCTACGTTGTCGCCATCAGGTGTTGAATAGGTGTTATTGGCATTCTTTTACCTACATTG  
TTTTACTCCGTTTTGATTACGGAATGAAATTTCCCTTGTTCACCAATTCAAGCAAGAGTCCCAATCCCCGGAATA  
TTTCGCCGCCAAAAGCCGTAAGAGTAGTGCCCGCCCGCCAGCTTGAAAACGAACCGCTCATGTGCGGATA  
CGAACCCCTTGGCGATTACGTAGGTAAGGTTGAATTGCGTGAGCAGCCCTTAAATTAAGTCTGCGGAAATTCCTG  
TCTTATCACCTGCGTTGAGCATTTCGACGCAAGCAGACAGGAACAGGCGGTCATGTGTCCGACGTTATCGGG  
GTCGTTCCGCGTAGGGTCCGGGTTGTAGCCGTCGTTGAACGGCGGCGAATTGCGCGGAACTTGGATGGAGTC  
AGTCCGCGGTTGTTCTGCTGGAAGTCCAGTAGGAACCTGGCCAGTTCGTCGAGACCGTAACGAGCTTCGGATT  
GACGGGCTTACCTTGGCGCACAGCTCGGCCACAGGTGACACGCTGACCACATACCGCGCGCTCATATCCC  
GCCCACGGATACGGTTGGTCCGAACCCCAATCGAAGGTATTGATTGCTCCTTCGCCTAGGTTGTCGTATCGCGG  
CCAAACGTACAGCGGGCAGAACGGACCCAAAACCCCAATGCGTTGCTGATACGCAATCTGCGCGTACGACAA  
AACTGAATCACGTTGTCGTAGTAGTCCGGCAGACCGGCGTTGAGCCAATTAACCGGATACTGGTAGCCCGTGT  
ACGGGGTCCGCGCCAGAATTCGTTACGGCGTAGGGTCCGGGAGTACTGATTGGAGAACGGCACACACCCGG  
CGTG

>locus\_IS407-m\_LMG21461

GTGTCGTGATGGACTGCATCGTGGCTCAAGAATCGATCAAGAAATCAGGGGATTACCGCGGGAAGTCGAGGTC  
TGCTACGAGAGTAGACGACATTATTTTGGTGAGATTTTCAATAAAGCCGTCAACTCGTGTCGGAGTGCCCTGG  
TCTACGGGGATTCTGGAGTGACGGAATTTTGTATGCTACGCCGGGACGTGGAGTACGGCGTCAACCCGAT  
ATCCACCCCGGAGATTCGAATAGCCCCGCCATCCGCCCCCGCGTTGCGCGAGTTCGGACATCTGTTCCGGCCGT  
GGGACAGATAGACCGCAATCTCTGCCCCAGCGCCCCAGCCCTGAATCTCCAGTTCGAGCCGAATCGGCAGC  
CGCTGATTTGTCGGCGCTTCTGCGAGCAATTCAGCTGACGTACCCGCGTCACGCAGCGCGGCGTAGTTT  
CCCGTTTACCG

>locus\_IS407-n\_LMG12614

CGCTCGTATGAATCGGACCCGTAGCGCCGGCCACAGTAGATGTCGCGCAGACCCGCTTCCGGAGATACGTGCA  
GCTTTGACGAAAGTGCCACGCGGCCCCAGGCTGTTGCTTTTCGTGCGCCACACGATGGAATGACGTAGAAGTC  
TGCGATTGACGGATGTTGGCTCAGGTGTTGATTTAGCTTGACGTTTCATGAAAGCACCTCCCGCCACCGCAAGCG  
AGTCAGTTCGAGTGACCTGCCCCATCAATAGGGCCAATGGGCTTCTAGCAAAGTCCCTTTAAACCAACTCCT  
GGAAAGCGGCAGGAGTATCCGCGGTTGCCGATGTTTCGCCGCAAACCTCTGACGGCGCAAGGTAGTTTCAGTGC  
GCTGTGCGGCCTTTGCTCGTTGTAGTCTGACGCCATGCCGCGATGACTGCCCGAGCATGCGCGAGCGTCTGTA  
ACCAGTGCTCGTTAAGGCATTGCTCGCGGAACCTGCCGTTGAACGATTGATGACGATTCTGCGTTGGCTTG  
CCCGCTGAATCAACTTCAGAGCGACCGGTTGCGATACGCCACTGGTCAAGCGCGCGGCTCGTAAATTCGG  
GGCCCTGGTCTGTTGCGACCGCCTTGGGATAGCCACGGAAGCGAGCTGCGCGTCCAATGCCCGAGCGACATA  
CAAACCTGAGATGCCATGGTTCGACAACGATGTGACAGCCCTTTTCGTGAAGTCTGTCGACGACGGTCAGGCAC  
TTCACGCGCCGGCGGTTGGAAAGCGCATCCATCACGAAATCGATTGACCACACCTCGTTGGGGGCGCCCGCA  
ATGCCAGTTGCTCGCGCTCAATCATTACGCCGTGGCGCTTGGCAGCGCGCCGCACAGCCAGCCCTGCCTCACGG  
TACAGGGGATAGATGGCCTTGTGATTGGCGTGCCTTCCGCTTCCACTAGGGCGTGAATCGGGCGTAGCC  
GAATCGACGACGTTGCTGTGCCAACTCCACCAGACGCGCCGCTAGCACCTCATTCTGTTGCTGCTTCCGCT  
CGTAATGCAGCACGCTGCGAGAAAGCCCGACAAGCCGGCAGGCGCGGCTCGGAGATGTTGACCTTCCCCCG  
AATCGCCGATACTGCTTCGCGTTTGGCTTGGGGCTCAGGGCTTCCCTTGACGACGACCTTCAACGCTTCCAT  
ATCGAGCATTGCTTCGGCCAGCAGCTTCTTACCGGGCATTCTCCACCTCGAGGTTCTTGAGCCGGCGGGCTT  
CCGAGACCTCCATGCCGCCAACTTCGCACGCCAAGTGTAGAACGACGCATCACTGAACCCGTGCTTCTGCA  
CAGTTCTTGACCGGCATACCGGCTCGGCTTCTTCAAGAACCCGATGATTTGCTGTTCCGTAAGCGCTTCTT  
CATGTTCTTCTTCTCCGAAAACGAACTTACTAGACTCCGGCTGGCCCTGTTTGCAGGGGGCAGGTCAGTC  
ACAACGGTGCTCGGCCACGCTCCGAGCGCGATTGGGCATCTGAGTGAGC

>locus\_IS407-n\_LMG18826

AGCGCACGCTCGTATGAATCGGACCCGTAGCGCCGGCCACAGTAGATGTCGCGCAGACCCGCTTCCGGAGATA  
CGTCGAGCTTTGACGAAAGTGCCACGCGGCCCCAGGCTGTTGCTTTTCGTGCGCCACACGATGGAATGACGTA

GAAGTCTGCGATTGACGGATGTTGGCTCAGGTGTTGATTTAGCTTGACGTTTCATGAAAGCACCTCCCGCCACCG  
CAAGCGAGTCAGTTCCGAGTGACCTGCCCCATCAATAGGGCCAATGGGCTTCTAGCAAAGTCCCTTTAAACC  
AACTCCTGAAAGCGGCAGGAGTATCCGCGGTTGCCGATGTTTCGCGCAAACCTCTGACGGCGCAAGGTAGT  
TCAGTGCGCTGTGCGGCCCTTTGCTCGTTGTAGTCCTGACGCCATGCCGCGATGACTGCCCGAGCATGCGCGAGC  
GTCGTGAACCAGTGCTCGTTAAGGCATTTCGTGCGGAACTTGCCGTTGAACGATTTCGATGTACGCATTCTGCGT  
TGGCTTGCCCCGCTGAATCAACTTCAGAGCGACGCCGTTTCGCATACGCCCACTGGTCAAGCGCGCGGCTCGTA  
AATTCGGGGCCCTGGTCTGTTTCGCACCGCCTTGGGATAGCCACGGAAGCGAGCTGCGCGGTCCAATGCCCGAG  
CGACATAAAACATGAGATGCCATGGTCGACAACGATGTGACAGCCTCTTTCGTGAAGTCGTGACGACGGT  
CAGGCACTTCACGCGCCGGCCGTTGGAAAGCGCATCCATCACGAAATCGATTGACCACACCTCGTTGGGGGCG  
CCCGGCAATGCCAGTTGCTCGCGCTCAATCATTACGCCGTGGCGCTTTCGACGCGCCGACAGCCAGCCCTG  
CCTCACGGTACAGGCGATAGATGCGCTTGTGATTGGCGTGCCTTTCGCGTTCCACTAGGGCGTGCAATCGG  
CGGTAGCCGAATCGACGACGTTTCGTGTGCCAACTCCACCAGACGCGCCGCTAGCACCTCATTCTCGTGGTCTGG  
CTTCGCGTCGTAATGCAGCACGCTGCGAGAAAGCCCGACAAGCCGGCAGGCGCGGCTCGGAGATGTTGACC  
TTCCCCGAATCGCCGATACTGCTTTCGCGTTTGGCTTTCGCGGCTCAGGGCTTTCCTTGACGACGACCTTCAAC  
GCTTCCATATCGAGCATTGCTTCGGCCAGCAGCTTCTTCAGCCGGCATTCTCCACCTCGAGGTTCTTGAGCCG  
GCGGGCTTCCGAGACCTCCATGCCGCCAACTTCGCACGCCAAGTGTAGAACGACGCATCACTGAACCCGTGC  
TTCCTGCACAGTTCTTGACCGGCATACCGCCTCGGCTTCTTCAGAAACCCGATGATTTGCTGTTCCGTA  
GCGCTTCTTCATGTTTCGTCTTCTTCCGAAAACGAACTTTACTAGACTCCGGCTGGCCCTGTTTGCAGGGGGCA  
GGTCAGTCAACCGGTGCTCGGCCAGCCTCCGAGCGCGATTGGGCATCTGAGGAGCGGT

>locus\_IS407-n\_LMG19240

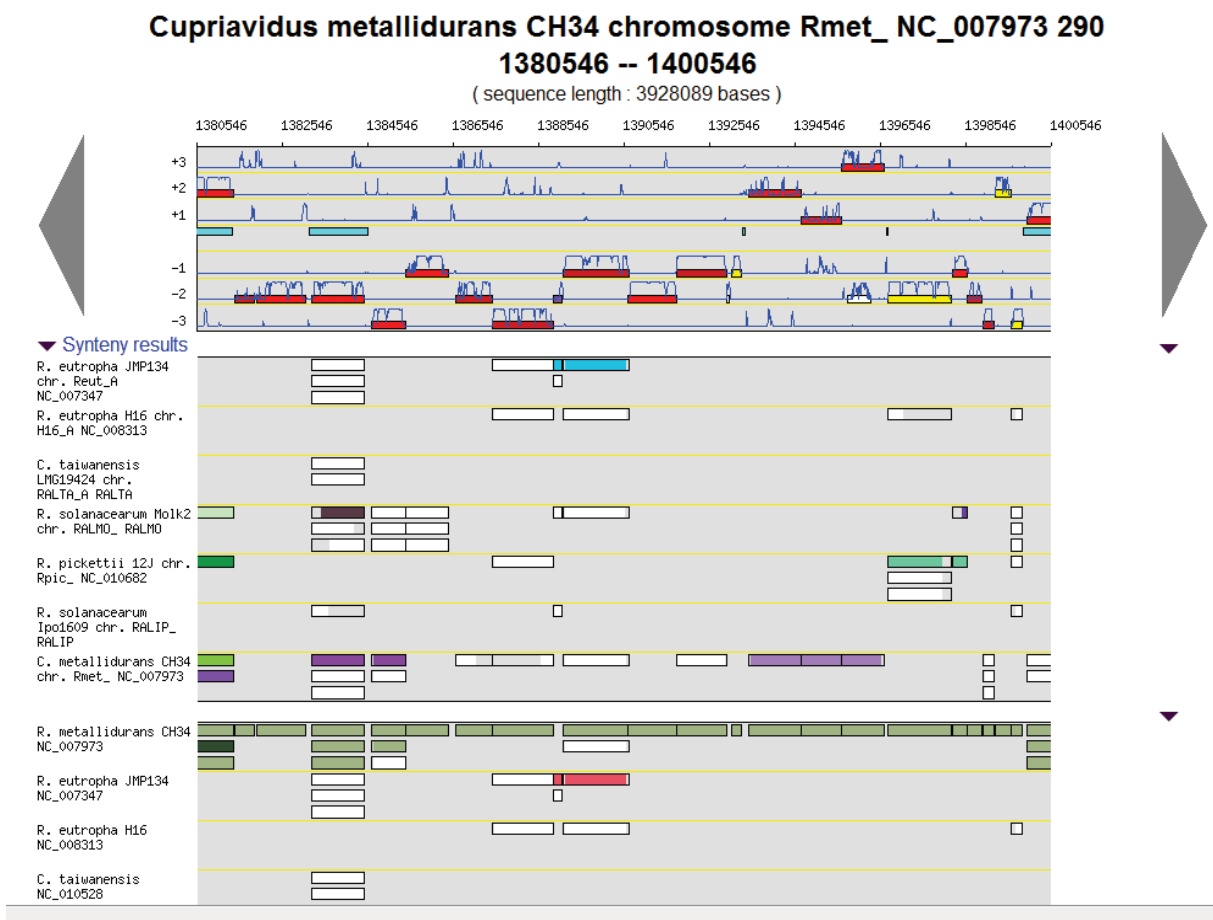
TTCAGAGCGACGCCGTTTCGCATACGCCCACTGGTCAAGCGCGCGGCTCGTAAATTCGGGGCCCTGGTCTGTTCCG  
CACCGCCTTGGGATAGCCACGGAAGCGAGCTGCGCGTCCAATGCCCGAGCGACATAAAACCTGAGATGCCA  
TGGTGACAACGATGTGACAGCCTCTTTCGTGAAGTCGTGACGACGGTCAGGCACTTCACGCGCCGGCCGTT  
GGAAAGCGCATCCATCACGAAATCGATTGACCACACCTCGTTGGGGGCGCCCGGCAATGCCAGTTGCTCGCGC  
TCAATCATTACGCCGTGGCGCTTTCGACGCGCCGACAGCCAGCCCTGCCTCACGGTACAGGCGATAGATGC  
GCTTGTGATTGGCGTGCCTTCGCGTTCCTACTAGGGCGTG

>locus\_IS407-n\_LMG21461

CGCTCGTATGAATCGGACCCGTAGCGCCGGCCACAGTAGATGTGCGCGACACCGCCTTCCGGAGATACGTGCA  
GCTTTGACGAAAGTGCCACGCGGCCCCAGGCTGTTGCTTTCGTGCGCCACACGATGGAATGACGTAGAAGTC  
TGCGATTGACGGATGTTGGCTCAGGTGTTGATTTAGCTTGACGTTTCATGAAAGCACCTCCCGCCACCGCAAGCG  
AGTCAGTTCGGAGTGACCTGCCCCATCAATAGGGCCAATGGGCTTCTAGCAAAGTCCCTTTAAACCAACTCCT  
GGAAAGCGGCAGGAGTATCCGCGGTTGCCGATGTTTCGCGCAAACCTCTGACGGCGCAAGGTAGTTTCAGTGC  
GCTGTGCGGCCTTTGCTCGTTGTAGTCCTGACGCCATGCCGCGATGACTGCCCCGAGCATGCGCGAGCGTCTGTA  
ACCAAGTCTCGTTAAGGCATTTCGTGCGGAACTTGCCGTTGAACGATTTCGATGTACGCATTCTGCGTTGGCTTG  
CCCGCCTGAATCAACTTCAGAGCGACGCCGTTTCGCATACGCCCACTGGTCAAGCGCGCGGCTCGTAAATTCGG  
GGCCCTGGTCTGTTTCGCACCGCCTTGGGATAGCCACGGAAGCGAGCTGCGCGTCCAATGCCCGAGCGACATA  
CAAACCTGAGATGCCATGGTCGACAACGATGTGACAGCCTCTTTCGTGAAGTCGTGACGACGCGTCAGGCAC  
TTCACGCGCCGGCCGTTGGAAAGCGCATCCATCACGAAATCGATTGACCACACCTCGTTGGGGGCGCCCGCA  
ATGCCAGTTGCTCGCGCTCAATCATTACGCCGTGGCGCTTTCGACGCGCCGCGCACAGCCAGCCCTGCCTCACGG  
TACAGGCGATAGATGCGCTTGTGATTGGCGTGCCTTTCGCGTTCCACTAGGGCGTGCAATCGGCGGTAGCC  
GAATCGACGACGTTTCGTGTGCCAACTCCACCAGACGCGCCGCTAGCACCTCATTCTCGTGGTCTGGCTTCGCGT  
CGTAATGCAGCAGCTGCGAGAAAGCCCGACAAGCCGGCAGGCGCGGCTCGGAGATGTTGACCTTCCCCCG  
AATCGCCGACTGCTTCGCGTTTGGCTTTCGCGGCTCAGGGCTTTCCTTGACGACGACCTTCAACGCTTCCAT  
ATCGAGCATTGCTTCGGCCAGCAGCTTCTTCAGCCGGCATTCTCCACCTCGAGGTTCTTGAGCCGCGGGCTT  
CCGAGACCTCCATGCCCGCAACTTCGCACGCCAAGTGTAGAACGACGCATCACTGAACCCGTGCTTCTGCA  
CAGTTCTTGACCGGCATACCGCCTCGGCTTCTTCAGAAACCCGATGATTTGCTGTTCCGTAAGCGCTTCTT  
CATGTTTCGTCTTCTTCCGAAAACGAACTTTACTAGACTCCGGCTGGCCCTGTTTGCAGGGGGCAGGTCAGTC  
ACAACGGTCTCGGCCAGCCTCCGAGCGCGATTGGGCATCTGAGTGAGC

## D. Chapitre 4

### 1. Synténie de *R. metallidurans* CH34



### 2. Synténie de *P. aeruginosa* PA14

