



HAL
open science

Approches cumulées de phylogénie et d'écologie pour déterminer les bases génétiques de la spécificité d'hôte des bactéries phytopathogènes, cas des *Xanthomonas* spp.

Nadia Mhedbi

► **To cite this version:**

Nadia Mhedbi. Approches cumulées de phylogénie et d'écologie pour déterminer les bases génétiques de la spécificité d'hôte des bactéries phytopathogènes, cas des *Xanthomonas* spp.. Biologie végétale. Université d'Angers, 2010. Français. NNT: . tel-00648729

HAL Id: tel-00648729

<https://theses.hal.science/tel-00648729>

Submitted on 6 Dec 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.

**Approches cumul es de phylog nie et d' cologie
pour d terminer les bases g n tiques de la
sp cificit  d'h te des bact ries phytopathog nes,
cas des *Xanthomonas* spp.**

TH SE DE DOCTORAT

Sp cialit  : Biologie des organismes

ECOLE DOCTORALE VENAM

Pr sent e et soutenue publiquement

Le : 21 d cembre 2010

  : Angers

par : Nadia MHEDBI-HAJRI

Devant le jury ci-dessous :

Xavier NESME (Rapporteur), Ing nieur de Recherche, INRA, Lyon

Matthieu ARLAT (Rapporteur), Professeur, Universit  Paul Sabatier, Toulouse

Monique ROYER (Examinateur), Chercheur, CIRAD, Montpellier

Philippe SIMONEAU (Examinateur), Professeur, Universit  d'Angers, Angers

Marie-Agn s JACQUES (Encadrant), Charg e de Recherche, INRA, Angers

Directeur de th se : Charles MANCEAU, (Ing nieur de Recherche), INRA, Angers

Remerciements

Un moment émouvant pour moi est celui d'écrire les remerciements.

Je tiens à remercier en premier lieu mon directeur de thèse, Charles Manceau, de m'avoir accueillie au sein de l'équipe EDTa-BP. Un grand merci pour ta disponibilité, ton suivi et tes suggestions au cours de ces trois années de thèse.

Je tiens également à remercier Xavier Nesme et Matthieu Arlat d'avoir accepté d'être les rapporteurs de mes travaux de thèse. Je remercie également Monique Royer et Philippe Simoneau qui ont bien voulu examiner ce travail.

Il me tient à cœur de remercier Marie-Agnès Jacques. Je tiens à t'exprimer toute ma reconnaissance pour la confiance et l'amitié que tu m'as témoignées. Je te remercie également pour ton encadrement, ta rigueur, ton soutien sans faille et tes encouragements durant ces trois années. Je tiens à te dire que travailler avec toi est un réel plaisir. Merci pour ta bonne humeur et tes éclats de rire...ô que oui très expressifs !

Un grand merci à Armelle Darrasse. Je te remercie pour tes encouragements et ton aide aussi. Tu as toujours été là quand j'avais besoin de toi. Je tiens à dire que c'était avec un grand plaisir que d'avoir partagé le bureau avec toi. Et n'oublie pas de prendre soin des plantes !

Un grand merci à Tristan Boureau et Stéphane Poussier. Je vous remercie tous les deux pour les discussions que nous avons pu avoir.

J'exprime également mes remerciements à Christophe Lemaire avec qui j'ai collaboré.

Je remercie sincèrement Ralf Koebnik et Olivier Pruvost d'avoir suivi mon travail de thèse et pour les suggestions qu'ils ont apportées.

Je voudrais également remercier Perrine David pour sa relecture et les conseils qu'elle m'a apportés concernant les analyses et la rédaction. Je te remercie pour ton soutien et tes encouragements.

Je remercie Pascal Poupard pour la relecture du manuscrit et ses corrections.

Je tiens à remercier Karine Durand et Chrystelle Brin pour leurs conseils et leur aide au laboratoire. Merci Sophie Bonneau pour ton aide précieuse surtout en période de rédaction de thèse. Un grand merci à Martial Briand et Alain Huart pour leur disponibilité et surtout leur aide en informatique.

Je remercie également Sandrine Pigné d'avoir contribué à ce travail lors de son stage de M2.

Mes remerciements s'adressent à tous mes collègues de l'UMR PaVé en particulier Marion, Perrine Portier, Sophie Gironde, Sophie Cesbron, Anne, Joël, Arnaud, Christine, Marie-Noël, Martine, Matthieu, Roland, Chrystelle, Thomas, Jean-Pierre, Michel, Bruno, Natalia, Thibaut, Baba, Salwa, Valérie, Amandine, Jacky, Frédérique, Pascale, Elisabeth Dineau et Rachelle. Un merci particulier à Elisabeth Deloménie pour sa disponibilité.

J'adresse un grand merci ma maman et mon papa pour leur soutien indéfectible. Merci pour les fleurs que vous m'avez envoyées quotidiennement. MERCI pour tout ! Un grand Merci à mes deux frères pour votre amour et vos encouragements. Vous êtes ma force !

Je voudrais remercier ma grand-mère pour les encouragements, les conseils et l'amour qu'elle m'apporte.

Je remercie sincèrement mes beaux parents, ma belle sœur et son mari pour leur soutien sans faille et leur amour.

Enfin, un grand MERCI à Ahmed. Merci pour ton soutien inconditionnel, la force et le bonheur que tu m'apportes chaque jour. Merci d'avoir si bien pris soin de moi durant cette période et pour ta patience. Je te remercie également pour les différentes relectures de ce manuscrit.

Table des matières

Introduction et objectifs	10
I Introduction	14
1 La phyllosphère	15
1.1 Topographie de la phyllosphère et propriétés physicochimiques de la cuticule	15
1.2 Caractéristiques environnementales de la phyllosphère	18
1.2.1 Paramètres physiques de l'environnement	18
1.2.2 Disponibilité en nutriments et autres métabolites d'origine végétale	19
1.2.3 Composés organiques volatiles émis par la feuille	22
2 Perception et adaptation à l'hôte	24
2.1 Chimiotactisme : Perception, reconnaissance spécifique et adaptation à l'hôte	24
2.1.1 Methyl-accepting Chemotaxis Proteins	25
2.1.2 Transduction du signal chimiotactique et mobilité	27
2.1.3 Temps d'adaptation de la réponse chimiotactique	28
2.1.4 Amplification du signal externe	29
2.2 Autres systèmes de perception et de transduction de signaux	30
2.2.1 les systèmes de régulation à deux composants	30
2.2.2 Quelques cas particulier de systèmes de régulation à deux composants étudiés chez les <i>Xanthomonas</i> spp.	33
2.2.3 Transporteurs TonB-dépendants	37
2.3 Adhésion bactérienne	39
Chapitre d'ouvrage: Adhesion mechanisms of plant-pathogenic <i>Xanthomonadaceae</i>	41
3 Évolution adaptative des génomes bactériens et répertoires de gènes	62
3.1 Évolution adaptative des génomes bactériens	63
3.1.1 Gain de gènes par transfert horizontal et duplication	63
3.1.2 Perte de gènes	66
3.1.3 Changements des séquences de gènes	67
3.2 Notion de répertoire de gènes	70
4 Bactéries modèles de l'étude : les <i>Xanthomonas</i> spp.	73
4.1 Vie épiphyte des <i>Xanthomonas</i> spp.	74
4.2 Les hôtes des <i>Xanthomonas</i> spp.	75
4.3 Symptômes et maladies causés par les bactéries appartenant au genre <i>Xanthomonas</i>	76
4.4 Caractéristiques cellulaires et physiologiques des <i>Xanthomonas</i> spp.	77
4.5 Caractéristiques génomiques des <i>Xanthomonas</i> spp.	78

4.6 Classification et évolution de la taxonomie des <i>Xanthomonas</i> spp.	80
4.6.1 Une taxonomie évolutive du genre <i>Xanthomonas</i> spp.	80
4.6.2 Une taxonomie basée sur des techniques moléculaires	81
4.6.3 Diversité infraspécifique de <i>X. axonopodis</i> et <i>X. campestris</i>	83
II Résultats	88
1 Rôle des senseurs et des adhésines dans la spécificité d'hôte des bactéries phytopathogènes, le cas des <i>Xanthomonas</i> spp.	89
1.1 Présentation de l'article 1 et des résultats complémentaires	89
1.2 Article 1: Sensing and adhesion are adaptive functions in the plant pathogenic xanthomonads	91
1.3 Résultats complémentaires	135
1.3.1 Matériels et Méthodes	135
1.3.2 Résultats	138
1.4 Discussion	143
2 Etude des relations phylogénétiques et généalogiques entre les souches de <i>X. axonopodis</i>	148
2.1 Présentation de l'étude phylogénétique et de l'article 2	148
2.2 Relations phylogénétiques entre les souches de <i>X. axonopodis</i>	150
2.2.1 Matériels et Méthodes	150
2.2.2 Résultats	153
2.3 Article 2: Evidence of recombination and migration among pathovars of <i>Xanthomonas axonopodis</i> species	158
2.4 Discussion	193
Conclusions générales et perspectives	198
Annexes	208
Références bibliographiques	232

Liste des figures

Partie I.

- Figure 1.** Composition de la cire cuticulaire des deux premières feuilles de plantes de *Phaseolus vulgaris* en fonction de l'exposition aux rayons UV-B.
- Figure 2.** Représentation schématique de la structure de la cuticule.
- Figure 3.** Surfaces foliaires supérieures de *Crassula ovata* (A), d'*Eucalyptus gunnii* (B), d'*Euphorbia characias* (C) en vue microscopique à balayage.
- Figure 4.** Fluctuations de la température, de l'humidité relative et de l'humidité de surface à la surface foliaire du poirier au cours de la journée.
- Figure 5.** Variations des quantités de sucres et d'acide malique collectés au niveau de la surface inférieure des feuilles de *Nicotiana plumbaginifolia*, en fonction de l'âge et de la position des feuilles.
- Figure 6.** Photographie des trichomes des feuilles de tomate colonisés par *X. axonopodis* pv. *vesicatoria*, réalisée par microscopie électronique à balayage.
- Figure 7.** Observations directes de *Erwinia herbicola* (Eh299R) marquée à la GFP dans la phyllosphère par microscopie à épifluorescence
- Figure 8.** Représentation schématique des différentes voies empruntées par les protéines pour se déposer sur la phyllosphère.
- Figure 9.** Trichomes à la surface des feuilles de *Nicotiana tabacum*.
- Figure 10.** Expression de la protéine GUS dans les trichomes glandulaires courts (A) et longs (B) de la feuille de *N. tabacum*.
- Figure 11.** Voies de diffusion des composées volatiles à partir des tissus foliaires.
- Figure 12.** Représentation schématique de la voie de transduction du signal chimiotactique chez les bactéries.
- Figure 13.** Organisation des gènes codant les MCPs dans les génomes de *Xanthomonas* spp.
- Figure 14.** Ilots polaires et latéraux de chémorécepteurs chez *E. coli*.
- Figure 15.** Modèle de la structuration d'un ilot de chémorécepteurs.
- Figure 16.** Structure d'un chémorécepteur dimérique.
- Figure 17.** Activité fonctionnelle des nanodisques comportant un seul homodimère ou des trimères de chémorécepteurs.
- Figure 18.** Voie de signalisation des chémorécepteurs chez *E. coli*.
- Figure 19.** Chimiotactisme chez *E. coli*.
- Figure 20.** Temps d'adaptation de la réponse chimiotactique.
- Figure 21.** Représentation schématique de l'organisation modulaire des histidines kinases.
- Figure 22.** Représentation schématique de l'organisation des régulateurs.
- Figure 23.** Représentation schématique de la perception de DSF et de la transduction du signal chez *X. campestris*.
- Figure 24.** Représentation schématique des voies de régulation contrôlant l'expression des gènes impliqués dans le pouvoir pathogène de *X. campestris* pv. *campestris*.
- Figure 25.** Régulation des gènes *hrp* chez les bactéries phytopathogènes.
- Figure 26.** Représentation du modèle de TBDT spécifique de la vitamine B₁₂.
- Figure 27.** Domaines fonctionnels des TBDTs.
- Figure 28.** Présence des gènes TBDT chez 226 souches séquencées à Gram-négatif en fonction de leur caractéristiques écologiques (A) et phylogénétiques (B).
- Figure 29.** Pili de type IV observés chez une souche sauvage de *Xylella fastidiosa* en microscopie électronique à balayage.
- Figure 30.** Représentation schématique de la sécrétion des adhésines autotransporteurs.
- Figure 31.** Adhésines de type autotransporteur trimérique (BadA et YadA) observées en microscopie électronique à transmission et modèle de la structure de l'adhésine YadA.
- Figure 32.** Représentation schématique de la sécrétion bipartite des adhésines non fibrillaires de type hémagglutinine.
- Figure 33.** Alignement de neuf génomes de souches séquencées de *Xanthomonas*.
- Figure 34.** Organisation et polymorphisme d'une région génomique présente dans les génomes des *Xanthomonas* spp. comprenant des gènes codant des senseurs couplés à leurs régulateurs (histidines kinases hybrides, HyHKs).

- Figure 35.** Dendrogramme obtenu à partir de la matrice présence/absence de 35 ET3s chez 132 souches appartenant à *X. axonopodis*.
- Figure 36.** Les styles de vie des *Xanthomonas* spp.
- Figure 37.** Symptômes de la graisse commune du haricot sur feuilles (A) et sur gousses (B), des taches foliaires du manioc (C), de la galle bactérienne du poivron (D), des taches noires du manguier sur fruit et sur feuilles* (E), du chancre européen du peuplier (F), du chancre asiatique des agrumes sur tiges et fruit (G) et de la nervation noire du chou (H)
- Figure 38.** Bactéries du genre *Xanthomonas* pourvues d'un flagelle polaire.
- Figure 39.** Diversité des colonies bactériennes du genre *Xanthomonas* sur un milieu gélosé riche.
- Figure 40.** Représentation schématique des réarrangements taxonomiques au sein du genre *Xanthomonas* sur la base des hybridations ADN-ADN.
- Figure 41.** Arbre phylogénétique construit avec la méthode Neighbour-Joining sur la base des séquences partielles de l'ADNr 16S et représentant les relations entre des souches de *Xanthomonas* spp.
- Figure 42.** Analyse UPGMA combinées des profils BOX, ERIC et REP des souches de *Xanthomonas* spp.
- Figure 43.** Arbre phylogénétique construit avec la méthode Neighbour-Joining sur la base des séquences partielles concaténées des gènes *dnaK*, *fyuA*, *gyrB* et *rpoD*.
- Figure 44.** Remaniement taxonomique au sein de l'espèce *X. axonopodis*.
- Figure 45.** Symptômes de la graisse commune du haricot observés sur les différents organes.
- Figure 46.** Arbre phylogénétique construit avec la méthode Neighbour-Joining sur la base des empreintes AFLP et représentant les relations entre les souches appartenant aux 4 lignées génétiques de *X. axonopodis* pv. *phaseoli* (*fuscans*, GL1, GL2 et GL3) et des souches appartenant à d'autres pathovars de *X. axonopodis*.

Partie II.

Chapitre 1.

- Figure 1.** Matrice de présence/absence de 24 gènes codant les MCPs dans les génomes de 10 souches séquencées de *Xanthomonas* spp.
- Figure 2.** Dendrogramme obtenu par la méthode Neighbor Joining sur la base de l'alignement de 165 séquences codant les MCPs provenant de 10 souches séquencées de *Xanthomonas* spp.

Chapitre 2.

- Figure 1.** Arbres phylogénétiques construits par Neighbor Joining à partir des données des séquences partielles de sept gènes de ménage (*atpD*, *dnaK*, *efp*, *glnA*, *gyrB*, *fyuA* et *rpoD*) et indication des changements de positionnement de certaines souches illustrant des événements de recombinaison au sein de l'espèce *X. axonopodis*.
- Figure 2.** Positionnement de cinq souches mal identifiées dans les arbres phylogénétiques des *Xanthomonas* spp.
- Figure 3.** Réseaux obtenus avec la méthode NeighborNet sur la base des séquences types (ST) des sept gènes de ménage (*atpD*, *dnaK*, *efp*, *fyuA*, *glnA*, *gyrB* et *rpoD*).
- Figure 4.** Représentation de la diversité de la population chez *X. axonopodis* par une analyse MLST.

Liste des tableaux

Partie I.

- Tableau 1.** Rendement en huiles essentielles et concentration des composés phénoliques extraits des surfaces foliaires de 19 espèces végétales.
- Tableau 2.** Activité antibactérienne des métabolites secondaires provenant de différentes espèces végétales contre les bactéries *E. herbicola* et *P. syringae*.
- Tableau 3.** Liste non exhaustive des maladies causées par les bactéries membres du genre *Xanthomonas* spp.
- Tableau 4.** Caractéristiques générales des génomes séquencés de *Xanthomonas* spp.
- Tableau 5.** Récapitulatif des différences pathologiques, phénotypiques et génétiques documentées entre les lignées *fuscans* et non-*fuscans* (GL1, GL2 et GL3) de *X. axonopodis* pv. *phaseoli*.

Partie II.

Chapitre 1.

- Tableau 1.** Séquences nucléotidiques et position des amorces utilisées pour l'amplification des orthologues des gènes XCV1933, XCV3230 et XCV1940 dans les 131 souches de *X. axonopodis*, taille des amplifiats et températures d'hybridation des amorces.
- Tableau 2.** Les paramètres des différents modèles *site* du programme CODEML implémenté dans le logiciel PAML (Yang, 1997).
- Tableau 3.** Identification des sites soumis à une pression de sélection positive dans l'orthologue de XCV1933 dans la souche B100 de *X. campestris* pv. *campestris*.
- Tableau 4.** Résultats des tests de vraisemblance des analyses de recombinaison effectuées avec le programme ClonalFrame (Didelot et Falush, 2007) et appliquées aux haplotypes des orthologues de XCV1933 et XCV3230 et de la région N-terminale du gène XCV1940.
- Tableau 5.** Résultats des tests de vraisemblance des analyses de pressions de sélection effectuées avec les six modèles *sites* (NSSites 0 à 8) et le modèle *branch* (M1) du programme CODEML et appliquées aux haplotypes des orthologues codant les MCPs et identification des sites probablement soumis à sélection positive.
- Tableau 6.** Valeurs des estimateurs inférés par les modèles *sites* non rejetés du programme CODEML pour les 35 haplotypes d'orthologues de XCV3230.
- Tableau 7.** Tableau de contingence comprenant les résultats du test McDonald et Kreitman (McDonald et Kreitman, 1991) réalisées avec les orthologues de XCV1933 et XCV3230 et de la région N-terminale du gène XCV1940 et à l'aide du logiciel DnaSP (Rozas *et al.*, 2003).

Chapitre 2.

- Tableau 1.** Gènes de ménage utilisés dans différentes études des relations phylogénétiques entre des souches appartenant à *Xanthomonas* spp.
- Tableau 2.** Taille des fragments amplifiés des sept gènes de ménage et positionnement dans la séquence concaténée.
- Tableau 3.** Probabilités associées aux résultats du test Shimodaira Hasegawa (SH) pour l'analyse de la congruence des topologies des arbres phylogénétiques construits à partir des alignements des sept gènes de ménage (*atpD*, *dnaK*, *efp*, *fyuA*, *glnA*, *gyrB* et *rpoD*) et des données concaténées. L'hypothèse testée stipule que les valeurs de vraisemblance associées aux topologies des arbres sont dans le même intervalle de confiance.

Partie III.

- Tableau 1.** Quotient intellectuel de 24 bactéries pathogènes de plantes, de l'Homme et des animaux.

Liste des Annexes

Partie II.

Chapitre 1.

- Annexe 1.** Résultats des tests de vraisemblance des analyses de pressions de sélection effectuées avec le modèle *branch-site* du programme CODEML et appliquées aux haplotypes des orthologues codant les MCPs.
- Annexe 2.** Liste des haplotypes des orthologues de XCV1933, XCV3230 et de la région N-terminale de XCV1940 amplifiés par PCR dans des souches de *X. axonopodis*.
- Annexe 3.** Résultats des tests de vraisemblance des analyses de pressions de sélection effectuées avec le modèle *branch-site* du programme CODEML et appliquées aux haplotypes des orthologues de XCV1933 et identification des sites probablement soumis à sélection positive.
- Annexe 4.** Résultats des tests de vraisemblance des analyses de pressions de sélection effectuées avec le modèle *branch-site* du programme CODEML et appliquées aux haplotypes des orthologues de la région N-terminale du gène XCV1940 et identification des sites probablement soumis à sélection positive.

Chapitre 2.

- Annexe 5.** Liste des souches complémentaires aux 131 souches utilisées dans l'article 2.
- Annexe 6.** Arbres phylogénétiques obtenus par Neighbor Joining avec 1000 simulations à partir des séquences partielles des gènes *atpD* (6.1), *dnaK* (6.2), *efp* (6.3), *fyuA* (6.4), *glnA* (6.5), *gyrB* (6.6) et *rpoB* (6.7) dans 138 souches de *Xanthomonas* spp. (*sensu* Vauterin *et al.*, 1995).
- Annexe 7.** Diversité des pathovars en fonction des gènes de ménage étudiés.
- Annexe 8.** Types de séquences par gène et profils alléliques définis pour les 131 souches de *X. axonopodis* de la collection.
- Annexe 9.** Liste des 41 séquences recombinantes détectées par six ou sept des sept méthodes de détection de la recombinaison implémentées dans le logiciel RDP et identification des souches de type parental et des sites de recombinaison.

Introduction et objectifs

Une meilleure compréhension des étapes clés du cycle infectieux d'une bactérie sur son hôte et des facteurs régissant la spécificité d'hôte est primordiale pour la recherche de nouvelles méthodes de lutte respectueuses de l'environnement et pour l'élaboration de textes réglementaires adaptés. Ceci est d'autant plus important qu'en l'absence de lutte chimique efficace et autorisée, la lutte contre les phytobactérioses est restreinte aux méthodes prophylactiques ou à des opérations effectuées *a posteriori* pouvant conduire à la destruction des lots contaminés (sélection sanitaire des semences).

Actuellement, il est communément admis que la spécificité parasitaire, caractérisant les pathovars de nombreuses espèces bactériennes, s'exprime par l'interaction compatible ou incompatible aboutissant respectivement à l'expression ou non de la maladie. Il est à noter que le terme *spécificité d'hôte* est couramment utilisé comme synonyme de spécificité parasitaire. Selon Poulin (2007), la notion de *spécificité d'hôte* est relative à l'ensemble des hôtes qu'un agent pathogène colonise au cours de son cycle biologique. Cette gamme d'hôtes renferme les hôtes cultivés sensibles, qualifiés d'hôtes principaux, mais aussi les hôtes auxiliaires, également appelés réservoirs. Ces plantes permettent aussi bien à l'agent pathogène de survivre, de s'installer, de se multiplier et éventuellement de se transmettre.

L'une des thématiques de l'équipe « Ecologie, Diversité et Taxonomie des Bactéries Phytopathogènes » (EDTa-BP) consiste à étudier les bases génétiques de la spécificité d'hôte chez les bactéries phytopathogènes. Les travaux de thèse d'Arnaud Darsonval (Thèse encadrée par Marie-Agnès Jacques et soutenue en avril 2008) ont pu montrer, auparavant, que plusieurs familles de gènes sont impliquées dans la colonisation de l'hôte par la bactérie. En effet, la survie des bactéries épiphytes dépend des adhésines non fibrillaires impliqués dans l'agrégation des cellules bactériennes en biofilm et des fonctions régulées par les régulateurs du système de sécrétion de type 3 (SST3), HrpG et HrpX. L'installation des bactéries à la surface foliaire dépend des adhésines fibrillaires et non fibrillaires. Enfin, la multiplication, la dispersion et la

transmission à la graine par voie vasculaire dépendent d'un SST3 fonctionnel et partiellement des adhésines (Darsonval *et al.*, 2008, 2009). La colonisation de la plantule par des bactéries transmises par les semences est par contre de type commensale, ne dépendant pas d'interaction étroite entre elles et la plantule (Darrasse *et al.*, 2010).

Pour déterminer les bases génétiques de la spécificité d'hôte, l'hypothèse émise est que la spécificité d'hôte reflète une adaptation étroite des bactéries à la plante. Selon cette hypothèse, la spécificité d'hôte est le résultat de plusieurs phases d'interaction : *i*) attraction de la bactérie par les exsudats de la plante hôte, *ii*) installation dans les tissus de l'hôte, *iii*) multiplication *in planta* et éventuellement expression du pouvoir pathogène et *iv*) dispersion et transmission des bactéries à un nouvel hôte. L'idée inhérente est que la spécificité d'hôte d'une bactérie phytopathogène peut être expliquée par un répertoire de gènes associés à l'interaction avec la plante encore appelés VA pour « virulence-associated genes ». L'approche choisie au sein de l'équipe a consisté à caractériser les répertoires de gènes candidats codant *i*) les chémorécepteurs -Methyl-accepting Chemotaxis Proteins (MCPs)- impliqués dans le comportement chimiotactique de la bactérie, *ii*) les senseurs des systèmes de régulation à deux composants et les transporteurs TonB-dépendants (TBDTs) impliqués dans la perception de signaux chimiques et environnementaux, *iii*) les adhésines fibrillaires et non fibrillaires impliquées dans l'adhésion de la bactérie à la surface des tissus végétaux et l'agrégation en biofilm et *iv*) les effecteurs de type III (ET3s) impliqués dans la capacité de la bactérie à se multiplier *in planta* et à se transmettre.

Dans le cadre du projet « Xanthost » portant sur la caractérisation des déterminants de la spécificité d'hôte des bactéries phytopathogènes, deux thèses complémentaires dirigées par Charles Manceau ont été initiées dans l'équipe EDTa-BP. Les bactéries modèles de cette étude appartiennent au genre *Xanthomonas*. Les travaux de thèse d'Ahmed Hajri (Thèse co-encadrée par Tristan Boureau et Stéphane Poussier et soutenue en décembre 2009) ont porté sur les ET3s qui interviennent lors des phases tardives de l'interaction *i.e* lors de la multiplication et l'expression des symptômes sur l'hôte. Ces travaux ont montré que les répertoires d'ET3s sont de bons candidats pour

expliquer la spécificité parasitaire des bactéries phytopathogènes du genre *Xanthomonas*. En effet, les répertoires d'ET3s corrélerent bien avec la structuration en pathovars de *X. axonopodis* (Hajri *et al.*, 2009). Au sein de cette espèce, les différentes lignées génétiques du pathovar *phaseoli* (*fuscans*, GL1, GL2 et GL3) pathogènes du haricot, bien qu'elles soient phylogénétiquement distinctes voire même éloignées, possèdent des répertoires d'ET3s similaires mais différenciables. Ceci peut traduire une convergence pathologique sur le haricot (Hajri *et al.*, 2009). Ces résultats sont en accord avec les travaux de Sarkar *et al.* (2006) qui ont porté sur l'étude des déterminants génétiques de la spécificité d'hôte chez *Pseudomonas syringae*. Selon ces auteurs, les ET3s présentent une distribution variable entre les souches de *P. syringae* et sont par conséquent, les meilleurs candidats pour expliquer la spécificité parasitaire.

Au cours de ma thèse, je me suis focalisée sur l'étude des premières étapes de l'interaction : attraction et installation. J'ai ainsi caractérisé les répertoires de gènes codant les MCPs, les senseurs des systèmes de régulation à deux composants, les transporteurs TonB-dépendants (TBDTs) et les adhésines fibrillaires et non fibrillaires dans une vaste collection de souches représentant la diversité de *X. axonopodis* et *X. campestris* (*sensu* Vauterin *et al.*, 1995). L'hypothèse que nous avons émise était que la spécificité d'hôte aboutissant à la survie et à la multiplication des bactéries voire même à la déclaration de la maladie commence dès les étapes précoces de l'interaction entre une bactérie et sa plante hôte. L'idée est que la spécificité d'hôte des bactéries phytopathogènes se dessine progressivement au cours de l'interaction avec l'hôte. La spécificité d'hôte serait d'abord large ; elle correspond à une gamme de plantes hôtes sensibles et de plantes réservoirs permettant la survie et la multiplication de la bactérie. Ensuite, la spécificité d'hôte se restreint ; elle se cantonne à une gamme de plantes hôtes qui expriment les symptômes de maladie provoqués par les bactéries pathogènes. À ce niveau, la spécificité d'hôte correspond de plus en plus à une spécificité parasitaire. Dans la première partie de ce manuscrit, je présente l'environnement auquel sont confrontés les *Xanthomonas* : la phyllosphère, puis les mécanismes bactériens de perception et d'adaptation à l'hôte et enfin leur évolution. Est-ce qu'au cours de l'évolution et des divergences écologiques des bactéries phytopathogènes, cette adaptation à l'hôte a été soumise à une pression de sélection éventuellement exercée par

l'hôte, est la question à laquelle je me suis proposée de répondre dans le premier chapitre de la partie Résultats.

Dans le deuxième chapitre, je me suis intéressée à l'étude des relations entre les souches de *X. axonopodis*. Dans un premier temps, j'ai entrepris une étude phylogénétique des souches de notre collection par le séquençage partiel de sept gènes de ménage (*atpD*, *dnaK*, *efp*, *fyuA*, *glnA*, *gyrB* et *rpoD*). L'objectif de ce travail vise à nous renseigner sur les relations qui existent entre les souches de notre collection. Au vu de l'importance de la recombinaison observée entre des souches de *X. axonopodis*, nous avons décidé en conséquence, de nous orienter vers une approche basée sur la coalescence. La coalescence permet en effet de retracer la généalogie *i.e* la dynamique des populations ancestrales et récentes des souches de *X. axonopodis* sur la base des séquences contemporaines des gènes de ménage. L'avantage de cette approche est qu'elle prend en compte des scénarios avec ou sans recombinaison ce qui n'est pas le cas des méthodes classiques de phylogénie qui ne considèrent que des scénarios sans recombinaison. La généalogie présente par ailleurs l'avantage d'estimer plusieurs paramètres de génétique des populations tels que les taux de migration, de recombinaison et de mutations et les temps de divergence. À ma connaissance, cette approche n'a encore jamais été appliquée sur les bactéries phytopathogènes. En examinant le cas des *Xanthomonas*, cette approche permettrait d'étudier autrement la diversité génétique et de mieux comprendre les processus démographiques et écologiques régissant la diversité de ces bactéries phytopathogènes et par conséquent l'émergence des maladies infectieuses.

Je terminerai enfin ce manuscrit par une conclusion générale portant sur les travaux que j'ai menés au cours de mes trois années de thèse et des perspectives qui s'ouvrent.

Première partie

Introduction

1. La phyllosphère

Les microorganismes capables de survivre et de se multiplier dans la phyllosphère sont souvent nommés épiphytes bien qu'ils colonisent à la fois la surface et les espaces intercellulaires des parenchymes (Lindow et Brandl, 2003). La phyllosphère est constituée par la feuille et l'environnement qui est soumis à son influence (Hirano et Upper, 2000). Les feuilles constituent un très vaste habitat pour les microorganismes. L'estimation de la surface globale des feuilles sur terre susceptible d'être colonisée par les microorganismes est de l'ordre de $6,4 \times 10^8 \text{ km}^2$. De plus, la population planétaire de bactéries épiphytes dans les régions tempérées et tropicales a été estimée à plus de 10^{26} cellules (Morris et Kinkel, 2002). Par ailleurs, le nombre de cellules bactériennes immigrantes déposées sur les feuilles est estimé à environ 10^4 cellules par mois (Lindow, 1996). L'examen d'empreintes de feuilles colonisées a révélé que les cellules bactériennes ne sont pas dispersées de façon aléatoire mais sont généralement localisées au niveau de quelques sites (Leveau et Lindow, 2001). En outre, ces cellules s'agrègent en microcolonies et forment des biofilms composés de plusieurs espèces bactériennes interagissant entre elles (Morris *et al.*, 1997, 1998 ; Lindow, 2002 ; Jacques *et al.*, 2005).

1.1 Topographie de la phyllosphère et propriétés physicochimiques de la cuticule

Le premier contact entre les bactéries immigrantes et la surface foliaire se fait au travers de la cuticule. Celle-ci présente une topographie variée tout le long du limbe. A l'échelle d'une bactérie, la surface foliaire est loin d'être uniforme et régulière. Elle peut être comparée à un relief qui se compose de vallées, de cavités, de plaines et de pics (Hirano et Upper, 2000 ; Beattie, 2002 ; Lindow et Brandl, 2003 ; Shepherd et Wagner, 2007). La topographie de la surface foliaire est en effet, marquée par la présence de stomates, de trichomes, d'hydathodes, de glandes et de crevasses dessinées par les nervures (Hirano et Upper, 2000). Des études en microscopie électronique ont montré

que les sites les plus communément colonisés par les bactéries sont situés à la base des trichomes, aux niveaux des stomates, des hydathodes et des dépressions de la cuticule et le long des nervures (Lindow et Brandl, 2003 ; Whipps *et al.*, 2008). De façon générale, les bactéries se localisent en grand nombre au niveau de la surface inférieure de la feuille qui possède une cuticule plus fine que celle de la surface supérieure et une plus grande densité en stomates et trichomes (Beattie et Lindow, 1999 ; Whipps *et al.*, 2008). Par exemple, les cellules de *P. syringae* s'agrègent en forte densité exclusivement à la base des trichomes glandulaires. Ces derniers offrent en effet des conditions optimales pour la croissance bactérienne car ils exsudent de fines gouttelettes d'eau riches en nutriments (Monier et Lindow, 2003). Par ailleurs, une étude a montré que la taille de la population bactérienne est positivement corrélée avec la densité de trichomes glandulaires et non-glandulaires (Yadav *et al.*, 2005).

La cuticule est composée essentiellement de cutine et de cire. La cutine est une matrice formée d'un biopolymère de type polyester. Elle est principalement constituée de deux familles d'acides gras en C₁₆ et en C₁₈, les acides ω -hydroxy et ω -hydroxyepoxy. Les différents monomères d'acides gras sont généralement fixés les uns aux autres par des liaisons esters. La composition et la quantité de cutine par unité de surface, peuvent varier considérablement entre les espèces végétales. Ces variations favorisent une reconnaissance spécifique des espèces végétales par les microbes (Beattie, 2002 ; Müller et Reiderer, 2005 ; Pollard *et al.*, 2008). La cire est présente dans la fraction lipophile de la cuticule. Il s'agit d'un ester dont les composés majeurs sont des *n*-alcanes ayant une longueur de chaîne comprise entre C₂₁ et C₃₅. D'autres composés en moindre proportion entrent dans la composition de la cire tels que les alcools primaires (avec une longueur de chaîne comprise entre C₂₂ et C₄₀), les acides gras (avec une longueur de chaîne comprise entre C₂₀ et C₂₄), les aldéhydes (avec une longueur de chaîne comprise entre C₂₄ et C₃₆), les alcools secondaires (avec une longueur de chaîne comprise entre C₂₁ et C₃₅), des cétones (avec une longueur de chaîne comprise entre C₂₁ et C₃₅), des β -dicétones (avec une longueur de chaîne comprise entre C₂₂ et C₃₆) et d'esters *n*-alkyle (avec une longueur de chaîne comprise entre C₃₆ et C₆₀) (Müller et Reiderer, 2005 ; Koch et Ensikat, 2008). Les acides triterpènes pentacycliques, les alcools, les cétones et les esters triterpènes et les dérivés

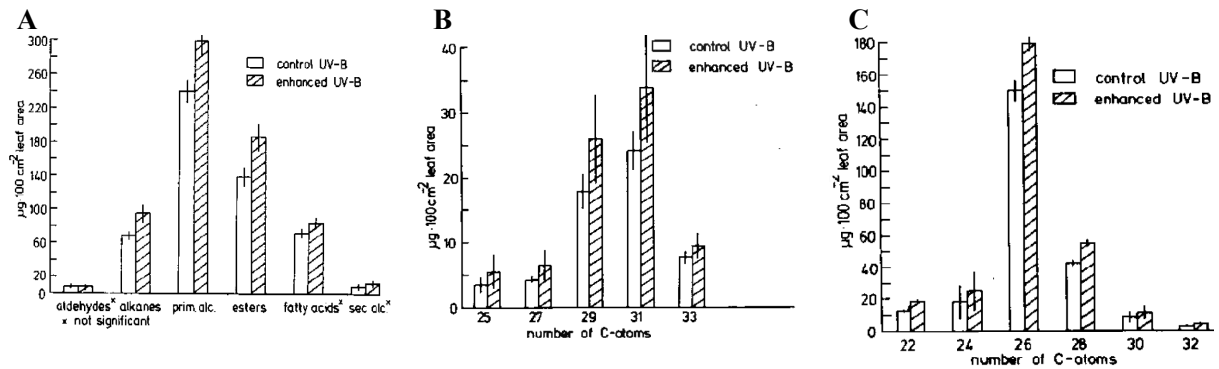


Figure 1. Composition de la cire cuticulaire des deux premières feuilles de plantes de *Phaseolus vulgaris* en fonction de l'exposition aux rayons UV-B (d'après Steinmüller et Tevini, 1985). Composition générale des deux premières feuilles de haricot (A). Composition en alcanes des deux premières feuilles de haricot (B). Composition en alcools primaires des deux premières feuilles (C).

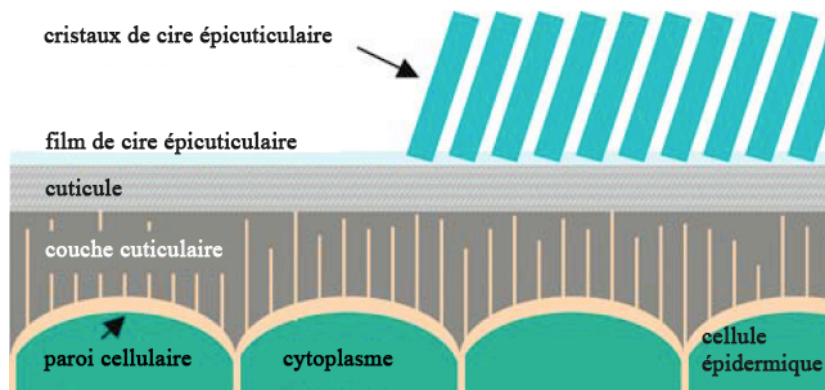


Figure 2. Représentation schématique de la structure de la cuticule (d'après Müller et Reiderer, 2005).

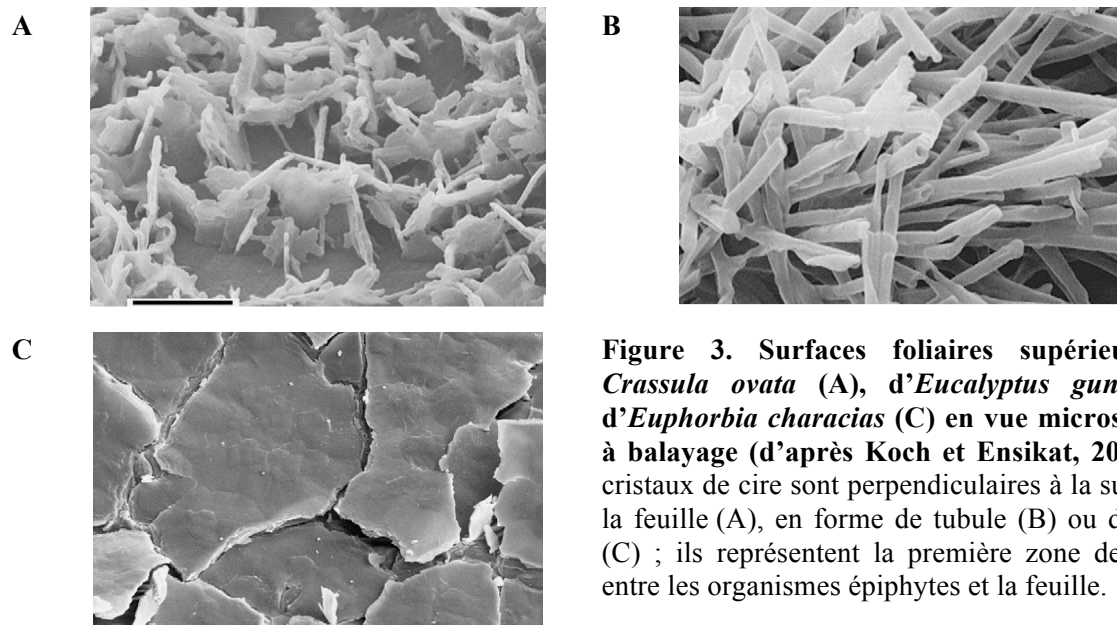


Figure 3. Surfaces foliaires supérieures de *Crassula ovata* (A), d'*Eucalyptus gunnii* (B), d'*Euphorbia characias* (C) en vue microscopique à balayage (d'après Koch et Ensikat, 2007). Les cristaux de cire sont perpendiculaires à la surface de la feuille (A), en forme de tubule (B) ou de croute (C) ; ils représentent la première zone de contact entre les organismes épiphytes et la feuille.

hydroxycinnamique, les flavonoïdes et les glucosides peuvent également entrer en quantités variables dans la composition de la cire (Figure 1) (Müller et Reiderer, 2005 ; Koch et Ensikat, 2008).

Au plan structural, la cuticule est une matrice composée de deux couches. Celles-ci se distinguent par leur ultra structure, leur composition chimique et leurs propriétés intrinsèques (Figure 2). En effet, la couche la plus externe, dite « *cuticule sensu stricto* », forme une pellicule continue et se compose d'un mélange de cutine et de cire. Elle est elle même recouverte par un film de cristaux de cire épicuticulaire (Figures 2 et 3), responsable de l'aspect visuel de la feuille. Les cristaux de cire sont principalement composés par de *n*-alcanes et sont spécifiques de l'espèce végétale. Le recouvrement et la composition du film de cristaux de cire épicuticulaire varient en fonction de plusieurs facteurs tels que les saisons, la lumière, le vent, la pluie (Müller et Reiderer, 2005 ; Koch et Ensikat, 2008).

Les propriétés physicochimiques de la cuticule influencent la transpiration et les échanges gazeux. La cuticule forme une barrière hydrophobe qui limite la perméabilité de la feuille. Elle réduit en conséquence les diffusions passives des nutriments et de l'eau de l'intérieur vers la surface de la feuille. Ainsi, les cuticules les plus épaisses peuvent fortement influencer la colonisation bactérienne à travers la limitation de la diffusion des nutriments nécessaires à la survie des microbes épiphytes (Hirano et Upper, 2000 ; Beattie, 2002 ; Lindow et Brandl, 2003 ; Müller et Reiderer, 2005 ; Yadav *et al.*, 2005). De plus, c'est au niveau de la cuticule que l'adhésion des microbes épiphytes se fait. Les adhésines reconnaissent spécifiquement des composants moléculaires qui se trouvent à la surface de leur hôte (Niemann *et al.*, 2003). Les propriétés physicochimiques (hydrophobie, disponibilité en nutriments) de la cuticule vont donc influencer l'étape d'installation des bactéries lors de leur interaction avec la plante hôte (Beattie, 2002 ; Müller et Reiderer, 2005). Il a été montré que *X. axonopodis* pv. *mangiferaeindicae* a une faible capacité à se maintenir à la surface des feuilles de manguier contrairement à *X. axonopodis* pv. *phaseoli*. Ceci pourrait être expliqué par la nature des feuilles de manguier qui contrairement aux feuilles de haricot, présentent une cuticule plus riche en cire et une faible densité de trichomes (Pruvost *et al.*, 2009).

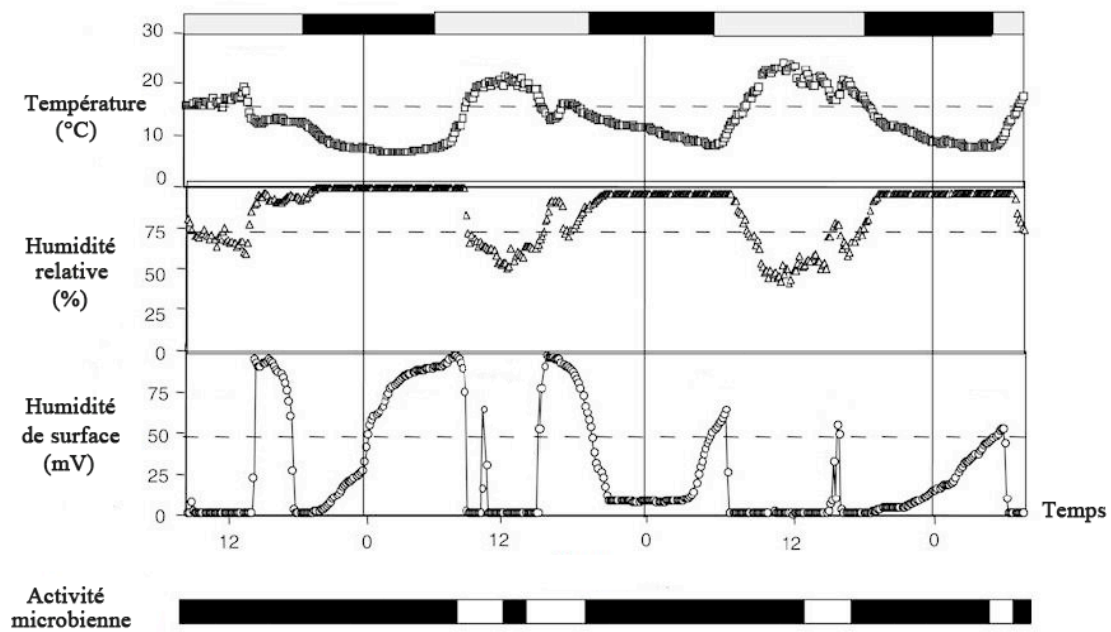


Figure 4. Fluctuations de la température, de l'humidité relative et de l'humidité de surface à la surface foliaire du poirier au cours de la journée (d'après Montesinos *et al.*, 2002). L'activité microbienne est représentée par les bandes horizontales blanches. Elle est importante lorsque la température est supérieure à 15°C et l'humidité de surface est supérieure à 50 mV.

1.2 Caractéristiques environnementales de la phyllosphère

La phyllosphère est un environnement hostile pour les bactéries épiphytes. Exposée à l'atmosphère et au soleil, la surface foliaire est sujette à des fluctuations rapides et extrêmes de la température, de l'humidité relative, du vent, des rayonnements solaires et de la disponibilité en eau et en nutriments (Hirano et Upper, 2000 ; Montesinos *et al.*, 2002 ; Lindow et Brandl, 2003). Ces fluctuations peuvent se produire en quelques secondes ou sur une échelle de temps plus longue comme le cycle jour/nuit (Figure 4) (Hirano et Upper, 2000 ; Montesinos *et al.*, 2002). La capacité des bactéries épiphytes à survivre et à se multiplier dans cet environnement instable dépend en partie des caractéristiques physicochimiques de la surface foliaire (Whipps *et al.*, 2008).

1.2.1 Paramètres physiques de l'environnement

Les paramètres physiques environnementaux affectent considérablement la vie des bactéries épiphytes ; ils conditionnent la survie, le taux de multiplication, la réponse chimiotactique et d'autres processus essentiels pour les bactéries épiphytes (Montesinos *et al.*, 2002). Les rayonnements ultraviolets du soleil et plus particulièrement les courtes longueurs d'onde UV-B (290-320 nm) endommagent l'ADN et d'autres constituants de la cellule bactérienne et sont donc particulièrement nuisibles pour les bactéries épiphytes (Gunasekera et Paul, 2007). Le manque de disponibilité en eau libre au niveau de la surface foliaire peut également altérer les caractéristiques chimiques de l'environnement des bactéries épiphytes. En effet, la dessiccation provoque l'augmentation de la concentration des solutés au niveau de la phyllosphère et par conséquent, l'augmentation de la pression osmotique. Cette augmentation peut être dommageable pour l'intégrité des cellules bactériennes épiphytes (Beattie et Lindow, 1995). En outre, des chutes de pluie importantes peuvent lessiver complètement les parties aériennes des plantes infectées et contribuer négativement à la colonisation de la phyllosphère (Lindemann et Upper, 1985 ; Butterworth et McCartney, 1991 ; Hirano *et al.*, 1994). Dans le cas de *P. syringae* pv. *tomato*, les fortes chutes de pluie réduisent

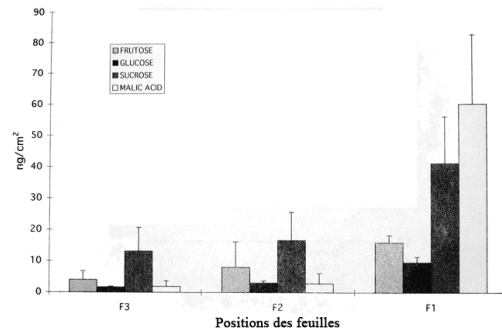


Figure 5. Variations des quantités de sucres et d'acide malique collectés au niveau de la surface inférieure des feuilles de *Nicotiana plumbaginifolia*, en fonction de l'âge et de la position des feuilles (d'après Derridj, 1996). Les feuilles sont divisées en trois groupes de trois feuilles (F1, F2 et F3) avec F1 correspondant aux feuilles les plus âgées.



Figure 6. Photographie des trichomes des feuilles de tomate colonisés par *X. axonopodis* pv. *vesicatoria*, réalisée par microscopie électronique à balayage (d'après Ojanen-Reuhs *et al.*, 1997). Les flèches précisent la localisation préférentielle des bactéries.

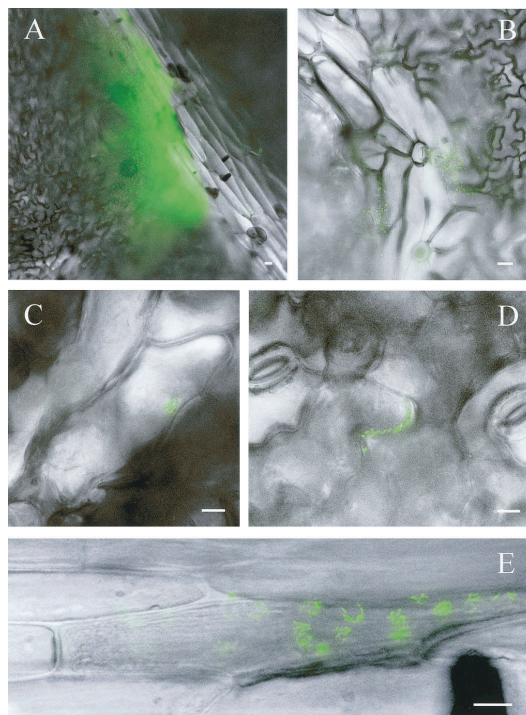


Figure 7. Observations directes de *Erwinia herbicola* (Eh299R) marquée à la GFP dans la phyllosphère par microscopie à épifluorescence (d'après Leveau et Lindow, 2001). De nombreuses bactéries fluorescentes se trouvent dans les gouttes d'eau (A) et sont associées aux trichomes (B). Elles sont également localisées au niveau des nervures (C) et des crevasses formées par les cellules épidermiques (D). Vingt quatre heures après inoculation, une forte fluorescence a été enregistrée au niveau des petites nervures et des couches de cellules épidermiques (E).

effectivement la population épiphyte (Pietrarelli *et al.*, 2006). En revanche, la dissémination de *X. axonopodis* pv. *allii* est favorisée par la succession de deux événements climatiques : les épisodes venteux et pluvieux suivis d'une augmentation de la température (Roumagnac *et al.*, 2004a).

1.2.2 Disponibilité en nutriments et autres métabolites d'origine végétale

La disponibilité en nutriments, sources de carbone et d'azote, à la surface foliaire est un déterminant majeur de la colonisation par les bactéries épiphytes. Les composés azotés limitent de façon secondaire la croissance des bactéries (Lindow et Brandl, 2003 ; Whipps *et al.*, 2008). La composition et la disponibilité en nutriments comprenant les carbohydrates, les acides organiques et les acides aminés nécessaires à la croissance bactérienne varient selon l'espèce de la plante hôte, l'âge de la feuille (Figure 5), l'état physiologique de la feuille et la présence de tissus endommagés (Blakeman, 1972 ; Fiala *et al.*, 1990). Les sucres simples comme le glucose, le fructose et le saccharose sont les sources de carbone dominantes au niveau de la phyllosphère. Néanmoins, d'autres sources exogènes de nutriments tels que le miellat sécrété par les pucerons, le pollen et les débris cellulaires contribuent positivement à la colonisation des bactéries épiphytes (Mercier et Lindow, 2000 ; Stadler et Müller, 2000).

Au niveau de la surface foliaire, les nutriments présentent une distribution spatiale hétérogène. En effet, une étude utilisant des cellules d'*Erwnia herbicola* marquées à la GFP et pulvérisées sur des plants de haricots, a montré qu'une heure après inoculation, la majorité des cellules bactériennes pulvérisées sont engagées dans un processus de consommation des sucres. En revanche, 24h après inoculation, moins de 1% des bactéries inoculées continuent à consommer les sucres. Celles-ci ne sont pas aléatoirement dispersées au niveau de la phyllosphère mais se localisent au niveau de sites particuliers tels que la base des trichomes, le long des nervures et près des stomates où les nutriments sont relativement abondants (Figures 6 et 7) (Leveau et Lindow, 2001). Ainsi, la plupart des bactéries immigrantes se retrouvent dans un environnement pauvre en nutriments avec des conditions de croissance défavorables alors que quelques

Tableau 1. Rendement en huiles essentielles et concentration des composés phénoliques extraits des surfaces foliaires de 19 espèces végétales (d'après Karamanoli *et al.*, 2005).

Espèces végétales¹	Huile essentielle (ml/100g de ps)²	Composés phénoliques (µg/g de pf)³
Cèdre	0,2 ± 0,04	-
Oranger	0,4 ± 0,1	18,6 ± 3,1
Lavande	0,8 ± 0,2	23,4 ± 1,2
Laurier	1,2 ± 0,1	23,8 ± 2,1
Tabac	-	25,3 ± 1,7
Ail	-	33,0 ± 5,8
Olivier	-	36,6 ± 2,7
Eucalyptus	1,3 ± 0,1	52,6 ± 2,9
Origan	2,9 ± 0,3	370 ± 3,1
Romarin	1,9 ± 0,3	487 ± 3,6
Sauge	2,0 ± 0,2	545 ± 3,8

¹: Aucune huile essentielle et aucun composé phénolique n'a été détecté dans les autres plantes testées : oignon, chou-fleur, laitue, poireau, pomme de terre, céleri, chou et aneth. ²: poids sec. ³: poids frais ; (moyenne ± écart type).

cellules rencontrent des sites favorables relativement riches en nutriments (Lindow et Brandl, 2003 ; Monier et Lindow, 2005 ; Whipps *et al.*, 2008).

Des métabolites secondaires tels que les alcaloïdes, les terpènes (huiles essentielles) et les acides phénoliques (flavone, quercétine et naringénine) ainsi que des protéines de défense peuvent être présents à la surface foliaire et avoir une activité antimicrobienne (Kubo *et al.*, 1995 ; Rauha *et al.*, 2000 ; Karamanoli *et al.*, 2005, Yadav *et al.*, 2005). Généralement, ces molécules sont secrétées par les trichomes glandulaires, les hydathodes et les cellules épidermiques puis se dispersent sur toute la surface foliaire (Shepherd et Wagner, 2007). Karamanoli *et al.* (2005) ont pu mettre en évidence que ces métabolites secondaires sont impliqués dans les défenses basales des plantes contre les invasions bactériennes. En effet, les métabolites secondaires à activité antimicrobienne peuvent agir directement sur la croissance des microorganismes (Stasiuk et Kozubek, 2010). Les composés phénoliques comme par exemple les acides phénoliques et les quinones, sont très réactifs et constituent par conséquent une source importante de radicaux libres. Ces radicaux peuvent se complexer irréversiblement aux protéines bactériennes provoquant ainsi leur inactivation et éventuellement, la perte de leur fonction (Mason et Wasserman, 1987). De plus, les composés phénoliques comme l'acide anacardique, peuvent inhiber la chaîne respiratoire (Kaprelyants *et al.*, 1987). L'acide anacardique peut également provoquer des désordres au niveau de la membrane bactérienne (Green *et al.*, 2007). Selon Cowan (1999), les structures microbiennes exposées au milieu externe, telles que les adhésines, peuvent constituer des cibles potentielles des composés phénoliques tels que les quinones et les tanins. Il a été récemment démontré qu'une huile essentielle, la cinnamone, agit sur les cellules planctoniques et les biofilms de la bactérie à Gram positif, *Staphylococcus epidermidis* (Nuryastuti *et al.*, 2009). Ces auteurs ont pu montrer moyennant l'imagerie par microscopie confocale, que la cinnamone est même capable de détacher et de tuer les biofilms préformés.

L'analyse des métabolites secondaires de 19 espèces végétales a permis de montrer que leur concentration et composition varient en fonction des espèces (Tableau 1) (Karamanoli *et al.*, 2005). Ces auteurs ont de plus mis en évidence que les

Tableau 2. Activité antibactérienne des métabolites secondaires provenant de différentes espèces végétales contre les bactéries *E. herbicola* et *P. syringae* (d'après Karamanoli *et al.*, 2005).

Plant species	Essential oils		Polar compounds		Phenolic compounds		Average antibacterial activity ²
	<i>E. herbicola</i>	<i>P. syringae</i>	<i>E. herbicola</i>	<i>P. syringae</i>	<i>E. herbicola</i>	<i>P. syringae</i>	
Eucalypt	2	2	2	2	0	0	1.33
Rosemary	2	1	1	1	2	2	1.45
Oregano	3	2	1	1	2	2	1.65
Tobacco ¹	–	–	3	3	0	0	1.50
Bay	2	2	2	0	0	0	0.95
Sage	2	1	1	1	2	2	1.40
Onion	–	–	1	0	0	0	0.50
Oleander	–	–	0	0	0	0	0.00
Cauliflower	–	–	0	0	0	0	0.00
Garlic ¹	–	–	3	3	0	0	1.50
Orange	0	0	0	0	0	0	0.00
Cedar	0	0	2	0	0	0	0.30
Lettuce	–	–	0	0	0	0	0.00
Cabbage	–	–	2	0	0	0	0.50
Lavender	1	0	1	0	0	0	0.60
Dill	–	–	2	0	0	0	0.50
Potato	–	–	0	0	0	0	0.00
Leek	–	–	2	0	0	0	0.50
Celery	–	–	0	0	0	0	0.00

0 = pas d'inhibition, 1 = zone d'inhibition <8 mm, 2 = zone d'inhibition comprise entre 8 et 15 mm, 3 = zone d'inhibition comprise entre 15 et 25 mm, – = l'essai n'a pas été réalisé. ¹: l'activité antimicrobienne est attribuée au composant principal, ²: l'activité antimicrobienne moyenne de l'espèce est la somme des scores divisée par le nombre de bioessais réalisés pour chaque espèce.

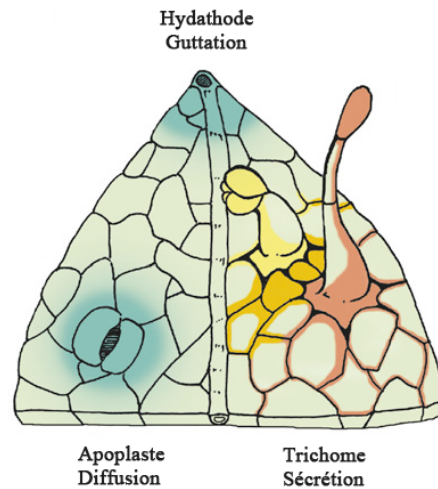


Figure 8. Représentation schématique des différentes voies empruntées par les protéines pour se déposer sur la phyllosphère (d'après Shepherd et Wagner, 2007). Les protéines (en bleu) se trouvant au niveau de l'apoplaste, suivent des voies hydrophiles qui favorisent leur diffusion à travers les stomates ou leur dépôt lors de la guttation au niveau des hydathodes. D'autres protéines (en jaune), sont synthétisées et exsudées par les trichomes glandulaires. Elles s'accumulent à la base de ces structures foliaires. Certains trichomes produisent et secrètent des métabolites secondaires (en rouge), favorisant la formation d'une matrice hydrophobe sur laquelle les protéines du phylloplan peuvent de disperser et se dissoudre.

métabolites secondaires extraits à partir des feuilles de l'origan, du tabac, de l'ail, du romarin, de sauge, de l'eucalyptus et de laurier ont une forte activité antimicrobienne. En revanche, les métabolites secondaires extraits des feuilles de chou-fleur, céleri, laitue, olivier, oranger et pomme de terre n'exercent aucune activité antimicrobienne. Enfin, les métabolites secondaires extraits des feuilles de lavande, de chou, d'oignon, de cèdre, de laurier, de poireau et de l'aneth ont une activité antimicrobienne intermédiaire. Il a ainsi été observé que les espèces dont les métabolites secondaires ont une forte activité microbienne, sont faiblement colonisées par les bactéries épiphytes et inversement (Karamanoli *et al.*, 2005). En outre, les bactéries *E. herbicola* et *P. syringae* ont toutes deux été isolées à partir des feuilles des 19 plantes testées. L'application des différents extraits sur ces deux bactéries a montré que *E. herbicola* est plus sensible que *P. syringae* (Tableau 2). En effet, les extraits de feuilles d'eucalyptus, d'origan, de romarin et de sauge inhibent la croissance des deux bactéries alors que les extraits de feuilles de chou, de poireau, d'oignon, de l'aneth, et du laurier ont une activité antibactérienne seulement sur *E. herbicola*. En revanche, les extraits de feuilles d'oranger, de céleri, de pomme de terre, de laitue, d'olivier et de chou-fleur sont totalement inactifs sur les deux bactéries. Ceci illustre bien le fait que *P. syringae* est plus largement distribuée dans la phyllosphère et explique en partie, sa participation massive dans les communautés microbiennes épiphytes (Karamanoli *et al.*, 2005). Par ailleurs, une autre étude a également montré que la densité bactérienne est négativement corrélée aux concentrations élevées d'un glucosinolate, le glucosinolate 2-phényléthyle, présent à la surface de la feuille (Ruppel *et al.*, 2008).

Les protéines antimicrobiennes peuvent diffuser dans la phyllosphère à partir de l'apoplaste et à travers les stomates. Ceci est le cas des défensines, des chitinases et des lipases (Figure 8) (Ancillo *et al.*, 2003 ; Oh *et al.*, 2005 ; Silverstein *et al.*, 2005). Les protéines antimicrobiennes peuvent également être déposées à la surface foliaire lors de la guttation au niveau des hydathodes. En effet, une peroxydase cationique a été découverte dans les gouttes de guttation du riz au cours d'une infection causée par *X. oryzae* pv. *oryzae* (Young *et al.*, 1995). De même, des gouttes de guttation collectées à partir de feuilles d'orge contiennent des protéines antimicrobiennes inhibant la pénétration de bactéries mobiles par les hydathodes. Par ailleurs, il a été montré que les

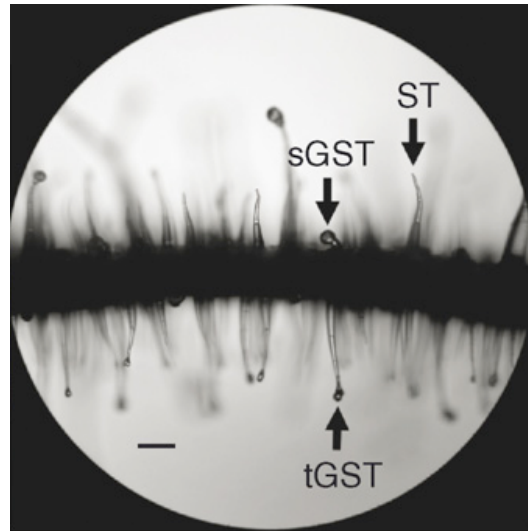


Figure 9. Trichomes à la surface des feuilles de *N. tabacum* (d'après Shepherd et Wagner, 2007). Il existe les trichomes longs et glandulaires (tGST) qui secrètent d'importantes quantités de métabolites secondaires et les trichomes courts et glandulaires qui secrètent des protéines (sGST). De simples trichomes (ST) peuvent également être présents. La barre d'échelle représente 0,01 mm.

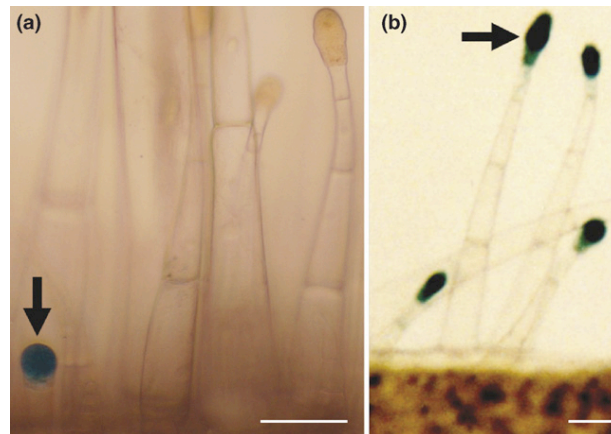


Figure 10. Expression de la protéine GUS dans les trichomes glandulaires courts (a) et longs (b) de la feuille de *N. tabacum* (d'après Shepherd et Wagner, 2007). Le gène GUS s'exprime spécifiquement sous le contrôle du promoteur du gène *T-phylloplanine* pour les trichomes courts et d'un promoteur spécifique CYP71D16 pour les trichomes longs. La phylloplanine est une protéine qui inhibe la germination des spores de l'oomycète *Peronospora tabacina*. La barre d'échelle représente 0,05 mm.

protéines de défense telles que les peroxydases dismutases et les chitinases, peuvent se localiser au niveau des trichomes de feuilles de tabac (Figures 9 et 10) (Amme *et al.*, 2005).

1.2.3 Composés organiques volatiles émis par la feuille

Toutes les plantes émettent des quantités importantes de composés organiques volatiles (Peñuelas et Llusà, 2008). Ces composés comprennent des alcanes, des alcènes, des alcools, des aldéhydes, des éthers, des esters et des acides carboxyliques (Peñuelas et Llusà, 2008). Les composés volatiles jouent un rôle important dans l'attraction et la répulsion des microorganismes interagissant avec la plante (Müller et Reiderer, 2005). Les alcools en C₆ saturés et insaturés et les aldéhydes volatils responsables de l'odeur verte des feuilles, peuvent en effet, inhiber la prolifération des microorganismes (Schönherr et Baur, 1996). Le *trans*-2-hexanal et le *cis*-3-hexenol sont principalement responsables de l'odeur fraîche du thé vert (Takéo, 1992). Ces deux composés sont aussi produits par les feuilles de haricot lors d'une interaction incompatible avec *P. syringae* pv. *phaseolicola*. D'ailleurs, les feuilles de haricot montrant des symptômes d'hypersensibilité (HR) émettent une odeur comparable à celle du thé (Croft *et al.*, 1993). En outre, il a été montré que le *trans*-2-hexanal a une forte activité bactéricide à de faibles concentrations alors que le *cis*-3-hexenol a une activité bactéricide à de fortes concentrations (Croft *et al.*, 1993). Ces mêmes auteurs ont également montré que les souches de *P. syringae* pv. *phaseolicola* exposées à 2,2 mM de *trans*-2-hexanal semblent être lysées. Les cellules bactériennes présentent en effet une membrane cytoplasmique interrompue et leur ADN est précipité. La densité optique de la culture à 600 nm diminue après le traitement avec *trans*-2-hexanal (Croft *et al.*, 1993).

Il existe deux types de composés organiques volatiles. Le premier type correspond aux composés organiques volatiles qui peuvent s'accumuler au niveau de structures foliaires spécifiques comme les trichomes glandulaires. Ces composés influencent fortement la colonisation de la surface foliaire par les microorganismes épiphytes (Peñuelas et Llusà, 2008). Le second type de composés organiques volatiles

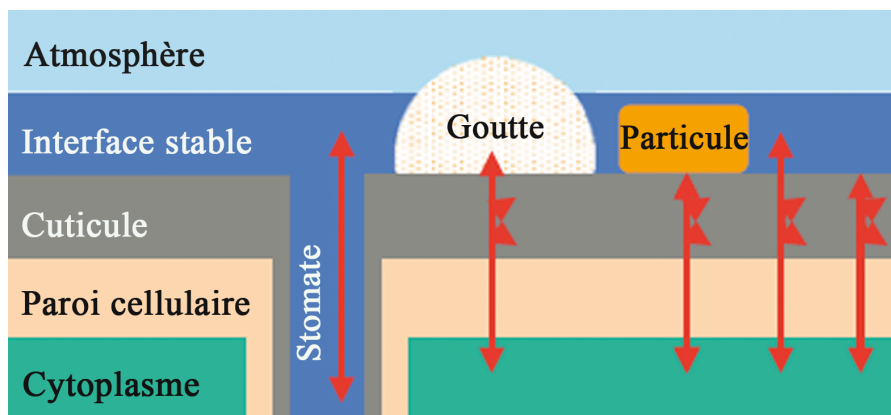


Figure 11. Voies de diffusion des composés volatiles à partir des tissus foliaires (d'après Müller et Reiderer, 2005).

ne s'accumule pas au niveau de structures spécifiques. Selon l'hypothèse émise par Rosenstiel *et al.* (2004), ces composés sont émis par la plante en tant que « purge métabolique de sécurité ». Cette purge permet d'éliminer l'excès de carbone qui n'est pas utilisé par la plante. Parmi ces composés volatiles, les isoprènes sont les plus abondants. Ces dérivés isopréniques sont synthétisés à partir du diphosphate de diméthylallyle (DMAPP) au niveau des chloroplastes comme d'ailleurs la chlorophylle et les caroténoïdes. Le DMAPP doit être toujours disponible au niveau de ces organites mais son accumulation provoque la séquestration du phosphate. L'enzyme isoprène synthétase se charge de prévenir cette accumulation en catalysant la synthèse d'isoprènes. L'excès d'isoprènes est émis par la feuille, sous forme de composés organiques volatiles (Rosenstiel *et al.*, 2004). Les isoprénoides émis peuvent éventuellement réagir avec les composés de la cuticule pour protéger les feuilles d'éventuels dommages dus à l'oxydation (Peñuelas et Llusà, 2008). Niinemets *et al.* (2004) proposent que les émissions de composés organiques sont principalement provoquées par les changements des conditions environnementales et varient en fonction des caractéristiques physicochimiques (solubilité, diffusibilité et volatilité) des composés (Niinemets *et al.*, 2004). Les composés très volatiles comme les gaz, les monoterpènes et les dérivés volatiles du glycosinolate sont émis aussi longtemps que les stomates sont ouverts. Dans le cas où ces derniers sont fermés, les composés très volatiles diffusent à travers la cuticule (Figure 11) (Müller et Reiderer, 2005).

2. Perception et adaptation à l'hôte

Dans un environnement tel que la phyllosphère, où les conditions physicochimiques varient constamment, il est essentiel que les microorganismes soient capables de détecter les changements et d'ajuster en conséquence leur métabolisme, leur physiologie et leur comportement (Bijlsma et Groisman, 2003 ; Wadhams et Armitage, 2004 ; Galperin, 2005). Les réponses possibles correspondent à une modification de l'expression de gènes et/ou à un déplacement actif dans l'habitat (Wadhams et Armitage, 2004). Les bactéries épiphytes vivant en association très étroite avec leurs plantes hôtes bénéficient de cette association pour acquérir des nutriments. Les événements conduisant à l'établissement de l'interaction entre ces bactéries et leurs hôtes commencent par la perception, la reconnaissance de molécules spécifiques d'origine végétale et par l'adhésion aux tissus de l'hôte (Brenner et Winans, 2005). Le chimiotactisme est un mécanisme par lequel les bactéries reconnaissent leur habitat *i.e* leur plante hôte, et par lequel elles répondent efficacement aux changements de la composition chimique de leur environnement.

2.1 Chimiotactisme : Perception, reconnaissance spécifique et adaptation à l'hôte

Le chimiotactisme est une réponse adaptative des bactéries à leur environnement (Parkinson *et al.*, 2005 ; Rao et Ordal, 2009, Vladimirov et Sourjik, 2009). Il s'agit d'un trait microbien essentiel permettant d'optimiser l'accès aux nutriments (molécules chémoattractives) et l'évitement des molécules toxiques (molécules chémorépulsives). De plus, il requiert un système sophistiqué de reconnaissance des stimuli, d'amplification de signaux et de traitement de l'information (Parkinson *et al.*, 2005 ; Rao et Ordal, 2009 ; Vladimirov et Sourjik, 2010). Les voies de signalisation impliquées dans le chimiotactisme ont beaucoup été étudiées dans le cas des bactéries entériques comme *Escherichia coli* et *Salmonella enterica* serovar Typhimurium (Armitage, 1999 ; Bren et Eisenbach, 2000). Les chémorécepteurs associés à la

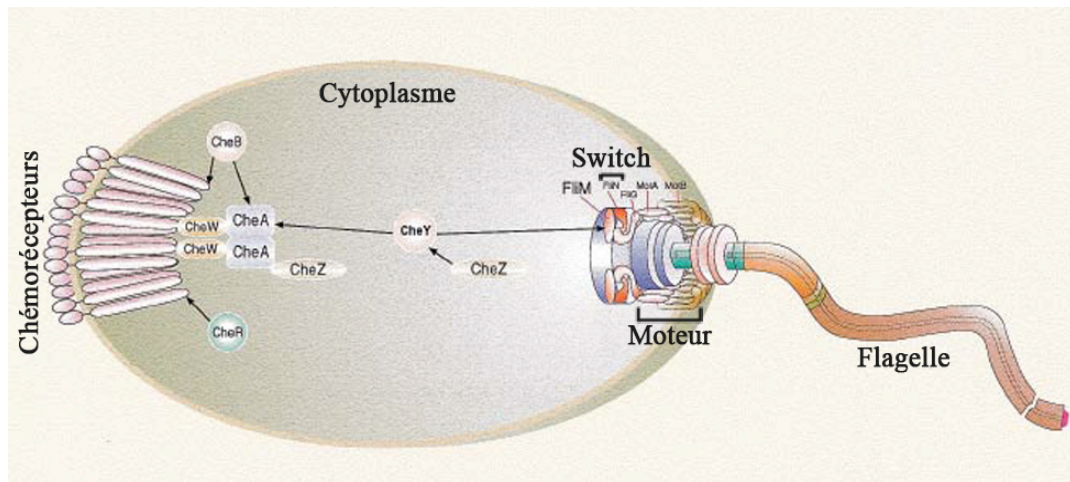


Figure 12. Représentation schématique de la voie de transduction du signal chimiotactique chez les bactéries (d'après Bren et Eisenbach, 2000). Les chémorécepteurs (MCPs) sont ancrés dans la membrane ; ils détectent les signaux chimiotactiques et interagissent avec les protéines cytoplasmiques Che. Celles-ci se chargent de transmettre, *via* une cascade de phosphorylation, le signal à l'intérieur de la cellule. La protéine CheY phosphorylée interagit avec le moteur du flagelle pour contrôler la mobilité.

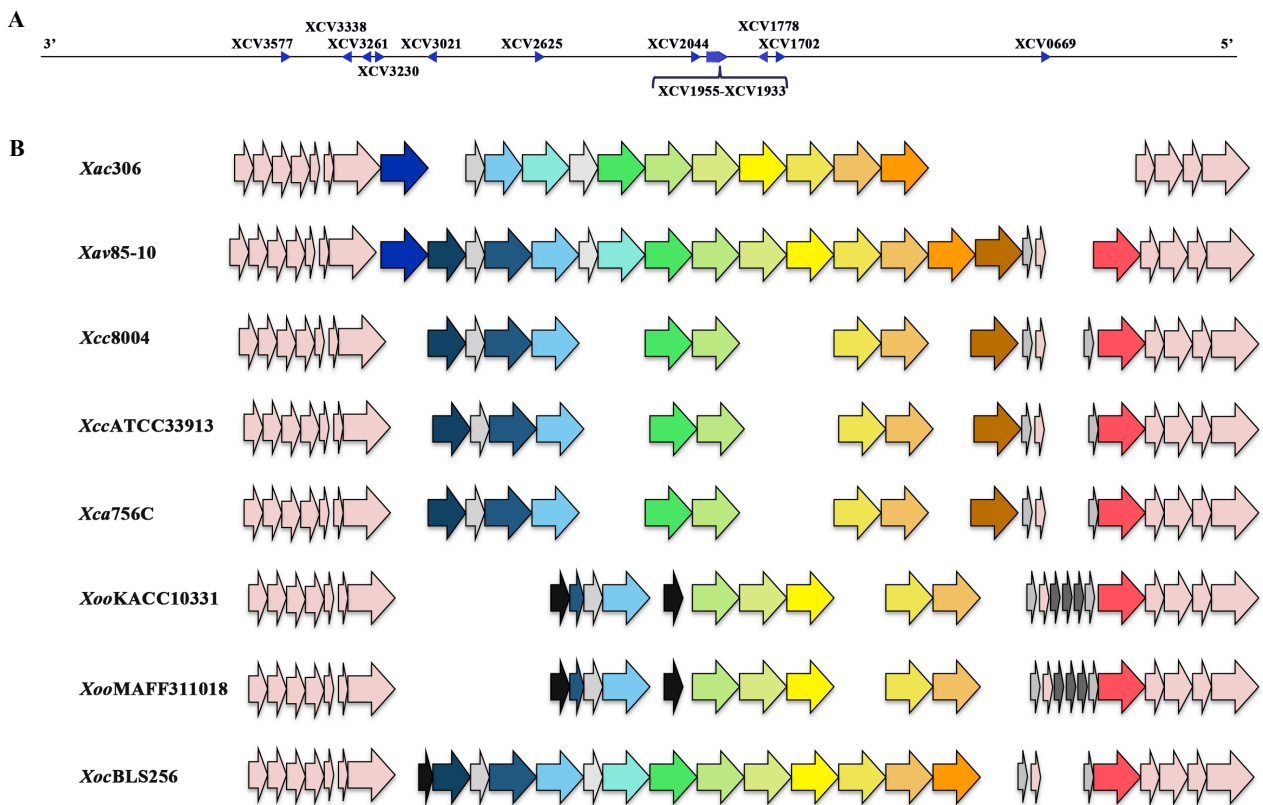


Figure 13. Organisation des gènes codant les MCPs dans les génomes de *Xanthomonas* spp. A : Organisation des gènes codant les MCPs dans le génome de la souche 85-10 de *X. axonopodis* pv. *vesicatoria*. L'encart renvoie à l'îlot génomique détaillé en B. B : Alignements des gènes codant les MCPs localisés dans l'îlot génomique dédié à la réponse chimiotactique et à la mobilité dans huit souches séquencées de *Xanthomonas* (d'après Koebnik, 2007). En rose clair, sont représentés les gènes codant les protéines Che du chimiotactisme ; en gris, les protéines hypothétiques, en noir les séquences d'insertions. Les gènes colorés en nuances de bleu, vert, jaune, orange et rouge représentent les gènes codant les MCPs. Pour faciliter la lecture des alignements, les gènes sont représentés de 3' en 5' car la majorité des gènes sont positionnés sur le brin -.

membrane, les Methyl-accepting Chemotaxis Proteins (MCPs), constituent le principal composant du système chimiotactique. Ces senseurs détectent spécifiquement les stimuli environnementaux et réagissent en conséquence, par un changement de conformation. Celui-ci provoque dans le cytoplasme, une cascade de transduction de signal qui repose sur des réactions de phosphorylation et aboutit au contrôle de la fréquence et du sens de rotation du flagelle (Figure 12) (Szurmant et Ordal, 2004).

2.1.1 Methyl-accepting Chemotaxis Proteins

Quatre et cinq gènes codant des MCPs ont été identifiés respectivement dans les génomes d'*E. coli* et de *Salmonella*. Ces MCPs peuvent détecter différentes molécules. Les MCPs de type Tsr reconnaissent la sérine, ceux de type Tar reconnaissent l'aspartate et le maltose, les MCPs de type Tcp reconnaissent le ribose, le galactose et le glucose, ceux de type Trg reconnaissent les citrates et ceux de type Tap, les dipeptides (Manson *et al.*, 1998 ; Hazelbauer *et al.*, 2007).

Chez *E. coli*, la duplication en tandem de gènes codant les MCPs a été rapportée pour la première fois par Wang *et al.* en 1982. Les deux gènes identifiés codent des chémorécepteurs (MCPs) similaires mais non identiques. En effet, le premier gène code un MCP de type Tar alors que le deuxième gène code un MCP de type Tap (Wang *et al.*, 1982 ; Slocum et Parkinson, 1983). La duplication des gènes permet l'acquisition de nouvelles fonctions au cours de l'évolution.

Pour ce qui est des *Xanthomonas*, les gènes codant les MCPs sont regroupés et certains d'entre eux sont en tandem dans l'îlot génomique dédié au chimiotactisme et à la mobilité. Cet îlot comprend 14 gènes chez la souche 85-10 de *X. axonopodis* pv. *vesicatoria*, 10 chez la souche 306 de *X. axonopodis* pv. *citri* et 8 chez la souche ATCC33913 de *X. campestris* pv. *campestris* (Figure 13). D'autres gènes codant des MCPs sont également dispersés dans les génomes de ces bactéries. Selon Moreira *et al.* (2004), les copies de gènes codant les MCPs se sont d'abord dupliquées en tandem au niveau d'une même région et ensuite répandues dans des régions génomiques plus distantes. La majorité des gènes codant les MCPs se situent au niveau d'îlots

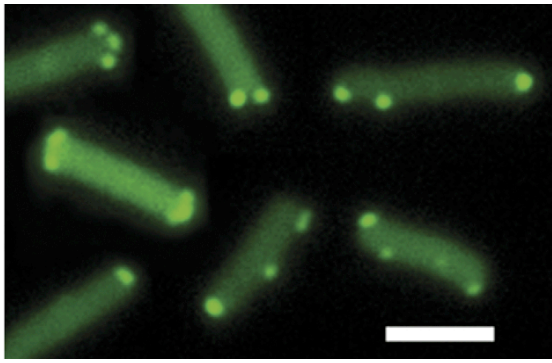


Figure 14. Ilots polaires et latéraux de chémorécepteurs chez *E. coli* (d'après Kentner et Sourjik, 2006). Les îlots de MCPs sont marqués par l'expression de la fluorescence de la protéine GFP fusionnée à la protéine cytoplasmique CheR. Les îlots polaires forment généralement un spot bien individualisé mais peuvent aussi apparaître sous la forme de deux ou trois spots. Quelques îlots sont latéraux et forment des spots bien distincts. La barre d'échelle représente 2 μm .

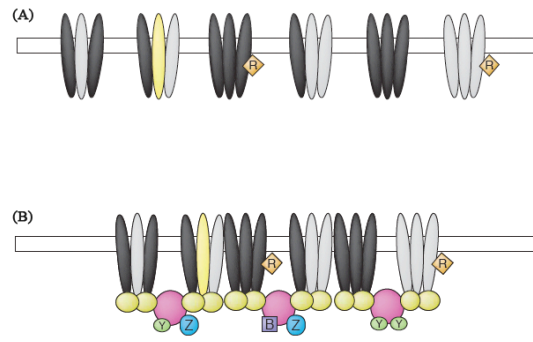


Figure 15. Modèle de la structuration d'un îlot de chémorécepteurs (d'après Sourjik, 2004). les chémorécepteurs (Tsr en noir ; Tar en gris ; Trg en jaune) sont organisés en trimères en absence de CheA et CheW et forment par conséquent, plusieurs spots polaires (A). CheA (en rose) et CheW (en jaune) interagissent avec les trimères de MCPs et forment ainsi, un seul spot individualisé (B).

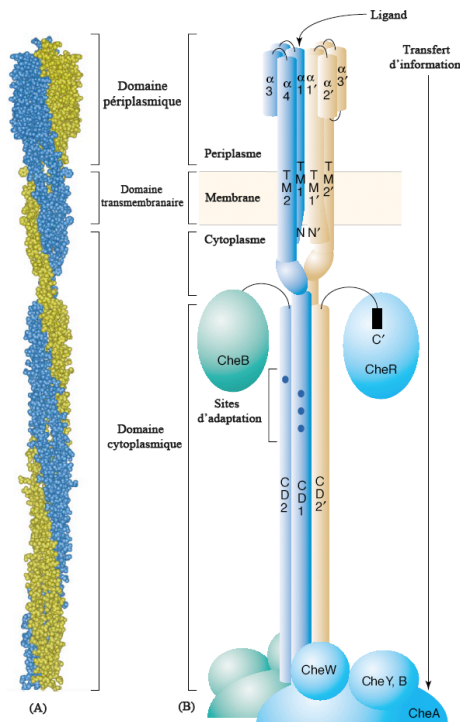
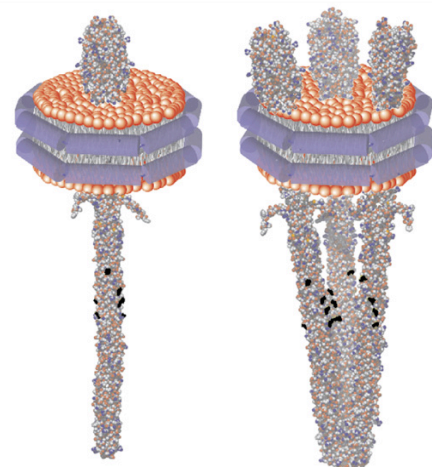


Figure 16. Structure d'un chémorécepteur dimérique (d'après Falke et Hazelbauer, 2001). Structure cristallographique d'un homodimère (A). Les deux unités symétriques de l'homodimère sont en bleu et jaune. Représentation schématique de la structure et des régions fonctionnelles de l'homodimère (B).



Fixation du ligand	yes	yes
Transduction du signal à travers la membrane	yes	yes
Changement de la conformation	yes	yes
Activation / control de l'autophosphorylation	no	yes

Figure 17. Activité fonctionnelle des nanodisques comportant un seul homodimère ou des trimères de chémorécepteurs (d'après Hazelbauer *et al.*, 2007). Les nanodisques sont solubles et sont formés d'une bicouche lipidique entourée d'un anneau amphiphile. Le nanodisque comportant un seul homodimère fixe bien le ligand ; il est bien méthylé et change de conformation mais est incapable de contrôler l'autophosphorylation de l'histidine kinase CheA contrairement aux nanodisques comportant des trimères d'homodimères.

génomiques consacrés au chimiotactisme et à la mobilité (Moreira *et al.*, 2004 ; Thieme *et al.*, 2005). Pour la souche 85-10 de *X. axonopodis* pv. *vesicatoria*, 14 gènes codant des MCPS sont regroupés dans une région de 147 kb. Cette organisation semble être plus ou moins conservée chez les différentes souches séquencées de *Xanthomonas* (Moreira *et al.*, 2004). Il est important de mentionner que la conservation en terme de séquences et d'organisation physique de groupes de gènes dans le génome de bactéries phylogénétiquement distantes n'est pas très fréquente (Lathe *et al.*, 2000). Ces auteurs suggèrent que la conservation des gènes et de leur ordre serait due au fait que les produits de ces gènes interagissent ensemble sur le plan fonctionnel.

La majorité des chémorécepteurs sont organisés en îlots polaires (Figure 14) (Hazelbauer *et al.*, 2007 ; Vladimirov et Sourjik, 2009). En effet, trois techniques différentes *i*) marquage avec des anticorps couplés à l'or colloïdal appliquée sur des cellules fixées et sectionnées, *ii*) fluorescence sur des cellules vivantes et *iii*) tomographie cryo-électronique sur des cellules intactes, ont pu détecter des îlots de chémorécepteurs et montrer que la majorité d'entre eux (environ 80%) se localisent au niveau du pôle de la cellule bactérienne (Maddock et Shapiro, 1993 ; Sourjik et Berg, 2000 ; Zhang *et al.*, 2007). Ces îlots sont mobiles au niveau de la membrane ; ils forment en effet un seul îlot en présence de CheA et de CheW mais en forment plusieurs en absence de ces protéines (Figure 15). Cette caractéristique permet aux MCPs de contrôler plusieurs histidines kinases (CheA) à la fois et surtout d'interagir ensemble et par conséquent, d'être hautement sensibles (Hazelbauer *et al.*, 2007).

Un MCP correspond à un homodimère de protéines et comporte trois domaines opérationnels (Figure 16) : un domaine périplasmique senseur, un domaine transmembranaire pour l'ancrage dans la membrane interne de la cellule bactérienne et un domaine cytoplasmique de signalisation organisé en quatre hélices antiparallèles enroulées (Kentner et Sourjik, 2006 ; Hazelbauer *et al.*, 2007 ; Vladimirov et Sourjik, 2009). En outre, il a été montré que ces dimères sont eux-mêmes organisés en trimères. Les résidus favorisant cette organisation sont souvent très conservés au sein de la famille des MCPs et sont même identiques chez les quatre MCPs identifiés chez *E. coli*. Cette structure trimérique est importante pour le fonctionnement des chémorécepteurs

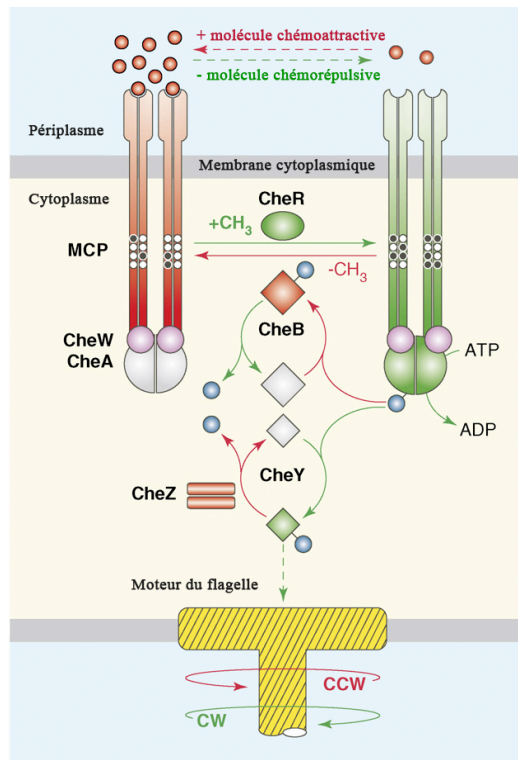


Figure 18. Voie de signalisation des chémorécepteurs chez *E. coli* (d'après Hazelbauer *et al.*, 2007). Les chémorécepteurs contrôlent la mobilité de la bactérie via un complexe ternaire MCP-CheA-CheW. Le fonctionnement du système chimiotactique est basé sur l'équilibre entre deux états conformationnels du complexe. L'état « kinase-on » stimule l'autophosphorylation de CheA alors que l'état « kinase-off » désactive l'autophosphorylation de CheA. L'activité nette reflète la proportion de complexes entre ces deux états. La présence de stimuli provoque un changement dans cet équilibre. Une augmentation de la concentration d'une molécule chémoattractive entraîne les chémorécepteurs dans un état « kinase-off ». Au contraire, une diminution de la concentration d'une molécule chémoattractive entraîne les chémorécepteurs dans un état « kinase-on ». Le changement de l'équilibre entre les états « kinase-on-kinase-off » module la phosphorylation de CheY (CheYp) et par conséquent, le contrôle du moteur du flagelle et la phosphorylation de CheB pour l'adaptation. Lorsque CheYp se lie au moteur, elle augmente la probabilité de rotation horaire (CW) et provoque un changement aléatoire du flagelle. Le comportement du flagelle par défaut correspond à la rotation antihoraire (CCW). La phosphatase CheZ hydrolyse CheYp entraînant la rotation horaire du flagelle.

puisque'elle constitue l'unité minimum de fonctionnement. En effet, l'étude de nanodisques comprenant un dimère isolé a montré que ce dernier est bien méthylé, qu'il change de conformation lors de la transduction du signal mais qu'il est incapable d'activer l'autophosphorylation de l'histidine kinase CheA contrairement aux nanodisques comportant des trimères de dimères (Figure 17) (Boldog *et al.*, 2006). Amin et Hazelbauer (2010) ont également étudié la fixation du ligand, le niveau de méthylation et la modification de la conformation au niveau d'un seul dimère et de dimères disposés ensemble sur des nanodisques. Ils ont montré que la fixation du ligand, le niveau de méthylation et la modification de la conformation se produisent avec la même sensibilité et intensité aussi bien au niveau d'un dimère isolé que des dimères disposés ensemble. Amin et Hazelbauer (2010) en ont conclu que le dimère reste l'unité fondamentale de la signalisation transmembranaire et que l'interaction au niveau des trimères est une étape ultérieure dans la signalisation. Par ailleurs, les MCPs peuvent former des clusters homogènes ou hétérogènes. Chez *E. coli*, les différents MCPs (Tar, Tcp, Trg, Tap) peuvent s'associer *in vivo* et former des trimères formés de différents dimères ce qui favorise la coopération entre les différents chémorécepteurs (Hazelbauer *et al.*, 2007 ; Vladimirov et Sourjik, 2009).

2.1.2 Transduction du signal chimiotactique et mobilité

La fréquence de rotation du flagelle est contrôlée par le chimiotactisme (Figure 18). En effet, le changement du sens de rotation du flagelle est induit par la phosphorylation de la protéine CheY (CheYp). Celle-ci se fixe au moteur du flagelle au niveau de la protéine FliM et change le sens de sa rotation. L'activité d'autophosphorylation de l'histidine kinase CheA est régulée par la fixation du ligand sur le chémorécepteur ; elle est inhibée en présence de molécules chémoattractives et activée en absence de molécules chémoattractives. L'histidine kinase CheA liée aux chémorécepteurs par l'intermédiaire d'un adaptateur CheW, transfère son groupement phosphore à CheY. CheYp diffuse ensuite librement dans le cytoplasme puis interagit avec le moteur du flagelle en modifiant son sens de rotation d'antihoraire à horaire. Chaque chémorécepteur peut être en conformation soit active soit inactive et ceci en fonction de sa fixation avec le ligand (chémoattractif ou chémorépulsif) et de son niveau

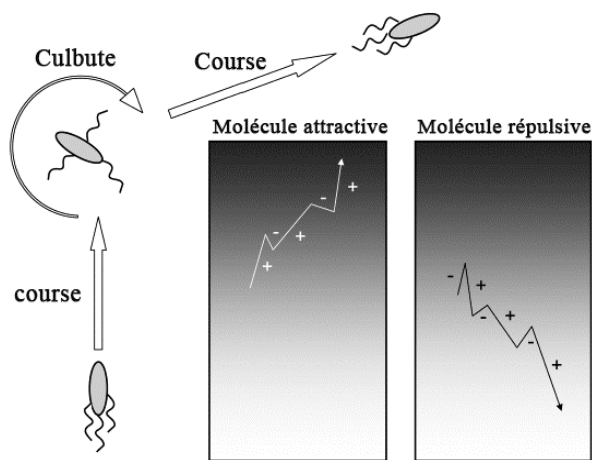


Figure 19. Chimiotactisme chez *E. coli* (d'après Vladimirov et Sourjik, 2009). L'hypothèse est que le choix de la direction d'une nouvelle course est aléatoire. Lorsque la bactérie rencontre des concentrations croissantes de molécules attractives, la course devient longue et rectiligne. Les courses rectilignes sont indiquées par le signe '+' et une course aléatoire normale, par un signe '-'

de méthylation. Les chémorécepteurs actifs stimulent l'autophosphorylation de CheA et les chémorécepteurs inactifs l'inhibent. Les MCPs sont méthylés par l'enzyme méthyltransférase CheR et déméthylés par l'enzyme méthylstérerase CheB (Szurmant et Ordal, 2004 ; Wadhams et Armitage, 2004 ; Kentner et Sourjik, 2006 ; Vladimirov et Sourjik, 2009).

Le déplacement des bactéries dans leur milieu est assuré par le changement du sens de la rotation du flagelle (Figures 18 et 19). Dans le cas d'*E. coli*, la rotation antihoraire du moteur (counterclockwise, CCW) arrange les flagelles en un seul faisceau ce qui pousse la cellule bactérienne dans une longue course rectiligne. Au contraire, la rotation horaire du flagelle (clockwise, CW) entraîne la désolidarisation des flagelles les uns des autres et la bactérie culbute. Celle-ci peut ensuite reprendre une course rectiligne vers une nouvelle direction (Belas *et al.*, 2008). Récemment, il a été démontré que le changement du sens de rotation du flagelle (entre CCW et CW) peut être partiel. Cet état intermédiaire entraîne des mouvements de retournement transitoire du flagelle sur lui même. De plus, le changement du sens de rotation du flagelle peut se faire plus ou moins rapidement (Bai *et al.*, 2010).

2.1.3 Temps d'adaptation de la réponse chimiotactique

L'adaptation à des stimulations continues et variables est contrôlée par deux enzymes : CheB, une méthylstérerase et CheR, une méthyltransférase (Kentner et Sourjik, 2006). Dans le cas de l'absence de molécules chémoattractives, les chémorécepteurs sont actifs ; ils induisent l'autophosphorylation de CheA ce qui augmente la concentration de CheYp. La protéine CheB également phosphorylée par CheA, entre en compétition avec la protéine CheR. Ceci entraîne l'accroissement de l'activité méthylstérerase de CheB et augmente ainsi l'état non méthylé des chémorécepteurs. Les MCPs non méthylés ont une faible capacité à induire l'autophosphorylation de CheA (même en cas de l'absence de molécules attractives). Ainsi, le rythme d'autophosphorylation de CheA et par conséquent, le rythme de changement du sens de rotation du flagelle retrouvent leurs états initiaux. Dans le cas contraire, l'augmentation de la concentration en molécule chémoattractrice rend les

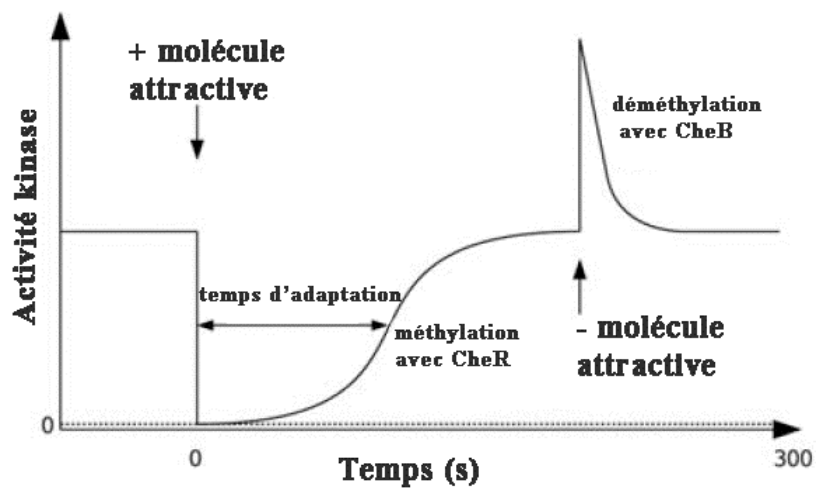


Figure 20. Temps d'adaptation de la réponse chimiotactique (d'après Vladimirov et Sourjik, 2009). L'addition d'une quantité saturante de molécule attractive provoque une inhibition très rapide (0,1 s) de l'activité kinase de CheA. Cette inhibition est suivie par une adaptation lente dépendant de l'activité méthylestérase CheR. Le temps d'adaptation est proportionnel au taux d'occupation des MCPs par le ligand. La suppression de la molécule attractive provoque une augmentation rapide de l'activité kinase. Cette augmentation est suivie par une adaptation dépendant de l'activité méthyltransférase CheB.

chémorecepteurs inactifs ce qui inhibe l'autophosphorylation de CheA et réduit en conséquence la concentration de CheYp et la fréquence de rotation du moteur du flagelle. La phosphorylation de CheB et par conséquent, son activité méthylestérase sont également inhibées. Ainsi, l'activité constitutive de la protéine méthyltransférase CheR augmente l'état méthylé des MCPs qui sont aptes à stimuler l'autophosphorylation de CheA (même en présence de la molécule attractive). Ceci permet au final, le retour du rythme d'autophosphorylation de CheA et la fréquence de changement de rotation du flagelle à leurs états initiaux de pré-stimulation (Wadhams et Armitage, 2004 ; Kentner et Sourjik, 2006). Le niveau de méthylation des chémorecepteurs joue le rôle d'une mémoire chimique permettant à la cellule bactérienne de comparer les concentrations actuelles de ligand avec celles rencontrées par la bactérie dans le passé proche. En effet, les réactions de méthylation et de déméthylation se produisent plus lentement que les réactions d'autophosphorylation impliquées dans cette voie de signalisation (Figure 20). L'intervalle de temps compris entre ces réactions correspond au temps d'adaptation pendant lequel la bactérie peut encore se rappeler de son état passé et comparer son état présent à celui du passé (Figure 20). Ainsi, le chimiotactisme chez les bactéries représente probablement le système comportemental le plus simple (Koshland, 1980).

2.1.4 Amplification du signal externe

Sourjik et Berg (2002) ont mesuré la sensibilité de l'amplification au début de la cascade de transduction du signal chimiotactique. Ils ont montré que la majorité de cette amplification se fait au niveau du complexe de signalisation formé par le chémorecepteur et les protéines Che qui lui sont associées. Ainsi, l'augmentation de la concentration en molécules chémoattractives provoquant l'inactivation d'un seul chémorecepteur revient à amplifier 36 fois le signal environnemental. Ceci indique qu'un MCP seul peut contrôler environ trois douzaines de protéines kinases CheA (Sourjik et Berg, 2002). En plus, la présence de différents types de MCPs au sein des trimères et des îlots peut modifier la coopération entre ces senseurs, leur sensibilité aux différents stimuli et le niveau net de l'activité d'autophosphorylation de la kinase CheA (Figure 15) (Hazelbauer *et al.*, 2007).

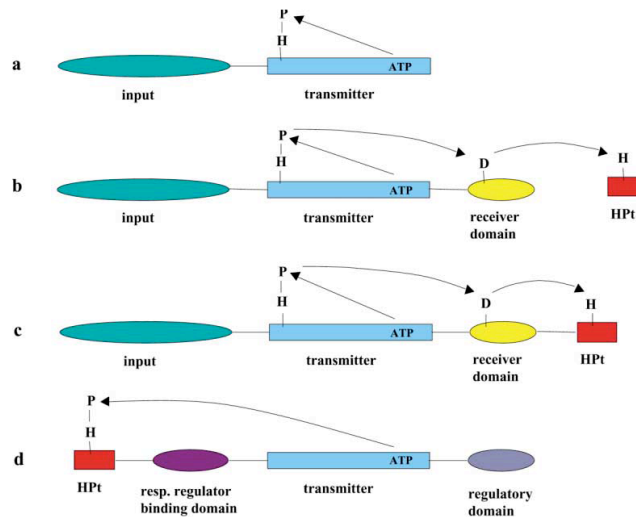


Figure 21. Représentation schématique de l'organisation modulaire des histidines kinases (d'après Foussard *et al.*, 2001). Les senseurs possèdent un domaine dit « *input* » impliqué dans la détection des signaux (a, b et c). Le domaine dit « *transmitter* » est impliqué dans l'émission du signal vers l'intérieur de la cellule. Généralement, il est conservé entre tous les senseurs des systèmes régulateurs à deux composants (a, b, c et d). L'histidine kinase CheA impliquée dans le chimiotactisme ne possède pas de domaine « *input* » ; elle est couplée au chémorécepteur membranaire (d). Les senseurs et les régulateurs peuvent fusionner et former des histidines kinases hybrides (b et c). Dans ce cas, ils possèdent un module de réception du signal et parfois un domaine transmetteur de groupement phosphate (c).

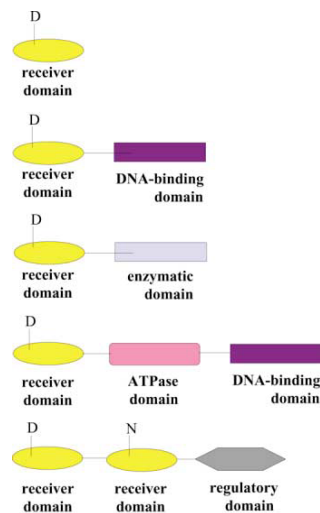


Figure 22. Représentation schématique de l'organisation des régulateurs (d'après Foussard *et al.*, 2001). Ces protéines contiennent toutes un domaine un module de réception du signal. Le deuxième domaine peut être constitué d'un ou plusieurs modules ayant des fonctions variables telles que la fixation à l'ADN, une activité enzymatique ou régulatrice.

2.2 Autres systèmes de perception et de transduction de signaux

2.2.1 Les systèmes de régulation à deux composants

Le système de régulation à deux composants comme son nom l'indique, comporte classiquement deux protéines : une histidine kinase jouant le rôle d'un senseur transmembranaire et un régulateur cytoplasmique. La transduction du signal par le système de régulation à deux composants se base sur une cascade de réactions de phosphorylation comme c'est d'ailleurs le cas de la voie de signalisation chimiotactique (Galperin, 2004 ; Qian *et al.*, 2008 ; Gooderham et Hancock, 2009 ; Whitworth et Cock, 2009). Plusieurs fonctions importantes comme la mobilité et celles liées au pouvoir pathogène des bactéries, sont régulées par des systèmes de régulation à deux composants. Chez les *Xanthomonas* spp., les principaux systèmes régulateurs étudiés sont RpfC/RpfG, RavS/RavR et ColS/ColR mais aussi le couple régulateur HrpG/HrpX (Qian *et al.*, 2008 ; Whitworth et Cock, 2009).

2.2.1.a. Les senseurs des systèmes de régulation à deux composants

Les senseurs histidines kinases sont typiquement organisés en homodimères comportant deux domaines fonctionnels. Le premier domaine en position N-terminale est extracytoplasmique et est proprement impliqué dans la détection des signaux externes (Figure 21a). La séquence protéique de ce domaine senseur peut varier en longueur ou en composition en acides aminés ce qui lui confère une spécificité de perception de différents stimuli environnementaux. Le second domaine en position C-terminale est cytoplasmique ; il s'agit d'un domaine émetteur du signal externe vers l'intérieur de la cellule et sa séquence protéique est généralement très conservée (Figure 21) (Galperin, 2004 ; Galperin et Gomelsky, 2005 ; Lavín *et al.*, 2007 ; Qian *et al.*, 2008 ; Whitworth et Cock, 2009).

Les senseurs des systèmes de régulation à deux composants sont généralement constitués d'histidine kinase. Néanmoins, certaines histidines kinases ne détectent pas

par elles mêmes les signaux externes ; elles dépendent des interactions avec d'autres senseurs transmembranaires. Ceci est le cas par exemple, de l'histidine kinase CheA (Figure 21d) du système de signalisation chimiotactique qui nécessite l'interaction avec plusieurs MCPs (*cf.* paragraphe 2.1.2) (Wadhams et Armitage, 2004 ; Galperin et Gomelsky, 2005). Par ailleurs, de nombreuses histidines kinases possèdent plusieurs domaines senseurs comme les domaines PAS et GAF (Galperin et Gomelsky, 2005). Certains gènes codant des senseurs et des régulateurs peuvent fusionner pour former des histidines kinases hybrides *i.e* des senseurs couplés à leurs régulateurs (Figure 21b et 21c) (Galperin et Gomelsky, 2005 ; Lavín *et al.*, 2007 ; Whitworth et Cock, 2009).

2.2.1.b. Les régulateurs des systèmes de régulation à deux composants

Les régulateurs cytoplasmiques contiennent deux domaines fonctionnels. Le domaine en position N-terminale assure la réception du signal émis par l'histidine kinase via une réaction de phosphorylation (Figure 22). En position C-terminale, il y a le second domaine qui peut être constitué d'un ou plusieurs modules variables (Figure 22). A titre d'exemple, le régulateur VieA de *Vibrio cholerae* contient un module EAL (impliqué dans la dégradation de messenger secondaire di-GMPc) et un domaine de fixation à l'ADN. La modulation de l'état phosphorylé du régulateur cytoplasmique contrôle l'expression de gènes cibles ou le comportement de la cellule bactérienne (Galperin, 2004 ; Galperin et Gomelsky, 2005 ; Lavín *et al.*, 2007 ; Qian *et al.*, 2008 ; Whitworth et Cock, 2009).

2.2.1.c. Mesure du potentiel d'adaptation des bactéries

Le nombre total de protéines impliquées dans la signalisation et présentes dans le génome d'une bactérie peut constituer une mesure du potentiel d'adaptation d'une bactérie à différents environnements. En d'autres termes, il s'agirait du quotient d'intelligence (**QI**) des bactéries (Galperin, 2005 ; Qian *et al.*, 2008). Galperin (2005) exprime le QI d'une bactérie selon la formule suivante :

$$\mathbf{QI} = \mathbf{5 \times 10^4 (n-5)^{1/2} \times L^{-1}} \quad \mathbf{[1]}$$

n étant le nombre total de protéines impliquées dans la transduction de signaux. **L** étant la taille du génome complet en kb additionnée éventuellement, de celle des plasmides. Les valeurs de **5 x 10⁴** et de **5** sont des coefficients empiriques choisis arbitrairement. Un QI = 100 est relatif à la présence de 9 protéines de transduction dans un génome de 1000 kb ou à la présence de 105

protéines de transduction dans un génome de 5000 kb. Aucune valeur de QI n'est définie pour un génome comportant moins de 6 protéines de transduction (Galperin, 2005).

D'autre part, le ratio du nombre total de récepteurs transmembranaires sur le nombre total de senseurs intracellulaires peut être employé pour définir si l'organisme est extraverti *i.e* capable de percevoir diligemment les signaux environnementaux, ou au contraire introverti comme c'est le cas des bactéries autotrophes, les cyanobactéries photosynthétiques par exemple (Galperin, 2005).

Les gènes codant les différents composants du système de régulation (senseurs et régulateurs) sont généralement organisés en opéron (Lavín *et al.*, 2007). Dans le génome de trois souches appartenant à différents pathovars de *P. syringae* (pvs. *tomato*, *syringae* et *phaseolicola*), 10 à 12% des gènes sont impliqués dans la régulation (Feil *et al.*, 2005). Ceci reflète la nécessité qu'ont les bactéries de s'adapter rapidement à divers environnements rencontrés lors de leur multiplication dans la phyllosphère, la colonisation de la plante hôte et éventuellement lors du processus infectieux (Lavín *et al.*, 2007). Dans le cas des *Xanthomonas*, approximativement 92 à 121 gènes codent des protéines du système de régulation à deux composants. Le nombre total de nucléotides des séquences des gènes codant ces protéines représentent 2,38 à 3,24 % des nucléotides totaux du chromosome (Qian *et al.*, 2008). Ce résultat est en accord avec le style de vie des bactéries du genre *Xanthomonas*. En effet, elles peuvent être confrontées à des niches écologiques très diverses et doivent par conséquent s'adapter à des environnements très hétérogènes. Le terme *niche écologique* se réfère dans le cas des *Xanthomonas* (*cf.* paragraphe 4.2.) à la gamme de plantes hôtes symptomatiques et asymptomatiques et aux différents compartiments d'une plante (le parenchyme, le xylème et la surface foliaire) que colonise la bactérie. Ainsi, les *Xanthomonas* sont équipées de systèmes de signalisation cellulaire leur permettant de percevoir différents stimuli environnementaux et de répondre de façon adéquate via la régulation de l'expression des gènes impliqués dans le métabolisme et éventuellement dans le pouvoir pathogène des bactéries (Qian *et al.*, 2008).

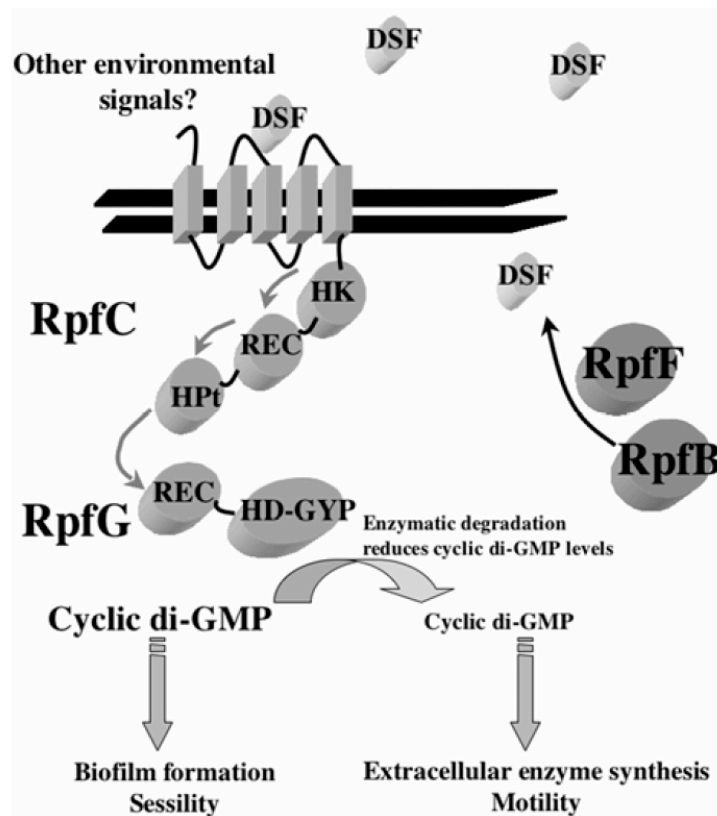


Figure 23. Représentation schématique de la perception de DSF et de la transduction du signal chez *X. campestris* (d'après Dow *et al.*, 2006). La synthèse du signal DSF se fait par RpfF et dépend de RpfB. Le signal DSF est détecté par le système de régulation à deux composants RpfC/RpfG. Le régulateur RpfG est une phosphodiésterase qui hydrolyse le di-GMPc qui est un messenger secondaire. RpfC détecte d'autres signaux environnementaux. Les rectangles représentent les hélices transmembranaires, REC est un domaine CheY-like, HPT est un domaine transmetteur de groupement phosphate et HK, un domaine histidine kinase.

2.2.2 Quelques cas particuliers de systèmes de régulation à deux composants étudiés chez les *Xanthomonas* spp.

2.2.3.a. le système *Rpf* et sa cascade de régulation

Les bactéries produisent des molécules signal qui s'accumulent dans le milieu avec l'augmentation de la population et régulent l'expression des gènes via des récepteurs protéiques correspondants. Les N-acyl-homosérine lactones (N-AHLs) sont les molécules signal le plus souvent synthétisées par les bactéries à Gram négatif comme *Erwinia* et *Pseudomonas* (Atkinson et Williams, 2009). Par contre, les *Xanthomonas* utilisent des facteurs diffusibles de nature chimique différente, DF (diffusible factor, butyrolactone) et DSF (diffusible signaling factor, acide cis-11-méthyl-2-dodecenoïque). Ce mécanisme taille de population-dépendant par lequel les bactéries communiquent de cellule à cellule est nommé quorum sensing (QS). Il leur permet de réguler et de coordonner par l'intermédiaire d'un régulateur transcriptionnel plusieurs fonctions cellulaires comme la mobilité, la formation de biofilms, la synthèse d'exopolysaccharides (EPS), la production d'enzymes extracellulaires, le transfert horizontal d'ADN et la résistance aux rayons UV (Diggle *et al.*, 2007). Par exemple, *E. carotovora* utilise la molécule N-AHL pour initier un processus infectieux dès que la population bactérienne dépasse le seuil critique. Cette stratégie permet à la bactérie de surmonter la résistance de l'hôte (Brencic et Winans, 2005). Le QS peut se produire au sein d'une population telle que les agrégats bactériens en biofilms, entre des membres appartenant à une même espèce ou à des espèces différentes (Lindow et Brandl, 2003).

Dans le cas des *Xanthomonas*, DF est impliqué dans la régulation de la production de la xanthomonadine et des EPS. Le mutant *pigB* non pigmenté de *X. campestris* pv. *campestris* ne synthétise pas le facteur diffusible DF. Il est de plus moins agressif, produit moins d'EPS et sa survie épiphyte est affectée. La production de xanthomonadine et d'EPS peut être restaurée par l'ajout de DF exogène (He et Zhang, 2008). Les *Xanthomonas* possèdent un deuxième système de QS avec une autre molécule diffusible, DSF. DSF est supposé pouvoir diffuser librement à travers les membranes bactériennes grâce à sa nature lipophile et s'accumule au cours de la phase stationnaire de croissance dans le milieu externe (Figure 23). Le signal DSF est détecté

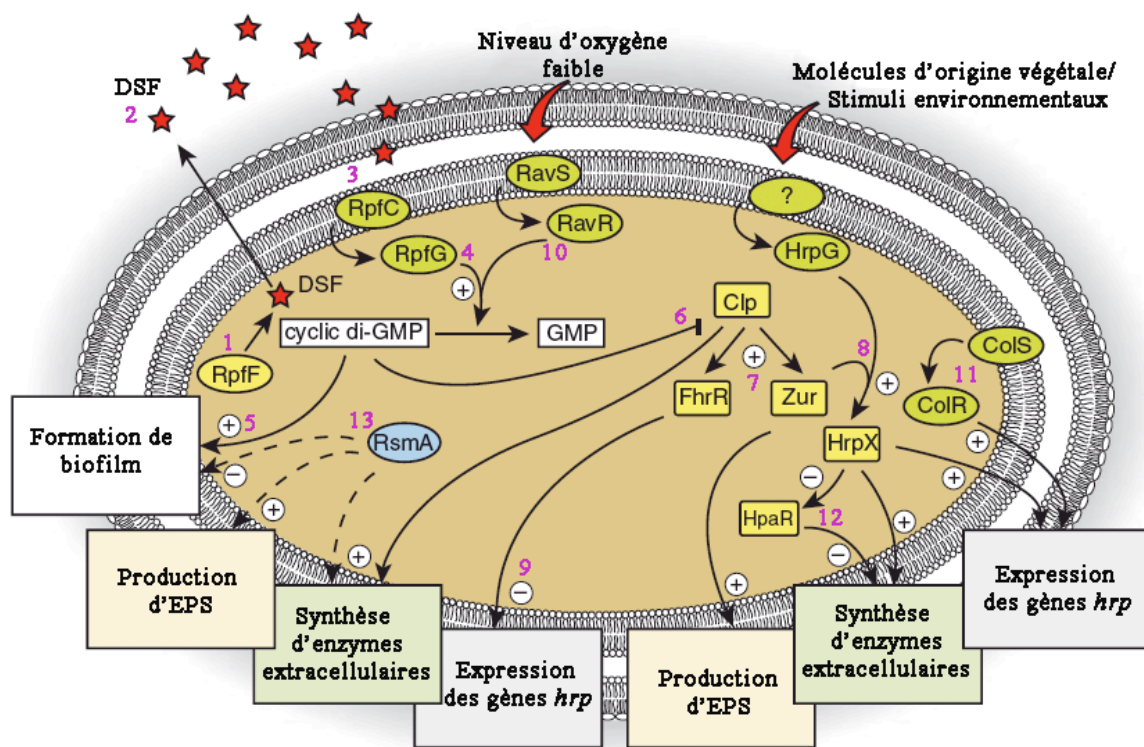


Figure 24. Représentation schématique des voies de régulation contrôlant l'expression des gènes impliqués dans le pouvoir pathogène de *X. campestris* pv. *campestris* (d'après Büttner et Bonas, 2010). 1. La protéine RpfF qui est une hydratase enoyl-CoA putative, est impliquée dans la synthèse de DSF, une molécule diffusible du QS 2, qui s'accumule durant la phase stationnaire de croissance dans le milieu extracellulaire. 3. DSF est détecté par le système de régulation à deux composants RpfC/RpfG. 4. Le régulateur RpfG est une phosphodiesterase impliquée dans l'hydrolyse d'un messenger secondaire intracellulaire, le di-GMP cyclique. 5. Un niveau faible du di-GMPc active l'expression des gènes impliqués dans le pouvoir pathogène alors qu'un niveau élevé du di-GMPc favorise la formation de biofilms et 6. réprime la fixation du régulateur transcriptionnel Clp aux promoteurs des gènes cibles. Clp active l'expression des gènes codant les enzymes extracellulaires, les composants du SST2 et SST3, des gènes impliqués dans la synthèse des EPS et 7. l'expression des gènes régulateurs *fhrR* et *zur*. 8. L'activateur transcriptionnel Zur et le régulateur HrpG activent l'expression du gène *hrpX* qui code un régulateur contrôlant l'expression des gènes *hrp*. 9. Le régulateur FhrR régule l'expression des gènes codant le flagelle et les gènes *hrp*. 10. Un autre système de signalisation à deux composants RavS/RavR contribue à l'hydrolyse du di-GMPc. Le senseur RavS possède deux domaines PAS et est probablement activé par de faibles niveaux d'oxygène. 11. L'expression des gènes des opérons *hrpC* et *hrpE* est contrôlée en plus de HrpG et HrpX par le système de régulation à deux composants ColS/ColR. Le couple ColS/ColR est également impliqué dans l'adaptation aux stress environnementaux. 12. HrpX régule l'expression d'un autre gène *hpaR* codant un régulateur transcriptionnel qui inhibe la synthèse de protéase extracellulaire. 13. Le régulateur post-transcriptionnel RsmA a été récemment rapporté chez *X. campestris* pv. *campestris*. RsmA active la synthèse d'EPS, d'endoglucanases et d'amylases extracellulaires et inhibe la formation de biofilms.

et transduit par un système de régulation à deux composants. Ce système comprend un senseur histidine kinase RpfC et un régulateur RpfG. De plus, il a été montré que la signalisation *rpf*-dépendante est nécessaire à la réouverture des stomates de l'hôte par *X. campestris* pv. *campestris* ; la fermeture des stomates étant une réaction de défense de la plante hôte contre les invasions des agents pathogènes (Gudesblat *et al.*, 2009).

Le régulateur RpfG est impliqué dans l'hydrolyse du di-GMP cyclique (di-GMPc) intracellulaire. Ce dernier est un messager secondaire régulant un panel de fonctions cellulaires comme la formation de biofilms, la mobilité et l'expression de gènes impliqués dans le pouvoir pathogène des *Xanthomonas*. Il a été montré que des niveaux élevés de di-GMPc favorisent la formation de biofilms alors que des niveaux faibles de di-GMPc activent la mobilité et l'expression de gènes impliqués dans le pouvoir pathogène (Figures 23 et 24) (Ryan *et al.*, 2007).

Le signal DSF et le système de régulation RpfC/RpfG activent le régulateur transcriptionnel Clp (CRP [cAMP receptor protein]-like protein) (Figure 24). Ce régulateur global contient des nucléotides et un domaine de fixation à l'ADN lui permettant de se lier aux promoteurs des gènes cibles tels que les gènes codant une pectate lyase et une endoglycanase chez *X. campestris* pv. *campestris* (He *et al.*, 2007 ; Hsiao *et al.*, 2009). Dans le cas de *X. campestris* pv. *campestris*, le régulateur global Clp active l'expression des gènes codant les enzymes extracellulaires, les composants du système de sécrétion de type 2 et de type 3 et les gènes impliqués dans la synthèse des EPS. En revanche, le régulateur Clp n'est pas impliqué dans la régulation de la formation de biofilms qui est DSF-dépendante (He *et al.*, 2007). Il a été de plus montré que chez *X. campestris* pv. *campestris*, Clp active l'expression de gènes *fhrR* et *zur* codant d'autres régulateurs (Figure 24). Le régulateur FhrR régule l'expression de gènes codant le flagelle et le système de sécrétion de type 3 alors que le régulateur Zur est impliqué dans le contrôle de l'assimilation du fer et la résistance aux antibiotiques (He *et al.*, 2007). Le régulateur Zur de la souche 8004 de *X. campestris* pv. *campestris*, active l'expression des gènes *hrp* via le régulateur transcriptionnel HrpX (Huang *et al.*, 2009). Par contre, les gènes *hrp* de la souche XC1 de *X. campestris* pv. *campestris* semblent être réprimés par Clp et FhrR (Figure 24) (He *et al.*, 2007). Ceci montre que

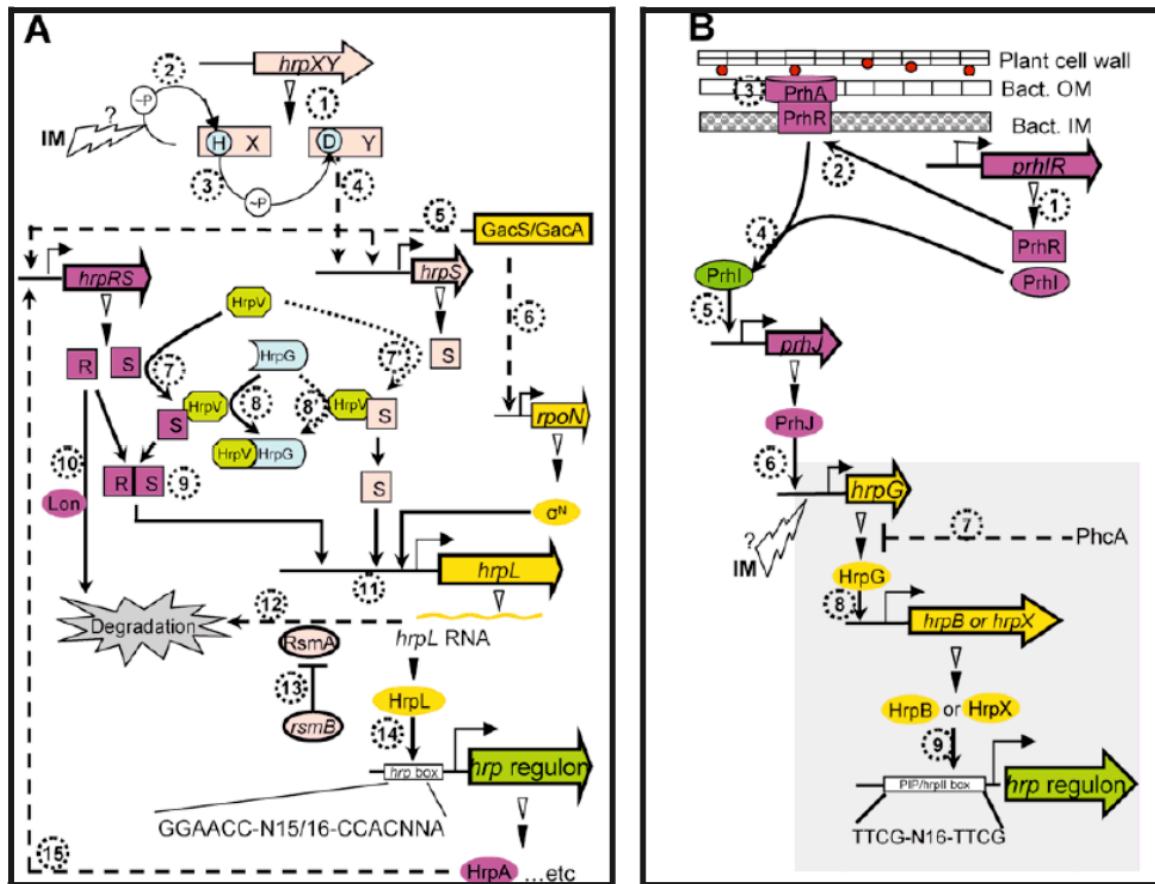


Figure 25. Régulation des gènes *hrp* chez les bactéries phytopathogènes (d'après Tang *et al.*, 2006). **A.** Bactéries du groupe I. 1. Chez *Erwinia* spp. l'opéron *hrpXY* est transcrit et traduit en HrpX et HrpY. 2. Un signal externe provoque l'autophosphorylation de HrpX. 3. HrpX transfère son groupement phosphate à HrpY. HrpY active la transcription de *hrpS* grâce à un mécanisme encore inconnu. 5. Le système de régulation à deux composants GacS/GacA active la transcription de *hrpS* chez *Erwinia* spp. et de *hrpRS* chez *P. syringae*. 6. Le couple GacS/GacA active la transcription de *rpoN* chez *Erwinia* spp. et *P. syringae*. 7. et 7'. HrpS est réprimé par HrpV via une interaction protéine-protéine. 8. et 8'. HrpG interagit avec HrpV et réactive HrpS. 9. Chez *P. syringae*, HrpS et HrpR forment un hétérodimère. 10. Chez *P. syringae*, HrpR est dégradée par la protéase Lon. 11. HrpR et HrpS chez *P. syringae* et HrpS chez *Erwinia* spp. s'associent avec RpoN au niveau du promoteur du gène *hrpL* pour activer son expression. 12. L'ARN issu de la transcription de *hrpL* peut être dégradé par le régulateur post-transcriptionnel RsmA. 13. L'ARN *rsmB* inhibe la fonction de RsmA. 14. HrpL reconnaît la boîte *hrp* et active la transcription du régulateur *hrp*. 15. HrpA agit en amont sur la transcription de *hrpRS* pour stimuler l'expression des gènes *hrp*. **B.** Bactéries du groupe II. 1. L'opéron *prhIR* est transcrit en PrhI et PrhR. 2. PrhR est localisée dans la membrane bactérienne inférieure. 3. PrhA est localisée dans la membrane bactérienne supérieure ; elle perçoit un signal d'origine végétale et transduit le signal à PrhR. 4. PrhR active PrhI. 5. PrhI active la transcription de l'opéron *prhJ*. 6. PrhJ active la transcription de *hrpG*. 7. Le régulateur post-transcriptionnel PhcA réprime *hrpG*. 8. HrpG active la transcription de *hrpX* chez les *Xanthomonas* spp. et de *hrpB* chez *R. solanacearum*. 9. HrpX chez *Xanthomonas* spp. et HrpB chez *R. solanacearum* reconnaissent la boîte PIP et activent les promoteurs des gènes *hrp*.

de multiples cascades de signalisation interviennent dans la régulation des gènes impliqués dans le pouvoir pathogène et que la régulation des gènes DSF-dépendante varierait parmi des différentes souches appartenant à une même espèce voire à un même pathovar (Büttner et Bonas, 2010).

Chez les *Xanthomonas*, un autre système de régulation à deux composants contribue à l'hydrolyse du messenger secondaire, le di-GMPc. Ce système se compose d'un senseur histidine kinase RavS et d'un régulateur RavR (Figure 24). Le couple RavS/RavR régule l'expression de plus de 206 gènes impliqués dans le pouvoir pathogène tels que les gènes codant les enzymes extracellulaires, les EPS, les gènes impliqués dans le chimiotactisme et la mobilité, les gènes du système de sécrétion de type 3, *etc* (He *et al.*, 2009). Cette régulation se fait via le régulateur global Clp (He *et al.*, 2009). La protéine RavR est impliquée dans l'hydrolyse du di-GMPc. Le senseur histidine kinase RavS possède quant à lui, deux domaines PAS qui sont également présents chez de nombreux senseurs d'oxygène de bactéries à Gram négatif (He *et al.*, 2009). Le second domaine PAS est nécessaire au fonctionnement de la protéine RavS suggérant que le couple RavS/RavR régule la production de facteurs du pouvoir pathogène en réponse à de faibles taux d'oxygène (He *et al.*, 2009).

2.2.3.c. La régulation de l'expression des gènes codant le SST3

Les gènes *hrp* codant le système de sécrétion de type 3 (SST3) ne sont pas constitutivement exprimés. En effet, leur expression est activée lorsque la bactérie entre dans la plante ou est mise en culture dans un milieu inducteur (Büttner et Bonas, 2010). En fonction du degré de conservation et des similarités dans l'organisation et la régulation des gènes *hrp*, les bactéries phytopathogènes à Gram négatif ont été classées en deux groupes. Ainsi, *E. amylovora* et *P. syringae* appartiennent au groupe I alors que les *Xanthomonas* et *Ralstonia solanacearum* font partie du groupe II (Figure 25) (Alfano et Colmer, 1997 ; Cornelis, 2006).

La régulation des gènes *hrp* des membres du groupe II comme les *Xanthomonas*, est contrôlée par le régulateur global HrpG. Ce dernier est membre de la famille OmpR et perçoit les signaux environnementaux via un TBSDT transducteur (PrhA)

(cf. paragraphe 2.2.3) actuellement identifié chez *R. solanacearum*, *X. oryzae* pv. *oryzae* et *X. axonopodis* pv. *citri* (Marenda *et al.*, 1998 ; Zou *et al.*, 2006 ; Aini *et al.*, 2010). Le gène *hrpG* est légèrement induit à la surface des plantes. En revanche, une induction très forte est observée lorsque la bactérie entre dans l'apoplaste de sa plante hôte (Boureau *et al.*, 2006 ; Zhang *et al.*, 2009). Le régulateur global HrpG active l'expression du gène *hrpX* codant un activateur transcriptionnel de type AraC. HrpX se fixe au niveau d'une séquence conservée, identifiée dans les régions promotrices (TTCGC-N₁₅-TTCGC) des gènes *hrp* cibles et appelée boîte PIP (plant-inductible promoter) (Wengelnik *et al.*, 1999 ; Koebnik *et al.*, 2006 ; Büttner et Bonas, 2010). Cependant, l'activation de l'expression de certains gènes par HrpX peut se faire sans la présence de la boîte PIP. Ceci suggère que la boîte PIP n'est donc pas essentielle à l'activation des gènes cibles par HrpX. C'est en effet le cas des promoteurs des gènes codant des ET3s, XopB, XopC1 et XopJ1, qui sont dépourvus de boîtes PIP et qui sont régulés par HrpX (Noël *et al.*, 2001). Par contre, le gène codant l'effecteur XopJ3 (anciennement nommé AvrRxv) qui possède une boîte PIP au niveau de son promoteur, est exprimé de façon constitutive (Ciesiolka *et al.*, 1999 ; White *et al.*, 2009). L'étude des mutants de *X. axonopodis* pv. *phaseoli* révèle que HrpG et HrpX sont essentiels au pouvoir pathogène et contribuent considérablement à la survie épiphyte de la bactérie. La survie épiphyte de *X. axonopodis* pv. *phaseoli* sur le haricot est plus affectée suite la mutation de *hrpG/hrpX* que suite à la mutation du gène codant le pilus du SST3, HrpE ou d'autres gènes codant la machinerie du SST3 (Darsonval *et al.*, 2008).

L'expression des gènes appartenant aux opérons *hrpC* et *hrpE* du cluster *hrp* de *X. campestris* pv. *campestris* est contrôlée par HrpG et HrpX mais également par un autre système de signalisation à deux composantes ColR/ColS (Figure 24). Ceci suggère que de nombreuses voies de transduction de signaux sont impliquées dans la régulation de l'expression des gènes *hrp* et que les opérons *hrp* peuvent être individuellement la cible de voies alternatives de signalisation (Zhang *et al.*, 2008). Le couple ColR/ColS est également impliqué dans l'adaptation aux stress environnementaux. En effet, les mutants de la souche 8004 de *X. campestris* pv. *campestris* affectés dans les gènes *colR* et *colS* présentent face à cinq agents stressants (ampicilline, rifampicine, NaCl, CdSO₄ et phénol) des CMI significativement inférieures à celle de la souche sauvage (Zhang

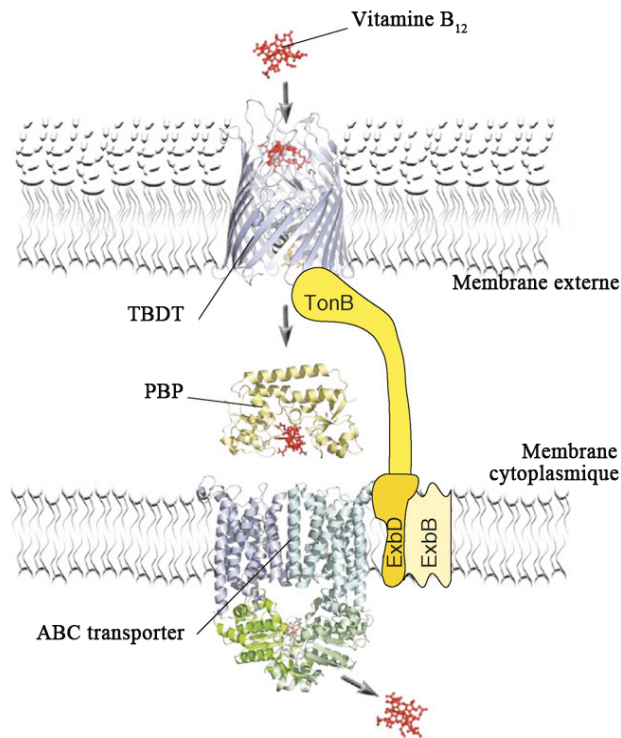


Figure 26. Représentation du modèle de TBDT spécifique de la vitamine B₁₂ (d'après Schauer *et al.*, 2008). Chez les bactéries à Gram-négatif, les TBDT transportent des molécules à travers la membrane externe. La vitamine B₁₂ en rouge, se lie spécifiquement au TBDT (BtuB) en bleu. Le complexe TonB-ExbB-ExbD en jaune, exploite la force proton motrice de la membrane cytoplasmique pour induire un changement de conformation du TBDT. Ceci provoque le relargage de la vitamine B₁₂ dans l'espace périplasmique. Dans le périplasme, la vitamine B₁₂ est capturée par PBP (periplasmic binding protein, ButF) en jaune et le tout est transmis au transporteur ABC (BtuCD).

et al., 2008). De plus, ces mutants n'ont aucune tolérance aux pHs de 4,5 et 8,5 contrairement à la souche sauvage (Zhang *et al.*, 2008).

2.2.3 Transporteurs TonB-dépendants

Les microorganismes ont développé des systèmes performants d'acquisition des nutriments rares afin de survivre dans leurs habitats. Les bactéries à Gram négatif possèdent deux membranes -externe et interne- à travers lesquelles les ressources doivent passer. La membrane externe possède différents types de porines permettant le transport passif de petites molécules. Pour les molécules les plus grosses et/ou la diffusion des molécules contre leur gradient de concentration, des transports actifs sont mis en place. Les Transporteurs TonB-dépendants (TBDTs) ancrés au niveau de la membrane externe des cellules des bactéries à Gram négatif, se chargent de ce type de transport actif. L'énergie nécessaire pour que le TBDT délivre le substrat vers le cytoplasme, est fournie par la protéine TonB ancrée dans la membrane interne. La protéine TonB interagit avec le TBDT au niveau de la TonB-box. L'action de TonB est accompagnée de deux autres protéines, ExbB et ExbD, également co-localisées dans la membrane interne. Le complexe TonB, ExbB et ExbD exploite la force proton-motrice de la membrane interne pour apporter de l'énergie au TBDT (Schauer *et al.*, 2008 ; Lim, 2010). Une fois que les molécules ont passé la membrane externe, les transporteurs ABC se chargent de les transporter à travers la membrane interne vers le cytoplasme de la cellule bactérienne (Figure 26) (Wiener, 2005).

Les TBDTs ont premièrement été identifiés comme des transporteurs de complexes Fe^{3+} -sidérophore. En effet, les TBDTs et les sidérophores sont des systèmes efficaces de captation et d'assimilation du fer ferrique (Fe^{3+}) développés par les bactéries pour acquérir cet élément essentiel pour leur métabolisme. Le Fe^{3+} est en plus très peu disponible à cause de sa faible solubilité en condition naturelle d'aérobiose. Plus tard, il a été montré que les TBDTs sont également impliqués dans le transport de composés non ferriques comme la vitamine B12 et les sucres (Schauer *et al.*, 2008 ; Lim, 2010). Par exemple, le gène *suxA* présent chez la souche ATCC33913 de *X. campestris* pv. *campestris*, code un TBDT transporteur de saccharose (Blanvillain *et*

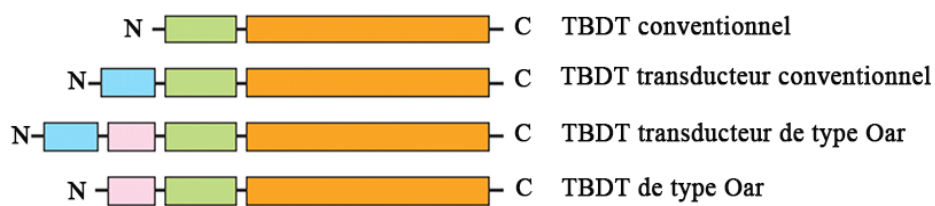


Figure 27. Domaines fonctionnels des TBDTs (d'après Koebnik, 2005). Tous les TBDTs comprennent un domaine β -barrel (orange) en position C-terminale, et un domaine plug (vert). Les TBDTs transducteurs possèdent une extension N-terminale (bleu). Les TBDTs de type Oar et les TBDTs transducteurs de type Oar possèdent un domaine Oar (rose). Ce domaine se trouve chez les bactéroïdes comme *M. xanthus* et chez les *Xanthomonas* et *Xylella*.

al., 2007). Les mêmes auteurs ont de plus montré que l'expression d'une quinzaine de gènes codant les TBDTs, sur les 72 identifiés *in silico* chez *X. campestris* pv. *campestris*, est induite par des molécules d'origine végétale telles que l'acide polygalacturonique, le maltose, l'arabinose, le xylose ou le xylane. Parmi les TBDTs identifiés chez la souche ATCC33913 de *X. campestris* pv. *campestris*, seulement 9 TBDTs sont impliqués dans l'acquisition du fer. En effet, leur expression est réprimée par la protéine régulatrice Fur (Ferric-Uptake Regulator) en présence de fer (Blanvillain *et al.*, 2007). Généralement, les TBDTs ont une très forte affinité avec leur ligand ce qui permet aux bactéries de se procurer des nutriments dont la concentration est faible et la disponibilité est limitée dans leur environnement.

Les TBDTs peuvent être impliqués dans le transport actif de différents éléments nutritifs et/ou dans la transduction de signaux. Ils sont classés en quatre catégories en fonction de leur composition en domaines fonctionnels. On distingue les TBDTs conventionnels, les TBDTs transducteurs conventionnels, les TBDTs de type Oar et les TBDTs transducteurs de type Oar (Figure 27). Tous les TBDTs (toutes catégories confondues) possèdent un domaine β -barrel en position C-terminale et un domaine plug en position N-terminale. Le domaine β -barrel confère au transporteur la forme de tonneau ancré dans la membrane externe de la cellule bactérienne. En absence de ligand, le domaine plug empêche le passage direct des nutriments vers l'intérieur de la cellule bactérienne. Le début de la séquence du domaine plug correspond à la TonB-box qui permet l'interaction avec la protéine TonB. D'autre part, les TBDTs transducteurs (conventionnels ou de type Oar) possèdent une extension N-terminale d'environ 70 acides aminés, impliquée dans la signalisation de la présence de nutriments *via* l'interaction avec un facteur anti-sigma. La cascade de signalisation active la transcription de gènes codant des sidérophores et des transporteurs spécifiques de ces systèmes. Enfin, les TBDTs de type Oar et les TBDTs transducteurs de type Oar possèdent en plus, un domaine de la protéine Oar de *Myxococcus xanthus*, qui est impliquée dans la morphogenèse au moment de la formation de myxospores (Koebnik, 2005).

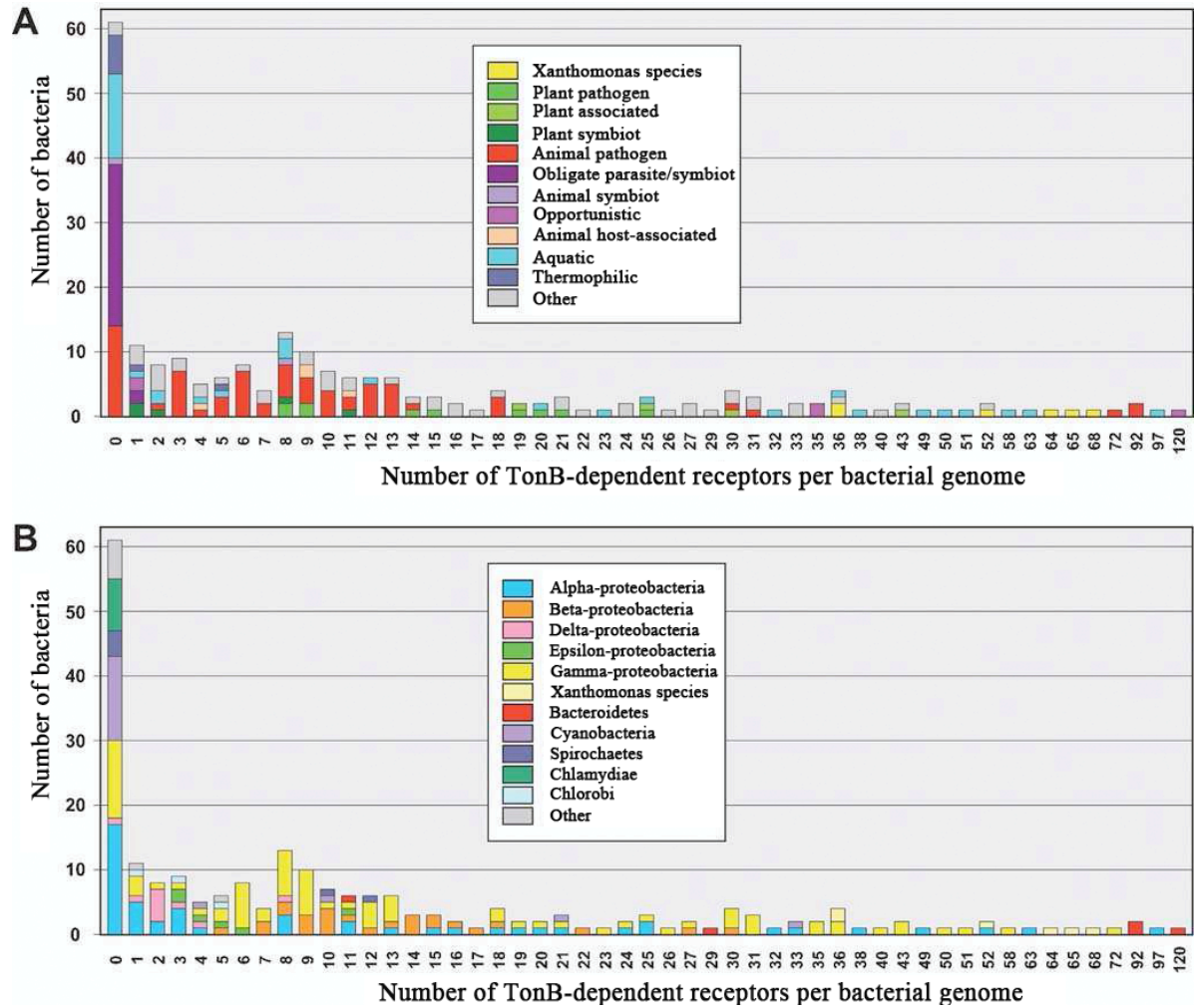


Figure 28. Présence des gènes TBBDT chez 226 souches séquencées à Gram-négatif en fonction de leur caractéristiques écologiques (A) et phylogénétiques (B) (d'après Banvillain *et al.*, 2007). Les gènes codant les TBBDTs ont été détectés par un screening sur les bases de données Pfam et Uniprot.

L'exploration des séquences génomiques complètes de 226 bactéries à Gram négatif a montré que la surreprésentation des TBDTs (plus de 30 TBDTs dans un génome) est une caractéristique restreinte à certaines bactéries (15,5%) (Figure 28) (Blanvillain *et al.*, 2007). En revanche, la surreprésentation des TBDTs est une caractéristique commune à toutes les souches séquencées de *Xanthomonas* spp. pour lesquelles 72 TBDTs ont été identifiés par une analyse *in silico*. La plupart des bactéries ayant cette particularité ont des styles de vies divers et sont capables d'exploiter différentes sources de carbone (Blanvillain *et al.*, 2007 ; Schauer *et al.*, 2008). Il est à noter que les souches phytopathogènes telles que *P. syringae* et *E. carotovora* subsp. *atroseptica* forment une classe intermédiaire avec un nombre de TBDTs inférieur à celui des *Xanthomonas* spp. La souche ATCC33913 de *X. campestris* pv. *campestris* possède des TBDTs transducteurs possédant des extensions N-terminales dont un est conventionnel et 7 sont de type Oar. Un autre TBDT transducteur a été également identifié mais il ne correspond à aucune des catégories présentées ci-dessus. La majorité des TBDTs (au nombre de 48) présents chez cette souche, sont conventionnels ; les 12 TBDTs restant sont identifiés comme étant des pseudo-TBDTs car ils ne présentent pas l'un des domaines β -barrel ou plug ou possèdent des domaines dégénérés (Blanvillain *et al.*, 2007).

2.3 Adhésion bactérienne

La détection d'un environnement favorable est suivie de l'adhésion des bactéries entre elles et au support. Ces étapes sont nécessaires à la formation d'un biofilm. De plus, l'adhésion aux tissus est essentielle pour une colonisation efficace de la plante hôte. Il s'agit d'un mécanisme conservé chez les bactéries à Gram négatif pathogènes de plantes, d'invertébrés et de mammifères (Cao *et al.*, 2001 ; Kline *et al.*, 2009 ; Amano *et al.*, 2010). Les adhésines ainsi que les polysaccharides impliqués dans l'adhésion comme les liposaccharides et les exopolysaccharides, sont présentés dans un chapitre d'ouvrage intitulé « *Adhesion Mechanisms of Plant-Pathogenic Xanthomonadaceae* » que j'ai coécrit avec Marie-Agnès Jacques et Ralf Koebnik. Ce chapitre est inclus dans la section 1 « *Biological systems* » de l'ouvrage « *Bacterial Adhesion: Biology, Chemistry, and Physics* ». Cet ouvrage est édité par Adrian Goldman et Dirk Linke. Des

figures (29 à 32) ont été ajoutées pour une meilleure illustration. Leur numérotation suit celle adoptée dans la partie introduction bibliographique de ce manuscrit.

ADHESION MECHANISMS OF PLANT-PATHOGENIC *XANTHOMONADACEAE*

Nadia Mhedbi-Hajri¹, Marie-Agnès Jacques¹, and Ralf Koebnik²

¹ *Pathologie Végétale (UMR077 INRA–Agrocampus Ouest–Université d'Angers),
Beaucouzé, France.*

² *Laboratoire Génome et Développement des Plantes (UMR5096 Université de
Perpignan–CNRS–IRD), Montpellier, France. Address correspondence to:
koebnik@gmx.de*

1. Introduction
2. Adhesive Structures – Polysaccharides
 - 2.1. Lipopolysaccharides
 - 2.2. Exopolysaccharides
3. Adhesive Structures – Proteins
 - 3.1. Pili (Fimbriae)
 - 3.1.1. *Chaperone/Usher Pili*
 - 3.1.2. *Type IV Pili*
 - 3.2. Non-Fimbrial Adhesins
 - 4.2.1. *Type V-Secreted Adhesins*
 - 4.2.2. *Other Outer Membrane Adhesins*
4. Adhesin Repertoires and Diversifying Selection
5. Adhesin Receptors
6. Concluding Remarks
7. Acknowledgments
8. References

1. INTRODUCTION

The family *Xanthomonadaceae* is a wide-spread family of bacteria belonging to the gamma subdivision of the Gram-negative proteobacteria. Current taxonomy lists 22 genera, among them the two plant-pathogenic genera *Xanthomonas* and *Xylella*, and the related genus *Stenotrophomonas*, isolates of which are increasingly recognised as an important cause of hospital-acquired infections.

Xanthomonas is a large genus of plant-associated bacteria, many of which cause severe diseases of crops and ornamentals, including economically important plants, such as rice, wheat, banana, beans, tomato, *Citrus* and cotton. Common diseases include vascular wilt, cankers, leaf spots, fruit spots, leaf blight and leaf streak. Strains of a certain pathovar only infect a restricted number of plant species and each strain colonizes a favoured ecological niche (tissue), some strains entering the vascular system and causing systemic disease while others stay more or less restricted to the site of infection. Several (putative) pathogenicity factors have been speculated to be responsible for this host and tissue specificity, among them the repertoire of type III effectors, presence/absence of adhesins, and/or their allelic diversity (Hajri *et al.*, 2009; Lu *et al.*, 2008).

Xylella fastidiosa is another important plant pathogen of the *Xanthomonadaceae* family, which is distributed worldwide. Strains of *X. fastidiosa* cause citrus variegated chlorosis or Pierce's disease in grapevine. Other strains cause disease on shrubs and trees (*e.g.* oleander, almond, oak, mulberry, peach, pear), but also on alfalfa. The bacteria form biofilms which contribute to the occlusion of xylem vessels, thus leading to water stress and wilting of the host plant. Leafhoppers are the vector for this xylem-limited fastidious bacterium. The spread of the disease requires the release of the bacterium from the animal vector and its transmission into new plants (Chatterjee *et al.*, 2008a). Four principal pathogenicity factors, most of them related to attachment, have been described for plant colonization and insect acquisition: (i) twitching motility by type IV pili, (ii) secretion of cell-wall degrading enzymes, (iii) cell-to-cell aggregation and surface attachment by fimbrial and non-fimbrial adhesins, and (iv) production of exopolysaccharides (Chatterjee *et al.*, 2008a).

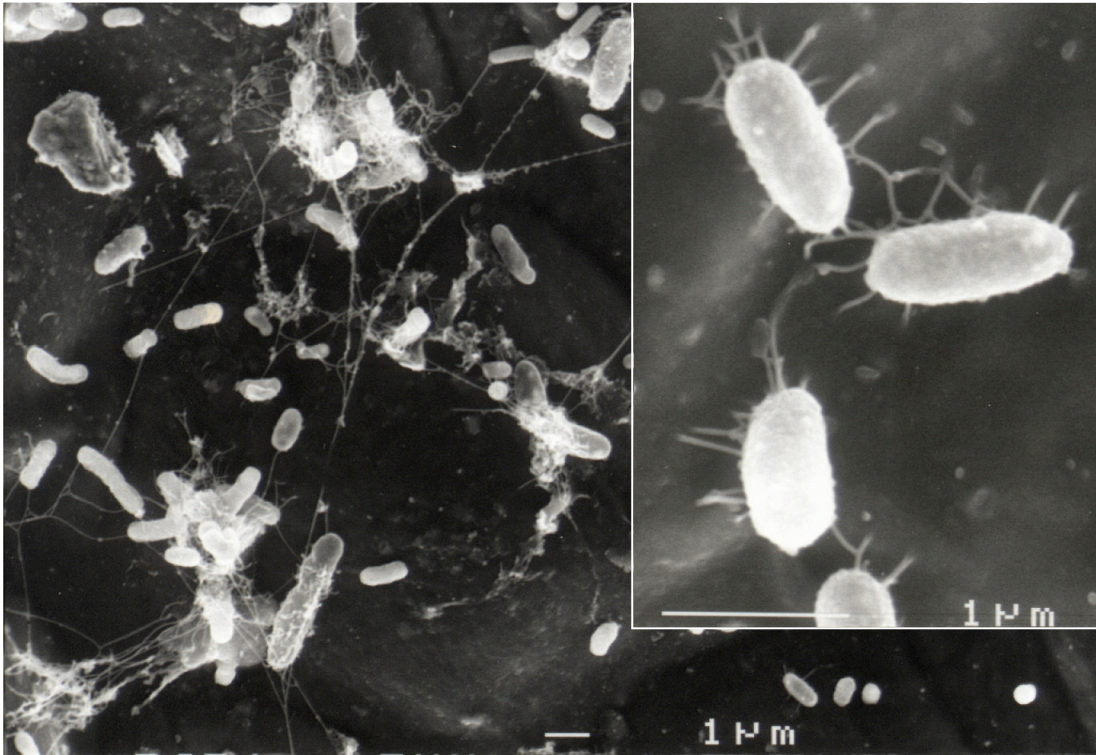


Figure 1. Scanning electron micrographs of field-grown bean leaf surface colonized by seed-borne *X. axonopodis* pv. *phaseoli*. Note the various cell surface appendages used by the bacterial cells to adhere to the leaf surface.

Adhesion is a widely conserved virulence mechanism among Gram-negative bacteria which is used by diverse pathogens to infect their plant, invertebrate and mammalian hosts (Amano, 2010; Cao *et al.*, 2001; Kline *et al.*, 2009). Attachment to a surface is also a prerequisite for aggregation in biofilms which enhance the resistance of bacteria to various biotic and abiotic stresses, favour the coordination of adapted responses to environmental changes and allow multiplication for successful colonization of host tissue (Amano, 2010; Danhorn and Fuqua, 2007). Sensing and adhesion are interconnected since biofilm formation is regulated by a chemosensory system revealing the interplay of the various systems mediating the adaptation to the environment (Dow *et al.*, 2003). The genome sequences of plant-pathogenic *Xanthomonadaceae* contain many genes for the formation of surface adhesive structures (Moreira *et al.*, 2004; Van Sluys *et al.*, 2002). Electron microscopy revealed the presence of surface appendages that are assumed to be involved in attachment to abiotic and/or biotic surfaces (Figure 1).

2. ADHESIVE STRUCTURES – POLYSACCHARIDES

2.1. Lipopolysaccharides

It is generally believed that lipopolysaccharides (LPS) play a role in adhesion, but it is difficult to ascertain whether this effect is direct or indirect (Hori and Matsumoto, 2010). Usually, LPS mutants are also affected in other traits, such as motility, production of type I pili and other adhesins, and type III secretion.

X. campestris pv. *campestris* is a vascular pathogen of cruciferous plants and enters the xylem *via* hydathodes. Analysis of *X. campestris* pv. *campestris* mutants affected in pathogenicity demonstrated that a mutation of *rfaX*, leading to an altered LPS, had the strongest impact on hydathode colonization (Hugouvieux *et al.*, 1998). However, this effect could be due to an attachment defect or due to higher sensitivity towards the antimicrobial defense response mounted in the hydathode. Later, other virulence-reduced mutants in LPS genes were isolated in *X. campestris* pv. *campestris* (Qian *et al.*, 2005). A loss-of-virulence screen of the non-vascular rice pathogen *X. oryzae* pv. *oryzicola* revealed a *wxocB* mutant that was assumed to be affected in the O chain

assembly (Wang *et al.*, 2007). This mutant was strongly impaired in its ability to spread in the mesophyll. Although the reasons for the reduced virulence of LPS mutants remain to be elucidated, these observations imply that an intact LPS is essential for full virulence of vascular and non-vascular xanthomonads.

2.2. Exopolysaccharides

Exopolysaccharide (EPS), also called xanthan gum for *Xanthomonas*, is another bacterial polysaccharide which is released as an extracellular slime (García-Ochoa *et al.*, 2000). EPS is involved in biofilm formation (Vu *et al.*, 2009), which in turn mediates close association of bacteria to abiotic and biotic surfaces (Danhorn and Fuqua, 2007).

More than ten years ago, the first EPS mutants of xanthomonads were described. A *gumD* mutant of *X. campestris* pv. *campestris* and a *gumG* mutant of *X. oryzae* pv. *oryzae* were reported to be reduced in virulence (Chou *et al.*, 1997; Dharmapuri and Sonti, 1999). An exhaustive transposon screen of *X. campestris* pv. *campestris* revealed further mutants in EPS biosynthesis genes which were affected in provoking disease of cabbage (Lu *et al.*, 2007; Qian *et al.*, 2005).

A *X. axonopodis* pv. *citri gumB* mutant, not only defective in production of EPS but also in the formation of biofilms, showed reduced growth and survival on leaf surfaces and reduced disease symptoms, suggesting an important role for the formation of biofilms in the epiphytic survival of *X. axonopodis* pv. *citri* (Rigano *et al.*, 2007). However, another xanthan-defective mutant of *X. axonopodis* pv. *citri*, mutated in *gumD*, remained fully pathogenic in citrus plants but displayed impaired survival under oxidative stress during stationary phase as well as impaired epiphytic survival on citrus leaves (Dunger *et al.*, 2007). It was concluded that xanthan does not play an essential role in citrus canker at the initial stages of infection, but facilitates the maintenance of bacteria on the host plant, possibly improving the efficiency of colonization of distant tissue (Dunger *et al.*, 2007).

EPS synthesis of *X. campestris* pv. *campestris* has been shown to be under control of a cell-cell signalling pathway, involving a small diffusible signal factor (DSF), the DSF synthetase RpfF and a two-component regulatory system comprising RpfC and RpfG (Fouhy *et al.*, 2006). DSF signalling is tightly linked to the intracellular second messenger cyclic dimeric GMP (c-di-GMP) (Jenal and Malone, 2006; Ryan *et al.*, 2006). Subsequent mutant analyses (Ryan *et al.*, 2007; Torres *et al.*, 2007) and microarray experiments (He *et al.*, 2006; He *et al.*, 2007) revealed that DSF/c-di-GMP signalling is not only finely balanced during biofilm formation but involves a hierarchical signalling network affecting many other pathogenicity-related traits, among them type III secretion, motility, extracellular hydrolytic enzymes, detoxification and resistance to oxidative stress. This way, a coordinated switch from a sessile aggregated lifestyle to motile exploratory lifestyle may be accomplished (Fouhy *et al.*, 2006).

Cell-to-cell signalling mediated by a structurally similar diffusible signal factor is also central to the regulation of the virulence of *X. fastidiosa*. *rpfC* mutants hyperexpress *rpfF*, overproduce DSF and are deficient in virulence and movement in the xylem vessels of grape, while *rpfF* mutants deficient in DSF production have the opposite phenotypes for these traits (Chatterjee *et al.*, 2008c). Using green fluorescent protein-labelled bacteria, the *rpfF*-deficient mutant was detected at a greater distance from the point of inoculation than the wild-type strain and attained up to 100-fold larger population sizes (Chatterjee *et al.*, 2008b). Hence, the hypervirulence phenotype of the *rpfF* mutant is likely due to an unrestrained multiplication and a more extensive spread of bacteria within the xylem vessels, thus leading to their occlusion. It was concluded that DSF-mediated cell-to-cell signalling, which restricts movement and colonization of *X. fastidiosa*, may be an adaptation to endophytic growth of the pathogen that prevents the excessive growth of cells in vessels (Chatterjee *et al.*, 2008b). Interestingly, *rpfC* mutants are able to colonize the mouthparts of insect vectors but are not transmitted as efficiently as wild-type bacteria to new host plants, apparently because of their strong adhesiveness (Chatterjee *et al.*, 2008c). Because of the opposing contributions of adhesiveness and other traits to movement within plants and transmission to new host plants, *X. fastidiosa* apparently coordinates these traits in a quorum-dependent fashion involving DSF.

3. ADHESIVE STRUCTURES – PROTEINS

3.1. Pili (Fimbriae)

Pili (fimbriae) are filamentous cell-surface appendages that have been classified by their potential to induce hemagglutination of erythrocytes, their morphology (size), their site of anchoring in the bacterium (polar *versus* omnidirectional), and by their biosynthesis. In *Xanthomonadaceae*, several types of pili have been observed or postulated to exist: chaperone/usher (CU) pili, type IV pili, and pili associated with several protein secretion systems (Büttner and Bonas, 2010). Virtually nothing is known whether or not pili of the type III, type IV and type VI protein secretion systems are specifically involved in adhesion of *Xanthomonadaceae*.

3.1.1. Chaperone/Usher Pili

CU pili are assembled by the relatively simple chaperone/usher pathway (Kline *et al.*, 2010). Genes for CU pili (sometimes also called type I pili) of *Xanthomonadaceae* were first discovered in the genome sequence of *X. fastidiosa* (Bhattacharyya *et al.*, 2002). Strain 9a5c was the first one to be completely sequenced (acc. no. NC_002488), followed by strain Temecula1 (acc. no. NC_004556). The gene cluster XF0077 to XF0083 (PD0058 to PD0062 in strain Temecula1) encodes a periplasmic chaperone, an outer membrane usher protein, and four candidate fimbrial subunits (three in strain Temecula1). The usher and chaperone proteins are almost identical among different strains of *X. fastidiosa* while the fimbrial subunits share 85 to 93 % sequence identity among orthologs but less identity among paralogs. The predicted protein product of *fimA* (PD0062) in *X. fastidiosa* shares 23% identity with the orthologous FimA in *Escherichia coli* K-12 (Li *et al.*, 2007); the usher (PD0060) and chaperone (PD0061) are 36 and 40 % identical to their *E. coli* orthologues.

By transmission electron microscopy two length classes of fimbriae were visualized, the shorter of which (0.4 to 1.0 μ m) corresponded to CU pili (Meng *et al.*, 2005). Scanning electron microscopy revealed that fimbriae size and number, cell aggregation, and cell size were reduced in *fimA* (PD0062) or *fimF* (PD0058) mutants of *X. fastidiosa* when compared with the parental strain (Feil *et al.*, 2003). Later, it was confirmed that a *fimA*

(PD0062) mutant produced no CU pili (Li *et al.*, 2007). Interestingly, CU pili were found to be predominantly located at only one pole of the cell (Li *et al.*, 2007; Meng *et al.*, 2005). Since *X. fastidiosa* has been found to attach polarly to xylem cells (Feil *et al.*, 2003), CU pili might therefore be involved in the polar attachment to plant cells. Mutants in *fimA* or *fimF* remained pathogenic to grapevines but exhibited significantly reduced virulence (Feil *et al.*, 2003; Feil *et al.*, 2007).

Mutants in *fimA* (PD0062), thus possessing only type IV pili, were found to be deficient in biofilm formation (Meng *et al.*, 2005). A role of CU pili in biofilm formation was also suggested by the upregulation of *fimA* during biofilm formation (de Souza *et al.*, 2004). Biofilm formation and adherence to surfaces might interfere with movement. Indeed, a *fimA* mutant was not only capable of moving *via* twitching motility (Meng *et al.*, 2005), but movement was in fact enhanced, suggesting that the presence of CU pili may partially restrict cell movement (Li *et al.*, 2007). Using a microfluidic device it was demonstrated that a *fimA* mutant moved six times faster than wild-type cells on a glass surface (De La Fuente *et al.*, 2007a). Along with this observation, CU pili were found to mediate stronger attachment to a glass surface than type IV pili (De La Fuente *et al.*, 2007b). Not only did *fimA* and *fimF* mutants adhere to glass surfaces less efficiently than the wild-type strain, but these cells occurred primarily as solitary cells and not as cell aggregates, unlike wild-type. In summary, CU pili of *X. fastidiosa* play an important role in cell-to-cell aggregation, biofilm formation and attachment to surfaces.

Complete genome sequencing led to the identification of a conserved gene cluster for CU pili in *Xanthomonas* (acc. no. YP_363214). This gene cluster is predicted to encode one usher protein (XCV1483), two chaperones (XCV1480, XCV1483) and two fimbrial subunits one of which is twice the size (XCV1481) of the other (XCV1484) due to a tandem duplication. The number, order and size of the genes of the CU pilus cluster of *Xanthomonas* is different to that found in *X. fastidiosa*. Moreover, the sequence similarity is very low with less than 25 % sequence identity between the chaperone and usher proteins from *Xanthomonas* and *X. fastidiosa*. *S. maltophilia*, another species of the *Xanthomonadaceae*, contains two CU pilus gene clusters, one corresponding to that

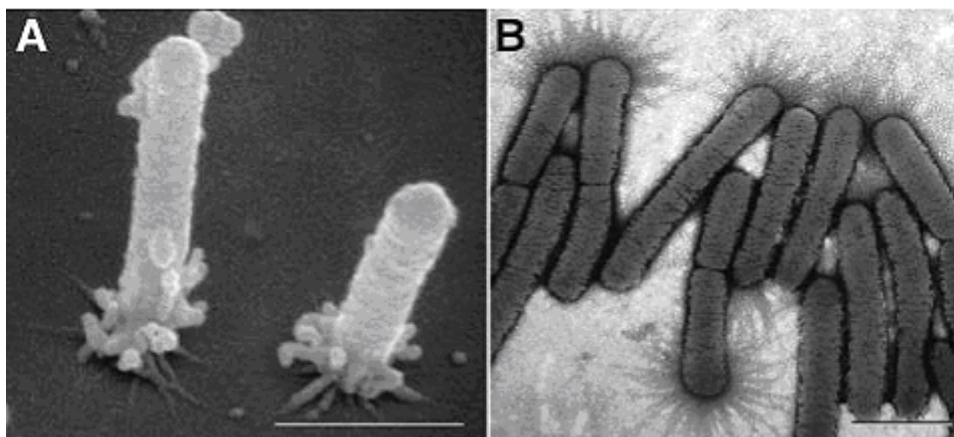


Figure 29. Pili de type IV observés chez une souche sauvage de *Xylella fastidiosa* en microscopie électronique à balayage (d'après Meng et Li, 2005). La barre d'échelle représente 1 μm

from *Xanthomonas* and one to that from *X. fastidiosa*. Hence, CU pili from *Xanthomonas* and from *X. fastidiosa* might play different roles.

3.1.2. Type IV Pili

Type IV pili are homopolymeric assemblies of a small pilin protein, called PilA in *Pseudomonas* and Pile in *Neisseria*, which are expressed by many Gram-negative bacteria, including *Xanthomonadaceae* (Figure 29) (Craig *et al.*, 2004). The type IV pilus assembly machinery is structurally related to the type II secretion system (Gerlach and Hensel, 2007). Once assembled, the long and flexible structure extends outward from the bacterial surface and establishes contact with host cell or abiotic surfaces. Type IV pili have the ability, once firmly attached to a surface, to retract through the bacterial cell wall, thus leading to movement of the bacterial cell – a process called twitching or gliding motility (Jarrell and McBride, 2008).

Type IV pilus-related genes are present in all sequenced strains of *Xanthomonas* and *Xylella*. The major pilin PilA (sometimes misleadingly called FimA) is always encoded on the opposite strand close to the pilus assembly gene *pilC*. The major pilin orthologs from *Xanthomonas* display an amazing diversity which does not correlate with the phylogenetic relationships of the strains. Some pilins are more closely related to Pile from *Neisseria* while others are more similar to homologs from *Legionella*. In a few lineages of *Xanthomonas* (e.g. *X. campestris* pv. *campestris* strains ATCC33913 and 8004, *X. axonopodis* pv. *vesicatoria* strain NCPPB3240, *X. axonopodis* pv. *citri* strain 306), a tandem duplication has occurred the consequences of which are not known (*i.e.* whether there is an antigenic switch due to differential expression of the two genes) (Ojanen-Reuhs *et al.*, 1997). The variability among the major pilins may be associated with specific interactions with their hosts, reminiscent of the situation in human and animal pathogens (Van Sluys *et al.*, 2002). The genome sequences of *X. fastidiosa* also reveal a gene duplication event which probably occurred before or along the speciation of *X. fastidiosa*. However, compared to *Xanthomonas*, there is much less sequence variation among the six completely sequenced strains of *X. fastidiosa*.

Experiments with several *Xanthomonas* species demonstrated an important role of type IV pili in attachment and aggressiveness on their host plants. For example, immunofluorescence studies showed that purified type IV pili of *X. campestris* pv. *hyacinthi* attached to stomata of hyacinth leaves, suggesting a role for these surface structures in the first stages of yellow disease (van Doorn *et al.*, 1994). A *fimA* mutant of *X. axonopodis* pv. *vesicatoria* strain NPPPB3240 was not significantly reduced in virulence upon infiltration or spraying of bacterial suspensions onto tomato leaves (Ojanen-Reuhs *et al.*, 1997). However, scanning electron microscopy revealed that the *fimA* mutant was affected in its ability to adhere and efficiently colonize trichomes of tomato leaves. The *fimA* mutant also was dramatically reduced in cell aggregation in laboratory cultures and on infected tomato leaves and exhibited decreased tolerance to UV light. When the tandemly arranged paralog *fimB* of strain NPPPB3240 was knocked out, compact cell aggregates similar to those formed by the wild-type strain were observed (Ojanen-Reuhs *et al.*, 1997). Mutants in *pilB* and *pilC* of *X. campestris* pv. *campestris* were reduced in virulence upon leaf clip inoculation (Qian *et al.*, 2005). A mutation in *pilA* of *X. axonopodis* pv. *phaseoli* decreased its aggressiveness on bean upon dip-inoculation of leaves (Darsonval *et al.*, 2009). Moreover, PilA was found to be involved in transmission of *X. axonopodis* pv. *phaseoli* to bean seeds *via* xylem vessels (Darsonval *et al.*, 2009).

A mutation in the *X. oryzae* pv. *oryzae* *pilQ* gene, which is predicted to encode the type IV pilus secretin, appeared to have no effect on leaf attachment or entry but reduces the virulence following wound inoculation and *in planta* migration (Das *et al.*, 2009; Lim *et al.*, 2008). These findings indicate that the type IV pilus is not only involved in initial attachment to organs of the leaf, but is also required for optimum virulence within the plant tissue, *i.e.* when *X. oryzae* pv. *oryzae* is multiplying or migrating within xylem vessels. This conclusion is supported by a transposon screen of *X. oryzae* pv. *oryzicola* which revealed that insertions in several *pil* genes (*pilM*, *pilQ*, *pilT*, *pilZ*, *pilY1*) reduce the colonization of rice leaves upon leaf infiltration (Wang *et al.*, 2007). Similarly, mutation in the *pilQ* and *pilB* genes in *X. fastidiosa* has been shown to affect colonization and upstream migration of the bacteria in xylem vessels (Meng *et al.*, 2005). Using a microfluidic flow chamber, the speed of twitching movement of a *pilY1*

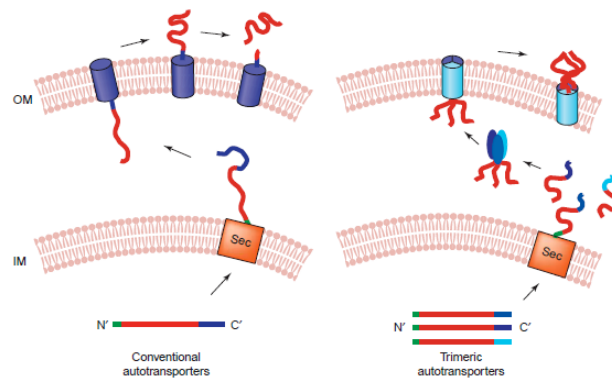


Figure 30. Représentation schématique de la sécrétion des adhésines autotransporteurs (d'après Cotter *et al.*, 2005). Les autotransporteurs monomériques sont clivés à la surface de la bactérie (à gauche) contrairement aux autotransporteurs trimériques (à droite).

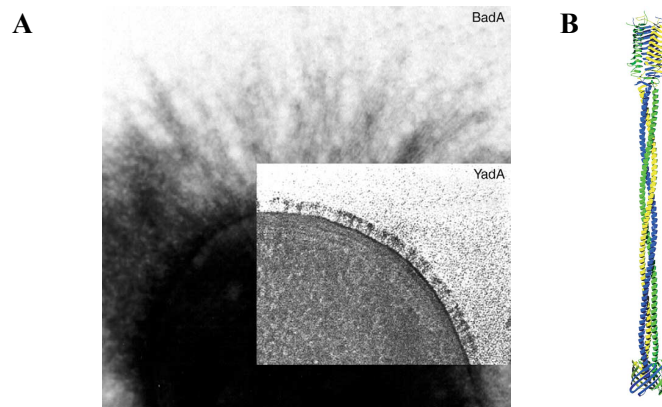


Figure 31. Adhésines de type autotransporteur trimérique (BadA et YadA) observées en microscopie électronique à transmission et modèle de la structure de l'adhésine YadA (d'après Linke *et al.*, 2005). A noter sur la figure A, la différence de longueur entre BadA (environ 300 nm) et YadA (environ 23 nm). Cette différence de longueur est due à la longueur de la « queue » de l'adhésine. Comme représentées sur la figure B, les adhésines de type autotransporteur trimérique se composent d'une tête, d'un cou, d'une très longue queue et d'un domaine d'ancrage.

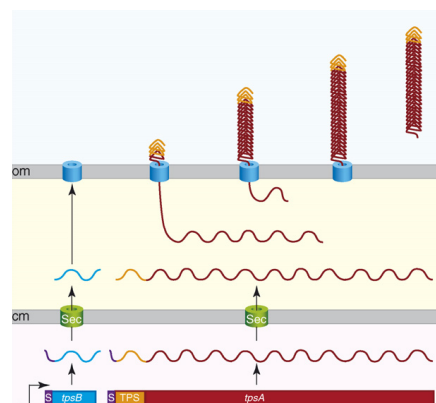


Figure 32. Représentation schématique de la sécrétion bipartite des adhésines non fibrillaires de type hémagglutinine (d'après Mazar et Cotter, 2007). om : membrane externe, cm : membrane cytoplasmique, *tpsA* : gène codant la structure adhésive, *tpsB* : gène codant la protéine canal de type β -barrel. Les gènes *tpsA* et *tpsB* sont schématisés en bas de la figure. Les protéines sont représentées sous la forme de lignes ondulées. Les protéines sont transloquées à travers la membrane cytoplasmique *via* le système sec (en vert). Lors de la translocation, la séquence signal (en violet) est clivée. Dans le périplasme, TpsA s'associe à TpsB grâce au domaine TPS (orange). TpsA est ensuite transloquée dans le milieu extracellulaire.

mutant of *X. fastidiosa* on a glass surface against the flow direction of media was found to be three times slower than that of wild-type cells (De La Fuente *et al.*, 2007a).

Type IV pili were also found to be required for optimal virulence, seed transmission, twitching motility and biofilm formation of several other Gram-negative plant pathogens, such as *Ralstonia solanacearum* and *Acidovorax avenae* subsp. *citrulli* (Bahar *et al.*, 2009; Kang *et al.*, 2002). In contrast, mutation of type IV pili of *Pseudomonas syringae* pv. *tomato* DC3000 had no apparent effect on pathogenicity, as judged by symptom formation upon leaf inoculation (Roine *et al.*, 1998). Yet, type IV pili were found to play a role in fitness during foliar colonization and survival, a finding that might be related to an enhanced tolerance against environmental conditions such as exposure to UV light (Roine *et al.*, 1998). Taken together, results from numerous studies support a role of type IV pili in attachment to certain plant surfaces, in formation of biofilms, and in colonization and migration within xylem vessels.

3.2. Non-Fimbrial Adhesins

Non-fimbrial adhesins are widely found in proteobacteria and contain different classes of proteins involved in attachment (Gerlach and Hensel, 2007; Pizarro-Cerdá and Cossart, 2006). Many of these proteins are presented on the cell surface upon transport *via* one of the three subtypes of type V secretion systems: (i) monomeric autotransporters (T5SS[a]) (Figure 30) (Dautin and Bernstein, 2007), (ii) trimeric autotransporters or oligomeric coiled-coil adhesins (Figures 30 et 31) (T5SS[b]) (Cotter *et al.*, 2005; Linke *et al.*, 2006), and (iii) two-partner secretion systems incl. filamentous hemagglutinins (T5SS[c]) (Figure 32) (Mazar and Cotter, 2006). Other non-fimbrial adhesins are secreted *via* a type I secretion system (*e.g.* LapA from *P. fluorescens*) or represent small integral outer membrane proteins (Amano, 2010; Gerlach and Hensel, 2007).

3.2.1. Type V-Secreted Adhesins

Xanthomonadaceae encode type V-secreted proteins which could act as non-fimbrial adhesins (Table 1) (Moreira *et al.*, 2004; Van Sluys *et al.*, 2002). These proteins are related to the prototypic yersinial adhesins YadA and YapH and to well-studied

Name ¹	Prototype ²	Annotation ³	Pfam domains ⁴	RGD ⁵	Xac ⁶	Xav	XooK	XooM	XooP	Xoc	XccA	Xcc8	XccB	Xca	Xal	Xf 9	Xf T	Xf A	Xf D	Xf 12	Xf 23
	XCV4444	hemagglutinin-related protein	Calx_beta; He_PIG; Autotransporter	YES	Ω	+	3Ω	+	+	Ω	+	+	+	+	-	-	-	-	-	-	-
YapH	XCV2103	filamentous hemagglutinin-related protein; YapH protein	-	YES	+	+	+	2Ω	+	+	+	+	+	+	-	-	-	-	-	-	-
	XCAORF_0904	YapH protein, putative	Autotransporter	no	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-
XadA1	XOO0842	adhesin-like protein	Hep-Hag; HIM	no	+	+	+	+	+	+	+	+	+	+	-	+	+	+	2Ω	+	+
XadA2	XCV3672	adhesin-like protein	Hep_Hag; HIM; YadA	no	+	+	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
XadB	XOO0681	adhesin-like protein	Hep_Hag; HIM; YadA	no	-	-	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
	XALc_1305	hypothetical adhesin	Hep_Hag; HIM; YadA	YES	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
	XALc_1884	hypothetical adhesin	Hep_Hag; HIM; YadA	YES	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
	XF1981	YadA-like surface protein	Hep_Hag; HIM; YadA	no	-	-	-	-	-	-	-	-	-	-	-	+	+	2+	+	+	+
	XF1529	YadA-like surface protein	Hep_Hag; HIM; X_fast-SP_rel; YadA	no	-	-	-	-	-	-	-	-	-	-	-	+	+	2+	+	+	+
FhaC	XAC1814	outer membrane hemolysin activator protein	POTRA_2; ShlB	/	+	+	-	-	+	+	+	+	-	+	+	+	+	2+	+	+	+
FhaB	XAC1815	filamentous hemagglutinin	Haemagg_act; Fil haemagg	YES	+	2Ω	-	-	+	+	+	+	-	+	+	3+	2+	2+	3Ω	-	3+
	XAC4114	hemolysin activator protein	POTRA_2; ShlB	/	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	XAC4113	filamentous hemagglutinin-like protein; [YapH protein]	Haemagg_act	no	+	+	Ω	-	-	Ω	-	-	-	-	-	-	-	-	-	-	-

Table 1. Distribution of non-filamentous adhesins and accessory proteins in completely sequenced strains of *Xanthomonadaceae*.

¹ Common name given to the proteins of *Xanthomonas* spp.

² Locus tag of the prototypical gene/protein given at <http://www.ncbi.nlm.nih.gov/>.

³ Common annotation of protein function at <http://www.ncbi.nlm.nih.gov/>.

⁴ Predicted protein domains at <http://pfam.sanger.ac.uk/>.

⁵ Presence of conserved RGD motifs.

⁶ Presence of proteins in the predicted proteome. +, presence; -, absence; W, protein fragments predicted from putative pseudogenes; numbers indicate numbers of paralogs or protein fragments. Xac, *X. axonopodis* pv. *citri* str. 306 (RefSeq accession number [acc. no.] NC_003919); Xav, *X. axonopodis* pv. *vesicatoria* str. 85-10 (acc. no. NC_007508); XooK, *X. oryzae* pv. *oryzae* KACC10331 (acc. no. NC_006834); XooM, *X. oryzae* pv. *oryzae* MAFF 311018 (acc. no. NC_007705); XooP, *X. oryzae* pv. *oryzae* PXO99^A (acc. no. NC_010717); Xoc, *X. oryzae* pv. *oryzicola* BLS256 (acc. no. NZ_AAQN000000000); XccA, *X. campestris* pv. *campestris* str. ATCC 33913 (acc. no. NC_003902); Xcc8, *X. campestris* pv. *campestris* str. 8004 (acc. no. NC_007086); XccB, *X. campestris* pv. *campestris* str. B100 (acc. no. NC_010688); Xca, *X. campestris* pv. *armoraciae* 756C (available at the Comprehensive Microbial Resource <http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi>); Xal, *X. albilineans* GPE PC73 (acc. no. NC_013722); Xf_9, *X. fastidiosa* 9a5c (acc. no. NC_002488); Xf_T, *X. fastidiosa* Temecula1 (acc. no. NC_004556); Xf_A, *X. fastidiosa* Ann-1 (acc. no. NZ_AAAM000000000); Xf_D, *X. fastidiosa* Dixon (acc. no. NZ_AAAL000000000); Xf_12, *X. fastidiosa* M12 (acc. no. NC_010513); Xf_23, *X. fastidiosa* M23 (acc. no. NC_010577).

adhesins from *Bordetella pertussis* (pertactin, filamentous hemagglutinin FHA) and *Haemophilus influenzae* (HMW1, HMW2) (Hoiczuk *et al.*, 2000; Smith *et al.*, 2001; St Geme 3rd and Yeo, 2009).

Comparative genomics revealed that predicted adhesion genes of some strains have frameshift mutations and/or in-frame stop codons, thus most likely leading to truncated, probably inactive protein variants (Table 1) (Bhattacharyya *et al.*, 2002; Van Sluys *et al.*, 2003). A pronounced case of DNA duplication and decay is found for the *fhaB* gene which encodes the passenger domain that is translocated across the outer membrane by the FhaC protein. Most strains appear to have a full-length *fhaB* gene downstream of which closely related gene fragments are present. It would be interesting to study whether these fragments can serve as a genetic reservoir to evolve new adhesin gene variants. Unfortunately, such genetic variability had the side-effect of misleading genome annotation, and many predicted genes are most likely not functional. For instance, two ORFs downstream of *fhaB* [XAC1815] from *X. axonopodis* pv. *citri*, XAC1816 and XAC1818, have been annotated as hemagglutinin-like proteins but they correspond to internal protein fragments of FhaB. Since XAC1816 and XAC1818 lack the conserved two-partner secretion domain at the N terminus, these proteins (genes) are most likely not functional (Gottig *et al.*, 2009). Similarly, *fhaB1* [XCV1860] and *fhaB2* [XCV1861] from *X. axonopodis* pv. *vesicatoria* might not be functional because these predicted genes correspond to 5' and 3' fragments of the full-length *fhaB* gene of *X. axonopodis* pv. *citri*. Another source of confusion comes from the annotation of several two-partner secreted proteins as YapH although they don't display any sequence homology to the yersinial YapH protein.

The first evidence that a non-fimbrial adhesin is involved in the infection process of a plant-pathogenic bacterium came from mutagenesis and expression studies of its gene (Ray *et al.*, 2002). The *xadA1* gene of *X. oryzae* pv. *oryzae*, a homologue of the yersinial *yadA* gene, was shown to be induced in minimal medium, and *xadA1* mutants were strongly affected for virulence upon epiphytic infection, suggesting an important virulence function somewhere along the path from the leaf surface, *via* the hydathodes, into the xylem vessels (Ray *et al.*, 2002). It should be noted that at the same time, *xadA1*

mutants produced less EPS (Ray *et al.*, 2002). Later, the role of XadA1 and, to a lesser extent, of its paralog XadB in attachment to and entry into rice leaves was elegantly confirmed by confocal fluorescence microscopy using eGFP-tagged *X. oryzae* pv. *oryzae* (Das *et al.*, 2009). Interestingly, when *X. axonopodis* pv. *passiflorae* cells were treated with leaf extracts, the *xadA* homolog was only induced with extracts from its host plant, passion fruit, but not with extracts from a non-host plant (tomato) (Tahara *et al.*, 2003). Darsonval and colleagues did not observe any role of XadA1 in *X. axonopodis* pv. *phaseoli* aggressiveness on bean (Darsonval *et al.*, 2009). Instead, the paralog *xadA2* was found to be required for vascular transmission to bean seeds (Darsonval *et al.*, 2009). Despite remaining pathogenic on grapevine, *xadA* mutants of *X. fastidiosa* also were affected in aggressiveness (Feil *et al.*, 2007). Apparently, different paralogs play specific roles in the infection process and efficient infection and dissemination may require an intricate balance of these and other adhesins.

DNA microarray expression studies of 279 selected genes from *X. axonopodis* pv. *citri* showed that more adhesin genes, *i.e.* *fhaB* and XAC4113, are induced in minimal medium that mimics *in planta* conditions (Astua-Monge *et al.*, 2005), and synthesis of the two-partner secreted adhesin FhaB was also found to be induced during citrus leaf colonization (Gottig *et al.*, 2009). Mutant analysis showed that this adhesin is not only involved in epiphytic fitness, attachment to plant surfaces and colonization of citrus leaves, but also in cell-to-cell-attachment and biofilm formation. Opposite to what was observed with a *xadA1* mutant, the *fhaB* mutant secreted more EPS (Gottig *et al.*, 2009). The homologous FhaB protein of *X. axonopodis* pv. *phaseoli* is involved in attachment to seeds of beans but not to leaves (Darsonval *et al.*, 2009). However, FhaB does not seem to be important in aggressiveness in *X. axonopodis* pv. *phaseoli*.

Conversely, the hemagglutinin-like YapH orthologue appears to limit aggressiveness of *X. axonopodis* pv. *phaseoli* on bean since a *yapH* mutant led to more severe symptoms when inoculated onto bean leaves at high concentration (Darsonval *et al.*, 2009). Thus, *yapH* fulfills the hallmarks of an ‘anti-virulence’ gene, mutations of which result in a hypervirulent phenotype as measured by a lower lethal dose, a colonization advantage, reduced clearance or decreased survival time of the host (Foreman-Wykert and Miller,

2003). On the other hand, YapH is the only adhesin of *X. axonopodis* pv. *phaseoli* which was found to be required for the complete phyllosphere colonization process and for seed transmission through the vascular pathway, probably due to its requirement for adhesion to biotic and abiotic surfaces and for biofilm formation (Darsonval *et al.*, 2009). A hypervirulent phenotype was also observed upon knockout of the haemagglutinin-like proteins HfxA (PD2118) and HxfB (PD1792) of *X. fastidiosa* strain Temecula1 (Guilhabert and Kirkpatrick, 2005). Surprisingly, Feil and colleagues reported the opposite effect when they inoculated a *hxfB* mutant of strain Temecula1 into grapevine, following the same procedure of needle inoculation at high concentration (Feil *et al.*, 2007). The expression of adhesion genes, *i.e.* *fimA*, *hxfA* and *hxfB*, is much higher in *rpfC* mutants of *X. fastidiosa*, which also exhibit a hyperattachment phenotype in culture that is associated with their inability to migrate in xylem vessels and cause disease (Chatterjee *et al.*, 2008c). Conversely, *rpfF* mutants synthesize less FimA, HxfA, and HxfB and are hypervirulent, thus looking like phenocopies of *hxfA* and *hxfB* mutants.

3.2.2. Other Outer Membrane Adhesins

Beside type V-secreted proteins, other integral outer membrane proteins have been identified as adhesins in human and animal pathogens. These proteins are embedded in the outer membrane as antiparallel beta-barrels (Koebnik *et al.*, 2000). The eight-stranded OmpA of *E. coli* is the archetype of this class of proteins and has been shown to function as an adhesin (Smith *et al.*, 2007). Despite the presence of similar small beta-barrel proteins in phytopathogenic bacteria, it is unknown whether or not they are involved in attachment to plant cells. Given the thick plant cell wall, such a mechanism seems unlikely to be exploited by plant pathogens. However, bacterial plant pathogens, such as *Xanthomonas* spp., secrete a plethora of cell-wall degrading enzymes that might help the bacteria to approach the plant cell plasma membrane (Moreira *et al.*, 2004; Van Sluys *et al.*, 2002).

4. ADHESIN REPERTOIRES AND DIVERSIFYING SELECTION

Recently, repertoires of genes involved in early stages of interaction between host and pathogenic bacteria were characterized in a large collection of *Xanthomonas* strains

(Mhedbi-Hajri *et al.*, 2010). Three categories of adhesin genes were described: (i) ubiquitous genes that are widespread among different pathovars and genetic lineages (type IV pilus genes *pilL*, *pilS*, *pilU* and *pilA*); (ii) less conserved genes that were not detected in several strains (*e.g.* XCV2103, *phaB* [XAC1815] and XAC1816); and (iii) species-specific genes that were detected only in *X. axonopodis* (*e.g.* *xadA2*, *phaB1* [XCV1860] and *phaB2* [XCV1861]). For instance, XAC1816 is present in ten pathovars of *X. axonopodis* and is also present in non-pathogenic strains of *X. campestris* whereas *phaB1* and *phaB2* were only detected in six pathovars of *X. axonopodis* (Mhedbi-Hajri *et al.*, 2010).

Using the McDonald-Kreitman test, Mhedbi-Hajri and colleagues analyzed adhesin genes from three strains of *X. axonopodis* (pathovars *vesicatoria*, *citri* and *phaseoli*) and from three strains of *X. campestris* pv. *campestris* for molecular signatures of selection pressure (Mhedbi-Hajri *et al.*, 2010). Adaptive divergence by diversifying selection was found to affect *xadA1* and *phaB*, particularly within the *X. axonopodis* clade. It was concluded that adhesion should be considered as a selective step for settlement of bacteria on plant tissue (Mhedbi-Hajri *et al.*, 2010).

5. ADHESIN RECEPTORS

Adhesins are thought to interact with a variety of host components, either unspecifically *via* hydrophobic or electrostatic interactions, or by binding to specific host cell receptor moieties (Hori and Matsumoto, 2010). Adhesion-binding may trigger the activation of complex signal-transduction cascades in host cells, thus inducing innate defense responses which can be either harmful or beneficial to the bacteria (Amano, 2010).

A few examples may illustrate the diversity of adhesin-receptor interactions that have been discovered for human pathogens (Amano, 2010; Gerlach and Hensel, 2007; Kline *et al.*, 2009). Tip adhesins of type I pili have been found to interact with glycolipids and glycoprotein receptors while type IV PAK pili from *P. aeruginosa* bind to gangliosides. Nonfimbrial adhesins recognize many different elements on host cell surfaces, including components of the extracellular matrix (ECM). For instance, the prototypic trimeric autotransporter YadA from *Yersinia* spp. mediates adhesion to several ECM proteins,

such as collagen, laminin, and fibronectin (Nummelin *et al.*, 2004). Other outer membrane-associated adhesins (*e.g.* neisserial Opa proteins) bind to integral host cell membrane receptors, such as integrins, cadherins, selectins, and carcinoembryonic antigen-related cell adhesion molecules (CEA-CAM). In contrast to these examples, nothing is known for plant-pathogenic *Xanthomonadaceae*.

Type V-secreted adhesins, such as the *E. coli* antigen 43 and pertactin from *B. pertussis*, bind to integrins *via* conserved RGD motifs (Junker *et al.*, 2006; Takagi 2004; van der Woude and Henderson, 2008). Similar to their role in animal tissue, integrin-like molecules of plants have been found to be involved in connecting the plant cell wall to the plasma membrane (Lü *et al.*, 2007), and this interaction could be disrupted by treatment with RGD-containing peptides (Canut *et al.*, 1998; Lü *et al.*, 2007; Schindler *et al.*, 1989). Candidate integrin-like RGD-binding receptors have been identified by immunolabelling with integrin-specific antisera, by cross-linking of iodinated RGD-heptapeptides and by affinity chromatography using immobilized RGD-containing peptides (Faik *et al.*, 1998; Labouré *et al.*, 1999; Senchou *et al.*, 2004).

The plant-pathogenic oomycete *Phytophthora infestans* secretes an RGD-containing protein which might bind to a specific receptor protein of *Arabidopsis thaliana*, thus enabling/promoting the infection process *via* a loosened cellular organization (Senchou *et al.*, 2004). Subsequently, a RGD-reactive receptor-like kinase with a lectin-like extracellular domain was identified, and it was speculated that RGD-dependent protein-protein interactions could play a structural and signalling role at the plant cell surface (Gouget *et al.*, 2006). Whether bacterial plant pathogens take advantage of RGD motifs for binding to plant cells is still unknown. However, conserved RGD motifs have been found in the monomeric autotransporters XCV2103 and XCV4444 but not in the trimeric autotransporters XadA or XadB. RGD motifs are also present in the FhaB homologs. Even if RGD-mediated binding to integrin-like proteins is involved in attachment to host cells, it is not clear whether these bacterial proteins determine host specificity due to the evolution of new binding specificities. Rather, diversifying selection might lead to host adaptation *via* escape from detection by the plant's immune

system, and these adhesins would thus constitute a further class of pathogen-associated molecular patterns (PAMPs) (Postel and Kemmerling, 2009; Silipo *et al.*, 2010).

6. CONCLUDING REMARKS

Our current knowledge about the adhesion process and the involved molecular structures of *Xanthomonadaceae* resembles a kaleidoscope but how all the different components interact synergistically or antagonistically is still very much unknown. In recent years, phytopathologists mainly used genetics and genomics to gain insight into the initial steps of the pathogen-host interaction. Thereby, they often relied on genetic screens which were not well-suited to elucidate the individual roles of candidate adhesion molecules. New assays, mimicking the infection process and using *in situ* fluorescence microscopy (Das *et al.*, 2009; Newman *et al.*, 2003), will allow a better understanding of the adaptive phase of infection of vascular and non-vascular pathogens, *i.e.* attraction and attachment. Moreover, structural biology, biochemistry and biophysics are expected to provide detailed insights into specific molecular interactions between pathogen-derived adhesins and their host cell receptors.

7. ACKNOWLEDGMENTS

We are grateful to Maurice Lesourd and Robert Filmon from the Service Commun d'Imagerie et d'Analyses Microscopiques, Faculté de Médecine, Université d'Angers, France, for help with scanning electron microscopy.

8. REFERENCES

- A. Amano. Bacterial adhesins to host components in periodontitis. *Periodontol 2000* **52**(1):12-37 (2010).
- G. Astua-Monge, J. Freitas-Astua, G. Bacocina, J. Roncoletta, S.A. Carvalho and M.A. Machado. Expression profiling of virulence and pathogenicity genes of *Xanthomonas axonopodis* pv. *citri*. *J Bacteriol* **187**(3):1201–1205 (2005).
- O. Bahar, T. Goffer and S. Burdman. Type IV Pili are required for virulence, twitching motility, and biofilm formation of *Acidovorax avenae* subsp. *citrulli*. *Mol Plant Microbe Interact* **22**(8):909–920 (2009).
- A. Bhattacharyya, S. Stilwagen, N. Ivanova, M. D'Souza, A. Bernal, A. Lykidis, V. Kapatral, I. Anderson, N. Larsen, T. Los, G. Reznik, E. Selkov Jr, T.L. Walunas, H. Feil, W.S. Feil, A. Purcell, J.L. Lassez, T.L. Hawkins, R. Haselkorn, R. Overbeek, P.F. Predki and N.C. Kyrpides. Whole-genome comparative analysis of three phytopathogenic *Xylella fastidiosa* strains. *Proc Natl Acad Sci U S A* **99**(19):12403–12408 (2002).
- D. Büttner and U. Bonas. Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiol Rev* **34**(2):107–133 (2010).
- H. Canut, A. Carrasco, J.P. Galaud, C. Cassan, H. Bouyssou, N. Vita, P. Ferrara and R. Pont-Lezica. High affinity RGD-binding sites at the plasma membrane of *Arabidopsis thaliana* links the cell wall. *Plant J* **16**(1):63–71 (1998).
- H. Cao, R.L. Baldini and L.G. Rahme. Common mechanisms for pathogens of plants and animals. *Annu Rev Phytopathol* **39**:259–284 (2001).
- S. Chatterjee, R.P. Almeida and S. Lindow. Living in two worlds: the plant and insect lifestyles of *Xylella fastidiosa*. *Annu Rev Phytopathol* **46**:243–271 (2008a).
- S. Chatterjee, K.L. Newman and S.E. Lindow. Cell-to-cell signaling in *Xylella fastidiosa* suppresses movement and xylem vessel colonization in grape. *Mol Plant Microbe Interact* **21**(10):1309–1315 (2008b).
- S. Chatterjee, C. Wistrom and S.E. Lindow. A cell-cell signaling sensor is required for virulence and insect transmission of *Xylella fastidiosa*. *Proc Natl Acad Sci U S A* **105**(7):2670–2675 (2008c).
- F.L. Chou, H.C. Chou, Y.S. Lin, B.Y. Yang, N.T. Lin, S.F. Weng and Y.H. Tseng. The *Xanthomonas campestris gumD* gene required for synthesis of xanthan gum is involved in normal pigmentation and virulence in causing black rot. *Biochem Biophys Res Commun* **233**(1):265–269 (1997).
- S.E. Cotter, N.K. Surana and J.W. St Geme 3rd. Trimeric autotransporters: a distinct subfamily of autotransporter proteins. *Trends Microbiol* **13**(5):199–205 (2005).
- L. Craig, M.E. Pique and J.A. Tainer. Type IV pilus structure and bacterial pathogenicity. *Nat Rev Microbiol* **2**(5):363–378 (2004).
- T. Danhorn and C. Fuqua. Biofilm formation by plant-associated bacteria. *Annu Rev Microbiol* **61**:401–422 (2007).
- A. Darsonval, A. Darrasse, K. Durand, C. Bureau, S. Cesbron and M.A. Jacques. Adhesion and fitness in the bean phyllosphere and transmission to seed of *Xanthomonas fuscans* subsp. *fuscans*. *Mol Plant Microbe Interact* **22**(6):747–757 (2009).
- A. Das, N. Rangaraj and R.V. Sonti. Multiple adhesin-like functions of *Xanthomonas oryzae* pv. *oryzae* are involved in promoting leaf attachment, entry, and virulence on rice. *Mol Plant Microbe Interact* **22**(1):73–85 (2009).
- N. Dautin and H.D. Bernstein. Protein secretion in gram-negative bacteria via the autotransporter pathway. *Annu Rev Microbiol* **61**:89–112 (2007).
- L. De La Fuente, T.J. Burr and H.C. Hoch. Mutations in type I and type IV pilus biosynthetic genes affect twitching motility rates in *Xylella fastidiosa*. *J Bacteriol* **189**(20):7507–7510 (2007a).

- L. De La Fuente, E. Montanes, Y. Meng, Y. Li, T.J. Burr, H.C. Hoch and M. Wu. Assessing adhesion forces of type I and type IV pili of *Xylella fastidiosa* bacteria by use of a microfluidic flow chamber. *Appl Environ Microbiol* **73**(8):2690–2696 (2007b).
- A.A. de Souza, M.A. Takita, H.D. Coletta-Filho, C. Caldana, G.M. Yanai, M.H. Muto, R.C. de Oliveira, L.R. Nunes and M.A. Machado. Gene expression profile of the plant pathogen *Xylella fastidiosa* during biofilm formation in vitro. *FEMS Microbiol Lett* **237**(2):341–353 (2004).
- S. Dharmapuri and R.V. Sonti. A transposon insertion in the *gumG* homologue of *Xanthomonas oryzae* pv. *oryzae* causes loss of extracellular polysaccharide production and virulence. *FEMS Microbiol Lett* **179**(1):53–59 (1999).
- J.M. Dow, L. Crossman, K. Findlay, Y.Q. He, J.X. Feng and J.L. Tang. Biofilm dispersal in *Xanthomonas campestris* is controlled by cell-cell signaling and is required for full virulence to plants. *Proc Natl Acad Sci U S A* **100**(19):10995–11000 (2003).
- G. Dunger, V.M. Relling, M.L. Tondo, M. Barreras, L. Ielpi, E.G. Orellano and J. Ottado. Xanthan is not essential for pathogenicity in citrus canker but contributes to *Xanthomonas* epiphytic survival. *Arch Microbiol* **188**(2):127–135 (2007).
- A. Faik, A.M. Labouré, D. Gulino, P. Mandaron and D. Falconet. A plant surface protein sharing structural properties with animal integrins. *Eur J Biochem* **253**(3):552–559 (1998).
- H. Feil, W.S. Feil, J.C. Detter, A.H. Purcell and S.E. Lindow. Site-directed disruption of the *fimA* and *fimF* fimbrial genes of *Xylella fastidiosa*. *Phytopathology* **93**(6):675–682 (2003).
- H. Feil, W.S. Feil and S.E. Lindow. Contribution of fimbrial and afimbrial adhesins of *Xylella fastidiosa* to attachment to surfaces and virulence to grape. *Phytopathology* **97**(3):318–324 (2007).
- A.K. Foreman-Wykert and J.F. Miller. Hypervirulence and pathogen fitness. *Trends Microbiol* **11**(3):105–108 (2003).
- Y. Fouhy, J.F. Lucey, R.P. Ryan and J.M. Dow. Cell-cell signaling, cyclic di-GMP turnover and regulation of virulence in *Xanthomonas campestris*. *Res Microbiol* **157**(10):899–904 (2006).
- F. García-Ochoa, V.E. Santos, J.A. Casas and E. Gómez. Xanthan gum: production, recovery, and properties. *Biotechnol Adv* **18**(7):549–579 (2000).
- R.G. Gerlach and M. Hensel. Protein secretion systems and adhesins: the molecular armory of Gram-negative pathogens. *Int J Med Microbiol* **297**(6):401–415 (2007).
- N. Gottig, B.S. Garavaglia, C. Garofalo, E.G. Orellano and J. Ottado. A filamentous hemagglutinin-like protein of *Xanthomonas axonopodis* pv. *citri*, the phytopathogen responsible for citrus canker, is involved in bacterial virulence. *PLoS One* **4**(2):e4358 (2009).
- A. Gouget, V. Senchou, F. Govers, A. Sanson, A. Barre, P. Rougé, R. Pont-Lezica and H. Canut. Lectin receptor kinases participate in protein-protein interactions to mediate plasma membrane-cell wall adhesions in *Arabidopsis*. *Plant Physiol* **140**(1):81–90 (2006).
- M.R. Guilhabert and B.C. Kirkpatrick. Identification of *Xylella fastidiosa* antivirulence genes: hemagglutinin adhesins contribute to *X. fastidiosa* biofilm maturation and colonization and attenuate virulence. *Mol Plant Microbe Interact* **18**(8):856–868 (2005).
- A. Hajri, C. Brin, G. Hunault, F. Lardeux, C. Lemaire, C. Manceau, T. Boureau and S. Poussier. A "repertoire for repertoire" hypothesis: repertoires of type three effectors are candidate determinants of host specificity in *Xanthomonas*. *PLoS One* **4**(8):e6632 (2009).
- Y.W. He, M. Xu, K. Lin, Y.J. Ng, C.M. Wen, L.H. Wang, Z.D. Liu, H.B. Zhang, Y.H. Dong, J.M. Dow and L.H. Zhang. Genome scale analysis of diffusible signal factor regulon in *Xanthomonas campestris* pv. *campestris*: identification of novel cell-cell communication-dependent genes and functions. *Mol Microbiol* **59**(2):610–622 (2006).
- Y.W. He, A.Y. Ng, M. Xu, K. Lin, L.H. Wang, Y.H. Dong and L.H. Zhang. *Xanthomonas campestris* cell-cell communication involves a putative nucleotide receptor protein Clp and a hierarchical signalling network. *Mol Microbiol* **64**(2):281–292 (2007).

- E. Hoiczuk, A. Roggenkamp, M. Reichenbecher, A. Lupas and J. Heesemann. Structure and sequence analysis of *Yersinia* YadA and *Moraxella* UspAs reveal a novel class of adhesins. *EMBO J* **19**(22):5989–5999 (2000).
- K. Hori and S. Matsumoto. Bacterial adhesion: From mechanism to control. *Biochem Engineer J* **48**(3):424–434 (2010).
- V. Hugouvieux, C.E. Barber and M.J. Daniels. Entry of *Xanthomonas campestris* pv. *campestris* into hydathodes of *Arabidopsis thaliana* leaves: a system for studying early infection events in bacterial pathogenesis. *Mol Plant Microbe Interact* **11**(6):537–543 (1998).
- K.F. Jarrell and M.J. McBride. The surprisingly diverse ways that prokaryotes move. *Nat Rev Microbiol* **6**(6):466–476 (2008).
- U. Jenal and J. Malone. Mechanisms of cyclic-di-GMP signaling in bacteria. *Annu Rev Genet* **40**:385–407 (2006).
- M. Junker, C.C. Schuster, A.V. McDonnell, K.A. Sorg, M.C. Finn, B. Berger and P.L. Clark. Pertactin β -helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. *Proc Natl Acad Sci U S A* **103**(13): 4918–4923 (2006).
- Y. Kang, H. Liu, S. Genin, M.A. Schell and T.P. Denny. *Ralstonia solanacearum* requires type 4 pili to adhere to multiple surfaces and for natural transformation and virulence. *Mol Microbiol* **46**(2):427–437 (2002).
- K.A. Kline, S. Falkner, S. Dahlberg, S. Normark and B. Henriques-Normark. Bacterial adhesins in host-microbe interactions. *Cell Host Microbe* **5**(6):580–592 (2009).
- K.A. Kline, K.W. Dodson, M.G. Caparon and S.J. Hultgren. A tale of two pili: assembly and function of pili in bacteria. *Trends Microbiol* **18**(5):224–232 (2010).
- R. Koebnik, K.P. Locher and P. Van Gelder P. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol Microbiol* **37**(2):239–253 (2000).
- A.M. Laboure, A. Faik, P. Mandaron and D. Falconet. RGD-dependent growth of maize calluses and immunodetection of an integrin-like protein. *FEBS Lett* **442**(2-3):123–128 (1999).
- Y. Li, G. Hao, C.D. Galvani, Y. Meng, L. De La Fuente, H.C. Hoch and T.J. Burr. Type I and type IV pili of *Xylella fastidiosa* affect twitching motility, biofilm formation and cell-cell aggregation. *Microbiology* **153**(3):719–726 (2007).
- S.H. Lim, B.H. So, J.C. Wang, E.S. Song, Y.J. Park, B.M. Lee and H.W. Kang. Functional analysis of *pilQ* gene in *Xanthomonas oryzae* pv. *oryzae*, bacterial blight pathogen of rice. *J Microbiol* **46**(2):214–220 (2008).
- D. Linke, T. Riess, I.B. Autenrieth, A. Lupas and V.A. Kempf. Trimeric autotransporter adhesins: variable structure, common function. *Trends Microbiol* **14**(6):264–270 (2006).
- B. Lu, F. Chen, Z.H. Gong, H. Xie, J.H. Zhang and J.S. Liang. Intracellular localization of integrin-like protein and its roles in osmotic stress-induced abscisic acid biosynthesis in *Zea mays*. *Protoplasma* **232**(1-2):35–43 (2007).
- G.T. Lu, Z.F. Ma, J.R. Hu, D.J. Tang, Y.Q. He, J.X. Feng and J.L. Tang. A novel locus involved in extracellular polysaccharide production and virulence of *Xanthomonas campestris* pathovar *campestris*. *Microbiology* **153**(3):737–746 (2007).
- H. Lu, P. Patil, M.A. Van Sluys, F.F. White, R.P. Ryan, J.M. Dow, P. Rabinowicz, S.L. Salzberg, J.E. Leach, R. Sonti, V. Brendel and A.J. Bogdanove. Acquisition and evolution of plant pathogenesis-associated gene clusters and candidate determinants of tissue-specificity in *Xanthomonas*. *PLoS One* **3**(11):e3828 (2008).
- J. Mazar and P.A. Cotter. Topology and maturation of filamentous haemagglutinin suggest a new model for two-partner secretion. *Mol Microbiol* **62**(3):641–654 (2006).
- Y. Meng, Y. Li, C.D. Galvani, G. Hao, J.N. Turner, T.J. Burr and H.C. Hoch. Upstream migration of *Xylella fastidiosa* via pilus-driven twitching motility. *J Bacteriol* **187**(16):5560–5567 (2005).

- N. Mhedbi-Hajri, A. Darrasse, S. Pigné, K. Durand, S. Fouteau, V. Barbe, C. Manceau, C. Lemaire and M.A. Jacques. Bacterial sensing and adhesion at early stages of host colonization is adaptive. Submitted for publication.
- L.M. Moreira, R.F. de Souza, N.F. Almeida Jr, J.C. Setubal, J.C. Oliveira, L.R. Furlan, J.A. Ferro and A.C. da Silva. Comparative genomics analyses of citrus-associated bacteria. *Annu Rev Phytopathol* **42**:163–184 (2004).
- K.L. Newman, R.P. Almeida, A.H. Purcell and S.E. Lindow. Use of a green fluorescent strain for analysis of *Xylella fastidiosa* colonization of *Vitis vinifera*. *Appl Environ Microbiol* **69**(12):7319–7327 (2003).
- H. Nummelin, M.C. Merckel, J.C. Leo, H. Lankinen, M. Skurnik and A. Goldman. The *Yersinia* adhesin YadA collagen-binding domain structure is a novel left-handed parallel beta-roll. *EMBO J* **23**(4):701–711 (2004).
- T. Ojanen-Reuhs, N. Kalkkinen, B. Westerlund-Wikström, J. van Doorn, K. Haahtela, E.L. Nurmiaho-Lassila, K. Wengelnic, U. Bonas and T.K. Korhonen. Characterization of the *fimA* gene encoding bundle-forming fimbriae of the plant pathogen *Xanthomonas campestris* pv. *vesicatoria*. *J Bacteriol* **179**(4):1280–1290 (1997).
- J. Pizarro-Cerdá and P. Cossart. Bacterial adhesion and entry into host cells. *Cell* **124**(4):715–727 (2006).
- S. Postel and B. Kemmerling. Plant systems for recognition of pathogen-associated molecular patterns. *Semin Cell Dev Biol* **20**(9):1025–1031 (2009).
- W. Qian, Y. Jia, S.X. Ren, Y.Q. He, J.X. Feng, L.F. Lu, Q. Sun, G. Ying, D.J. Tang, H. Tang, W. Wu, P. Hao, L. Wang, B.L. Jiang, S. Zeng, W.Y. Gu, G. Lu, L. Rong, Y. Tian, Z. Yao, G. Fu, B. Chen, R. Fang, B. Qiang, Z. Chen, G.P. Zhao, J.L. Tang and C. He. Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome Res* **15**(6):757–767 (2005).
- S.K. Ray, R. Rajeshwari, Y. Sharma and R.V. Sonti. A high-molecular-weight outer membrane protein of *Xanthomonas oryzae* pv. *oryzae* exhibits similarity to non-fimbrial adhesins of animal pathogenic bacteria and is required for optimum virulence. *Mol Microbiol* **46**(3):637–647 (2002).
- L.A. Rigano, F. Siciliano, R. Enrique, L. Sendín, P. Filippone, P.S. Torres, J. Qüesta, J.M. Dow, A.P. Castagnaro, A.A. Vojnov M.R. Marano. Biofilm formation, epiphytic fitness, and canker development in *Xanthomonas axonopodis* pv. *citri*. *Mol Plant Microbe Interact* **20**(10):1222–1230 (2007).
- E. Roine, D.M. Raineri, M. Romantschuk, M. Wilson and D.N. Nunn. Characterization of type IV pilus genes in *Pseudomonas syringae* pv. *tomato* DC3000. *Mol Plant Microbe Interact* **11**(11):1048–1056 (1998).
- R.P. Ryan, Y. Fouhy, J.F. Lucey, L.C. Crossman, S. Spiro, Y.W. He, L.H. Zhang, S. Heeb, M. Cámara, P. Williams and J.M. Dow. Cell-cell signaling in *Xanthomonas campestris* involves an HD-GYP domain protein that functions in cyclic di-GMP turnover. *Proc Natl Acad Sci U S A* **103**(17):6712–6717 (2006).
- R.P. Ryan, Y. Fouhy, J.F. Lucey, B.L. Jiang, Y.Q. He, J.X. Feng, J.L. Tang and J.M. Dow. Cyclic di-GMP signalling in the virulence and environmental adaptation of *Xanthomonas campestris*. *Mol Microbiol* **63**(2):429–442 (2007).
- M. Schindler, S. Meiners and D.A. Cheresch. RGD-dependent linkage between plant cell wall and plasma membrane: consequences for growth. *J Cell Biol* **108**(5):1955–1965 (1989).
- V. Senchou, R. Weide, A. Carrasco, H. Bouyssou, R. Pont-Lezica, F. Govers and H. Canut. High affinity recognition of a *Phytophthora* protein by *Arabidopsis* via an RGD motif. *Cell Mol Life Sci* **61**(4):502–509 (2004).
- A. Silipo, G. Erbs, T. Shinya, J.M. Dow, M. Parrilli, R. Lanzetta, N. Shibuya, M.A. Newman and A. Molinaro. Glyco-conjugates as elicitors or suppressors of plant innate immunity. *Glycobiology* **20**(4):406–419 (2010).
- A.M. Smith, C.A. Guzmán and M.J. Walker. The virulence factors of *Bordetella pertussis*: a matter of control. *FEMS Microbiol Rev* **25**(3): 309–333 (2001).

- S.G. Smith, V. Mahon, M.A. Lambert and R.P. Fagan. A molecular Swiss army knife: OmpA structure, function and expression. *FEMS Microbiol Lett* **273**(1):1–11 (2007).
- J.W. St Geme 3rd and H.J. Yeo. A prototype two-partner secretion pathway: the *Haemophilus influenzae* HMW1 and HMW2 adhesin systems. *Trends Microbiol* **17**(8):355–360 (2009).
- S.T. Tahara, A. Mehta and Y.B. Rosato. Proteins induced by *Xanthomonas axonopodis* pv. *passiflorae* with leaf extract of the host plant (*Passiflorae edulis*). *Proteomics* **3**(1):95–102 (2003).
- I. Takagi. Structural basis for ligand recognition by RGD (Arg-Gly-Asp)-dependent integrins. *Biochem Soc Trans* **32**(3):403–406 (2004).
- P.S. Torres, F. Malamud, L.A. Rigano, D.M. Russo, M.R. Marano, A.P. Castagnaro, A. Zorreguieta, K. Bouarab, J.M. Dow and A.A. Vojnov. Controlled synthesis of the DSF cell-cell signal is required for biofilm formation and virulence in *Xanthomonas campestris*. *Environ Microbiol* **9**(8):2101–2109 (2007).
- M.W. van der Woude and I.R. Henderson. Regulation and function of Ag43 (flu). *Annu Rev Microbiol* **62**:153–169 (2008).
- J. van Doorn, P.M. Boonekamp and B. Oudega. Partial characterization of fimbriae of *Xanthomonas campestris* pv. *hyacinthi*. *Mol Plant Microbe Interact* **7**(3):334–344 (1994).
- M.A. Van Sluys, M.C. de Oliveira, C.B. Monteiro-Vitorello, C.Y. Miyaki, L.R. Furlan, L.E. Camargo, A.C. da Silva, D.H. Moon, M.A. Takita, E.G. Lemos, M.A. Machado, M.I. Ferro, F.R. da Silva, M.H. Goldman, G.H. Goldman, M.V. Lemos, H. El-Dorry, S.M. Tsai, H. Carrer, D.M. Carraro, R.C. de Oliveira, L.R. Nunes, W.J. Siqueira, L.L. Coutinho, E.T. Kimura, E.S. Ferro, R. Harakava, E.E. Kuramae, C.L. Marino, E. Giglioti, I.L. Abreu, L.M. Alves, A.M. do Amaral, G.S. Baia, S.R. Blanco, M.S. Brito, F.S. Cannavan, A.V. Celestino, A.F. da Cunha, R.C. Fenille, J.A. Ferro, E.F. Formighieri, L.T. Kishi, S.G. Leoni, A.R. Oliveira, V.E. Rosa Jr, F.T. Sasaki, J.A. Sena, A.A. de Souza, D. Truffi, F. Tsukumo, G.M. Yanai, L.G. Zaros, E.L. Civerolo, A.J. Simpson, N.F. Almeida Jr, J.C. Setubal and J.P. Kitajima. Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. *J Bacteriol* **185**(3):1018–1026 (2003).
- M.A. Van Sluys, C.B. Monteiro-Vitorello, L.E.A. Camargo, C.F.M. Menck, A.C.R. da Silva, J.A. Ferro, M.C. Oliveira, J.C. Setubal, J.P. Kitajima, and A.J. Simpson. Comparative genomic analysis of plant-associated bacteria. *Annu Rev Phytopathol* **40**:169–189 (2002).
- B. Vu, M. Chen, R.J. Crawford and E.P. Ivanova. Bacterial extracellular polysaccharides involved in biofilm formation. *Molecules* **14**(7):2535–2554 (2009).
- L. Wang, S. Makino, A. Subedee and A.J. Bogdanove. Novel candidate virulence factors in rice pathogen *Xanthomonas oryzae* pv. *oryzicola* as revealed by mutational analysis. *Appl Environ Microbiol* **73**(24):8023–8027 (2007).

3. Evolution adaptative des génomes bactériens et répertoires de gènes

Lors des différentes étapes de l'interaction avec leur plante hôte *i.e* l'attraction, l'adhésion, la multiplication et la dispersion, les bactéries font face à des challenges rigoureux pour survivre, se multiplier et se transmettre à un nouvel hôte (Moxon *et al.*, 1994, 2006 ; Falush, 2009). Les mécanismes d'adaptation développés par les bactéries agissent au plan phénotypique et/ou au plan génotypique. Le premier mécanisme d'adaptation consiste à percevoir les variations et les stimuli concernant l'environnement externe et à réguler en conséquence, les réponses comportementales et physiologiques de la bactérie. Ainsi, les senseurs bactériens détectent les signaux externes puis, se chargent de l'activation des voies de transduction cytoplasmiques régulant au final l'expression des gènes cibles. La régulation des réponses offre un avantage clair au fitness des bactéries mais le répertoire des réponses régulées est souvent limité. Le deuxième mécanisme correspond à l'adaptation génétique. La population évolue par la sélection de variants (mutants) les mieux adaptés (Moxon *et al.*, 1994, 2006). Contrairement au premier mécanisme d'adaptation reposant sur la régulation de l'expression des gènes, les mutations et l'acquisition de fragments d'ADN de souches qui sont déjà adaptées à un environnement donné, génèrent une diversification au sein de la population et une évolution de l'organisme vis-à-vis de son environnement sur un plus long terme (Moxon *et al.*, 1994, 2006 ; Falush, 2009). Dans les génomes, les mutations touchent préférentiellement les gènes associés à l'interaction avec l'hôte tels que les gènes impliqués dans le chimiotactisme, l'adhésion aux cellules de l'hôte, la sensibilité aux antibiotiques *etc* (Moxon *et al.*, 1994 ; Falush, 2009). Ainsi, par ce mécanisme d'adaptation, les populations bactériennes deviennent très adaptées à leurs hôtes voire même spécialistes de niches écologiques particulières (Moxon *et al.*, 1994, 2006).

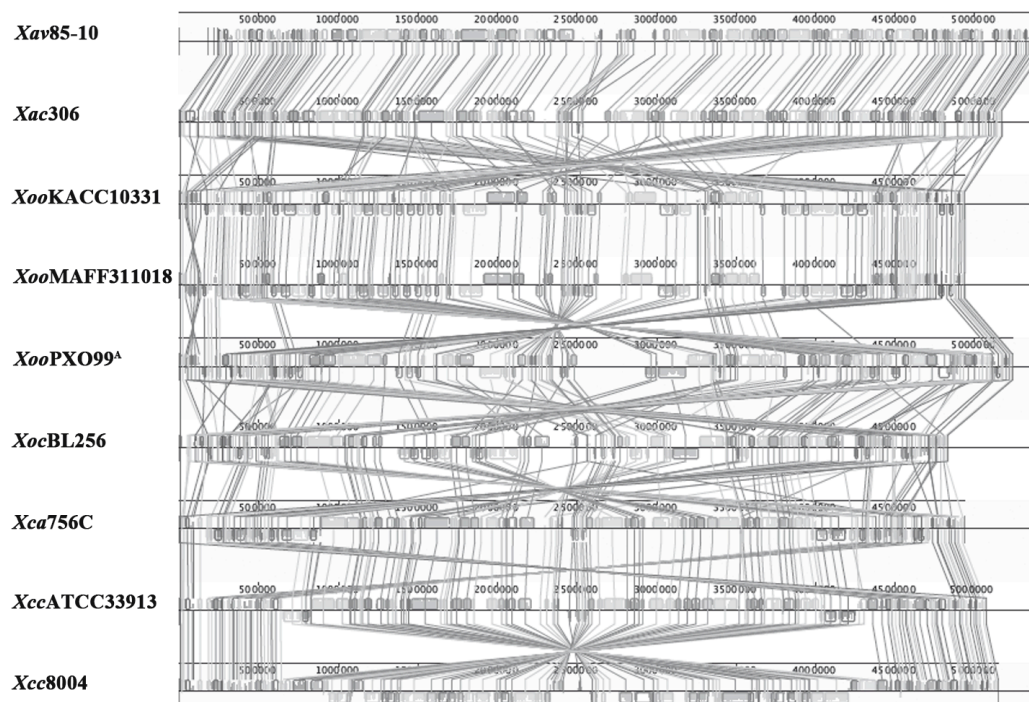


Figure 33. Alignement de neuf génomes de souches séquencées de *Xanthomonas* (d'après Meyer et Bogdanove, 2009). Les régions qui n'ont pas été affectées par des réarrangements sont représentées par des rectangles. L'orientation de ces régions est indiquée par leur position sur le génome. Les lignes entre les génomes représentent des régions orthologues. Des lignes parallèles illustrent une synténie alors que les lignes croisées révèlent la présence d'inversions.

3.1 Évolution adaptative des génomes bactériens

Chez les eucaryotes, la variabilité génétique résulte de la reproduction sexuée ; des recombinaisons chromosomiques se font au cours de la méiose *i.e* lors de la formation des gamètes. Chez les procaryotes, cette forme de reproduction n'existe pas. Cependant, l'analyse des séquences génomiques de souches bactériennes pathogènes de l'Homme, des animaux et des végétaux a révélé une grande variabilité. Cette variabilité génétique est due à l'action de forces modelant l'évolution de ces agents pathogènes. Les trois forces majeures qui contrôlent la dynamique évolutive des génomes bactériens et leur adaptation sont principalement : le gain, la perte et les changements des séquences de gènes (Pallen et Wren, 2007). Il est à noter que de petites variations peuvent se produire dans un génome au niveau d'un seul nucléotide ; c'est ce qu'on appelle les SNPs (single-nucleotide polymorphisms) (Pallen et Wren, 2007).

3.1.1 Gain de gènes par transfert horizontal et duplication

La plus grande surprise résultant de l'application de la génomique à la bactériologie a été la découverte de la grande variabilité des génomes bactériens. A titre d'exemple, une région génomique de 1,4 Mb diffère entre les souches K-12 et O157:H7 d'*E. coli* (Hayashi *et al.*, 2001). Le contenu des gènes dans les génomes des souches séquencées appartenant à différentes espèces et pathovars du genre *Xanthomonas* est largement conservé. En revanche, les alignements complets de ces séquences génomiques révèlent l'existence de nombreuses inversions, insertions, délétions et réarrangements génétiques (Figure 33) (Thieme *et al.*, 2005 ; Qian *et al.*, 2008 ; Meyer et Bogdanove, 2009). Ces variations montrent qu'au sein des *Xanthomonas*, les génomes bactériens se caractérisent par une grande plasticité (Meyer et Bogdanove, 2009).

La majorité des bactéries présentent deux catégories de gènes : les gènes du « core genome » et les gènes du génome accessoire. Les gènes du « core genome » sont

ubiquistes chez toutes les souches d'une même espèce. Ils codent des protéines dont le rôle est essentiel au bon fonctionnement cellulaire de la bactérie tels que le métabolisme, la transcription, *etc.* Ces gènes possèdent un G+C% homogène. Par contre, les gènes du génome accessoire codent généralement des fonctions qui ne sont pas primordiales à la multiplication bactérienne. Ces gènes confèrent un avantage sélectif sous des conditions particulières. La majorité des fragments d'ADN transférés horizontalement appartient au génome accessoire dont le G+C% est différent du reste du génome (Hacker et Carniel, 2001 ; Pallen et Wren, 2007). La proportion de gènes appartenant au génome accessoire d'une bactérie varie de 18% (chez *E. coli*) à une proportion inférieure à 1% (chez *Mycoplasma*) (Lawrence et Ochman, 1998 ; Ochman *et al.*, 2000).

Le transfert horizontal de gènes est la principale source d'« innovation » et de variation au niveau d'un génome. Les transferts horizontaux de gènes se produisent généralement entre des souches bactériennes de même espèce ou d'espèces différentes partageant la même niche écologique (Pallen et Wren, 2007). Les échanges de gènes se produisent dans les sites qualifiés de hot spots où les conditions sont favorables pour la croissance, la colonisation et l'agrégation des cellules. C'est effectivement le cas des sites confinés qui se trouvent au niveau de la phyllosphère (van Elsas *et al.*, 2003). Les conditions qui influencent les échanges de gènes au niveau des hot pots sont abiotiques (ex. la température, le pH, l'humidité relative, la disponibilité en nutriments et en oxygène) et biotiques (ex. interactions au sein des communautés microbiennes d'une même niche écologique) (van Elsas *et al.*, 2003). Les éléments génétiques mobiles comme les plasmides, les bactériophages et les îlots génomiques (ex. les îlots de pathogénicité) sont impliqués dans les transferts horizontaux de gènes (Hacker et Carniel, 2001 ; Maurelli, 2007 ; Pallen et Wren, 2007). Ces éléments comportent des gènes codant des facteurs impliqués dans la colonisation et l'infection (Pallen et Wren, 2007). Les îlots génomiques renfermant des clusters de gènes ayant des fonctions spécifiques, confèrent un avantage sélectif sous des conditions stressantes. Ils améliorent la capacité d'adaptation des bactéries ; ils augmentent leur taux de survie, de colonisation et de transmission. Ces îlots peuvent être subdivisés en différentes catégories en fonction du style de vie et de la niche de la bactérie. Les îlots écologiques

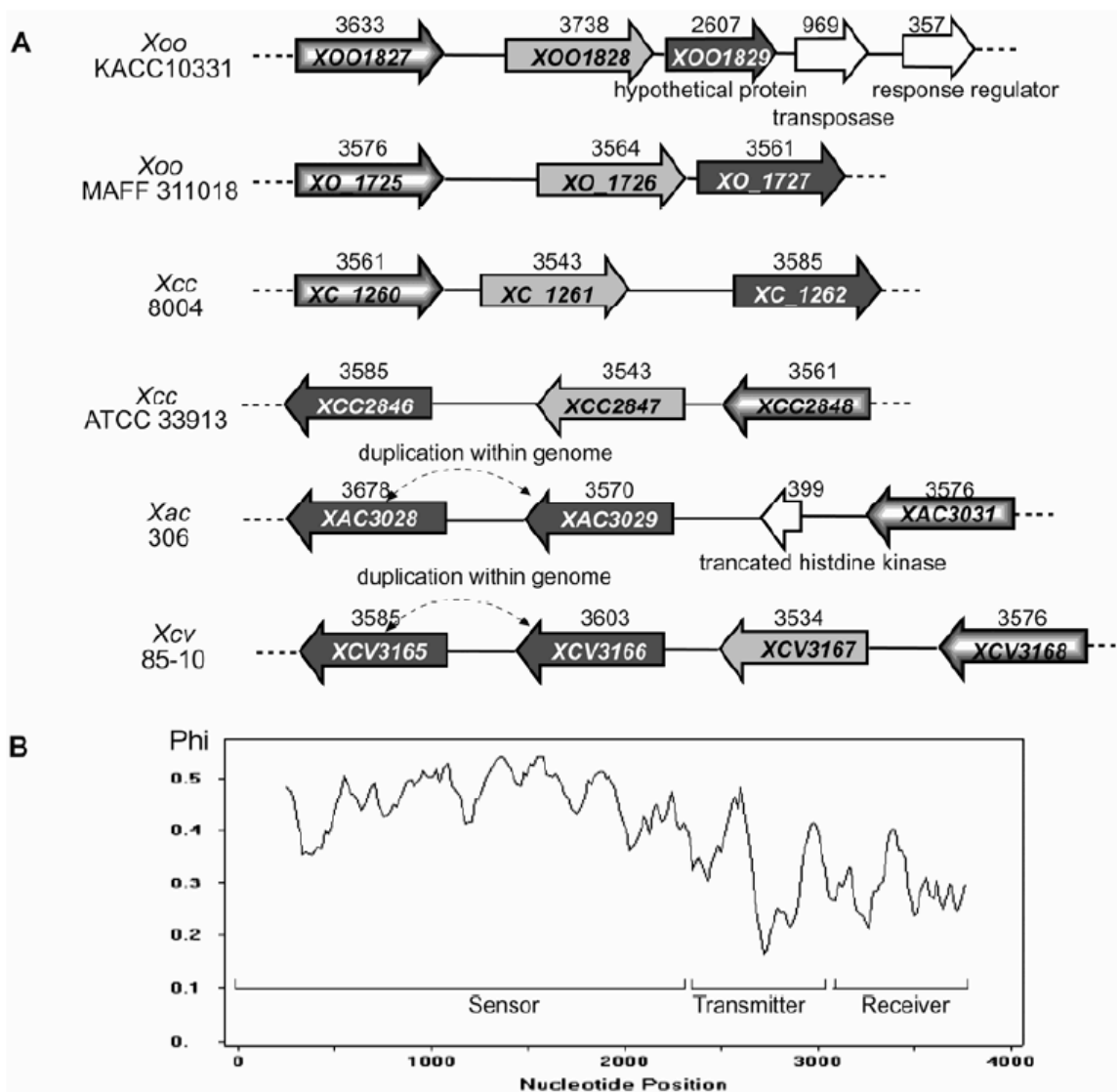


Figure 34. Organisation et polymorphisme d'une région génomique présente dans les génomes des *Xanthomonas* spp. comprenant des gènes codant des senseurs couplés à leurs régulateurs (histidines kinases hybrides, HyHKs) (Qian *et al.*, 2008). La figure A illustre l'organisation des gènes codant les HyHKs dans six génomes de *Xanthomonas*. Les flèches représentent les gènes codant les HyHKs et leurs directions. Les flèches portant la même nuance de gris indiquent que les gènes sont des orthologues ou des paralogues. La longueur des gènes est indiquée au dessus des flèches. Les accessions des gènes sont données en fonction des souches : 85-10 de *X. axonopodis* pv. *vesicatoria*, 306 de *X. axonopodis* pv. *citri*, ATCC33913 et 8004 de *X. campestris* pv. *campestris* et KACC10331 et MAFF311018 de *X. oryzae* pv. *oryzae*. La figure B illustre l'analyse du polymorphisme nucléotidique des HyHKs en utilisant une fenêtre coulissante. La taille de la fenêtre est de 50 bp et le pas est de 9 bp.

favorisent en effet la survie des bactéries dans un habitat particulier alors que les îlots de pathogénicité sont impliqués dans le pouvoir pathogène des bactéries (Hacker et Carniel, 2001). Un même îlot peut être écologique quand la bactérie réside à la surface de la plante hôte et devenir un îlot de pathogénicité quand la bactérie engage un processus infectieux (Hacker et Carniel, 2001). C'est le cas des adhésines qui peuvent jouer un double rôle. Elles favorisent en effet l'installation et l'adhésion des bactéries sur les tissus de l'hôte et participent à leur pouvoir pathogène (Hacker et Carniel, 2001 ; Darsonval *et al.*, 2009).

La duplication de gènes contribue à l'expansion de familles de protéines favorisant l'adaptation des bactéries à des environnements différents (Gevers *et al.*, 2004). C'est d'ailleurs le cas des gènes codant les MCPs (da Silva *et al.*, 2002 ; Moreira *et al.*, 2004 ; Thieme *et al.*, 2005). Les gènes codant le système de régulation à deux composants chez les *Xanthomonas* sont soumis également à des duplications (Qian *et al.*, 2008). En effet, un segment génomique de 13 kb contient deux à quatre répétitions en tandem de senseurs couplés à leurs régulateurs. En outre, une analyse des sites synonymes et non synonymes par fenêtre coulissante, indique que le polymorphisme nucléotidique de la région 5' correspondant à la région senseur, est supérieur à celui de la région 3' correspondant à la région de transmission du signal (Figure 34). Ce résultat suggère que la région senseur évolue plus rapidement que la région de transmission du signal. Ceci est probablement avantageux pour les bactéries du genre *Xanthomonas* pour détecter de nouveaux stimuli externes avec des cascades de signalisation déjà établies (Qian *et al.*, 2008). Jordan *et al.* (2002) ont montré que les gènes issus de duplications présentent des taux d'évolution (Ka/Ks) plus élevés (Ka étant le taux de substitutions non synonymes et Ks étant le taux de substitutions synonymes). Les copies dupliquées peuvent diverger suite à l'action de mutations ponctuelles et de pressions de sélection exercée par l'habitat ou l'hôte (Wang *et al.*, 1982 ; Qian *et al.*, 2008 ; Whitworth et Cock, 2009). Après duplication d'un régulateur transcriptionnel par exemple, le nouveau régulateur dupliqué peut initialement reconnaître les mêmes sites de fixation et réguler les mêmes gènes cibles comme le régulateur transcriptionnel d'origine. Au cours de la divergence, le régulateur transcriptionnel dupliqué peut continuer soit de réguler les mêmes gènes cibles que son ancêtre mais en réponse à un

signal externe différent soit de réguler d'autres gènes cibles (Teichmann et Madan Babu, 2004).

3.1.2 Perte de gènes

Malgré les transferts horizontaux de gènes, les bactéries conservent en général un génome de même taille. Le gain de gènes est en effet compensé par la perte d'autres gènes. Les gènes qui ne sont pas maintenus par une pression de sélection naturelle sont perdus (Pallen et Wren, 2007). La perte de gènes correspond en quelque sorte à une purge progressive à partir des génomes de gènes inutiles au style de vie des bactéries (Pallen et Wren, 2007). Le séquençage de plusieurs génomes bactériens a permis de recueillir des « clichés » montrant directement les processus dynamiques de perte de gènes. A titre d'exemple, les génomes des souches KACC10331 et MAFF311018 de *X. oryzae* pv. *oryzae* ont perdu 37 gènes codant des protéines du système de régulation à deux composants sur les 70 conservées entre les différentes souches séquencées de *Xanthomonas* (Qian *et al.*, 2008). Parmi les gènes perdus par *X. oryzae* pv. *oryzae*, des gènes codent des régulateurs qui contrôleraient l'expression du facteur sigma σ^B (Qian *et al.*, 2008). La perte du facteur sigma et de ses régulateurs a également été observée chez d'autres bactéries comme *Mycobacterium leprae*, l'agent causal de la lèpre. Les facteurs sigma sont utilisés par l'ARN polymérase pour contrôler l'expression des gènes en réponse à des stimuli environnementaux et physiologiques (Tyagi et Saini, 2004). Ces facteurs confèrent à l'ARN polymérase, une spécificité de fixation aux promoteurs des gènes cibles (Madan Babu, 2003). La perte des gènes codant les facteurs sigma déclenche en aval, l'accumulation de mutations au niveau des gènes cibles et la création de pseudogènes (Madan Babu, 2003 ; Tyagi et Saini, 2004 ; Qian *et al.*, 2008). La perte de ces gènes uniquement chez *X. oryzae* pv. *oryzae* indique que ces gènes sont inutiles pour leur style de vie mais qu'ils restent importants pour le style de vie des autres *Xanthomonas* (Qian *et al.*, 2008). L'évolution réductive des génomes bactériens se faisant par perte de gènes vise donc à éliminer les gènes qui ne sont pas indispensables au style de vie de la bactérie (Maurelli *et al.*, 2007).

3.1.3 Changements des séquences de gènes

Les changements des séquences sont relatifs à tout changement affectant les séquences nucléotidiques ou l'ordre des gènes préexistant (Pallen et Wren, 2007). Ils sont principalement dus à des mutations adaptatives ou l'insertion d'éléments transposables au niveau des gènes. Chez les bactéries pathogènes, les mutations pathoadaptatives peuvent améliorer le fitness de ces bactéries et augmenter les taux de croissance et de survie dans une nouvelle niche (Sokurenko *et al.*, 1999).

Des études basées sur l'analyse de séquences ont permis de montrer que les structures de surface tels que les transporteurs membranaires (les transporteurs ABC et les transporteurs de la famille Omp, les TBDTs et les adhésines fibrillaires FimH) sont soumis à une pression de sélection positive et en particulier, chez les souches d'*E. coli* uropathogènes (Chen *et al.*, 2006, 2009). Ceci n'est pas le cas des souches colonisant des environnements autres que l'appareil urinaire. Ce type d'approche démontre l'importance de la variabilité allélique dans l'adaptation des bactéries à leur environnement (Chen *et al.*, 2006). Des travaux ont démontré que chez *E. coli*, des pressions de sélection positive sont exercées sur la séquence codant l'adhésine FimH (Chen *et al.*, 2009). Ces pressions visent à sélectionner des variants de FimH ayant deux spécificités écologiques différentes. En effet, les variants nommés M₁L sont identifiés chez les souches d'*E. coli* colonisant les tissus intestinaux et les variants M₁H, des souches d'*E. coli* colonisant les tissus de l'appareil urinaire. Ainsi, les pressions de sélection positive permettent à ces bactéries de s'adapter à des niches écologiques différentes et de présenter par conséquent, des tropismes tissulaires différents (Sokurenko *et al.*, 1998). Il est important de mentionner, que les variants M₁H d'*E. coli* présentent une plus grande spécificité tissulaire. En effet, ils adhèrent à des substrats cellulaires d'origine différente (globule rouge, bouche et appareil urinaire) contrairement aux variants M₁L qui ne colonisent que les tissus intestinaux (Sokurenko *et al.*, 1998). Cependant, l'adaptation des variants M₁H à un plus grand nombre de récepteurs cellulaires s'accompagne d'un compromis. Ces variants sont en effet, plus sensibles aux composés mannosylés qui inhibent leur adhésion. Ces composés sont principalement présents dans les muqueuses de l'appareil digestif. En revanche,

l'adhésion des variants M₁L est optimisée au niveau des surfaces présentant de fortes concentrations en composés mannosylés (Sokurenko *et al.*, 1998). Par ailleurs, une étude réalisée sur une collection de *P. syringae* a démontré l'importance de la diversification allélique des ET3s (ex. HopZ1a, HopZ1b, HopZ1c) dans l'adaptation à l'hôte. Cette diversification allélique est due aux pressions de sélection exercées par les hôtes d'isolement (Ma *et al.*, 2006 ; Zhou *et al.*, 2009).

Le taux de mutation est le déterminant majeur de la diversité au sein d'une population. Il augmente suite à l'action d'une pression de sélection lorsque par exemple, la bactérie est sujette à des fluctuations importantes dans son habitat (Tenailon *et al.*, 2001). Le principal inconvénient d'une stratégie adaptative moyennant l'augmentation du taux de mutation chez les bactéries est que des mutations délétères peuvent se produire à des fréquences plus importantes que celles des mutations bénéfiques (Radman *et al.*, 2000 ; Moxon *et al.*, 2006). Il est important de mentionner que les bactéries paient généralement le prix des mutations qui améliorent leur fitness en devenant spécialistes d'une niche écologique bien spécifique (Moxon *et al.*, 2006). Ainsi, l'adaptation des bactéries à des niches différentes entraîne des différences écologiques au sein des populations (Cohan, 2002). Ces différences écologiques peuvent séparer les populations sur le plan spatiotemporel (Cohan, 2002). De plus, les pressions de sélection exercées par l'hôte peuvent provoquer des divergences extrêmes et rapides entre les individus et jouer un rôle prééminent dans la spéciation (Orr et Smith, 1998). Le ratio Ka/Ks permet d'étudier les taux d'évolution des gènes et de connaître la nature de la pression de sélection exercée sur les gènes étudiés. Chez les souches appartenant à différents pathovars et espèces de *Xanthomonas*, la majorité des gènes impliqués dans le système de régulation à deux composants présentent des ratios Ka/Ks inférieurs à 1 (en moyenne 0,406) suggérant que le taux de remplacement des acides aminés est faible à cause de l'action de pression de sélection purificatrice. En revanche, cinq gènes codant un senseur, deux régulateurs et deux senseurs couplés à leurs régulateurs présentent des ratios Ka/Ks supérieurs à 2 suggérant que ces gènes sont soumis à une pression de sélection positive. Cette pression favorise l'accumulation de mutations génétiques à caractère adaptatif (Qian *et al.*, 2008).

Les bactéries pathogènes peuvent évoluer par perte de fonctions. Cette force évolutive favorise l'élimination et la répression de l'expression d'un gène dont le produit est incompatible avec la survie et la croissance d'un agent pathogène dans une niche écologique. Lorsque ce gène est impliqué dans le pouvoir pathogène de l'agent pathogène, il est qualifié de gène d'avirulence puisqu'il est incompatible avec l'expression du pouvoir pathogène du microorganisme. Selon le modèle de Flor (1971), les produits des gènes d'avirulence sont reconnus par des protéines de résistance impliquées dans les défenses de l'hôte. Cette reconnaissance aboutit à une interaction incompatible avec l'expression d'une réaction d'hypersensibilité par l'hôte. Dans ce cas, la pathoadaptation de la bactérie peut s'effectuer par l'élimination du gène d'avirulence par l'intermédiaire de mutations ponctuelles modifiant le gène (Maurelli, 2007). Cette forme de pathoadaptation se fait en réponse à une pression de sélection exercée par l'hôte qui entraîne la perte d'une fonction pour améliorer le fitness d'une bactérie dans son habitat (Maurelli, 2007 ; Pallen et Wren, 2007).

L'insertion d'éléments transposables comme les séquences d'insertion (IS) peut être à l'origine de réarrangements génomiques. Plusieurs bactéries pathogènes se caractérisent par la prolifération dans leurs génomes, d'éléments transposables et particulièrement les ISs (Pallen et Wren, 2007). L'insertion des ISs dans les génomes bactériens peut être aléatoire ou spécifique. Au niveau d'un gène, elle peut entraîner son inactivation (Mahillon et Chandler, 1998). En outre, l'abondance des ISs dans un génome favorise les recombinaisons homologues au sein du même génome ce qui peut engendrer des réarrangements bouleversant l'ordre ancestral des gènes préexistants (Pallen et Wren, 2007). Dans le cas de *Yersinia pestis*, une recombinaison entre des ISs engendre au niveau du segment génomique, des anomalies dans la composition en bases G et C (Parkhill *et al.*, 2001). En revanche, les effets de ces réarrangements sur la biologie et le pouvoir pathogène de l'organisme restent encore inconnus (Parkhill *et al.*, 2001 ; Pallen et Wren, 2007). Dans le cas de *R. solanacearum*, les ISs interviennent dans la variation de phase avec le passage d'une forme pathogène à une forme non pathogène et réciproquement (Poussier *et al.*, 2003). Plusieurs familles d'ISs ont été identifiées chez les *Xanthomonas*. La souche KACC10331 de *X. oryzae* pv. *oryzae* en possède 207 alors que la souche 306 de *X. axonopodis* pv. *citri*, la souche ATCC33913

de *X. campestris* pv. *campestris* et la souche 85-10 de *X. axonopodis* pv. *vesicatoria* en possèdent respectivement 109, 108 et 66 (da Silva *et al.*, 2002 ; Thieme *et al.*, 2005 ; Lee *et al.*, 2005). Une étude menée dans notre laboratoire, sur les répertoires d'ET3s de souches appartenant à *X. axonopodis*, a montré que la majorité des réarrangements identifiés (21/23) au niveau des gènes codant les ET3s, correspond à des insertions d'ISs (Hajri *et al.*, 2009). Dans le cas de *X. axonopodis* pv. *vesicatoria*, l'insertion d'une IS dans un gène codant un ET3 (*avrBs1*) peut s'accompagner d'un passage d'une interaction incompatible à une interaction compatible avec l'hôte (Kearney et Staskawicz, 1990).

3.2 Notion de répertoire de gènes

Lors de chacune des étapes de l'interaction (attraction, installation, multiplication et transmission) avec leurs hôtes, les bactéries font appel à différents répertoires de gènes. En effet, les gènes codant les MCPs et les adhésines interviennent dans les phases d'attraction et d'installation de la bactérie sur son hôte. Les gènes codant les facteurs du pouvoir pathogène interviennent lorsque la bactérie engage un processus infectieux (Sarkar *et al.*, 2006). Le séquençage de génomes bactériens et l'analyse par génomique comparative a montré que les répertoires de gènes sont variables entre les souches séquencées (da Silva *et al.*, 2002 ; Lavín *et al.*, 2007 ; Qian *et al.*, 2008 ; Hajri *et al.*, 2009). Ces répertoires peuvent en effet évoluer par le gain, la perte et les changements des séquences des gènes (Ochman et Moran, 2001 ; Hajri *et al.*, 2009).

La majorité des études sur les bactéries phytopathogènes est surtout axée sur les phases tardives de colonisation de l'hôte par la bactérie et non pas sur les phases précoces d'installation. Cet intérêt est dû en partie à la généralisation des procédures d'inoculation qui délivrent artificiellement et directement les bactéries dans les tissus de l'épiderme de la plante (Melotto *et al.*, 2006). De plus, l'expression ou non des symptômes par la plante hôte renseigne sur la nature de l'interaction entre la bactérie phytopathogène et son hôte (Sarkar *et al.*, 2006). En conséquence, les répertoires de gènes les plus étudiés concernent principalement les gènes impliqués dans les phases

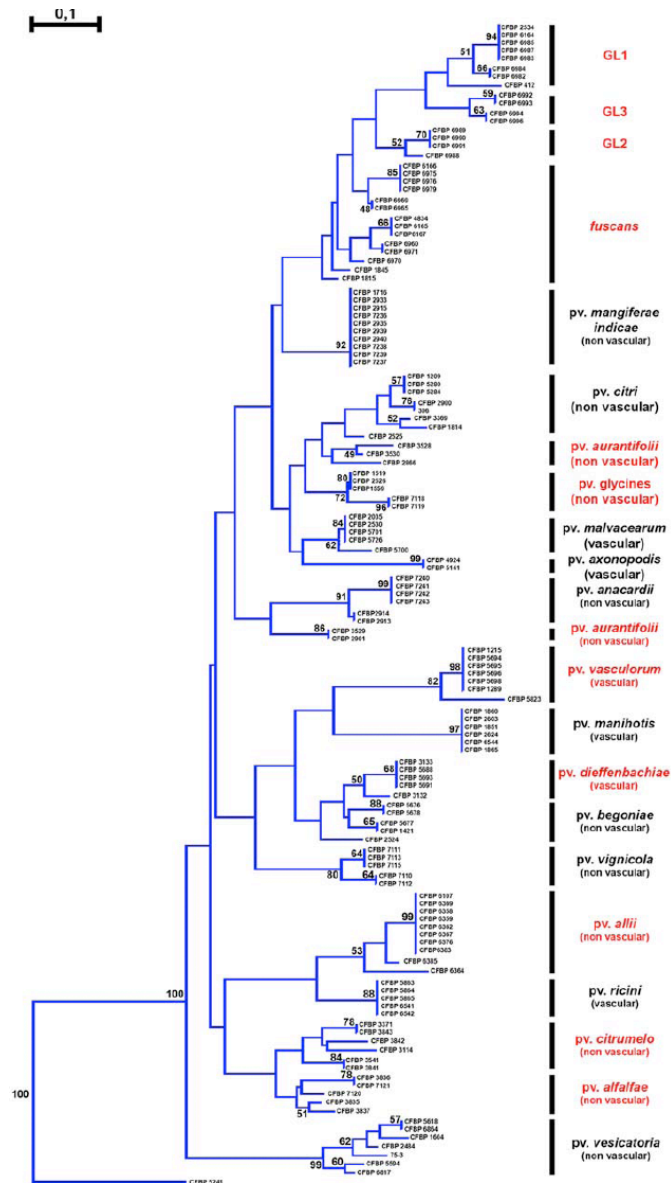


Figure 35. Dendrogramme obtenu à partir de la matrice présence/absence de 35 ET3s chez 132 souches appartenant à *X. axonopodis* (Hajri *et al.*, 2009). le dendrogramme a été construit par la méthode de Neighbor Joining utilisant les distances de Jaccard et enraciné avec la souche CFBP5241 de *X. campestris* pv. *campestris*. Les valeurs significatives (>50) de bootstraps (1000 simulations) sont mentionnées aux nœuds des branches du dendrogramme. Les pathovars polyphylétiques sont en rouge et les monophylétiques, en noir. GL1, GL2, GL3 et *fuscans* correspondent aux quatre lignées génétiques définies au sein du pathovar *phaseoli*.

tardives de l'interaction entre la bactérie et son hôte. C'est particulièrement le cas des répertoires des gènes codant les effecteurs du système de sécrétion de type 3 (ET3s). En effet, les ET3s sont considérés comme les facteurs les plus importants dans la spécificité d'hôte des bactéries phytopathogènes comme *P. syringae* et les *Xanthomonas* (Sarkar *et al.*, 2006 ; Hajri *et al.*, 2009) car ils sont impliqués dans la suppression ou l'induction des réactions de défenses et favorisent la multiplication et la transmission des bactéries (Sarkar *et al.*, 2006).

Les répertoires d'ET3s ont été déterminés par génomique comparative suite au séquençage de nombreux génomes bactériens (Joardar *et al.*, 2005 ; Almeida *et al.*, 2009). A titre d'exemple, les répertoires d'ET3s des souches DC3000 de *P. syringae* pv. *tomato* et de 1448A de *P. syringae* pv. *phaseolicola* sont différents et renferment des ET3s aussi bien fonctionnels que non fonctionnels (Schechter *et al.*, 2004 ; Joardar *et al.*, 2005). Par ailleurs, la comparaison des génomes des souches pathogènes de la tomate, DC3000 et T1 de *P. syringae* pv. *tomato*, révèle que leurs répertoires d'ET3s sont différents. Les répertoires de ces deux souches ont en communs 14 ET3s (Almeida *et al.*, 2009). L'étude de la distribution des ET3s a également été réalisée sur des collections de souches. Il a été démontré que les répertoires d'ET3s de *P. syringae*, *R. solanacearum* et *X. axonopodis*, combinent des gènes à distribution large et d'autres à distribution variable (Sarkar *et al.*, 2006 ; Guidot *et al.*, 2007 ; Hajri *et al.*, 2009). Concernant les *Xanthomonas*, les répertoires d'ET3s corrélaient bien avec la structuration en pathovars de *X. axonopodis* (Hajri *et al.*, 2009). En effet, le dendrogramme construit à partir de la matrice de présence/absence des ET3s chez les souches de *X. axonopodis* permet de regrouper les souches selon les pathovars auxquels elles appartiennent (Figure 35). Les résultats obtenus par Hajri *et al.* (2009) mettent en évidence chez des souches phylogénétiquement éloignées, des similarités dans les répertoires d'ET3s. Selon les auteurs, ce résultat pourrait traduire des convergences pathologiques. Ceci est illustré par l'exemple des 4 lignées génétiques (*fuscans*, GL1, GL2 et GL3) de *X. axonopodis* pv. *phaseoli*, pathogènes du haricot, qui sont phylogénétiquement distinctes mais qui présentent des répertoires d'ET3s semblables mais toutefois différenciables (Hajri *et al.*, 2009).

Une étude de génomique comparative de gènes codant des protéines du système de régulation à deux composants a également été entreprise sur trois souches de *P. syringae* ayant des gammes d'hôte différentes : la souche B728a du pathovar *syringae*, la souche DC3000 du pathovar *tomato* et la souche 1448A du pathovar *phaseolicola* (Lavín *et al.*, 2007). Cette étude a montré que les répertoires de gènes codant les protéines du système de régulation à deux composants sont différents entre les trois souches étudiées. Cette différence pourrait contribuer à la colonisation de gamme d'hôtes différentes par ces trois souches. A l'instar des répertoires d'ET3s, les répertoires de gènes codant des protéines du système de régulation à deux composants renferment des gènes présents chez les trois souches et d'autres gènes à distribution plus variable (Lavín *et al.*, 2007). Dans le cas de *R. solanacearum*, Guidot *et al.* (2007) ont montré que les répertoires des gènes codant les adhésines et les ET3s sont variables entre les souches qui sont phylogénétiquement très proches (appartenant à un même phylotype) et ayant des gammes d'hôtes différentes. De façon surprenante, la distribution de ces gènes variables corrèle bien avec la phylogénie des souches étudiées (Guidot *et al.*, 2007). Les auteurs suggèrent *i*) que ces gènes auraient été acquis par l'ancêtre et auraient eu la même évolution que le reste des gènes ou *ii*) que ces gènes auraient été acquis indépendamment par les différents phylotypes durant leur évolution et qu'ils n'auraient jamais été échangés par transfert horizontal. Guidot *et al.* (2007) présentent les adhésines comme étant de bons candidats pour la spécificité d'hôte. En effet, l'adhésion aux tissus de la plante hôte est la première étape dans l'interaction entre l'hôte et la bactérie pathogène et il est raisonnable de penser qu'il existe un certain degré de spécificité entre les adhésines et la surface des cellules de la plante hôte (Guidot *et al.*, 2007).

4. Bactéries modèles de l'étude : les *Xanthomonas* spp.

Le genre *Xanthomonas* regroupe 27 espèces de bactéries phytopathogènes attaquant plus de 400 plantes hôtes dont plusieurs sont d'intérêt économique majeur tels que le riz, les agrumes, la banane, la canne à sucre, le haricot, le chou, le coton, les céréales, etc. Ces plantes constituent l'alimentation de base et une source de revenus importante pour de nombreuses populations de par le monde. Par le biais des échanges commerciaux de semences et suite à l'émergence soudaine de populations phytopathogènes, les *Xanthomonas* sont à l'origine sur tous les continents d'épidémies récurrentes dont l'incidence socio-économique est majeure (Brown, 2001 ; Bafana, 2008). A titre d'exemple, en 1967, une forte épidémie de la graisse commune du haricot s'est déclarée dans l'état du Michigan aux Etats-Unis d'Amérique. L'agent causal de cette maladie, *X. axonopodis* pv. *phaseoli*, a infecté plus de 75% des 263 000 hectares plantés et a provoqué une baisse de rendement de l'ordre de 20% (Gilbertson et Maxwell, 1992). Le flétrissement bactérien du bananier dû à *X. vasicola* pv. *musacearum* a émergé en Ethiopie et s'est progressivement disséminé en Afrique de l'Est puis en Afrique centrale (Ouganda) où il s'est implanté avec force. En Ouganda, entre 2001 et 2004, 5 à 7% des agriculteurs ont abandonné la culture de banane suite aux épidémies de *X. vasicola* pv. *musacearum*. Les récoltes ont chuté de 65 à 80%. Des estimations ont prédit des pertes cumulées de 4 milliards de dollars américains, entre 2001 et 2010, si l'incidence de la maladie augmente de 8 à 15% par an selon les cultivars (Smith *et al.*, 2008). En Europe, ces observations ont conduit à l'inscription de plusieurs de ces organismes sur la liste des agents de quarantaine (annexes de la directive européenne 2000/29/CE) (*X. axonopodis* pv. *phaseoli*, *X. arboricola* pv. *pruni*, etc) et dans la réglementation des biens à double usage : civil et militaire (*X. albilineans*, les *Xanthomonas* pathogènes des *Citrus* et *X. oryzae* pv. *oryzae*) (CE 394/2006 du 27/02/2006 modifiant le règlement CE 1334/2000).

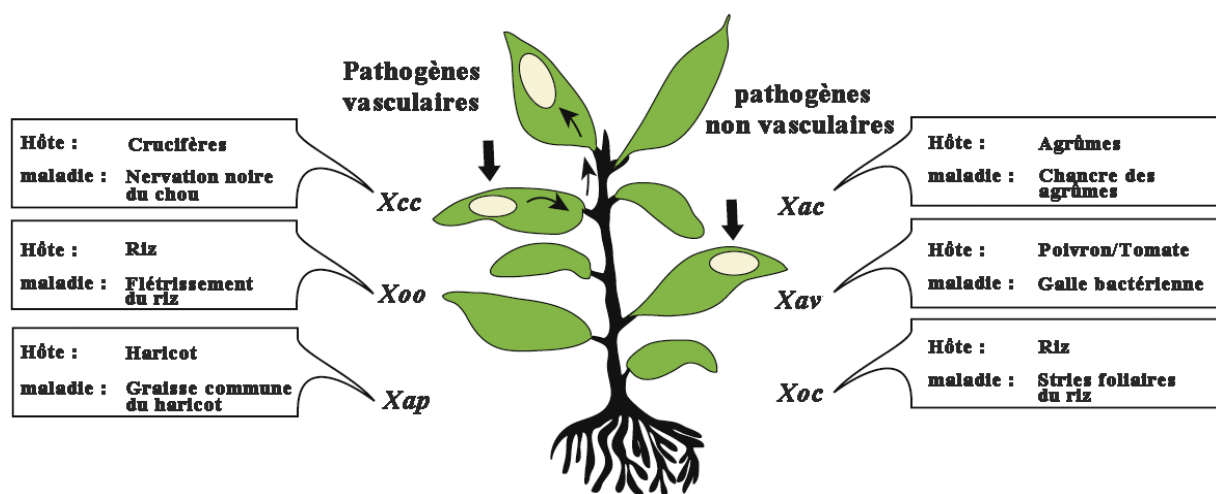


Figure 36. Les styles de vie des *Xanthomonas* spp. (d'après Büttner et Bonas, 2010). Exemples de bactéries appartenant au genre *Xanthomonas*, leurs plantes hôtes et les maladies qu'elles provoquent. *X. campestris* pv. *campestris* (*Xcc*), *X. oryzae* pv. *oryzae* (*Xoo*) et *X. axonopodis* pv. *phaseoli* (*Xap*) se propagent de manière systémique dans la plante hôte alors que *X. axonopodis* pv. *citri* (*Xac*), *X. axonopodis* pv. *vesicatoria* (*Xav*) et *X. oryzae* pv. *oryzicola* (*Xoc*) se multiplient localement au niveau des tissus foliaires (mésophile).

Au sein des *Xanthomonas*, chaque souche bactérienne possède une gamme d'hôte restreinte et ne provoque la maladie que sur un faible nombre de plantes hôtes. Toutes les souches responsables d'un même type de symptôme et sur une même gamme de plantes hôtes sont regroupées au sein d'un même pathovar (Young *et al.*, 1978). L'existence de plusieurs pathovars au sein de chaque espèce du genre *Xanthomonas* traduit globalement l'existence de plusieurs styles de vie au sein de ce genre et surtout une forte adaptation des souches à leurs hôtes (Figure 36). La majorité des *Xanthomonas* présente une colonisation épiphyte puis pénètrent dans la plante *via* les ouvertures naturelles (hydathodes, stomates et blessures). A l'intérieur de la plante, les *Xanthomonas* se multiplient localement dans les espaces intercellulaires ou colonisent les vaisseaux conducteurs du xylème et se propagent de manière systémique dans la plante (Figure 36). Les *Xanthomonas* constituent ainsi un bon modèle pour l'étude des bases génétiques et moléculaires responsables *in fine* de l'adaptation à l'hôte.

4.1 Vie épiphyte des *Xanthomonas* spp.

Les cellules de *Xanthomonas* spp. sont généralement déposées à la surface foliaire par l'action du vent, des éclaboussures d'eau et les bioaérosols (Rudolph, 1993). Les populations épiphytes de *Xanthomonas* peuvent également provenir de l'exsudation des bactéries à travers les ouvertures naturelles et des lésions qu'elles peuvent provoquer comme c'est le cas pour *X. axonopodis* pv. *citri* (Timmer *et al.*, 1991). Nombre d'entre elles, transmises par les semences, colonisent les jeunes plantules au moment de la levée (Maude, 1996 ; Weller et Saettler, 1980 ; Darrasse *et al.*, 2010). Les *Xanthomonas* colonisent efficacement la surface foliaire de la plante hôte lorsque les conditions sont favorables. Elles colonisent préférentiellement les sites riches en nutriments. Par exemple, les cellules de *X. oryzae* pv. *oryzae* se localisent autour des hydathodes 24h après inoculation. Par ailleurs, l'adhésion des *Xanthomonas* aux tissus de leur plante hôte, favorise le maintien des populations épiphytes. En effet, les souches appartenant à *X. oryzae* pvs. *oryzae* et *oryzicola* adhèrent bien aux trichomes des feuilles de maïs (Timmer *et al.*, 1987). De même, les souches de *X. axonopodis* pvs. *citri* et *phaseoli* forment des biofilms à la surface des feuilles de limetier et de haricot (Jacques *et al.*, 2005 ; Rigano *et al.*, 2007). Les populations épiphytes de *Xanthomonas*

varient considérablement en fonction des conditions environnementales. Les populations de *X. axonopodis* pv. *citri* oscillent selon le cycle jour/nuit. En effet, la taille des populations est comprise entre 10^4 et 10^6 cellules par feuille tôt le matin et diminue pour être comprise entre 10^2 et 10^5 à la mi-journée (Timmer, 1989). Il a été montré que suite aux événements pluvieux, les populations de cellules solitaires de *X. axonopodis* pv. *phaseoli* augmentent (Jacques *et al.*, 2005). Les *Xanthomonas* spp. peuvent survivent à travers les saisons en association avec des plantes pérennes et les semences et par conséquent, constituer une source potentielle d'inoculum (Rudolph, 1993). Les souches appartenant à *X. axonopodis* pv. *vignicola* peuvent se maintenir en association avec les débris végétaux, les organes infectés, les semences et les hôtes alternatifs (Sikirou et Wydra, 2004). Les *Xanthomonas* sont aussi capables de coloniser la surface foliaire sans provoquer l'expression de symptômes (Stall *et al.*, 1993 ; Weller et Saettler, 1980 ; Jacques *et al.*, 2005 ; Darrasse *et al.*, 2007). Enfin, il a été reporté qu'en conditions naturelles, *X. axonopodis* pvs. *allii* et *phaseoli*, peuvent se maintenir au niveau de la phyllosphère de plantes non hôtes, respectivement le haricot et l'oignon (Gent *et al.*, 2005). En conditions contrôlées, *X. campestris* pv. *campestris* peut également survivre et se maintenir sur les feuilles du haricot avec des densités de populations élevées (1×10^5 CFU (colony-forming units) /g) (Darsonval *et al.*, 2008).

4.2 Les hôtes des *Xanthomonas* spp.

Les plantes cultivées sont les hôtes d'isolement majoritaires des *Xanthomonas* phytopathogènes et sont les principaux hôtes connus de ces bactéries. Dans quelques cas, d'autres hôtes -adventices- peuvent avoir été recensés. C'est le cas de *Strophostyles helvola* assurant la survie de *X. axonopodis* pv. *phaseoli* (Vidaver, 1993). Des plantes peuvent de plus avoir été identifiées comme hôtes potentiels à la suite d'inoculations conduites en conditions contrôlées (Cafati et Saettler, 1980). Les validations de ces informations en situations naturelles font alors généralement défaut. Il est ainsi très difficile d'établir des listes exhaustives des hôtes des différentes souches des *Xanthomonas* spp.

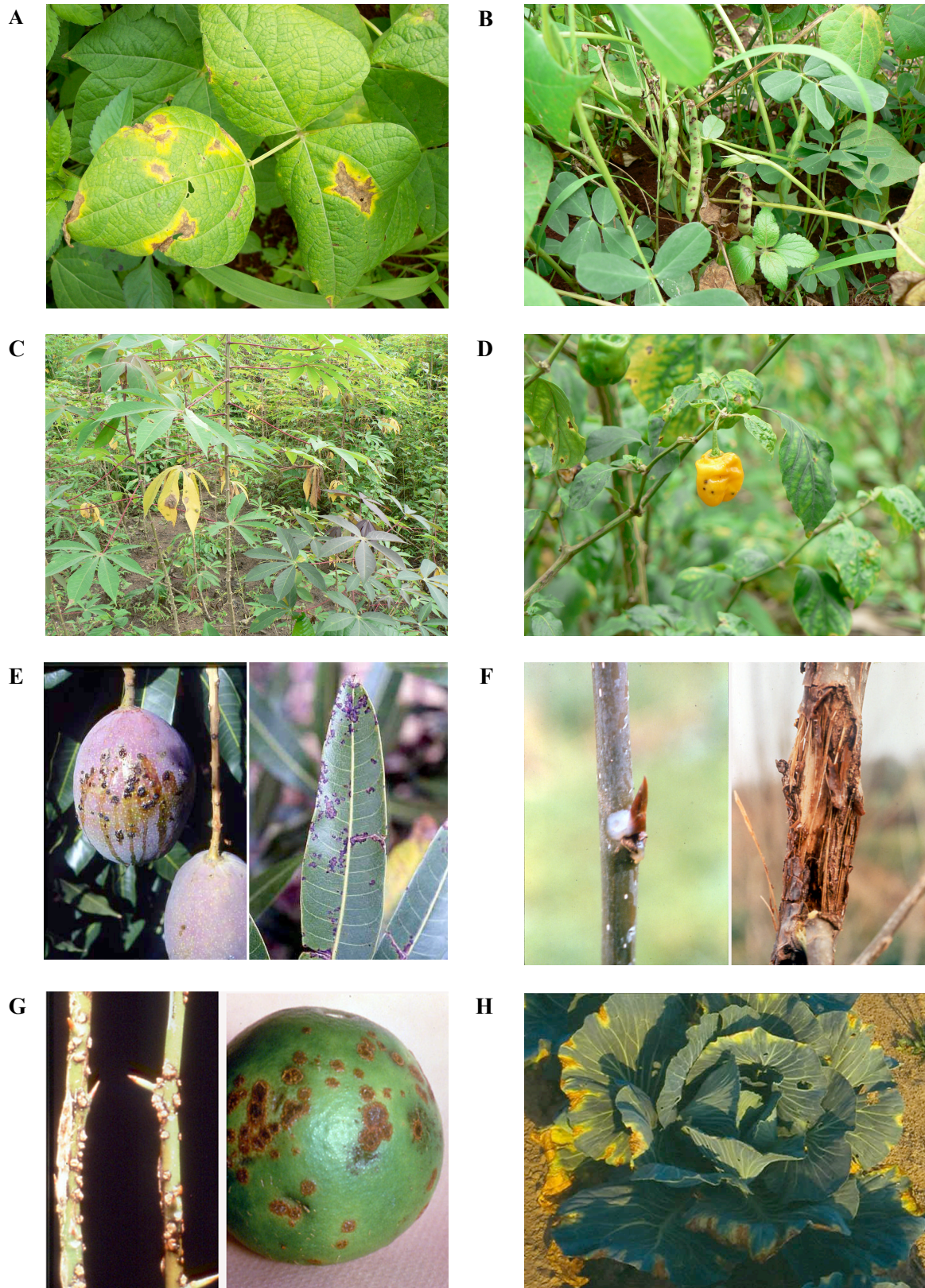


Figure 37. Symptômes de la graisse commune du haricot sur feuilles (A) et sur gousses (B), des taches foliaires du manioc (C), de la galle bactérienne du poivron (D), des taches noires du manguiers sur fruit et sur feuilles* (E), du chancre bactérien du peuplier (F), du chancre asiatique des agrumes sur tiges et fruit (G) et de la nervation noire du chou (H) (Photos INRA sauf * CIRAD).

Les *Xanthomonas* spp. sont des bactéries exclusivement associées aux plantes. Il n'existe pas à ma connaissance de rapport documentant une association étroite avec des cellules animales ou une aptitude à la colonisation du sol nu et à la survie libre en eau douce. Cependant, il a été rapporté la vécation de *X. oryzae* et *X. campestris* par des tardigrades (*Macrobiotus hufelandi*), celle de *X. axonopodis* pv. *citri* par la mineuse des agrumes (*Phyllocnistis citrella*), celle de *X. axonopodis* pv. *phaseoli* par des coléoptères (*Ceratoma ruficornis* et *Diaprepes abbreviatus*) et celle de *X. vasicola* pv. *musacearum* par des himénoptères (*Apis mellifera*) (Kaiser et Vakili, 1978 ; Krantz *et al.*, 1999 ; van Der Wolf *et al.*, 2006). Néanmoins, ces cas de dispersion semblent être ponctuels et passifs ne faisant pas appel à une interaction entre l'insecte et la bactérie. L'insecte ne semble pas non plus être un hôte permettant la survie de la bactérie. La récente découverte du SST3 de type SPI-1 chez *X. albilineans* et la lignée génétique GL1 de *X. axonopodis* pv. *phaseoli* (Alavi *et al.*, 2008 ; Pieretti *et al.*, 2009 ; Marguerettaz *et al.*, 2010) généralement impliqués dans les interactions entre les bactéries et les cellules eucaryotes animales pourraient néanmoins apporter un éclairage nouveau sur le rôle des insectes dans le cycle de vie des *Xanthomonas* spp. A l'instar de ce qui est connu chez les bactéries pathogènes de l'homme, le rôle de ces environnements alternatifs pourrait être majeur en terme d'acquisition de facteurs de virulence et de résistance aux bactéricides (Cangelosi *et al.*, 2004).

4.3 Symptômes et maladies causés par les bactéries appartenant au genre *Xanthomonas*

Les souches du genre *Xanthomonas* se caractérisent par une grande diversité pathologique. En effet, elles infectent approximativement 124 monocotylédones et 268 dicotylédones. En outre, elles provoquent de nombreux symptômes tels que des nécroses, chancres, tâches foliaires, chloroses, défoliations, flétrissement, hypertrophies et hyperplasies et affectent différents organes aériens de la plante hôte tels que feuilles, tiges et fruits (Figure 37) (Rudolph, 1993 ; Meyer et Bogdanove, 2009). Individuellement, chaque souche du genre *Xanthomonas* infecte une gamme étroite de plantes hôtes. Les *Xanthomonas* sont généralement des bactéries épiphytes colonisant la phyllosphère de leur hôte avant de pénétrer dans la plante via les ouvertures naturelles

Tableau 3. Liste non exhaustive des maladies causées par les bactéries membres du genre *Xanthomonas* spp.

Souches de <i>Xanthomonas</i> spp.	Maladies provoquées
<i>X. axonopodis</i> pv. <i>phaseoli</i>	Graisse commune du haricot
<i>X. axonopodis</i> pv. <i>malvacearum</i>	Taches foliaires du coton
<i>X. axonopodis</i> pv. <i>alfalfae</i>	Taches foliaires des légumineuses
<i>X. axonopodis</i> pv. <i>allii</i>	Dépérissement bactérien des alliacées
<i>X. axonopodis</i> pv. <i>citri</i>	Chancre citrique
<i>X. axonopodis</i> pv. <i>vesicatoria</i>	Galle bactérienne du poivron et de la tomate
<i>X. axonopodis</i> pv. <i>manihotis</i>	Taches foliaires du manioc
<i>X. axonopodis</i> pv. <i>mangiferaeindicae</i>	Taches noires du manguiier
<i>X. axonopodis</i> pv. <i>vasculorum</i>	Gommose de la canne à sucre
<i>X. axonopodis</i> pv. <i>glycines</i>	Maladie des pustules bactériennes du soja
<i>X. campestris</i> pv. <i>campestris</i>	Nervation noire du chou
<i>X. campestris</i> pv. <i>incanae</i>	Dépérissement des giroflées
<i>X. campestris</i> pv. <i>raphani</i>	Taches foliaires du radis
<i>X. oryzae</i> pv. <i>oryzae</i>	Stries foliaires et flétrissement du riz
<i>X. oryzae</i> pv. <i>oryzicola</i>	Stries foliaires et flétrissement du riz
<i>X. vasicola</i> pv. <i>musacearum</i>	Flétrissement bactérien du bananier
<i>X. vasicola</i> pv. <i>vasculorum</i>	Gommose de la canne à sucre
<i>X. vasicola</i> pv. <i>holcicola</i>	Taches foliaires du sorgho
<i>X. arboricola</i> pv. <i>pruni</i>	Taches bactériennes du prunier, de l'abricotier, du cerisier et du pêcher
<i>X. arboricola</i> pv. <i>juglandis</i>	Bactériose classique du noyer et Chancre vertical suintant
<i>X. arboricola</i> pv. <i>corylina</i>	Maladie de Filbert du noisetier
<i>X. translucens</i> pv. <i>hordei</i>	Taches foliaires de l'orge
<i>X. translucens</i> pv. <i>translucens</i>	Taches foliaires du blé
<i>X. translucens</i> pv. <i>undulosa</i>	Taches foliaires du blé dur
<i>X. translucens</i> pv. <i>secalis</i>	Taches foliaires du seigle
<i>X. fragariae</i>	Taches foliaires et flétrissement du fraisier
<i>X. cynareae</i>	Graisse de l'artichaut



Figure 38. Bactéries du genre *Xanthomonas* pourvues d'un flagelle polaire (<http://nsw-rsc.ricethailand.go.th/>).

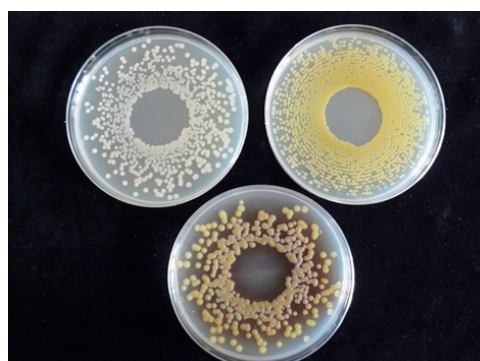


Figure 39. Diversité des colonies bactériennes du genre *Xanthomonas* sur un milieu gélosé riche. A droite, des colonies jaunes typiques du genre *Xanthomonas*. A gauche, des colonies blanches de *X. axonopodis* pv. *mangiferaeindicae* ne produisant pas de xanthomonadine. Au centre, des colonies de *X. axonopodis* pv. *phaseoli* GL. *fuscans* produisant un pigment brun dit fuscans (photo INRA).

comme les hydathodes, les stomates ou les blessures. Certaines souches de *Xanthomonas* spp. ont cependant des compétences épiphytes très limitées sur leur hôte (Pruvost *et al.*, 2009). A l'intérieur de la plante hôte, les *Xanthomonas* se caractérisent par une spécificité tissulaire ; elles se multiplient en effet dans les espaces intercellulaires et/ou colonisent les vaisseaux conducteurs (Meyer et Bogdanove, 2009 ; Büttner et Bonas, 2010). Le tableau 3 présente des maladies causées par différentes souches appartenant au genre *Xanthomonas*.

4.4 Caractéristiques cellulaires et physiologiques des *Xanthomonas* spp.

La diversité pathologique des *Xanthomonas* s'oppose à l'uniformité observée sur les plans morphologique et physiologique au sein de ce genre. Les *Xanthomonas* spp. sont phénotypiquement très homogènes. Les cellules ont une forme bacille de 0,4 à 0,7 µm de large et de 0,7 à 2 µm de long. Elles sont mobiles et pourvues d'un unique flagelle polaire (Figure 38). La plupart des colonies sont pigmentées grâce à la production d'un pigment jaune photo-protecteur et antioxydant à base de caroténoïdes, appelé xanthomonadine (Rajagopal *et al.*, 1997 ; Meyer et Bogdanove, 2009 ; Büttner et Bonas, 2010). D'ailleurs, le nom « *Xanthomonas* » attribué à ces souches provient du grec *Xanthos* qui veut dire jaune et *monas*, entité (Figure 39). Cependant, quelques souches de *Xanthomonas* spp. ne produisent pas de xanthomonadine et forment par conséquent des colonies blanches comme c'est par exemple le cas de *X. axonopodis* pv. *mangiferaeindicae* (Figure 39) (Gagnevin et Pruvost, 2001). Sur milieu de culture, les colonies bactériennes présentent un aspect muqueux qui est dû à la production massive d'un exopolysaccharide (EPS) appelé xanthane (Sutherland, 1993 ; Meyer et Bogdanove, 2009 ; Büttner et Bonas, 2010). Cet EPS est associé à l'agressivité de *X. campestris* pv. *campestris* ; il protège les cellules bactériennes contre les stress environnementaux tels que les radiations UV, le froid et la dessiccation. Par ailleurs, le xanthane est énormément utilisé en industrie agroalimentaire, cosmétique et pharmaceutique pour les propriétés qu'il détient. Le xanthane est en effet, une gomme visqueuse présentant des propriétés d'agent épaississant et stabilisant (Meyer et Bogdanove, 2009 ; Büttner et Bonas, 2010).

Tableau 4. Caractéristiques générales des génomes séquencés de *Xanthomonas* spp. (d'après Meyer et Bogdanove, 2009 ; Pieretti *et al.*, 2009 ; Moreira *et al.*, 2010).

Nom de la souche	Taille du génome (Mb)	G+C (%)	Nombre de plasmides	Régions codantes (%)	Séquences codantes (nombre de CDS)	Nombre d'ISs	Références
<i>X. axonopodis</i> pv. <i>citri</i> 306	5,27	64,8	2	85,6	4313	108	da Silva <i>et al.</i> , 2002
<i>X. axonopodis</i> pv. <i>vesicatoria</i> 85-10	5,42	64,6	4	86,6	4487	66	Thieme <i>et al.</i> , 2005
<i>X. axonopodis</i> pv. <i>aurantifolii</i> ICPB11122	4,88	64,9	inconnu	NA	3804	NA	Moreira <i>et al.</i> , 2010
<i>X. axonopodis</i> pv. <i>aurantifolii</i> ICPB10535	5,01	64,8	inconnu	NA	3921	NA	Moreira <i>et al.</i> , 2010
<i>X. campestris</i> pv. <i>campestris</i> ATCC33913	5,08	65,1	0	90,1	4181	109	da Silva <i>et al.</i> , 2002
<i>X. campestris</i> pv. <i>campestris</i> 8004	5,15	65,0	0	87,2	4273	115	Qian <i>et al.</i> , 2005
<i>X. campestris</i> pv. <i>campestris</i> B100	5,08	65,0	0	70,0	4471	59	Vorhölter <i>et al.</i> , 2008
<i>X. campestris</i> pv. <i>armoraciae</i> 756C	4,94	65,3	0	85,3	4534	NA	A. Bogdanove, non publié
<i>X. oryzae</i> pv. <i>oryzae</i> KACC10331	4,94	63,7	0	87,6	4637	714	Lee <i>et al.</i> , 2005
<i>X. oryzae</i> pv. <i>oryzae</i> MAFF311018	4,94	63,7	0	83,9	4372	712	Ochiai <i>et al.</i> , 2005
<i>X. oryzae</i> pv. <i>oryzae</i> PX099 ^A	5,24	63,6	0	83,0	5083	683	Salzberg <i>et al.</i> , 2008
<i>X. oryzae</i> pv. <i>oryzicola</i> BLS256	4,83	64,0	0	86,0	4554	NA	A. Bogdanove, non publié
<i>X. albilineans</i> GPE PC73	3,77	62,0	3	84,0	3115	NA	Pieretti <i>et al.</i> , 2009

Sur le plan physiologique, les *Xanthomonas* ont un métabolisme respiratoire de type aérobie strict. Leur croissance optimale se fait à des températures comprises entre 25 et 30°C (Saddler et Bradbury, 2005). Ces bactéries sont catalase positive, oxydase négative ou faiblement positive, uréase négative et indole négative ; elles sont incapables de réduire les nitrates (Bradbury, 1984). Elles se distinguent des autres bactéries phytopathogènes par la présence de trois acides gras qui leur sont uniques : 11 :0 iso, 11 :0 iso 3OH et 13 :0 iso 3OH (Yang *et al.*, 1993).

4.5 Caractéristiques génomiques des *Xanthomonas* spp.

Les génomes de 18 souches (<http://www.xanthomonas.org/>) appartenant à 5 espèces et 11 pathovars ont été entièrement séquencés. Parmi ces génomes, 9 sont actuellement disponibles dans les bases de données publiques (National Center for Biotechnology Information, NCBI) (Tableau 4). Les séquences partielles des génomes des souches BLS256 de *X. oryzae* pv. *oryzicola* (A. Bogdanove, données non publiées), ICPB11122 et ICPB10535 de *X. axonopodis* pv. *aurantifolii* (Moreira *et al.*, 2010), NCPPB4381 de *X. vasicola* pv. *musacearum* et NCPPB702 de *X. vasicola* pv. *vasculorum* (Studholme *et al.*, 2010) sont également disponibles sur NCBI. Enfin, les génomes d'une autre souche de *X. vasicola* pv. *musacearum* et de la souche 756C de *X. campestris* pv. *armoraciae* -synonyme de *raphani*- ne sont pas encore disponibles (A. Bogdanove, données non publiées). Par ailleurs au moins deux autres génomes ont été séquencés et sont en cours d'annotation ; il s'agit des séquences génomiques des souches CFBP4834 de *X. axonopodis* pv. *phaseoli* (M-A. Jacques, données non publiées) et CIO151 de *X. axonopodis* pv. *manihotis* (A. Bernal, données non publiées).

Les caractéristiques des séquences génomiques complètes des souches séquencées de *Xanthomonas* sont généralement similaires (Tableau 4). La taille est comprise entre 4,83 Mb à 5,42 Mb, le contenu en G+C% est compris entre 63,6% et 65,3% et le nombre de gènes varie entre 3804 et 5083. En revanche, le génome de la souche GPE PC73 de *X. albilineans* est réduit. Il se distingue des autres génomes par sa petite taille (environ 3,77 Mb). De plus, il ne contient que 3115 gènes, ne possède pas de cluster *hrp* codant le système de sécrétion de type 3 et son contenu en G+C% est de

62% (Pieretti *et al.*, 2009). Le contenu des gènes dans les génomes des souches séquencées est largement conservé. Néanmoins, les alignements complets de ces séquences génomiques révèlent l'existence de nombreuses inversions, insertions, délétions et réarrangements génétiques (Figure 32) (Thieme *et al.*, 2005 ; Meyer et Bogdanove, 2009). L'abondance des séquences d'insertions (IS) dans les génomes de *Xanthomonas* suggère qu'elles joueraient un rôle important dans les transferts horizontaux de gènes et par conséquent dans l'évolution de leurs génomes (Meyer et Bogdanove, 2009).

Des plasmides ont été identifiés dans certains génomes de *Xanthomonas*. Ils sont au nombre de quatre dans la souche 85-10 de *X. axonopodis* pv. *vesicatoria* (Thieme *et al.*, 2005), deux dans la souche 306 de *X. axonopodis* pv. *citri* (da Silva *et al.*, 2002) et de trois dans les souches GPE PC73 de *X. albilineans* et CFBP4834 de *X. axonopodis* pv. *phaseoli* (Pieretti *et al.*, 2009 ; M-A. Jacques, données non publiées). Les souches séquencées ICPB11122 et ICPB10535 de *X. axonopodis* pv. *aurantifolii* possèdent des plasmides mais leur nombre reste inconnu (Moreira *et al.*, 2010). Par contre, les souches séquencées de *X. campestris* pv. *campestris* sont dépourvues de plasmides (da Silva *et al.*, 2002 ; Qian *et al.*, 2005 ; Vorhölter *et al.*, 2008).

La comparaison des génomes de souches appartenant à différents pathovars et espèces de *Xanthomonas* ayant des spécificités tissulaires et écologiques différentes a été réalisée afin d'élucider les déterminants génétiques impliqués dans l'adaptation à l'hôte (Qian *et al.*, 2005 ; Meyer et Bogdanove, 2009). La comparaison des génomes des souches appartenant aux pathovars *citri* et *aurantifolii* de *X. axonopodis* a permis d'identifier certains ET3s responsables de la pathogénie sur un même hôte (Moreira *et al.*, 2010). La comparaison des génomes des souches appartenant aux pathovars *campestris* et *raphani* de *X. campestris* d'une part, et celle des souches des pathovars *oryzae* et *oryzicola* de *X. oryzae* d'autre part, permettrait d'appréhender les déterminants des spécificités tissulaires. L'idée émergente de ces analyses est que l'adaptation à l'hôte des souches bactériennes très proches écologiquement se fait plutôt par le biais de mutations ponctuelles, d'insertions et de délétions que par des différences dans le contenu des gènes (Meyer et Bogdanove, 2009).

4.6 Classification et évolution de la taxonomie des *Xanthomonas* spp.

La première référence écrite à une maladie due à un *Xanthomonas* remonte à 1883 par la découverte de Wakker d'une maladie jaune sur les hyacinthes causée par *Bacterium hyacinthi*, une bactérie formant des colonies jaunes (Wakker, 1883). Plusieurs autres bactéries du même type ont été également isolées par la suite. Ces bactéries comme toutes les autres bactéries pathogènes de plantes, étaient regroupées dans la tribu des *Erwinieae* dans la première édition du *Bergey's Manual of Determinative Bacteriology* (1923) (Bergey *et al.*, 1923). Cette tribu était à son tour subdivisée en deux genres : *Erwinia* et *Phytomonas*. Une étude comparative des différentes bactéries appartenant au genre *Phytomonas* menée par Burkholder en 1930, a permis de proposer des regroupements d'espèces. Le genre *Xanthomonas* formait l'un de ces groupes (Dawson, 1939).

4.6.1 Taxonomie évolutive du genre *Xanthomonas* spp.

Pendant longtemps, la taxonomie bactérienne s'est basée sur le pouvoir pathogène des bactéries. Dans le cas des *Xanthomonas*, chaque variant ayant une gamme de plantes hôtes différente et/ou induisant des symptômes différents était considéré comme une espèce à part entière. Ceci a abouti à un accroissement considérable du nombre d'espèces (environ une centaine) au sein du genre *Xanthomonas* qui découle du concept « new host-new species » (Starr, 1981). En 1974, le nombre d'espèces au sein du genre *Xanthomonas* a été réduit à cinq sur la base des caractéristiques phénotypiques : *X. campestris* (espèce type), *X. fragariae*, *X. albilineans*, *X. axonopodis* et *X. ampelina* (Dye et Lelliott, 1974). En 1978, Ridé et Ridé ont intégré sur la base des caractères morphologiques, biochimiques et sérologiques, l'agent causal du chancre du peuplier, *Aplanobacter populi*, dans le genre *Xanthomonas* sous le nom de *X. populi*. Les travaux sur les caractères phénotypiques et génotypiques menés sur *P. maltophilia* ont conduit Swings *et al.* (1983) à classer cette espèce dans le genre *Xanthomonas* sous le nom de *X. ampelina*. La position taxonomique de *X. ampelina* a également été

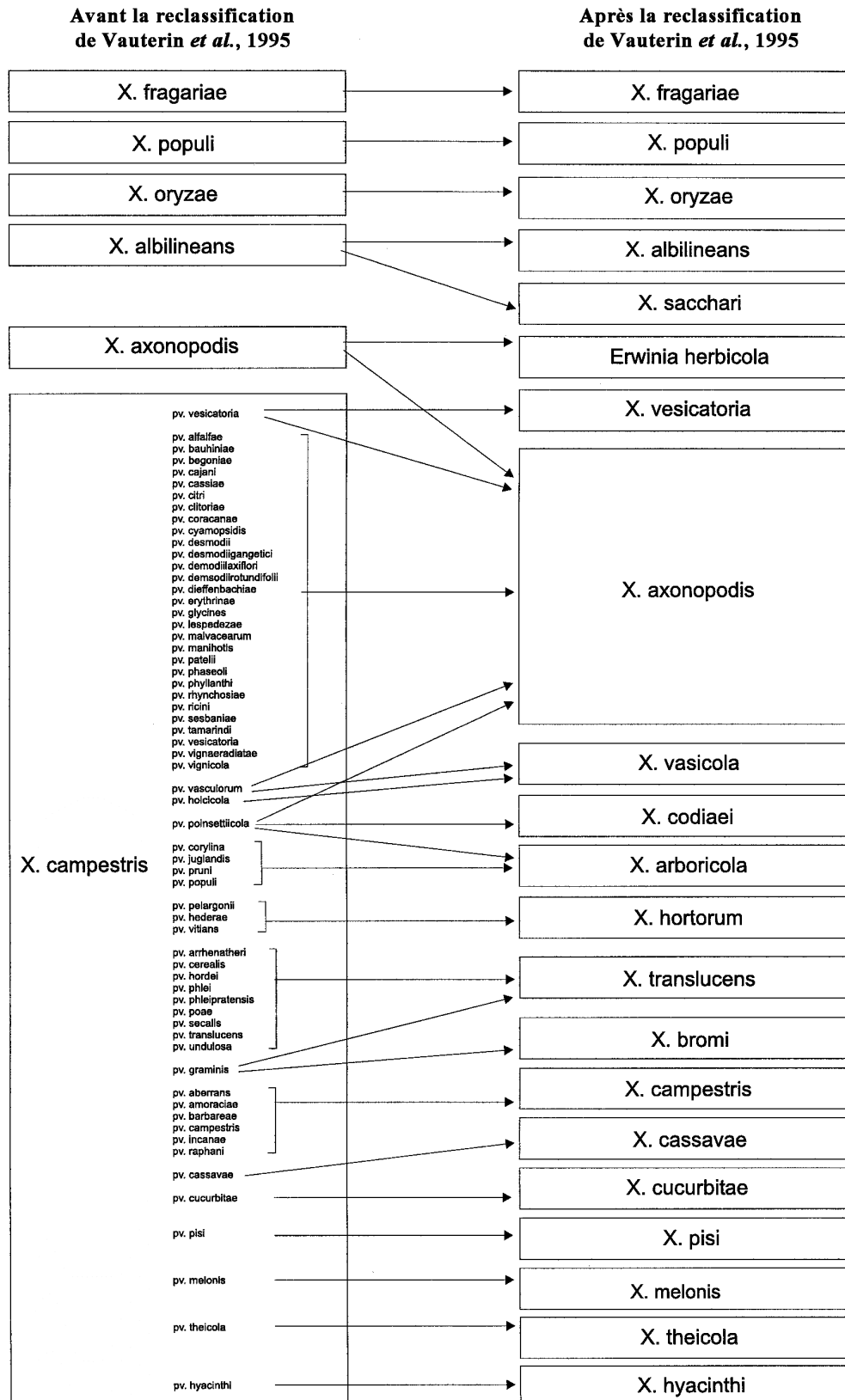


Figure 40. Représentation schématique des réarrangements taxonomiques au sein du genre *Xanthomonas* sur la base des hybridations ADN-ADN (d'après Vauterin *et al.*, 2000). Les espèces *X. fragariae*, *X. albilineans*, *X. oryzae* et *X. populi* sont réapprouvées. L'espèce *X. campestris* a été scindée en 16 espèces.

changée. Par une approche polyphasique (électrophorèse de protéines et hybridations ADN-ADN et ADN-ARNr), elle a été élevée à un rang taxonomique supérieur formant le nouveau genre *Xylophilus* sous le nom de *Xylophilus ampelinus* (Willems *et al.*, 1987). En 1990, Van den Mooter et Swings ont répertorié sur la base d'une étude des caractères morphologiques, physiologiques et biochimiques, huit espèces au sein du genre *Xanthomonas* : *X. campestris*, *X. fragariae*, *X. albilineans*, *X. axonopodis*, *X. graminis*, *X. oryzae*, *X. populi* et *X. maltophilia* (Van den Mooter et Swings, 1990). Parmi ces espèces, *X. maltophilia* a été reclassée dans le genre *Stenotrophomonas* sous le nom de *S. maltophilia* (Palleroni et Bradbury, 1993). L'espèce type *X. campestris* est restée la plus complexe ; elle regroupe entre autre des espèces nouvellement décrites mais non validées par les *Approved Lists of Bacterial Names*. L'espèce *X. campestris* a été subdivisée en 140 pathovars sur la base de leur spectre d'hôte suite à l'adoption du terme pathovar (Young *et al.*, 1978 ; Hayward, 1993). Il est à noter que la nomenclature utilisant le terme pathovar n'a pas été acceptée dans les *Approved Lists of Bacterial Names*. Elle a en revanche été adoptée comme une solution provisoire en attente d'une classification établie sur des principes acceptés (Vauterin et Swings, 1997).

4.6.2 Une taxonomie basée sur des techniques moléculaires

Suite à l'expansion rapide des techniques de génétique moléculaire, une nouvelle classification des *Xanthomonas* employant des approches polyphasiques intégrant des données phénotypiques et génotypiques a été proposée. Les techniques d'hybridation ADN-ADN (Vauterin *et al.*, 1995 ; Rademaker *et al.*, 2000), de génotypage par AFLP (amplified fragment-length polymorphism) (Rademaker *et al.*, 2000), d'amplification de séquences répétées (rep-PCR) (Rademaker *et al.*, 2000, 2005) et l'étude de la phylogénie basée sur l'amplification de fragments de gènes de ménage (*atpD*, *dnaK*, *efp*, *fyuA*, *gyrB* and *rpoD*) (Parkinson *et al.*, 2007, 2009 ; Young *et al.*, 2008 ; Ah-You *et al.*, 2009 ; Hajri *et al.*, 2009 ; Bui Thi Ngoc *et al.*, 2010) ont été appliquées au genre *Xanthomonas*. Elles ont permis l'obtention d'une classification des souches de *Xanthomonas* plus compréhensible.

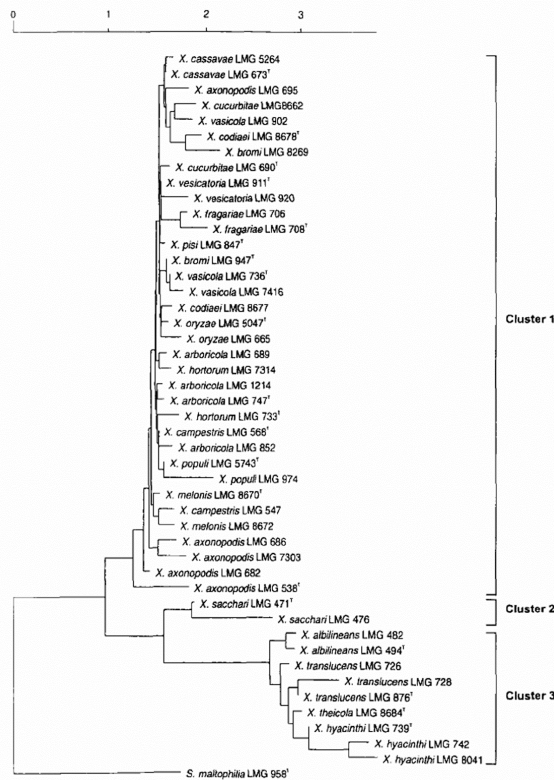


Figure 41. Arbre phylogénétique construit avec la méthode Neighbour-Joining sur la base des séquences partielles de l'ADNr 16S et représentant les relations entre des souches de *Xanthomonas* spp. (d'après Hauben *et al.*, 1997). Trois clusters ont été identifiés. Le premier est majoritaire ; il comprend 15 espèces. Les deux autres clusters (2 et 3) regroupent les 5 espèces restantes (*X. translucens*, *X. hyacinthi*, *X. theicola*, *X. sacchari* et *X. albilineans*).

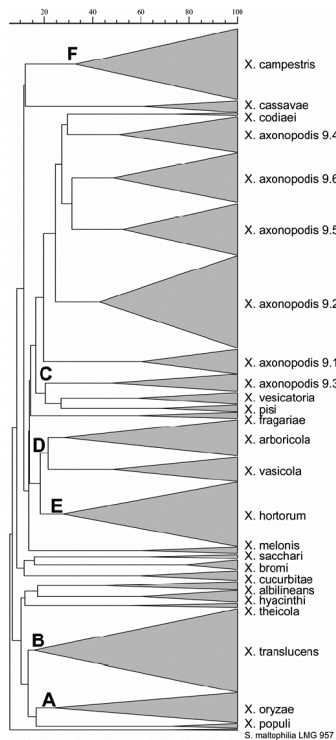


Figure 42. Analyse UPGMA combinées des profils BOX, ERIC et REP des souches de *Xanthomonas* spp. (d'après Rademaker *et al.*, 2005). Les 339 souches ont été groupées dans les différentes espèces définies par Vauterin *et al.* (1995) sur la base des hybridations ADN-ADN. Les clusters 9.1 à 9.6 se réfèrent aux 6 groupes stables au sein de *X. axonopodis*.

Vauterin et ses associés ont réformé en 1995 la taxonomie des *Xanthomonas* spp. moyennant les hybridations ADN-ADN. Ils ont identifié 20 espèces au sein du genre *Xanthomonas* (Figure 40). Ils confirment quatre espèces décrites auparavant : *X. fragariae*, *X. albilineans*, *X. oryzae* et *X. populi*. L'espèce type *X. campestris* a été scindée en 16 espèces et ne compte plus que six pathovars (pvs. *aberrans*, *armoraciae*, *barbareae*, *campestris*, *incaneae* et *raphani*) au lieu des 140 qui lui étaient attribués. Par ailleurs, environ 34 pathovars ont été assignés à l'espèce *X. axonopodis* (Vauterin *et al.*, 1995). Les travaux de Vauterin *et al.* (1995) ont permis d'obtenir la classification la plus représentative de la diversité génomique et des relations génétiques au sein du genre *Xanthomonas* en élevant de nombreux pathovars de *X. campestris* au rang d'espèce ou en les plaçant dans des espèces préexistantes (Vauterin *et al.*, 1995). En outre, des pathovars préexistants se retrouvent scindés dans des espèces différentes (ex. *vesicatoria*, *vasculorum*, *graminis* et *poinsetticola*). Néanmoins, ces travaux n'ont pas permis l'assignation exacte de tous les pathovars. En effet, une soixantaine de pathovars qualifiés d'orphelins n'ont pas été assignés à une espèce donnée et sont restés attribués par défaut à l'espèce *X. campestris*.

Par ailleurs, l'analyse de séquences partielles de l'ADN ribosomal 16S montre une faible résolution taxonomique en raison de la conservation de ces séquences au sein des *Xanthomonas* (Figure 41). Seulement trois groupes ont pu être distingués ; le premier groupe majoritaire regroupe 15 espèces tandis que les autres groupes comprennent les espèces *X. sacchari*, *X. albilineans*, *X. translucens*, *X. theicola* et *X. hyacinthi* (Hauben *et al.*, 1997).

Pour affiner davantage la taxonomie du genre *Xanthomonas*, une étude basée sur le typage moléculaire d'éléments répétés : BOX (conserved repetitive sequences), ERIC (enterobacterial repetitive intergenic consensus), et REP (repetitive extragenic palindromic) a été entreprise (Figure 42) (Rademaker *et al.*, 2000, 2005). Parmi les 339 souches typées, 160 souches appartiennent à des pathovars orphelins ; les autres appartiennent aux 20 espèces identifiées auparavant par Vauterin *et al.* (1995). L'analyse des profils rep-PCR obtenus a permis de regrouper toutes les souches dans les 20 espèces de *Xanthomonas* et de caractériser la diversité des espèces *X. hortorum*,

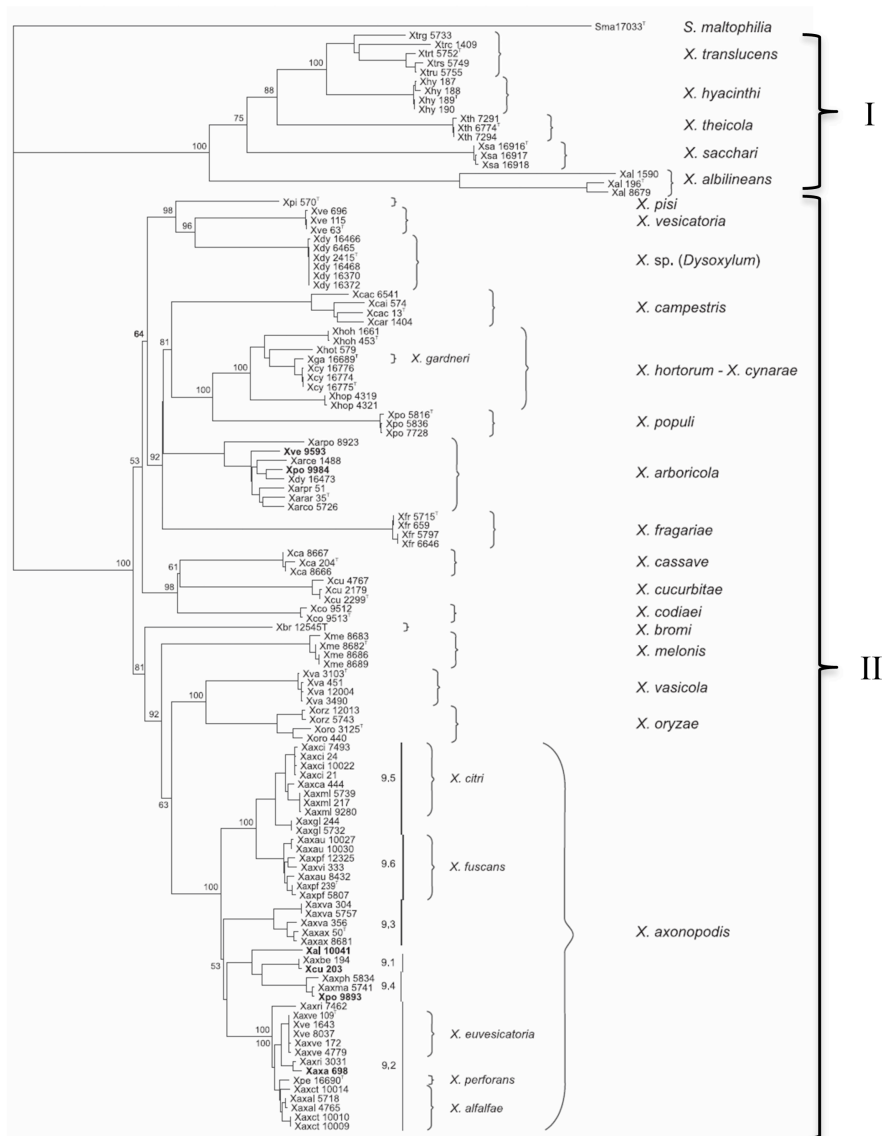


Figure 43. Arbre phylogénétique construit avec la méthode Neighbour-Joining sur la base des séquences partielles concaténées des gènes *dnaK*, *fyuA*, *gyrB* et *rpoD* (d'après Young *et al.*, 2008). Les souches sont regroupées en deux clusters. Le premier cluster comprend les 5 espèces *X. translucens*, *X. hyacinthi*, *X. theicola*, *X. sacchari* et *X. albilineans*. Le deuxième cluster qui est majoritaire regroupe les 15 espèces restantes. Ces clusters corroborent les regroupements établis par Hauben *et al.* (1997) sur la base des séquences partielles de l'ADNr 16S et pourraient correspondre à deux genres différents. Les groupes 9.1 à 9.6 définis au sein de *X. axonopodis* sont représentés. Sur cette figure, *X. alfalfae*, *X. citri*, *X. euvesicatoria*, *X. fuscans* et *X. perforans* constituent des sous-espèces au sein de *X. axonopodis*.

X. arboricola, *X. axonopodis*, *X. oryzae*, *X. campestris* et *X. translucens* (Rademaker *et al.*, 2005).

Récemment, une analyse MLSA (Multi Locus Sequence Analysis) utilisant quatre gènes de ménage, *dnaK*, *fyuA*, *gyrB* et *rpoD*, a été appliquée sur une collection de souches représentatives d'espèces déjà décrites au sein du genre *Xanthomonas* (Figure 43) (Young *et al.*, 2008). Ce schéma MLSA rapporte deux groupes correspondant à ceux identifiés par Hauben *et al.* (1997) sur la base des séquences partielles de l'ADN ribosomal 16S (Figure 43). La différenciation entre ces deux groupes est appuyée par l'incapacité à obtenir des amorces pour les gènes *dnaK* et *fyuA* pour les espèces du groupe 1. Ceci indique un manque d'homologie dans les gènes entre ces deux groupes. D'après ces résultats, Young *et al.* (2008) suggèrent que le genre *Xanthomonas* tel qu'il est actuellement défini, pourrait correspondre à deux genres et concluent que des études polyphasiques sont nécessaires pour tester cette hypothèse. Par ailleurs, Young *et al.* (2008) montrent par cette méthode que les deux espèces *X. cynarae* et *X. gardnerii* semblent être des synonymes. En outre, les nouvelles espèces *X. alfalfae*, *X. citri*, *X. euvesicatoria*, *X. fuscans* et *X. perforans* proposées par Jones *et al.* (2004) et Schaad *et al.* (2005, 2006, 2007) ne se différencient pas clairement de *X. axonopodis* comme des espèces à part entière. Ces résultats sont en accord avec ceux de Parkinson *et al.* (2009).

4.6.3 Diversité infraspécifique de *X. axonopodis* et *X. campestris*

La majorité des espèces définies au sein du genre *Xanthomonas* (*sensu* Vauterin *et al.*, 1995) renferme des souches qui infectent une gamme limitée de plantes hôtes. L'espèce *X. campestris* renferme des souches qui attaquent les Brassicacées alors que l'espèce *X. oryzae* renferme les souches qui attaquent le riz. En revanche, au sein de l'espèce *X. axonopodis*, les souches affectent plusieurs plantes hôtes différentes. Elles présentent également une forte hétérogénéité infraspécifique pour les hybridations ADN-ADN, l'AFLP et les rep-PCR (Vauterin *et al.*, 1995 ; Rademaker *et al.*, 2005).

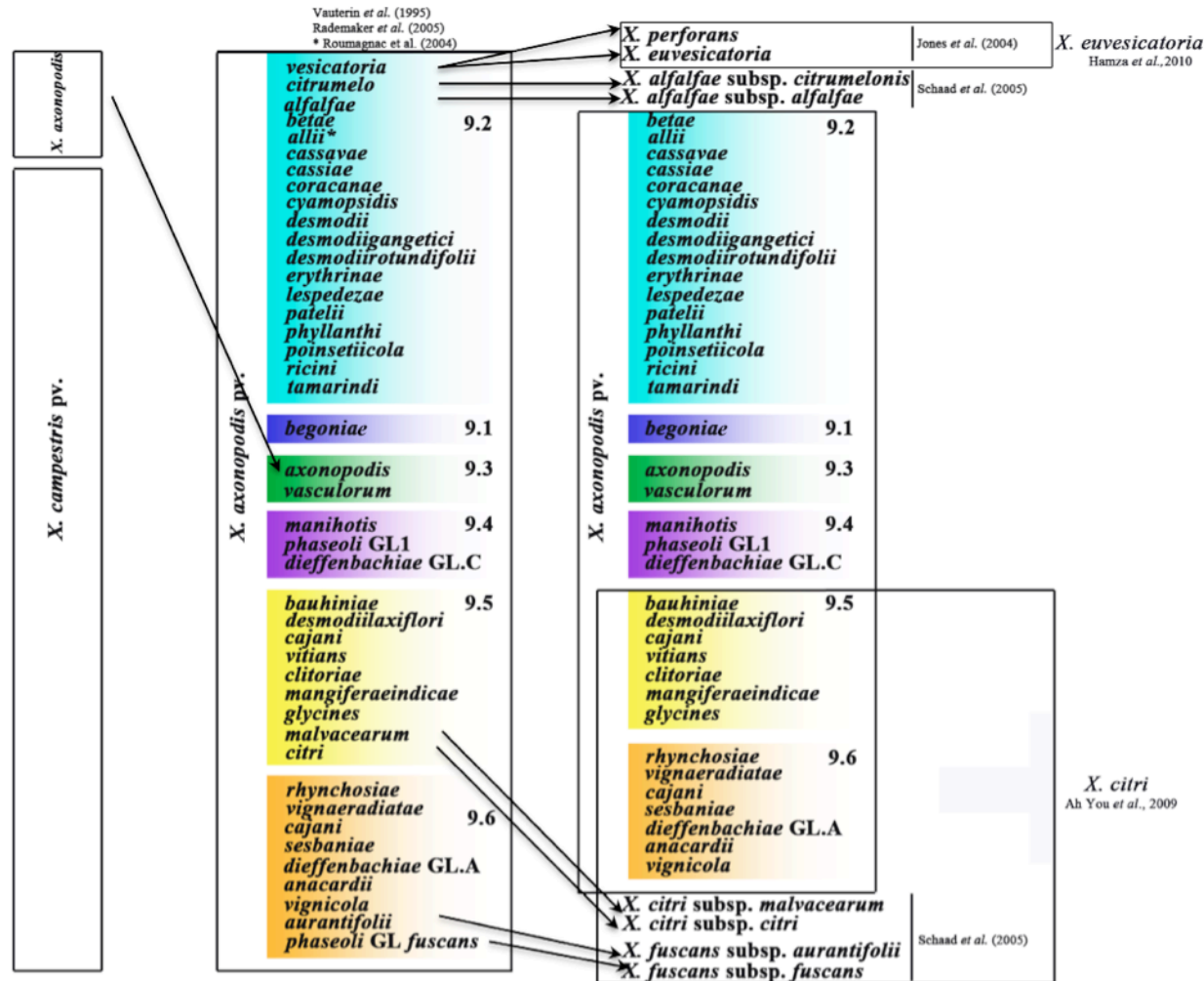


Figure 44. Remaniements taxonomiques au sein de l'espèce *X. axonopodis*. D'après les travaux de Vauterin *et al.*, 1995, Jones *et al.*, 2004, Rademaker *et al.*, 2005, Schaad *et al.*, 2005, Ah You *et al.*, 2009 et Hamza *et al.*, 2010.

4.6.3.a Diversité de l'espèce *X. axonopodis*

Selon le travail de Vauterin et ses collaborateurs (1995), l'espèce *X. axonopodis* est hétérogène et comprend au moins six groupes (Figure 44). Les travaux de Rademaker *et al.* (2005) confirment l'existence des six groupes (nommés 9.1 à 9.6) dans cette espèce sur la base du polymorphisme d'éléments répétés. Depuis, de nombreux travaux ont tenté de clarifier le statut de ces groupes ou de certains pathovars appartenant à ces groupes. Ainsi, Schaad et ses associés (2005) ont proposé d'élever les pathovars *alfalfae* et *citrumelo* appartenant au groupe 9.2, *citri* et *malvacearum* appartenant au groupe 9.5 et *fuscans* et *aurantifolii* appartenant au groupe 9.6, au rang de sous-espèces de trois espèces nouvellement créées : *X. alfalfae*, *X. citri* et *X. fuscans* (Schaad *et al.*, 2006). Néanmoins, ce travail a laissé non modifié le statut des autres pathovars des groupes étudiés. Jones *et al.* (2004) ont de manière similaire proposé d'élever les membres du pathovar *vesicatoria* au rang de deux espèces distinctes : *X. euvesicatoria* et *X. perforans*. Ils ont pour la suite décrit une nouvelle espèce pathogène de la tomate, *X. gardnerii*. Young *et al.* (2008) indiquent que *X. euvesicatoria* et *X. perforans* sont probablement des synonymes et qu'ils formeraient une sous-espèce au sein de *X. axonopodis*. La synonymie entre *X. euvesicatoria* et *X. perforans* a récemment été réaffirmée par l'étude menée par Hamza *et al.* (2010) sur la diversité génétique et les relations phylogénétiques entre les souches responsables de la galle bactérienne de la tomate et du poivron. Selon Young *et al.* (2008), *X. citri* et *X. fuscans* seraient également des synonymes pouvant être reclassés comme sous-espèce de *X. axonopodis*. Une étude récente se basant sur une approche polyphasique (AFLP, MLSA et hybridations ADN-ADN) a confirmé la synonymie entre *X. citri* et *X. fuscans* et propose de considérer ces deux espèces comme une seule espèce nommée *X. citri* (Ah-You *et al.*, 2009). Les travaux de Young *et al.* (2008) confirment l'hétérogénéité de l'espèce *X. axonopodis sensu* Vauterin *et al.* (1995). Ils indiquent que l'élévation au rang d'espèce des groupes 9.1 à 9.6 n'est pas supportée par les critères de diversité génétique et encore moins de l'homogénéité protéique observée.

Au sein de l'espèce *X. axonopodis*, certains pathovars sont monophylétiques *i.e* les souches sont toutes regroupées au sein d'un même groupe rep-PCR comme par exemple, le pathovar *citri* (Hajri *et al.*, 2009 ; Bui Thi Ngoc *et al.*, 2010) ; au contraire,

d'autres pathovars sont polyphylétiques *i.e* les souches appartiennent à différents groupes rep-PCR comme c'est le cas des pathovars *dieffenbachiae* et *phaseoli* (Rademaker *et al.*, 2005 ; Robène-Soustrade *et al.*, 2006 ; Alavi *et al.*, 2008 ; Hajri *et al.*, 2009).

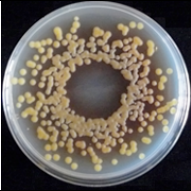
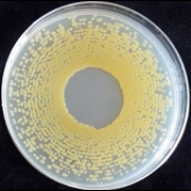
Le pathovar monophylétique *citri* renferme des souches pathogènes des *Citrus* appartenant au groupe 9.5 (Rademaker *et al.*, 2005 ; Ah-You *et al.*, 2009 ; Hajri *et al.*, 2009 ; Bui Thi Ngoc *et al.*, 2010). En revanche, trois pathotypes, A, A* et A^w, ont été décrits au sein de ce pathovar (Bui Thi Ngoc *et al.*, 2010). Le pathotype A présente une large gamme de plantes hôtes appartenant au genre *Citrus*. Par contre, les pathotypes A* et A^w présentent une gamme de plantes hôtes plus restreinte (Bui Thi Ngoc *et al.*, 2010). Les pathotypes A* et A^w sont phylogénétiquement très proches du pathotype A. Sur la base des tests de pouvoir pathogène, de l'AFLP et de la MLSA, le pathotype A^w semble être le synonyme du pathotype A* (Bui Thi Ngoc *et al.*, 2010).

Le pathovar polyphylétique *dieffenbachiae* comprend des souches qui appartiennent à des groupes rep-PCR différents, les groupes 9.4 et 9.6 (Rademaker *et al.*, 2005). Des tests sérologiques et de pouvoir pathogène par inoculations de différentes souches de *X. axonopodis* pv. *dieffenbachiae* sur différentes espèces d'Aracées révèlent l'existence de quatre groupes. Le groupe I rassemble les souches pathogènes de l'*Anthurium*, le groupe II rassemble les souches pathogènes du *Syngonium* et non pathogènes de l'*Anthurium*, le groupe III comprend les souches pathogènes d'Aracées autres que l'*Anthurium* et le groupe IV comprend les souches isolées de l'*Anthurium* sans être pathogènes de cet hôte. Les souches du groupe IV sont, de plus, faiblement pathogènes de *Dieffenbachia maculata* (Robène-Soustrade *et al.*, 2006). Par ailleurs, Hajri *et al.* (2009), ont différencié 3 groupes au sein du pathovar *dieffenbachiae*. Le groupe A composé d'une seule souche isolée de *Dieffenbachia*, appartient au groupe 9.6. Le groupe B également composé d'une seule souche isolée de philodendron, se place dans le groupe 9.2. Le groupe C rassemble des souches isolées d'*Anthurium* appartient au groupe 9.4. La souche isolée de philodendron ne s'avère pathogène sur aucun Aracées (O. Pruvost, communication personnelle). Ainsi, le groupe A pourrait correspondre au groupe III, et le groupe C au groupe I.



Figure 45. Symptômes de la graisse commune du haricot observés sur les différents organes. A gauche, nécroses huileuses sur la face inférieure d'une foliole. Au centre, taches huileuses et nécroses entourées d'un halo jaune sur une gousse et une feuille au champ. A droite, symptômes de graisse sur graines (photo INRA et OEPP, 2004).

Tableau 5. Récapitulatif des différences pathologiques, phénotypiques et génétiques documentées entre les lignées *fuscans* et non *fuscans* (GL1, GL2 et GL3) de *X. axonopodis* pv. *phaseoli*.

Caractéristiques	<i>fuscans</i>		non <i>fuscans</i>		Références
Symptômes		Identiques			
Agressivité	Plus grande		Plus faible		Opio <i>et al.</i> , 1996
Voie de catabolisme de la tyrosine	Incomplète, stoppée à l'acide homogénistique		Complète		Goodwin et Sopher, 1994
Virulence sur <i>Medicago truncatula</i>	Virulent		Avirulent		M-A. Jacques, non publié
Virulence sur <i>P. vulgaris</i> cv. Masai	Virulent		Avirulent		T. Boureau, non publié
Agressivité sur <i>P. vulgaris</i> par trempage et pulvérisation	Forte		Réduite		M-A. Jacques, non publié
Diversité génétique	Plus faible (0,184±0,209)		Plus élevée, (0,233±0,226) Trois lignées génétiques distinctes (GL1 appartient à 9.4, GL2 et GL3 appartiennent à 9.6)		Mahuku <i>et al.</i> , 2006 Mkandawire <i>et al.</i> , 2004 Alavi <i>et al.</i> , 2007
SST3 de type <i>hrp</i>			Présent		Alavi <i>et al.</i> , 2008
SST3 de type SPI-1	Absent		Présent chez uniquement GL1		
Efficacité de transmission à la graine par voie florale	80%		GL1 : 10% GL2 : 30% GL3 : 70%		M-A. Jacques, non publié

Le pathovar polyphylétique *phaseoli* regroupe des souches responsables de la graisse commune du haricot (Figure 45). Selon l'aptitude des souches à produire un pigment brun diffusible (nommé *fuscans*) sur les milieux de culture contenant de la tyrosine, on distingue les souches *fuscans* et les souches non *fuscans*. Les souches *fuscans* sont plus agressives sur *Phaseolus vulgaris* mais globalement les souches *fuscans* et non *fuscans* engendrent les mêmes symptômes sur le haricot (Opio *et al.*, 1996). Différentes techniques moléculaires comme par exemple, les hybridations ADN-ADN (Hildebrand *et al.*, 1990), l'électrophorèse en champ pulsé (Chan et Godwin, 1999), la RFLP et la rep-PCR (Mahuku *et al.*, 2006), ont pu différencier les deux variants. Par ailleurs, une analyse par AFLP de la diversité génétique d'une collection mondiale de souches de *X. axonopodis* pv. *phaseoli* a confirmé la diversité au sein de ce pathovar. Cette analyse a révélé que les souches *fuscans* sont homogènes tandis que les souches non *fuscans* sont plus diverses (Tableau 5). Celles-ci sont distribuées en trois lignées génétiques GL1, GL2 et GL3. La lignée génétique GL1 est phylogénétiquement éloignée des trois autres lignées (Figure 46) (Alavi *et al.*, 2008). Elle se trouve en effet dans le groupe rep-PCR 9.4 alors que les lignées génétiques GL2, GL3 et *fuscans* se trouvent dans le groupe rep-PCR 9.6 (Rademaker *et al.*, 2005 ; Ah-You *et al.*, 2009 ; Hajri *et al.*, 2009).

4.6.3.b. Diversité de l'espèce *X. campestris*

L'espèce *X. campestris* a été réduite à 6 pathovars (pvs. *aberrans*, *armoraciae*, *barbareae*, *campestris*, *incanae* et *raphani*), suite aux travaux de Vauterin *et al.* (1995). Plus récemment, Fargier et Manceau (2007) ont restreint le nombre de pathovars de *X. campestris* à trois : *campestris*, *incanae* et *raphani* sur la base de tests de pouvoir pathogène. En effet, les souches du pathovar *barbareae* se sont avérées non pathogènes ; elles ne sont donc plus assignées à un pathovar et ne portent plus que le nom *X. campestris*. La souche 756C du pathovar *armoraciae* provoque le même symptôme que les souches du pathovar *raphani*, agent causal de la maladie des taches foliaires. Les noms *armoraciae* et *raphani* sont donc synonymes. De façon similaire, les souches du pathovar *aberrans* provoquent les mêmes symptômes que les souches du pathovar *campestris*, agent causal de la maladie de la nervation noire. Les noms *aberrans* et *campestris* sont aussi des synonymes. Le pathovar *incanae* est quant à lui

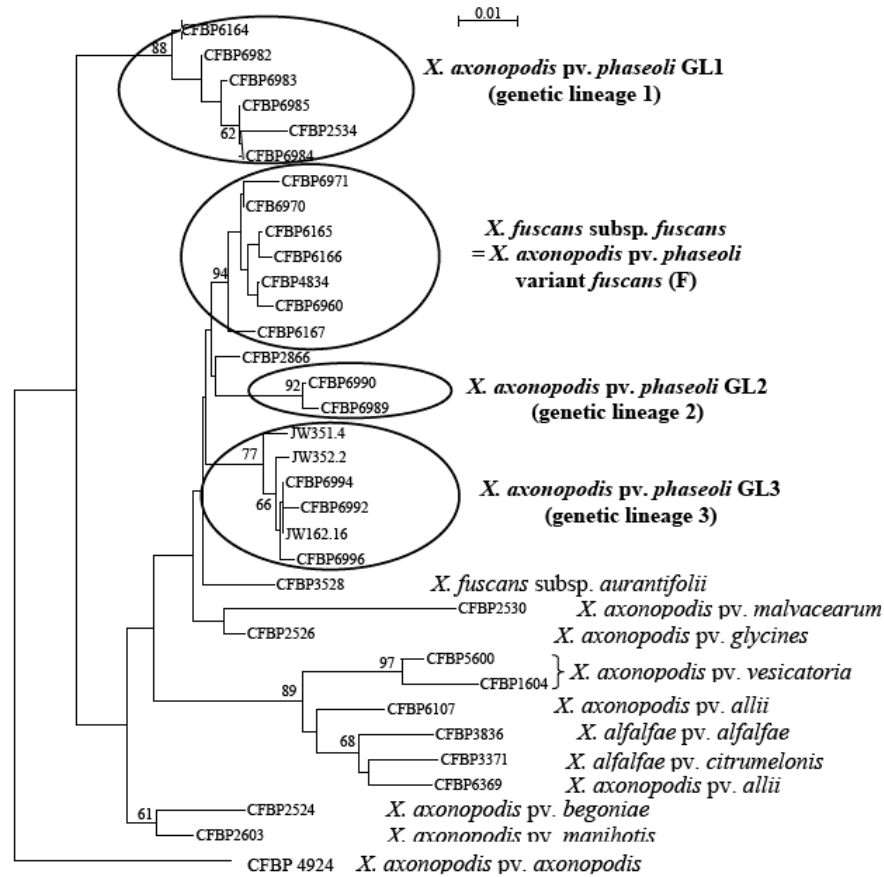


Figure 46. Arbre phylogénétique construit avec la méthode Neighbour-Joining sur la base des empreintes AFLP et représentant les relations entre les souches appartenant aux 4 lignées génétiques de *X. axonopodis* pv. *phaseoli* (*fuscans*, GL1, GL2 et GL3) et des souches appartenant à d'autres pathovars de *X. axonopodis* (d'après Alavi *et al.*, 2008). Sur cette figure, la lignée génétique GL1 est éloignée des trois autres lignées génétiques (*fuscans*, GL2 et GL3) de *X. axonopodis* pv. *phaseoli*.

responsable du dépérissement des giroflées (Fargier et Manceau, 2007). L'étude phylogénétique et de la structuration des souches au sein de *X. campestris* confirme l'existence de trois pathovars au sein de cette espèce. Elle montre également qu'il existe une grande diversité au sein de *X. campestris* due à l'accumulation de mutations et d'événements de recombinaison entre les membres de l'espèce *X. campestris* (Fargier, 2007). Au sein du pathovar *campestris*, neuf races ont été identifiées en fonction de l'aptitude des souches à provoquer des réactions différentielles sur certaines variétés de Brassicacées (Kamoun *et al.*, 1992 ; Vicente *et al.*, 2001 ; Fargier et Manceau, 2007). De façon similaire, le pathovar *raphani* comprend également trois races (Vicente *et al.*, 2006).

Deuxième partie

Résultats

1. Rôle des senseurs et des adhésines dans la spécificité d'hôte des bactéries phytopathogènes, cas des *Xanthomonas* spp.

1.1 Présentation de l'article 1 et des résultats complémentaires

Afin d'établir la liste des gènes candidats codant des facteurs de virulence potentiellement impliqués dans les étapes précoces d'adaptation à l'hôte, une analyse *in silico* par *genome mining* a été entreprise. Cette analyse repose sur l'exploitation des séquences génomiques de quatre souches de *Xanthomonas* ayant des gammes d'hôtes différentes : la souche 85-10 de *X. axonopodis* pv. *vesicatoria*, responsable de la galle bactérienne sur tomate et poivron, la souche 306 de *X. axonopodis* pv. *citri*, responsable du chancre asiatique des agrumes, la souche ATCC33913 de *X. campestris* pv. *campestris*, responsable de la nervation noire des crucifères et la souche KACC10331 de *X. oryzae* pv. *oryzae*, responsable de la maladie des stries foliaires et du flétrissement des feuilles de riz. La recherche des domaines fonctionnels et l'utilisation de mots-clés ont permis d'identifier les gènes candidats dans chacun des génomes de référence. Cette recherche a été complétée par une analyse BLAST pour identifier les orthologues dans les autres génomes. Ce travail nous a permis dans un premier temps de nous renseigner sur la distribution et le niveau de conservation des séquences nucléotidiques des gènes chez les quatre souches de référence et d'autre part, d'établir une liste quasi-exhaustive des gènes candidats impliqués dans les étapes précoces de l'interaction entre la bactérie et son hôte. A partir de cette première liste, une liste restreinte de gènes présentant un polymorphisme de présence/absence et nucléotidique a été établie.

Nous avons caractérisé par PCR les répertoires de ces gènes candidats dans une vaste collection de souches représentant la diversité des espèces *X. axonopodis* et *X. campestris* et analysé leur distribution au regard de la gamme d'hôte de ces souches.

Enfin, nous avons étudié le type de pression de sélection régissant l'histoire évolutive des gènes candidats (MCPs et adhésines). Cette analyse a été réalisée dans un premier temps, sur un jeu de gènes candidats présents dans les génomes des souches séquencées de *Xanthomonas*.

L'ensemble des résultats obtenus est présenté ci-après sous forme d'un article intitulé « *Sensing and adhesion are adaptive functions in the plant pathogenic xanthomonads* » et soumis pour publication à la revue « *BMC evolutionary biology* ».

Je présente ensuite des résultats complémentaires visant à :

- Etudier les relations phylogénétiques entre les séquences de gènes codant les MCPs.
- Etudier l'histoire évolutive des gènes codant les MCPs présents dans les génomes de quatre souches de *X. campestris* par un élargissement de l'étude des pressions de sélection.
- Confirmer les résultats de pression de sélection obtenus *in silico* sur les trois génomes disponibles de *X. axonopodis*, aux souches de notre collection pour trois gènes codant les MCPs après amplification et séquençage.

1.2 Article 1:

Sensing and adhesion are adaptive functions in the plant pathogenic xanthomonads

Sensing and adhesion are adaptive functions in the plant pathogenic xanthomonads

Nadia Mhedbi-Hajri¹, Armelle Darrasse¹, Sandrine Pigné¹, Karine Durand¹, Stéphanie Fouteau², Valérie Barbe², Charles Manceau¹, Christophe Lemaire³, and Marie-Agnès Jacques¹

¹UMR077 PaVé, INRA, 42, rue Georges Morel, F-49071 Beaucouzé, France; ²CEA/DSV/IG/Genoscope, 2 rue Gaston Cremieux, 91057 Evry Cedex 06, France and ³UMR077 PaVé, Université d'Angers, 42, rue Georges Morel, F-49071 Beaucouzé, France

Authors' e-mail addresses: nadia.mhedbi-hajri@angers.inra.fr,
armelle.darrasse@angers.inra.fr, sandrine.pigne@yahoo.fr,
Karine.Durand@angers.inra.fr, svincent@genoscope.cns.fr, vbarbe@genoscope.cns.fr,
charles.manceau@angers.inra.fr, christophe.lemaire@univ-angers.fr, marie-agnes.jacques@angers.inra.fr

Correspondence: M-A Jacques INRA, UMR077 PaVé, 42, rue Georges Morel, F-49071 Beaucouzé, France. E-mail: marie-agnes.jacques@angers.inra.fr and C Lemaire UA, UMR077 PaVé, 42, rue Georges Morel, F-49071 Beaucouzé, France. christophe.lemaire@univ-angers.fr

Abstract

Background: Bacterial plant pathogens belonging to the *Xanthomonas* genus are tightly adapted to their host plants and are not known to colonize other environments. Host range of each strain is usually restricted to a few host plant species. Bacterial strains responsible for the same type of symptoms on the same host range cluster in a pathovar. The phyllosphere is a highly stressful environment, but it provides a selective habitat and a source of substrates for these bacteria. Xanthomonads colonize host phylloplane before entering leaf tissues and engaging into an invasive pathogenic phase. Hence, these bacteria must have evolved strategies to adapt to life in this environment. We hypothesized that determinants responsible for bacterial host adaptation are expressed starting from chemotactic attraction and adhesion on host tissue.

Results: We established the distribution of 70 genes coding sensors and adhesins in a large collection of xanthomonad strains. These 173 strains belong to different pathovars of *Xanthomonas* spp. and display different host ranges. Candidate genes are involved in chemotactic attraction (25 genes), chemical environment sensing (35 genes) and adhesion (10 genes). Our study revealed that candidate gene repertoires comprised core and variable gene suites that likely have distinct roles in host adaptation. Most pathovars were characterized by unique repertoires of candidate genes highlighting a correspondence between pathovar clustering and repertoires of sensors and adhesins. To further challenge our hypothesis, we tested for molecular signatures of selective pressures on candidate genes extracted from sequenced genomes of strains belonging to different pathovars. We identified strong evidence of adaptive divergence acting on most candidate genes.

Conclusions: These data provide insight into the potential role played by sensors and adhesins in the adaptation of xanthomonads to their host plants. The correspondence between repertoires of sensor and adhesin genes and pathovars and the rapid evolution of sensors and adhesins shows that events leading to host specificity may occur as early as chemotactic attraction by host and adhesion to tissues for plant pathogenic xanthomonads.

Background

Deciphering how bacteria adapt to their hosts helps understand their spreading. Host specificity can be established by determining the genes coding virulence factors that are not conserved among strains, which differ in their host range [1]. Virulence-associated genes are expressed during initial host colonization, multiplication, development of symptoms, and dispersal. Sarkar and colleagues [1], and Hajri and associates [2] demonstrated that canonical virulence factors such as type III effectors (T3Es) play a critical role in host specificity. However, T3Es are injected into plant host cells once bacteria have already penetrated into host tissues [3]. Thus phases preceding infection could also be involved in host specificity *i.e.* be under selective pressures.

For bacteria to adapt specifically to their hosts, they sense favorable environmental stimuli and then they move toward favorable conditions [4,5]. Bacteria have evolved receptors and sensors in their cell wall to detect chemical and environmental signals such as the presence of chemoattractants, chemorepellents, and oxygen. They thereby integrate information on their biotic and abiotic environment [6]. Studies on *Rhizobia* revealed the importance of sensors in the perception of specific host signals early during symbiotic interaction with legumes [7,8]. Similarly, *Agrobacterium tumefaciens* and *Ralstonia solanacearum* specifically detect various components from root exudates that attract them toward their hosts [9,10].

Environmental signals are mainly detected by Methyl-accepting Chemotaxis Proteins (MCPs) and Sensors of Two-Component Regulatory System (STCRS). MCPs are the principal components of the chemotaxis system [4]. Detection of signals by these transmembrane chemoreceptors directs cell locomotion by regulating the histidine kinase CheA, which in turn communicates the information to the flagellar motor by phosphorylating its cognate response regulator CheY [4,5]. Changes in the direction or the speed of flagellar rotation modify swimming behavior, resulting in movement towards higher gradients of attractants and away from high concentrations of repellents [11,12]. In *Escherichia coli*, chemotaxis proteins cluster in membrane-associated patches [13,14]. Interactions within patches contribute to the notable features of this signaling system: high sensitivity, wide dynamic range, signal integration, memory, and adaptation [15]. Besides MCPs, bacteria sense their nutritional environment through

TonB-Dependent Transporters (TBDTs) [16]. A large proportion of TBDTs genes are related to plant scavenging and carbohydrate utilization. TBDTs are over-represented in various bacteria interacting with plants such as *Xanthomonas* spp. [17].

Adhesion to a surface is a prerequisite for aggregation in biofilm, which enhances the resistance of bacteria to various biotic and abiotic stresses, favors the coordination of adapted responses to environmental changes, and allows multiplication [18,19]. Sensing and adhesion mechanisms are interconnected since biofilm formation is regulated by a chemosensory system [20]. The adhesion step involves surface structures in a broad group of fimbrial and nonfimbrial adhesins. The fimbrial proteins include type IV pili (Tfp), which are polymeric assemblies of the protein pilin [21,22]. The nonfimbrial adhesins belong to the autotransporter family (*e.g.* XadA and YadA proteins) [23,24] and to the two-partner secretion system (*e.g.* FhaB and YapH proteins) [25].

Each plant pathogenic bacterium belonging to the *Xanthomonas* genus is able to colonize a restricted variety of plant hosts and microniches. *Xanthomonas* are exclusively plant-associated bacteria, mainly phyllosphere colonizers, and are not encountered in other environments [26]. Globally, they infect a huge range of economically important plants such as rice, banana, citrus, bean, tomato, pepper, sugarcane, and wheat [26]. The large host range of the genus strikingly contrasts with the typically narrow host range of individual strains restricted to one or several species of a botanical family [27]. Besides their very homogeneous phenotype, xanthomonads differ mainly by their host specificity. This is illustrated in the pathovar subspecific division, which clusters bacterial strains causing similar symptoms on a same host range [28]. A few pathovars are represented by polyphyletic genetic lineages *i.e.* pv. *phaseoli* [29] and pv. *dieffenbachiae* [30]. The two lineages of this last pathovar are pathogenic on different hosts (*Anturium* and *Dieffenbachia*) and hence may be considered as separate pathovars [30].

The 12 available *Xanthomonas* genomes (<http://www.genomesonline.org/>) revealed a huge number of genes encoding chemotactic sensors, systems sensing the nutritional environment of the cell (MCPs, TBDTs and STCRS) and attachment structures [17,31,32]. This large number reflects a high degree of adaptability and the presence of mechanisms and structures involved in the exploration of the bacterial environment and

adaptive colonization. This led us to hypothesize that determinants responsible for bacterial host specificity are expressed starting from chemotactic attraction by host tissues and adhesion on phylloplane. Thus, characterizing repertoires of genes encoding sensors and adhesins will provide information about the interaction and the adaptation of bacteria to their host plants. Here we characterized the distribution of genes encoding MCPs, STCRS, TBDTs, and adhesins in a large collection of strains belonging to different pathovars in several species of *Xanthomonas*. We also tested for molecular signatures of selective pressures on candidate genes. Two types of outcomes were expected: (i) strong purifying selection acting on genes involved in recognition of common structures of plant tissues and (ii) adaptive divergence on genes coding for sensors and adhesins used for colonization of specific niches. We identified a large variety of repertoires generally fitting with the pathovar clustering. Adaptive divergence was found to affect most candidate genes. These findings provide insight on the evolutionary importance of chemotactic attraction and adhesion in the host specificity of plant pathogenic bacteria.

Results

Identification and selection of genes under study

Based on data mining, we identified genes involved in bacterial attraction, sensing, and adhesion to host. We extracted sequences of genes from four complete genome sequences (*X. axonopodis* pv. *vesicatoria* (*Xav*) strain 85-10, *X. axonopodis* pv. *citri* (*Xac*) strain 306, *X. campestris* pv. *campestris* (*Xcc*) strain ATCC33913, and *X. oryzae* pv. *oryzae* (*Xoo*) strain KACC10331). These four bacteria have different host ranges and are phylogenetically distant. First, a list containing 320 genes involved in sensing, chemotaxis, motility, and adhesion was established using BLAST analysis. This list includes 30 genes encoding MCPs, 17 genes encoding chemotactic protein, 34 genes encoding flagellar components, 31 genes involved in Tfp biogenesis, 10 genes encoding nonfimbrial adhesins, 115 genes encoding TBDTs, and 83 genes encoding STCRS (See additional file 1: Table A1 for the complete list of genes involved in sensing and adhesion in xanthomonads). Second, based on presence/absence and nucleotidic polymorphism, we selected 70 candidate genes: 28 genes encoding TBDTs,

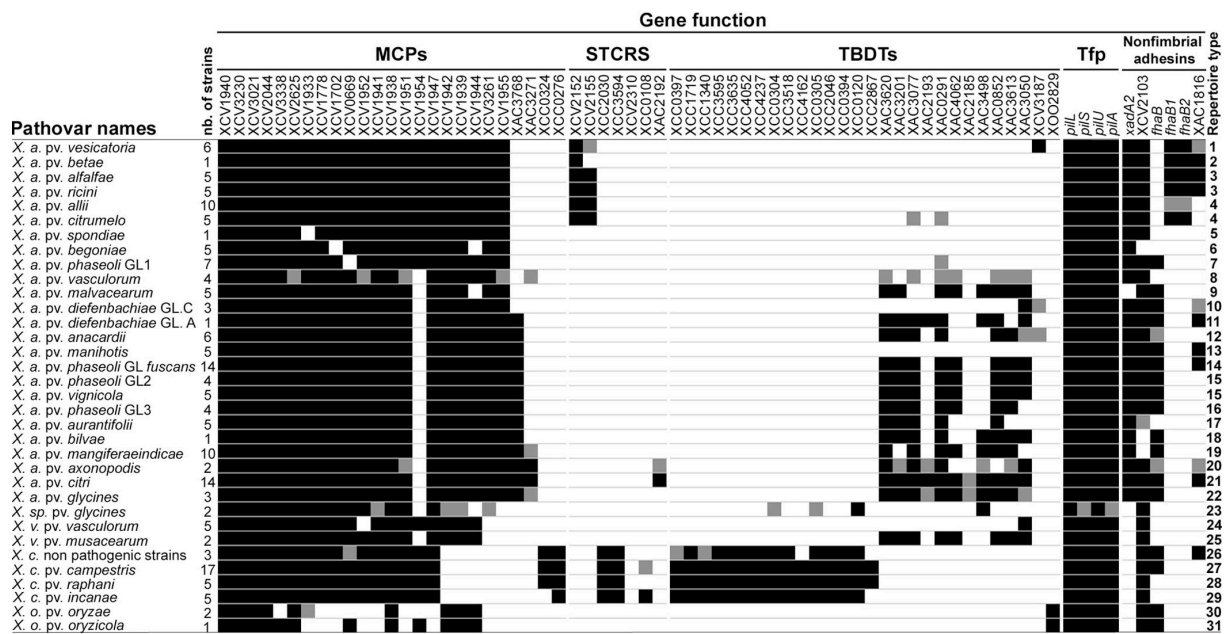


Figure 1 Distribution of genes involved in sensing and adhesion among lineages of *Xanthomonas* spp. Black and white squares represent the presence and the absence, respectively, of the corresponding gene in at least 80% of strains of a pathovar or a genetic lineage. Gray squares indicate that the presence of the gene is variable among strains of the lineage. Thirty one repertoire types were identified on the basis of the presence-absence of genes under study. At least three independent PCR reactions were performed to ascertain the presence/absence of each gene. *X. a.*: *X. axonopodis*; *X. v.*: *X. vasicola*; *X. c.*: *X. campestris* and *X. o.*: *X. oryzae*.

7 genes encoding STCRS, 25 genes encoding MCPs, 2 genes encoding Tfp sensors, 1 gene encoding Tfp assembly ATPase, and 7 genes encoding adhesins.

Characterization of repertoires of MCPs, STCRS, TBDTs, and adhesins

We investigated the distribution of the 70 selected genes in 173 strains belonging to different lineages in *Xanthomonas* spp. by PCR (See additional file 2: Table A2 for the list of bacterial strains used in this study). Three independent PCR reactions using two different sets of primers and three DNA batches were used to monitor the presence of genes in our strain collection. A gene was considered absent when no signal was obtained. Our approach cannot rule out the fact that some genes that we considered as potentially absent may have undergone several point mutations in the primer regions, which could result in divergence sufficient to avoid amplification through several PCRs. Every PCR result on the DNA of *X. axonopodis* pv. *phaseoli* GL *fuscans* strain CFBP4834 was confirmed by BLAST analysis on the draft genome sequence of this strain (our unpublished data). No discrepancies were observed for this strain in any of the 70 candidate genes.

MCPs, STCRS, TBDTs, and adhesin repertoires combined three categories of genes (Figure 1): first, ubiquitous genes showing a broad distribution among strains (e.g. XCV1940, detected in all strains, and *pilS*, detected in most strains); second, genes displaying a variable distribution since they were not detected in several strains (e.g. XCV1954 and XCV2103); third, genes that are found in only one species (e.g. XAC3768 and *xadA2*, detected only in *X. axonopodis*, and XCC0324 and XCC0276, found only in *X. campestris*).

Most bacterial lineages in *Xanthomonas* spp. displayed unique repertoires of MCPs, STCRS, TBDTs, and adhesins (Figure 1). Our results showed that 28 of the 34 pathovars and genetic lineages of *Xanthomonas* spp. had distinct repertoires. In contrast, different pathovars may display the same repertoire, as do pathovars *alfalfae* and *ricini*, pathovars *allii* and *citrumelo*, and pathovars *vignicola* and genetic lineage GL2 of pathovar *phaseoli*. A large variability in the number of genes constituting repertoires was observed among pathovars with *X. oryzae* strains harboring the smallest repertoires, whereas strains of *X. axonopodis* pv. *citri* displayed the largest repertoires.

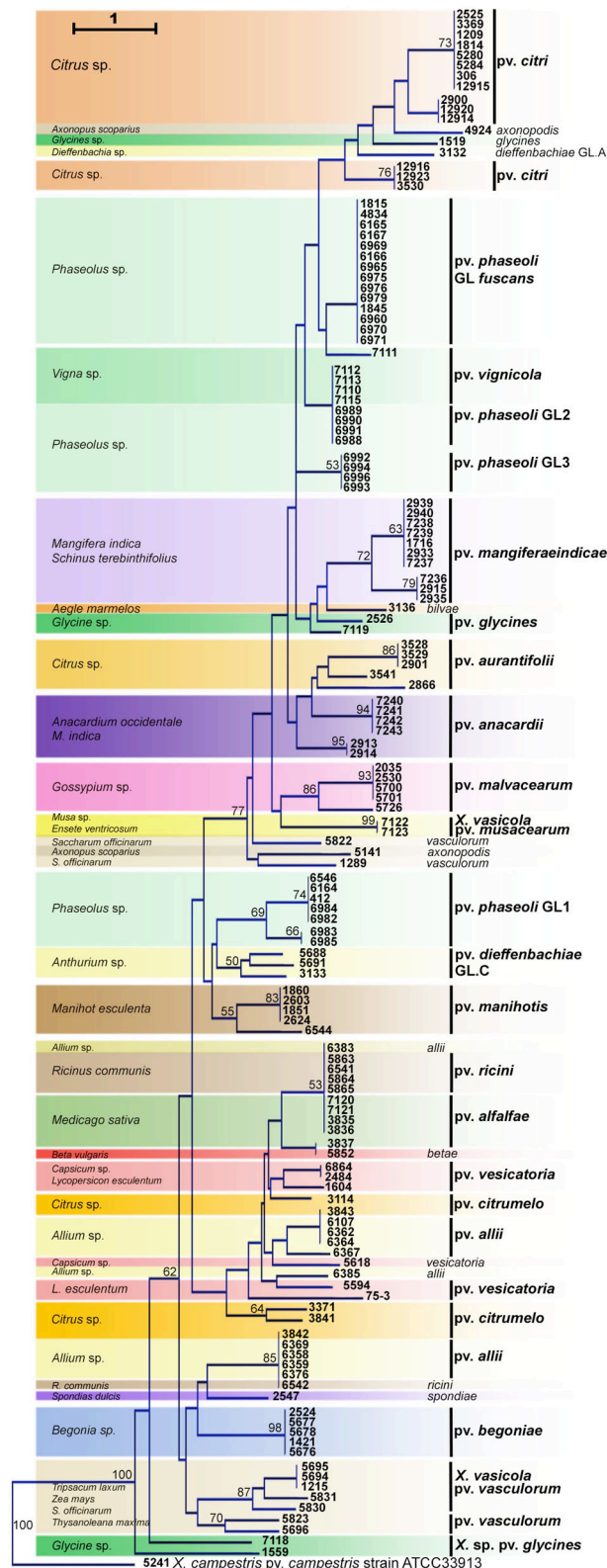


Figure 2 Dendrogram constructed based on presence/absence of candidate genes in strains of *Xanthomonas* spp. The dendrogram was constructed with the Neighbor Joining method using Euclidean distance and rooted with strain CFBP5241 of *X. campestris* pv. *campestris*. Confidence on nodes was established using 1000 bootstrap replicates. Only bootstrap values above 50% are reported.

Repertoires of MCPs, STCRS, TBDTs, and adhesins were almost identical within pathovars *alfalfae*, *begoniae*, *malvacearum*, *manihotis*, *musacearum*, and *oryzae*. Regarding pathovar *anacardii*, strains isolated from *Mangifera indica* (CFBP2913 and CFBP2914) displayed repertoires that differed by three characters (XAC3050, XCV3187 and *fhaB*) from the repertoires of the strains isolated from *Anacardium occidentale* (CFBP7240, CFBP7241, CFBP7242 and CFBP7243). The diversity observed among repertoires in pathovar *allii* could not be correlated by either the genetic diversity or the known host of isolation [33] (See additional file 2: Table A2 for the list of host of isolation of each bacterial strain used in this study).

Based on the presence/absence matrix, clustering of strains from different pathovars could be linked in some cases to the host plant (Figure 2). Grouped together were strains of the pathovar *vignicola*, which infect *Vigna unguiculata* and *V. sinensis*, and strains of GL2 of pathovar *phaseoli*, which infect *Phaseolus vulgaris*. The hosts of isolation of these strains belong to the same botanical family (*Fabaceae*) and are closely related [34]. Also clustered together were strains phylogenetically distant [30] and belonging to different species (CFBP5823 of *X. axonopodis* pv. *vasculorum* and CFBP5830, CFBP5831, and CFBP1215 of *X. vasicola* pv. *vasculorum*) but isolated from the same host, *Saccharum officinarum* (Figure 2).

Positive selection acting on *X. axonopodis* and *X. campestris* genes encoding MCPs and adhesins

The McDonald-Kreitman (MK) tests for adaptive divergence done on 24 candidate genes (MCPs, adhesins, and Tfp sensors) and four housekeeping genes from three sequenced strains of *X. axonopodis* and three of *X. campestris* revealed that 13 of the 24 candidate genes showed robust signatures of adaptive divergence (Figure 3 and see additional file 3: Table A3 for the results of the MK tests). The four housekeeping genes did not display any signal of positive selection after Bonferroni correction. Significant signatures of adaptation were found in the divergence between *X. axonopodis* and *X. campestris* sequences on 11 genes (7 MCPs, 2 adhesins, and 2 Tfp sensors) out of the 24 for which analysis was possible. Using *XocBLS256* as an outgroup showed that diversifying selection preferentially affected the *X. axonopodis* clade.

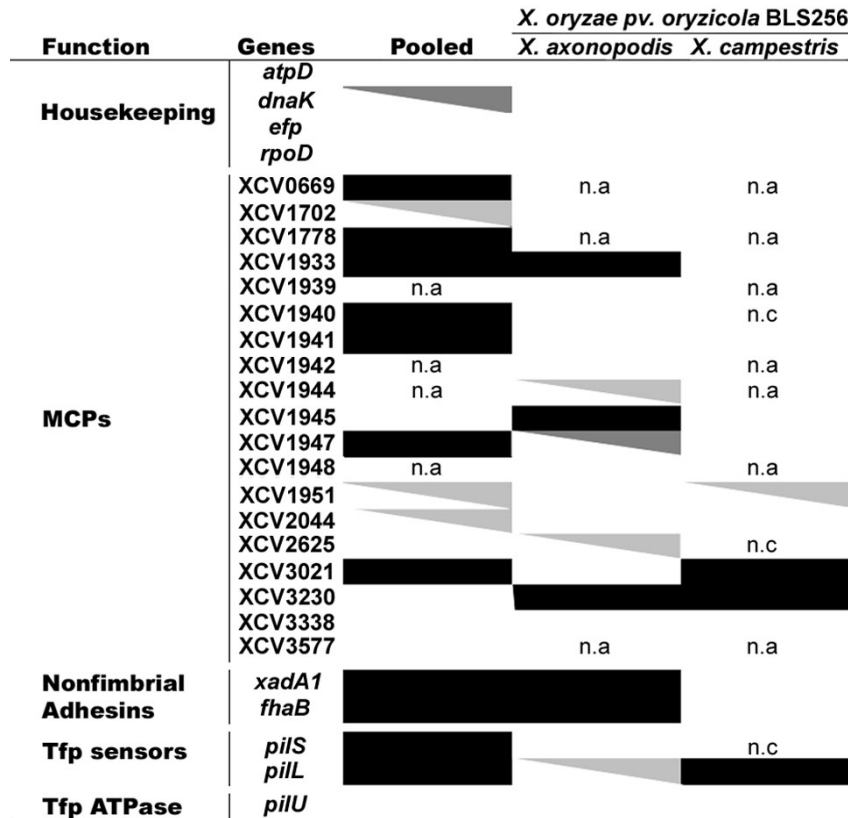


Figure 3 MK test results for adaptive divergence on genes involved in sensing and adhesion. Genes were extracted from sequenced genomes of three strains of *X. axonopodis* (*Xav85-10*, *Xac306*, and *XapCFBP4838*) and three strains of *X. campestris* (*XccATCC33913*, *Xcc8004*, and *XccB100*). The ratios of replacement and synonymous changes within a species were compared with the ratios of replacement and synonymous changes fixed between the species. Pooled refers to MK test done using both species, *X. axonopodis* and *X. campestris*. Using *X. oryzae* pv. *oryzicola* BLS256 as an outgroup allowed to ascertain which species has been affected by positive selection. A sequential Bonferroni correction for multiple tests was applied. Black squares correspond to high p values ($p < 0.0034$) and indicate robust evidence of positive selection before and after correction. Gray triangles indicate signals of positive selection, which are not considered after sequential Bonferroni correction: dark-gray and light gray triangles indicate p values ranging from 0.0034 to 0.01 and from 0.01 to 0.05, respectively. White squares correspond to non-significant p values ($p > 0.05$) and indicate no signal of positive selection. A non-available result (n.a) means the orthologous sequences were lacking on the genome sequence. Lack of polymorphism did not allow test computation (n.c).

Positive selection was also tested using the branch-site model implemented in PAML within *X. axonopodis* on candidate genes already tested with the MK test. The branch-site model allows detecting selection at a few codons (sites) on a specific lineage (branch). Strains of different pathovars in *X. axonopodis* (strains *Xav85-10*, *Xac306* and *XapCFBP4834*) occupy distinct habitats, exploit distinct niches, and cause different diseases on different host-plants (*Lycopersicon esculentum* and *Capsicum* sp., *Citrus* sp. and *Phaseolus vulgaris*, respectively). We hypothesized that these strains representing different pathovars would have undergone selection differently for correspondingly distinct adaptations. For each gene, three tests were applied, each considering a different foreground branch corresponding to each of the three pathovars. Results revealed that sets of genes under positive selection in *Xav85-10*, *Xac306*, and *XapCFBP4834* were distinct. Indeed in the case of pathovar *vesicatoria* (*Xav85-10*), the set of genes under selection included XCV1940 and XCV1942. Regarding pathovar *citri* (*Xac306*), the set included XCV1702, XCV1945, and *xadA1*. Finally, in pathovar *phaseoli* (*XapCFBP4834*), positive selection was detected on genes XCV1945 and XCV1951. Interestingly, XCV1945 was found to be under positive selection on both pathovars *citri* and *phaseoli*. However, the Bayes Empirical Bayes (BEB), which estimates the probabilities of each site on the foreground branch evolving under positive selection, identified different sites either in *Xac306* or *XapCFBP4834* (Table 1). For each gene, from 1 to 10 sites were found under selection. Sixteen of the 22 sites under selection were located in conserved domains found in the proteins (Table 1 and Figure 4).

Discussion

The early phases of host-colonization are crucial for pathogenic bacteria [35]. However, little is known about the importance of early phases in the determinism of host-specificity for plant pathogenic bacteria. Emphasis has been put on later phases of infection, as illustrated by the demonstration of the role of T3Es in host-specificity [1,2,36]. Host-recognition can be considered the first instance of host-pathogen interaction allowing bacteria to colonize. The importance of chemotaxis in plant-bacteria interactions has been clearly documented in some cases. For example *Ralstonia solanacearum* chemotactic mutant is unable to colonize its host when inoculated into the soil, whereas it remains fully pathogenic when infiltrated inside plant tissues [10].

Table 1 Sites under positive selection in candidate genes. Genes were extracted from genomic sequences of three strains belonging to three pathovars of *X. axonopodis* (*vesicatoria*, *citri* and *phaseoli*). Sites were identified using Branch-site model A (model = 2 NSites = 2) compared to the null model with ω fixed to 1.

Strain	Target name ^a	gene Residue	Residue position	$p(\omega>1)$ ^b	Interpro accession
<i>Xav</i> 85-10	XCV1940	P	79	0,92	-
	XCV1942	I	48	0,91	-
	XCV1942	R	131	0,948	-
<i>Xac</i> 306	XCV1702	Q	622	0,929	IPR004089
	XCV1945	D	59	0,909	-
	XCV1945	S	92	0,982	-
	<i>xadA1</i>	Y	390	0.903	-
	<i>xadA1</i>	Q	880	0.947	IPR008640
	<i>xadA1</i>	E	882	0.951	IPR008640
<i>Xap</i> CFBP4834	XCV1945	S	352	0,955	IPR004089
	XCV1945	S	658	0,954	IPR004089
	XCV1945	Q	685	0,945	IPR004089
	XCV1951	S	780	0,971	IPR004089
	XCV1951	V	814	0.924	IPR004089
	XCV1951	E	847	0.968	IPR004089
	XCV1951	V	851	0.984	IPR004089
	XCV1951	N	855	0.962	IPR004089
	XCV1951	V	858	0.975	IPR004089
	XCV1951	K	859	0.989	IPR004089
	XCV1951	R	872	0.982	IPR004089
	XCV1951	T	876	0.993	IPR004089
	XCV1951	A	880	0.982	IPR004089

^a: National Center for Biotechnology Information (NCBI) gene name. XCV indicates a gene in *X. axonopodis* pv. *vesicatoria* 85-10 genome.

^b: ω estimates the parameter D_n/D_s for each site and $p(\omega>1)$ represents posterior probability of sites with $\omega>1$.

Thus evolutionary processes of host-specificity would also be driven by selective forces at the very first steps of host-colonization. Our data show evidence of such adaptive processes for numerous genes involved in chemotactic attraction, environment sensing, and adhesion to surfaces. These mechanisms precede the infection of a plant by xanthomonads. The mechanisms facilitating plant penetration and thus allowing infection are chemotaxis, aerotaxis, and fast multiplication inside host tissue, which relies on adhesion.

Many candidate genes undergo adaptive constraints during bacterial colonization on plant tissue. Of the 24 genes common among repertoires of studied pathovars (*vesicatoria*, *citri*, and *phaseoli*) of *X. axonopodis* and *X. campestris* pv. *campestris*, nearly half are subject to positive selection. This proportion is almost equal among the two gene families (MCPs and adhesins). Charlesworth and Eyre-Walker [37] showed that about 50% of amino-acid substitutions surveyed in the enteric bacterial genomes were subject to adaptive evolution. Thus both recognition and adhesion should be considered as selective steps for bacterial colonization. To our knowledge this is the first report of positive selection acting on MCPs in plant-pathogenic bacterium. Our results are consistent with those published by Chen *et al.* [38] and Petersen *et al.* [39] showing that positive selection acts on genes encoding surface structures of *E. coli* cell. These genes encode regulators of LPS O-antigen chain length, putative adhesins that affect biofilm formation, ferrichrome-iron receptors, two outer membrane porins, and more [38,39]. In xanthomonads, the extracellular appendage of the Hrp pilus evolved under the constraint of positive selection likely to avoid recognition by plant defense surveillance systems [40].

Most sites that were found under selection in candidate genes using PAML analysis were located in conserved domains predicted to play a role in perception for MCPs and in adhesion for XadA1. Indeed, the Tar domain of chemoreceptors directly binds to aspartate and related amino acids [41]. The Hep-Hag motif is found in the passenger domain of adhesins [42]. This domain is known to contribute to the binding activity of invasins/agglutinins [23]. This result is another argument in favor of adaptive evolution acting on these genes in link with the ecological behavior of the strains.

Many bacteria assemble multifunctional proteic structures on their surface that serve for adhesion. This feature might be an adaptation to different environmental conditions

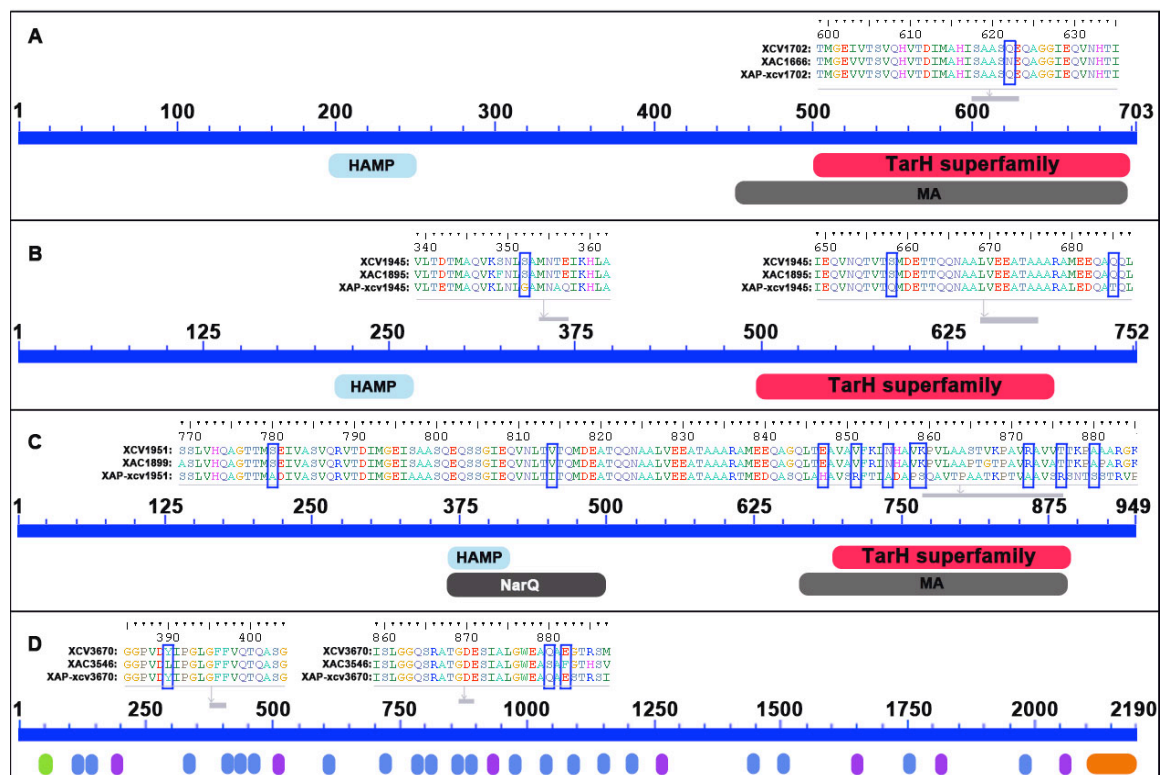


Figure 4 Schematic representations of examples of sites under selection and conserved domains in MCPs (A, B, and C) and adhesins (D). In the upper parts of each panel the sequence alignments of proteins are presented for parts containing sites under selection. The orthologous sequences collected from genome sequences of *Xcv*85-10, *Xac*306 and *Xap*CFBP4834 were aligned using ClustalW according to the translated amino-acid sequences that were manually performed using BIOEDIT. Sites that were found under selection are highlighted by green frames. In the lower parts of the panels, conserved domains are represented. HAMP: Histidine kinase, Adenylyl cyclase, Methyl-accepting protein, and Phosphatase domain. HAMP is a signalling domain that occurs in a wide-variety of signalling proteins. TarH superfamily: Taxis toward Aspartate and Related amino acids and Homologs. MA: Methyl-accepting chemotaxis-like domain (chemotaxis sensory transducer), thought to undergo reversible methylation in response to attractants or repellents during bacterial chemotaxis. NarQ: signal transduction histidine kinase, nitrate/nitrite-specific. Annotation bars coloured in grey refer to multi-domains that are excluded from domain-domain neighboring. In the D panel, green, blue, purple, and orange ovals symbolized TAT signal, Hep-Hag motif, HIM motif and YadAlike, C-terminal domain, respectively.

and, in the case of pathogenic bacteria, to different hosts or host tissues [43]. In fact adhesins are involved in various processes leading to host colonization and transmission to seed by plant-pathogenic bacteria. For example, Tfp serves remarkably diverse functions, including twitching motility, cell to cell adhesion, and thus microcolony and biofilm formation [21]. Tfp is an important virulence factor for vascular and non vascular plant pathogens [44,45]. Moreover, Darsonval and colleagues [45] showed that PilA is involved not only in adhesion but also in transmission to seed, and the mutation of *pilA* in strain CFBP4834 of *X. axonopodis* pv. *phaseoli* GL *fuscans* leads to lower pathogenicity on bean (*P. vulgaris*). Besides, YapH -an hemagglutinin- is required for adhesion to seed, leaves, and abiotic surfaces.

An interesting consequence of strong adaptive constraints is a specialization at some early steps (i.e. chemotactic attraction) of host colonization by xanthomonads. This early adaptation may limit the competition for habitat. In fact, xanthomonads are known to be phenotypically very homogeneous except for pathogenicity. Lack of selective constraints in host colonization would lead to colonization by a wide range of incompatible strains. Colonization of specialized pathovars would therefore be less successful. Indeed, resource allocation would be redirected in favor of competition detrimental to pathogenicity. Then, early adaptation should make infection more efficient. Finally, such host-isolation could act as an ecological isolating barrier to limit recombination between differentially adapted pathogenic strains. As in eukaryotic organisms, ecological differences in bacteria are known to promote speciation [46]. Indeed, habitat sharing would allow recombination between strains that belong to different pathovars and that consequently produce strains that may reveal genetic incompatibilities.

This comparative analysis of repertoires of MCPs, STCRS, TBBDTs and adhesins provide useful insight into bacterial behavior. First, the number of MCPs and more generally sensors is higher in *Xanthomonas* strains than in *E. coli* and *Salmonella* [47]. *E. coli* and *Salmonella* have only five MCPs whereas strains *Xav85-10*, *Xac306*, and *XccATCC339313* have about 20 MCPs; strain *XooKACC10331* has only about 10 MCPs. The large numbers (14 in *Xav85-10*, 10 in *Xac306*, 8 in *XccATCC339313*, and 7 in *XooKACC10331*) of MCP and other sensors repeated in tandem are unusual in bacteria, suggesting a prominent role in the life style of *Xanthomonas* [17,31,32].

Second, repertoires of MCPs, STCRS, TBDTs, and adhesins differed among the majority of pathovars and genetic lineages belonging to the tested *Xanthomonas* spp. and displaying different host range. Repertoires of genes coding sensors and adhesins comprised core and variable gene suites. Some genes under study were not infra-specifically conserved and hence belong to the accessory genome. Repertoires of genes involved in attraction and adhesion may evolve by gene gain or loss probably after duplication events. In the case of *Drosophila*, the size of repertoires of genes encoding olfactory and gustatory receptors varies through gene duplication, pseudogenization, and gene loss. The changes described among tested species of *Drosophila* show that these receptors have changed during species divergence and their evolution might reflect species' adaptation to their chemical environment [48]. Similarly, in xanthomonads the products of the variable set of sensor and adhesin genes may be involved in the recognition of specific components allowing strain adaptation to a particular set of hosts. Moreover, signals of adaptive divergence have been identified on such genes of the variable set.

This study showed that among the same genus, *Xanthomonas*, the majority of pathovars and genetic lineages belonging to different species (*X. axonopodis*, *X. campestris*, *X. vasicola*, and *X. oryzae*) displayed different and unique repertoires of MCPs, STCRS, TBDTs, and adhesins while they displayed different host range. It is important to note that the distribution of sensor and adhesin gene repertoire does not necessarily correlate with strain phylogeny. Indeed, bacteria as phylogenetically distant as *X. axonopodis* pv. *vasculorum* and *X. vasicola* pv. *vasculorum* [30] share a common repertoire of sensor and adhesin genes (Figure 2) and a common host: sugarcane. This case illustrates the sharing of a common ecological niche (symptomatic host) by two phylogenetically distant bacteria.

Our results suggest that adaptation to host embrace pathoadaptation but also asymptomatic colonization steps. Indeed, several pathovars and genetic lineages shared the same repertoires whereas they are known to infect different crops. This means that they could share the same asymptomatic host range but develop symptoms on a restricted number of different host plants. This is the case for the pathovar *vignicola* and the genetic lineage GL2 of pathovar *phaseoli*. *X. axonopodis* pv. *phaseoli* GL2 and pv. *vignicola* strains may detect similar plant-originated molecules potentially conserved

among their host plants. This hypothesis is supported by the fact that *X. axonopodis* pv. *phaseoli* GL2 and *vignicola* both infect legumes. *X. axonopodis* pv. *phaseoli* GL2 infects *Phaseolus* spp. and *X. vignicola* infects *Vigna unguiculata*; these legumes belong to the Milletioideae clade and are phylogenetically closely related [33]. Interestingly, cross inoculations would provide insights on the ecological behaviors of these two pathogens (survival, maintaining, and chemotaxis responses). The four genetic lineages (*fuscans*, GL1, GL2, GL3) of pathovar *phaseoli*, which all share a common host (*P. vulgaris*), present distinct repertoires of MCPs, STCRS, TBDTs, and adhesins. Genetic lineage *fuscans*, GL2, and GL3 are phylogenetically closely related and belong to rep-PCR group 9.6 whereas GL1 is distant and belongs to rep-PCR group 9.4 [2,30,49]. These four distinct genetic lineages have different T3E repertoires but clustered together on the dendrogram constructed on the matrix of presence/absence of T3Es genes, supporting the hypothesis of an adaptive pathological convergence on bean [2]. Our results suggest that, upstream of the invasive pathological stage, the four genetic lineages have different ecological behaviors. Colonization does not necessarily lead to infection and then may not be under the same adaptive processes as host infection. We can speculate that each lineage of *X. axonopodis* pv. *phaseoli* can be found on different asymptomatic hosts.

Plant pathogenic xanthomonads are associated with aerial parts of plants. They are not usually encountered in other environments. Plant pathogenic pseudomonads colonize non-host habitats such as snow or water [50]. Our attempts to isolate xanthomonads from such environments were, however, unsuccessful (our unpublished data). Saprophytic survival of xanthomonads in soil is very poor. Apart from their primary host, many xanthomonads can survive for long periods of time in association with weeds that grow naturally in crops. It is not yet known which weeds are susceptible to colonization by each plant pathogenic xanthomonad. We refer in our study to the main crop contaminated by each pathovar which certainly represent the major opportunity for bacterial multiplication.

Conclusions

This study showed that the majority of the tested pathovars belonging to different species of *Xanthomonas* (*X. axonopodis*, *X. campestris*, *X. vasicola*, and *X. oryzae*)

displayed unique repertoires of genes coding proteins involved in sensing (MCPs, STCRS, TBDTs), and adhesion. Our data show evidence of adaptive processes for numerous genes involved in chemotactic attraction, environment sensing, and adhesion to surfaces. Most sites that were found under selection in candidate genes were located in conserved domains predicted to play a role in perception or in adhesion. Xanthomonads are plant-associated bacterium and are not known to efficiently colonize other environments. Hence, we argue that the gene evolution we observed may reflect pathovar adaptation to the host-plant environment.

The molecular variation of genes involved in host recognition and adhesion to host tissues clearly shows adaptation begins at the very first steps of host colonization. We suggest that such adaptive divergence at early phases of host colonization would act as an ecological isolating barrier to promote speciation. However, such a process could have appeared after isolation by genes involved in infection, like T3Es. In this case, adaptation would rather be a consequence than a cause. The respective roles of host recognition and host infection in promoting ecological reproductive barriers remain unclear despite their high evolutionary relevance.

Methods

The xanthomonad strains used in this study named following the nomenclature proposed by Vauterin *et al.* [27,51] are presented in Additional file 2. Genes encoding MCPs, STCRS, TBDTs, and adhesin-related genes were selected from sequenced genomes of *Xav85-10*, *Xac306*, *XccATCC33913*, and *XooKACC10331* (<http://www.ncbi.nlm.nih.gov>). These genes were identified according to their function-based assignment in the NCBI website to cell motility (COG N), signal transduction mechanisms (COG T), and inorganic ion transport and metabolism (COG P) classes and according to Thieme *et al.* [32] for fimbrial and nonfimbrial adhesin genes. Using Blastx analysis with default parameters [52], we identified orthologous genes in each of the four reference genomes. Hence, a list containing 320 genes involved in sensing, chemotaxis, motility, and adhesion was established (See Additional file 1: Table A1). This list was refined by analyzing the functional-domain composition of the genes. We selected genes encoding MCPs containing periplasmic domains (*e.g.* TAR, PRK09793, PAS and HAMP domains). Regarding genes encoding TBDTs and STCRS, we selected

receptor and sensor parts, respectively, based on the gene description and gave priority to genes identified in only one of the four sequenced strains. Genes were discarded that encoded cytoplasmic proteins such as chemotactic proteins (CheA-Z) and genes encoding flagellar components characterized by a very low polymorphism of presence or absence and sequence conservation.

Primer design was based on the alignment of orthologous sequences collected from the four genomes in conserved fragments, taking care to avoid amplification of other genes that share domains with the target. Then, sets of primers were validated both by *in silico* specific gene amplifications (Amplify software version 3.1.4) and by specific gene amplification using PCR with extracted genomic DNA from the four sequenced genomes previously mentioned. PCR reactions were prepared as described previously [2]. To characterize repertoires, PCRs were performed using one set of extracted genomic DNA (1 ng) and two sets of boiled bacterial cells (3×10^8 cfu ml⁻¹) per strain. Preliminary dot blot hybridizations were done to validate PCR results. We designed probes for specific (XCC0324) and variable (XCV2103 and XAC1816) genes. Hybridizations were performed as described previously [2] on extracted DNA (250 ng) from reference strains. Positive signals were obtained for all strains in contradiction to the expected results based on bioinformatic analysis of presence/absence on the sequenced-genomes from the same strains. These discrepancies can be explained by the structure of these modular genes, which share motifs in functional domains (*e.g.* HAMP, hemagglutinin, or autotransporter domains) and consequently have many sequence similarities. As dot blot could not be used to confirm absence of signal we designed, whenever possible, another set of primers for each non-ubiquitous gene. Generally, two sets of primers were used for each gene on three batches of DNA (See Additional file 4: Table A4 for the sequences of primers). Finally, a fourth set for result validation was obtained using the draft genome sequence of *X. axonopodis* pv. *phaseoli* GL *fuscans* strain CFBP4834 (our unpublished data). Every PCR result on the DNA of this strain was confirmed by BLAST analysis [52].

Based on the presence/absence matrix of MCPs, STCRS, TBDTs, and adhesin genes obtained by PCRs for each of the 173 strains of *Xanthomonas* sp., we constructed a dendrogram using Euclidean distance and Neighbor-Joining method, which was visualized using PAST version 1.90 software [53].

Tests for positive selection were performed on orthologous sequences collected from the genome sequences of *Xav85-10*, *Xac306* and *XapCFBP4834*, *XccATCC33913*, *Xcc8004* and *XccB100*, and *XocBLS256*. Twenty four candidate genes, including 19 MCPs genes and 5 adhesin-related genes, were analyzed together with 4 housekeeping genes. The nucleotide sequences were aligned using ClustalW [54] according to the translated amino-acid alignment in order to keep the codon structure of the coding sequences. Adjustments on multiple alignments were manually performed using BIOEDIT version 7.0.9.0 program [55]. Annotation of domains was performed by using the Conserved Domain Database available [56] in NCBI web site. The DNASP software package [57] was used to perform several tests for positive selection using the MK test [58]. The MK test is a simple method to contrast the patterns of within-species polymorphism and between-species divergence at synonymous and non-synonymous (replacement) sites in the encoding region of a gene. It evaluates whether an excess of replacement mutations versus synonymous mutations had been fixed between the two species compared with replacement and synonymous polymorphisms within each species. A significant ratio of fixed replacement to fixed synonymous mutations leads to rejecting the neutral mutation hypothesis and indicates adaptive fixation of selectively advantageous mutations. Three MK test conditions were applied (i) to detect adaptive divergence signals using two-pooled species (*X. axonopodis* vs *X. campestris*) and (ii) to ascertain which species has been affected by positive selection by assigning the fixed replacement changes to *X. axonopodis* and/ or *X. campestris* using *X. oryzae* pv. *oryzicola* (*Xoc*) strain BLS256 as outgroup strain. The significance of MK results was established by the Fisher exact test. Bonferroni correction for multiple tests [59] was hand-computed and applied on MK test results. Testing for positive selection acting at a specific locus on a particular pathovar (*i.e.* branch) was performed using the branch-site model A of Yang and Nielsen [60] implemented in the CODEML program of the PAML package (version 3.14) [60,61]. Testing was performed on each gene previously analyzed for adaptive divergence. Three sequenced-genome strains of different pathovars in *X. axonopodis* (pvs. *vesicatoria*, *citri* and *phaseoli*) were used. For each gene, three tests were done by labeling one branch as foreground at a time. Model selection was performed using likelihood ratio test. In contrast to other methods of selection detection, the Branch-site model allows for small numbers of taxa.

Authors' contributions

NMH design and performed the experiments, analyzed the data and drafted the manuscript. AD participated in the design of the study, in the construction of gene repertoires, in the data analyses and in the writing. SP and KD participated in the construction of gene repertoires. SF and VB performed sequencing of CFBP4834 genome. CM participated in the design of the study and revised the manuscript. CL analyzed the data and wrote the paper. MAJ conceived the study, design the experiments, analyzed the data and wrote the paper. All authors read and approved the final manuscript.

Acknowledgments and funding

We thank Ralf Koebnik and Tristan Boureau for critical reading of the manuscript. N Mhedbi-Hajri was supported by a grant from the Tunisian Government. This study was a part of programs funded by the Region Pays de la Loire and by the Plant Health and Environment Department of INRA.

References

1. Sarkar SF, Gordon JS, Martin GB, Guttman DS: **Comparative genomics of host-specific virulence in *Pseudomonas syringae***. *Genetics* 2006, **174**(2):1041-1056.
2. Hajri A, Brin, C, Hunault, G, Lardeux, F, Lemaire, C, Manceau, C, Boureau, T, Poussier, S: **A «Repertoire for Repertoire» Hypothesis: Repertoires of Type Three Effectors are Candidate Determinants of Host Specificity in *Xanthomonas***. *PLoS ONE* 2009, **4**:e6632.
3. Boureau T, ElMaarouf-Bouteau H, Garnier A, Brisset MN, Perino C, Pucheu I, Barny MA: **DspA/E, a type III effector essential for *Erwinia amylovora* pathogenicity and growth *in planta*, induces cell death in host apple and nonhost tobacco plants**. *Mol Plant Microbe Interact* 2006, **19**(1):16-24.
4. Rao CV, Ordal GW: **The molecular basis of excitation and adaptation during chemotactic sensory transduction in bacteria**. *Contrib Microbiol* 2009, **16**:33-64.
5. Vladimirov N, Sourjik V: **Chemotaxis: how bacteria use memory**. *Biological Chemistry* 2009, **390**(11):1097–1104.

6. Szurmant H, Ordal GW: **Diversity in chemotaxis mechanisms among the bacteria and archaea.** *Microbiol Mol Biol Rev* 2004, **68**(2):301-319.
7. Yost CK, Rochepeau P, Hynes MF: ***Rhizobium leguminosarum* contains a group of genes that appear to code for methyl-accepting chemotaxis proteins.** *Microbiology* 1998, **144** (Pt 7):1945-1956.
8. Yost CK, Clark KT, Del Bel KL, Hynes MF: **Characterization of the nodulation plasmid encoded chemoreceptor gene mcpG from *Rhizobium leguminosarum*.** *BMC Microbiol* 2003, **3**:1.
9. Hawes MC, Smith LY: **Requirement for chemotaxis in pathogenicity of *Agrobacterium tumefaciens* on roots of soil-grown pea plants.** *J Bacteriol* 1989, **171**(10):5668-5671.
10. Yao J, Allen C: **Chemotaxis is required for virulence and competitive fitness of the bacterial wilt pathogen *Ralstonia solanacearum*.** *J Bacteriol* 2006, **188**(10):3697-3708.
11. Bai F, Branch RW, Nicolau DV, Pilizota T, Steel BC, Maini PK, Berry R: **Conformational spread as a mechanism for cooperativity in the bacterial flagellar switch.** *Science* 2010, **327**(5966):685-689.
12. Vladimirov N, Libiedz D, Sourjik V: **Predicted auxiliary navigation mechanism of peritrichously flagellated chemotactic bacteria.** *PLOS Compt Biol* 2010, **6**: e100717.
13. Maddock JR, Shapiro L: **Polar location of the chemoreceptor complex in the *Escherichia coli* cell.** *Science* 1993, **259**: 1717–1723.
14. Sourjik V, Berg HC: **Localization of components of the chemotaxis machinery of *Escherichia coli* using fluorescent protein fusions.** *Mol Microbiol* 2000, **37**(4):740-751.
15. Kentner D, Sourjik V: **Spatial organization of the bacterial chemotaxis system.** *Curr Opin Microbiol* 2006, **9**(6):619-624.
16. Koebnik R: **TonB-dependent trans-envelope signalling: the exception or the rule?** *Trends in Microbiology* 2005, **13**: 343–347.

17. Blanvillain S, Meyer D, Boulanger A, Lautier M, Guynet C, Denancé N, Vasse J, Lauber E, Arlat M: **Plant carbohydrate scavenging through tonb-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria.** *PLoS One* 2007, **2**(2):e224.
18. Danhorn T, Fuqua C: **Biofilm formation by plant-associated bacteria.** *Annu Rev Microbiol* 2007, **61**:401-422.
19. Amano A: **Bacterial adhesins to host components in periodontitis.** *Periodontol 2000* 2010, **52**:12-37.
20. Hickman JW, Tifrea DF, Harwood CS: **A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels.** *Proc Natl Acad Sci U S A* 2005, **102**(40):14422-14427.
21. Craig L, Pique ME, Tainer JA: **Type IV pilus structure and bacterial pathogenicity.** *Nat Rev Microbiol* 2004, **2**(5):363-378.
22. Pizarro-Cerdá J, Cossart P: **Bacterial adhesion and entry into host cells.** *Cell* 2006, **124**(4):715-727.
23. Cotter SE, Surana NK, St Geme JW: **Trimeric autotransporters: a distinct subfamily of autotransporter proteins.** *Trends Microbiol* 2005, **13**(5):199-205.
24. Koretke KK, Szczesny P, Gruber M, Lupas AN: **Model structure of the prototypical non-fimbrial adhesin YadA of *Yersinia enterocolitica*.** *J Struct Biol* 2006, **155**(2):154-161.
25. Mazar J, Cotter PA: **Topology and maturation of filamentous haemagglutinin suggest a new model for two-partner secretion.** *Mol Microbiol* 2006, **62**(3):641-654.
26. Hayward AC: **The hosts of *Xanthomonas*.** In: *Xanthomonas*. Edited by Swings JG and Civerolo EL. London UK: Chapman and Hall; 1993: 1–119.
27. Vauterin L, Hoste B, Kersters K, and Swings J: **Reclassification of *Xanthomonas*.** *Int J Syst Bacteriol* 1995, **45**:472–489.
28. Dye DW, Bradbury JF, Goto M, Hayward AC, Lelliott RA, Schroth MN: **International standards for naming pathovars of phytopathogenic bacteria and a list of pathovar names and pathotype strains.** *Rev Plant Pathol* 1980, **59**:153–168.

29. Alavi SM, Sanjari S, Durand F, Brin C, Manceau C, Poussier S: **Assessment of the genetic diversity of *Xanthomonas axonopodis* pv. *phaseoli* and *Xanthomonas fuscans* subsp. *fuscans* as a basis to identify putative pathogenicity genes and a type III secretion system of the SPI-1 family by multiple suppression subtractive hybridizations.** *Appl Environ Microbiol* 2008, **74**(10):3295-3301.
30. Rademaker JLW, Louws FJ, Schultz MH, Rossbach U, Vauterin L, Swings J, de Bruijn F: **A comprehensive species to strain taxonomic framework for *Xanthomonas*.** *Phytopathology* 2005, **95**(9):1098-1111.
31. da Silva AC, Ferro JA, Reinach FC, Farah FC, Furlan LR, Quaggio RB, Monteiro-Vitorello CC, Van Sluys MA, Almeida NF, Alves LMC do Amaral AM, Bertolini MC, Camargo LE, Camarotte G, Cannavan F, Cardozo J, Chambergo F, Ciapina LP, Cicarelli RM, Coutinho LL, Cursino-Santos JR, El-Dorry H, Faria JB, Ferreira AJ, Ferreira RC, Ferro MI, Formighieri EF, Franco MC, Greggio CC, Gruber A, Katsuyama AM, Kishi LT, Leite RP, Lemos EG, Lemos MV, Locali EC, Machado MA, Madeira AM, Martinez-Rossi NM, Martins EC, Meidanis J, Menck CF, Miyaki CY, Moon DH, Moreira LM, Novo MT, Okura VK, Oliveira MC, Oliveira VR, Pereira HA, Rossi A, Sena JA, Silva C, de Souza RF, Spinola LA, Takita MA, Tamura RE, Teixeira EC, Tezza RI, Trindade dos Santos M, Truffi D, Tsai SM, White FF, Setubal JC, Kitajima JP: **Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities.** *Nature* 2002, **417**(6887):459-463.
32. Thieme F, Koebnik R, Bekel T, Berger C, Boch J, Büttner D, Caldana C, Gaigalat L, Goesmann A, Kay S Kirchner O, Lanz C, Linke B, McHardy AC, Meyer F, Mittenhuber G, Nies DH, Niesbach-Klosgen U, Patschkowski T, Ruckert C, Rupp O, Schneiker S, Schuster SC, Vorholter FJ, Weber E, Puhler A, Bonas U, Bartels D, Kaiser O: **Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* revealed by the complete genome sequence.** *J Bacteriol* 2005, **187**(21):7254-7266.
33. Roumagnac P, Gagnevin L, Gardan L, Sutra L, Manceau C, Dickstein ER, Jones J, Rott P, Pruvost O: **Polyphasic characterization of xanthomonads isolated from onion, garlic and Welsh onion (*Allium* spp.) and their relatedness to different *Xanthomonas* species.** *Int J Syst Evol Microbiol* 2004, **54**(Pt 1):15-24.

34. Wojciechowski MF, Lavin M, Sanderson MJ: **A phylogeny of legumes (*Leguminosae*) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family.** *Am J Bot* 2004, **91**:1846–1862.
35. Chen SL, Hung CS, Pinkner JS, Walker JN, Cusumano CK, Li Z, Bouckaert J, Gordon J, Hultgren S: **Positive selection identifies an in vivo role for FimH during urinary tract infection in addition to mannose binding.** *Proc Natl Acad Sci U S A* 2009, **106**(52):22439-22444.
36. He YQ, Zhang L, Jiang BL, Zhang ZC, Xu RQ, Tang DJ, Qin J, Jiang W, Zhang X, Liao J, Cao JR, Zhang SS, Wei ML, Liang XX, Lu GT, Feng JX, Chen B, Cheng J, Tang JL: **Comparative and functional genomics reveals genetic diversity and determinants of host specificity among reference strains and a large collection of Chinese isolates of the phytopathogen *Xanthomonas campestris* pv. *campestris*.** *Genome Biol* 2007, **8**(10):R218.
37. Charlesworth J, Eyre-Walker A: **The rate adaptive evolution in enteric bacteria.** *Mol Biol Evol* 2006, **23**:1348–1356.
38. Chen SL, Hung CS, Xu J, Reigstad CS, Magrini V, Sabo A, Blasiar D, Bieri T, Meyer RR, Ozersky P, Armstrong JR, Fulton RS, Latreille JP, Spieth J, Hooton TM, Mardis ER, Hultgren SJ, Gordon JI: **Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach.** *Proc Natl Acad Sci U S A* 2006, **103**(15):5977-5982.
39. Petersen L, Bollback JP, Dimmic M, Hubisz M, Nielsen R: **Genes under positive selection in *Escherichia coli*.** *Genome Res* 2007, **17**(9):1336-1343.
40. Weber E, Koebnik R: **Positive selection of the Hrp pilin HrpE of the plant pathogen *Xanthomonas*.** *J Bacteriol* 2006, **188**(4):1405-1410.
41. Manson MD, Armitage JP, Hoch JA, Macnab RM: **Bacterial locomotion and signal transduction.** *J Bacteriol* 1998, **180**: 1009–1022.
42. Bernstein HD: **Are ‘autotransporters’ really transporters?** *Trends in Microbiology* 2007, **15**: 441–447.
43. Gerlach RG, Hensel M: **Protein secretion systems and adhesins: the molecular armory of Gram-negative pathogens.** *Int J Med Microbiol* 2007, **297**(6):401-415.

- 44; Wang L, Makino S, Subedee A, Bogdanove AJ: **Novel candidate virulence factors in Rice pathogen *Xanthomonas oryzae* pv. *oryzicola* as revealed by mutational analysis.** *Appl Environ Microbiol* 2007, **73**: 8023–8027.
45. Darsonval A, Darrasse A, Durand K, Bureau C, Cesbron S, Jacques MA: **Adhesion and fitness in the bean phyllosphere and transmission to seed of *Xanthomonas fuscans* subsp. *fuscans*.** *Mol Plant Microbe Interact* 2009, **22**(6):747-757.
46. Cohan FM: **Sexual isolation and speciation in bacteria.** *Genetica* 2002, **116**(2-3):359-370.
47. Galperin MY: **A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts.** *BMC Microbiol* 2005, **5**:35.
48. Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG: ***Drosophila* chemoreceptor gene evolution: selection, specialization and genome size.** *Mol Ecol* 2008, **17**(7):1648-1657.
49. Young JM, Park DC, Shearman HM, Fargier E: **A multilocus sequence analysis of the genus *Xanthomonas*.** *Syst Appl Microbiol* 2008, **31**(5):366-377.
50. Morris CE, Sands DC, Vinatzer BA, Glaux C, Guilbaud C, Buffière A, Yan S, Dominguez H, Thompson BM: **The life history of the plant pathogen *Pseudomonas syringae* is linked to the water cycle.** *ISME J* 2008, **2**: 321–334.
51. Vauterin L, Rademaker J, Swings J: **Synopsis on the taxonomy of the genus *Xanthomonas*.** *Phytopathology* 2000, **90**(7):677-682.
52. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman D: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**(17):3389-3402.
53. Hammer Ø, Harper, DAT, Ryan, PD: **Past: paleontological statistics software package for education and data analysis.** *Palaeontologia Electronica* 2001, **4**:art. 4.
54. 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.

55. Hall TA: **Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucleic Acids Symposium Series* 1999, **41**:95–98.

56. Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Tasneem A, Thanki N, Yamashita RA, Zhang D, Zhang N, Bryant SH: **CDD: specific functional annotation with the Conserved Domain Database.** *Nucleic Acids Res* 2009, **37**: 205–10.

57. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R: **DnaSP, DNA polymorphism analyses by the coalescent and other methods.** *Bioinformatics* 2003, **19**(18):2496-2497.

58. McDonald JH, Kreitman M: **Adaptive protein evolution at the Adh locus in Drosophila.** *Nature* 1991, **351**(6328):652-654.

59. Rice WR: **Analyzing tables of statistical tests.** *Evolution* 1989, **43**(223-225).

60. Yang Z, Nielsen R: **Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages.** *Mol Biol Evol* 2002, **19**(6):908-917.

61. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood.** *Comput Appl Biosci* 1997, **13**(5):555-556.

Table A1 Genome mining: identification of genes involved in sensing and adhesion in xanthomonads. Genes were extracted from 4 complete sequenced genomes of xanthomonads available in NCBI website

	Gene product	<i>X. axonopodis</i> pv. <i>vesicatoria</i> str. 85-10			<i>X. axonopodis</i> pv. <i>citri</i> str. 306			<i>X. oryzae</i> pv. <i>oryzae</i> KACC10331			<i>X. campestris</i> pv. <i>campestris</i> ATCC33913		
		Gene name	Locus tag	Homology (%)	Gene name	Locus tag	Homology (%)	Gene name	Locus tag	Homology (%)	Gene name	Locus tag	Homology (%)
-accepting Chemotaxis Proteins (MCPs)	methyl-accepting chemotaxis protein	-	XCV0669	100%	<i>tsr</i>	XAC0611	97%	-	-	-	<i>tsr</i>	XCC3522	95%
	methyl-accepting chemotaxis protein	-	XCV1702	100%	<i>tsr</i>	XAC1666	95%	-	-	-	<i>tsr</i>	XCC2047	78%
	methyl-accepting chemotaxis protein	-	XCV1933	100%	-	-	-	-	-	-	<i>tsr</i>	XCC1869	84%
	methyl-accepting chemotaxis protein	-	XCV1938	100%	-	-	-	-	-	-	<i>tsr</i>	XCC1873 and 1874	86%
	methyl-accepting chemotaxis protein	-	XCV1939	100%	-	-	-	<i>tsr</i>	XOO2848	90%	-	-	-
	methyl-accepting chemotaxis protein	-	XCV1940	100%	<i>tsr</i>	XAC1891	93%	<i>tsr</i>	XOO2847	94%	<i>tsr</i>	XCC1875	82%
	methyl-accepting chemotaxis protein	-	XCV1941	100%	<i>tsr</i>	XAC1892	96%	-	-	-	<i>tsr</i>	XCC1876	84%
	methyl-accepting chemotaxis protein	-	XCV1942	100%	<i>tsr</i>	XAC1893	89%	-	-	-	-	-	-
	methyl-accepting chemotaxis protein	-	XCV1944	100%	<i>tsr</i>	XAC1894	96%	<i>tsr</i>	XOO2844	87%	-	-	-
	methyl-accepting chemotaxis protein	-	XCV1945	100%	<i>tsr</i>	XAC1895	91%	<i>tsr</i>	XOO2842	83%	<i>tsr</i>	XCC1878	81%
	putative methyl-accepting chemotaxis protein	-	XCV1948	100%	<i>tsr</i>	XAC1897	95%	<i>tsr</i>	XOO2840	87%	-	-	-
	methyl-accepting chemotaxis protein	-	XCV1947	100%	<i>tsr</i>	XAC1896	97%	-	-	-	<i>tsr</i>	XCC1879	89%
	methyl-accepting chemotaxis protein	-	XCV1951	100%	<i>tsr</i>	XAC1899	97%	-	-	-	<i>tsr</i>	XCC1880	83%
	methyl-accepting chemotaxis protein	-	XCV1952	100%	<i>tsr</i>	XAC1900	97%	-	-	-	<i>tsr</i>	XCC1881	89%
	methyl-accepting chemotaxis protein	-	XCV1954	100%	<i>tsr</i>	-	-	-	-	-	-	-	-
	methyl-accepting chemotaxis protein	-	XCV1955	100%	<i>tsr</i>	XAC1902	95%	-	-	-	-	-	-
	methyl-accepting chemotaxis protein	-	XCV2044	100%	<i>mcp</i>	XAC1996	100%	<i>mcp</i>	XOO2558	99%	<i>mcp</i>	XCC1962	98%
	methyl-accepting chemotaxis protein	-	XCV2625	100%	<i>mcp</i>	XAC2448	98%	-	-	-	<i>mcp</i>	XCC2315	90%
	methyl-accepting chemotaxis protein	-	XCV3021	100%	<i>mcp (tar)</i>	XAC2866	98%	<i>mcp (tar)</i>	XOO1469	95%	<i>mcp</i>	XCC2701	91%
	methyl-accepting chemotaxis protein	-	XCV3261	100%	<i>mcp</i>	XAC3132	97%	-	-	-	-	-	-
	methyl-accepting chemotaxis protein	-	XCV3338	100%	<i>mcp1</i>	XAC3213	97%	-	-	-	<i>mcp1</i>	XCC3084	98%
	putative methyl-accepting chemotaxis protein	-	XCV1778	100%	<i>mcpA</i>	XAC1746	97%	-	-	-	<i>mcpA</i>	XCC1727	82%
	methyl-accepting chemotaxis protein	-	XCV3577	100%	<i>tar</i>	XAC3449	97%	-	-	-	<i>tar</i>	XCC3321	91%
methyl-accepting chemotaxis protein	<i>pilJ</i>	XCV3230	100%	<i>pilJ</i>	XAC3099	96%	<i>pilJ</i>	XOO1749	93%	<i>pilJ</i>	XCC2922	96%	
chemotaxis transducer	-	-	-	<i>tcp</i>	XAC3271	100%	-	-	-	-	-	-	

Methyl-	chemotaxis protein	-	-	-	-	XAC3768	100%	-	-	-	-	-	-
	chemotaxis protein	-	-	-	-	-	-	-	XOO2840	100%	-	-	-
	chemotaxis protein	-	-	-	-	-	-	-	XOO2842	100%	-	-	-
	chemotaxis protein	-	-	-	-	-	-	-	-	-	tsr	XCC0324	100%
	chemotaxis protein	-	-	-	-	-	-	-	-	-	tsr	XCC0276	100%
Chemotactic proteins	chemotaxis protein CheA	<i>cheA1</i>	XCV1956	100%	<i>cheA</i>	XAC1903	98%	<i>cheA</i>	XOO2836	98%	<i>cheA</i>	XCC1885	95%
	chemotaxis protein CheA	<i>cheA2</i>	XCV1974	100%	<i>cheA</i>	XAC1930	99%	<i>cheA</i>	XOO2624	98%	<i>cheA</i>	XCC1903	93%
	chemotaxis protein CheA	<i>cheA3</i>	XCV3020	100%	<i>cheA</i>	XAC2865	98%	<i>cheA</i>	XOO1470	88%	-	-	-
	chemotaxis response regulator protein-	<i>cheB</i>	XCV1330	100%	<i>cheB</i>	XAC1280	95%	-	-	-	<i>cheB</i>	XCC1183	91%
	chemotaxis-specific methyltransferase	<i>cheB1</i>	XCV1930	100%	<i>cheB</i>	XAC1888	100%	<i>cheB</i>	XOO2859	98%	<i>cheB</i>	XCC1866	98%
	chemotaxis-specific methyltransferase	<i>cheB2</i>	XCV3025	100%	<i>cheB</i>	XAC2870	99%	<i>cheB</i>	XOO1465	94%	<i>cheB</i>	XCC2705	98%
	chemotaxis protein CheD	<i>CheD</i>	XCV1931	100%	<i>cheD</i>	XAC1889	99%	<i>cheD</i>	XOO2858	97%	<i>cheD</i>	XCC1865	97%
	methylase of chemotaxis methyl-accepting	<i>cheR1</i>	XCV1331	100%	<i>cheR</i>	XAC1281	99%	-	-	-	<i>cheR</i>	XCC1184	97%
	methylase of chemotaxis methyl-accepting	<i>cheR2</i>	XCV1932	100%	<i>cheR</i>	XAC1281	99%	<i>cheR</i>	-	-	<i>cheR</i>	XCC1184	97%
	methylase of chemotaxis methyl-accepting	<i>cheR3</i>	XCV3024	100%	<i>cheR</i>	XAC2869	98%	<i>cheR</i>	XOO1466	95%	<i>cheR</i>	XCC2704	90%
	methylase of chemotaxis methyl-accepting	<i>cheR4</i>	XCV3849	100%	<i>cheR</i>	XAC3730	92%	-	XOO0644	88%	<i>cheR</i>	XCC3686	81%
	chemotaxis signal transduction protein	-	XCV2034	100%	<i>cheV</i>	XAC1987	99%	<i>cheV</i>	XOO2568	99%	<i>cheV</i>	XCC1953	97%
	chemotaxis signal transduction protein	<i>cheW1</i>	XCV1935	100%	-	-	-	-	-	-	<i>cheW</i>	XCC1871	98%
	chemotaxis signal transduction protein	<i>cheW2</i>	XCV1959	100%	<i>cheW</i>	XAC1906	96%	<i>cheW</i>	XOO2833	66%	<i>cheW</i>	XCC1888	78%
	chemotaxis signal transduction protein	<i>cheW2</i>	XCV2624	100%	<i>cheW</i>	XAC2447	98%	-	XOO1468	66%	<i>cheW</i>	XCC2314	90%
	chemotaxis signal transduction protein	<i>cheW3</i>	XCV3022	100%	<i>cheW</i>	XAC2867	95%	<i>cheW</i>	XOO1468	91%	<i>cheW</i>	XCC2702	84%
	chemotaxis phosphatase CheZ	<i>cheZ</i>	XCV1975	100%	<i>cheZ</i>	XAC1931	100%	<i>cheZ</i>	XOO2623	96%	<i>cheZ</i>	XCC1904	91%
ients	flagellar motor protein	<i>motC</i>	XCV1962	100%	<i>motC</i>	XAC1909	98%	<i>motC</i>	XCC1891	97%	<i>motC</i>	XCC1891	97%
	flagellar motor protein	<i>motA</i>	XCV3814	100%	<i>motA</i>	XAC3693	99%	-	XOO0688	100%	<i>motA</i>	XCC3653	97%
	flagellar motor protein	<i>motB1</i>	XCV1961	100%	<i>motB</i>	XAC1908	98%	<i>motB</i>	XOO2831	96%	<i>motB</i>	XCC1890	94%
	flagellar motor protein	<i>motB2</i>	XCV3815	100%	<i>motB</i>	XAC3694	99%	<i>motB</i>	XOO0687	95%	<i>motB</i>	XCC3654	96%
	flagellar basal body P-ring biosynthesis	<i>flgA</i>	XCV2035	100%	<i>flgA</i>	XAC1988	97%	<i>flgA</i>	XOO2567	97%	<i>flgA</i>	XCC1954	96%
	flagellar basal body protein	<i>flgB</i>	XCV2033	100%	<i>flgB</i>	XAC1986	100%	<i>flgB</i>	XOO2569	95%	<i>flgB</i>	XCC1952	90%
	flagellar basal body rod protein	<i>flgC</i>	XCV2032	100%	<i>flgC</i>	XAC1985	99%	<i>flgC</i>	XOO2570	99%	<i>flgC</i>	XCC1951	98%
	flagellar hook capping protein	<i>flgD</i>	XCV2031	100%	<i>flgD</i>	XAC1984	99%	<i>flgD</i>	XOO2571	92%	<i>flgD</i>	XCC1950	94%
	flagellar hook protein FlgE	<i>flgE</i>	XCV2030	100%	<i>flgE</i>	XAC1983	99%	<i>flgE</i>	XOO2572	96%	<i>flgE</i>	XCC1949	96%
	flagella basal body rod protein	<i>flgF</i>	XCV2029	100%	<i>flgF</i>	XAC1982	98%	<i>flgF</i>	XOO2574	98%	<i>flgF</i>	XCC1948	98%
	flagella basal body rod protein	<i>flgG</i>	XCV2028	100%	<i>flgG</i>	XAC1981	100%	<i>flgG</i>	XOO2575	98%	<i>flgG</i>	XCC1947	99%
	flagellar L-ring protein precursor	<i>flgH</i>	XCV2027	100%	<i>flgH</i>	XAC1980	98%	<i>flgH</i>	XOO2576	85%	<i>flgH</i>	XCC1946	93%
	flagellar P-ring protein precursor	<i>flgI</i>	XCV2026	100%	<i>flgI</i>	XAC1979	99%	<i>flgI</i>	XOO2577	97%	<i>flgI</i>	XCC1945	97%

Flagellar compon	muramidase (flagellum-specific)	<i>flgJ</i>	XCV2025	100%	<i>flgJ</i>	XAC1978	98%	<i>flgJ</i>	XOO2578	96%	<i>flgJ</i>	XCC1944	90%
	flagellar hook-associated protein 1 (HAP1)	<i>flgK</i>	XCV2024	100%	<i>flgK</i>	XAC1977	99%	<i>flgK</i>	XOO2579	96%	<i>flgK</i>	XCC1943	91%
	flagellin and related hook-associated proteins	<i>flgL</i>	XCV2023	100%	<i>flgL</i>	XAC1976	98%	<i>flgL</i>	XOO2580	93%	<i>flgL</i>	XCC1942	91%
	flagellar biosynthesis pathway component	<i>flhA</i>	XCV1980	100%	<i>flhA</i>	XAC1936	99%	<i>flhA</i>	XOO2618	97%	<i>flhA</i>	XCC1909	96%
	flagellar biosynthesis pathway component	<i>flhB</i>	XCV1981	100%	<i>flhB</i>	XAC1937	98%	-	XOO2617	94%	<i>flhB</i>	XCC1910	93%
	flagellar GTP-binding protein FlhF	<i>flhF</i>	XCV1979	100%	<i>flhF</i>	XAC1935	97%	<i>flhF</i>	XOO2619	84%	<i>flhF</i>	XCC1908	81%
	flagellin and related hook-associated proteins	<i>fliC</i>	XCV2022	100%	<i>fliC</i>	XAC1975	88%	<i>fliC</i>	XOO2581	97%	<i>fliC</i>	XCC1941	88%
	flagellar capping protein	<i>fliD</i>	XCV2021	100%	<i>fliD</i>	XAC1974	96%	<i>fliD</i>	XOO2582	96%	<i>fliD</i>	XCC1940	88%
	flagellar hook-basal body complex protein	<i>fliE</i>	XCV1999	100%	<i>fliE</i>	XAC1955	96%	<i>fliE</i>	XOO2600	95%	<i>fliE</i>	XCC1927	94%
	flagellar biosynthesis lipoprotein FliF	<i>fliF</i>	XCV1998	100%	<i>fliF</i>	XAC1954	99%	<i>fliF</i>	XOO2601	94%	<i>fliF</i>	XCC1926	93%
	flagellar motor switch protein FliG	<i>fliG</i>	XCV1997	100%	<i>fliG</i>	XAC1953	100%	<i>fliG</i>	XOO2602	97%	<i>fliG</i>	XCC1925	97%
	flagellar biosynthesis/type III secretory	<i>fliH</i>	XCV1996	100%	<i>fliH</i>	XAC1952	99%	<i>fliH</i>	XOO2603	88%	<i>fliH</i>	XCC1924	90%
	flagellar biosynthesis/type III secretory	<i>fliI</i>	XCV1995	100%	<i>fliI</i>	XAC1951	99%	<i>fliI</i>	XOO2604	98%	<i>fliI</i>	XCC1924	97%
	flagellar hook-length control protein FliK	<i>fliK</i>	XCV1993	100%	<i>fliK</i>	XAC1949	92%	<i>fliK</i>	XOO2606	87%	-	XCC1921	72%
	flagellar basal body-associated protein FliL	<i>fliL</i>	XCV1992	100%	<i>fliL</i>	XAC1948	98%	<i>fliL</i>	XOO2607	91%	<i>fliL</i>	XCC1920	97%
	flagellar motor switch protein	<i>fliM</i>	XCV1991	100%	<i>fliM</i>	XAC1947	99%	<i>fliM</i>	XOO2608	97%	<i>fliM</i>	XCC1919	98%
	flagellar motor switch/type III secretory	<i>fliN</i>	XCV1990	100%	<i>fliN</i>	XAC1946	99%	<i>fliN</i>	XOO2609	98%	<i>fliN</i>	XCC1918	94%
	flagellar biosynthesis pathway component	<i>fliP</i>	XCV1988	100%	<i>fliP</i>	XAC1944	100%	<i>fliP</i>	XOO2611	97%	<i>fliP</i>	XCC1916	96%
	flagellar biosynthesis pathway component	<i>fliQ</i>	XCV1987	100%	<i>fliQ</i>	XAC1942	100%	<i>fliQ</i>	XOO2612	100%	<i>fliQ</i>	XCC1915	96%
	flagellar biosynthesis pathway component	<i>fliR</i>	XCV1986	100%	<i>fliR</i>	XAC1941	99%	<i>fliR</i>	XOO2613	97%	<i>fliR</i>	XCC1914	93%
flagellin-specific chaperone FliS	<i>fliS</i>	XCV2020	100%	<i>fliS</i>	XAC1973	97%	<i>fliS</i>	XOO2583	98%	<i>fliS</i>	XCC1939	94%	
fimbrial adhesins (Tfp	PilA-related fimbrial protein	-	XCV3930	62%	<i>pilA</i>	XAC3805	62%	<i>pilA</i>	XOO0581	60%	<i>pilA</i>	XCC3753	100%
	Tfp pilus assembly pathway ATPase PilB	<i>pilB</i>	XCV3352	94%	<i>pilB</i>	XAC3239	94%	<i>pilB</i>	XOO1589	95%	<i>pilB</i>	XCC3097	100%
	Fimbrial assembly protein	<i>pilC</i>	XCV3354	86%	<i>pilC</i>	XAC3242	87%	<i>pilC</i>	XOO1585	88%	<i>pilC</i>	XCC3100	100%
	type IV pre-pilin leader peptidase	<i>pilD</i>	XCV3355	97%	<i>pilD</i>	XAC3243	97%	<i>pilD</i>	XOO1584	98%	<i>pilD</i>	XCC3101	100%
	Tfp pilus assembly protein major pilin	<i>pilA</i>	XCV3353	50%	-	-	-	<i>pilE</i>	XOO1586	46%	<i>pilE</i>	XCC3099	100%
	type IV pilin PilE	<i>pilE</i>	XCV2814	72%	-	-	-	<i>pilE1</i>	XOO3195	55%	<i>pilE1</i>	XCC2488	100%
	putative secreted protein	-	XCV2068	86%	<i>pilF</i>	XAC2017	86%	<i>pilF</i>	XOO2533	84%	<i>pilF</i>	XCC1983	100%
	type IV pilus response regulator PilG	<i>pilG</i>	XCV3233	97%	<i>pilG</i>	XAC3102	97%	<i>pilG</i>	XOO1746	97%	<i>pilG</i>	XCC2925	100%
	PilH protein	<i>pilH</i>	XCV3232	97%	<i>pilH</i>	XAC3101	97%	<i>pilH</i>	XOO1747	95%	<i>pilH</i>	XCC2924	100%
	chemotaxis signal transduction protein	<i>pilI</i>	XCV3231	99%	<i>pilI</i>	XAC3100	98%	<i>pilI</i>	XOO1748	99%	<i>pilI</i>	XCC2923	100%
	two-component system sensor histidine	<i>pilL</i>	XCV3229	73%	<i>pilL</i>	XAC3098	69%	<i>pilL</i>	XOO1751	95%	<i>pilL</i>	XCC2921	100%
	Tfp pilus assembly protein	<i>pilM</i>	XCV3501	99%	<i>pilM</i>	XAC3385	98%	<i>pilM</i>	XOO1157	97%	<i>pilM</i>	XCC3232	100%

Type IV pilus biogenesis and	Tfp pilus assembly protein	<i>pilN</i>	XCV3500	97%	<i>pilN</i>	XAC3384	96%	<i>pilN</i>	XOO1158	95%	<i>pilN</i>	XCC3231	100%
	Fimbrial assembly membrane protein	<i>pilO</i>	XCV3499	97%	<i>pilO</i>	XAC3383	97%	<i>pilO</i>	XOO1159	96%	<i>pilO</i>	XCC3230	100%
	Tfp pilus assembly protein	<i>pilP</i>	XCV3498	99%	<i>pilP</i>	XAC3382	99%	<i>pilP</i>	XOO1160	98%	<i>pilP</i>	XCC3229	100%
	fimbrial assembly protein	<i>pilQ</i>	XCV3497	92%	<i>pilQ</i>	XAC3381	87%	<i>pilQ</i>	XOO1161	92%	<i>pilQ</i>	XCC3228	100%
	two-component system response regulator	<i>pilR</i>	XCV3351	95%	<i>pilR</i>	XAC3238	94%	<i>pilR</i>	XOO1591	95%	<i>pilR</i>	XCC3096	100%
	two-component system sensor protein PilS	<i>pilS</i>	XCV3350	90%	<i>pilS</i>	XAC3237	89%	<i>pilS</i>	XOO1592	90%	<i>pilS</i>	XCC3095	100%
	twitching motility protein	-	XCV3068	100%	<i>pilT</i>	XAC2924	100%	<i>pilT</i>	XOO1418	99%	<i>pilT</i>	XCC2755	100%
	Tfp pilus assembly protein ATPase	<i>pilU</i>	XCV3067	99%	<i>pilU</i>	XAC2923	99%	<i>pilU</i>	XOO1419	97%	<i>pilU</i>	XCC2754	100%
	putative type IV pilus assembly protein PilV	-	XCV2820	75%	-	-	-	<i>pilV</i>	XOO3201	63%	<i>pilV</i>	XCC2494	100%
	PilW-related protein	-	XCV2819	69%	-	XAC2667	25%	-	XOO3200	60%	<i>pilW</i>	XCC2493	100%
	PilX-related protein	-	XCV2818	68%	<i>pilX</i>	XAC2666	31%	<i>pilX</i>	XOO3199	56%	<i>pilX</i>	XCC2492	100%
	tip-associated adhesin	-	XCV2815	67%	<i>pilY1</i>	XAC2665	31%	<i>pilY1</i>	XOO3196	56%	<i>pilY1</i>	XCC2489	100%
	tip-associated adhesin	-	XCV2815	62%	-	-	-	-	-	-	-	XCC2491	100%
	type IV fimbriae assembly protein	-	XCV1152	97%	<i>pilZ</i>	XAC1133	97%	<i>pilZ</i>	XOO0887	95%	<i>pilZ</i>	XCC1024	100%
	pilin	-	-	-	-	-	-	-	-	-	-	XCC3098	100%
	pili assembly chaperone	<i>papD1</i>	XCV1480	100%	-	XAC1423	93%	-	XOO1978	90%	-	XCC1376	90%
	pili assembly chaperone	<i>papD2</i>	XCV1483	100%	<i>ecpD</i>	XAC1426	96%	<i>ecpD</i>	XOO1981	92%	<i>ecpD</i>	XCC1380	84%
	putative type IV pilus assembly protein FimT	-	XCV2813	100%	-	-	-	<i>fimT</i>	XOO3194	84%	-	-	-
	putative type IV pilus assembly protein FimT	-	XCV2821	100%	-	-	-	-	-	-	-	-	-
	Nonfimbrial adhesins	Xanthomonas adhesin XadA	<i>xadA1</i>	XCV3670	100%	<i>xadA</i>	XAC3546	94%	-	-	-	XCC0658	85%
		Xanthomonas adhesin XadA	<i>xadA2</i>	XCV3672	100%	<i>xadA</i>	XAC3548	96%	-	-	-	-	-
outer membrane protein		-	-	-	-	-	<i>xadA</i>	XOO0842	100%	-	-	-	
outer membrane protein		-	-	-	-	-	-	XOO0681	100%	-	-	-	
putative filamentous hemagglutinin-like		-	XCV4444	100%	-	-	-	XOO4475	95%	-	XCC4201	83%	
filamentous hemagglutinin-related protein		<i>yapH</i>	XCV4203	100%	<i>yapH</i>	XAC4113	92%	-	-	-	-	-	
putative filamentous hemagglutinin-like		-	XCV2103	100%	<i>yapH</i>	XAC2151	94%	<i>yapH</i>	XOO2380	90%	<i>yapH</i>	XCC2024	75%
filamentous hemagglutinin-related protein		<i>fhaB1</i>	XCV1860	100%	<i>fhaB</i>	XAC1815	90%	-	-	-	<i>fhaB</i>	XCC1794	80%
filamentous hemagglutinin-related protein		<i>fhaB2</i>	XCV1861	100%				-	-	-	-	-	-
hemagglutinin/hemolysin-related protein		-	-	-	-	XAC1816	100%	-	-	-	-	-	-
	TonB-dependent outer membrane receptor	-	XCV0050	100%	<i>cirA</i>	XAC0074	96%	-	-	-	<i>cirA</i>	XCC0050	85%
	TonB-dependent outer membrane receptor	-	XCV0102	100%	-	XAC0126	96%	<i>yncD</i>	XOO0017	93%	<i>yncD</i>	XCC0098	89%
	TonB-dependent outer membrane receptor	-	XCV0123	100%	<i>iroN</i>	XAC0144	99%	-	-	-	<i>iroN</i>	XCC0119	91%
	TonB-dependent outer membrane receptor	-	XCV0159	100%	<i>fpvA</i>	XAC0176	97%	-	-	-	<i>fpvA</i>	XCC0158	90%
	TonB-dependent outer membrane receptor	-	XCV0573	100%	-	XAC0544	98%	-	-	-	-	XCC0531	92%
	TonB-dependent outer membrane receptor	-	XCV0714	100%	<i>fepA</i>	XAC0653	93%	-	-	-	<i>fepA</i>	XCC3474	87%

TonB-dependent outer membrane receptor	-	XCV0751	100%	<i>fecA</i>	XAC0690	97%	-	-	-	-	-	-
TonB-dependent outer membrane receptor	-	XCV0754	100%	<i>fecA</i>	XAC0693	96%	-	-	-	<i>fecA</i>	XCC3427	91%
TonB-dependent outer membrane receptor	-	XCV0772	100%	<i>fyuA</i>	XAC0716	98%	<i>fyuA</i>	XOO3909	93%	<i>fyuA</i>	XCC3408	95%
TonB-dependent outer membrane receptor	-	XCV0863	100%	<i>cirA</i>	XAC0811	98%	<i>cirA</i>	XOO3793	93%	<i>cirA</i>	XCC0759	91%
TonB-dependent outer membrane receptor	-	XCV0875	100%	<i>phuR</i>	XAC0823	92%	-	-	-	<i>phuR</i>	XCC0768	89%
TonB-dependent outer membrane receptor	-	XCV1027	100%	<i>cirA</i>	XAC0999	97%	<i>cirA</i>	XOO0978	96%	-	-	-
TonB-dependent outer membrane receptor	-	XCV1053	100%	<i>fecA</i>	XAC1023	96%	<i>fecA</i>	XOO3683	95%	<i>fecA</i>	XCC0946	92%
TonB-dependent outer membrane receptor	-	XCV1161	100%	<i>fyuA</i>	XAC1143	98%	<i>fyuA</i>	XOO0897	97%	<i>fyuA</i>	XCC1037	94%
TonB-dependent outer membrane receptor	-	XCV1165	100%	<i>fecA</i>	XAC1146	93%	<i>fecA</i>	XOO0901	91%	<i>fecA</i>	XCC1041	85%
TonB-dependent outer membrane receptor	-	XCV1326	100%	<i>fyuA</i>	XAC1276	98%	-	-	-	<i>fyuA</i>	XCC1179	93%
TonB-dependent outer membrane receptor	-	XCV1361	100%	<i>btuB</i>	XAC1310	97%	-	-	-	<i>btuB</i>	XCC1258	90%
TonB-dependent outer membrane receptor	-	XCV1801	100%	<i>cirA</i>	XAC1769	99%	-	-	-	<i>cirA</i>	XCC1750	92%
TonB-dependent outer membrane receptor	-	XCV1963	100%	<i>cirA</i>	XAC1910	97%	<i>cirA</i>	XOO2826	94%	<i>cirA</i>	XCC1892	97%
TonB-dependent outer membrane receptor	-	XCV2075	100%	<i>cirA</i>	XAC2024	93%	-	-	-	<i>cirA</i>	XCC1990	86%
TonB-dependent outer membrane receptor	-	XCV2510	100%	-	XAC2312	99%	-	XOO2166	96%	-	XCC2208	94%
TonB-dependent outer membrane receptor	-	XCV2702	100%	<i>cirA</i>	XAC2520	98%	<i>cirA</i>	XOO2500	90%	<i>cirA</i>	XCC2385	92%
TonB-dependent outer membrane receptor	-	XCV2725	100%	<i>btuB</i>	XAC2531	97%	-	-	-	<i>btuB</i>	XCC2395	93%
TonB-dependent outer membrane receptor	-	XCV2730	100%	<i>btuB</i>	XAC2535	94%	<i>btuB</i>	XOO3123	90%	<i>btuB</i>	XCC2400	87%
TonB-dependent outer membrane receptor	-	XCV2801	100%	<i>btuB</i>	XAC2600	98%	-	-	-	<i>btuB</i>	XCC2469	94%
TonB-dependent outer membrane receptor	-	XCV2823	100%	<i>oar</i>	XAC2672	99%	<i>oar</i>	XOO3204	87%	<i>oar</i>	XCC2497	80%
TonB-dependent outer membrane receptor	-	XCV2894	100%	<i>btuB</i>	XAC2894	97%	-	-	-	<i>btuB</i>	XCC257	94%
TonB-dependent outer membrane receptor	-	XCV2989	100%	<i>phuR</i>	XAC2829	98%	-	-	-	<i>phuR</i>	XCC2658	93%
TonB-dependent outer membrane receptor	-	XCV2991	100%	<i>fhuA</i>	XAC2830	97%	<i>fhuA</i>	XOO1530	94%	<i>fhuA</i>	XCC2665	92%
TonB-dependent outer membrane receptor	-	XCV3085	100%	<i>fhuA</i>	XAC2941	98%	-	-	-	<i>fhuA</i>	XCC277	95%
TonB-dependent outer membrane receptor	-	XCV3147	100%	<i>fecA</i>	XAC2998	99%	<i>fecA</i>	XOO1260	97%	<i>fepA</i>	XCC2828	91%
TonB-dependent outer membrane receptor	-	XCV3187	100%	-	-	-	-	-	-	-	-	-
TonB-dependent outer membrane receptor	-	XCV3206	100%	<i>iroN</i>	XAC3071	84%	<i>iroN</i>	XOO1784	89%	<i>iroN</i>	XCC2887	84%
TonB-dependent outer membrane receptor	-	XCV3252	100%	<i>fepA</i>	XAC3121	97%	<i>fepA</i>	XOO1727	93%	<i>fepA</i>	XCC2944	95%
TonB-dependent outer membrane receptor	-	XCV3290	100%	<i>fhuA</i>	XAC3158	96%	-	-	-	<i>fhuA</i>	XCC3036	93%
TonB-dependent outer membrane receptor	-	XCV3297	100%	<i>bfeA</i>	XAC3166	97%	<i>bfeA</i>	XOO1366	95%	<i>bfeA</i>	XCC3043	96%
TonB-dependent outer membrane receptor	-	XCV3299	100%	<i>bfeA</i>	XAC3168	98%	-	XOO1364	96%	<i>bfeA</i>	XCC3045	95%
TonB-dependent outer membrane receptor	-	XCV3301	100%	<i>bfeA</i>	XAC3169	98%	<i>bfeA</i>	XOO1363	95%	<i>bfeA</i>	XCC3046	83%
TonB-dependent outer membrane receptor	-	XCV3306	100%	<i>fecA</i>	XAC3176	97%	<i>fecA</i>	XOO1359	92%	<i>fecA</i>	XCC3050	89%
TonB-dependent outer membrane receptor	-	XCV3333	100%	<i>bfeA</i>	XAC3207	98%	<i>bfeA</i>	XOO1604	93%	<i>bfeA</i>	XCC3079	95%
TonB-dependent outer membrane receptor	-	XCV3489	100%	<i>fhuE</i>	XAC3370	91%	-	-	-	<i>fhuE</i>	XCC3215	91%

TonB-Dependent Transporters (TBDTs)

TonB-dependent outer membrane receptor	-	XCV3542	100%	<i>fhuA</i>	XAC3427	95%	-	-	-	<i>fhuA</i>	XCC3279	92%
TonB-dependent outer membrane receptor	-	XCV3572	100%	<i>btuB</i>	XAC3444	99%	-	-	-	-	-	-
TonB-dependent outer membrane receptor	-	XCV3576	100%	<i>btuB</i>	XAC3448	98%	-	-	-	-	-	-
TonB-dependent outer membrane receptor	-	XCV3617	100%	<i>fyuA</i>	XAC3489	98%	<i>fyuA</i>	XOO1099	93%	<i>fyuA</i>	XCC335	93%
TonB-dependent outer membrane receptor	-	XCV3654	100%	<i>bfrC</i>	XAC3529	96%	<i>bfrC</i>	XOO0861	94%	-	XCC0674	88%
TonB-dependent outer membrane receptor	-	XCV3685	100%	<i>btuB</i>	XAC3560	98%	-	-	-	<i>btuB</i>	XCC3405	96%
TonB-dependent outer membrane receptor	-	XCV3875	100%	-	XAC3756	96%	-	XOO0626	92%	-	XCC3714	88%
TonB-dependent outer membrane receptor	-	XCV4138	100%	<i>iroN</i>	XAC4048	98%	<i>iroN</i>	XOO0394	96%	<i>iroN</i>	XCC396	93%
TonB-dependent outer membrane receptor	-	XCV4224	100%	-	XAC4131	96%	-	-	-	-	-	-
TonB-dependent outer membrane receptor	-	XCV4362	100%	<i>cirA</i>	XAC4256	98%	<i>cirA</i>	XOO4431	96%	<i>cirA</i>	XCC412	93%
TonB-dependent outer membrane receptor	-	XCV4374	100%	-	XAC4273	98%	-	XOO0250	96%	-	XCC4131	87%
TonB-dependent outer membrane receptor	-	XCV4468	100%	-	-	-	-	-	-	-	XCC4222	96%
TonB-dependent outer membrane receptor	-	XCV1492	100%	<i>fhuA</i>	XAC1435	96%	<i>fhuA</i>	XOO1992	92%	<i>fhuA</i>	XCC1391	90%
TonB-dependent outer membrane receptor	-	XCV1800	100%	<i>fhuA</i>	XAC1768	98%	-	-	-	<i>fhuA</i>	XCC1749	84%
TonB-dependent outer membrane receptor	-	XCV2896	100%	<i>oar</i>	XAC2743	84%	-	-	-	<i>oar</i>	XCC2573	80%
TonB-dependent outer membrane receptor	-	XCV3428	100%	<i>iroN</i>	XAC3311	99%	-	-	-	<i>iroN</i>	XCC316	93%
TonB-dependent outer membrane receptor	-	XCV3451	100%	<i>fecA</i>	XAC3334	97%	<i>fecA</i>	XOO3403	95%	<i>fecA</i>	XCC3177	86%
TonB-dependent outer membrane receptor	-	XCV3485	100%	<i>cirA</i>	XAC3366	98%	<i>cirA</i>	XOO1180	95%	<i>cirA</i>	XCC3209	90%
TonB-dependent outer membrane receptor	-	XCV4308	100%	-	-	-	-	XOO4396	79%	-	-	-
TonB-dependent outer membrane receptor	-	XCV4376	100%	-	XAC4274	99%	-	XOO0249	74%	-	-	-
TonB-dependent outer membrane receptor	-	XCV4480	100%	<i>fecA</i>	XAC4368	97%	-	-	-	<i>fecA</i>	XCC4235	97%
TonB-dependent outer membrane receptor	-	XCV1493	100%	<i>fhuA</i>	XAC1435	97%	<i>fhuA</i>	XOO1992	94%	<i>huA</i>	XCC1391	92%
TonB-dependent outer membrane receptor	-	XCV1799	100%	<i>fhuA</i>	XAC1768	99%	-	-	-	<i>fhuA</i>	XCC1749	80%
TonB-dependent outer membrane receptor	-	XCV3429	100%	<i>iroN</i>	XAC3311	98%	-	-	-	<i>iroN</i>	XCC3161	94%
TonB-dependent outer membrane receptor	-	XCV4307	100%	-	-	-	-	XOO4396	93%	-	-	-
TonB-dependent outer membrane receptor	-	XCV3534	100%	<i>oar</i>	XAC3418	98%	-	XOO1233	96%	-	XCC3270	96%
TonB-dependent outer membrane receptor	-	XCV3484	100%	<i>cirA</i>	XAC3366	91%	<i>cirA</i>	XOO1180	90%	<i>cirA</i>	XCC3209	82%
TonB-dependent outer membrane receptor	-	XCV4468	100%	-	-	-	-	-	-	-	XCC4222	94%
TonB-dependent outer membrane receptor	-	XCV4481	100%	<i>fecA</i>	XAC4368	99%	-	-	-	<i>fecA</i>	XCC4235	97%
TonB-dependent outer membrane receptor	-	XCV3323	100%	<i>btuB</i>	XAC3194	95%	<i>btuB</i>	XOO1345	92%	<i>btuB</i>	XCC3067	86%
TonB-dependent outer membrane receptor	-	XCV2895	100%	<i>oar</i>	XAC2743	95%	-	-	-	<i>oar</i>	XCC2573	93%
TonB-dependent outer membrane receptor	-	-	-	-	XAC0706	100%	<i>iroN</i>	XOO3924	96%	-	-	-
TonB-dependent outer membrane receptor	-	-	-	-	XAC0852	100%	-	-	-	-	-	-
TonB-dependent outer membrane receptor	-	-	-	-	XAC2193	100%	-	-	-	-	-	-
TonB-dependent outer membrane receptor	-	-	-	-	XAC3050	100%	-	-	-	-	-	-

TonB-dependent outer membrane receptor	-	-	-	-	XAC3077	100%	-	-	-	-	-	-
TonB-dependent outer membrane receptor	-	-	-	-	XAC3201	100%	-	-	-	-	-	-
TonB-dependent outer membrane receptor	-	-	-	-	XAC3613	100%	-	-	-	-	-	-
TonB-dependent outer membrane receptor	-	-	-	-	XAC4062	100%	-	-	-	-	-	-
TonB-dependent outer membrane receptor	-	-	-	-	XAC2185	100%	-	-	-	-	-	-
outer membrane receptor for transport of iron receptor	-	XCV3323	95%	-	XAC3194	100%	btuB	XOO1345	93%	btuB	XCC3067	88%
TonB-dependent outer membrane receptor	-	XCV3654	95%	-	XAC3530	100%	bfrC	XOO0861	93%	-	XCC0674	87%
outer membrane receptor for ferric iron	-	XCV2894	97%	-	XAC2742	100%	-	-	-	btuB	XCC2572	94%
Oar protein	-	-	-	-	XAC3498	100%	-	-	-	-	-	-
Oar protein	-	-	-	-	XAC0291	100%	-	-	-	-	-	-
outer membrane receptor FepA	-	-	-	-	XAC3620	100%	-	-	-	-	-	-
Oar protein	-	-	-	-	XAC2773	100%	oar	XOO3320	84%	-	-	-
TonB-dependent receptor	-	-	-	-	-	-	-	-	-	-	XCC0120	100%
TonB-dependent receptor	-	-	-	-	-	-	-	-	-	-	XCC0304	100%
TonB-dependent receptor	-	-	-	-	-	-	-	-	-	-	XCC0305	100%
TonB-dependent receptor	-	-	-	-	-	-	-	-	-	-	XCC0394	100%
TonB-dependent receptor	-	-	-	-	-	-	-	-	-	-	XCC0397	100%
TonB-dependent receptor	-	-	-	-	-	-	-	-	-	-	XCC1340	100%
TonB-dependent receptor	-	-	-	-	-	-	-	-	-	-	XCC1719	100%
TonB-dependent receptor	-	-	-	fhuA	XAC1768	75%	-	-	-	-	XCC1749	100%
TonB-dependent receptor	-	-	-	cirA	XAC1769	99%	-	-	-	-	XCC1751	100%
TonB-dependent receptor	-	-	-	-	-	-	-	-	-	-	XCC2046	100%
TonB-dependent receptor	-	-	-	-	-	-	-	-	-	-	XCC2867	100%
TonB-dependent receptor	-	-	-	-	-	-	btuB	XOO0950	89%	-	XCC3316	100%
OmpA-related protein	-	-	-	-	XAC4724	78%	-	XOO0249	95%	-	XCC4132	100%
ferric pseudobactin M114 receptor protein	-	-	-	-	-	-	-	-	-	-	XCC3595	100%
ferrichrome-iron receptor 3	-	-	-	-	-	-	-	-	-	-	XCC4162	100%
ferripyoverdine receptor	-	-	-	-	-	-	-	-	-	-	XCC3518	100%
hypothetical protein XCC3635	-	-	-	-	-	-	-	-	-	-	XCC3635	100%
outer membrane receptor for ferric iron	-	XCV3489	86%	fhuE	XAC3370	86%	-	-	-	-	XCC3216	100%
Oar protein	-	XCV3534	90%	oar	XAC3418	92%	-	XOO1233	90%	-	XCC3271	100%
Oar protein	-	-	-	-	-	-	-	-	-	-	XCC4237	100%
TonB-dependent receptor	-	-	-	-	-	-	-	-	-	-	XCC4052	100%
TonB-dependent receptor	-	XCV0123	89%	iroN	XAC0144	89%	-	XOO0029	100%	iroN	XCC0119	88%
TonB-dependent siderophore receptor,	-	-	-	-	-	-	-	XOO2829	100%	-	-	-

	TonB-dependent receptor	-	XCV2725	89%	<i>btuB</i>	XAC2531	90%	-	XOO3119	100%	<i>btuB</i>	XCC 2395	85%
	hypothetical protein XOO4285	-	XCV0573	94%	-	XAC0544	94%	-	XOO4285	100%	-	XCC0531	87%
	hypothetical protein XOO4286	-	XCV0574	84%	-	XAC0544	84%	-	XOO4286	100%	-	XCC0531	84%
	TonB-dependent receptor, putative	-	-	-	-	-	-	-	XOO3861	100%	-	-	-
/ system	sensor histidine kinase / response regulator	-	XCV3601	100%	-	XAC3473	98%	-	XOO1115	95%	-	XCC3348	88%
	sensory box histidine kinase	-	XCV2040	100%	-	XAC1994	98%	-	XOO2562	95%	-	XCC1960	88%
	sensory box histidine kinase/response	-	XCV2754	100%	<i>fixL</i>	XAC2555	96%	<i>fixL</i>	XOO3145	93%	<i>fixL</i>	XCC2421	87%
	sensory box histidine kinase/response	-	XCV1705	100%	-	XAC1669	98%	-	-	-	-	XCC1652	80%
	sensory/regulatory protein RpfC	-	XCV1919	100%	-	-	-	<i>rpfC</i>	XOO2870	92%	<i>rpfC</i>	XCC1856	89%
	Transmembrane sensor protein, FecR family	-	XCV4223	100%	-	XAC4130	93%	-	-	-	-	-	-
	two component system sensor protein	-	XCV2152	100%	-	-	-	-	-	-	-	-	-
	two-component system regulatory protein	-	XCV0115	100%	-	XAC0136	98%	-	-	-	-	XCC0109	91%
	two-component system sensor histidine	-	XCV0231	100%	-	XAC0225	98%	-	XOO4201	92%	-	-	-
	two-component system sensor histidine	-	XCV2188	100%	-	XAC2167	94%	-	-	-	-	-	-
	two-component system sensor histidine	-	XCV3761	100%	<i>styS</i>	XAC3643	98%	-	-	-	-	XCC0562	91%
	two-component system sensor histidine	-	XCV0746	100%	-	XAC0685	99%	-	XOO3934	98%	-	XCC3434	91%
	two-component system sensor histidine	-	XCV2215	100%	-	XAC2054	96%	-	XOO2323	90%	-	XCC2153	86%
	two-component system sensor histidine	-	XCV3165	100%	-	XAC3028	95%	-	-	-	-	-	-
	two-component system sensor histidine	-	XCV3166	100%	-	XAC3029	97%	-	-	-	-	-	-
	two-component system sensor histidine	-	XCV3167	100%	-	-	-	-	XOO1828	89%	-	XCC2847	81%
	two-component system sensor histidine	-	XCV3168	100%	-	XAC3031	96%	-	XOO1827	91%	-	-	-
	two-component system sensor histidine	-	XCV3793	100%	-	XAC3673	97%	-	-	-	-	-	-
	two-component system sensor histidine	-	XCV3804	100%	-	XAC3683	99%	-	XOO0698	97%	-	XCC3643	90%
	two-component system sensor histidine	-	XCV2671	100%	-	XAC2492	97%	-	XOO2797	92%	-	XCC2360	84%
	two-component system sensor kinase	-	XCV2105	100%	<i>nodV</i>	XAC2149	97%	<i>nodV</i>	XOO2378	96%	-	-	-
	two-component system sensor kinase	-	XCV4087	100%	-	XAC3994	93%	-	-	-	-	XCC3910	85%
	two-component system sensor kinase KdpD	-	XCV0811	100%	<i>kdpD</i>	XAC0759	98%	<i>kdpD</i>	XOO3843	96%	<i>kdpD</i>	XCC0705	93%
	two-component system sensor protein	-	XCV0114	100%	-	XAC0135	96%	-	-	-	-	-	-
	two-component system sensor protein	-	XCV0191	100%	<i>ntrB</i>	XAC0207	98%	<i>ntrB</i>	XOO4484	96%	<i>ntrB</i>	XCC0188	96%
	two-component system sensor protein	-	XCV0335	100%	<i>smeS</i>	XAC0326	98%	-	-	-	-	-	-
	two-component system sensor protein	-	XCV0528	100%	-	XAC0494	97%	-	XOO0519	93%	-	XCC0483	87%
	two-component system sensor protein	-	XCV0677	100%	<i>algZ</i>	XAC0620	98%	<i>algZ</i>	XOO4009	95%	<i>algZ</i>	XCC3513	92%
	two-component system sensor protein	-	XCV0744	100%	-	XAC0683	95%	-	XOO3936	94%	-	XCC3436	86%
	two-component system sensor protein	-	XCV0888	100%	<i>colS</i>	XAC0835	99%	<i>colS</i>	XOO3762	93%	<i>colS</i>	XCC0780	87%
two-component system sensor protein	-	XCV0934	100%	<i>torS</i>	XAC0897	98%	<i>torS</i>	XOO3709	88%	<i>torS</i>	XCC0825	94%	

Sensors of the two component regulatory

two-component system sensor protein	-	XCV1071	100%	<i>phoR</i>	XAC1041	99%	<i>phoR</i>	XOO3667	97%	<i>phoR</i>	XCC0962	96%
two-component system sensor protein	-	XCV1254	100%	<i>colS</i>	XAC1222	97%	<i>colS</i>	XOO3534	95%	<i>colS</i>	XCC1121	88%
two-component system sensor protein	-	XCV1260	100%	-	XAC1228	99%	-	XOO3528	99%	-	XCC1127	94%
two-component system sensor protein	-	XCV1828	100%	<i>regS</i>	XAC1798	99%	<i>regS</i>	XOO2228	97%	<i>regS</i>	XCC1780	93%
two-component system sensor protein	-	XCV2111	100%	<i>lytS</i>	XAC2142	93%	-	-	-	-	-	-
two-component system sensor protein	-	XCV2155	100%	-	-	-	-	-	-	-	-	-
two-component system sensor protein	-	XCV2310	100%	-	-	-	-	-	-	-	-	-
two-component system sensor protein	-	XCV3014	100%	<i>creC</i>	XAC2854	95%	-	-	-	<i>creC</i>	XCC2694	86%
two-component system sensor protein	-	XCV3267	100%	<i>exsG</i>	XAC3136	97%	-	-	-	<i>exsG</i>	XCC2959	77%
two-component system sensor protein	-	XCV3364	100%	<i>colS</i>	XAC3249	98%	<i>colS</i>	XOO1207	78%	<i>colS</i>	XCC3106	96%
two-component system sensor protein	-	XCV3610	100%	<i>tctE</i>	XAC3482	97%	<i>tctE</i>	XOO1106	95%	<i>tctE</i>	XCC3351	93%
two-component system sensor protein	-	XCV4115	100%	<i>phoQ</i>	XAC4022	99%	<i>phoQ</i>	XOO0424	98%	<i>phoQ</i>	XCC3942	94%
two-component system sensor/regulator	-	XCV4292	100%	-	XAC4193	99%	-	XOO0336	97%	-	XCC4076	95%
two-component system sensor/response	-	XCV1332	100%	-	XAC1283	97%	-	-	-	-	XCC1185	86%
VirA-like sensor histidine kinase	-	XCV2145	100%	-	-	-	<i>virA</i>	XOO3032	91%	-	-	-
putative sensor protein	-	XCV0667	100%	-	XAC0610	80%	-	XOO4021	90%	-	XCC3523	78%
putative sensor protein	-	XCV1610	100%	-	XAC1570	94%	-	-	-	-	XCC1520	84%
putative sensor protein	-	XCV1982	100%	-	XAC1938	98%	-	XOO2616	87%	-	XCC1911	83%
putative sensor protein	-	XCV1985	100%	-	XAC1940	96%	-	-	-	-	-	-
putative sensor protein	-	XCV2579	100%	-	XAC2382	99%	-	XOO2708	96%	-	XCC2274	95%
putative sensor protein	-	XCV2623	100%	-	XAC2446	92%	-	-	-	-	-	-
putative sensor protein	-	XCV2971	100%	-	XAC2810	97%	-	-	-	-	-	-
putative sensor protein	-	XCV3588	100%	-	XAC3460	96%	-	XOO0930	86%	-	XCC3333	84%
putative tryptophan-rich sensory protein	-	XCV1865	100%	<i>tspO</i>	XAC1819	87%	-	-	-	<i>tspO</i>	XCC1799	80%
putative two-component system sensor	-	XCV2038	100%	-	XAC1991	99%	-	XOO2564	97%	-	XCC1957	95%
putative two-component system sensor	-	XCV2965	100%	<i>baeS</i>	XAC2804	98%	<i>baeS</i>	XOO1558	93%	-	-	-
putative two-component system sensor	-	XCV3453	100%	-	XAC3335	96%	-	XOO3405	93%	-	XCC3178	84%
putative two-component system sensor	-	XCV4386	100%	-	XAC4283	89%	-	XOO0240	83%	-	XCC4144	78%
putative two-component system sensor	-	XCV1324	100%	<i>cvgSY</i>	XAC1274	97%	-	-	-	<i>cvgSY</i>	XCC1176	88%
putative two-component system sensor	-	XCV3840	100%	-	XAC3720	98%	-	XOO0650	97%	-	XCC3677	91%
putative two-component system sensor	-	XCV1323	100%	-	XAC1273	96%	-	-	-	-	XCC1175	88%
putative two-component system sensor	-	XCV1333	100%	-	XAC1282	95%	-	-	-	-	XCC1186	81%
putative two-component system sensor	-	XCV4069	100%	-	XAC3975	94%	-	-	-	-	-	-
putative two-component system sensor-	-	XCV2039	100%	-	XAC1992	99%	<i>pdeA</i>	XOO2563	98%	<i>pdeA</i>	XCC1958	99%
two-component system sensor protein	-	XCV0783	97%	-	XAC0729	100%	-	-	-	-	-	-

sensor histidine kinase	-	XCV1079	92%	-	XAC1075	100%	-	-	-	-	-	-
sensor histidine kinase	-	XCV1079	90%	-	XAC1074	100%	-	-	-	-	-	-
two-component system sensor protein	-	-	-	-	XAC2192	100%	-	-	-	-	-	-
"hypothetical protein"	-	XCV2041	99%	-	XAC1993	100%	-	XOO2561	96%	-	XCC1959	94%
"hypothetical protein"	-	XCV2973	97%	-	XAC2812	100%	-	XOO1551	92%	-	XCC2641	84%
"hypothetical protein"	-	XCV4470	97%	-	XAC4358	100%	-	-	-	-	XCC4224	85%
sensor histidine kinase	-	XCV1546	96%	-	XAC1488	100%	-	XOO2330	88%	-	XCC1445	79%
sensor histidine kinase	-	XCV1398	97%	-	XAC1345	100%	-	XOO1879	93%	-	XCC1294	83%
sensor histidine kinase		-	-		-	-		-	-		XCC2030	100%
sensor histidine kinase		-	-		-	-		-	-		XCC0108	100%
sensor histidine kinase		-	-		-	-		-	-		XCC3594	100%
two-component system sensor protein		-	-		XAC2167	82%		-	-		XCC2179	100%
sensor histidine kinase	-	XCV2754	93%	<i>fixL</i>	XAC2555	93%	<i>fixL</i>	XOO3147	100%	<i>fixL</i>	XCC2421	87%
two-component system sensor protein	-	XCV3014	80%	<i>creC</i>	XAC2854	80%	<i>creC</i>	XOO1477	100%	-	-	-
two-component system sensor protein	-	XCV0783	96%	-	XAC0729	95%	-	XOO3875	100%	-	XCC3395	91%
two-component system sensor protein	-	XCV0744	94%	-	XAC0683	91%	-	XOO3936	100%	-	XCC3436	84%
two-component system sensor protein	-	XCV4069	84%	<i>ygiY</i>	XAC3975	82%	-	XOO0057	100%	-	XCC3894	81%

<i>citri</i>	5284		<i>Citrus</i> sp.	Malaysia	1999
<i>citri</i>	2900		<i>Citrus</i> sp.	Japan	NA
<i>citri</i>		306		NA	NA
<i>citri</i>		EDTa-BP12916, JK2-20*	<i>C. aurantifolia</i>	Saudi Arabia	1988
<i>citri</i>		EDTa-BP12923, JS582*	<i>C. aurantifolia</i>	Iran	1997
<i>citri</i>		EDTa-BP12915, JJ60-1*	<i>C. aurantifolia</i>	India	1988
<i>citri</i>		EDTa-BP12920, JF90-8*	<i>C. aurantifolia</i>	Oman	1988
<i>citri</i>		EDTa-BP12914, LB302*	<i>C. aurantifolia</i> / <i>C. macrophylla</i>	USA	2002
<i>citri</i>	3530		<i>C. limon</i>	Uruguay	1984
<i>citrumelo</i>	3371	LMG 9325, ICPB 10483	NA	NA	1989
<i>citrumelo</i>	3841	LMG 9160	<i>Poncirus trifoliata</i> x <i>C. sinensis</i>	USA	NA
<i>citrumelo</i>	3842	LMG 9167	<i>P. trifoliata</i> x <i>C. paradisi</i>	USA	NA
<i>citrumelo</i>	3843	LMG 9172	<i>C. paradisi</i>	USA	NA
<i>citrumelo</i>	3114		<i>Citrumelo</i> cv. <i>Swingle</i>	USA	1984
<i>dieffenbachiae</i> GL.A	3132	NCPBP 985, ATCC 23379, LMG 7399	<i>Dieffenbachia</i> sp.	USA	1950
<i>dieffenbachiae</i> GL.C	3133	NCPBP 1833, ICMP 5727, LMG 695	<i>Anthurium</i> sp.	Brazil	1965
<i>dieffenbachiae</i> GL.C	5688		<i>A. andreaeanum</i>	Venezuela	NA
<i>dieffenbachiae</i> GL.C	5691		<i>Anthurium</i> sp.	Mauritius	NA
<i>glycines</i>	7119	NCPBP 3658	<i>Glycine max</i>	Brazil	1981
<i>glycines</i>	1519	NCPBP 1717	<i>G. hispida</i>	Zimbabwe	1962
<i>glycines</i>	2526	NCPBP 554, ATCC 43911, ICMP 5732, LMG 712	<i>G. hispida</i>	Sudan	1956
<i>malvacearum</i>	2035		<i>Gossypium hirsutum</i>	Argentina	1981
<i>malvacearum</i>	2530	NCPBP 633, ICMP 5739, LMG 761	<i>G. hirsutum</i>	Sudan	1958
<i>malvacearum</i>	5700		<i>G. hirsutum</i>	Senegal	1990
<i>malvacearum</i>	5701		<i>G. hirsutum</i>	Madagascar	1990
<i>malvacearum</i>	5726		<i>G. barbadense</i>	Sudan	1991
<i>mangiferaeindicae</i>	1716	ATCC 11637, ICMP 5740, LMG 941, NCPBP 490	<i>M. indica</i>	India	1957
<i>mangiferaeindicae</i>	2933		<i>M. indica</i>	Réunion Island	1981
<i>mangiferaeindicae</i>	7236	JN576	<i>M. indica</i>	Japan	1993
<i>mangiferaeindicae</i>	2915	NCPBP 2438	<i>M. indica</i>	South Africa	1971
<i>mangiferaeindicae</i>	2935		<i>M. indica</i>	Australia	1978
<i>mangiferaeindicae</i>	2939		<i>Schinus terebenthifolius</i>	Réunion Island	1987
<i>mangiferaeindicae</i>	2940		<i>S. terebenthifolius</i>	Réunion Island	1987
<i>mangiferaeindicae</i>	7238	JP742	<i>S. terebenthifolius</i>	Réunion Island	1994
<i>mangiferaeindicae</i>	7239	JP757	<i>S. terebenthifolius</i>	Réunion Island	1994
<i>mangiferaeindicae</i>	7237	JP740	<i>S. terebenthifolius</i>	Réunion Island	1994
<i>manihotis</i>	1860		<i>Manihot esculenta</i>	Nigeria	1978
<i>manihotis</i>	2603	NCPBP 2443	<i>M. esculenta</i>	Colombia	1972
<i>manihotis</i>	1851	CIAT111	<i>M. esculenta</i>	USA	NA
<i>manihotis</i>	2624		<i>M. esculenta</i>	Réunion Island	1986
<i>manihotis</i>	6544		<i>M. esculenta</i>	Brazil	1992
<i>phaseoli</i> GL fuscans	1815		<i>Phaseolus</i> sp.	Greece	1978
<i>phaseoli</i> GL fuscans	4834		<i>Phaseolus vulgaris</i> cv. Michelet	France	1998
<i>phaseoli</i> GL fuscans	6165	LMG 826, ATCC 19315, ICMP 239, NCPBP 381	<i>P. vulgaris</i>	Canada	1957
<i>phaseoli</i> GL fuscans	6167	LMG 7511, ICMP 242	<i>Phaseolus</i> sp.	USA	1964
<i>phaseoli</i> GL fuscans	6969	SNES C44529	<i>P. vulgaris</i>	Tanzania	2001

<i>phaseoli</i> GL <i>fuscans</i>	6166	LMG 837, NCPPB 1654	<i>P. vulgaris</i>	South Africa	1963
<i>phaseoli</i> GL <i>fuscans</i>	6965	SNES C44525	<i>P. vulgaris</i>	NA	NA
<i>phaseoli</i> GL <i>fuscans</i>	6975	JR51	<i>Phaseolus</i> sp.	France	1994
<i>phaseoli</i> GL <i>fuscans</i>	6976	JR52	<i>Phaseolus</i> sp.	Czech Rep.	1994
<i>phaseoli</i> GL <i>fuscans</i>	6979	SNES C44537	<i>P. vulgaris</i>	Tanzania	2001
<i>phaseoli</i> GL <i>fuscans</i>	1845		<i>Phaseolus</i> sp.	Greece	1978
<i>phaseoli</i> GL <i>fuscans</i>	6960	JW355.2	<i>P. vulgaris</i>	Réunion Island	2000
<i>phaseoli</i> GL <i>fuscans</i>	6970	JR7	<i>Phaseolus</i> sp.	USA	1990
<i>phaseoli</i> GL <i>fuscans</i>	6971	JR13	<i>Phaseolus</i> sp.	Tanzania	1992
<i>phaseoli</i> GL1	6546	ATCC 9563, NCPPB 3035, ICMP 5834, CFBP 2534	<i>P. vulgaris</i>	USA	NA
<i>phaseoli</i> GL1	6164	LMG 8014, NCPPB 1811	<i>P. vulgaris</i>	Romania	1966
<i>phaseoli</i> GL1	6984	JW355.1	<i>P. vulgaris</i>	Réunion Island	2000
<i>phaseoli</i> GL1	6982	JW351.5	<i>P. vulgaris</i>	Réunion Island	2000
<i>phaseoli</i> GL1	412		<i>P. vulgaris</i>	USA	NA
<i>phaseoli</i> GL1	6983	JW353.1	<i>P. vulgaris</i>	Réunion Island	2000
<i>phaseoli</i> GL1	6985	JW355.3	<i>P. vulgaris</i>	Réunion Island	2000
<i>phaseoli</i> GL2	6989	JW162.12	<i>P. vulgaris</i> cv. Marla	Réunion Island	2000
<i>phaseoli</i> GL2	6990	JW351.3	<i>P. vulgaris</i> cv. Marla	Réunion Island	2000
<i>phaseoli</i> GL2	6991	JW355.5	<i>P. vulgaris</i> cv. Marla	Réunion Island	2000
<i>phaseoli</i> GL2	6988	JW162.11	<i>P. vulgaris</i> cv. Marla	Réunion Island	2000
<i>phaseoli</i> GL3	6992	JW214.3	<i>P. vulgaris</i> cv. Marla	Réunion Island	2000
<i>phaseoli</i> GL3	6994	JR3	<i>P. vulgaris</i>	Tanzania	1990
<i>phaseoli</i> GL3	6996	JW162.7	<i>P. vulgaris</i> cv. Marla	Réunion Island	2000
<i>phaseoli</i> GL3	6993	JW162.1	<i>P. vulgaris</i> cv. Marla	Réunion Island	2000
<i>ricini</i>	5863	IBSBF 313	<i>Ricinus communis</i>	Brazil	1981
<i>ricini</i>	5864	IBSBF 1191	<i>R. communis</i>	Brazil	1995
<i>ricini</i>	5865	IBSBF 1192	<i>R. communis</i>	Brazil	1995
<i>ricini</i>	6541		<i>R. communis</i>	Brazil	1981
<i>ricini</i>	6542		<i>R. communis</i>	Brazil	1985
<i>spondiae</i>	2547	ICMP 17032	<i>Spondias dulcis</i>	Mauritius	1985
<i>vasculorum</i>	1289		<i>Saccharum officinarum</i>	Réunion Island	1970
<i>vasculorum</i>	5823	LMG 901, ATCC 35938, ICMP 5757, NCPPB 796	<i>S. officinarum</i>	Mauritius	1979
<i>vasculorum</i>	5696		<i>Thysanolaena maxima</i>	Réunion Island	NA
<i>vasculorum</i>	5822	LMG 895, ICMP 304, NCPPB 181	<i>S. officinarum</i>	Australia	1946
<i>vesicatoria</i>	1604		<i>Capsicum annuum</i>	Guadeloupe	NA
<i>vesicatoria</i>	5594		<i>Lycopersicon esculentum</i>	Guadeloupe	1993
<i>vesicatoria</i>		75-3, XV153	<i>L. esculentum</i>	NA	NA
<i>vesicatoria</i>	6864		<i>C. frutescens</i>	USA	1947
<i>vesicatoria</i>	2484		<i>L. esculentum</i>	Guadeloupe.	1980
<i>vesicatoria</i>	5618	XCV85-10	<i>C. annuum</i>	USA	NA
<i>vignicola</i>	7112	LMG 8752, NCPPB 1838	<i>Vigna unguiculata</i> cv. China Red	USA	1942
<i>vignicola</i>	7113	LMG 840, NCPPB 2061	<i>V. unguiculata</i>	Sudan	1966
<i>vignicola</i>	7110	LMG 831, NCPPB 638	<i>V. unguiculata</i>	Zimbabwe	NA
<i>vignicola</i>	7111	NCPPB 1633	<i>V. sinensis</i>	USA	1942
<i>vignicola</i>	7115	NCPPB 3187	<i>V. sinensis</i>	Brazil	1978
<i>betae</i>	5852	IBSBF 857, ICMP 8917, LMG 9040, NCPPB 2592	<i>Beta vulgaris</i>	Brazil	1973
<i>bilvae</i>	3136	NCPPB 3213, ICMP 8918	<i>Aegle marmelos</i>	India	1980
X. campestris					
non pathogenic strain	5824	LMG 7383, NCPPB 1930	<i>Armoracia</i> sp.	USA	1939
non pathogenic strain	5825	LMG 547, NCPPB 983	<i>Barbarea vulgaris</i>	USA	1939
non pathogenic strain	5826	LMG 7385	<i>B. vulgaris</i>	USA	1939

X. campestris pv.					
campestris	1119		<i>Brassica oleracea</i> var botrytis	France	1967
campestris	1869		<i>B. oleracea</i>	Ivory Coast	1976
campestris		EDTa-BP12824, 305 ^a	<i>B. oleracea</i>	Netherlands	NA
campestris	5241	LMG 568, ATCC33913, DSM 3586, ICMP 13, NCPBP 528	<i>B. oleracea</i> var. gemmifera	United Kingdom	1957
campestris	4956		<i>B. oleracea</i> var botrytis	Belgium	1999
campestris	5817		<i>B. oleracea</i> var botrytis	Chili	2001
campestris	1712		<i>B. oleracea</i> var. capitata	France	1975
campestris	1713		<i>B. oleracea</i> var botrytis	France	1975
campestris	6865	NCPBP 2986	<i>B. oleracea</i> var. capitata	Australia	1975
campestris	4954		<i>B. oleracea</i> var botrytis	Belgium	1999
campestris		EDTa-BP12825, HR16181F	<i>B. rapa</i>	Portugal	1996
campestris	4953		<i>B. oleracea</i> var botrytis	Belgium	1999
campestris	5130	SNES 3430	<i>B. oleracea</i>	NA	2000
campestris	1124		<i>B. oleracea</i> var botrytis	France	1967
campestris	6863	NCPBP 875	<i>B. oleracea</i> var botrytis	Germany	1958
campestris	4955		<i>B. oleracea</i> var botrytis	Belgium	1999
campestris	6650	LMG 8004, NCPBP 1145	<i>B. oleracea</i>	United Kingdom	1958
incanae	1371		<i>Matthiola incana</i>	France	1972
incanae	1438	NCPBP 1934	<i>M incana</i>	USA	1949
incanae	1606		Ornamental cruciferous plants	France	1974
incanae	2527	NCPBP 937, ATCC 13462, ICMP 574, LMG 7490	<i>M incana</i>	USA	1950
incanae	5686		<i>M. incana</i>	France	1981
raphani	5827	LMG 860, ICMP 1404, ICPB XV16, NCPBP 1946	<i>Raphanus sativus</i>	USA	1940
raphani	5828	LMG 7505, ICMP 1640	<i>R. sativus</i>	USA	NA
raphani	5829	LMG 8134, ICMP 1641	<i>R. sativus</i>	USA	NA
raphani	3838	ICMP 7, LMG 535, NCPBP 347	<i>Iberis</i> sp.	Tanzania	1954
raphani		756C	<i>B. oleracea</i> var. capitata	Est Asia	NA
X. oryzae pv.					
oryzae	7088	KACC10331	<i>Oryza sativa</i>	Korea	NA
oryzae		EDTa-BP12842, MAFF311018			
oryzicola	7109	BLS256	<i>O. sativa</i>	Phillipines	1984
X. vasicola pv.					
vasculorum	5830	LMG 900, NCPBP 795	<i>S. officinarum</i>	Malaysia	NA
vasculorum	5831	LMG 902, NCPBP 889	<i>S. officinarum</i>	South Africa	1960
vasculorum	1215		<i>S. officinarum</i>	Kenya	NA
vasculorum	5694		<i>Zea mays</i>	Réunion Island	NA
vasculorum	5695		<i>Tripsacum laxum</i>	Réunion Island	NA
musacearum	7122	NCPBP 2251	<i>Musa</i> sp.	Ethiopia	1967
musacearum	7123	NCPBP 2005, ATCC 49084, ICPB XM130, ICMP 2870, LMG 785	<i>Ensete ventricosum</i>	Ethiopia	1966
X. sp. pv.					
glycines	7118	NCPBP 1716	<i>Glycine javanica</i>	Zambia	1963
glycines	1559	NCPBP 1717	<i>G. hispida</i>	France	1974

a CFBP = Collection Française des Bactéries Phytopathogènes, INRA Angers-Nantes, France

b. Collection names: ICMP = International Collection of Micro-organisms from Plants Auckland, New Zealand
LMG = Belgian Coordinated Collections of Micro-organisms, University of Gent, Belgium
NCPBPB = National Collection of Plant Pathogenic Bacteria, York, United Kingdom
MAFF = Ministry of Agriculture, Forestry, and Fisheries of Japan, Okinawa
ICPB = International Collection of phytopathogenic Bacteria, Davis, CA, USA
IBSBF = Biological Institute Culture Collection of Phytopathogenic Bacteria, Instituto Biologico, Brazil
ATCC = American Type Culture Collection, Manassas
CIAT = Centro Internacional de Agricultura Tropical, Cali, Columbia
KACC = Korean Agricultural Culture Collection, Suwon, Republic of Korea
EDTa-BP = Collection of bacterial strains of Ecologie, Diversité et Taxonomie des Bactéries Phytopathogènes team, INRA Angers-Nantes, France
* Bacterial strains provided by O. Pruvost, Cirad, Reunion Island, France

Table A3 Results of the MK tests for positive selection on genes encoding MCPs and adhesins. Four housekeeping genes (*atpD*, *dnaK*, *efp* and *rpoD*) were used as controls. The cell entries represent the observed numbers of polymorphic synonymous sites (P_s), polymorphic nonsynonymous sites (P_n), divergent synonymous sites (D_s), and divergent nonsynonymous sites (D_n). A significant excess of nonsynonymous divergence ($D_n/D_s > P_n/P_s$) leads to rejecting the neutral mutation hypothesis and is interpreted as a signal of positive selection. Mk tests were performed (i) on pooled sequences from 6 strains (*Xav*85-10, *Xac*306 and *Xap*CFBP4834, *Xcc*ATCC33913, *Xcc*8004 and *Xcc*B100) belonging to two species (*X. axonopodis* and *X. campestris*) to detect adaptive divergence signals, and (ii) by assigning the fixed replacement changes to the pooled sequences from *X. axonopodis* (iia) or *X. campestris* (iib) using *X. oryzae* pv. *oryzicola* (*Xoc*) strain BLS256 as outgroup to ascertain which species has been affected by positive selection.

	P_s	P_n	D_s	D_n	p value ^a	G value ^b
<i>atpD</i>						
(i) pooled	76	9	23	2	1.00	0.15
(iia) <i>X. axonopodis</i> vs <i>Xoc</i> BLS256	64	7	19	2	1.00	0.00
(iib) <i>X. campestris</i> vs <i>Xoc</i> BLS256	18	2	44	6	1.00	0.06
<i>dnaK</i>						
(i) pooled	73	1	47	8	4.71×10^{-3} **	9.07**
(iia) <i>X. axonopodis</i> vs <i>Xoc</i> BLS256	69	1	36	3	0.13	2.65
(iib) <i>X. campestris</i> vs <i>Xoc</i> BLS256	4	0	81	11	1.00	- ^c
<i>efp</i>						
(i) pooled	14	1	17	6	0.21	2.56
(iia) <i>X. axonopodis</i> vs <i>Xoc</i> BLS256	12	1	14	2	1.00	0.18
(iib) <i>X. campestris</i> vs <i>Xoc</i> BLS256	2	0	31	6	1.00	-
<i>rpoD</i>						
(i) pooled	90	14	41	5	0.79	0.20
(iia) <i>X. axonopodis</i> vs <i>Xoc</i> BLS256	52	2	39	6	0.14	3.14
(iib) <i>X. campestris</i> vs <i>Xoc</i> BLS256	69	13	69	5	8.14×10^{-3} **	8.27**
XCV0669						
(i) pooled	107	12	153	47	2.76×10^{-3} **	9.58**
(iia) <i>X. axonopodis</i> vs <i>Xoc</i> BLS256	/ ^d	/	/	/	/	/
(iib) <i>X. campestris</i> vs <i>Xoc</i> BLS256	/	/	/	/	/	/
XCV1702						
(i) pooled	102	42	226	159	0.01*	6.71**
(iia) <i>X. axonopodis</i> vs <i>Xoc</i> BLS256	99	40	99	51	0.36	0.91
(iib) <i>X. campestris</i> vs <i>Xoc</i> BLS256	3	2	246	185	1.00	0.00
XCV1778						
(i) pooled	86	24	217	138	8.95×10^{-4} ***	11.37***
(iia) <i>X. axonopodis</i> vs <i>Xoc</i> BLS256	/	/	/	/	/	/
(iib) <i>X. campestris</i> vs <i>Xoc</i> BLS256	/	/	/	/	/	/
XCV1933						
(i) pooled	98	30	139	95	1.16×10^{-3} **	11.15
(iia) <i>X. axonopodis</i> vs <i>Xoc</i> BLS256	62	2	82	28	1.08×10^{-4} ***	17.37***
(iib) <i>X. campestris</i> vs <i>Xoc</i> BLS256	39	28	163	98	0.57	0.40
XCV1939						
(i) pooled	/	/	/	/	/	/
(iia) <i>X. axonopodis</i> vs <i>Xoc</i> BLS256	106	45	85	26	0.26	1.33
(iib) <i>X. campestris</i> vs <i>Xoc</i> BLS256	/	/	/	/	/	/
XCV1940						

(i) pooled	88	35	140	139	0.78×10^{-4} ***	16.33 ***
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	88	35	68	30	0.77	0.12
(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	0	0	170	158	-	-
XCV1941						
(i) pooled	83	24	154	96	3.36×10^{-3} **	8.94 **
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	91	29	93	37	0.47	0.59
(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	3	0	196	105	0.55	-
XCV1942						
(i) pooled	/	/	/	/	/	/
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	102	66	69	30	0.15	2.21
(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	/	/	/	/	/	/
XCV1944						
(i) pooled	/	/	/	/	/	/
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	77	63	72	95	0.04 *	4.32 *
(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	/	/	/	/	/	/
XCV1945						
(i) pooled	449	259	8	20	0.43	0.77
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	276	110	122	96	1.66×10^{-4} ***	14.77 ***
(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	224	136	151	112	0.25	1.46
XCV1947						
(i) pooled	107	25	136	89	0.57×10^{-4} ***	17.07 ***
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	100	22	117	58	5.06×10^{-3} **	8.62 **
(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	7	3	204	103	1.00	0.06
XCV1948						
(i) pooled	/	/	/	/	/	/
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	55	11	91	22	0.69	0.22
(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	/	/	/	/	/	/
XCV1951						
(i) pooled	157	63	228	140	0.02 *	5.46 *
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	132	58	136	43	0.20	1.97
(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	28	7	315	198	0.03 *	5.28 *
XCV2044						
(i) pooled	32	0	43	9	0.01 *	-
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	30	0	36	2	0.50	-
(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	2	0	67	11	1.00	-
XCV2625						
(i) pooled	59	7	134	34	0.09	3.30
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	59	7	52	21	0.01 *	7.42 **
(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	0	0	150	44	-	-
XCV3021						
(i) pooled	107	15	108	52	0.69×10^{-4} ***	16.49 ***
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	80	18	103	15	0.26	1.32
(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	33	1	158	61	5.62×10^{-4} ***	13.63 ***
XCV3230						
(i) pooled	75	24	102	2	3.93×10^{-4} ***	14.13 ***
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	54	0	65	29	0.23×10^{-4} ***	19.43 ***

(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	23	0	127	43	2.77×10^{-3} **	-
XCV3338						
(i) pooled	75	22	118	62	0.06	4.25 *
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	66	20	86	37	0.35	1.20
(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	10	3	165	84	0.56	0.67
XCV3577						
(i) pooled	82	14	150	38	0.27	1.39
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	/	/	/	/	/	/
(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	/	/	/	/	/	/
<i>xadA1</i>						
(i) pooled	156	26	252	107	0.55×10^{-4} ***	16.76 ***
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	141	22	156	105	0.00×10^{-4} ***	36.87 ***
(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	15	4	315	184	0.22	2.15
<i>fhaB</i>						
(i) pooled	536	263	1245	900	0.08×10^{-4} ***	20.22 ***
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	499	238	1443	3191	0.00×10^{-4} ***	353.32 ***
(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	0	0	1577	3246	-	-
<i>pilS</i>						
(i) pooled	64	9	122	53	2.334×10^{-3} **	9.76 **
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	51	8	105	28	0.24	1.58
(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	13	1	155	53	0.20	3.02
<i>pilL</i>						
(i) pooled	232	61	406	246	0.00×10^{-4} ***	27.65 ***
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	156	45	264	126	0.01 *	6.52 *
(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	79	17	485	277	2.27×10^{-4} ***	14.51 ***
<i>pilU</i>						
(i) pooled	37	2	60	3	1.00	0.01
(iia) <i>X. axonopodis</i> vs <i>XocBLS256</i>	36	1	41	6	0.13	3.09
(iib) <i>X. campestris</i> vs <i>XocBLS256</i>	1	1	71	8	0.21	1.94

^a: The significance of MK results was established by the Fisher's exact test (*: $0.01 < p < 0.05$; **: $0.001 < p < 0.01$; ***: $p < 0.001$).

^b: Value of the log-likelihood ratio test and significance of the associated probability (*: $0.01 < p < 0.05$; **: $0.001 < p < 0.01$; ***: $p < 0.001$).

^c: cannot be calculated due to an absence of polymorphism.

^d: unavailable data due to the absence of the gene in the sequences of *X. campestris*.

Table A4 Sequences of pairs of primers used to amplify candidate genes. The annealing temperature of the PCRs and the lengths of the amplified fragments are also indicated.

First pairs of primers			Second pairs of primers				
Primers' names	Reverse and Forward Primers	T°C	PCR fragment size (bp)	Primers' names	Reverse and Forward Primers	T°C	PCR fragment size (bp)
XCV1940-Rev1	GCCAGATCAGCTTCCGCATGGATG	60	417	NB. Ubiquitous gene	no second pair of primers was designed		
XCV1940-Fwd1	AGGTCGCCGGCTGCCATCGACT	60					
XCV3230-Rev1	TCGGCAGCGTCAGCACCAGCTTCTGG	60	871	NB. Ubiquitous gene	no second pair of primers was designed		
XCV3230-Fwd1	GCGCGTCTCGCGCACCGAACCGAA	60					
XCV3021-Rev1	GCGGGCTGCCAAGGTGCAGAG	65	964	NB. Ubiquitous gene	no second pair of primers was designed		
XCV3021-Fwd1	CATGGTGGTGTGACGTAGCGCTGCAT	65					
XCV2044-Rev1	AACGAACTGCGTACGTTCTGTCAGCAAC	60	738	NB. Ubiquitous gene	no second pair of primers was designed		
XCV2044-Fwd1	GCATGTCCTCGAACTGCAGGGCA	60					
XCV3338-Rev1	GCTGGACGATCTGGACACCATGATCC	60	338	XCV3338-Rev2	ACACCGATGATCTCGCCGATCTTGCGCGAGGC	65	839
XCV3338-Fwd1	ACCTGCGCGGTGACGTCCATCCAC	60		XCV3338-Fwd2	ATGCTCAAAGTGTGATCGAGCCGGATGTGCAGAGCTT	65	
XCV2625-Rev1	CTGGCGCAATTCGATCGCGCTGGCA	60	751	XCV2625-Rev2	AGCCGGAAGAACGCCATGCTCTGCTGCAG	65	1285
XCV2625-Fwd1	GCGATGTCGTCGATGATGCCGATCTTC	60		XCV2625-Fwd2	TGGCCAGCAGCGCAGAAGAGCGCCAGTC	65	
XCV1933-Rev1	GCTGGCGTTCGGCATCGTGTCTGAT	65	1142	XCV1933-Rev2	ACGCTGGCGACGATTTTCGCCACGGTG	65	2009
XCV1933-Fwd1	ACTTCGATGGTGGTGCCTCGGCCATTC	65		XCV1933-Fwd2	ATGCAATGGATCAACAATCTGAAACTGATGCC	65	
XCV1778-Rev1	GTGATGGCGGTTCATCGAGTTTTCGC	65	1020	XCV1778-Rev2	CAAGGCTACGCACTTCGGTCGCCACCAC	65	1483
XCV1778-Fwd1	TCACCTCGATGATCTGACCGATGCGGT	65		XCV1778-Fwd2	TTCGATCTGGACGGTGCATTCTGGATGC	65	
XCV1702-Rev1	GTCAACAGCGTCAAGCAACGCTACG	65	1352	XCV1702-Rev2	TTGGTCTGGAAGGCGATGCCGTCGATGAC	65	1296
XCV1702-Fwd1	TTGACGGTGGAGGTCAGCTTTCATC	65		XCV1702-Fwd2	CGAGCAGCGCACCTGCCGCTGATC	65	
XCV0669-Rev1	ATTACTTCTTCATCGTCGACCCGCAGATGC	60	1623	XCV0669-Rev2	GTCATTTCTGCCAGTGCATCACCGTGGTGCT	65	1586
XCV0669-Fwd1	TCTCGGCCATGATGTCGGTCACGCGCTT	60		XCV0669-Fwd2	ATCAAGACCGCCATCGATAGCGGCGA	65	
XCV1952-Rev1	CGCTTCTCCGCGCAGACCAAGGTGAT	65	857	XCV1952-Rev2	GAAGAGACGATCTCGGCCATGGTCTTGCC	65	1334
XCV1952-Fwd1	GTCGATGACGCTGATGATCTCGGCGATCTT	65		XCV1952-Fwd2	GACACCAATCGCAGGCGCTGGTTTTAC	65	
XCV1941-Rev1	TTCCGCCATGAGCTGGACAAGGCGA	65	1538	No other sets of primers were specific			
XCV1941-Fwd1	CACCAGCTGGTTGCCGTGTTTCGAC	65					
XCV1938-Rev1	AGCACCACAAACTGGGCAAGGAC	65	400	NB. Ubiquitous gene	no second pair of primers was designed		
XCV1938-Fwd1	GGGCGGGTGACCATCTTGCG	65					
XCV1951-Rev1	GAGTGGGTCGGCATGGGCACCTTGTG	65	2022	XCV1951-Rev2	GCCGACCACCGCATCGGCACTGAAGC	65	1213
XCV1951-Fwd1	TGCTCACCGAATCGTCGATCAGGTGCTTGA	65		XCV1951-Fwd2	CTACCAGGCCGAGCAGATCGGCAGCGA	65	
XCV1954-Rev1	TCGCGATCATCGTCAAGTCAATAACCG	60	400	XCV1954-Rev2	ATGAAATTCCTGCATTCCATGACCGTCGGGC	65	1337
XCV1954-Fwd1	ATTCTGCTGGTTCTGCAATTCACCG	60		XCV1954-Fwd2	CGCCCCACGATGGTGGAGAGCTGT	65	
XCV1947-Fwd1	GGACGAGGTTTCGATGCCGGACATGGTGGTGACCCT	65	1353	XCV1947-Fwd2	TCGATGATGGAAGTCAGCGACCGCAACC	65	434
XCV1947-Rev1	CTTGCGCATGGATGAAATCCAGGCCAAGTACGCC	65		XCV1947-Rev2	ACGAGGTTTCGATGCCGGACATGGTGG	65	
XCV1942-Rev1	GGATCACCCGACGCTGACCCAGCCCTTG	65	1217	XCV1942-Rev2	TCCACCAGCGCGGCTTCTGCTGGG	65	1976
XCV1942-Fwd1	CGCTGTGACCAGCAGCGAGCCT	65		XCV1942-Fwd2	CGCCTTCCGCTGCGTTCGGCTCCTG	65	
XCV1939-Rev1	ATGACCAAGATCCGCCTCTCCAACGACAT	65	774	XCV1939-Rev2	ATCTGCGTGACGGTGAGGTGACCTGTTTCGAT	65	1976
XCV1939-Fwd1	TGTGGCTGTCGACCAGCGCATTGGTGT	65		XCV1939-Fwd2	ATGACATCTCTTTACTGCGTTCAATGTAGGGCC	65	
XCV1944-Rev1	CCGACAACCCGGTGCAGCAGGAACG	60	524	XCV1944-Rev2	CTGCCAGCTGGTCTCGACGGTCGTCGAT	65	2061
XCV1944-Fwd1	CCTGATCGTGCCGGCGCTTCATTTCCT	60		XCV1944-Fwd2	ATGTCCATTGCCACGGTCAACATCGAGAC	65	
XCV3261-Rev1	CCTGATCACCATGTCCAATGCGCGTTATGA	65	507	No other sets of primers were specific			
XCV3261-Fwd1	AGCGATGCCCGCTGGCCAGGAT	65					
XCV1955-Rev1	ACCGAGTTGTACGACGAGGTCAAACCTGAAACTG	65	1190	XCV1955-Rev2	TGGTGTGGTCTGCTGCTGGCTGCATTT	65	1613
XCV1955-Fwd1	GCCGAGTTGGTTGGCCTGGCGCGCTT	65		XCV1955-Fwd2	ACCGGCCGATACAGGGGCGCGCTT	65	
XAC3768-Rev1	TGATCGCCACCTGCTGAGAGCC	65	2134	XAC3768-Rev2	TGATCGCCACCTGCTGAGAGCC	65	573
XAC3768-Fwd1	TCAGGCGACGTGTTTTCCAGGTGCTG	65		XAC3768-Fwd2	TGGCGGTGACCGAGCTGTCCACCG	65	
XAC3271-Fwd1	CCGAGGCGGGCATGCGCCAAGCTGGTGGAGATAG	65	454	XAC3271-Fwd2	TCAGAGCCGCGGCCAGCGAAAC	65	461

XAC3271-Rev1	GAGCCGCGGCCAGCGAACCCCTGGTGCGCCACTGCGT	65		XAC3271-Rev2	GTGACCGAGGCGGGGCATGG	65	
XCC0324-Rev1	ATGCACTTCGGCGTCTTCGTGATCCTG	65	1770	XCC0324-Rev2	ATGCACTTCGGCGTCTTCGTGATCCTGG	65	1320
XCC0324-Fwd1	TGGGCGTGTGCTTCCACAAAAAGG	65		XCC0324-Fwd2	GGCCAACCCATCGATGACGGCGATGATG	65	
XCC0276-Rev1	GTGGCGCAGGCATCGGCGAGCGAA	65	1890	XCC0276-Rev2	GTGGCGCAGGCATCGGCGAGCGAA	65	645
XCC0276-Fwd1	TCGCACGTGCCGCGGCCGAGGCTT	65		XCC0276-Fwd2	GGAAACATCGGCGAGCGTGCTGACCAC	65	
XCV2152-Fwd1	TGACACCCTTATGGATGCGCTTCATCCTG	65	1996		No other sets of primers were specific		
XCV2152-Rev1	CAATTGCAGCAACGCTCCGGGCAAAAAC	65			No other sets of primers were specific		
XCV2155-Fwd1	AACTCGACTGGCGCTGTACCACCAAG	65	1220		No other sets of primers were specific		
XCV2155-Rev1	GCAACGCCTTGGCGCCGTAATGCAGT	65			No other sets of primers were specific		
XCC2030-Fwd1	TAAAGGTAAAGGTGGTGCCGGCCAACA	65	926		No other sets of primers were specific		
XCC2030-Rev1	CGGTCATCTTCAGCCATGCCAGCCA	65			No other sets of primers were specific		
XCC3594-Fwd1	CGGCGCAATTACGTGACTGGGAACA	65	633		No other sets of primers were specific		
XCC3594-Rev1	AACGCGGCATTGAACACGCATTGCTG	65			No other sets of primers were specific		
XCV2310-Fwd1	ATTCCAATCTCGACCGAACGCTCGACG	65	1203		No other sets of primers were specific		
XCV2310-Rev1	CCTCCATCCCTTGGCCTCAATTCAAGACT	65			No other sets of primers were specific		
XCC0108-Fwd1	CAACTCTCCTGCGTGATCTGCTCAAA	58	399		No other sets of primers were specific		
XCC0108-Rev1	GCCACCATCAGGAAGGGACG	58			No other sets of primers were specific		
XAC2192-Fwd1	AACGGCATAATCGCGCTACAACGAGGC	65	380		No other sets of primers were specific		
XAC2192-Rev1	TGTTGAACCGGTTGAACTCGGCCACC	65			No other sets of primers were specific		
XCC0397-Fwd1	CGTCGATCTGCTGCAGATGCTCAAGGT	65	1579		No other sets of primers were specific		
XCC0397-Rev1	AAATGTTGTTTCGACGTTTCAGGCTTGAGATCGGT	65			No other sets of primers were specific		
XCC1719-Fwd1	CATCGTTTCGGTGCTGCCGAGCAA	65	1045		No other sets of primers were specific		
XCC1719-Rev1	GGTACGACTCGTTGCGATGCACGGTTT	65			No other sets of primers were specific		
XCC1340-Fwd1	TCGGTCACGTTTCAGCGCTGCGAGATT	65	1438		No other sets of primers were specific		
XCC1340-Rev1	ATCAACGCGATCAATGTCACCCGCAATAGC	65			No other sets of primers were specific		
XCC3595-Fwd1	CTGCAGACCGGCGGCTACGTATTG	65	800		No other sets of primers were specific		
XCC3595-Rev1	GGTGTAACTCGGCAGGCCGCAAAAAAAT	65			No other sets of primers were specific		
XCC3635-Fwd1	ATACATGTTCTGCACCTACCGGTCAGGTT	65	1178		No other sets of primers were specific		
XCC3635-Rev1	AGTGCGCTCAAAGCTGGGCGCCAAAATA	65			No other sets of primers were specific		
XCC4052-Fwd1	TAATAGGACCGGCCGTTGTCGAGCTGC	65	1307		No other sets of primers were specific		
XCC4052-Rev1	CAAGCTGGATTACAGCCAGAACGCCG	65			No other sets of primers were specific		
XCC4237-Fwd1	AAGGTCGGTAGCGGCTCGCAGGTATT	65	1163		No other sets of primers were specific		
XCC4237-Rev1	CACCCATTCGCCATCCACGCGGAA	65			No other sets of primers were specific		
XCC0304-Fwd1	TCAGAAGTGGTAGGTACCACCAGCC	65	452		No other sets of primers were specific		
XCC0304-Rev1	CGCCTGGATGCCAATGTCACCTACAAC	65			No other sets of primers were specific		
XCC3518-Fwd1	CTCCCCTTCGTTGCCGAGTGTGCCAT	65	1300		No other sets of primers were specific		
XCC3518-Rev1	GGCCACATCGCCATTCCAGTCGAAG	65			No other sets of primers were specific		
XCC4162-Fwd1	CGCAGCAGCAGCGCCAGTTGGTA	65	912		No other sets of primers were specific		
XCC4162-Rev1	ACACCGGCCGTTGCGCCATCA	65			No other sets of primers were specific		
XCC0305-Fwd1	ATGTAGGGCGCCAGCATGCTGGTGTA	65	1211		No other sets of primers were specific		
XCC0305-Rev1	CGGCATCCCGATGAAGGCTCGTT	65			No other sets of primers were specific		
XCC2046-Fwd1	ACTTCAACAACAAGCTGTACTACAACAGCTACGAA	65	784		No other sets of primers were specific		
XCC2046-Rev1	ATGCGATGCCACAGCACCCAGTG	65			No other sets of primers were specific		
XCC0394-Fwd1	AAGTTGCAACGTATTGGCAGTGCCGCATA	65	837		No other sets of primers were specific		
XCC0394-Rev1	CGTAGATGCCAACCAGCCACTTC	65			No other sets of primers were specific		
XCC0120-Fwd1	GACATCGGCAAGTTTCCCGATCTGAATCTG	60	1414		No other sets of primers were specific		
XCC0120-Rev1	GTGTCGAAGGTGTTGCTCACCGACTG	60			No other sets of primers were specific		
XCC2867-Fwd1	AACTGCGCAATGGTGTTCCGGTCTGCGTTT	65	729		No other sets of primers were specific		

XCC2867-Rev1	GACGTCCAGGTGAAGGGCTCGAGTG	65		
XAC3620-Fwd1	ATCGTAGGGCTCGCGGTGTCGGA	65	830	No other sets of primers were specific
XAC3620-Rev1	CTCCATCGCCACCAACCTGCTCGATT	65		
XAC3201-Fwd1	GCAACGGCTTGCCATCCACGTTGCAAT	65	1886	No other sets of primers were specific
XAC3201-Rev1	TTTCTGCTTCCCGATGGTGGGGGAC	65		
XAC3077-Fwd1	CGTATCGCAGGTACCGCCTATAGTGC	65	1355	No other sets of primers were specific
XAC3077-Rev1	GCCATAGGCCACCAGGTATTACTGGA	65		
XAC2193-Fwd1	GCCTGGCGCGGTCTCCCAACAA	65	790	No other sets of primers were specific
XAC2193-Rev1	GGCGCGGCGCAAGGGCGTATAGA	65		
XAC0291-Fwd1	AGCTTGCCATCGGCAATGTCGGGACG	65	1231	No other sets of primers were specific
XAC0291-Rev1	GAAGACGGCTATTACGTGCGCAGATACATC	65		
XAC4062-Fwd1	CCGCCTGCGTTCGCTGGAGCAATTCA	65	1442	No other sets of primers were specific
XAC4062-Rev1	CCAAAATCTGCCCGCATTGACGATGAAGATATT	65		
XAC2185-Fwd1	CAACGCAAGGCGCGACCAACGTGAAAGAA	60	1417	No other sets of primers were specific
XAC2185-Rev1	ATCGAAGATCGCCGCGCTGTAGCTCTG	60		
XAC3498-Fwd1	ATTCGCATCAGCGGTTCTGAAGCCGATTGGAA	65	1089	No other sets of primers were specific
XAC3498-Rev1	TTGAGCTGCGCCGACCATGCCTGGTT	65		
XAC0852-Fwd1	CAAGTTTCTGTTCGACGGTCTGCCGATCA	65	1470	No other sets of primers were specific
XAC0852-Rev1	AAAGCCGCTGCCATGCGCGTATTGCAACT	65		
XAC3613-Fwd1	TATGCCTTATCTGTGTCTCTTGCCGTGTTGT	65	1374	No other sets of primers were specific
XAC3613-Rev1	TGTCGCCGCTGGATCAAGCAGGTTGTAATC	65		
XAC3050-Fwd1	GCTTTTTTCGGAACCCCGGCAGGCAAA	65	831	No other sets of primers were specific
XAC3050-Rev1	GGTGCCATTGCAAAGCGCAAAGCGAC	65		
XCV3187-Fwd1	TCCAACCGCAGGCTCGCCGCAATTA	65	1285	No other sets of primers were specific
XCV3187-Rev1	GGCCGGGCTGATACAGGTAAGCC	65		
XOO2829-Fwd1	CGTCACCGTGGCAAGGTCCGCGAT	65	824	No other sets of primers were specific
XOO2829-Rev1	ATCGATGACGTCGGCCGCAAGCTA	65		
<i>pilL</i> -Fwd1	CAGGTCATACAGTTCGCGATCAGATCCAAC	65	1221	No other sets of primers were specific
<i>pilL</i> -Rev1	CCGCTGGAACACATGCTGCGCAACTC	65		
<i>pilS</i> -Rev1	TGGTGCCTGGCATTGCTGGTGGATGATCT	65	739	No other sets of primers were specific
<i>pilS</i> -Fwd1	ATCCTGCTGCGGCTGCGCTACGGCA	65		
<i>pilU</i> -Rev1	TCGAACAGGCTCTGGTCTGAAGGTGCG	65	977	No other sets of primers were specific
<i>pilU</i> -Fwd1	ATGAGCACCATCGACTCACCTCCTTCC	65		
<i>pilA</i> -Rev1	GAAGTTGGGTGTCGTCGATTTTCGGAACTG	65	360	No other sets of primers were specific
<i>pilA</i> -Fwd1	GCGTGTTCGGCATCTCGCTGCGAT	65		
<i>xadA2</i> -Rev1	ATGAAGCGAATTTATCTGGAGGCTGGTCCG	65	1015	No other sets of primers were specific
<i>xadA2</i> -Fwd1	TTCGAAACACGTTGGAATCGTAGCCAATG	65		
XCV2103-Fwd1	TTCCGGGACAGCGTGGACATCAGCCAG	65	1570	No other sets of primers were specific
XCV2103-Rev1	CCGATGTCGGTGATGTTGAGCGTGCC	65		
<i>phaB</i> -Fwd1	TGATCCTCAACAACAGTGCAGATCTCCAAG	65	2596	No other sets of primers were specific
<i>phaB</i> -Rev1	ATCGTGCCTGCTATCCAATGCAATCGCACCGGCG	65		
<i>phaB1</i> -Rev1	CGCAGTGGTGAGGTTGGCAGCCTA	65	1354	No other sets of primers were specific
<i>phaB1</i> -Fwd1	TGATCCTCAACAACAGTGCAGATCTCCAAG	65		
<i>phaB2</i> -Fwd1	GTCTGCGACACGCGATACGCAAGTG	65	1238	No other sets of primers were specific
<i>phaB2</i> -Rev1	ATCGTGCCTGCTATCCAATGCAATCGCACCG	65		
XAC1816-Rev1	TTCAGCCAGACTGCTGCAACTCGGT	65	424	No other sets of primers were specific
XAC1816-Fwd1	TACAGCCAGCAGAAGGTCAACAGCAGCTA	65		

1.3 Résultats complémentaires

1.3.1 Matériels et Méthodes

Recherche de séquences codant les MCPs dans les banques de données

L'identification des orthologues des gènes codant les MCPs dans les génomes de six autres souches de *Xanthomonas* a été poursuivie selon la même démarche que celle décrite dans l'article 1. Une analyse BLASTX en date de juin 2009, a été réalisée sur la souche CFBP4834 de *X. axonopodis* pv. *phaseoli*, les souches 8004 et B100 de *X. campestris* pv. *campestris*, les souches MAFF311018 et PXO99^A de *X. oryzae* pv. *oryzae* et la souche BLS256 de *X. oryzae* pv. *oryzicola*. Le génome de la souche CFBP4834 de *X. axonopodis* pv. *phaseoli* a été séquencé par le Génoscope en 2008. La séquence obtenue a été annotée automatiquement grâce à iANT (integrated ANnotation Tool) (Thébaut *et al.*, 2000). La souche 756C du pathovar *armoraciae* a été utilisée dans l'étude des pressions de sélection au sein de l'espèce *X. campestris*. Sa séquence génomique n'est pas disponible dans les bases de données publiques (NCBI) ; elle m'a été rendue accessible par Ralf Koebnik (IRD Montpellier). La recherche des gènes candidats a été réalisée par une analyse BLASTX sur le site CMR (<http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi>) afin d'identifier les positions des gènes dans le génome de la souche 756C. L'extraction des séquences nucléotidiques des gènes cibles a été réalisée à l'aide du logiciel Artemis (Rutherford *et al.*, 2000).

Amplification et séquençage des orthologues des gènes XCV1933, XCV3230 et XCV1940 à partir des souches de X. axonopodis

Les amorces ont été dessinées manuellement sur la base des alignements des séquences des gènes issues des génomes des 10 souches séquencées de *Xanthomonas*. Trois couples d'amorces spécifiques couvrant la totalité des orthologues des gènes XCV1933 et XCV3230 ont été obtenus (Tableau 1). Concernant les orthologues du gène XCV1940, un seul couple d'amorces couvrant la partie N-terminale a été obtenu malgré nos efforts pour dessiner des couples d'amorces spécifiques couvrant la totalité du gène. Ces différents couples d'amorces ont été validés par une amplification

Tableau 1. Séquences nucléotidiques et position des amorces utilisées pour l'amplification des orthologues des gènes XCV1933, XCV3230 et XCV1940 dans les 131 souches de *X. axonopodis*, taille des amplifiats et températures d'hybridation des amorces.

Gènes	Taille (pb)	Amorces	Séquences nucléotidiques (5' → 3')	Position sur le gène		Taille de l'amplifiat (pb)	T (°C)
XCV1933	2382	F1-1933	ATGCAATGGATCAACAATCTGAA ACTGATGCCC	1	33		
		R1-1933	CAGGTCGCGCTGCATGCGCTGCAT GGA	751	780	780	64
		F2-1933	TGGTGGTGTGTTGATCGGCATCG CCGGCGGCCTGGGCTT	581	620		
		R2-1933	TGAGCTCCTCCATCGAGGCGGCG GTTTCTTCCAGATT	1564	1600	1020	67
		F3-1933	TCGCGTGGCGACCTGACCGTGCGC ATGCAT	1360	1389		
		R3-1933	TCAGAACTCTTGCCAGTTACCGTC GGCCAG	3253	2382	1023	64
XCV3230	2037	FWD3230-1	ATGAGTACCGCCCCGGACGCCCCG CAAGACCAACAAGCTC	1	39		
		REV3230-1	ACCACGCCAGCCAGAAGTTTCGG GAACAGGCG	904	935	935	67
		FWD3230-2	AACCCGGCCGCGTTGTCGGCGC	697	718		
		REV3230-2	TCCATCGAACTGACCGCTTCGTTG GTATCGG	1712	1742	1046	67
		FWD3230-3	CTGATCAACGACATTTCCGAGCAG ACCAACATCCT	1534	1568		
		REV3230-3	TCACGCCGGCAGCTTGAAGTCGG CG	2013	2037	504	67
XCV1940	2262	F1-1940	ATGATTGCTCTTCTTCAACGCTAC AACGTCGGCAAGC	1	37		
		R1-1940	GTGATGGCCGCCTTCTCGCCGGGC AAC	992	1022	1022	67

spécifique *in silico* à l'aide du logiciel Amplify version 3.1.4 (<http://engels.genetics.wisc.edu/amplify/>), et par l'amplification spécifique des gènes cibles à partir des souches séquencées. L'amplification des trois gènes a été réalisée sur les 131 souches de la collection de *X. axonopodis* comme décrit dans l'article 1. La taille des amplifiats a été vérifiée sur gel d'agarose (1.5%) par électrophorèse. Les amplifiats ont ensuite été envoyés au plateau technique Séquençage et Génotypage/Ouest-Génopole (Nantes). Le nettoyage des séquences brutes issues du séquençage et l'assemblage des différents fragments ont été effectués à l'aide du logiciel Geneious version 4.8.5 (<http://www.geneious.com/>).

Alignement des séquences obtenues

L'alignement des séquences nucléotidiques a été fait en fonction de l'alignement protéique, avec les paramètres définis par défaut dans le programme CLUSTAL W disponible dans la version 7.0.9.0 du logiciel BioEdit (Hall, 1999).

Analyse des relations phylogénétiques entre les séquences provenant des 10 souches séquencées de Xanthomonas

Le dendrogramme retraçant les relations phylogénétiques entre les différentes séquences a été construit avec la méthode Neighbor Joining disponible dans le logiciel MEGA version 4.0.2 (Tamura *et al.*, 2007) et en utilisant le modèle Jukes-Cantor (Jukes et Cantor, 1969). La robustesse des branches du dendrogramme est évaluée par les valeurs de bootstrap obtenues avec 1000 simulations.

Analyse des pressions de sélection

Test de recombinaison

L'analyse des pressions de sélection avec le logiciel PAML (Yang, 1997) requiert des séquences non recombinantes. Pour vérifier cette condition, un test de recombinaison a été réalisé à l'aide d'un programme disponible sur le site internet http://hammerlab.biosci.arizona.edu/imgc_online.html pour les séquences provenant des souches de *X. campestris*. Concernant les orthologues de XCV1933, XCV3230 et XCV1940, amplifiées chez les souches de *X. axonopodis*, le test de recombinaison a été fait à l'aide du programme ClonalFrame, en comparant les valeurs de vraisemblances

Tableau 2. Les paramètres des différents modèles *site* du programme CODEML implémenté dans le logiciel PAML (Yang, 1997).

Modèle	NSsites	Nombre de paramètres estimés	Paramètres
<i>One ratio</i>	0	1	ω
<i>Neutral</i>	1	2	p_0 ($p_1 = 1 - p_0$) $\omega_0 < 1, \omega_1 = 1$
<i>Selection</i>	2	4	p_0, p_1 ($p_2 = 1 - p_0 - p_1$) $\omega_0 < 1, \omega_1 = 1, \omega_2 > 1$
<i>Discrete</i>	3	5	p_0, p_1 ($p_2 = 1 - p_0 - p_1$) $\omega_0, \omega_1, \omega_2$
<i>Beta</i>	7	2	p, q
<i>Beta&ω</i>	8	4	p_0 ($p_1 = 1 - p_0$) $p, q, \omega_S > 1$

ω est le ratio d_N/d_S

ω_i est une classe de ω . p_i et q correspondent à des proportions de sites dont le ω appartient à la classe ω_i .

Les LRTs (likelihood ratio test) ont été réalisés en comparant le NSsites 2 au NSsites 1, le NSsites 8 au NSsites 7 et le NSsites 3 au NSsites 0.

des modèles prenant en compte des scénarios avec ou sans recombinaison (Didelot et Falush, 2007).

Constitution des haplotypes

Avec la fonction non redondant databases (NRDB) disponible sur le site internet <http://pubmlst.org/perl/mlstanalyse/mlstanalyse.pl?site=pubmlst>, les haplotypes ont été générés à partir des séquences alignées pour les analyses de pressions de sélection avec le logiciel PAML.

Tests des pressions de sélection

Le rapport des substitutions non synonymes (d_N) sur les substitutions synonymes (d_S) est un indicateur de la pression de sélection qui s'exerce sur une séquence codante. Il est estimé par le programme CODEML implémenté dans le logiciel PAML (Yang, 1997). La sélection positive peut affecter différemment les sites (codons) et les branches (souches). Plusieurs tests ont été appliqués pour mettre en évidence des signaux de pressions de sélection affectant les séquences cibles (modèles *site*, *branch* et *branch-site*). Dans le programme CODEML, six modèles *site* permettent à ω (estimateur de d_N/d_S) de varier en fonction des sites pour tester l'existence de pression de sélection et identifier les codons soumis à cette pression de sélection (Tableau 2). Trois modèles testent les hypothèses de neutralité. Le modèle *one ratio* (NSsites 0) suppose que ω est constant entre tous les sites analysés. Le modèle *neutral* (NSsites 1) suppose deux classes de ω : $\omega_0 < 1$ pour les sites maintenus par une pression de sélection purificatrice et $\omega_1 = 1$ pour les sites évoluant de manière neutre. Le modèle *beta* (NSsites 7) suppose que ω suit une distribution β variant dans l'intervalle (0, 1). Trois autres modèles testent les pressions de sélection positive en prenant en compte les sites pour lesquels $\omega > 1$. Le modèle *selection* (NSsites 2) considère les trois classes pour ω : $\omega_0 < 1$, $\omega_1 = 1$ et $\omega_2 > 1$. Le modèle *discrete* (NSsites 3) se base sur une distribution discrète avec trois classes de sites ayant trois ratios ω (ω_0 , ω_1 et ω_2) et trois proportions. Le modèle *beta& ω* (NSsites 8) considère en plus du modèle *beta* (NSsites 7), une classe de ω pouvant être supérieur à 1. Les LRTs (likelihood ratio tests, $LRT = 2 \times (\ln_1 - \ln_2)$) permettent de comparer les modèles testant la sélection positive aux modèles nuls. Ainsi, le modèle NSsites 2 a été comparé au modèle NSsites 1, le modèle NSsites 8 a été comparé au modèle NSsites 7

Nom de la souche ^a	Nom du gène ^b																								
	XCV1940	XCV1945	XCV2044	XCV3021	XCV3230	XCV1702	XCV1933	XCV1941	XCV1947	XCV1951	XCV1954	XCV2625	XCV3338	XCV1939	XCV1942	XCV1944	XCV1948	XCV0669	XCV1778	XCV1938	XCV1952	XCV3577	XCV1955	XCV3261	
<i>Xcv</i> 85-10	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
<i>Xac</i> 306	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
<i>Xap</i> CFBP4834	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
<i>Xoc</i> BLS256	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
<i>Xoo</i> KACC10331	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
<i>Xoo</i> MAFF311018	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
<i>Xoo</i> PXO99 ^A	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
<i>Xcc</i> 33913	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
<i>Xcc</i> 8004	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
<i>Xcc</i> B100	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

Figure 1. Matrice de présence/absence de 24 gènes codant les MCPs présents dans les génomes de 10 souches séquencées de *Xanthomonas* spp. Les orthologues des 24 gènes codant les MCPs chez la souche 85-10 de *X. axonopodis* pv. *vesicatoria* ont été recherchés par BLASTX dans les génomes des deux autres souches de *X. axonopodis*, quatre souches de *X. oryzae* et trois souches de *X. campestris*. Sont indiqués en noir, les gènes présents et en blanc, les gènes absents dans les génomes des souches étudiées. ^a *Xcv* : *X. axonopodis* pv. *vesicatoria* ; *Xac* : *X. axonopodis* pv. *citri* ; *Xap* : *X. axonopodis* pv. *phaseoli* ; *Xoc* : *X. oryzae* pv. *oryzicola* ; *Xoo* : *X. oryzae* pv. *oryzae* ; *Xcc* : *X. campestris* pv. *campestris*. ^b Les noms attribués aux 25 groupes obtenus sont relatifs aux accessions de la souche 85-10 de *X. axonopodis* pv. *vesicatoria*. Ψ : pseudogène XOO2855.

et le modèle NSsites 3 au modèle NSsites 0 comme indiqué par Yang (1997).

Pour tester si la pression de sélection s'exerce de manière différente entre les branches, nous avons examiné les variations de ω en fonction des branches *i.e* entre les souches de notre analyse. Pour cela, nous avons comparé le modèle *branch free ratio* (M1) qui permet à ω de varier en fonction des branches mais pas en fonction des sites, au modèle nul (Nsites 0). Enfin, le modèle *branch-site* dans lequel ω peut varier en fonction des codons et des branches a été testé. Pour toutes les analyses réalisées, lorsque les LRTs sont significatifs, les sites qui sont probablement affectés par la sélection positive sont identifiés par la méthode Bayes empirical Bayes (BEB) qui calcule aussi leurs probabilités postérieures (Yang *et al.*, 2005).

La divergence adaptative des orthologues des gènes XCV1933, XCV3230 et XCV1940, issues des souches de *X. axonopodis* a été évaluée par rapport aux séquences issues des génomes de *X. oryzae* et *X. campestris* à l'aide du test McDonald et Kreitman (McDonald et Kreitman, 1991) implémenté dans le logiciel DnaSP (Rozas *et al.*, 2003). A cet effet, deux conditions ont été testées en assignant en tant que séquence externe, la séquence du gène candidat présente dans la souche BLS256 de *X. oryzae* pv. *oryzicola* ou dans la souche ATCC33913 de *X. campestris* pv. *campestris*.

1.3.2 Résultats

Relations phylogénétiques entre les séquences nucléotidiques de 24 MCPs provenant des 10 souches séquencées de Xanthomonas spp.

Cent soixante cinq séquences nucléotidiques ont été extraites de 10 génomes séquencés de *Xanthomonas* appartenant à des espèces différentes, à l'aide de l'analyse BLASTX. Cette analyse a révélé la présence de 65, 54 et 46 orthologues respectivement dans les différents génomes de *X. axonopodis*, *X. campestris* et *X. oryzae*. Cinq gènes sont ubiquistes étant présents dans les 10 génomes analysés (XCV1940, XCV1945, XCV2044, XCV3021 et XCV3230) (Figure 1). Huit autres gènes sont présents dans le génome d'au moins une souche de chacune des trois espèces *X. axonopodis*,

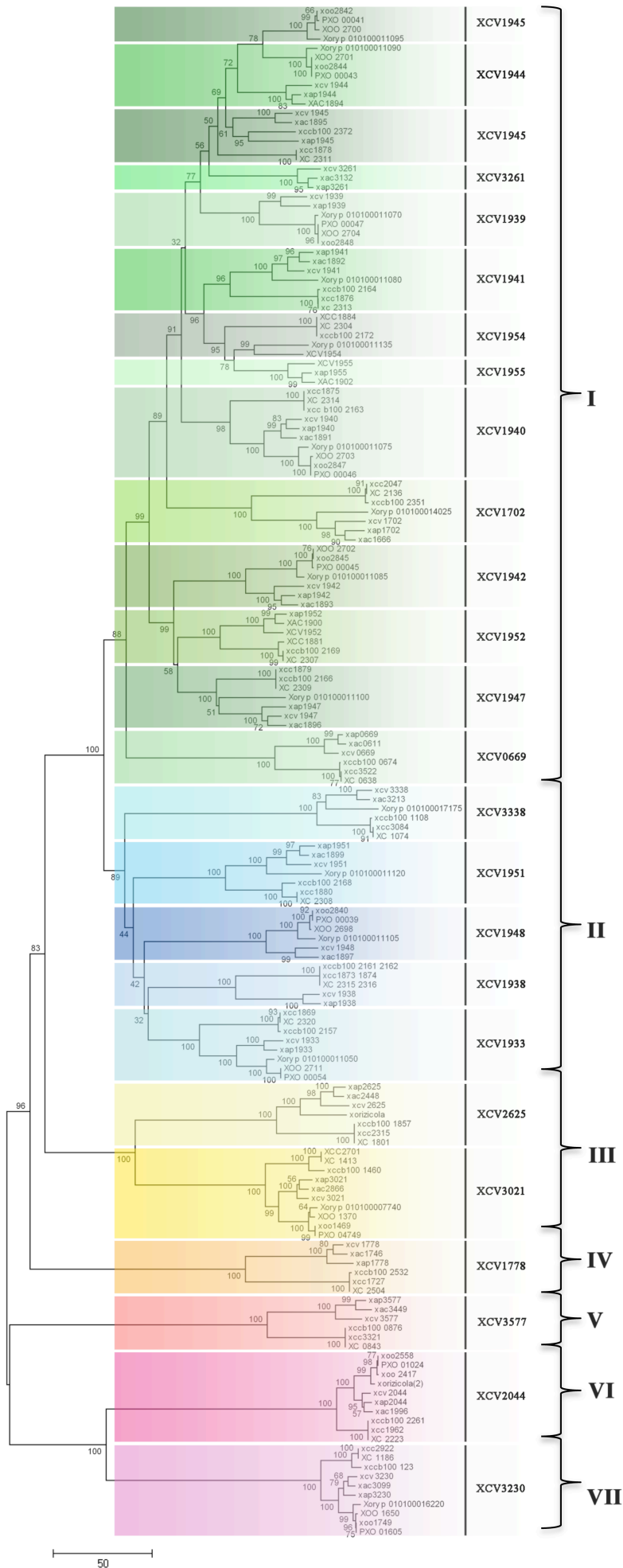


Figure 2. Dendrogramme obtenu par la méthode Neighbor Joining sur la base de l'alignement de 165 séquences codant les MCPs provenant de 10 souches séquencées de *Xanthomonas* spp. Les noms attribués aux 25 groupes obtenus sont relatifs aux accessions de la souche 85-10 de *X. axonopodis* pv. *vesicatoria*. Les 25 groupes s'agrègent en sept grands groupes (I à VII). Les valeurs de bootstraps sont obtenues avec 1000 simulations.

X. campestris et *X. oryzae* (XCV1702, XCV1933, XCV1941, XCV1947, XCV1951, XCV1954, XCV2625 et XCV3338) (Figure 1). Neuf gènes ont un orthologue dans les souches de deux espèces : XCV1939, XCV1942, XCV1944 et XCV1948 sont présents chez *X. axonopodis* et *X. oryzae* alors que XCV0669, XCV1778, XCV1938, XCV1952, XCV3577 sont présents chez *X. axonopodis* et *X. campestris*. Deux gènes sont présents que dans une seule espèce (XCV1955 et XCV3261). Six gènes (XCV1939, XCV1942, XCV1944, XCV1948, XCV1955 et XCV3261) ne présentent aucun orthologue chez *X. campestris* et sept gènes (XCV0669, XCV1778, XCV1938, XCV1952, XCV1955, XCV3261 et XCV3577) sont absents chez *X. oryzae*. Certains gènes peuvent présenter une distribution variable à l'intérieur même de *X. oryzae* et *X. axonopodis* (Figure 1). C'est le cas de XCV1702, XCV1941, XCV1947, XCV1951, XCV1954 XCV2625 et XCV3338 dont les orthologues sont absents dans les trois souches du pathovar *oryzae* alors qu'ils sont présents dans la souche du pathovar *oryzicola*. En ce qui concerne *X. axonopodis*, le gène XCV1954 présent dans la souche 85-10 du pathovar *vesicatoria* ne présente aucun orthologue dans les souches 306 et CFBP4834 des pathovars *citri* et *phaseoli* (Figure 1).

Le dendrogramme généré à partir de l'alignement de ces 165 séquences nucléotidiques a permis de rassembler ces séquences en 25 groupes (Figure 2). Ces groupes comprennent des orthologues d'un même gène provenant des différentes souches de *Xanthomonas*. Seuls les orthologues de XCV1945 sont séparés en deux groupes distincts : l'un regroupe les orthologues de *X. oryzae* distants des orthologues provenant des génomes des autres génomes (Figure 2). Au sein de chaque groupe, les séquences nucléotidiques des gènes codant les MCPs sont généralement rassemblées selon la phylogénie des souches. Ainsi, les séquences issues de *X. axonopodis* sont plus proches des séquences de *X. oryzae* que de celles de *X. campestris* (Figure 2). Ceci concerne les clusters rassemblant les orthologues de XCV1940, XCV2625, XCV3021, XCV2044 et XCV3230. Les 25 groupes sont agrégés en sept grands groupes phylogénétiques (Figure 2), soutenus par de fortes valeurs de bootstraps (de 88 à 100). Les deux premiers grands groupes (I et II) rassemblent des gènes qui sont localisées dans l'îlot chromosomique dédié au chimiotactisme dans chacun des 10 génomes de *Xanthomonas* à l'exception de XCV0669, XCV1702, XCV3261 et XCV3338 qui sont

Tableau 3. Identification des sites soumis à une pression de sélection positive dans l'orthologue de XCV1933 dans la souche B100 de *X. campestris* pv. *campestris*.

Codon	Position du résidu	$p(\omega > 1)$^a
K	86	0,956
N	88	0,973
G	90	0,992
K	94	0,982
G	98	0,991
V	99	0,951
D	102	0,974

^a : ω est le ratio d_N/d_S au niveau des sites et $p(\omega > 1)$, la probabilité postérieure de sites ayant un $\omega > 1$. L'hypothèse testée est celle d'une pression de sélection agissant sur les codons.

isolés dans les génomes. Les cinq derniers groupes composés de un à deux groupes d'orthologues, renferment les autres gènes qui sont dispersés dans les génomes des *Xanthomonas*.

Différents événements de duplication peuvent être identifiés sur le dendrogramme. Ainsi, XCV1944 et XCV1945 sont des paralogues ayant évolué différemment. Dans le cas de XCV1944, les orthologues ont évolué selon la phylogénie des souches mais semblent avoir été perdus par les trois souches de *X. campestris*. En revanche, pour ce qui est de XCV1945, les orthologues présents dans les souches de *X. oryzae* sont plus proches des orthologues de XCV1944 que des orthologues de XCV1945 présents dans les souches de *X. axonopodis* et *X. campestris*. Ce résultat suggère que les orthologues de XCV1945 présents dans *X. oryzae* n'auraient pas beaucoup évolué par rapport à XCV1944. Par contre, les orthologues de XCV1945 présents dans les génomes des souches de *X. axonopodis* et *X. campestris* auraient subi une divergence adaptative plus importante.

Analyse des pressions de sélection exercées sur les gènes candidats provenant de quatre souches de X. campestris

Sur l'ensemble des 19 gènes codant les MCPs analysés dans l'article 1, 16 sont également présents dans les trois génomes des souches séquencées de *X. campestris* pv. *campestris* et dans le génome de la souche 756C de *X. campestris* pv. *armoraciae*. Parmi ces 16 gènes, les orthologues de XCV1940, XCV2625 et XCV3577 ont été écartés de l'analyse des pressions de sélection en raison de l'absence de polymorphisme dans les trois souches de *X. campestris* pv. *campestris*. L'analyse des pressions de sélection avec le modèle *branch-site* a donc été appliquée sur les haplotypes des 13 gènes candidats qui ne montrent aucun signal de recombinaison. Il est important de mentionner qu'à l'exception des orthologues de XCV1952, les séquences des gènes provenant des souches ATCC33913 et 8004 sont identiques. Sur les 13 gènes candidats, 12 ne sont pas soumis à des pressions de sélection positive que ce soit dans le pathovar *campestris* ou dans le pathovar *armoraciae* (Annexe 1). Seul l'orthologue du gène XCV1933 présent dans la souche B100 du pathovar *campestris* est affecté par une pression de sélection positive s'exerçant au niveau de sept sites probables (Tableau 3).

Tableau 4. Résultats des tests de vraisemblance des analyses de recombinaison effectuées avec le programme ClonalFrame (Didelot et Falush, 2007) et appliquées aux haplotypes des orthologues de XCV1933 et XCV3230 et de la région N-terminale du gène XCV1940.

haplotypes d'orthologues de	Valeur de la vraisemblance		LRT ^a	$p(\chi^2)$ ^b
	Modèle avec recombinaison	Modèle sans recombinaison		
XCV1933	-2983,42	-2615,12	-736,60	na
XCV3230	-1964,64	-1929,49	-70,29307	na
XCV1940	-1533,40	-1366,40	-334,001486	na

^a : likelihood ratio test. Les tests LRTs ont été réalisés en testant l'hypothèse de recombinaison. Le degré de liberté du test est égal à 1.

^b : probabilité associée à la loi de Khideux.

na : non applicable

Aucun de ces sites n'est inclus dans un domaine fonctionnel reconnu par BLASTP dans la base de données InterPro.

Etude des pressions de sélection exercées sur trois gènes codant des MCPs dans 131 souches de X. axonopodis

Les orthologues de trois gènes codant des MCPs (XCV1933, XCV3230 et XCV1940) ont été amplifiés dans les souches de *X. axonopodis* de notre collection. En dépit de nos efforts pour amplifier les gènes candidats dans les 131 souches de la collection, des orthologues n'ont été spécifiquement amplifiés pour XCV1933 que dans 42 souches, pour XCV3230 que dans 66 souches et pour la région N-terminale (924 sur 2262 bp) de XCV1940 que dans 64 souches (Annexe 2). Pour ces souches, un signal PCR unique a été obtenu. Par contre, pour les autres souches de la collection, soit l'absence d'amplifiat soit des profils multi-bandes ont été obtenus. Le test de recombinaison a révélé qu'il n'existe pas d'événements de recombinaison au sein de nos jeux de données (Tableau 4).

A partir des séquences amplifiées des gènes XCV1933, XCV3230 et la partie N-terminale de XCV1940 dans les souches de *X. axonopodis* (Annexe 2), 30, 35 et 23 haplotypes ont été respectivement obtenus. Pour tous ces gènes, les haplotypes représentent 13 pathovars sur les 21 étudiés. Néanmoins, ces derniers diffèrent en fonction du gène. Les pathovars *axonopodis*, *betae*, *bilvae*, *vasculorum*, *vignicola* et *spondiae* ne figurent pour aucun des gènes candidats. Par ailleurs, pour chacun des trois gènes, le nombre d'haplotypes par pathovar est variable. Il est compris entre un et sept haplotypes pour XCV1933 et un et neuf haplotypes pour XCV3230 et la partie N-terminale de XCV1940. Généralement, les haplotypes sont propres à chaque pathovar et c'est le pathovar *allii* qui renferme le plus d'haplotypes (Annexe 2).

Au sein de X. axonopodis

Avec le modèle *site*, l'hypothèse selon laquelle les orthologues de XCV1933 évoluent de manière neutre est rejetée. Un codon en position 740 est identifié comme probablement ($p=0,957$) soumis à une pression de sélection positive dans toutes les séquences analysées (Tableau 5). Le modèle *branch* indique que les branches

Tableau 5. Résultats des tests de vraisemblance des analyses de pressions de sélection effectuées avec les six modèles *sites* (NSsites 0 à 8) et le modèle *branch* (M1) du programme CODEML et appliquées aux haplotypes des orthologues codant les MCPs et identification des sites probablement soumis à sélection positive.

Séquences d'orthologues de	Modèle testé	Valeur de la vraisemblance	np ^a	LRT ^b	ddl ^c	$p(\chi^2)$ ^d	codon sous pression de sélection
XCV1933	NSsites 0	-4966,71	59				
	NSsites 3	-4927,45	65	78,52	6	7,19E-15	
	NSsites 1	-4927,53	60				
	NSsites 2	-4927,32	62	0,42	2	0,81	
	NSsites 7	-4947,45	60				
	NSsites 8	-4927,42	62	40,05	2	2,01E-09	A 740 ($p=0,957^*$)
	M1	-4926,22	115	80,98	56	0,02	
XCV3230	NSsites 0	-3966,00	69				
	NSsites 3	-3964,68	73	2,63	4	0,62	
	NSsites 1	-3965,25	70				
	NSsites 2	-3965,25	72	0	2	1	
	NSsites 7	-3964,69	70				
	NSsites 8	-3964,69	72	-2E-3	2	na	
	M1	-3943,13	135	45,73	66	0,97	
XCV1940	NSsites 0	-2309,12	45				
	NSsites 3	-2289,86	49	38,53	4	8,73E-08	
	NSsites 1	-2291,15	46				
	NSsites 2	-2291,15	48	0	2	1	
	NSsites 7	-2290,10	46				
	NSsites 8	-2290,10	48	-3,6E-05	2	na	
	M1	-2274,57	87	69,01	42	5,00E-3	

^a: nombre de paramètres estimés.

^b: likelihood ratio test. Les LRTs des modèles *branch* et *site* ont été réalisés en comparant le modèle NSsites 3 au modèle NSsites 0, le modèle NSsites 2 au modèle NSsites 1 et le modèle NSsites 8 au modèle NSsites 7. Les tests LRTs du modèle M1 ont été effectués en comparant ce modèle au modèle NSsites 0.

^c: degré de liberté.

^d: probabilité associée à la loi de Khideux. L'hypothèse testée est celle d'une évolution neutre.

* indique que la probabilité p que le codon soit soumis à une pression de sélection est significative $p>0.95$.

na : non applicable.

Tableau 6. Valeurs des estimateurs inférés par les modèles *sites* non rejetés du programme CODEML pour les 35 haplotypes d'orthologues de XCV3230.

Modèle	Paramètres estimateurs
<i>one ratio</i> (NSsites 0)	$\omega=0,01260$
<i>neutral</i> (NSsites 1)	$p_0=0,99649, \omega_0=0,01033$ $p_1=0,00351, \omega_1=1$
<i>discrete</i> (NSsites 3)	$p_0=0,58392, \omega_0=0$ $p_1=0,29523, \omega_1=0$ $p_2=0,12085, \omega_2=0,10712$
<i>beta</i> (NSsites 7)	$p=0,04604, q=2,31140$

n'évoluent pas de manière similaire dans notre jeu de données. Les analyses de pression de sélection ont été poursuivies avec le modèle *branch-site*. Un acide aminé (A) en position 578 de la séquence haplotype H24 a été alors identifié comme probablement ($p=0,980$) soumis à une pression de sélection positive (Annexe 3). Cet haplotype correspond aux séquences amplifiées chez les souches CFBP5863 et CFBP5864 du pathovar *ricini* (Annexe 2).

Concernant le gène XCV3230, l'hypothèse testant la neutralité de l'évolution n'est pas rejetée (Tableau 5). Ce résultat indique que d'une part, les codons des orthologues de XCV3230 provenant des souches de *X. axonopodis* ne sont pas soumis à une pression de sélection positive et d'autre part, que les branches évoluent de la même manière (Tableau 5). Le modèle *one ratio* (NSsites 0) indique que la valeur moyenne des ratios ω est de 0,01260 ; ω est très inférieur à 1. Avec le modèle *neutral* (NSsites 1), 99,65% des sites analysés ont une valeur de ω égale à 0,01033. Le modèle *discrete* (NSsites 3) montre que 87,9% des sites analysés présentent des valeurs de ω égale à zéro (Tableau 6). Dans la mesure où la majorité des sites présentent des valeurs de ω très inférieures à 1, cela suggère fortement qu'au sein de *X. axonopodis*, les orthologues de XCV3230 évoluent sous l'action d'une pression de sélection purificatrice.

Pour les orthologues de la région N-terminale du gène XCV1940, le modèle *site* indique que ω n'est pas constant entre les codons. En revanche, aucun site potentiellement soumis à une pression de sélection positive n'a été identifié (Tableau 5). Selon le modèle *branch*, l'hypothèse selon laquelle les branches évoluent de la même manière est rejetée (Tableau 5). Pour ces séquences, l'analyse des pressions de sélection a été poursuivie avec le modèle *branch-site*. Un acide aminé (N) en position 239 a été identifié comme étant probablement ($p=0,960$) affecté par une pression de sélection positive (Annexe 4). Ce site se trouve sur l'haplotype H13 correspondant aux séquences amplifiées chez les souches CFBP3133, CFBP5688 et CFBP5691 de la lignée génétique C du pathovar *dieffenbachiae* (Annexe 2).

Dans la divergence de *X. axonopodis* par rapport à *X. campestris* et *X. oryzae*

La divergence adaptative des orthologues de XCV1933, XCV3230 et de la

Tableau 7. Tableau de contingence comprenant les résultats du test McDonald et Kreitman (McDonald et Kreitman, 1991) réalisées avec les orthologues de XCV1933 et XCV3230 et de la région N-terminale du gène XCV1940 et à l'aide du logiciel DnaSP (Rozas *et al.*, 2003).

Orthologues du gène ^a	Conditions du test ^b	Rapport du polymorphisme non synonyme sur le polymorphisme synonyme ^c	Rapport des mutations non synonymes fixées / les mutations synonymes fixées ^d	Valeur de <i>p</i> ^e	Valeur de <i>G</i> ^f
XCV1933	<i>X. axonopodis</i> vs <i>XocBLS256</i>	20/180	39/72	0,000***	28,21***
	<i>X. axonopodis</i> vs <i>XccATCC33913</i>	20/180	150/174	0,000***	82,69***
XCV3230	<i>X. axonopodis</i> vs <i>XocBLS256</i>	13/150	28/51	0,000***	26,40***
	<i>X. axonopodis</i> vs <i>XccATCC33913</i>	13/150	22/81	0,002**	8,90**
XCV1940	<i>X. axonopodis</i> vs <i>XocBLS256</i>	48/114	19/24	0,099	3,12
	<i>X. axonopodis</i> vs <i>XccATCC33913</i>	48/114	105/82	0,000***	25,10***

^a : les orthologues de XCV1933, XCV3230 et de la région N-terminale du gène XCV1940 ont été amplifiées respectivement dans 42, 65 et 64 souches de *X. axonopodis* de la collection.

^b : le test McDonald et Kreitman a été réalisé en assignant en tant que séquence externe, la séquence du gène candidat présente dans le génome de la souche BLS256 de *X. oryzae* pv. *oryzicola* ou dans celui de la souche ATCC33913 de *X. campestris* pv. *campestris*.

^c : au sein de l'espèce *X. axonopodis*

^d : entre les espèces

^e : la significativité du test McDonald et Kreitman est déterminée par le test exact de Fisher (* 0.01 < *p* < 0,05; ** 0.001 < *p* < 0,01; *** *p* < 0,001).

^f : valeur de la vraisemblance et significativité associée à la probabilité du test G d'indépendance (* 0.01 < *p* < 0,05; ** 0.001 < *p* < 0,01; *** *p* < 0,001).

région N-terminale de XCV1940 amplifiés respectivement chez 42, 66 et 64 souches de *X. axonopodis*, a été analysée par rapport à deux séquences outgroup chacune provenant de la souche BLS256 de *X. oryzae* pv. *oryzicola* et de la souche ATCC33913 de *X. campestris* pv. *campestris*. Pour les orthologues de XCV1933 et XCV3230, des signaux robustes d'adaptation ont été détectés au niveau de la divergence entre d'une part, *X. axonopodis* et la souche BLS256 de *X. oryzae* pv. *oryzicola* et d'autre part, *X. axonopodis* et la souche ATCC33913 de *X. campestris* pv. *campestris* (Tableau 7). En ce qui concerne les orthologues de la région N-terminale du gène XCV1940, les signatures d'adaptation ont seulement été détectées au niveau de la divergence entre *X. axonopodis* et la souche ATCC33913 de *X. campestris* pv. *campestris* (Tableau 7).

1.4 Discussion

L'étude de l'implication des senseurs et des adhésines dans la spécificité d'hôte des *Xanthomonas* a été réalisée en caractérisant les répertoires de gènes candidats dans une vaste collection de souches, en analysant les relations phylogénétiques entre les différents gènes codant les MCPs et les histoires évolutives de ces gènes avec plusieurs jeux de données.

Pour étudier les relations phylogénétiques entre 165 séquences de gènes codant les MCPs provenant de 10 souches séquencées de *Xanthomonas*, nous avons construit un dendrogramme. Lorsque les groupes comprennent des orthologues de MCPs provenant des 10 souches séquencées, les séquences provenant des souches de *X. axonopodis* sont plus proches des séquences de *X. oryzae* que celles de *X. campestris*. Ces regroupements sont en accord avec la phylogénie des souches. Selon le schéma MLSA établi par Young *et al.* (2008), *X. axonopodis* est plus proche de *X. oryzae* que de *X. campestris*. Par ailleurs, le dendrogramme construit dans le cadre de l'étude des répertoires de gènes codant les senseurs et les adhésines dans les souches de *X. axonopodis*, regroupe globalement les souches en fonction des groupes 9.1 à 9.6 définis sur la base du polymorphisme d'éléments répétés par Rademaker *et al.* (2005) au sein de *X. axonopodis* (*cf.* Figure 2 de l'article 1). Ces résultats suggèrent que les séquences de gènes codant les MCPs auraient évolué par transfert vertical *i.e* suivant la

phylogénie des souches.

La phylogénie des gènes codant les MCPs suggère que plusieurs événements de duplication sont à l'origine de l'expansion du nombre de copies de ces gènes dans les génomes des souches de *Xanthomonas*. Ces événements de duplication se sont probablement produits dans l'ancêtre commun des *Xanthomonas* avant les phénomènes de spéciation. En effet, 13 gènes sont présents dans les trois espèces étudiées : *X. axonopodis*, *X. campestris* et *X. oryzae*. De plus, l'organisation des gènes codant les MCPs est conservée dans les souches séquencées des *Xanthomonas*. Ceci généralise les résultats obtenus par Moreira *et al.* (2004) concernant les souches séquencées 306 de *X. axonopodis* pv. *citri* et ATCC33913 de *X. campestris* pv. *campestris*. Certains orthologues ne sont présents que dans deux des trois espèces, soit *X. axonopodis* et *X. oryzae* soit *X. axonopodis* et *X. campestris*. Ceci suggère fortement qu'après les événements de duplication qui ont eu lieu dans l'ancêtre commun, des gènes codant les MCPs auraient été perdus par réduction du génome dans certaines espèces. Ceci est en accord avec les travaux de Qian *et al.* (2008) qui montrent que les génomes des souches KACC10331 et MAFF311018 de *X. oryzae* pv. *oryzae* ont effectivement perdu 37 gènes codant des protéines du système de régulation à deux composants sur les 70 conservées entre les différentes souches séquencées de *Xanthomonas*. En revanche, la perte des gènes codant les MCPs peut être due à des pseudogénéisations. En effet, le génome de la souche KACC10331 de *X. oryzae* pv. *oryzae* possède un pseudogène XOO2855 (identifié par TBLASTX) dont la séquence nucléotidique présente une forte identité avec le gène XCV1933 qui code un chémorécepteur. Avec notre démarche expérimentale, les séquences de pseudogènes n'ont pas été recueillies par l'analyse BLASTX puisqu'elle ne tient compte que des séquences codantes. Enfin, quelques MCPs (ex. les orthologues de XCV1941 et XCV1702), auraient probablement été perdus dans certaines souches (MAFF311018, KACC10331 et PXO99^A) du pathovar *oryzae*.

Par ailleurs, le fait que les paralogues qui co-localisent dans le génome soient phylogénétiquement éloignés des paralogues dispersés dans le génome, suggère que ces séquences auraient subi des histoires évolutives différentes. Ceci pourrait être en partie

expliqué par les contextes génomiques auxquels appartiennent ces différents gènes. En effet, nous avons pu montrer avec l'étude des pressions de sélection exercées sur trois gènes codant des MCPs dans les souches de *X. axonopodis* de notre collection, que les orthologues des gènes XCV1933 et XCV1940 regroupés dans l'îlot génomique dédié à la réponse chimiotactique et à la mobilité des bactéries, sont soumis à des pressions de sélection positive. Cet îlot comprend notamment des gènes impliqués dans la transduction du signal chimiotactique (*cheA-cheZ*) et les gènes codant le flagelle (*flgA-L, flhA-F, fliC* et *motA-B*). En revanche, le gène XCV3230 qui est isolé dans le génome est soumis à une pression de sélection purificatrice. Ces gènes se localisent dans des contextes relativement différents du contexte des gènes regroupés même s'ils sont en présence de quelques autres gènes impliqués dans la signalisation. Ces résultats sont en accord avec ceux de Moreira *et al.* (2004). Ces auteurs rapportent que chez *X. axonopodis* pv. *citri*, les copies de MCPs regroupés présentent entre elles une plus forte identité qu'avec les copies qui sont dispersées dans le reste du génome.

Les résultats des tests de McDonald et Kreitman présentés dans l'article 1, montrent que pour la majorité des gènes codant les MCPs (ex. XCV0669, XCV1778, XCV1933, *etc*), des signaux robustes d'adaptation sont détectés au niveau de la divergence entre *X. axonopodis* et *X. campestris*. Il était donc intéressant d'examiner l'histoire évolutive de ces gènes chez *X. campestris*. Nos résultats montrent que 12 des 13 gènes candidats analysés ne sont pas soumis à des pressions de sélection dans les quatre souches de *X. campestris*. Il est important de mentionner qu'à l'intérieur de *X. campestris*, ces souches auraient la même gamme d'hôtes, seule la souche du pathovar *armoraciae* induit une pathologie différente (Fargier et Manceau, 2007). Cette situation est différente de celle observée chez *X. axonopodis* présentée dans l'article 1. Ainsi, ces résultats confortent notre hypothèse concernant l'importance de la pression de sélection exercée par l'hôte. Cette pression de sélection n'est visible qu'à l'intérieur de *X. axonopodis* pour laquelle les séquences des gènes appartiennent à trois souches qui attaquent des plantes hôtes appartenant à des familles botaniques différentes (ex. Solanacées, Rutacées, Fabacées).

L'étude de la divergence adaptative des orthologues de XCV1933 et de la région

N-terminale de XCV1940 confirme les résultats obtenus dans l'article 1. En effet, pour les orthologues de XCV1933, des signaux d'adaptation robustes sont détectés au niveau de la divergence entre *X. axonopodis* et *X. oryzae* et entre *X. axonopodis* et *X. campestris*. Par ailleurs, les orthologues de XCV1933 sont soumis à une pression de sélection positive au sein de *X. campestris* et *X. axonopodis*. En ce qui concerne les orthologues de la région N-terminale de XCV1940, nous retrouvons bien les signaux d'adaptation entre *X. axonopodis* et *X. oryzae*. En accord avec les résultats obtenus dans l'article 1, aucune pression de sélection positive n'agit au niveau de la divergence entre *X. axonopodis* et *X. campestris*. Nous avons pu vérifier pour les orthologues du gène XCV3230, que des pressions de sélection positive agissent au niveau de la divergence entre *X. axonopodis* et *X. oryzae*. En revanche, nous avons détecté en plus, des signaux forts d'adaptation dans la divergence de *X. axonopodis* et *X. campestris* pour les orthologues de XCV3230. La détection de signaux de pression de sélection positive indique que les MCPs seraient impliqués dans la divergence adaptative au sein des *Xanthomonas*.

Nous avons également analysé les pressions de sélection qui affectent l'évolution des orthologues de XCV1933, XCV3230 et de la partie N-terminale de XCV1940 au sein des souches de *X. axonopodis* de notre collection. Deux types de résultats étaient attendus *i)* une pression de sélection purificatrice agissant sur les gènes impliqués dans la reconnaissance de molécules végétales communes et/ou *ii)* une pression de sélection positive agissant sur les gènes codant les MCPs impliqués dans la reconnaissance de molécules spécifiques d'une niche particulière. Ces deux cas de figure sont présents dans notre jeu de données. En effet, les orthologues de XCV3230 sont soumis à une pression de sélection purificatrice alors que certains orthologues de XCV1933 et de la région N-terminale de XCV1940 sont soumis à une pression de sélection positive. Ces résultats suggèrent que les orthologues de XCV3230 pourraient être impliqués dans la détection d'une molécule d'origine végétale communément reconnue par les *Xanthomonas* (ex. des acides gras composant la cuticule ou des aldéhydes volatiles) alors que les orthologues de XCV1933 et XCV1940 seraient impliqués dans la détection de molécules végétales plus spécifiques d'une niche (ex. sucres et dipeptides).

Nous n'avons pas réussi à amplifier les gènes dans l'ensemble des souches de la collection alors que ces gènes devaient être présents. En effet, leur présence ubiquiste avait été révélée lors des caractérisations des répertoires de senseurs et d'adhésines (article 1). Par contre, dans cette étude, les signaux d'amplification ont été obtenus suite à l'amplification d'un fragment soit qui ne recouvrait pas la totalité du gène soit qui était trop grand pour être séquencé (*cf* Table S3 de l'article 1). L'amplification n'a pas été très efficace car pour certaines souches, aucun signal ou des signaux multiples ont été obtenus. Ceci peut être expliqué par : *i*) le défaut de spécificité des amorces au niveau des sites d'accrochages et/ou *ii*) par la présence de domaines répétées au niveau des gènes codant les MCPs. Il serait intéressant de faire séquencer les différentes bandes présentes dans les profils multi-bandes obtenus pour de nombreuses souches. Ainsi, en complétant le jeu de données, des tests McDonald et Kreitman pourront être réalisés pour étudier les pressions de sélection au niveau de la divergence entre les pathovars définis au sein des groupes 9.1 à 9.6 de l'espèce *X. axonopodis*. En outre, une représentation plus complète des souches de *X. axonopodis* permettrait probablement d'identifier d'autres sites différemment soumis à une pression de sélection positive chez d'autres pathovars de *X. axonopodis*.

2. Etude des relations phylogénétiques et généalogiques entre les souches de *X. axonopodis*

2.1 Présentation de l'étude phylogénétique et de l'article 2

Les études phylogénétiques visent à étudier l'évolution des organismes vivants en vue d'établir leur parenté. Elles se basent généralement sur un arbre phylogénétique qui représente le degré de parenté entre différents organismes ayant un ancêtre commun (Gevers *et al.*, 2005 ; Morrison, 2005). Les caractères actuellement utilisés pour la construction des arbres phylogénétiques sont présents dans le génome des organismes étudiés ; il s'agit principalement des séquences codant une protéine. Ces données moléculaires permettent d'inférer une histoire pour les espèces pour lesquelles on dispose de peu de caractères morphologiques. La comparaison entre les individus repose sur la recherche d'homologie entre les séquences. Ainsi, les longueurs des branches de l'arbre phylogénétique, construit à partir de l'alignement des séquences obtenues, reflètent les quantités de changements évolutifs accumulés et sont exprimées généralement en nombre de substitutions par site (Didelot, 2010). De plus, les valeurs de bootstraps indiquées sur l'arbre phylogénétique permettent d'évaluer la robustesse statistique de chaque branche (Morrison, 2005).

La MLSA est une méthode phylogénétique compilant l'analyse simultanée de plusieurs gènes. Elle se base sur la concaténation des séquences de différents gènes dits de ménage assurant des fonctions indispensables au bon fonctionnement de la cellule (Wertz *et al.*, 2003 ; Christensen *et al.*, 2004 ; Holmes *et al.*, 2004 ; Gevers *et al.*, 2005). Le but de l'utilisation de plusieurs gènes est de se rapprocher au maximum du génome complet des individus étudiés (Mignard, 2008). Néanmoins, les gènes peuvent être affectés par des événements de recombinaison et avoir des histoires évolutives

différentes. Par conséquent, ils risquent individuellement de ne pas retracer des phylogénies congruentes (Tenaillon *et al.*, 2010). En effet, de nombreuses données accumulées récemment montrent que la recombinaison est un processus fondamental dans l'évolution des bactéries (Posada *et al.*, 2002 ; Rosenberg et Nordborg, 2002). Plusieurs études ont révélé que la recombinaison se produit au sein de nombreux genres et espèces bactériennes pathogènes comme *Neisseria* ou *Agrobacterium tumefaciens* (Feil *et al.*, 1996 ; Costechareyre *et al.*, 2009). Il est à noter que les événements de recombinaison peuvent provoquer des erreurs de placements de souches dans les arbres phylogénétiques (Hanage *et al.*, 2005). En revanche, selon Tenaillon *et al.* (2010), la recombinaison n'affecte que des petits fragments génomiques ce qui n'altère pas la topologie globale de l'arbre construit avec les données concaténées qui sont de longueur relativement importantes. Le test Shimodaira-Hasegawa (S-H) permet d'évaluer la significativité des incongruences des topologies des arbres phylogénétiques. Actuellement, il ne semble pas y avoir de consensus quant aux effets de la recombinaison dans les études phylogénétiques. Par exemple, les études menées sur les *Xanthomonas* sont poursuivies bien que le S-H test montre des incongruences entre les arbres phylogénétiques. Dans ce cas (Fargier, 2007 ; Bui Thi Ngoc *et al.*, 2010), les arbres individuels sont néanmoins généralement congruents avec l'arbre construit sur la base des données concaténées.

Une approche alternative basée sur la coalescence prend en compte des scénarios avec et sans recombinaison. Elle retrace, sur la base des séquences contemporaines de gènes de ménage, la généalogie des souches. La coalescence présente l'avantage d'estimer des paramètres de génétique des populations comme les taux de mutation, de migration, de déterminer les ancêtres communs et de préciser les temps de divergence (Rosenberg et Nordborg, 2002). Contrairement à la phylogénie qui ne considère qu'une seule histoire évolutive, la coalescence explore plusieurs scénarios évolutifs probables ou généalogies qui incorporent ou non les événements de recombinaison et de migration. Le choix du modèle évolutif repose sur des critères statistiques et sur des paramètres biologiques (épidémiologie, évolution, *etc*) (Rosenberg et Nordborg, 2002 ; Didelot et Falush, 2007 ; Kuhner, 2008).

Les résultats obtenus par l'approche classique de phylogénie sont présentés dans une première partie de ce chapitre et ceux obtenus par l'approche basée sur la coalescence sont présentés dans l'article intitulé « *Evidence of recombination and migration among pathovars of X. axonopodis species* » qui sera très prochainement soumis pour publication dans la revue « *Molecular Ecology* ».

2.2 Relations phylogénétiques entre les souches de *X. axonopodis*

2.2.1 Matériels et Méthodes

Cent trente huit souches de l'espèce *X. axonopodis sensu* Vauterin *et al.* (1995) ont été sélectionnées (Annexe 5 et *cf.* Tableau 1 de l'article 2). Ces souches, disponibles dans la Collection Française des Bactéries Phytopathogènes (CFBP) - Centre INRA Angers - Nantes, appartiennent à 21 pathovars de *X. axonopodis*. Dix neuf de ces 21 pathovars sont monophylétiques *i.e* composés de souches phylogénétiquement proches. Des travaux précédents ont montré que deux des pathovars (pvs. *dieffenbachiae* et *phaseoli*) sont polyphylétiques (Rademaker *et al.*, 2005 ; Alavi *et al.*, 2008 ; Hajri *et al.*, 2009). Les deux lignées du pathovar *dieffenbachiae* devraient être chacune considérée comme un pathovar authentique puisque les souches ont des gammes d'hôtes différentes. Quant aux souches du pathovar *phaseoli*, elles se regroupent en quatre lignées génétiques distinctes. Ainsi, les 138 souches appartiennent à 25 lignées génétiques.

L'étude phylogénétique repose sur l'analyse des séquences de sept gènes de ménage : le gène *atpD* codant la sous-unité β de l'ATP synthétase, le gène *dnaK* codant la chaperone DnaK, le gène *efp* codant le facteur d'élongation P, le gène *fyuA* codant un transporteur TonB-dépendant, le gène *glnA* codant la glutamine synthétase, le gène *gyrB* codant la subunité B de l'ADN gyrase et le gène *rpoD* codant le facteur σ de l'ARN polymérase. La majorité de ces gènes a déjà été utilisée dans d'autres études

Tableau 1. Gènes de ménage utilisés dans différentes études des relations phylogénétiques entre des souches appartenant à *Xanthomonas* spp.

Nombre de gènes étudiés	Gènes utilisés	Cibles de l'étude	Références
4	<i>dnaK, fyuA, gyrB, rpoD</i>	Le genre <i>Xanthomonas</i>	Young <i>et al.</i> , 2008
3	<i>atpD, gyrB, dnaK</i>	L'espèce <i>X. axonopodis</i>	Ah-You <i>et al.</i> , 2009
1	<i>rpoD</i>	L'espèce <i>X. axonopodis</i>	Hajri <i>et al.</i> , 2009
4	<i>atpD, dnaK, efp, gyrB</i>	L'espèce <i>X. axonopodis</i>	Bui Thi Ngoc <i>et al.</i> , 2010
4	<i>atpD, dnaK, efp, gyrB</i>	Les espèces <i>X. euvesicatoria</i> , <i>X. gardneri</i> , <i>X. perforans</i> et <i>X. vesicatoria</i>	Hamza <i>et al.</i> , 2010
8	<i>atpD, dnaK, efp, fyuA, glnA, gyrB, rpoD, tpiA</i>	L'espèce <i>X. campestris</i>	Fragier, 2007

Tableau 2. Taille des fragments amplifiés des sept gènes de ménage et positionnement dans la séquence concaténée.

Locus	Taille du fragment	Position le concaténat (bp)
<i>atpD</i>	777	1-777
<i>dnaK</i>	897	778-1674
<i>efp</i>	360	1675-2034
<i>fyuA</i>	864	2035-2898
<i>glnA</i>	984	2899-3882
<i>gyrB</i>	801	3883-4683
<i>rpoD</i>	783	4684-5466

phylogénétiques sur les *Xanthomonas* ; c'est le cas de *atpD*, *dnaK*, *efp*, *fyuA*, *glnA*, *gyrB* et *rpoD* (Tableau 1). Ces gènes sont distribués de manière homogène dans le génome de la souche ATCC33913 de *X. campestris* pv. *campestris* (Fargier, 2007). Nous avons amplifié ces gènes dans les 138 souches de *X. axonopodis* de la collection et les avons séquencés. Une fois les séquences obtenues, nous les avons nettoyées et les avons alignées comme décrit dans l'article 2. Les séquences des 7 gènes ont été concaténées (Tableau 2) en suivant l'ordre alphabétique de leur nom avec le logiciel Geneious version 4.8.5 (<http://www.geneious.com/>).

Etude phylogénétique par MLSA

Les analyses phylogénétiques ont été réalisées sur les séquences nucléotidiques de chaque gène de ménage et également sur les données concaténées. Les arbres phylogénétiques ont été construits avec la méthode Neighbor Joining disponible dans le logiciel MEGA version 4.0.2 (Tamura *et al.*, 2007) en utilisant le modèle Jukes-Cantor (Jukes et Cantor, 1969). La méthode Neighbor Joining se base sur le principe de l'évolution minimale et attribue le même poids à tous les caractères étudiés. Les séquences nucléotidiques des gènes de ménage de trois souches séquencées (la souche ATCC33913 de *X. campestris* pv. *campestris* et les souches KACC10331 et MAFF311018 de *X. oryzae* pv. *oryzae*) ont été utilisées pour enraciner l'arbre. Enfin, la robustesse des branches des arbres phylogénétiques est renseignée par les valeurs de bootstrap obtenues avec 1000 simulations.

Afin de tester la congruence des topologies des différents arbres phylogénétiques, le test Shimodaira-Hasegawa (SH) du programme DNAML implémenté dans le logiciel Phylip, a été appliqué sur nos données (Shimodaira et Hasegawa, 1999 ; Felsenstein, 2005). Ce test vérifie si les valeurs de vraisemblance associées aux topologies des arbres phylogénétiques sont dans le même intervalle de confiance (Shimodaira et Hasegawa, 1999). Une différence significative entre les valeurs de vraisemblance de deux arbres phylogénétiques indique que les gènes ont subi des évolutions différentes probablement dues à des événements de recombinaison.

Afin de mettre en évidence d'éventuels événements de recombinaison au sein des souches de la collection, nous avons réalisé deux analyses. La première est de type graphique permettant de visualiser les éventuels signaux conflictuels dans nos données à l'aide de réseaux. A cet effet, nous avons généré à partir de l'alignement de chaque gène, des types de séquences (STs, de l'anglais « sequence types ») avec la fonction non redundant databases (NRDB) disponible sur le site <http://pubmlst.org/perl/mlstanalyse/mlstanalyse.pl?site=pubmlst>. Pour chaque gène, l'alignement des STs a été analysé avec la méthode NeighborNet disponible dans la version 4.11.3 du logiciel SplitsTree (Huson et Bryant, 2006), NeighborNet étant basée sur les distances qui représentent les connections entre les taxa étudiés (Bryant et Moulton, 2004 ; Morrison, 2005). La seconde méthode est algorithmique. Nous avons employé le logiciel RDP version 3.38 (Martin et Rybicki, 2000) pour détecter et localiser les éventuels sites de recombinaison, calculer les événements minimums de recombinaison et identifier les séquences recombinantes et parentales. Nous avons utilisé sept méthodes de détection disponibles dans ce logiciel: RDP, GENECONV, MaxChi, Chimaera, BootScan, SiScan et 3Seq. Les analyses ont été réalisées sur les données concaténées avec les paramètres définis par défaut. Avec ce logiciel, les événements de recombinaison sont acceptés lorsqu'ils sont détectés par au moins trois des sept méthodes.

Etude de la diversité des souches au sein de X. axonopodis par MLST

L'analyse MLST (Multilocus sequence typing) a été développée pour étudier les relations des souches au niveau infraspécifique (Maiden *et al.*, 1998). Conventionnellement, sept gènes de ménage sont analysés pour caractériser les souches à l'intérieur d'une espèce. Chaque allèle est assigné à un nombre et chaque souche est définie par un profil allélique correspondant à une combinaison de différents allèles des gènes étudiés. L'étude de la diversité et de la structure de la population se base sur la constitution de groupes possédant des profils alléliques différents (Maiden *et al.*, 1998). Ainsi, pour les souches de *X. axonopodis* de la collection, des profils alléliques ont été définis arbitrairement. Chaque profil allélique est représenté par un ST. Les STs ont été analysés par eBURST (<http://eburst.mlst.net/>) pour étudier la diversité de la population au sein de *X. axonopodis* à travers la formation de complexes clonaux et de singletons.

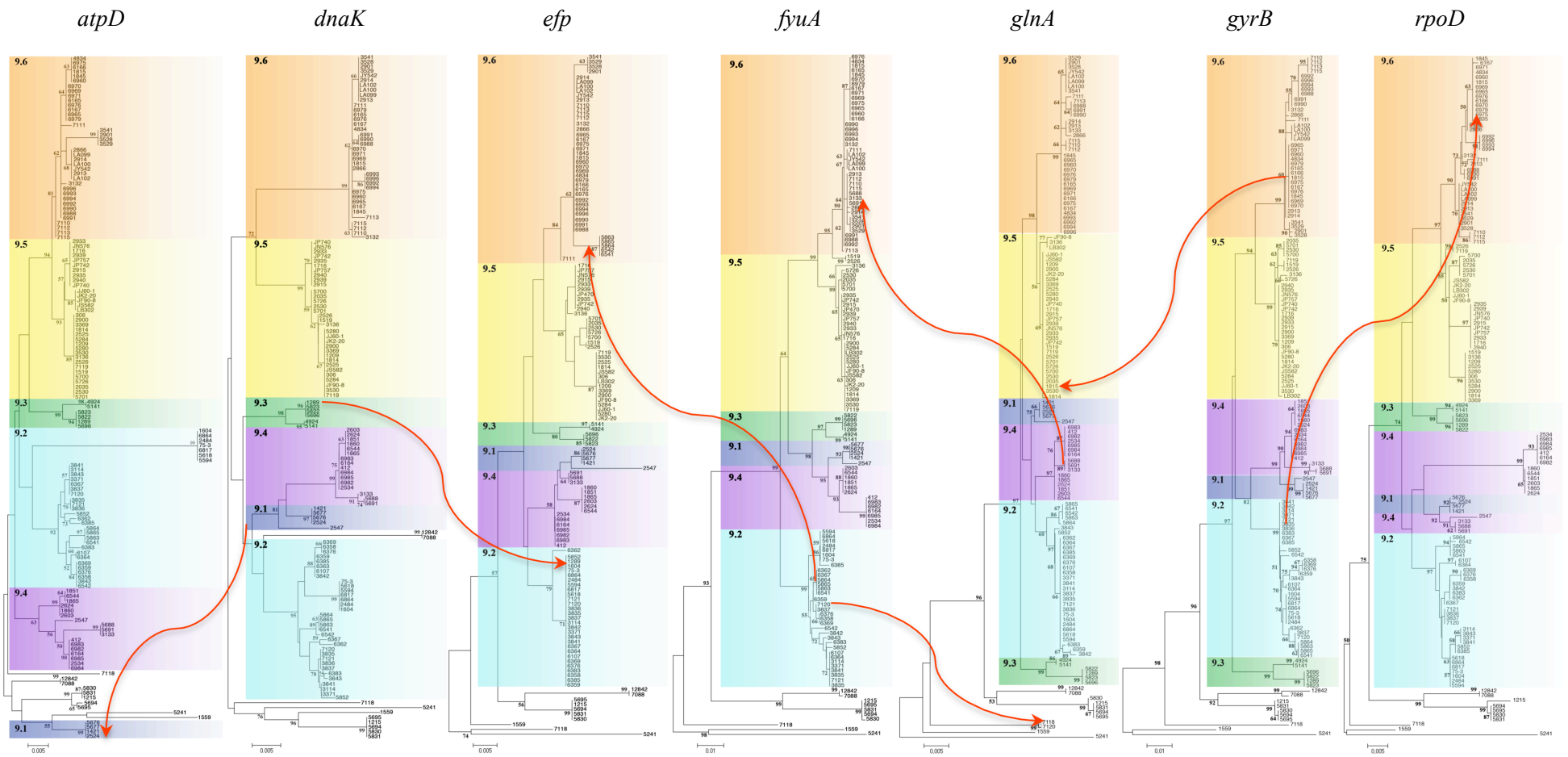


Figure 1. Arbres phylogénétiques construits par Neighbor Joining à partir des données des séquences partielles de sept gènes de ménage (*atpD*, *dnaK*, *efv*, *glnA*, *gyrB*, *fyvA* et *rpoD*) et indication des changements de positionnement de certaines souches illustrant des événements de recombinaison au sein de l'espèce *X. axonopodis*.

Un complexe clonal rassemble les souches dont les profils alléliques sont très proches *i.e* se différenciant par au maximum deux loci. Les singletons correspondent aux souches dont les profils alléliques diffèrent par plus de trois loci.

2.2.2 Résultats

L'agencement des groupes 9.1 à 9.6 n'est pas complètement conservé entre les différents arbres phylogénétiques. Ainsi, nous observons que pour chacun des sept arbres, la position de la branche du groupe 9.3 diffère (Figure 1, Annexe 6). En effet, pour les arbres construits à partir des alignements des séquences de *atpD*, *efp*, *fyuA* et *rpoD*, le nœud portant le groupe 9.3 porte également la branche regroupant 9.6 et 9.5. Par contre, pour l'arbre construit à partir des alignements des séquences de *dnaK*, le nœud portant le groupe 9.3 porte les branches des groupes 9.1, 9.2 et 9.4. Sur les arbres générés à partir des alignements de *glnA* et *gyrB*, le groupe 9.3 s'embranché au niveau du nœud portant les deux super clades rassemblant les groupes 9.5 et 9.6 et les groupes 9.1, 9.4 et 9.2. Par ailleurs, nous relevons que le positionnement des groupes 9.1 et 9.4 varie sur les arbres générés à partir de l'alignement des séquences de *atpD* et *fyuA* (Figure 1, Annexe 6). En effet, le groupe 9.1 s'embranché au niveau du nœud portant les souches qui se trouvent à l'extérieur du groupe formé par les souches de *X. axonopodis* sur l'arbre basé sur les séquences de *atpD*. Sur l'arbre basé sur les séquences de *fyuA*, les groupes 9.1, 9.4 et 9.3 s'embranchent au niveau du même nœud.

La grande majorité des pathovars sont monophylétiques *i.e* les souches d'un même pathovar sont regroupées au sein d'un même groupe phylogénétique (Annexe 6). Dans le cas contraire, les souches appartenant à un même pathovar peuvent être séparées et former des groupes polyphylétiques comme c'est le cas des pathovars *phaseoli* et *dieffenbachiae* (Annexe 6). Quelque soit le gène étudié, les souches du pathovar *phaseoli* sont toujours regroupées en quatre lignées génétiques ; l'une appartient au groupe 9.4, les trois autres lignées appartiennent au groupe 9.6. Les souches du pathovar *dieffenbachiae*, sont séparées selon leur hôte d'isolement. En effet, la souche isolée de *Dieffenbachia* (CFBP3132) se trouve dans le groupe 9.6 alors que les souches isolées de l'*Anthurium* (CFBP3133, CFBP5688 et CFBP5691) se trouvent

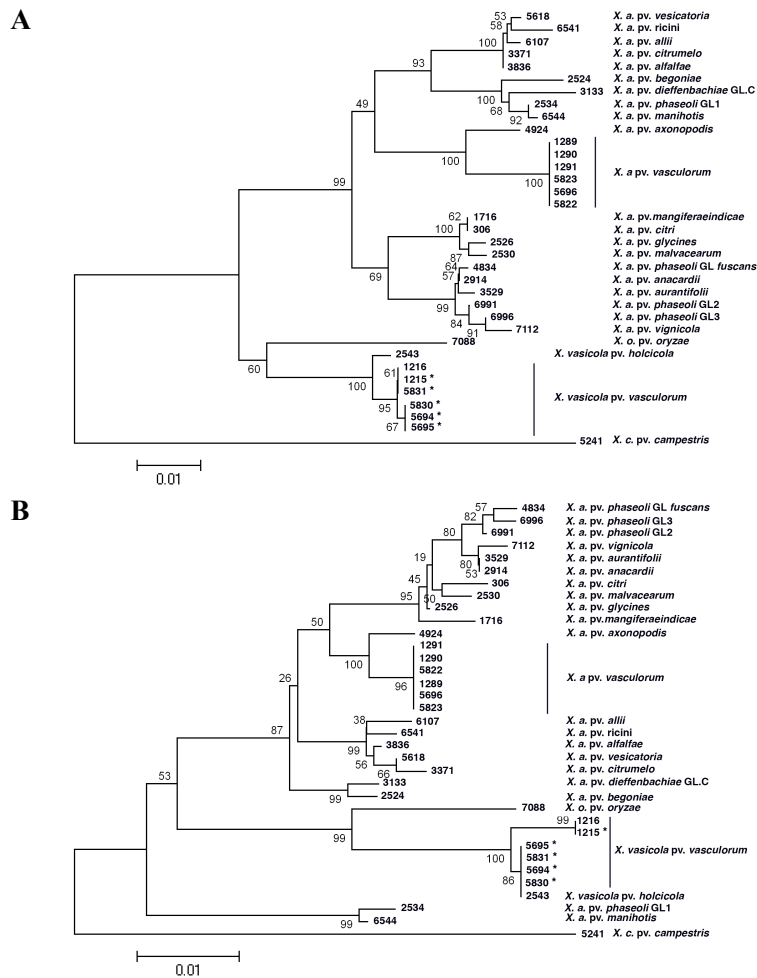


Figure 2. Positionnement de cinq souches mal identifiées dans les arbres phylogénétiques des *Xanthomonas* spp. Les arbres phylogénétiques obtenus par Neighbor Joining avec 1000 simulations à partir des séquences partielles des gènes *gyrB* (A) et *rpoD* (B) obtenus de 26 souches représentatives des pathovars de *X. axonopodis* (*X.a.*), de 2 souches de *X. vasicola* et de 1 souche de *X. oryzae* (*sensu* Vauterin *et al.*, 1995). Les arbres ont été enracinés avec la séquence partielle de la souche CFBP5241 de *X. campestris* pv. *campestris*.

Tableau 3. Probabilités associées aux résultats du test Shimodaira Hasegawa (SH) pour l'analyse de la congruence des topologies des arbres phylogénétiques construits à partir des alignements des sept gènes de ménage (*atpD*, *dnaK*, *efp*, *fyuA*, *glnA*, *gyrB* et *rpoD*) et des données concaténées. L'hypothèse testée stipule que les valeurs de vraisemblance associées aux topologies des arbres sont dans le même intervalle de confiance.

Arbre phylogénétique	<i>atpD</i>	<i>dnaK</i>	<i>efp</i>	<i>fyuA</i>	<i>glnA</i>	<i>gyrB</i>	<i>rpoD</i>	Données concaténées
<i>atpD</i>		0,004	0,003	0,000	0,000	0,000	0,000	0,000
<i>dnaK</i>	0,000		0,155	0,000	0,000	0,000	0,000	0,000
<i>efp</i>	0,000	0,000		0,000	0,000	0,000	0,000	0,000
<i>fyuA</i>	0,000	0,000	0,006		0,000	0,000	0,000	0,000
<i>glnA</i>	0,000	0,000	0,000	0,000		0,000	0,000	0,000
<i>gyrB</i>	0,001	0,000	0,000	0,000	0,000		0,000	0,000
<i>rpoD</i>	0,000	0,001	0,018	0,000	0,000	0,000		0,000
données concaténées	0,293	0,304	0,392	0,069	0,000	0,347	0,179	

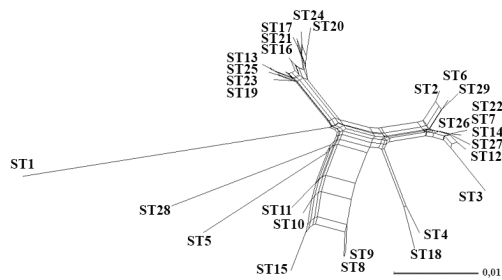
Sont indiqués en gras, les cas de rejet de l'hypothèse nulle.

dans le groupe 9.4. Les souches du pathovar *allii* peuvent être éloignées les unes des autres au sein du groupe 9.2. Elles ne forment jamais toutes ensemble un groupe cohésif (Annexe 7).

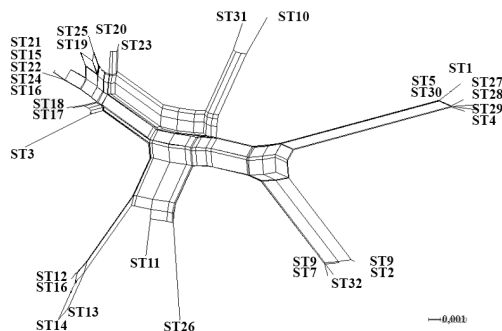
Par ailleurs, certaines souches initialement identifiées comme *X. axonopodis* pv. *glycines* ou *X. axonopodis* pv. *vasculorum* ne se regroupent jamais avec les autres souches appartenant au même pathovar (Annexe 6). De plus, ces souches n'appartiennent pas au groupe formé par l'ensemble des souches de *X. axonopodis* (Annexe 6). C'est le cas des souches CFBP1559 et CFBP7118 du pathovar *glycines* et des souches CFBP1215, CFBP5694, CFBP5696, CFBP5830 et CFBP5831 du pathovar *vasculorum*. Ces sept souches présentent la particularité d'être proches des souches externes : ATCC33913 (CFBP5241) de *X. campestris* pv. *campestris* et KACC10331 (CFBP7088) et MAFF311018 (CFBP12842) de *X. oryzae* pv. *oryzae*. Par comparaison de leurs séquences avec celles disponibles à la CFBP pour *gyrB* et *rpoD* pour l'ensemble des souches de *Xanthomonas* spp., les souches CFBP1215, CFBP5694, CFBP5696, CFBP5830 et CFBP5831 se regroupent avec les souches CFBP1216 et CFBP2543 respectivement identifiées comme *X. vasicola* pv. *vasculorum* et *X. vasicola* pv. *holcicola* (Figure 2). Par contre, les cinq autres souches identifiées comme *X. axonopodis* pv. *vasculorum* se regroupent bien avec la souche CFBP1290 qui est également identifiée comme *X. axonopodis* pv. *vasculorum* (Figure 2). En raison de leur non appartenance à l'espèce *X. axonopodis*, les sept souches externes ont été écartées pour la suite de l'étude (étude de la recombinaison et étude généalogique basée sur la coalescence).

Non seulement la topologie globale des arbres est assez variable entre les sept arbres phylogénétiques obtenus sur la base de chacun des gènes de ménage analysé, mais des incongruences majeures sont révélées pour plusieurs souches (Figure 1). Dix sept souches appartenant aux pathovars *begoniae*, *ricini*, *spondiae*, *vasculorum*, *dieffenbachiae*, *alfalfae* et *fuscans* ont une position très différente selon l'arbre considéré (Figure 1). Ces résultats suggèrent fortement des signaux phylogénétiques conflictuels qui sont probablement le résultat d'événements de recombinaison. Le test SH révèle que la valeur de vraisemblance associée à la topologie de chaque arbre

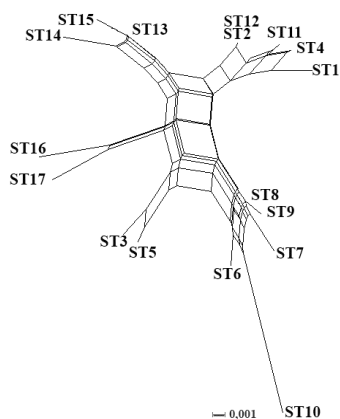
atpD



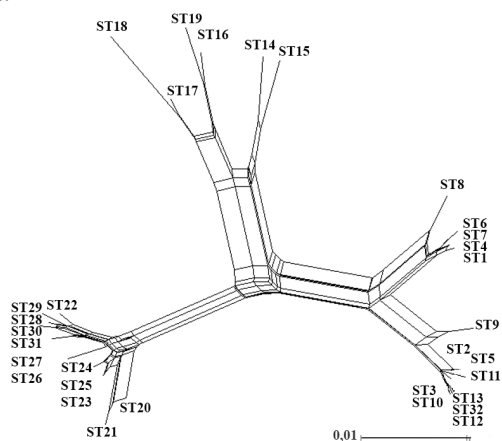
dnaK



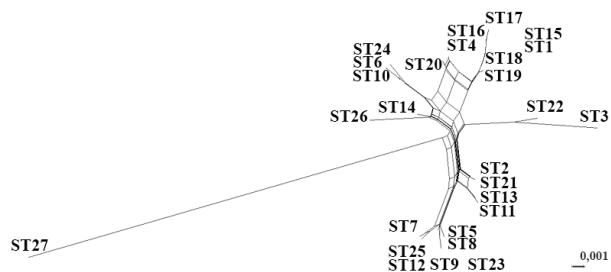
efp



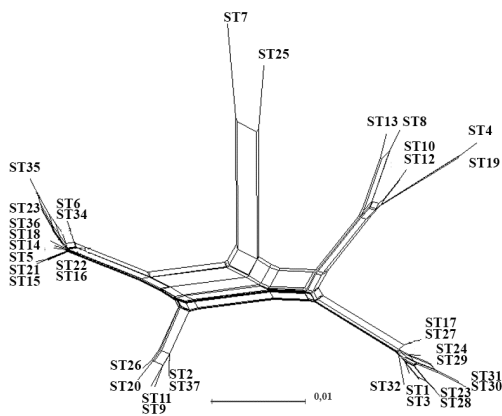
fyuA



glnA



gyrB



rpoD

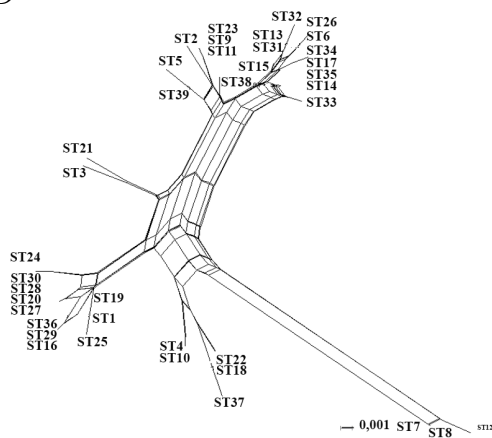


Figure 3, Réseaux obtenus avec la méthode NeighborNet implémentée dans le logiciel SplitsTree (Huson et Bryant, 2006) sur la base des types de séquences (ST) des sept gènes de ménage (*atpD*, *dnaK*, *efp*, *fyuA*, *glnA*, *gyrB* et *rpoD*). cf. Annexe 7 pour les souches et leurs STs correspondants.

phylogénétique est différente des valeurs de vraisemblance de chacun des autres arbres. Par contre, dans la majorité des cas, cette valeur n'est pas significativement différente de la valeur de vraisemblance associée à la topologie de l'arbre des données concaténées (Tableau 3).

L'analyse des séquences de chaque gène a permis de générer 29, 32, 18, 32, 27, 37 et 39 STs respectivement pour *atpD*, *dnaK*, *efp*, *fyuA*, *glnA*, *gyrB* et *rpoD* (Annexe 8). Parmi les STs générés, la majorité est spécifique d'un pathovar mais certains peuvent être partagés par de nombreux pathovars. Généralement, le pathovar *allii* détient le plus de STs (1 à 7 STs) pour chacun des gènes. De plus, il partage des STs avec des souches des pathovars *alfalfae*, *citrumelo* et/ou *vesicatoria* (Annexe 8).

Pour chacun des gènes étudiés, des réseaux d'haplotypes ont été obtenus. Tous les réseaux présentent de nombreuses arrêtes formant des parallélogrammes entre les STs. La présence de plusieurs arrêtes indique l'existence de nombreux signaux conflictuels qui peuvent être dus à des événements de recombinaison au sein de l'espèce *X. axonopodis* (Figure 3). Chaque réseau constitue une combinaison de plusieurs arbres phylogénétiques possibles. Dans certains réseaux, des branches présentant des parallélogrammes à leur base sont plus longues que les autres. Ceci traduit un degré de divergence plus élevé comparativement aux autres branches plus courtes.

Au moins, six des sept méthodes implémentées dans RDP ont permis de détecter 41 séquences recombinantes (Annexe 9). Cette analyse révèle un échange de matériel génétique respectivement entre les groupes 9.2 et 9.6, 9.4 et 9.6 et au sein des groupes 9.2 et 9.4 (Annexe 9). Un premier groupe comprend 24 séquences recombinantes dont les séquences de type parental correspondent à celles des souches CFBP6976 (*X. axonopodis* pv. *phaseoli* GL. *fuscans*) et CFBP5594 (*X. axonopodis* pv. *vesicatoria*). Dans ce groupe, la recombinaison se produit au niveau de différents sites, soit en position 324 bp et 5454 bp, affectant les gènes *atpD* et *rpoD* soit entre 1333 bp et 5331 bp affectant les gènes *efp* et *rpoD*. Les souches concernées appartiennent aux pathovars *citrumelo*, *alfalfae*, *betae*, *allii* et *ricini*. L'ensemble des souches testées d'un même pathovar est concerné par ces événements de recombinaison, sauf la souche CFBP6383 du pathovar *allii*. Seule une méthode (MaxChi) détecte le même événement de

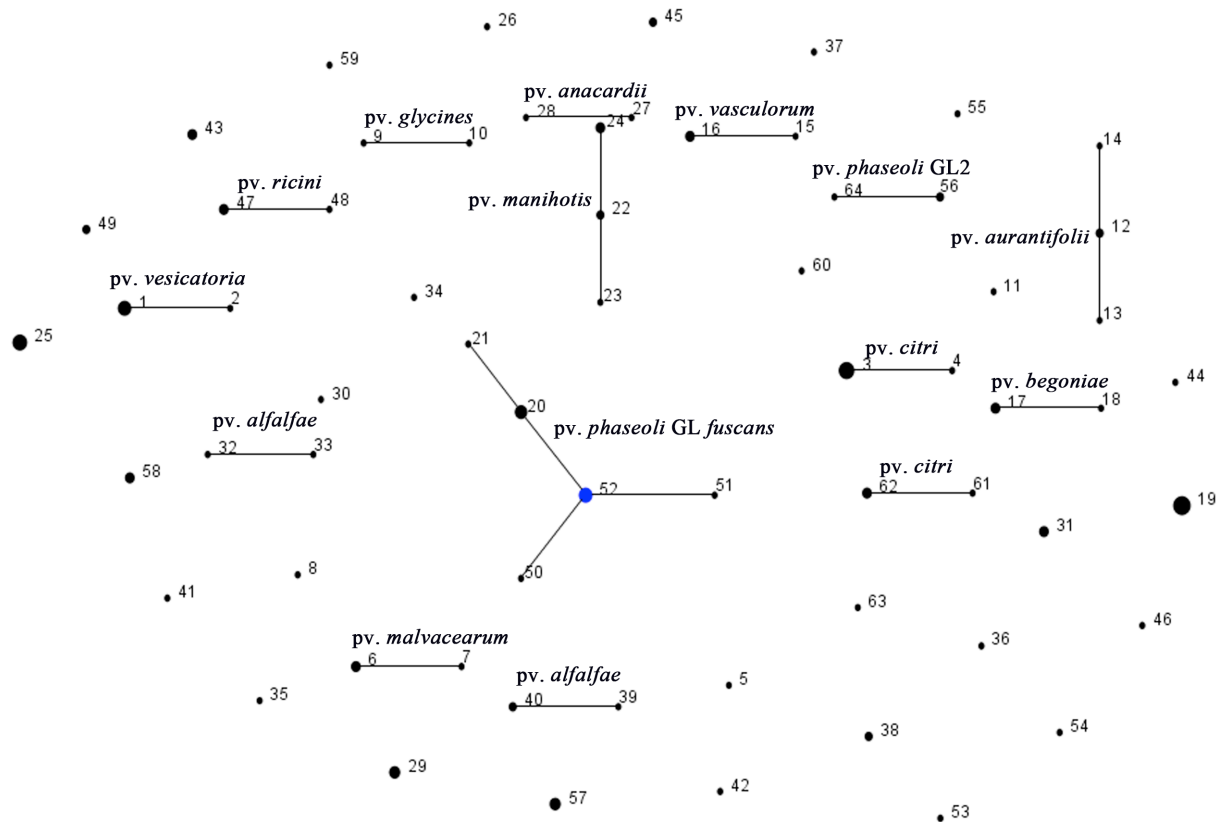


Figure 4. Représentation de la diversité de la population chez *X. axonopodis* par une analyse MLST. L'analyse a été réalisée par eBURST avec 64 profils alléliques représentant 131 souches de *X. axonopodis*. Chaque cercle représente un type de séquence (ST) unique, la taille du cercle est proportionnelle au nombre de souches possédant le ST, Les complexes clonaux sont représentés par les lignes connectant des STs qui diffèrent par au plus deux loci. Les cercles isolés correspondent à des singletons. cf. Annexe 7 pour les STs correspondant aux différents profils alléliques des souches.

recombinaison dans cette souche, cet événement n'est pas significatif avec les autres tests. Le deuxième groupe renferme 13 séquences recombinantes dont les séquences de type parental sont celles des souches CFBP5691 (*X. axonopodis* pv. *dieffenbachiae* GL.C) et CFBP2547 (*X. axonopodis* pv. *spondiae*). Dans ce groupe, la recombinaison a lieu au niveau des sites 4871 bp et 5343 bp n'affectant que le gène *rpoD*. Les souches concernées correspondent à l'ensemble des souches du pathovar *manihotis* et de la lignée génétique GL1 du pathovar *phaseoli*. Le troisième groupe comprend trois séquences recombinantes avec des séquences de type parental correspondant à celles des souches CFBP6546 (*X. axonopodis* pv. *phaseoli* GL1) et CFBP7110 (*X. axonopodis* pv. *vignicola*). La recombinaison se produit au niveau des sites 2043 bp et 2896 bp affectant le gène *fyuA*. Les souches concernées sont les trois souches de la lignée génétique GL.C du pathovar *dieffenbachiae*. Le dernier groupe ne renferme qu'une seule séquence recombinante entre les séquences de type parental des souches CFBP3837 (*X. axonopodis* pv. *alfalfae*) et CFBP5618 (*X. axonopodis* pv. *vesicatoria*). Dans ce cas, la recombinaison affecte les gènes *fyuA* et *gyrB* ; elle s'est produite au niveau des sites 2862 à 3958 bp dans la souche CFBP7120 du pathovar *alfalfae*.

Concernant l'analyse MLST, 62 profils alléliques (STs) ont été obtenus pour les 131 souches de *X. axonopodis* (Annexe 8). Parmi ces STs, 39 sont représentées par une seule souche. Le pathovar le plus polymorphe est *allii*, il renferme le plus de STs, soit sept (ST35, ST42, ST43, ST49, ST53, ST54 et ST55) (Annexe 8). Par contre, les souches du pathovar *mangiferaeindicae* sont toutes représentées par le même ST (ST19) (Annexe 8 et Figure 4). Quinze complexes clonaux et 29 singletons ont été identifiés (Figure 4). Les complexes clonaux sont formés à partir de STs représentant des souches des pathovars *vesicatoria* (ST1-ST2), *ricini* (ST47-ST48), *glycines* (ST9-ST10), *anacardii* (ST27-ST28), *manihotis* (ST24-ST22-ST23), *vasculorum* (ST15-ST16), *aurantifolii* (ST12-ST13-ST14), *begoniae* (ST17-ST18), *malvacearum* (ST6-ST7), *alfalfae* (ST32-ST33 et ST39-ST40) et *citri* (ST3-ST4 et ST61-ST62). Concernant le pathovar *citri*, les deux complexes clonaux correspondent aux deux pathotypes A et A*/A^w. D'autre part, les quatre lignées génétiques du pathovar *phaseoli* se distinguent également par leurs profils alléliques. La lignée *fuscans* et GL2 possèdent respectivement cinq et deux STs (ST21-ST20-ST52-ST51-ST50 et ST64-ST56) formant

des complexes clonaux distincts. Par contre, les lignées GL1 et GL3 possèdent chacune un ST unique (respectivement ST25 et ST57) commun à toutes les souches de la lignée génétique. Elles sont par conséquent chacune représentées par un singleton (Annexe 8). Les deux lignées génétiques (A et C) de *dieffenbachiae* sont représentées par trois STs (ST36, ST37 et ST38) formant trois singletons.

2.3 Article 2:

Evidence of recombination and migration among pathovars of *Xanthomonas axonopodis* species

Evidences of recombination and migration among pathovars of *Xanthomonas axonopodis*

NADIA MHEDBI-HAJRI¹, AHMED HAJRI¹, TRISTAN BOUREAU², ARMELLE DARRASSE¹, KARINE DURAND¹, CHRYSTELLE BRIN¹, MARION FISCHER-LE SAUX¹, CHARLES MANCEAU¹, STEPHANE POUSSIER³, OLIVIER PRUVOST⁴, CHRISTOPHE LEMAIRE² and MARIE-AGNES JACQUES¹.

¹ UMR077 PaVé, INRA, 42, rue Georges Morel, F-49071 Beaucouzé, France

² UMR077 PaVé, Université d'Angers, 42, rue Georges Morel, F-49071 Beaucouzé, France

³ UMR077 PaVé, Agrocampus Ouest centre d'Angers INHP, 42, rue Georges Morel, F-49071 Beaucouzé, France

⁴ UMR PVBMT, CIRAD F-97410 Saint Pierre, Cedex Réunion, France

Keywords: gene genealogy, population genetics, MLSA, pathogen, Host adaptation, MCP, adhesin, sensor, T3E

§Corresponding authors: Marie-Agnès Jacques, Fax +33 2 41 22 57 05, E-mail: marie-agnes.jacques@angers.inra.fr & Christophe Lemaire, Fax +33 2 41 22 57 05, E-mail: christophe.lemaire@univ-angers.fr

Running title: Evolutionary history of *Xanthomonas axonopodis*

Abstract

Deciphering mechanisms shaping bacterial diversity can provide insights into pathogen biology and in turn lead to predict the emergence of infectious diseases. Xanthomonads are worldwide distributed plant pathogenic bacteria. *X. axonopodis* is a genetically heterogeneous species grouping in six subgroups strains that are collectively pathogenic on a large number of plants. However, every strain belonging to this species has a very narrow host range. We aimed at examining the evolutionary, ecological and demographic processes that shape this diversity. We inferred relationships between strains using fragments of seven housekeeping genes on a collection of 131 strains encompassing the known sources of diversity in *X. axonopodis*. Using ClonalFrame and coalescent approaches we estimated parameters of isolation between genetic subgroups in *X. axonopodis*. We show that the six subgroups coalesced over a period of several millions of years. Recombination occurred as frequently as point mutation, but its impact was about three times more important than mutation in the diversity observed in the whole dataset. Genetic exchanges concerned most of the genetic subgroups. We could associate many recombination events to transfers of some virulence-associated genes *i.e.* sensors of chemotaxis and of two-component regulators, non-fibrillar adhesins and effectors of the type three secretion system. We argue that despite their long isolation the recent occurrence and the extent of recombination along with the agricultural intensification would not lead to speciation of these subgroups within *X. axonopodis* but would rather contribute to increase the gene flow among subgroups abrogating their limits.

Introduction

Delineating population unit boundaries remains a great challenge in evolutionary biology especially for micro-organisms. For bacteria, population units are more difficult to define than for most of the eukaryotes (Maynard Smith et al. 2000; Guttman & Stavriniades 2010; Balloux 2010). Clonality and horizontal gene transfer represent two major difficulties in the analysis of population structure of prokaryotes. Most of the models of population genetics assume sexual transmission of characters (Didelot & Falush 2007). However, the development of high-throughput DNA sequencing and phylogenetic methods based on the DNA sequence polymorphism such as MLST, MLSA and MLVA allowed the analysis of infraspecific variation during the last decade. These methods assign the clustering of several related strains in populations. Despite their high power of clustering, classical phylogenetic methods could not allow to infer population genetic parameters such as effective size, migration and recombination rates (Didelot 2010). Recombination is a key factor in the evolution of bacteria as it reflects parasexual events and then gene flow between groups of strains (Cohan 2002). Coalescent theory (Kingman 1982) presents the advantage to derive estimators of population parameters from gene genealogies, such as mutation or migration rates and to devise statistical tests of models of evolution (Rosenberg & Nordborg 2002; Drummond et al. 2005; Kuhner, 2008; Pybus & Rambaut 2009; Hey 2010). This approach considers all possible genealogies including those with recombination (*i.e.* ClonalFrame, Didelot & Falush 2007) and their probabilities under different models of evolution: demographic expansion *vs.* stability, and host subdivision (Rosenberg & Nordborg 2002).

The question of population structure parameters has been largely examined for human pathogenic bacteria for obvious clinical reasons using MLST and/or coalescent approaches (Moodley et al. 2009; Robinson et al. 2010). Such studies contributed remarkably to the understanding of the evolution and pathogenesis of bacterial pathogens such as vibrios (for review, see Bisharat 2010). Less attention has been paid to plant pathogenic bacteria. However, the social and environmental impacts of plant diseases are becoming great concerns as world population continues to increase and arable land and other natural resources still decrease.

Table 1 List of *X. axonopodis* strains used in the study

Genetic lineage (GL)	Strain code ^a	Host of isolation	Geographic origin	Year of isolation	Sub-group ^b
pv. <i>alfalfae</i>	3835	<i>Medicago sativa</i>	Australia	1972	9.2
	3836	<i>M. sativa</i>	Sudan	NA ^c	
	3837	<i>M. sativa</i>	USA	1965	
	7120	<i>M. sativa</i>	Japan	1962	
	7121	<i>M. sativa</i>	India	NA	
pv. <i>allii</i>	6107	<i>Allium fistulosum</i>	Japan	1998	9.2
	6358	<i>A. sativum</i>	Reunion Island	1994	
	6359	<i>A. cepa</i>	USA	1980	
	6362	<i>A. cepa</i>	Brazil	1986	
	6364	<i>A. sativum</i>	Cuba	1986	
	6367	<i>A. cepa</i>	Barbados	NA	
	6369	<i>A. cepa</i>	Reunion Island	1996	
	6376	<i>A. cepa</i>	Mauritius	1997	
	6383	<i>A. cepa</i>	USA	1983	
	6385	<i>A. cepa</i>	South Africa	NA	
pv. <i>anacardii</i>	2913	<i>Mangifera indica</i>	Brazil	NA	9.6
	2914	<i>M. indica</i>	Brazil	NA	
	7240	<i>Anacardium occidentale</i>	Brazil	2001	
	7241	<i>A. occidentale</i>	Brazil	2004	
	7242	<i>A. occidentale</i>	Brazil	2004	
	7243	<i>A. occidentale</i>	Brazil	2004	
pv. <i>aurantifolii</i>	2866	<i>Citrus aurantiifolia</i>	Brazil	1982	9.6
	2901	<i>C. limon</i>	Argentina	NA	
	3528	<i>C. limon</i>	Argentina	1988	
	3529	<i>C. limon</i>	Uruguay.	1983	
	3541	<i>C. aurantiifolia</i>	Mexico	NA	
pv. <i>axonopodis</i>	4924	<i>Axonopus scoparius</i>	Colombia	1949	9.3
	5141	<i>A. scoparius</i>	Colombia	1949	
pv. <i>begoniae</i>	1421	<i>Begonia</i> sp.	France	NA	9.1
	2524	<i>Begonia</i> sp.	New Zealand	1962	
	5676	<i>B. rugosa</i>	Antilles	1988	
	5677	<i>B. pendula</i>	France	1991	
pv. <i>betae</i>	5852	<i>Beta vulgaris.</i>	Brazil	1973	9.2
pv. <i>bilvae</i>	3136	<i>Aegle marmelos.</i>	India	1980	9.5
pv. <i>citri</i>	1209	<i>C. grandis</i>	Hong-Kong	1963	9.5
	1814	<i>Citrus</i> sp.	Reunion Island	1978	
	2525	<i>C. limon</i>	New Zealand	1956	
	2900	<i>Citrus</i> sp.	Japan	NA	
	3369	<i>C. aurantifolia</i>	USA	1989	
	3530	<i>C. limon</i>	Uruguay.	1984	
	5280	<i>C. hystrix</i>	Thailand	1998	
	5284	<i>Citrus</i> sp.	Malaysia	1999	
	JK2-20*	<i>C. aurantifolia</i>	Saudi Arabia	1988	
	JS582*	<i>C. aurantifolia</i>	Iran	1997	
	JJ60-1*	<i>C. aurantifolia</i>	India	1988	
	JF90-8*	<i>C. aurantifolia</i>	Oman	1988	
	LB302*	<i>C. aurantifolia</i> x <i>C. macrophylla</i>	USA	2002	
	306 **	NA	NA	NA	
	pv. <i>citrumelo</i>	3114	<i>Poncirus trifoliata</i> x <i>C. paradisi</i>	USA	
3371		NA	NA	1989	
3841		<i>P.trifoliata</i> x <i>C. sinensis</i>	USA	NA	

Xanthomonads are major plant pathogenic bacteria. Bacteria belonging to the genus *Xanthomonas* are collectively responsible for diseases on more than 400 different host plants, among which many economically important crops (Hayward 1993). Besides very homogeneous physiological characteristics, diversity among xanthomonads strains is expressed regarding host range. Individually, strains are characterized by a narrow host range and exhibit a marked tissue specificity infecting either the mesophyll parenchyma or the vascular system. Most species defined within the genus *Xanthomonas* encompass strains with limited host range whereas one species, *X. axonopodis*, comprises strains with different host ranges (Vauterin et al. 1995). Differentiation of species depends, by convention, on DNA-DNA hybridization studies, the application of threshold criteria (70% DNA-DNA reassociation and $\Delta T_m < 5^\circ\text{C}$) and availability of phenotypic properties (Wayne et al. 1987). Moreover, on the basis of several DNA-based methods this species, *X. axonopodis*, displays high infraspecific heterogeneity (Vauterin et al. 1995 and 2000; Rademaker et al. 2000 and 2005). Rep-PCR and AFLP genomic fingerprinting shows that *X. axonopodis* does not form a coherent group but comprises six subgroups named 9.1 to 9.6 (Rademaker et al. 2000 and 2005; Ah You et al. 2009). Within each of these six subgroups, numerous pathovars (infrasubspecific division that groups strains causing the same pathological facies on the same host range, Dye et al. 1980) are defined, making each subgroup pathologically heterogeneous. For example, the subgroup 9.6 groups strains that are pathogenic on at least *Citrus* sp., *Anacardium occidentale*, *Mangifera indica*, *Dieffenbachia* sp., *Phaseolus* sp. or *Vigna* sp. (Rademaker et al. 2000). There is no obvious ecological rationale to these clusterings in *X. axonopodis* since neither host range, tissue specificity, nor geographical origin seems to explain the classification of strains in the 6 subgroups. Conversely, the congeneric species *X. campestris* clusters all xanthomonads strains that are pathogenic on Brassicaceae (Fargier & Manceau 2007; Vauterin et al. 1995 and 2000).

In order to identify the processes leading to *X. axonopodis* subdivision in 6 subgroups, phylogeny based analyses were performed. Phylogenetic relatedness among strains was assessed by sequencing fragments of several housekeeping genes among which *atpD*, *dnaK*, *efp*, *gyrB*, and *rpoD* (Ah You et al. 2009; Bui Thi Ngoc et al. 2010; Hamza et al. 2010; Parkinson et al. 2007 and 2009; Young et al. 2008 and 2010).

Table 1 (continued)

	3842	<i>P. trifoliata</i> x <i>C. paradisi</i>	USA	NA	
	3843	<i>C. paradisi</i>	USA	NA	
pv. <i>dieffenbachiae</i> GL A	3132	<i>Dieffenbachia</i> sp.	USA	1950	9.6
pv. <i>dieffenbachiae</i> GL C	3133	<i>Anthurium</i> sp.	Brazil	1965	9.4
	5688	<i>A. andreanum</i>	Venezuela.	NA	
	5691	<i>Anthurium</i> sp.	Mauritius	NA	
pv. <i>glycines</i>	1519	<i>Glycine hispida</i>	Zimbabwe	1962	9.5
	2526	<i>G. hispida</i>	Sudan	1956	
	7119	<i>G. max</i>	Brazil	1981	
pv. <i>malvacearum</i>	2035	<i>Gossypium hirsutum</i>	Argentina	1981	9.5
	2530	<i>G. hirsutum</i>	Sudan	1958	
	5700	<i>G. hirsutum</i>	Senegal	1990	
	5701	<i>G. hirsutum</i>	Madagascar	1990	
	5726	<i>G. barbadense</i>	Sudan	1991	
pv. <i>mangiferaeindicae</i>	1716	<i>M. indica</i>	India	1957	9.5
	2915	<i>M. indica</i>	South Africa	1971	
	2933	<i>M. indica</i>	Reunion Island	1981	
	2935	<i>M. indica</i>	Australia	1978	
	2939	<i>Schinus terebenthifolius</i>	Reunion Island	1987	
	2940	<i>S. terebenthifolius</i>	Reunion Island	1987	
	7236	<i>M. indica</i>	Japan	1993	
	7237	<i>S. terebenthifolius</i>	Reunion Island	1994	
	7238	<i>S. terebenthifolius</i>	Reunion Island	1994	
	7239	<i>S. terebenthifolius</i>	Reunion Island	1994	
pv. <i>manihotis</i>	1851	<i>Manihot esculenta</i>	USA	NA	9.4
	1860	<i>M. esculenta</i>	Nigeria	1978	
	1865	<i>M. esculenta</i>	Congo	1977	
	2603	<i>M. esculenta</i>	Colombia	1972	
	2624	<i>M. esculenta</i>	Reunion Island	1986	
	6544	<i>M. esculenta</i>	Brazil	1992	
pv. <i>phaseoli</i> GL <i>fuscans</i>	1815	<i>Phaseolus</i> sp.	Greece	1978	9.6
	1845	<i>Phaseolus</i> sp.	Greece	1978	
	4834	<i>Phaseolus vulgaris</i>	France	1998	
	6165	<i>P. vulgaris</i>	Canada	1957	
	6166	<i>P. vulgaris</i>	South Africa	1963	
	6167	<i>Phaseolus</i> sp.	USA	1964	
	6960	<i>P. vulgaris</i>	Reunion Island	2000	
	6965	<i>P. vulgaris</i>	NA	NA	
	6969	<i>P. vulgaris</i>	Tanzania	2001	
	6970	<i>Phaseolus</i> sp.	USA	1990	
	6971	<i>Phaseolus</i> sp.	Tanzania	1992	
	6975	<i>Phaseolus</i> sp.	France	1994	
	6976	<i>Phaseolus</i> sp.	Czech Republic	1994	
	6979	<i>P. vulgaris</i>	Tanzania	2001	
pv. <i>phaseoli</i> GL1	412	<i>P. vulgaris</i>	USA	NA	9.4
	6164	<i>P. vulgaris</i>	Romania	1966	
	6546	<i>P. vulgaris</i>	USA	NA	
	6982	<i>P. vulgaris</i>	Reunion Island	2000	
	6983	<i>P. vulgaris</i>	Reunion Island	2000	
	6984	<i>P. vulgaris</i>	Reunion Island	2000	
	6985	<i>P. vulgaris</i>	Reunion Island	2000	
pv. <i>phaseoli</i> GL2	6988	<i>P. vulgaris</i>	Reunion Island	2000	9.6
	6990	<i>P. vulgaris</i>	Reunion Island	2000	

Results of these studies support the delineation in 6 subgroups within *X. axonopodis*. However, whether each or some subgroups should be elevated to the status of species remains still debated. For instance, Schaad and colleagues (2005, 2006 and 2007) and Jones and colleagues (2004 and 2006) elevated separate members of the 9.2 subgroup to the species rank whereas Young et al. (2008) considered they should remain in the same subgroup. For example, strains belonging to pv. *alfalfae* and pv. *citrumelo* have been proposed as separate subspecies in the newly created *X. alfalfae* species (Schaad et al. 2005 and 2006). Additionally Jones et al. (2006) proposed to elevate strains of the pv. *vesicatoria* into two new species: *X. euvesicatoria* and *X. perforans*. Status of the remaining strains from the 9.2 subgroups (*i.e.* from pv. *allii* and pv. *ricini*) was not studied. Another example lies in the proposition of Ah You et al. (2009) to pool 9.5 and 9.6 subgroups whereas Schaad et al. (2005 and 2006) elevated members of these subgroups into two separate species. In prokaryotes, speciation to be achieved needs sexual and/or ecological isolation and decline in recombination rate between differentiated genetic lineages (Cohan 2002; Costechareyre et al. 2009). Despite the clear clustering established by phylogenetic based analyses, parameters of isolation between the genetic subgroups in *X. axonopodis* still remain to be estimated. Divergence time and rates of genetic exchanges (recombination) are thus key parameters needed to state about the nature of these subgroups.

To decipher mechanisms leading to the emergence of infectious diseases, it is important to examine the genetic diversity of bacterial pathogens and the evolutionary, ecological and demographic processes that shape this diversity. This study was designed to gain insights into the evolutionary processes operating at the population-species interface of the *X. axonopodis* species. We used coalescent approaches to estimate parameters of isolation between genetic subgroups in *X. axonopodis*. First, we estimated the importance of recombination in the overall diversity of *X. axonopodis*. Second, divergence time and migration rates were inferred between genetic subgroups. We show that, despite a long time of isolation, genetic exchanges recently occurred between most of the genetic subgroups. These findings lead us to conclude that according to the biological species concept, species status seems not appropriate to these genetic subgroups.

Table 1 (continued)

	6991	<i>P. vulgaris</i>	Reunion Island	2000	
pv. <i>phaseoli</i> GL3	6992	<i>P. vulgaris</i>	Reunion Island	2000	9.6
	6994	<i>P. vulgaris</i>	Tanzania	1990	
	6996	<i>P. vulgaris</i>	Reunion Island	2000	
	6993	<i>P. vulgaris</i>	Reunion Island	2000	
pv. <i>ricini</i>	5863	<i>Ricinus communis</i>	Brazil	1981	9.2
	5864	<i>R. communis</i>	Brazil	1995	
	5865	<i>R. communis</i>	Brazil	1995	
	6541	<i>R. communis</i>	Brazil	1981	
	6542	<i>R. communis</i>	Brazil	1985	
pv. <i>spondiae</i>	2547	<i>Spondias dulcis</i>	Mauritius	1985	9.1
pv. <i>vasculorum</i>	1289	<i>Saccharum officinarum</i>	Reunion Island	1970	9.3
	5696	<i>Thysanolaena maxima</i>	Reunion Island	NA	
	5822	<i>S. officinarum</i>	Australia	1946	
	5823	<i>S. officinarum</i>	Mauritius	1979	
pv. <i>vesicatoria</i>	75-3**	<i>Lycopersicon esculentum</i>	NA	NA	9.2
	1604	<i>Capsicum annuum</i>	Guadeloupe	NA	
	2484	<i>L. esculentum</i>	Guadeloupe	1980	
	5594	<i>L. esculentum</i>	Guadeloupe	1993	
	5618	<i>C. annuum</i>	USA	NA	
	6817	NA	Thailand	1997	
	6864	<i>C. frutescens</i>	USA	1947	
pv. <i>vignicola</i>	7110	<i>Vigna unguiculata</i>	Zimbabwe	NA	9.6
	7111	<i>V. sinensis</i>	USA	1942	
	7112	<i>V. unguiculata</i>	USA	1942	
	7113	<i>V. unguiculata</i>	Sudan	1966	
	7115	<i>V. sinensis</i>	Brazil	1978	

^aStrain code in the French Collection of Plant Pathogenic Bacteria (CFBP), except * which were provided by O. Pruvost, Cirad, Reunion Island, France and **the reference strains whose genome are publicly available. All strains were provided by the CFBP.

^bRademaker et al. 2005.

^cNot available.

Material and methods

Bacterial strains and DNA extraction

Strains of *X. axonopodis sensu* Vauterin et al. (1995, 2000) selected for this study are listed in Table 1. On a pathological point of view, these 131 strains span over 21 pathovars. Most of these pathovars are monophyletic, *i.e.* composed of a unique genetic lineage. In contrast, pathovars *dieffenbachiae* and *phaseoli* are polyphyletic, *i. e.* pv. *dieffenbachiae* is composed of 2 distinct genetic lineages (GL A and GL C) (Rademaker et al. 2005; Hajri et al. 2009), whereas pv. *phaseoli* is composed of 4 genetic lineages GL 1 to GL 3, and GL fuscans (Alavi et al. 2008). Thus, on a phylogenetic point of view, our collection of 131 strains spans over 25 distinct genetic lineages. Strains were provided by the French collection of plant pathogenic bacteria (CFBP, Angers France, http://www.angers.inra.fr/cfbp/index_e.html). Strains were selected in order to maximize the diversity in terms of geographical origin, host and year of isolation and genetic diversity within pathovars. Each pathovar or genetic lineage is represented by at least 5 strains, except for the pathovars *betae*, *bilvae*, *spondiae*, and *axonopodis* for which only one or two strains were available. Cultures were stored in 40% glycerol at -80°C for long term preservation. They were checked for purity and routinely cultivated on YPGA (yeast extract, 7 g liter⁻¹; peptone, 7 g liter⁻¹; glucose, 7 g liter⁻¹; agar, 18 g liter⁻¹) for 2-4 days at 28°C.

Suspensions made from fresh cultures (overnight growth at 28°C under agitation in YP broth: yeast extract, 7 g liter⁻¹; peptone, 7 g liter⁻¹; pH 7.2) were used for DNA extraction using the standard hexadecyltrimethylammonium bromide method (Ausubel et al. 1992). Quality and quantity of DNA were spectrophotometrically evaluated (Nanodrop ND-100, Nanodrop Technologies). Then, the extracted DNAs were aliquoted and stored at -20°C in microtubes.

Gene amplification and sequencing

Primers for partial sequencing of seven housekeeping genes (*atpD*: ATP synthase beta chain, *dnaK*: encoding the 70-kDa heat shock protein, *efp*: elongation factor P, *fyuA* coding a transmembrane protein (Ton-B dependent transporter), *glnA*: glutamine synthetase I, *gyrB*: DNA gyrase subunit B, and *rpoD*: RNA polymerase sigma 70

Table 2 Sequences of primers used for PCR amplification of fragments of the seven genes, sizes of generated fragments, and corresponding annealing temperature

Locus	Sequence of forward (F) and reverse (R) primers (5' → 3')	Fragment length (bp)	Annealing temperature (°C)
<i>atpD</i>	F: GGGCAAGATCGTTCAGAT R: GCTCTTGGTCGAGGTGAT	868	60
<i>dnaK</i>	F: GGTATTGACCTCGGCACCAC R: ACCTTCGGCATAACGGGTCT	1034	60
<i>efp</i>	F: TCATCACCGAGACCGAATA R: TCCTGGTTGACGAACAGC	445	62
<i>fyuA</i>	F: ACCATCGACATGGACTGGACC R: GTCGCCGAACAGGTTACC	963	60
<i>glnA</i>	F: ATCAAGGACAACAAGGTCG R: GCGGTGAAGGTCAGGTAG	1094	60
<i>gyrB</i>	F: TGC GCGGCAAGATCCTCAAC R: GCGTTGTCCTCGATGAAGTC	1051	60
<i>rpoD</i>	F: ATGGCCAACGAACGTCCTGC R: AACTTGTAACCGCGACGGTATTCG	1313	60

factor) were designed (Table 2) from genomic *Xanthomonas* sequences available in GenBank: *X. axonopodis* pv. *vesicatoria* AM039952 and *X. axonopodis* pv. *citri* AE008923. PCR amplifications were performed in a 50 µl reaction mixture containing 1X Go Taq Buffer (Promega), 200 µM dNTP, 0.5 µM of each primer (Table 2), 0.4 U of Go Taq Polymerase and 3 ng of template genomic DNA in a Applied Biosystems thermocycler with an initial denaturation at 94°C for 2 min, 30 cycles of denaturation for 1 min at 94°C, annealing for 1 min 30 s at a gene specific temperature (Table 2), extension for 1 min 30 s at 72°C, and a final extension for 10 min at 72°C. Purity and yield of PCR products were checked by running 8 µl reaction mixture in 1.2% agarose gel and post-staining using ethidium bromide. The remaining amplified PCR products were sequenced with reverse and forward primers at Biogenouest platform (Nantes, France).

Sequence acquisition and alignment

Forward and reverse nucleotide sequences were edited and assembled by using PREGAP4 and GAP4 of the Staden Package (Staden 1996) then translated and aligned using CLUSTAL W-based subalignment tool with default parameters (Thompson et al. 1997) available in MEGA 4.0.2 program (Tamura et al. 2007) to obtain high quality sequences. Multiple alignments were manually edited using BIOEDIT program (Hall 1999). Amino acid alignments were transposed back to nucleotide sequence level to gain a codon-based alignment (Hall 1999). Sequences were concatenated following the alphabetic order of the genes ending in a sequence of 5466 bp (1-777 for *atpD*, 778-1674 for *dnaK*, 1675-2034 for *efp*, 2035-2898 for *fyuA*, 2899-3882 for *glnA*, 3883-4683 for *gyrB* and 4684-5466 for *rpoD*).

The partial sequences generated in this study are deposited in the GenBank/EMBL/DDBJ databases.

Statistical analyses

Occurrence of intragenic and intergenic recombination was analyzed using the four gametes test of Hudson and Kaplan (1985) implemented in DnaSP v.5 (Librado & Rosas 2009) and seven non-parametric detection programs implemented in RDP version 3.38 (Martin et al., 2005a): RDP (Martin & Rybicki 2000), GENECONV (Padidam et

al. 1999), MaxChi (Maynard Smith 1992), Chimaera (Posada & Crandall 2001), BootScan (Martin et al. 2005b), SiScan (Gibbs et al. 2000), and 3Seq (Boni et al. 2007). The analysis was performed with default settings for the different detection methods and the Bonferroni-corrected P-value cut-off was set at 0.05. Recombination events were accepted when detected with at least three detection methods out of seven. The Web-based service GARD (genetic algorithm for recombination detection) was also used to detect and locate recombination breakpoints (Kasakovskiy Pond et al. 2006). All the sequences of the 131 strains of *X. axonopodis* were analyzed under a coalescent framework, in order to take into account all the nucleotidic variation between sequences (Didelot & Falush 2007) and to permit the inference of genetic parameters between populations (Kuhner 2008; Hey 2010). Populations inferred from the ClonalFrame analysis did match the 6 rep-PCR subgroups (9.1 to 9.6). These groups define monophyletic units for which parameters such as effective size, migration rates or Time to Most Recent Common Ancestor (TMRCA) make sense. Haplotype number (Hap), haplotype diversity (Hd , Nei 1987), nucleotidic diversities (θ_π , Nei 1987, and θ_w , Watterson 1975), Neutrality tests (Tajima's D, Tajima 1989, and Fu and Li's D* and F*, Fu & Li 1993) were estimated for each of the seven genes for each population using DnaSP. Bonferroni's correction for multiple tests (Rice 1989) was applied.

Clonal genealogy of the 131 strains was estimated using ClonalFrame (Didelot & Falush 2007). ClonalFrame is well suited for bacterial genealogy as it estimates clonal relationships between sampled strains and estimates both mutations and positions of homolog recombination events in a neutral coalescent model. Parameters are estimated using a Bayesian framework. Parameter space was explored using a Markov Chain Monte Carlo (MCMC) simulation of 100,000 iterations, of which 30,000 were considered as burn-in. Interval of genealogy sampling was set to 50, which ensured good independence between successive sampling. A 50% consensus tree was produced. Parameters estimated were the mutational rate θ , the intragenic recombination rate R , the average length of a recombination event δ , the rate ν of new polymorphism introduced by recombination and the TMRCA. Two additional statistics r/m and ρ/θ represent a measure of the impact of the recombination (r) relative to mutation (m) in the diversity of the sample, and a measure of the rate of recombination per site (ρ)

relative to the mutation rate (θ), respectively. A first analysis was performed on the global dataset. Additional analyses were performed on each population dataset.

Migration rates and divergence time were estimated using the Isolation-with-Migration model of Hey implemented in IMA2 (Hey 2010). In this model, an ancestral population of size N_A splits at a time t into two populations of respective sizes N_1 and N_2 , which may occasionally exchange a small number of migrants after divergence. IMA2 provides the opportunity to analyze divergence times, effective sizes and migration rates for more than two populations. The Isolation-with-Migration model not only permits estimates of migration rates but also direction and age of migration events since divergence. Relationships between all populations need to be provided in order to represent ancestors of populations. Genealogy given by ClonalFrame permits to set ancestral nodes for ancestral effective size and divergence time estimations. Our dataset violates the assumption of no recombination within loci. We used the four gametes-test implemented in DnaSP (Librado & Rosas 2009) and identified block undergoing recombination. These blocks were discarded and the remaining dataset was divided into a set of 94 non-recombinant loci. IMA2 uses a Bayesian framework in which user should define prior input values of parameters. Priors for θ parameter within each population are provided by estimates obtained using ClonalFrame. Parameter space is explored by a heuristic MCMC exploration and posterior distributions of each parameter are then generated. We used 30,000,000 MCMC iterations with a genealogy sampling performed every 100 iterations. The first 1,000,000 iterations were discarded as burn-in. Mixing and convergence were obtained by running 100 Metropolis coupled-chains with 100 chain swapping attempts between iterations. Number of migrants between populations per generation was obtained using m estimated by IMA2 and θ estimated by ClonalFrame as it assumes clonal inheritance. A mutation rate of 10^{-8} per site per year was used as reference for demographic estimation of parameters (Guttman & Dykhuizen 1994) and generation times of 0.003, 1 and 10 years were used.

Inference of repertoires of virulence associated (VA) genes at ancestor nodes.

Repertoires of 107 genes coding different VA proteins (Methyl accepting Chemotaxis Proteins MCPs, sensory proteins of two component regulatory systems, adhesins and Type III effectors T3Es) were determined for each strain of this study by PCR and Dot

Table 3 Descriptive statistics for the overall polymorphism in *X. axonopodis* inferred from the analysis of seven genes from 131 strains

Locus	Sites ^a	Hap ^b	Hd ^c	S ^d	θ_{π} ^e	θ_w ^f	GC%	Tajima's D ^g	Fu & Li's D* ^g	Fu & Li's F* ^g
<i>atpD</i>	777	29	0.951	103	0.02226	0.02504	65.34	-0.3564	1.5092	0.8210
<i>dnaK</i>	897	32	0.948	88	0.02469	0.01944	63.15	0.8652	0.6533	0.8958
<i>efp</i>	360	18	0.867	25	0.01563	0.01491	63.17	0.1431	1.2266	0.9525
<i>fyuA</i>	864	32	0.950	111	0.02925	0.02481	64.39	0.5769	0.6556	0.7445
<i>glnA</i>	984	27	0.894	85	0.01050	0.01679	63.62	-1.1970	-3.9457** ^h	-3.2901
<i>gyrB</i>	801	37	0.941	115	0.03014	0.02864	65.64	0.1689	1.3641	0.9969
<i>rpoD</i>	783	34	0.959	84	0.02215	0.02196	64.34	0.0284	1.1777	0.8083
Total	5466	62	0.978	611	0.02235	0.02202	64.28	0.0493	0.4488	0.3146

^aNumber of analyzed sites.

^bNumber of haplotypes.

^cHaplotype diversity (Nei 1987).

^dNumber of polymorphic sites.

^eNucleotide diversity (Nei 1987).

^fNucleotide diversity from S (Watterson 1975).

^gNeutrality tests: Tajima's D (Tajima 1989) and Fu & Li's D* and F* (Fu & Li 1993). ^hAssociated P-value for neutrality tests: ** $p < 0.01$. Fu & Li's D* on *glnA* did not remain significant after a correction of Bonferroni for multiple tests was applied with threshold $K=0.05/8 = 0.00625$.

Table 4 Descriptive statistics for the overall polymorphism in the pathovars of *X. axonopodis* inferred from the analysis of seven genes from 131 strains. Only pathovars represented by more than one strain are indicated. Clade refers to genetic subgroups 9.1 to 9.6.

Clade	Pathovar name	n ^a	S ^b	Hap ^c	Hd ^d	θ_{π} ^e	θ_w ^f	TajimaD ^g	FuLiD* ^g	FuLiF* ^g
9.1	<i>begoniae</i>	4	1	2	0.50	0.000	0.000	-0.61	-0.61	-0.48
	<i>citrumelo</i>	5	33	3	0.70	0.002	0.003	-1.25	-1.25	-1.35
	<i>ricini</i>	5	24	3	0.70	0.002	0.002	-1.24	-1.24	-1.33
9.2	<i>vesicatoria</i>	7	0	1	0.00	0.000	n a ^h	n c ⁱ	n c	n c
	<i>alfalfae</i>	5	76	4	0.90	0.006	0.007	-0.64	-0.64	-0.70
	<i>allii</i>	10	64	7	0.91	0.004	0.004	0.15	0.16	0.18
9.3	<i>axonopodis</i>	2	0	1	0.00	0.000	n a	n c	n c	n c
	<i>vasculorum</i>	4	7	2	0.50	0.001	0.001	-0.82	-0.82	-0.80
	<i>diffenbachiae</i>	3	3	2	0.67	0.000	0.000	n c	n c	n c
9.4	<i>manihotis</i>	6	1	2	0.60	0.000	0.000	1.45	1.05	1.16
	<i>phaseoli GL1</i>	7	0	1	0.00	0.000	n a	n c	n c	n c
	<i>citri</i>	14	12	5	0.66	0.001	0.001	0.68	0.38	0.53
9.5	<i>glycines</i>	3	24	3	1.00	0.003	0.003	n c	n c	n c
	<i>malvacearum</i>	5	6	3	0.70	0.000	0.001	-1.15	-1.15	-1.19
	<i>mangiferaeindicae</i>	10	0	1	0.00	0.000	n a	n c	n c	n c
	<i>phaseoli fuscans</i>	14	9	5	0.73	0.000	0.001	-1.57	-2.27	-2.38
9.6	<i>phaseoli GL2</i>	3	2	2	0.67	0.000	0.000	n c	n c	n c
	<i>phaseoli GL3</i>	4	0	1	0.00	0.000	n a	n c	n c	n c
	<i>vignicola</i>	5	26	3	0.70	0.002	0.002	-0.17	-0.17	-0.18
	<i>anacardii</i>	6	7	3	0.60	0.001	0.001	0.89	1.15	1.17
	<i>aurantifolii</i>	5	22	4	0.90	0.002	0.002	-1.24	-1.24	-1.33

^aNumber of strains.

^bNumber of polymorphic sites.

^cNumber of haplotypes.

^dHaplotype diversity, Nei 1987.

^eNucleotide diversity, Nei 1987.

^fNucleotide diversity from S, Watterson 1975.

^gNeutrality tests: Tajima's D (Tajima 1989) and Fu & Li's D* and F* (Fu & Li 1993).

^hNon available

ⁱNot computable

Blot experiments in previous studies by Hajri et al. (2009) and Mhedbi-Hajri et al. (submitted). Using the Mesquite 2.5 system for phylogenetic computing (Maddison & Maddison 2009), the matrix of presence/absence of these genes for each strain was associated to the scenario for genealogy obtained in ClonalFrame. Repertoires of genes at nodes of the ClonalFrame genealogy were inferred by parsimony using the function Trace Character History implemented in Mesquite.

We explored the possibility that recombination events identified by IMA2 on the basis of seven housekeeping gene analysis may have occurred with a concomitant transfer of some genes belonging to the repertoires of VA genes. Indeed, for each recombination events between strains identified by IMA2, we assigned a node considered as donor and a node considered as recipient. We considered that a character may have been transferred when it is present in the donor, absent in the recipient but present in descendants of the recipient node.

Results

Overall polymorphism

One hundred and thirty one strains of *X. axonopodis* representing the 25 genetic lineages were sequenced at seven housekeeping loci. Analyzed sequence lengths ranged from 360 bp (*efp*) to 984 bp (*glnA*), leading to a total of 5466 bp for the sequence of concatenated data set (Table S1, Supporting Information). The G+C content for all loci ranged from 63.15% (*dnaK*) to 65.64% (*gyrB*). Loci had between 18 (*efp*) and 37 (*gyrB*) alleles. The *gyrB* and *rpoD* genes were the most polymorphic. The overall haplotype diversity was 0.978 (Table 3). The nucleotide θ_s and Watterson θ_w diversity estimators among all genes were 0.0223 and 0.0220, respectively. Tajima's D and Fu and Li's D* and F* tests indicated that no departures from mutation/drift equilibrium were detected on the overall dataset. However, significant departure was observed for *glnA* with the Fu and Li's D* test but not with other tests (Tajima's D and Fu and Li's F* tests). Nevertheless no test remained significant after Bonferroni's correction for multiple tests with threshold $K=0.00625$.

A set of 62 haplotypes was obtained for the *X. axonopodis* strains (Table 3, Table S1, Supporting Information). Each haplotype belonged to a single pathovar. The

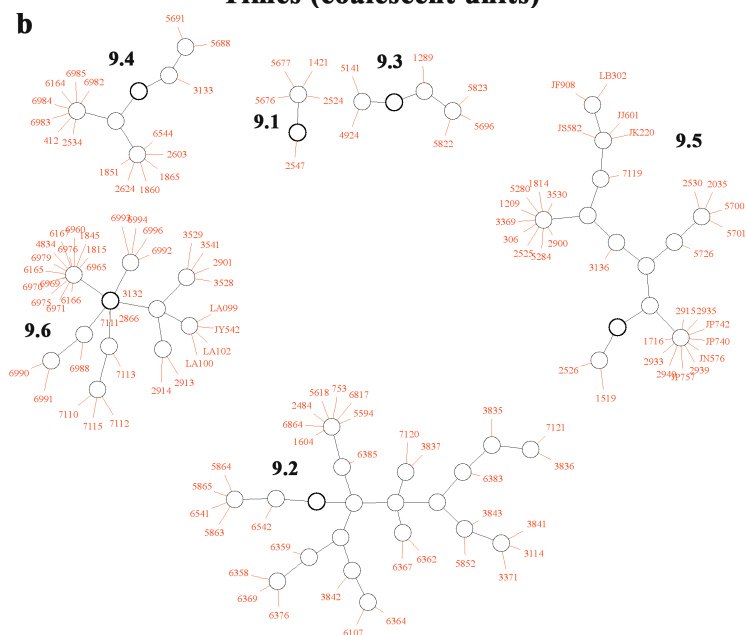
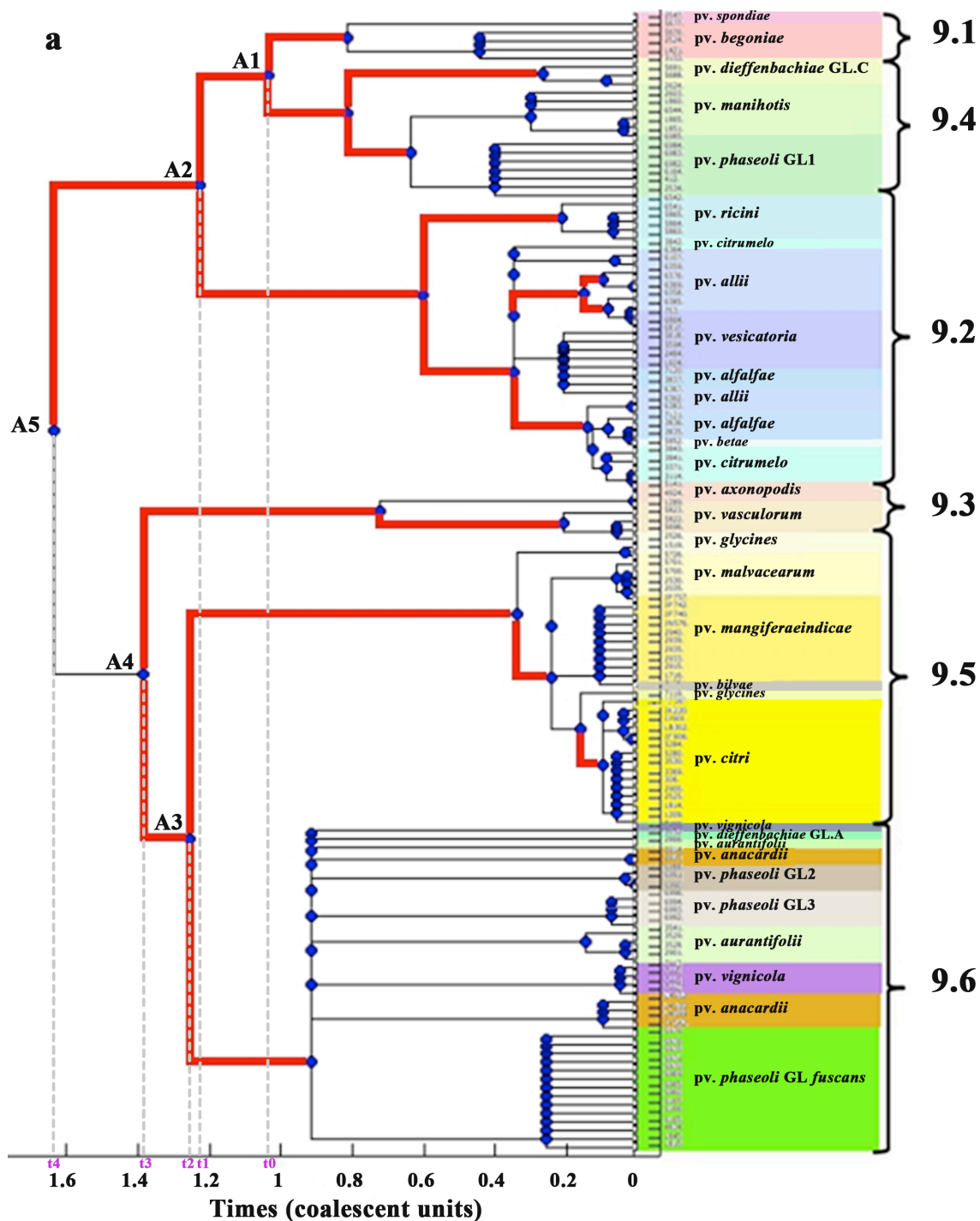


Fig. 1 Genealogy inferred by ClonalFrame. a) Majority-rule consensus genealogy. Captions 9.1 to 9.6 refer to genetic subgroups within *X. axonopodis*. Populations A1 to A5 represent ancestral populations. Red branches indicates significant occurrence of recombination ($p > 0.9$). Despite being grouped in a same rake strains may contain different closely related haplotypes. Parameters were estimated from the sampling of 100 000 iterations of 131 strains typed on 94 non-recombinant loci. b) Dot tree obtained from the genealogy. Limit of node age was set to 1.2 in coalescence unit.

number of haplotype per pathovar ranged from 1 to 7 (Table 4). No obvious trend between sample size and level of polymorphism was observed. Pathovars *vesicatoria*, *axonopodis* and *mangiferaeindicae* were monomorphic despite attempts were made to maximize the diversity of their host and geographical origins. No representatives of the *X. perforans* species were included in this study. Conversely, pathovars *alfalfae*, *allii* and *glycines* presented a high polymorphism evaluated by the number of haplotypes, (respectively 4, 7 and 3) the haplotype diversity (respectively 0.90, 0.91 and 1.00) as well as the high nucleotidic diversity θ_x (respectively 0.006, 0.004 and 0.003). The nucleotidic diversity θ is null for 11 out of 21 pathovars. No departure from mutation/drift equilibrium was observed at neutrality tests.

In order to screen for evidences of recombination within *X. axonopodis*, three different methods were employed: the web server GARD, RDP and DnaSP programs. All tested methods detected recombination signals in the concatenated data set. The web-based service GARD found breakpoints at positions 726 (*atpD*), 2028 (*efp*), 2895 (*fyuA*), 3897 (*gyrB*) and 4725 (*rpoD*) in the concatenated sequence. Six of the seven detection programs implemented in RDP3 program found four unique events and 41 recombinant sequences. These recombination events affected strains from pathovars *alfalfae*, *allii*, *betae*, *citrumelo*, *dieffenbachiae GL C*, *manihotis*, *phaseoli GL1*, and *ricini*. The DnaSP program estimated the minimum number of recombination events to 91 for the concatenated data set. Analysis of recombination on loci individually gave different results according to the program used. Indeed the web-based service GARD detected recombination signals for *atpD* and *fyuA* sequences, and found breakpoints at position 360 (to 777 bp) and 485 (to 864 bp) respectively. However DnaSP detected recombination for all the loci and the minimum number of recombination events was estimated to 12 for *atpD* locus, 16 for *dnaK* locus, 4 for *efp* locus, 18 for *fyuA* locus, 7 for *glnA* locus, 18 for *gyrB* locus and 12 for *rpoD* locus.

Coalescence-based analysis using clonal inference

Clustering of strains into 6 subgroups corresponding to rep-PCR subgroups (Rademaker et al. 2005) was strongly supported by the majority-rule genealogy inferred from ClonalFrame analysis (Fig. 1). The genealogy produced indicated a first divergence event that may have separated subgroups 9.1, 9.2 and 9.4 on one side and subgroups

Table 5 ClonalFrame parameters for the six clades and the whole *X. axonopodis* dataset. Clade refers to genetic subgroups 9.1 to 9.6

Parameter ^a	<i>X. axonopodis</i> clade						Whole set
	9.1	9.2	9.3	9.4	9.5	9.6	
n	5	33	6	16	33	38	131
θ	0.009		0.21	0.24	0.368		
	[0.001-0.057]	13.13	[0.002-1.282]	[0.006-0.903]	[0.002-1.705]	7.21	14.10
ν	0.0111	[4.52-25.79]	0.009	0.02	0.005	[1.33-19.98]	[6.67-22.21]
	[0.008-0.015]	[0.015-0.028]	[0.006-0.013]	[0.018-0.036]	[0.003-0.007]	[0.001-0.004]	[0.008-0.011]
R	1.868		0.23				
	[0.009-7.225]	4.56	[0.002-1.743]	1.09	3.43	2.21	12.43
TMRCA	1.053	[2.46-7.93]	2.296	[0.57-1.83]	[1.614-6.01]	[0.123-6.55]	[9.12-15.59]
	[0.253-2.758]	1.98	[0.746-4.335]	5.72	1.74	0.97	1.67
		[0.82-4.19]		[2.83-10.58]	[0.89-3.42]	[0.52-1.8]	[1.23-2.17]

^an: number of strains, θ: number of mutation events confidence interval, ν: rate of substitution *via* recombination confidence interval, R: the number of recombination events confidence interval, and TMRCA: the estimate of time to most recent common ancestor. Values in square brackets figure the confidence interval.

9.3, 9.5 and 9.6 on the other side. In the former group, clade 9.2 seems to have diverged first, and then 9.1 and 9.4. In the latter group, 9.3 diverged first and then 9.5 and 9.6 (Fig. 1). However, branching of 9.3 with 9.5 and 9.6 subgroups was less robust than other nodes (data not shown). More strict consensus (*i.e.* with valid nodes that appeared more than 60% in the posterior genealogy distribution) genealogies put the subgroup 9.3 at the same level than the common ancestor, whereas other nodes remained valid. Subgroups 9.1 and 9.4 have diverged the most recently followed (backward in time) by the divergence between subgroup 9.2 and the ancestor of subgroups 9.1 and 9.4 (A1 population). Furthermore, divergence times between subgroups 9.5 and 9.6, and then 9.3 were older.

Within pathovars, strains may be grouped together in monophyletic groups or separated in several phylogenetically distant clusters (Fig. 1). Pathovar *dieffenbachiae* and pathovar *phaseoli* displayed two and four genetic lineages, respectively. These genetic lineages belonged to subgroups 9.4 and 9.6. Within a subgroup, some pathovars coalesced separately as it is the case for pathovars *allii* (four groups), *alfalfae* (two groups), *citrumelo* (two groups), *glycines* (two groups), *anacardii* (two groups), and *aurantifolii* (two groups).

Analysis of the diversity was performed within each of the 6 subgroups at the seven loci and in the concatenated sequences. Analysis of polymorphism showed a number of haplotypes ranging from 3 (9.1 and 9.3 subgroups) to 19 (9.2 and 9.6 subgroups) (Table S1, Supporting Information). Regarding nucleotidic diversity, subgroup 9.2 had the highest estimate whereas subgroup 9.6 had the lowest one. Significant departures from mutation/drift were detected with the three neutrality tests in five of the six clades. However, these tests did not remain significant after Bonferroni correction for multiple tests for each clade with threshold $K=0.00625$.

ClonalFrame includes the effect of recombination in the reconstruction of the genealogy. Significant recombination was found in the coalescents of the whole dataset ($R=12.43$) and of the six clades separately (Table 5). Recombination is found to occur as frequently as point mutation in the history of the 131 strains ($\rho/\theta=0.97$ [0.42-1.84]). The mean tract length of the recombinant fragment (δ) was 554 [403-742] bp. However impact of recombination is about three times more important than mutation in the diversity observed in the whole dataset ($r/m=3.18$ [1.71-5.38]). Due to a lack of

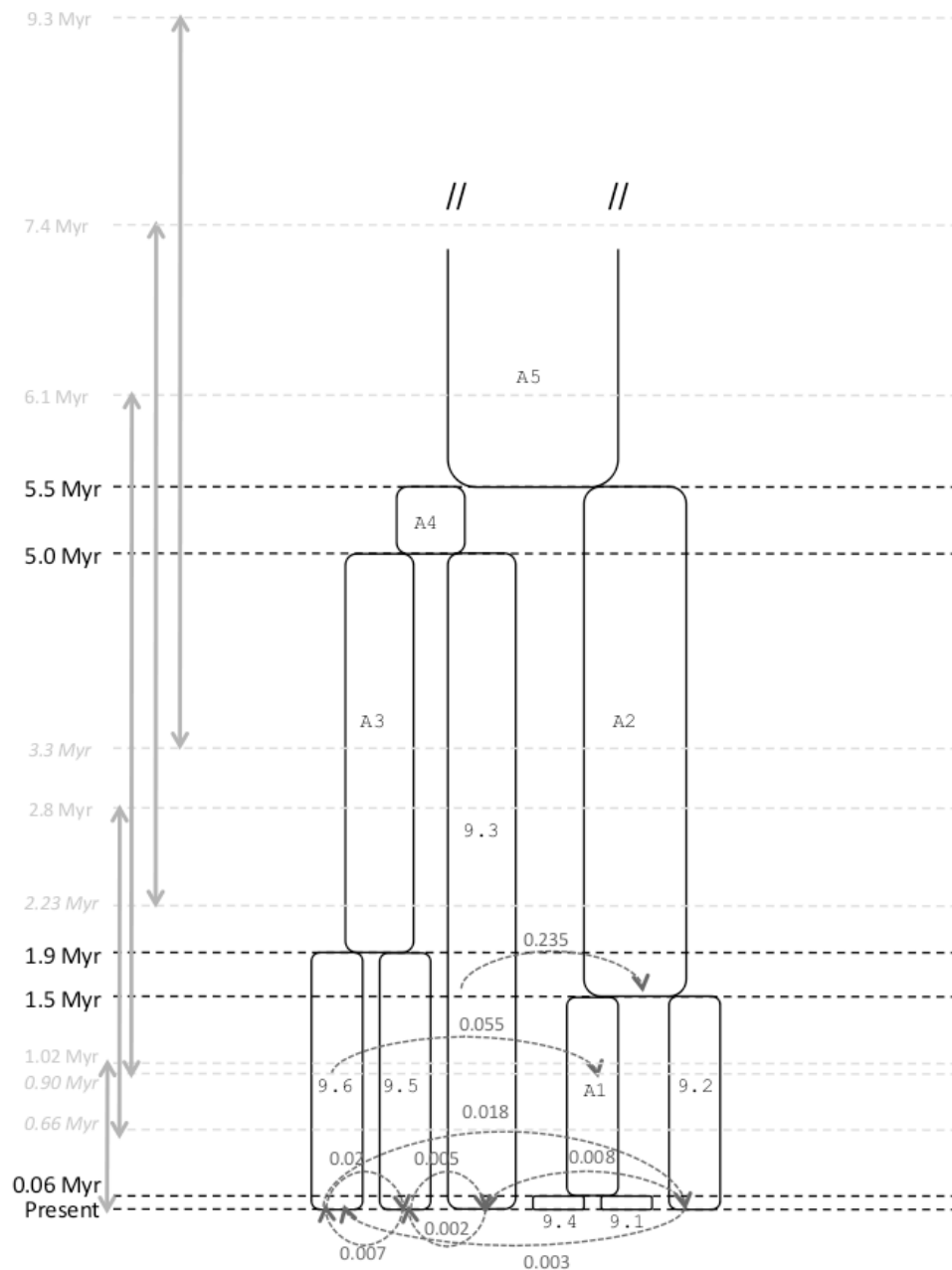


Fig. 2 Evolutionary history set in IMA2 of 131 strains of *X. axonopodis* belonging to 25 genetic lineages. Divergence time estimates are given in Myr at the left (in black). Plain grey double arrows indicate 95% Interval Confidence. Values in italic and normal grey respectively indicate lower and upper limits of 95% highest posterior density on divergence times. Directions of migration are represented by dotted grey arrows, numerical values represent the effective number of migrants. Only migrations significantly different from zero are represented ($p < 0.05$). Populations 9.1 to 9.6 refer to genetic subgroups in *X. axonopodis* and A1 to A5 refer to ancestral populations.

convergence of the MCMC sampling process parameters δ , r/m and ρ/θ for each of the six clades individually were not reported. Rate of new polymorphism introduced by recombination is about 1% for the whole dataset and ranges from 0.003 in clade 9.6 to 0.02 in clade 9.2 (Table 5). Recombination likely (*i.e.* $p>90\%$) occurred in some branches (Fig. 1).

Estimation of parameters under Isolation-with-Migration model

The Isolation-with-Migration model was used to jointly estimate effective size of contemporary and ancestral populations, divergence times and migration rates between populations. An evolutionary scenario was derived from the genealogical analysis of our strains under ClonalFrame. This scenario (Fig. 1) was provided to IMA2 for parameter estimation (Fig. 2 and Table S2, Supporting Information). Divergence times range from 0.06 [Present-1.02] millions of years (Myr) for subgroups 9.1 and 9.4, to 5.5 [3.3 – 6.3] Myr for the Most Recent Common Ancestor (MRCA). Subgroups 9.5 and 9.6 diverged 1.9 [0.9 – 6.1] Myr ago, and subgroup 9.3 diverged from A3 population (common ancestor of 9.5 and 9.6) 5 [2.2 – 7.4] Myr ago, while subgroups 9.2 diverged from A1 population (common ancestor of 9.1 and 9.4) 1.5 [0.7 – 2.8] Myr ago. For the shorter generation time (0.003 year) that we tested effective size ranges from 1×10^6 (subgroup 9.1) to 3.2×10^8 (A2). Modifying the generation time to 1 and 10 years influenced in an inversely proportional way the effective sizes ($3.33 \cdot 10^{-2}$ and $3.33 \cdot 10^{-3}$ compared to those obtained with a generation time of 0.003 year). Only nine migration rates were significantly different from zero. Seven of them have occurred since 0.06 Myr (Fig. 2). Recent genetic exchanges occurred mainly between subgroups 9.5, 9.6 and 9.3, which belonged to the same super-clade. In the other super-clade of subgroups 9.1, 9.2 and 9.4, no genetic exchanges occurred. Population 9.2 exchanged with populations 6.3 and 9.6.

Estimation of transfer of VA genes concomitant to recombination events

Repertoires of VA genes in ancestor strains were inferred by parsimony analysis at all the nodes of the genealogy provided by the ClonalFrame analysis. Based on these inferred repertoires of VA genes we estimated the possibility that transfer of some genes of the repertoires may have occurred concomitantly to any of the 9 recombination

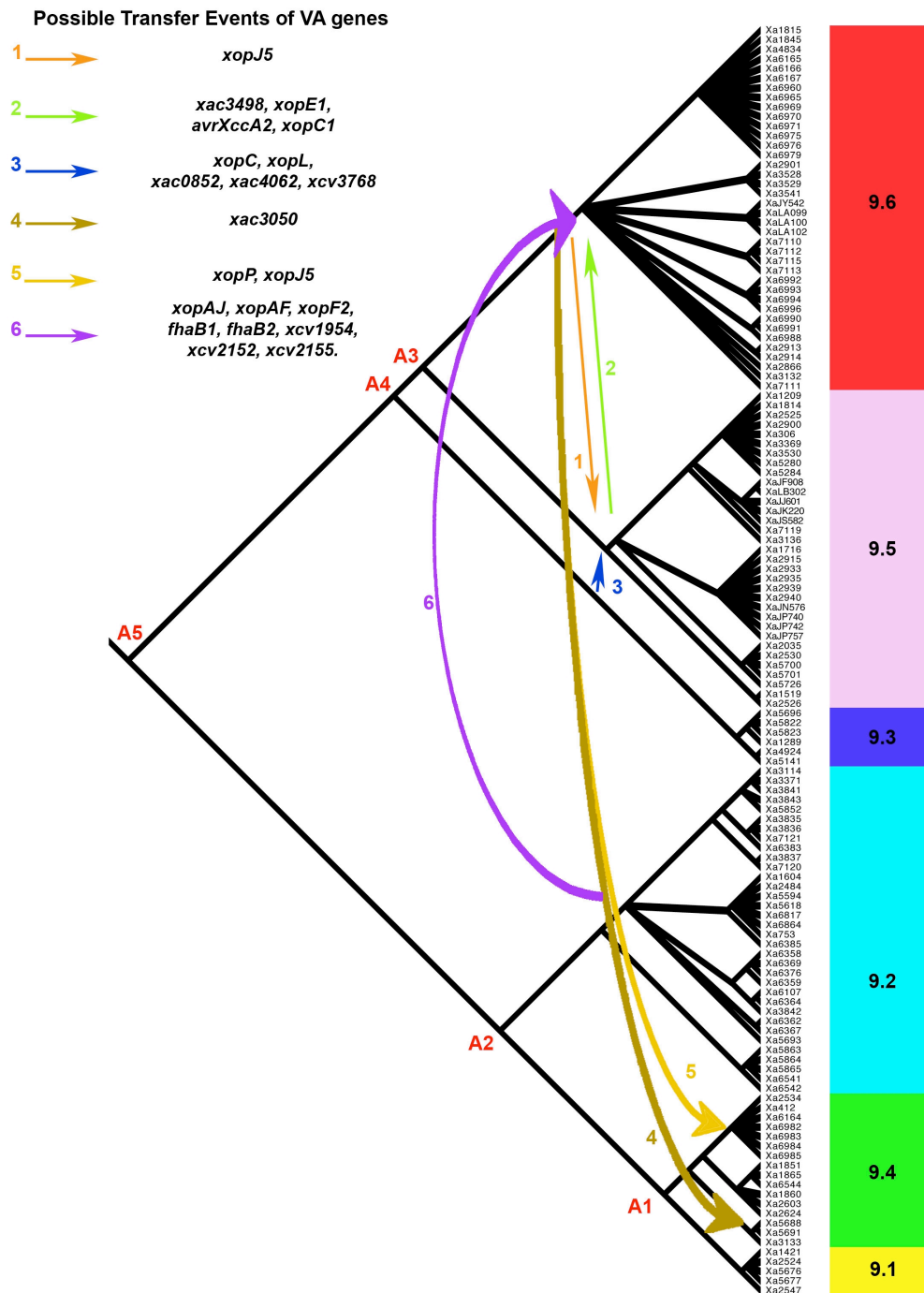


Fig.3 Inferences on transfer of VA genes possibly involved in host specificity. The repertoires of 107 genes encoding MCPs, sensors, adhesins and T3Es were previously determined for each strain of the collection (Hajri et al. 2009; Mhedbi-Hajri et al. submitted). Using a parsimony approach implemented in Mesquite 2.5 (Maddison and Maddison, 2009), repertoires of genes at nodes of the ClonalFrame genealogy were inferred. Comparison of repertoires at ancestral nodes provided hypotheses on horizontal transfers concomitant to recombination events identified in Fig. 2. Populations 9.1 to 9.6 refer to genetic subgroups in *X. axonopodis* and A1 to A5 refer to ancestral populations. Four of the recombination events detected by IMA2 may be associated to transfer of genes encoding MCP (*xac3768, xcv1954*), adhesins (*fhaB1, fhaB2*), sensors (*xac0852, xac2152, xac2155, xac3050, xac3498, xac4062*) and T3Es (*avrXccA2, xopC1, xopE1, xopF2, xopJ5, xopL, xopP, xopAJ, xopAF*).

events (Fig. 3). Transfer events detected in our approach involved transfer of T3Es, sensor proteins, MCPs and adhesins. Transfer of genes encoding T3Es occurred between ancestors of groups 9.5 and 9.6 (Fig. 3, transfer events #1 and 2), between ancestors of groups 9.3 and 9.5 (Fig. 3, transfer event #3), and between ancestors of groups 9.2 and 9.6 (Fig. 3, transfer event #6). As well, transfer of the T3Es genes *xopP* and *xopJ5* occurred between the ancestor of group 9.6 and the ancestor of the Genetic Lineage 1 of pathovar *phaseoli* (Fig. 3, transfer event #5). Transfer events involving genes encoding components of sensory systems occurred between ancestors of groups 9.2 and 9.6 (Fig. 3, transfer event #6), between ancestors of groups 9.3 and 9.5 (Fig. 3, transfer event #3) and between ancestors of groups 9.5 and 9.6 (Fig. 3, transfer event #2). As well, the gene *xac3050* encoding a sensor protein was transferred from the ancestor of group 9.6 to the ancestor of genetic lineage C of pathovar *dieffenbachiae* (Fig. 3, transfer event #4). Finally, genes *fhaB1* and *fhaB2* encoding non-fibrillar adhesins were transferred between ancestors of groups 9.2 and 9.6 (Fig. 3, transfer event #6).

Discussion

Genealogy confirms the existence of 6 subgroups in X. axonopodis

The main goal of our study was to decipher the genetic structure within the bacterial phytopathogenic species *X. axonopodis* (*sensu* Vauterin et al. 1995, 2000). Individuals from this species can be found associated with a large variety of hosts and are present over a large geographical area. Genealogical analysis of 131 strains encompassing 25 genetic lineages and 21 pathovars revealed that *X. axonopodis* could be divided into six clusters matching with those described by Rademaker et al. (2000 and 2005) on the basis of AFLP and rep-PCR polymorphism. This result also confirmed those obtained by MLSA scheme with 4 housekeeping genes developed by Young et al. (2008) on 36 *X. axonopodis* strains. Our analysis based on coalescence approach supports previous clusterings within *X. axonopodis* and infers parameter estimates that make sense about ecological isolation.

Our coalescence-based approach permitted to estimate the divergence times between these groups in order to decipher the evolutionary history of this species.

According to the Isolation-with-Migration model, the six subgroups coalesced over a period of several millions of years (Fig 2). Subgroup 9.3 was found to be the oldest with 5 [2.71-7.29] Myr. The strains of this latter subgroup seemed to have coalesced at a time close to the one of the TMRCA (*ca.* 5.5 [3.68-9.47] Myr). The subgroups 9.1 and 9.4 are supposed to have the shortest history with a divergence time estimated to be around 60,000 years [present-1.03 Myr] whereas divergence times of subgroups 9.2, 9.5 and 9.6 fell around 1.5 [0.66-2.83] to 1.9 [0.90-6.09] Myr.

Divergence and ecological specificities

This deep lineage history of 9.3 could be related to host specificity. In our collection, the subgroup 9.3 is represented only by strains pathogenic on monocots whereas all other subgroups contain only strains pathogenic on dicots, the only exception being the pathovar *allii* in subgroup 9.2. Parkinson et al. (2007) proposed that *Xanthomonas* genus first arose as a monocot pathogen. Subgroup 9.3 could then be considered as the direct descendant of the ancestor of *X. axonopodis*. The long divergence time of subgroup 9.3 could also explain the lack of robustness of the node of the 9.3 branch in genealogical and phylogenetical analyses. In our study, the subgroup 9.3 fell at the base of super-clade which also contains subgroups 9.5 and 9.6. Young et al. (2008) also observed a poorly robust branching (bootstrap value of 53%) of strains of subgroup 9.3 with subgroups 9.1, 9.2 and 9.4. Again, the lack of robustness in the position of subgroup 9.3 could account for its ancestral status. The subgroup 9.3 seemed to have rapidly known an independent evolution in the history of *X. axonopodis*.

Until now, there is no clear rationale accounting for both topology and divergence times for the five other genetic subgroups in *X. axonopodis*. However it is interesting to note that host-specificity may not have played an important role in the pattern of divergence of these subgroups. In one hand, bacterial strains that are pathogenic on phylogenetically distant hosts cluster in the same subgroup (Fig. 1, Rademaker et al. 2005). In the other hand, strains that are pathogenic on a same plant species or family are displayed in very divergent subgroups. For example pathogens of *Citrus* spp. are displayed in subgroups 9.5 and 9.6 that may have diverged around 2 Myr ago, but also in subgroup 9.2 which diverged from the ancestor of 9.5 and 9.6 around 5.5 Myr ago. Accordingly, strains that are pathogenic on *Anacardiaceae* are found in

divergent subgroups: 9.1 (on *Spondias dulcis*), 9.5 (on *Mangifera indica*) and 9.6 (on *M. indica* and *Anacardium occidentale*). It is also the case for strains infecting legumes found in subgroups 9.2 (on *Medicago sativa*), 9.4 (on *Phaseolus vulgaris*) and 9.6 (on *P. vulgaris*). In addition, strains that are pathogenic on common bean (*P. vulgaris*) cluster as previously mentioned by Alavi et al. (2008) in four lineages. Bean pathogenic strains belonged to subgroups 9.4 and 9.6 that coalesced a long time ago, at a time close to the one of the TMRCA (*ca.* 5.5 Myr), but display pathological convergence today.

An evolutionary history that is apparently independent from pathological specificities could be a surprising result as evolution of pathogenic bacteria seems to be mainly driven by host-specificity (Deitsch et al. 2009, Ding et al. 2009, Zdziarski et al. 2010). One could expect to have the evolutionary history of the strains of *X. axonopodis* matching the evolution of their respective hosts, as it is often the case for other pathogenic organisms (Moodley et al. 2009). Three kinds of arguments may explain why it is not the case here. First, xanthomonads could interact with a wider host range than symptomatic crops only. Our sampling is only based on pathogenic strains isolated from cultivated host plants. We did not analyze strains from wild hosts while some xanthomonads are known to inhabit and infect wild species (Sikirou & Wydra 2004; Barak et al. 2001; Schaad & Dianese 1981), and we limited our collection to strains isolated from cultivated hosts exhibiting symptoms. *X. axonopodis* also encompasses strains that are commensals on plants and for example found in association with seeds (Darrasse et al. 2010, Haworth & Spiers 1992). However, xanthomonads are known to colonize only plant-associated environments and until now were not isolated from other environments such as water or raw soil. Hence, a part of the diversity might have been neglected in our study, and these other niches could have shaped the strain diversity. Second, pathogenicity may have been acquired secondarily in the evolution of *X. axonopodis* by a few individuals of each lineage following horizontal gene transfers. Hajri et al. (2009) showed that repertoires of Type III Secretion System effectors could account for such a phenotypic convergence. Third, we cannot rule out an alternative hypothesis stating that strains belonging to the various pathovars all derived from an ancestor carrying a very broad host range. Specialization of strains into pathovars displaying distinct host ranges may account for agricultural development which arose only recently, beginning about 10,000 years ago. Under such hypothesis, a polyphagous

pathogenic bacterium would have been more fit in natural environments, whereas development of agriculture would have selected pathogenic bacteria specialized on a limited host range. This is consistent with a recent theory on the origin and evolution of plant pathogens in agro-ecosystems (Stuckenbrock & McDonald 2008). The evolutionary scenario we provided here (Figure 2) derived from the genealogical analysis of our strains under ClonalFrame. Such analysis provided a clustering of strains into 6 subgroups. Hence, we were not able to date the individualization of pathovars within our dataset and to face this dating to periods of domestication of host plants and intensification of agriculture.

The evolutionary history of strains pathogenic on cultivated plants might have been influenced by human practices rather than for strains specific of wild plants. Bacterial populations inhabiting cultivated crops have known a tremendous expansion related to the large diffusion of agriculture and to other human practices like commercial exchanges and extensive migration (Hancock 2006). Cultivated crops have been subjected to many genetic changes due to domestication and breeding. In addition, the massive increase of agricultural surfaces might have allowed a large demographic expansion of bacterial pathogens of cultivated plant. However, no signal of such a demographic change has been detected by neutrality tests like Tajima's D or Fu and Li's D* and F* tests. Nevertheless, the power of these tests is known to decrease with time after expansion (Achaz 2009). Expansion of several centuries coupled with a short generation time as expected in bacteria, would lead to a high number of generations until present and may preclude neutrality tests from powerful detection.

Our demographic estimates are based on values of mutation rate and generation time determined for *E. coli* (Guttman & Dykhuizen 1994). Such hypotheses may induce biases on the absolute estimates of parameters but do not change their relative values. Nübel et al. (2010) reviewed substitution rates of several human bacterial pathogens. Estimates ranged from 3×10^{-8} for *E. coli* to 6.6×10^{-5} for *Campylobacter jejuni*. This latter substitution rate is overestimated by the effect of recombination known to be high in this species. Hence actual mutation rates should be close to the one of *E. coli*. Our demographic estimates (divergence time, migration rates and dating) should then be considered as conservative and may have occurred more recently.

Migration is related to ecological niches

Gene flow between subgroups estimated under the Isolation-with-Migration model is rare and low. Under this model, migration was allowed between populations present at the same time. Estimates of migration ranged between 0.002 (from 9.3 to 9.5) to 0.235 (from 9.3 to A2). Using the same model, Moodley et al. (2009) studied the gene flow between populations of *Helicobacter pylori* living in human and found estimates ranging from 0 to 6.6. In our study, subgroups in *X. axonopodis* seem to have evolved like populations isolated from each other. Lower gene flow than for *H. pylori* can easily be explained by the fact that strains of *Xanthomonas* studied here are found on several hosts whereas only human *H. pylori* were considered by Moodley et al. (2009). Thus it is not excluded to find such gene flow estimates between populations of the same pathovar.

A major result of the analysis under the Isolation-with-Migration model is the posterior estimate of time of gene flow. Among the nine significant migration events detected, seven of them occurred between 60,000 years and present. The two others concern exchanges that occurred between 9.6 and A1 one million of year ago and between 9.3 and A2 1.6 Myr ago (Fig. 2). Recent migration events may account for agricultural development. Under such hypothesis, plants coming from different geographical areas and hosting different pathogenic strains have been cultivated in the same places. Intensive cultivation has probably led to increase the local effective size of bacterial strains. Then recombination events with strains from other subgroups could have been effective even if they were rare. Following the theory developed by Stuckenberg & McDonald (2008) ancestral strains of *X. axonopodis* probably were pathologically less specialized than contemporary strains. Consequently, these populations would have a higher probability to share identical or spatially close ecological niches that could account for the observed pattern of gene flow. The genetic and environmental uniformity of monocultures coupled with the large size of intensified agro-ecosystems would have then selected for highly specialized and aggressive pathogens.

Migration seems to preferentially occur between certain subgroups while others remain isolated. Subgroups 9.5 and 9.6, 9.5 and 9.3 as well as 9.2 and 9.6 exchanged genes symmetrically. Subgroup 9.2 has sent migrants unilaterally toward subgroup 9.3.

Conversely, subgroups 9.1 and 9.4 are isolated from other subgroups. Proximity of ecological niches may explain the observed pattern of gene flow. Subgroups that have exchanged genes for less than 60,000 years could have colonized the same families of host-plants. For example, both subgroups 9.5 and 9.6 contain strains pathogenic of *Anacardiaceae* (pv. *mangiferaeindicae* in 9.5 and pv. *anacardii* in 9.6, Ah You et al. 2009), *Rutaceae* (pvs. *citri* and *bilvae* in 9.5 and pv. *aurantifolii* in 9.6, Bui Thi Ngoc et al. 2010) and *Fabaceae* (pv. *glycines* in 9.5 and pathovars *phaseoli fuscans*, *GL2* and *GL3* in 9.6). Thus, strains from subgroups 9.5 and 9.6 share common hosts or phylogenetically closely related hosts at least. Subgroups 9.6 and 9.2 share pathogens on *Rutaceae* (pv. *citrumelo* in 9.2 and pv. *aurantifolii* in 9.6, Rademaker et al. 2000), and *Fabaceae* (pv. *alfalfae* in 9.2 and pv. *phaseoli* and pv. *vignicola* in 9.6, Rademaker et al. 2000). This niche overlapping is potentially a cause of genetic exchanges between pathologically specialized pathogens. Here again, extension of agriculture and secondary contacts of crops might have favored gene flows between strains infecting the same host-plant families. Genetic exchanges between subgroups could also be due to the ability of certain pathovars to survive as epiphytic on other hosts. Gent et al. (2005) reported epiphytic survival of pv. *allii* on legumes as well as epiphytic survival of pv. *phaseoli* on onions. This is another opportunity for strains coming from different subgroups to co-exist on the same host and exchange genes.

Genetic exchange may also be caused by spatial proximity of crops. Pathovars *vesicatoria* (subgroup 9.2) and *phaseoli* (essentially subgroup 9.6) colonize hosts (*Lycopersicon esculentum* and *P. vulgaris*, respectively) coming from the same domestication area in Central and South America. It was already observed that these two pathovars share common mobile elements not present in other pathovars of the species. Indeed, a common IS element belonging to a family rarely found in Xanthomonadaceae was identified only in the four genetic lineages of *X. axonopodis* pv. *phaseoli* and in *X. axonopodis* pv. *vesicatoria* (Alavi et al. 2007). This could indicate that somehow they share ecological niches allowing horizontal gene transfer. In the same way, observed gene flow between subgroups 9.3 and 9.5 may also reflect the sharing of cultivation zones rather than ecological niches. As discussed above, 9.3 is a group of strains specific of monocots and 9.5 groups strains that are specific of dicots. However, host-plants (mango, sugar cane and citrus, for example) of these two

subgroups of strains share the same environment being cultivated in the same agropedoclimatic areas indeed in the same fields.

At last, our data suggests that VA genes possibly involved in host specificity (MCPs, adhesins, sensors and T3Es) were transferred concomitantly to recombination events identified on the basis of the analysis of housekeeping genes between groups 9.3 and 9.5, between groups 9.5 and 9.6, and between groups 9.2 and 9.6. Since Hajri and his colleagues (2009) showed that there is a correlation between repertoires of T3Es and pathological specificity for strains in *X. axonopodis*, such transfer may have impacted host range of strains and promoted the emergence of strains carrying new pathological specificities. Along the same lines, transfer of MCPs and sensors may have modified the range of hosts perceived by pathogenic bacteria, while acquisition of new adhesins may have enhanced the ability of strains to colonize new hosts. On the basis of the analysis of housekeeping genes, we did not observe any recombination events between strains of the three lineages of pathovar *phaseoli* belonging to subgroup 9.6 and the genetic lineage (GL1) falling in subgroup 9.4. Hypotheses discussed above would have predicted high migration rates between strains colonizing the same host. However, earlier migration that occurred between 9.6 and A1 (ancestor of 9.1 and 9.4) may account for the pathological convergence between strains of the different lineages of the pathovar *phaseoli*. One could hypothesize the transmission from the strains of pv. *phaseoli* of subgroup 9.6 to A1 of one or several virulence factor(s) which allowed some strains to infect beans. Indeed, our approach detected the transfer of the T3E genes *xopP* and *xopJ5* between the ancestor of strains belonging to the group 9.6 and the ancestor of the GL1 of pathovar *phaseoli* (Fig.3, transfer event #5). The same rationale could be applied to pathovar *dieffenbachiae*, which is also present in subgroups 9.6 and 9.4. Our approach detected the transfer of the gene *xac3050* encoding a sensor protein from the ancestor of group 9.6 to the ancestor of genetic lineage C of pathovar *dieffenbachiae* (Fig. 3, transfer event #4).

Recombination and speciation

Recombination occurred in our whole dataset as in each of the six clades. Estimate of r/m shows that recombination impacted the diversity of our dataset three times more than mutation. Vos and Didelot (2009) estimated r/m on 48 species of bacteria and

archae. Estimates of r/m ranged from 0.02 (*Leptospira interrogans*) to 63.6 (*Flavobacterium psychrophilum*) (Vos and Didelot, 2009). Our estimate of $r/m = 3.18$ for *X. axonopodis* is comparable to that of *Campylobacter insulaenigrae* ($r/m=3.2$) which is ranked 17th highest value. However r/m could be underestimated by our sampling that mixed several populations (Vos & Didelot, 2009). Estimates on distinct clades were not reliable. Nevertheless we could suggest that recombination in *X. axonopodis* has shaped most of the diversity. Recent migration events detected by IMA2 should account for high recombination rates (see below). Inspection of evolutionary events along the branches of the genealogy showed that recombination is widespread in the genealogy (Fig. 1). In addition, recombination affects the seven genes used for the study. Such a high recombination rate could reflect genetic exchanges among subgroups. For instance, data presented here show that the divergence time of subgroups 9.5 and 9.6 felt around 1.9 Myr. This suggests that the divergence between these two subgroups is very ancient. Genetic exchange has occurred since 0.06 Myr and should be linked to the recent extension of agriculture. According to these results, conditions of sexual and ecological isolation are not met to elevate independently these two subgroups at the species level. In the same way, significant recent gene flow between other subgroups should reflect that genetic isolation is not complete, that would prevent to consider these subgroups as separate species.

Recombination events due to genetic exchanges have been found to occur recently in *X. axonopodis* strains probably along agricultural intensification and expansion. This finding leads us to question about the evolution of these subgroups. Would divergence acquired since several Myr lead to speciation despite the recent occurrence of migration? In one hand, the seven housekeeping genes we used are involved in the basal metabolism that can be shared within the bacterial community. Genetic barriers associated with speciation may be located elsewhere in non-recombinant regions of the genome. In the other hand, secondary contacts due to agricultural activities could be sufficient to enhance homogenization among genomes. One cannot exclude that increasing genetic exchanges among strains belonging to different subgroups would promote emergence of strains with new pathological competences.

Acknowledgments and funding

We thank Perrine David for helpful discussion and critical reading of the manuscript. N Mhedbi-Hajri and A Hajri were supported by grants from the Tunisian Government. A Hajri was supported by an additional grant from the Conseil Général du Maine et Loire and from INRA. This study was part of programs funded by the Region Pays de la Loire and by the Plant Health and Environment Department of INRA.

References

- Achaz G (2009) Frequency spectrum neutrality tests: one for all and all for one. *Genetics* **183**, 249-258.
- Ah-You N, Gagnevin L, Grimont PAD *et al.* (2009) Polyphasic characterization of xanthomonads pathogenic to members of the Anacardiaceae and their relatedness to species of *Xanthomonas*. *International Journal of Systematic and Evolutionary Microbiology*, **59**, 306–318.
- Alavi MS, Poussier S, Manceau C (2007) Characterization of IS*Xax1*, a novel insertion sequence restricted to *Xanthomonas axonopodis* pv. *phaseoli* (variants *fuscans* and *non-fuscans*) and *Xanthomonas axonopodis* pv. *vesicatoria*. *Applied and Environmental Microbiology*, **73**, 1678–1682.
- Alavi SM, Sanjari S, Durand F *et al.* (2008) Assessment of the genetic diversity of *Xanthomonas axonopodis* pv. *phaseoli* and *Xanthomonas fuscans* subsp. *fuscans* as a basis to identify putative pathogenicity genes and a type III secretion system of the SPI-1 family by multiple suppression subtractive hybridizations. *Applied and Environmental Microbiology*, **74**, 3295–3301.
- Ausubel FM, Brent R, Kingston *et al.* (1992) *Current protocols in molecular biology*. Greene Publishing associates and Wiley Interscience, New york, USA.
- Balloux F (2010) Demographic influences on bacterial population structure. In: *Bacterial population genetics in infectious disease*, (eds Robinson DA, Falush D, Feil E), pp. 103–120. John Wiley & Sons Inc., Hoboken, New Jersey.

- Barak JD, Koike ST, Gilbertson RL (2001) The role of crop debris and weeds in the epidemiology of bacterial leaf spot of lettuce in California. *Plant Disease*, **85**, 169-178.
- Bisharat N (2010) Population genetics of Vibrios. In: *Bacterial population genetics in infectious disease*, (eds Robinson DA, Falush D, Feil E), pp. 103–120. John Wiley & Sons Inc., Hoboken, New Jersey.
- Boni MF, Posada D, Feldman MW (2007) An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics*, **176**, 1035–1047.
- Bui Thi Ngoc L, Vernière C, Jouen E *et al.* (2009) Amplified fragment length polymorphism and multilocus sequence analysis-based genotypic relatedness among pathogenic variants of *Xanthomonas citri* pv. *citri* and *Xanthomonas campestris* pv. *bilvae*. *International Journal of Systematic and Evolutionary Microbiology*, **60**, 515–525.
- Cohan FM (2002) Sexual isolation and speciation in bacteria. *Genetica*, **116**, 359–370.
- Costechareyre D, Bertolla F, Nesme X (2009) Homologous recombination in *Agrobacterium*: potential implications for the genomic species concept in bacteria. *Molecular Biology and Evolution*, **26**, 167–176.
- Darrasse A, (2010) Transmission of plant-pathogenic bacteria by nonhost seeds without induction of an associated defense reaction at emergence. *Applied and Environmental Microbiology*, **76**, 6787-6796.
- Deitsch KW, Lukehart SA, Stringer JR (2009) Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nature Reviews Microbiology*, **7**, 493–503
- Didelot X, (2010) Demographic influences on bacterial population structure. In: *Bacterial population genetics in infectious disease*, (eds Robinson DA, Falush D, Feil E), pp. 103–120. John Wiley & Sons Inc., Hoboken, New Jersey.

- Didelot X, Falush D (2007) Inference of bacterial microevolution using multilocus sequence data. *Genetics*, **175**, 1251–1266.
- Ding F, Tang P, Hsu M-H, Cui P, Hu S, Yu J, Chiu C-H (2009) Genome evolution driven by host adaptations results in a more virulent and antimicrobial-resistant *Streptococcus pneumoniae* serotype 14. *BMC Genomics*, **10**, 158 doi:10.1186/1471-2164-10-158
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185–1192.
- Dye DW, Bradbury JF, Goto M, Hayward AC, Lelliott RA, Schroth MN (1980) International standards for naming pathovars of phytopathogenic bacteria and a list of pathovar names and pathotype strains. *Review of Plant Pathology*, **59**, 153–168.
- Fargier E, Manceau C (2007) Pathogenicity assays restrict the species *Xanthomonas campestris* into three pathovars and reveal nine races within *X. campestris* pv. *campestris*. *Plant Pathology*, **56**, 805–818.
- Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.
- Gent DH, Al-Saadi A, Gabriel DW *et al.* (2005) Pathogenic and genetic relatedness among *Xanthomonas axonopodis* pv. *allii* and other pathovars of *X. axonopodis*. *Phytopathology*, **95**, 918–925.
- Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics*, **16**, 573–582.
- Guttman DS, Stavrinos J (2010) Population genomics of bacteria. In: *Bacterial population genetics in infectious disease*, (eds Robinson DA, Falush D, Feil E), pp. 121–151. John Wiley & Sons Inc., Hoboken, New Jersey.
- Guttman DS, Dykhuizen DE (1994) Clonal divergence in *Escherichia coli* as a result of

recombination, not mutation. *Science*, **266**, 1380–1383.

Hajri A, Brin C, Hunault G *et al.* (2009) A «Repertoire for Repertoire» Hypothesis: Repertoires of type three effectors are candidate determinants of host specificity in *Xanthomonas*. *PLoS ONE*, **4**, e6632.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hancock JF (2006) *Plant Evolution*, 2nd edn. CABI, Cambridge, USA.

Hamza AA, Robène-Soustrade I, Jouen E, Gagnevin L, Lefeuvre P (2010) Genetic and pathological diversity among *Xanthomonas* strains responsible for bacterial spot on tomato and pepper in the Southwest Indian Ocean Region. *Plant Disease*, **94**, 993–999.

Hayward AC (1993) The hosts of *Xanthomonas*. In: *Xanthomonas*, (eds Swings JC, Civerolo EL), pp. 1–119. Chapman and Hall, London.

Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, **27**, 905–920.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

Jones JB, Lacy GH, Bouzar H, Stall RE, Schaad NW (2004) Reclassification of the xanthomonads associated with bacterial spot disease of tomato and pepper. *Systematic and Applied Microbiology*, **27**, 755–762.

Jones JB, Lacy GH, Bouzar H, Stall RE, Schaad, NW (2006) *Xanthomonas euvesicatoria* sp. nov., *Xanthomonas gardneri* nom. rev., comb. nov. and *Xanthomonas perforans* sp. nov. In List of new names and new combinations previously effectively, but not validly, published, validation list no. 109.

International Journal of Systematic and Evolutionary Microbiology, **56**, 925–927.

Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW (2006) GARD: a genetic algorithm for recombination detection. *Bioinformatics*, **22**, 3096–3098.

Kingman JFC (1982) The coalescent. *Stochastic Processes and their Applications*, **13**, 235–248.

Kuhner MK (2008) Coalescent genealogy samplers: windows into population history. *Trends in Ecology and Evolution*, **24**, 86–93.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Maddison WP, Maddison DR (2009) Mesquite: a modular system for evolutionary analysis. Version 2.6 <http://mesquiteproject.org>.

Martin D & Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. *Bioinformatics*, **16**, 562–563.

Martin DP, Williamson C, Posada D (2005a) RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics*, **21**, 260–262.

Martin DP, Posada D, Crandall KA, Williamson C (2005b) A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Research and Human Retroviruses*, **21**, 98–102.

Maynard Smith J (1992) Analysing the mosaic structure of genes. *Journal of Molecular Evolution*, **34**, 126–129.

Moodley Y, Linz B, Yamaoka Y *et al.* (2009) The peopling of the Pacific from a bacterial perspective. *Science*, **323**, 527–530.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia Univ. Press, New York.

- Nübel U, Dordel J, Kurt K, Strommenger B, Westh H, et al (2010) A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant *Staphylococcus aureus*. *PLoS Pathogens*, **6**, e1000855. doi:10.1371/journal.ppat.1000855.
- Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. *Virology*, **265**, 218–225.
- Parkinson N, Aritua V, Heeney J *et al.* (2007) Phylogenetic analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. *International Journal of Systematic and Evolutionary Microbiology*, **57**, 2881–2887.
- Parkinson N, Cowie C, Heeney J, Stead D (2009) Phylogenetic structure of *Xanthomonas* determined by comparison of *gyrB* sequences. *International Journal of Systematic and Evolutionary Microbiology*, **59**, 264–274.
- Posada D, Crandall K (2001) Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences of the USA*, **24**, 13757–13762.
- Pybus OG, Rambaut A (2009) Evolutionary analysis of the dynamics of viral infectious disease. *Nature Reviews Genetics*, **10**, 540–550.
- Rademaker JLW, Hoste B, Louws FJ *et al.* (2000) Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: *Xanthomonas* as a model system. *International Journal of Systematic and Evolutionary Microbiology*, **50**, 665–677.
- Rademaker JLW, Louws FJ, Schultz MH *et al.* (2005) A comprehensive species to strain taxonomic framework for *Xanthomonas*. *Phytopathology*, **95**, 1098–1111.
- Rice W (1989) Analysis tables of statistical tests. *Evolution*, **43**, 223–225.
- Robinson DA, Falush D, Feil E (2010) *Bacterial population genetics in infectious disease*. John Wiley & Sons Inc., Hoboken, New Jersey.

- Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics*, **3**, 380–390.
- Schaad NW, Dianese JC (1981) Cruciferous weeds as sources of inoculums of *Xanthomonas campestris* in black rot of crucifers. *Phytopathology*, **71**, 1215–1220.
- Schaad NW, Postnikova E, Lacy GH *et al.* (2005) Reclassification of *Xanthomonas campestris* pv. *citri* (ex Hasse 1915) Dye 1978 forms A, B/C/D, and E as *X. smithii* subsp. *citri* (ex Hasse) sp. nov. nom. rev. comb. nov., *X. fuscans* subsp. *aurantifolii* (ex Gabriel 1989) sp. nov. nom. rev. comb. nov., and *X. alfalfae* subsp. *citrumelo* (ex Riker and Jones) Gabriel *et al.*, 1989 sp. nov. nom. rev. comb. nov.; *X. campestris* pv. *malvacearum* (ex Smith 1901) Dye 1978 as *X. smithii* subsp. *smithii* nov. comb. nov. nom. nov.; *X. campestris* pv. *alfalfae* (ex Riker and Jones, 1935) Dye 1978 as *X. alfalfae* subsp. *alfalfae* (ex Riker *et al.*, 1935) sp. nov. nom. rev.; and "var. *fuscans*" of *X. campestris* pv. *phaseoli* (ex Smith, 1987) Dye 1978 as *X. fuscans* subsp. *fuscans* sp. nov. *Systematic and Applied Microbiology*, **28**, 494–518.
- Schaad NW, Postnikova E, Lacy GH *et al.* (2006) Emended classification of xanthomonads pathogens on citrus. *Systematic and Applied Microbiology*, **29**, 690–695.
- Schaad NW, Postnikova E, Lacy GH, *et al.* (2007) *Xanthomonas alfalfae* sp. nov., *Xanthomonas citri* sp. nov. and *Xanthomonas fuscans* sp. nov. In list of new names and new combinations previously effectively, but not validly, published, validation list no. 115. *International Journal of Systematic and Evolutionary Microbiology*, **57**, 893–897.
- Sikirou R, Wydra K, (2004) Persistence of *Xanthomonas axonopodis* pv. *vignicola* in weeds and crop debris and identification of *Sphenostylis stenocarpa* as a potential new host. *European Journal of Plant Pathology*, **110**, 939–947

- Staden R (1996) The Staden sequence analysis package. *Molecular Biotechnology*, **5**, 233–241.
- Stuckenbrock, EH, and McDonald, BA (2008) The origins of plant pathogens in agroecosystems. *Annual Review of Phytopathology*, **46**,75-100.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876–4882.
- Vauterin L, Hoste B, Kersters K, Swings J (1995) Reclassification of *Xanthomonas*. *International Journal of Systematic Bacteriology*, **45**, 472–489.
- Vauterin L, Rademaker J, Swings J (2000) Synopsis on the taxonomy of the genus *Xanthomonas*. *Phytopathology*, **90**, 677–682.
- Vos M, Didelot X (2009) A comparison of homologous recombination rates in bacteria and archaea. *ISME Journal*, **3**, 199-208.
- Watterson GA (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256-276.
- Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O et al. (2010) Report of the Ad Hoc committee on reconciliation of approaches to bacterial systematics. *International Journal of Systematic Bacteriology*, **37**, 463–484.
- Young JM, Park DC, Shearman H, Fargier E (2008) A multilocus sequence analysis of the genus *Xanthomonas*. *Systematic and Applied Microbiology*, **31**, 366–377.

Young JM, Wilkie JP, Park DC, Watson DRW (2010) New Zealand strains of plant pathogenic bacteria classified by multi-locus sequence analysis; proposal of *Xanthomonas dyei* sp. nov. *Plant Pathology*, **59**, 270–281.

Zdziarski J, Brzuszkiewicz E, Wullt B, Liesegang H, Biran D, et al. (2010) Host imprints on bacterial genomes—rapid, divergent evolution in individual patients. *PLoS Pathogens*, **6**, e1001078. doi:10.1371/journal.ppat.1001078

Table S1 Descriptive statistics of the polymorphism in seven housekeeping genes for the 6 clades defined in *X. axonopodis*. Clade refers to genetic subgroups 9.1 to 9.6

Clade	Locus	n ^a	Sites ^b	Hap ^c	Hd ^d	S ^e	q _p ^f	q _w ^g	Tajima's D ^h	Fu & Li's D* ^h	Fu & Li's F* ^h
9.1	<i>atpD</i>	5	777	2	0.400	20	0.01030	0.01236	-1.2321	-1.2321*	-1.3200
9.1	<i>dnaK</i>	5	897	2	0.400	12	0.00535	0.00642	-1.2054	-1.2054	-1.2776
9.1	<i>efp</i>	5	360	2	0.400	6	0.00667	0.00800	-1.1455	-1.1455	-1.1870
9.1	<i>fyuA</i>	5	864	2	0.400	7	0.00326	0.00392	-1.1617	-1.1617	-1.2109
9.1	<i>glna</i>	5	984	2	0.400	6	0.00244	0.00293	-1.1455	-1.1455	-1.1870
9.1	<i>gyrB</i>	5	801	2	0.400	5	0.00250	0.00300	-1.1240	-1.1240	-1.1558
9.1	<i>rpoD</i>	5	783	3	0.700	10	0.00511	0.00613	-1.1927	-1.1927	-1.2578
9.1	Total	5	5466	3	0.700	66	0.00484	0.00580	-1.2620	-1.2620**	-1.3693
9.2	<i>atpD</i>	33	777	10	0.873	58	0.02168	0.01839	0.6612	0.2215	0.4377
9.2	<i>dnaK</i>	33	897	12	0.900	35	0.00813	0.00961	-0.5535	-1.6268	-1.5015
9.2	<i>efp</i>	33	360	4	0.366	12	0.00645	0.00821	-0.6867	-0.9416	-1.0107
9.2	<i>fyuA</i>	33	864	12	0.871	51	0.01021	0.01522	-1.2110	-3.1191*	-2.9325*
9.2	<i>glna</i>	33	984	9	0.551	61	0.00530	0.01552	-2.4402**	-4.3451**	-4.3842**
9.2	<i>gyrB</i>	33	801	11	0.854	51	0.00777	0.01569	-1.8539*	-3.5897**	-3.5585**
9.2	<i>rpoD</i>	33	783	12	0.907	38	0.00718	0.01230	-1.5069	-2.0230	-2.1885
9.2	Total	33	5466	19	0.939	306	0.00958	0.01399	-1.2099	-2.6447*	-2.5525*
9.3	<i>atpD</i>	6	777	2	0.533	7	0.00480	0.00395	1.2674	1.5773*	1.6214
9.3	<i>dnaK</i>	6	897	2	0.533	6	0.00357	0.00293	1.2465	1.5507*	1.5891
9.3	<i>efp</i>	6	360	3	0.733	9	0.01130	0.01217	-0.4291	-0.2308	-0.2949
9.3	<i>fyuA</i>	6	864	2	0.533	11	0.00684	0.00561	1.3171	1.6404**	1.6996*
9.3	<i>glna</i>	6	984	2	0.533	9	0.00488	0.00401	1.2971	1.6150*	1.6679
9.3	<i>gyrB</i>	6	801	2	0.533	16	0.01065	0.00875	1.3468	1.6784**	1.7474*
9.3	<i>rpoD</i>	6	783	2	0.533	6	0.00409	0.00336	1.2465	1.5507*	1.5891
9.3	Total	6	5466	3	0.733	64	0.00612	0.00521	1.1147	1.4406	1.5045
9.4	<i>atpD</i>	16	777	4	0.750	15	0.00866	0.00582	1.9104	1.5088**	1.8666**
9.4	<i>dnaK</i>	16	897	4	0.692	6	0.00215	0.00202	0.2211	0.6123	0.5809
9.4	<i>efp</i>	16	360	3	0.675	3	0.00368	0.00251	1.3496	1.0443	1.2863
9.4	<i>fyuA</i>	16	864	3	0.675	44	0.01914	0.01581	0.8907	1.6674**	1.6720*
9.4	<i>glna</i>	16	984	3	0.675	4	0.00173	0.00123	1.2835	1.1414	1.3501
9.4	<i>gyrB</i>	16	801	4	0.692	13	0.00526	0.00489	0.2942	1.4782*	1.3241
9.4	<i>rpoD</i>	16	783	3	0.675	27	0.01208	0.01049	0.6255	1.6091**	1.5375
9.4	Total	16	5466	5	0.767	112	0.00765	0.00625	0.9696	1.6783**	1.7071*
9.5	<i>atpD</i>	33	777	3	0.606	5	0.00256	0.00159	1.6358	1.1348	1.4956
9.5	<i>dnaK</i>	33	897	5	0.695	5	0.00218	0.00137	1.5621	0.2416	0.7407
9.5	<i>efp</i>	33	360	5	0.695	4	0.00544	0.00274	2.4714*	1.0474	1.7053**
9.5	<i>fyuA</i>	33	864	6	0.703	23	0.00362	0.00661	-1.5689	-0.3323	-0.8673
9.5	<i>glna</i>	33	984	4	0.229	4	0.00047	0.00100	-1.3299	-1.0136	-1.2864
9.5	<i>gyrB</i>	33	801	6	0.504	11	0.00244	0.00338	-0.8750	-0.6208	-0.8214
9.5	<i>rpoD</i>	33	783	6	0.778	12	0.00504	0.00378	1.0671	0.9890	1.1895
9.5	Total	33	5466	13	0.850	64	0.00282	0.00289	-0.0937	0.1469	0.0777

9.6	<i>atpD</i>	38	777	8	0.862	14	0.00385	0.00429	-0.3229	0.1890	0.0268
9.6	<i>dnaK</i>	38	897	7	0.748	9	0.00149	0.00239	-1.1077	-1.1137	-1.3043
9.6	<i>efp</i>	38	360	3	0.240	2	0.00069	0.00133	-0.9247	-0.8125	-0.9782
9.6	<i>fyuA</i>	38	864	8	0.799	8	0.00226	0.00222	0.0548	-0.7016	-0.5447
9.6	<i>glnA</i>	38	984	8	0.750	13	0.00156	0.00314	-1.5890	-2.2835	-2.4236
9.6	<i>gyrB</i>	38	801	12	0.835	16	0.00466	0.00475	-0.0632	-0.0360	-0.0529
9.6	<i>rpoD</i>	38	783	10	0.824	12	0.00402	0.00370	0.2683	0.4691	0.4758
9.6	Total	38	5466	19	0.940	74	0.00273	0.00323	-0.5647	-0.7498	-0.8143

^a Sample size.

^b Number of analyzed sites.

^c Number of haplotypes.

^d Haplotype diversity, Nei 1987.

^e Number of polymorphic sites

^f Nucleotide diversity, Nei 1987.

^g Nucleotide diversity from S, Waterson 1975.

^h Neutrality tests: Tajima's D (Tajima 1989), Fu & Li's D* and F* (Fu & Li 1993) and associated P-value: (* $p < 0.05$ and ** $p < 0.01$). No neutrality test remained significant after a correction of Bonferroni for multiple tests was applied for each clade with threshold $K = 0.05/8 = 0.00625$.

Table S2 IMA2 parameter estimates

Parameters ^a	HiPt ^b	HPD95Lo ^c	HPD95Hi ^c
t0	0.060	0.000	1.025
t1	1.507	0.663	2.834
t2	1.869	0.904	6.090
t3	5.005	2.231	7.417
t4	5.487	3.316	9.346
Ne _{0.1}	1.005	0.000	17.085
Ne _{0.2}	65.324	37.184	91.453
Ne _{0.3}	35.174	21.105	67.334
Ne _{0.4}	3.015	0.000	19.095
Ne _{0.5}	29.144	15.075	53.264
Ne _{0.6}	37.184	21.105	63.314
Ne _{A1}	137.682	71.353	314.558
Ne _{A2}	322.598	157.781	500.000
Ne _{A3}	5.025	0.000	1500.000
Ne _{A4}	7.035	3.015	2000.000
Ne _{A5}	171.851	19.095	326.618
Nm9.1>9.2	0.0001375	0	0.0021125
Nm9.1>9.3	0.0000115	0	0.0019435
Nm9.1>9.4	0.0000125	0	0.0020875
Nm9.1>9.5	0.0000115	0	0.0019205
Nm9.1>9.6	0.0000105	0	0.0019425
Nm9.2>9.1	0.0032775	0	0.05915
Nm9.2>9.3	0.0018105	0	0.011295
Nm9.2>9.4	0.0012285	0	0.0608
Nm9.2>9.5	0.0000265	0	0.012325
Nm9.2>9.6	0.018795	0.00894	0.03201
Nm9.2>A1	0.00529	0	0.02398
Nm9.3>9.1	0.0000515	0	0.042485
Nm9.3>9.2	0.008865	0.0010695	0.022875
Nm9.3>9.4	0.0000515	0	0.041765
Nm9.3>9.5	0.00546	0.0000915	0.01656
Nm9.3>9.6	0.0000265	0	0.01492
Nm9.3>A1	0.0000335	0	0.022745
Nm9.3>A2	0.0000455	0	0.041635
Nm9.3>A3	0.02525	0	0.045245
Nm9.4>9.1	0.0000105	0	0.0033495
Nm9.4>9.2	0.0000095	0	0.0032395
Nm9.4>9.3	0.0000105	0	0.0032235
Nm9.4>9.5	0.0000095	0	0.0033345
Nm9.4>9.6	0.0006195	0	0.0040215
Nm9.5>9.1	0.0000295	0	0.02646
Nm9.5>9.2	0.0016465	0	0.00727
Nm9.5>9.3	0.0019575	0.0000435	0.00712
Nm9.5>9.4	0.0000295	0	0.02304
Nm9.5>9.6	0.021175	0.007975	0.03241
Nm9.5>A1	0.0000185	0	0.008195
Nm9.5>A2	0.01531	0	0.03077
Nm9.6>9.1	0.0000395	0	0.026665

Nm9.6>9.2	0.0031265	0.0003515	0.009085
Nm9.6>9.3	0.0000135	0	0.0038205
Nm9.6>9.4	0.0000395	0	0.02919
Nm9.6>9.5	0.007115	0.0012495	0.02272
Nm9.6>A1	0.0000825	0	0.00619
Nm9.6>A2	0.0008625	0	0.03491
NmA1>9.2	0.0002405	0	0.0623
NmA1>9.3	0.010685	0	0.08735
NmA1>9.5	0.0014625	0	0.04924
NmA1>9.6	0.05475	0.019885	0.1284
NmA2>9.3	0.2349	0.07695	0.5165
NmA2>9.5	0.07895	0	0.3828
NmA2>A1	0.0009995	0	0.3528
NmA2>A3	0.026985	0	0.3728
NmA2>A4	0.1669	0	0.4288
NmA3>9.3	0.0009995	0	0.6945
NmA3>A2	0.0009995	0	0.6705
NmA4>A2	0.0029985	0	1.1565

^a t_i represents the i^{th} divergence time (Myr) according to the scenario provided in Figure 1, N_{e_i} represents the effective size of the population (in 10^6 individuals) calculated with a generation time equal to 0.003 year, $N_{mi>j}$ represents the number of migrants from population j to i . Estimates of $N_{mi>j}$ significantly different from zero are indicated in bold.

^b H_{ipt} is the point with the highest value in the histogram of posterior distribution.

^c HPD_{95Low} and HPD_{95Hii} are respectively the lower and the higher bound of the estimated 95% highest posterior density.

2.4 Discussion

A travers les études phylogénétiques et généalogiques menées sur les souches de *X. axonopodis* de notre collection, nous avons pu mettre en évidence des signaux conflictuels probablement dus à des événements de recombinaison. Nous avons utilisé pour cela plusieurs méthodes de détection de la recombinaison. La méthode la plus simple consiste à construire les différents arbres phylogénétiques par Neighbor Joining et de constater s'il existe des changements de position des souches. Ces changements engendrent généralement des incongruences entre les topologies des différents arbres. Avec cette méthode, nous avons observé sur certains arbres phylogénétiques, que des changements de positions de souches ont eu lieu. Pour les arbres construits sur la base des séquences de *efp*, *fyuA*, *glnA* et *rpoD*, des mouvements se produisent au sein de *X. axonopodis*, entre les groupes 9.2 et 9.3 (*efp*), 9.2 et 9.6 (*efp* et *rpoD*), et 9.4 et 9.6 (*fyuA*) et 9.5 et 9.6 (*glnA*). En outre, certaines souches appartenant aux groupes 9.1 et 9.2 se placent même à l'extérieur de *X. axonopodis* sur les arbres construits sur la base des séquences de *atpD* et *glnA*. Il n'est donc pas surprenant de noter que les valeurs de bootstrap renseignant sur la robustesse des branches ne soient pas toujours significatives. Les événements conflictuels dans la phylogénie peuvent être révélés par des valeurs de bootstrap non significatives témoignant une faible robustesse (Morrison, 2005).

L'analyse RDP a quant à elle permis d'identifier 41 séquences recombinantes et les souches de type parental. Elle confirme que des échanges de matériel génétique se sont produits au sein de *X. axonopodis* et notamment entre des souches de type parental appartenant aux groupes 9.2 et 9.6 (affectant les gènes *atpD* et *rpoD* des souches du groupe 9.2) et aux groupes 9.4 et 9.6 (affectant le gène *fyuA* des souches du groupe 9.4). En outre, l'analyse RDP a détecté d'autres événements de recombinaison entre des souches de type parental appartenant au groupe 9.4 (affectant le gène *rpoD* des souches du groupe 9.4), et entre des souches du groupe 9.2 (affectant les gènes *fyuA* et *gyrB*).

La méthode SplitStree qui consiste à construire des réseaux phylogénétiques, a permis de générer des réseaux densément maillés pour chacun des sept gènes de ménage

étudiés. Ces réseaux témoignent de la présence de signaux conflictuels probablement dus à des événements de recombinaison. Ce résultat est en accord avec l'analyse DnaSP qui détecte également des événements de recombinaison pour tous les gènes étudiés (cf. article 2).

Le test SH testant la congruence des arbres phylogénétiques confirme l'ensemble de ces résultats. Il montre que les sept arbres phylogénétiques construits à partir des alignements de *atpD*, *dnaK*, *efp*, *fyuA*, *glnA*, *gyrB* et *rpoD* ont des topologies significativement différentes. En revanche, en comparant chaque arbre avec l'arbre des données concaténées, le test SH montre une congruence des topologies. Ce résultat est prévisible car l'arbre des données concaténées est « le reflet » de tous ces arbres phylogénétiques. Les mêmes résultats ont été obtenus avec le test SH réalisé d'une part, en comparant différents arbres phylogénétiques et d'autre part, en comparant l'arbre des données concaténées avec chacun des arbres phylogénétiques (Fargier, 2007 ; Bui Thi Ngoc *et al.*, 2010). Ceci confirme l'existence de signaux conflictuels dans la phylogénie des *Xanthomonas*. Chez *X. axonopodis*, il a déjà été mentionné que le gène *atpD* est affecté par la recombinaison (Bui Thi Ngoc *et al.*, 2010 ; Hamza *et al.*, 2010).

Par ailleurs, dans l'article 2, nous avons mis en évidence par une approche basée sur la coalescence, que des échanges bilatéraux de matériel génétique se produisent entre les groupes 9.5 et 9.6, 9.3 et 9.5 et 9.2 et 9.6. Avec cette approche, nous reconfirmons la migration observée entre les groupes 9.2 et 9.6 sur les arbres phylogénétiques construits à partir des alignements de séquences *efp* et *rpoD*. Cette migration mise en évidence par l'analyse IMA2 peut être associée à des événements de recombinaison détectés par l'analyse RDP entre ces deux groupes. D'autres échanges détectés uniquement par IMA2 se produisent unilatéralement du groupe 9.6 vers l'ancêtre A1 des groupes 9.4 et 9.1, du groupe 9.2 vers 9.3 et du groupe 9.3 vers l'ancêtre A2 des groupes 9.2, 9.1 et 9.4. Ces deux derniers échanges sont cohérents avec le changement de position d'une souche (CFBP1289 identifiée comme *X. axonopodis* pv. *vasculorum*) du groupe 9.3 vers le groupe 9.2.

Concernant le pathovar *citri* (*sensu* Vauterin *et al.*, 1995), toutes les souches sont

assignées au groupe 9.5 de *X. axonopodis*. Au sein de ce pathovar, il existe deux sous-groupes correspondant aux pathotypes A et A*/A^w. Bui thi Ngoc *et al.* (2010) indiquent que ces deux pathotypes se distinguent par les séquences du gène *atpD*. Nos données confirment cette différenciation et montrent par ailleurs que ces deux sous-groupes se distinguent également par leurs profils alléliques. Les souches nommées A*/A^w ne présentent généralement pas de polymorphisme pour la majorité des gènes. Néanmoins, deux des cinq souches testées (JF90-8 et LB302) présentent des types de séquences (STs) différents pour les gènes *fyuA* et *gyrB*.

Comme l'ont déjà souligné Ah-You *et al.* (2007), les souches pathogènes des *Anacardiaceae* se répartissent en trois groupes. Le premier correspond aux souches du pathovar *mangiferaeindicae* dont les souches sont assignées au groupe 9.5 de *X. axonopodis*. Sur les arbres phylogénétiques, les branches portant les souches du pathovar *mangiferaeindicae* sont soutenues par des valeurs de bootstrap supérieures à 50. Le deuxième groupe renferme les souches du pathovar *anacardii* dont les souches se placent au sein du groupe 9.6. Celles-ci se répartissent en deux sous-groupes en fonction de la plante hôte (*Mangifera indica* ou *Anacardium occidentale*). Le dernier groupe est représenté par la souche CFBP2547 du pathovar *spondiae*. D'après les résultats de l'étude phylogénétique, cette souche se place soit au sein du groupe 9.1 soit au sein du groupe 9.4 en fonction du gène de ménage. Dans l'étude basée sur la coalescence, la souche *spondiae* est assignée au groupe 9.1. En revanche, Ah-You *et al.* (2009) indiquent que sur l'arbre phylogénétique construit avec les données concaténées (*atpD*, *dnaK* et *gyrB*), cette souche se place de manière équidistante des groupes 9.1 et 9.4. Néanmoins, d'après les données AFLP, cette souche se place clairement dans le groupe 9.4 (Ah You *et al.*, 2009).

Les souches des pathovars *vesicatoria*, *allii*, *alfalfae*, *citrumelo* et *ricini* se rassemblent au sein du groupe 9.2 de *X. axonopodis*. A l'exception des souches des pathovars *vesicatoria* et *ricini*, les souches de chacun des autres pathovars de ce groupe ne forment pas toujours des groupes cohésifs. Par exemple, les souches du pathovar *allii* possèdent le plus grand nombre de profils alléliques. Ceci est en accord avec les résultats de l'étude polyphasique réalisée sur une collection de souches du pathovar *allii*

(Roumagnac *et al.*, 2004b). Celle-ci révèle une grande diversité au sein de ce pathovar qui n'est corrélée ni à l'hôte ni à l'année d'isolement. Par ailleurs, des études récentes ont mis en évidence l'existence d'événements de recombinaison affectant les souches de *X. axonopodis* pv. *vesicatoria* (synonyme de *X. euvesicatoria*) au niveau du gène *atpD* (Bui Thi Ngoc *et al.*, 2010 ; Hamza *et al.*, 2010). Selon les auteurs, la recombinaison est à l'origine du placement de ces souches dans le groupe 9.2 de *X. axonopodis*. D'après ces mêmes auteurs, l'importance de la longueur de la branche portant les souches de *X. euvesicatoria* sur l'arbre construit avec les données concaténées (*atpD*, *dnaK*, *efp* et *gyrB*) est due à l'existence d'événements de recombinaison. Ce résultat est en accord avec les longueurs de branches portant les souches du pathovar *vesicatoria* qui sont importantes sur les arbres construits à partir des alignements de séquences partielles de *atpD* et *dnaK*.

Young *et al.* (2008) indiquent que *X. axonopodis* est une espèce hétérogène. Ils précisent aussi qu'aucun groupe rep-PCR ne se différencie clairement de *X. axonopodis* en tant qu'espèce. Les propos de Young *et al.* (2008) sont en accord avec nos résultats. En effet, d'après l'étude généalogique, un premier événement de divergence a permis de séparer les groupes 9.3, 9.5 et 9.6 des groupes 9.1, 9.2 et 9.4. Dans le premier groupement, les groupes 9.5 et 9.6 semblent avoir divergé il y a environ 1,9 millions d'années. Plusieurs données semblent indiquer que le groupe 9.3 pourrait être le descendant direct de l'ancêtre des *Xanthomonas* (Parkinson *et al.*, 2007). Dans le deuxième groupement, le groupe 9.2 semble avoir divergé depuis environ 1,5 millions d'années. Ensuite, ce sont les groupes 9.1 et 9.4 qui ont divergé relativement récemment, il y a environ 0,06 millions d'années. Des échanges de matériel génétique se produisent relativement récemment, il y a environ 0.06 millions d'années, entre différents groupes de *X. axonopodis*. Par ailleurs, l'analyse MLST réalisée sur la même collection de souches de *X. axonopodis*, révèle également une très grande diversité au sein de cette espèce.

D'après l'étude menée par Schaad *et al.* (2005), les pathovars *citri* et *malvacearum* d'une part, et le pathovar *aurantifolii* et la lignée génétique *fuscans* du pathovar *phaseoli* d'autre part, ont été indépendamment élevés au rang d'espèce

respectivement sous le nom de *X. citri* et *X. fuscans*. Selon Ah You *et al.* (2009), les noms *X. citri* et *X. fuscans* (*sensu* Schaad *et al.*, 2005) sont des synonymes. Ces auteurs proposent que ces deux espèces soient élevées ensemble au rang d'espèce. Ainsi, l'espèce *X. citri* serait composée par les souches appartenant aux groupes 9.5 et 9.6. D'après l'étude basée sur la coalescence, les groupes 9.5 et 9.6 ont divergé et indépendamment évolué depuis très longtemps. De plus, les échanges de matériel génétique entre ces deux groupes ne sont produits que relativement récemment. Ces migrations relativement récentes auraient ainsi brouillé les relations de parenté existant actuellement entre les souches de ces deux groupes. Sur la base de ces résultats, nous ne pouvons pas confirmer la synonymie entre *X. citri* et *X. fuscans* (*sensu* Schaad *et al.*, 2005) alors que du point de vue phylogénétique, les séquences contemporaines de *X. citri* (*sensu* Schaad *et al.*, 2005) semblent être assez proches de celles de *X. fuscans* (*sensu* Schaad *et al.*, 2005).

Les arbres construits dans le cadre des études phylogénétiques regroupent les souches en fonction des liens de parenté existant entre les séquences contemporaines. En revanche, ces approches sont incapables d'analyser rétrospectivement l'histoire évolutive des souches sur la base de séquences contemporaines contrairement aux approches basées sur la coalescence. L'étude généalogique basée sur la coalescence permet de renseigner sur l'histoire démographique des populations ancestrales et contemporaines et sur les niveaux d'échanges de matériel génétique entre les individus. Elle a surtout permis de clarifier les histoires évolutives de chacun des groupes 9.1 à 9.6 définis au sein de *X. axonopodis*. La connaissance de tous ces paramètres en plus de la connaissance des liens phylogénétiques et de la diversité au sein de l'espèce, a permis de mettre en évidence qu'après un isolement pouvant être long, les groupes 9.1 à 9.6 ont échangé récemment du matériel génétique ce qui est en opposition à l'isolement sexuel auquel on pouvait d'attendre pour des groupes en voie de spéciation.

Troisième partie
Conclusions générales et perspectives

L'hypothèse émise dans le cadre de ce travail est que la spécificité d'hôte reflète une adaptation étroite des bactéries à leur hôte, qu'elle se façonne dès les étapes précoces et se dessine progressivement au cours de l'interaction entre la bactérie et son hôte. Pour tester cette hypothèse, nous avons étudié le rôle des répertoires de gènes potentiellement impliqués dans les étapes précoces de détection et d'adhésion, dans la spécificité d'hôte des bactéries. A cet effet, nous avons caractérisé les répertoires de gènes codant les senseurs et les adhésines dans une vaste collection de souches appartenant à deux espèces de *Xanthomonas* (*X. axonopodis* et *X. campestris*) et étudié l'histoire évolutive des gènes codant les MCPs et les adhésines. Nous avons également analysé les relations phylogénétiques et généalogiques entre les souches de *X. axonopodis*.

Répertoires : différences et rapprochements

Par une démarche de génétique d'association, nous avons pu montrer que la majorité des pathovars et des lignées génétiques (28 des 34) de *Xanthomonas* possèdent des répertoires uniques. Ces pathovars sont par définition différenciés entre eux par leur gamme d'hôte et/ou les symptômes qu'ils provoquent. Ainsi, à ces différences de pathologie, on peut associer des différences de répertoires de senseurs et d'adhésines. Ce résultat constitue ainsi une première validation de notre hypothèse de travail. La validation définitive proviendra de la démonstration du rôle des protéines impliquées dans la perception et l'adhésion dans la spécificité d'hôte par une approche classique de génétique fonctionnelle. Six pathovars non individualisés partagent chacun leur répertoire avec seulement un autre pathovar. Ceci pourrait refléter l'une des limites de notre étude, limite liée au nombre de gènes et de souches étudiés. Il est important de mentionner qu'en ce qui concerne les pathovars dont les répertoires ne sont pas différenciés, nous ne disposons d'aucun génome séquencé. La disponibilité de leurs séquences génomiques permettrait l'identification de gènes qui leur seraient spécifiques et sans doute la différenciation de leurs répertoires. Les similarités entre leurs répertoires pourraient refléter des spécificités écologiques communes à ces souches prises deux à deux. C'est le cas du pathovar *vignicola* et de la lignée génétique GL2 du pathovar

phaseoli qui partagent le même répertoire de senseurs et d'adhésines. Le pathovar *vignicola* et la lignée génétique GL2 sont respectivement pathogènes de *Vigna unguiculata* et *P. vulgaris*. Ils détecteraient des molécules conservées entre ces deux Fabacées appartenant à la sous famille des Faboideae qui sont très proches phylogénétiquement puisqu'elles appartiennent à la même clade des Milletioïdes (Wojciechowski *et al.*, 2004). Dans ce cas, il est possible que les souches GL2 du pathovar *phaseoli* et celles du pathovar *vignicola* soient capables de détecter les mêmes molécules leur permettant de survivre et de se maintenir respectivement sur *V. unguicula* ou sur *P. vulgaris*. Ainsi, il n'est pas à exclure que le rapprochement des répertoires permettrait à certaines souches de survivre et coloniser des plantes hôtes asymptomatiques. Ces plantes constitueraient en effet des plantes réservoirs.

Par ailleurs, nous avons montré que les quatre lignées génétiques (*fuscans*, GL1, GL2 et GL3) du pathovar *phaseoli* qui sont toutes pathogènes du haricot, possèdent des répertoires différents. En revanche, d'après l'étude des répertoires d'ET3s menées sur *X. axonopodis*, il a été démontré que ces quatre lignées ont des répertoires proches pouvant traduire une convergence pathologique (Hajri *et al.*, 2009). Au regard de nos résultats, nous pouvons poser deux hypothèses complémentaires : *i*) les différences de répertoires constitueraient des indices de l'existence de variations dans les gammes d'hôtes des lignées génétiques de *X. axonopodis* pv. *phaseoli* et *ii*) ces différences révéleraient des comportements écologiques différents entre ces quatre lignées en amont de la phase pathologique. Les hypothèses proposées ne sont pas mutuellement exclusives. En effet, la distance entre les répertoires de ces quatre lignées génétiques reflète leur éloignement phylogénétique. En outre, des différences en termes de virulence et d'agressivité entre des souches appartenant aux quatre lignées génétiques du pathovar *phaseoli* ont été montrées à l'aide de tests de pouvoir pathogène effectués sur une gamme de plante hôtes appartenant à des genres, espèces et variétés différentes. En effet, les représentants testés de la lignée génétique GL1 semblent globalement moins agressifs que ceux des trois autres lignées (Hajri, 2009 ; M-A. Jacques données non publiées). La vérification de notre hypothèse se heurte à l'écueil de notre méconnaissance de l'ensemble des hôtes symptomatiques et asymptomatiques des souches de la collection. En effet, la gamme d'hôtes recensés pour chacun des différents

pathovars étudiés est limitée aux plantes cultivées généralement symptomatiques, sur lesquelles les souches ont été isolées. Par ailleurs, la présence d'un système de sécrétion de type SPI-1 (*Salmonella* pathogenicity island-1) en plus du système du SST3 de type *hrp* dans les souches de la lignée génétique GL1 exclusivement, conforte l'hypothèse selon laquelle les quatre lignées ont des comportements écologiques différents (Alavi *et al.*, 2008 ; Marguerettaz *et al.*, 2010). Ainsi, la lignée génétique GL1 possède deux SST3s alors que les trois autres lignées génétiques (*fuscans*, GL2 et GL3) n'en possèdent qu'un seul, de type *hrp*. Auparavant, la présence du SST3 de type SPI-1 a été rapportée dans plusieurs bactéries comme la souche bsa de *Burkholderia pseudomallei*, Ea273 de *E. amylovora*, la souche Et1/99 de *E. tasmaniensis* et plus récemment la souche GPE PC73 de *X. albilineans* (Attree et Attree, 2001; Triplett *et al.*, 2006; Kube *et al.*, 2008; Pieretti *et al.*, 2009). Le rôle du SST3 de type SPI-1 n'est pas encore élucidé chez les bactéries phytopathogènes. Il est suggéré que ce système aurait un rôle dans la transmission des bactéries par des insectes (ex. *E. amylovora*) (Triplett *et al.*, 2006).

Quotient intellectuel des bactéries et reflet de leur potentiel adaptatif

De façon générale, l'existence d'un grand nombre de gènes codant les MCPs et plus généralement les senseurs chez les *Xanthomonas* révèle l'importance du chimiotactisme et de la détection des caractéristiques de l'environnement dans le style de vie de ces bactéries. Dans un génome donné, le nombre total de protéines impliquées dans la signalisation peut être une mesure du potentiel adaptatif de l'organisme. Cette mesure définie par Galperin en 2005, traduit le quotient d'intelligence (QI) chez les bactéries. La formule exprimant le QI en fonction du nombre total de protéines impliquées dans la transduction de signaux et de la taille du génome complet, est présentée dans le paragraphe 2.2.1 de l'introduction bibliographique. Ainsi, nous avons donc calculé le QI de 24 bactéries ayant des spécificités écologiques et parasitaires différentes. Ces bactéries sont pathogènes de plantes (ex. *Xanthomonas* spp., *Xylella fastidiosa*) ou de l'Homme et des animaux (ex. *Yersinia pestis*). Parmi ces bactéries, certaines sont intracellulaires obligatoires (ex. *Rickettsia conorii*, *Chlamydia trachomatis*) ou facultatives (ex. *Brucella melitensis*) ; d'autres bactéries présentent une distribution ubiquiste (ex. *Caulobacter crescentus*, *Stenotrophomonas maltophilia*,

Tableau 1. Quotient intellectuel de 24 bactéries pathogènes de plantes, de l'Homme et des animaux.

Organisme	Souche	Phylum	Taille du génome (kb)	Nb de protéines impliquées dans la transduction	QI
<i>Xanthomonas axonopodis</i> pv. <i>vesicatoria</i>	85-10	Gamma	5420	380	178
<i>Xanthomonas axonopodis</i> pv. <i>citri</i>	306	Gamma	5274	359	178
<i>Xanthomonas campestris</i> pv. <i>campestris</i>	8004	Gamma	5148	345	179
<i>Xanthomonas campestris</i> pv. <i>campestris</i>	ATCC33913	Gamma	5076	344	181
<i>Xanthomonas campestris</i> pv. <i>campestris</i>	B100	Gamma	5079	343	180
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i>	KACC10331	Gamma	4941	244	156
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i>	MAFF311018	Gamma	4940	259	161
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i>	PXO99 ^A	Gamma	5240	274	156
<i>Xylella fastidiosa</i>	9a5c	Gamma	2731	103	181
<i>Stenotrophomonas maltophilia</i>	K279a	Gamma	4851	399	204
<i>Pseudomonas syringae</i> pv. <i>phaseolicola</i>	1448A	Gamma	6112	516	184
<i>Pseudomonas syringae</i> pv. <i>syringae</i>	DC3000	Gamma	6538	563	180
<i>Pseudomonas syringae</i> pv. <i>syringae</i>	B728a	Gamma	6093	532	188
<i>Ralstonia solanacearum</i>	GMI1000	Beta	5810	446	180
<i>Erwinia amylovora</i>	ATCC49946	Gamma	3905	232	192
<i>Agrobacterium tumefaciens</i>	C58	Alpha	5674	538	203
<i>Clavibacter michiganensis</i> subsp. <i>michiganensis</i>	NCPPB382	Actino	3395	231	221
<i>Caulobacter crescentus</i>	CB15	Alpha	4016	332	225
<i>Brucella melitensis</i>	ATCC23457	Alpha	3311	197	208
<i>Rickettsia conori</i>	Malish7	Alpha	1268	21	157
<i>Escherichia coli</i>	536	Gamma	4938	352	188
<i>Yersinia pestis</i>	Antiqua	Gamma	4879	329	184
<i>Chlamydia trachomatis</i>	434/Bu	Chlamydiae	1038	13	136

Pseudomonas syringae). Nous avons recherché pour chaque souche bactérienne, le nombre total de protéines impliquées dans la signalisation à partir de la base de données Microbial Signal Transduction Database (<http://www.mistdb.com>) (Ulrich et Zhulin, 2010) et calculé manuellement leur QI (Tableau 1). Les bactéries ayant les QI les plus élevés sont *Caulobacter crescentus*, *Clavibacter michiganensis*, *Brucella melitensis*, *Stenotrophomonas maltophilia* et *Agrobacterium tumefaciens*. Par ailleurs, *E. coli*, *Yersinia pestis*, *Ralstonia solanacearum* et les bactéries des genres *Pseudomonas* et *Xanthomonas* présentent des QI élevés et similaires. Parmi les *Xanthomonas*, les souches de *X. oryzae* pv. *oryzae* ont des QI plus faibles et équivalents au QI de *Rickettsia conorii*, une bactérie intracellulaire obligatoire. Enfin, le QI le plus faible est celui de la bactérie intracellulaire *Chlamydia trachomatis*. Ces données indiquent que les organismes colonisant des niches écologiques diverses (sol, eau, hôte eucaryote) et des environnements instables présentent des systèmes de transduction assez complexes. Dans le cas inverse, les bactéries comme par exemple *Chlamydia trachomatis* et *Rickettsia conorii*, qui colonisent des environnements riches (l'intérieur des cellules eucaryotes) où les conditions sont stables, n'auraient besoin que des systèmes de transduction simples. Ces conclusions corroborent celles de Galperin (2005). Selon cet auteur, le nombre de protéines impliquées dans la signalisation est positivement corrélé à la taille du génome. Il est assez surprenant de noter que les cinq souches de *Xanthomonas* et les *Pseudomonas* possèdent des QI équivalents. Les *Pseudomonas* sont connues pour coloniser durablement différentes niches écologiques comme les végétaux, le sol, la neige et les cours d'eau (Morris *et al.*, 2008). Les *Xanthomonas* sont, quant à eux, pour l'instant connus pour être exclusivement associés aux végétaux. La valeur du QI des *Xanthomonas* qui est aussi élevée que celui des *Pseudomonas*, suggère fortement que ces bactéries qui sont associées aux plantes, auraient une distribution plus large que celle qui est connue aujourd'hui. Il n'est pas exclu que nous ne connaissions qu'une partie des niches écologiques colonisées par ces bactéries.

Pistes pour des études fonctionnelles

La caractérisation des répertoires de senseurs et d'adhésines couplée à l'étude des pressions de sélection régissant l'histoire évolutive des MCPs et des adhésines fournit des hypothèses à tester par des études fonctionnelles. La première piste concerne

l'étude de l'effet de la mutation d'un gène codant un MCP sur le comportement chimiotactique et la capacité de reconnaissance de l'hôte par les souches appartenant à des pathovars dont les répertoires ne se différencient que par un seul MCP. Ceci est le cas des souches des pathovars *campestris* et *incanae* de l'espèce *X. campestris*. En effet, les répertoires de ces pathovars se différencient par un seul MCP (XCC0324). Dans ce cas, il s'agira d'étudier *in vitro* le comportement chimiotactique de la souche mutée envers des exsudats et des extraits végétaux prélevés de *Brassica oleracea* (plante hôte des souches du pathovar *campestris*), de *Matthiola incana* (plante hôte des souches du pathovar *incanae*) et d'*Arabidopsis thaliana*. Par ailleurs, d'autres tests *in planta* permettraient de tester la capacité de survie et de maintien de la souche mutée dans le gène XCC0324 dans la phyllosphère et dans les tissus des feuilles de *B. oleracea* et *M. incana*. Un changement au niveau de la réponse chimiotactique chez la souche mutée du pathovar *campestris*, indiquerait que les MCPs sont impliqués dans la reconnaissance spécifique de l'hôte par la bactérie, et par conséquent, dans la spécificité d'hôte. La deuxième piste concerne la validation fonctionnelle par mutagenèse dirigée de la contribution du chimiotactisme dans le pouvoir pathogène des bactéries du genre *Xanthomonas*. Chez les bactéries, le chimiotactisme est impliqué dans l'interaction avec leurs hôtes (animal et/ou végétal). Chez *E. coli*, la mutation de *cheA*, *cheY* ou de *cheW* conduit à un phénotype non-mobile (Parkinson, 1976 ; Conley *et al.*, 1989). En outre, il a été démontré que les mutants de *R. solanacearum* affectés dans les gènes *cheA* ou *cheW*, sont non-mobiles, très diminués dans leur capacité de colonisation des racines de tomate alors qu'injectés directement dans la tige, leur agressivité est totale. Ils sont, de plus, faiblement compétitifs lors d'une co-inoculation avec le parent sauvage. Ces résultats montrent que chez *R. solanacearum*, une colonisation efficace de l'hôte dépend fortement du chimiotactisme (Yao et Allen, 2006). La même approche a été réalisée chez la souche 3937 de *Dickeya dadantii* (synonyme *Erwinia chrysanthemi*). Les gènes cibles qui ont été mutés sont des gènes impliqués dans la transduction du signal chimiotactique *cheW*, *cheB*, *cheY* et *cheZ* et dans la structure du moteur du flagelle *motA*. Il a été montré que tous les mutants affectés dans chacun de ces gènes, présentent une diminution de leur capacité à pénétrer leur plante hôte *A. thaliana*. En revanche, la diminution du pouvoir pathogène de ces mutants dépend du gène muté et de l'hôte infecté. Les mutants générés peuvent être ainsi classés en fonction de leur agressivité

comme suit : $motA < cheY < cheB = cheW < cheZ$. Tous ces résultats indiquent que le chimiotactisme et la mobilité jouent un rôle important dans le pouvoir pathogène de cette bactérie (Antúnez-Lamas *et al.*, 2009). Dans le cas des *Xanthomonas*, il serait intéressant d'examiner l'effet de la mutation de gènes impliqués dans le chimiotactisme dans le pouvoir pathogène (réponse chimiotactique, mobilité, capacité de colonisation et agressivité) sur la souche modèle du laboratoire CFBP4834 de *X. axonopodis* pv. *phaseoli* GL *fuscans*. Dans les souches séquencées de *Xanthomonas* spp., les gènes *cheA*, *cheB*, *cheW* sont présents en plusieurs copies (Thieme *et al.*, 2005). Par contre, les souches séquencées ne possèdent qu'une seule copie des gènes *cheY* et *cheZ*. Ces deux gènes constituent donc de bons candidats pour cette étude. Après vérification de l'abolition de la réponse chimiotactique chez le mutant, il serait intéressant de tester son pouvoir pathogène sur *P. vulgaris*. Pour cela, il faudrait examiner si après les inoculations par trempage ou par blessure, la souche mutée est encore capable de pénétrer dans les tissus de l'hôte et de provoquer des symptômes de maladie.

Concept d'espèce

Selon Ernst Mayr (1942), le concept d'espèce biologique tel qu'il est établi pour les eucaryotes, définit l'espèce comme un ensemble d'organismes interféconds et reproductivement isolés des autres organismes. Selon ces règles de taxonomie, une espèce bactérienne authentique serait définie par la similarité génomique de ses membres. Ainsi, la cohésion génomique qui existe au sein de l'espèce peut être apparentée à un isolement reproducteur limitant les flux de gènes entre des organismes divergents. Dans ce cas, la recombinaison homologe est l'un des mécanismes favorisant l'isolement génétique d'une espèce. En ce qui concerne les bactéries, l'échange de matériel génétique peut être considéré comme de la reproduction sexuée. Les échanges de gènes entre des bactéries phylogénétiquement éloignées sont importants dans l'évolution de ces organismes. Ainsi, le concept d'espèce biologique tel qu'il est définit pour les eucaryotes, n'est pas transposable directement pour les procaryotes. L'étude menée par Costechareyre *et al.* (2009) sur le complexe d'espèces *A. tumefaciens* a démontré que la fréquence de la recombinaison homologe ne chute pas brutalement entre des individus divergents, c'est-à-dire appartenant à des espèces génomiques différentes. Le rapport entre diversité des séquences et taux de

recombinaison ne constitue pas un indicateur fiable de la distance génétique et ne permet donc pas de générer des espèces. Les auteurs concluent que « la définition de l'espèce génomique chez les procaryotes n'est définitivement pas réconciliable avec le concept d'espèce biologique défini chez les eucaryotes ». Il n'existe pas de consensus quant à la définition de l'espèce procaryote. Selon Cohan (2002), les différences écologiques peuvent isoler les populations au plan spatiotemporel. Ainsi, les échanges de matériel génétique sont rares entre des individus colonisant des niches écologiques distinctes. Cohan (2002) a donc introduit la notion d'écotype qui regroupe des populations d'organismes colonisant les mêmes niches écologiques et dont la divergence est maintenue par des pressions de sélection liée à l'occupation d'une niche particulière. Dans le cas de notre modèle d'étude, l'espèce *X. axonopodis sensu* Vauterin *et al.* (1995), nous avons mis en évidence par les approches phylogénétiques et généalogiques, des signaux conflictuels entre les différents groupes constituant cette espèce, suite à des échanges de matériel génétique. Ces échanges de matériel se produisent *via* des recombinaisons entre les différents groupes et touchent aussi bien le core génome que le génome accessoire. Ces résultats montrent que ces groupes ne sont pas isolés. Dans notre modèle, la recombinaison homologue concerne les gènes de ménage et a été observée entre plusieurs sous-groupes à différents moments de l'évolution. Elle a sans doute également été associée à des transferts horizontaux de gènes associés à la virulence (ET3s) en provenance d'organismes phylogénétiquement éloignés (Hajri *et al.*, 2009).

Les échanges de matériels génétiques au sein de *X. axonopodis* concernent aussi bien des gènes de ménage que des gènes associés à l'interaction avec la plante hôte. Sur les 21 gènes associés à l'interaction avec la plante affectés par les transferts horizontaux de gènes, dix gènes codent des senseurs et des adhésines. Ainsi, les transferts horizontaux de gènes ont contribué au modelage des répertoires de gènes impliqués dans les phases précoces de l'interaction entre les bactéries et leurs hôtes. Par le biais de ces transferts, la gamme de plante hôte reconnue par la bactérie peut être modifiée. L'acquisition de nouvelles adhésines peut en effet favoriser la capacité des bactéries à se maintenir et à coloniser de nouvelles niches écologiques et/ou de nouvelles plantes hôtes. Morris *et al.* (2007) ont suggéré que les écosystèmes sauvages pourraient jouer

un rôle important dans l'acquisition de nouveaux déterminants pouvant améliorer le fitness, l'agressivité et le taux de survie des bactéries phytopathogènes. Dans le cas des *Xanthomonas* phytopathogènes, le rôle de ces niches n'est pas connu.

En 2009, Parkinson et ses collaborateurs ont séquencé le gène de ménage *gyrB* dans une collection comprenant 189 souches de *Xanthomonas* isolées de plantes d'intérêt agro-économique majeur et mineur et d'adventices indigènes de l'Inde. Ces auteurs ont montré que les espèces *X. euvesicatoria*, *X. fuscans* et *X. citri* présentent une faible diversité de séquences. En revanche, ces espèces sont phylogénétiquement proches des souches appartenant à 86 pathovars infectant diverses espèces végétales échantillonnées dans des régions arables en Inde. D'après les auteurs, la forte similarité de séquences au sein de ces groupes indique qu'un grand nombre de pathovars isolés en Inde se seraient diversifiés relativement récemment à partir de ces trois espèces. L'importance du nombre de pathovars isolés suggère que cette région est un centre de diversification de ces bactéries. Dans cette région, l'agriculture est pratiquée de façon extensive et la densité des adventices est élevée. Ces conditions favoriseraient la dissémination et la diversification des pathovars.

Dans le cadre de ce travail, nous nous sommes interrogés sur l'importance des étapes pré-infectieuses dans l'adaptation à l'hôte. Les résultats obtenus montrent que la perception de l'environnement est une fonction importante pour l'adaptation à l'hôte. Elle ne permet néanmoins pas d'expliquer totalement la spécificité parasitaire. Ce résultat est cohérent avec l'hypothèse émise dans notre étude des relations généalogiques entre les souches de *X. axonopodis*. Selon cette hypothèse, l'ancêtre des souches contemporaines aurait pu être un pathogène généraliste capable de coloniser une vaste gamme de plante hôtes. La domestication des plantes, l'agriculture puis l'intensification de l'agriculture se traduisant par des monocultures sur de vastes zones auraient conduit à la spécialisation des pathogènes. La diminution de la virulence observée dans les souches contemporaines se serait accompagnée d'une augmentation de l'agressivité par la sélection de variants (Stuckenbrock et McDonald, 2008). Ainsi, il est probable que certaines de nos souches ne soient pas encore parvenues à un stade très poussé de spécialisation. Le partage de répertoires de gènes codant des protéines

impliquées dans la perception de l'environnement entre souches montrant des spécialisations parasitaires différentes refléteraient la capacité des souches à coloniser une gamme d'hôte plus large que celle sur laquelle elles sont capables de provoquer des épidémies, leur permettant d'utiliser différentes composantes du milieu, ce qui peut se révéler utile dans les zones d'agriculture extensive.

Annexes

Annexe 1. Résultats des tests de vraisemblance des analyses de pressions de sélection effectuées avec le modèle *branch-site* du programme CODEML et appliquées aux haplotypes des orthologues codant les MCPs.

Nom du gène analysé ^a	modèles <i>Branch-site</i>	Branche testée ^b	Valeur de la vraisemblance	LRT ^c	$p(\chi^2)$ ^d
XCV0669	M2	<i>Xca</i>	-2740,16	0	1
	M2nul	<i>Xca</i>	-2740,16		
	M2	B100	-2740,40	0	1
	M2nul	B100	-2740,40		
	M2	<i>Xcc+ 8004</i>	-2740,30	0	1
	M2nul	<i>Xcc+ 8004</i>	-2740,30		
XCV1702	M2	<i>Xca</i>	-2780,62	0	1
	M2nul	<i>Xca</i>	-2780,62		
	M2	B100	-2778,78	0,09	0,76
	M2nul	B100	-2778,83		
	M2	<i>Xcc+8004</i>	-2780,62	0	1
	M2nul	<i>Xcc+8004</i>	-2780,62		
XCV1778	M2	<i>Xca</i>	-2484,34	0	1
	M2nul	<i>Xca</i>	-2484,34		
	M2	B100	-2484,34	0	1
	M2nul	B100	-2484,34		
	M2	<i>Xcc+8004</i>	-2481,78	0,43	0,51
	M2nul	<i>Xcc+8004</i>	-2482,00		
XCV1933	M2	<i>Xca</i>	-2922,60	0	1
	M2nul	<i>Xca</i>	-2922,60		
	M2	<u>B100</u> ^e	-2903,00	32,77	1E-08
	M2nul	B100	-2919,39		
	M2	<i>Xcc+8004</i>	-2922,60	0	1
	M2nul	<i>Xcc+8004</i>	-2922,60		
XCV1941	M2	<i>Xca</i>	-2312,89	0	1
	M2nul	<i>Xca</i>	-2312,89		
	M2	B100	-2313,78	0	1
	M2nul	B100	-2313,78		
	M2	<i>Xcc+8004</i>	-2313,78	0	1
	M2nul	<i>Xcc+8004</i>	-2313,78		
XCV1945	M2	<i>Xca</i>	-3205,19	0,05	0,83
	M2nul	<i>Xca</i>	-3205,21		
	M2	B100	-3194,39	0,55	0,46
	M2nul	B100	-3194,67		
	M2	<i>Xcc+8004</i>	-3205,21	0	1
	M2nul	<i>Xcc+8004</i>	-3205,21		
XCV1947	M2	<i>Xca</i>	-2832,83	0	1
	M2nul	<i>Xca</i>	-2832,83		
	M2	B100	-2832,74	0	1
	M2nul	B100	-2832,74		
	M2	<i>Xcc+8004</i>	-2831,14	0,36	0,55
	M2nul	<i>Xcc+8004</i>	-2831,32		
XCV1951	M2	<i>Xca</i>	-3411,13	0	1
	M2nul	<i>Xca</i>	-3411,13		
	M2	B100	-3411,28	0	1
	M2nul	B100	-3411,28		
	M2	<i>Xcc+ 8004</i>	-3411,28	-0,06	-

Suite Annexe 1

Nom du gène analysé ^a	modèles <i>Branch-site</i>	Branche testée ^b	Valeur de la vraisemblance	LRT ^c	$p(\chi^2)$ ^d
XCV1952	M2nul	<i>Xcc+</i> 8004	-3411,25		
	M2	<i>Xca</i>	-2114,14	4E-06	0,99
	M2nul	<i>Xca</i>	-2114,14		
	M2	B100	-2114,14	0	1
	M2nul	B100	-2114,14		
	M2	<i>Xcc</i>	-2112,58	0	1
	M2nul	<i>Xcc</i>	-2112,58		
XCV2044	M2	8004	-2114,14	0	1
	M2nul	8004	-2114,14		
	M2	<i>Xca</i>	-1460,04	0	1
	M2nul	<i>Xca</i>	-1460,04		
	M2	B100	-1460,04	0	1
	M2nul	B100	-1460,04		
	M2	<i>Xcc+</i> 8004	-1460,04	0	1
XCV3021	M2nul	<i>Xcc+</i> 8004	-1460,04		
	M2	<i>Xca</i>	-2727,51	0,50	0,48
	M2nul	<i>Xca</i>	-2727,76		
	M2	B100	-2727,76	0	1
	M2nul	B100	-2727,76		
	M2	<i>Xcc+</i> 8004	-2727,76	0	1
	M2nul	<i>Xcc+</i> 8004	-2727,76		
XCV3230	M2	<i>Xac</i>	-2708,82	0	1
	M2nul	<i>Xac</i>	-2708,82		
	M2	B100	-2708,82	0	1
	M2nul	B100	-2708,82		
	M2	<i>Xcc+</i> 8004	-2708,82	0	1
	M2nul	<i>Xcc+</i> 8004	-2708,82		
	XCV3338	M2	<i>Xca</i>	-2493,11	0
M2nul		<i>Xca</i>	-2493,11		
M2		B100	-2493,11	0	1
M2nul		B100	-2493,11		
M2		<i>Xcc+</i> 8004	-2491,12	0	1
M2nul		<i>Xcc+</i> 8004	-2491,12		

^a : nom de l'orthologue dans la souche 85-10 de *X. axonopodis* pv. *vesicatoria*.

^b *Xca* : orthologue du gène cible dans la souche 756C de *X. campestris* pv. *armoraciae* ; B100 : orthologue du gène cible dans la souche B100 de *X. campestris* pv. *campestris* ; *Xcc+*8004 : orthologues du gène cible identiques dans les souches ATCC33913 et 8004 de *X. campestris* pv. *campestris* et 8004 : orthologue du gène cible dans la souche 8004 de *X. campestris* pv. *campestris*.

^c likelihood ratio test. Les tests LRTs du modèle *branch-site* sont réalisés en testant l'hypothèse de neutralité (M2nul), $LRT = 2 \times (\ln M2 - \ln M2nul)$. Le nombre de degré de liberté du test est égal à 1.

^d : probabilité associée à la loi de Khideux.

^e : La séquence orthologue du gène XCV1933 présente dans la souche B100 de *X. campestris* pv. *campestris* a été soulignée en raison de l'existence de sept sites soumis à une pression de sélection positive identifiés dans le Tableau 3.

Annexe 2. Liste des haplotypes des orthologues de XCV1933, XCV3230 et de la région N-terminale de XCV1940 amplifiés par PCR dans des souches de *X. axonopodis*.

Nom de pathovar	Code ^a	Hôte d'isolement	XCV1933	XCV3230	XCV1940
<i>alfalfae</i>	3835	<i>Medicago sativa</i>	H17	H14	H14
	3836	<i>M. sativa</i>	H18	H14	-
	3837	<i>M. sativa</i>	H19	H15	H15
	7120	<i>M. sativa</i>	-	-	H15
	7121	<i>M. sativa</i>	-	-	-
<i>allii</i>	6107	<i>Allium fistulosum</i>	H25	H20	H17
	6358	<i>A. sativum</i>	H26	H23	H18
	6359	<i>A. cepa</i>	-	-	-
	6362	<i>A. cepa</i>	H27	H24	H17
	6364	<i>A. sativum</i>	H25	H20	H17
	6367	<i>A. cepa</i>	H28	H25	H19
	6369	<i>A. cepa</i>	H29	H23	H18
	6376	<i>A. cepa</i>	-	H23	H18
	6383	<i>A. cepa</i>	-	H26	H20
	6385	<i>A. cepa</i>	H30	H35	H21
<i>anacardii</i>	2913	<i>Mangifera indica</i>	-	-	H10
	2914	<i>M. indica</i>	-	-	H10
	7240	<i>Anacardium occidentale</i>	-	-	H23
	7241	<i>A. occidentale</i>	-	-	H23
	7242	<i>A. occidentale</i>	-	-	H23
	7243	<i>A. occidentale</i>	-	-	H23
<i>aurantifolii</i>	2866	<i>Citrus aurantiifolia</i>	-	-	H8
	2901	<i>C. limon</i>	-	-	H9
	3528	<i>C. limon</i>	-	H13	H9
	3529	<i>C. limon</i>	-	H13	-
	3541	<i>C. aurantiifolia</i>	-	-	-
<i>axonopodis</i>	4924	<i>Axonopus scoparius</i>	-	-	-
	5141	<i>A. scoparius</i>	-	-	-
<i>begoniae</i>	1421	<i>Begonia</i> sp.	H5	H2	H4
	2524	<i>Begonia</i> sp.	H11	H2	H4
	5676	<i>B. rugosa</i>	H22	H2	H4
	5677	<i>B. pendula</i>	H11	H2	H4
	5678	<i>Begonia</i> sp.	H11	H2	H4
<i>betae</i>	5852	<i>Beta vulgaris</i> .	-	-	-
<i>bilvae</i>	3136	<i>Aegle marmelos</i> .	-	-	-

Suite Annexe 2

Nom de pathovar	Code ^a	Hôte d'isolement	XCV1933	XCV3230	XCV1940
<i>citri</i>	1209	<i>C. grandis</i>	H4	H1	H2
	1814	<i>Citrus</i> sp.	H7	H1	H2
	2525	<i>C. limon</i>	H7	H1	H2
	2900	<i>Citrus</i> sp.	-	-	-
	3369	<i>C. aurantifolia</i>	H7	H1	-
	3530	<i>C. limon</i>	H16	H32	H2
	5280	<i>C. hystrix</i>	H7	H1	-
	5284	<i>Citrus</i> sp.	-	H33	-
	JK2-20*	<i>C. aurantifolia</i>	-	-	-
	JS582*	<i>C. aurantifolia</i>	-	-	-
	JJ60-1*	<i>C. aurantifolia</i>	-	-	-
	JF90-8*	<i>C. aurantifolia</i>	-	-	-
	LB302*	<i>C. aurantifolia</i> / <i>C. macrophylla</i>	-	-	-
	306	NA	H2	H1	H2
<i>citrumelo</i>	3114	<i>P. trifoliata</i> x <i>C. paradisi</i>	-	-	H12
	3371	NA	H15	H12	-
	3841	<i>Poncirus trifoliata</i> x <i>C. sinensis</i>	H20	H16	-
	3842	<i>P. trifoliata</i> x <i>C. paradisi</i>	H21	H16	-
	3843	<i>C. paradisi</i>	-	-	H1
<i>dieffenbachiae</i> GL.A	3132	<i>Dieffenbachia</i> sp.	H14	-	-
<i>dieffenbachiae</i> GL.C	3133	<i>Anthurium</i> sp.	-	H11	H13
	5688	<i>A. andreanum</i>	H23	H17	H13
	5691	<i>Anthurium</i> sp.	-	-	H13
<i>glycines</i>	1519	<i>Glycine hispida</i>	H6	H3	-
	2526	<i>G. hispida</i>	H12	H8	H7
	7119	<i>G. max</i>	-	H31	H22
<i>malvacearum</i>	2035	<i>Gossypium hirsutum</i>	H10	H7	H6
	2530	<i>G. hirsutum</i>	H10	H7	H6
	5700	<i>G. hirsutum</i>	-	H18	H6
	5701	<i>G. hirsutum</i>	-	-	H6
	5726	<i>G. barbadense</i>	-	-	H16
<i>mangiferaeindicae</i>	1716	<i>M. indica</i>	-	-	H5
	2915	<i>M. indica</i>	-	-	H5
	2933	<i>M. indica</i>	-	-	H5
	2935	<i>M. indica</i>	-	-	H11
	2939	<i>Schinus terebenthifolius</i>	-	-	H5

Suite Annexe 2

Nom de pathovar	Code ^a	Hôte d'isolement	XCV1933	XCV3230	XCV1940
	2940	<i>S. terebenthifolius</i>	-	-	H5
	7236	<i>M. indica</i>	-	-	H5
	7237	<i>S. terebenthifolius</i>	-	-	H5
	7238	<i>S. terebenthifolius</i>	-	-	-
	7239	<i>S. terebenthifolius</i>	-	-	H5
<i>manihotis</i>	1851	<i>Manihot esculenta</i>	H8	H6	-
	1860	<i>M. esculenta</i>	H9	H6	-
	1865	<i>M. esculenta</i>	-	-	-
	2603	<i>M. esculenta</i>	-	H10	-
	2624	<i>M. esculenta</i>	H9	H6	-
	6544	<i>M. esculenta</i>	-	-	-
<i>phaseoli</i> GL <i>fuscans</i>	1815	<i>Phaseolus</i> sp.	-	H5	H3
	1845	<i>Phaseolus</i> sp.	-	-	H3
	4834	<i>Phaseolus vulgaris</i>	H3	H5	H3
	6165	<i>P. vulgaris</i>	-	H22	H3
	6166	<i>P. vulgaris</i>	-	H5	-
	6167	<i>Phaseolus</i> sp.	-	H5	-
	6960	<i>P. vulgaris</i>	-	H5	H3
	6965	<i>P. vulgaris</i>	-	H5	H3
	6969	<i>P. vulgaris</i>	-	H5	-
	6970	<i>Phaseolus</i> sp.	-	-	H3
	6971	<i>Phaseolus</i> sp.	-	-	H3
	6975	<i>Phaseolus</i> sp.	-	H5	H3
	6976	<i>Phaseolus</i> sp.	-	H5	H3
	6979	<i>P. vulgaris</i>	-	-	H3
<i>phaseoli</i> GL1	412	<i>P. vulgaris</i>	-	-	-
	6164	<i>P. vulgaris</i>	-	H21	-
	6546	<i>P. vulgaris</i>	H13	H9	-
	6982	<i>P. vulgaris</i>	-	-	-
	6983	<i>P. vulgaris</i>	-	-	-
	6984	<i>P. vulgaris</i>	-	H27	-
	6985	<i>P. vulgaris</i>	-	-	-
<i>phaseoli</i> GL2	6988	<i>P. vulgaris</i>	-	-	-
	6990	<i>P. vulgaris</i>	-	H28	-
	6991	<i>P. vulgaris</i>	-	H29	-
<i>phaseoli</i> GL3	6992	<i>P. vulgaris</i>	-	-	-

Suite Annexe 2

Nom de pathovar	Code ^a	Hôte d'isolement	XCV1933	XCV3230	XCV1940
	6994	<i>P. vulgaris</i>	-	-	-
	6996	<i>P. vulgaris</i>	-	H30	-
	6993	<i>P. vulgaris</i>	-	-	-
<i>ricini</i>	5863	<i>Ricinus communis</i>	H24	H19	-
	5864	<i>R. communis</i>	H24	H34	-
	5865	<i>R. communis</i>	-	H19	-
	6541	<i>R. communis</i>	-	H19	-
	6542	<i>R. communis</i>	-	-	-
<i>spondiae</i>	2547	<i>Spondias dulcis</i>	Absent	-	-
<i>vasculorum</i>	1289	<i>S. officinarum</i>	-	-	-
	5696	<i>Thysanolaena maxima</i>	-	-	-
	5822	<i>S. officinarum</i>	-	-	-
	5823	<i>S. officinarum</i>	-	-	-
<i>vesicatoria</i>	75-3	<i>L. esculentum</i>	H1	H4	-
	1604	<i>Capsicum annuum</i>	H1	H4	-
	2484	<i>Lycopersicon esculentum</i>	-	-	-
	5594	<i>L. esculentum</i>	H1	H4	-
	5618	<i>C. annuum</i>	H1	H4	H1
	6817	NA	-	-	-
	6864	<i>C. frutescens</i>	-	-	-
<i>vignicola</i>	7110	<i>Vigna unguiculata</i>	-	-	-
	7111	<i>V. sinensis</i>	-	-	-
	7112	<i>V. unguiculata</i>	-	-	-
	7113	<i>V. unguiculata</i>	-	-	-
	7115	<i>V. sinensis</i>	-	-	-

^a : numéro CFBP des souches, à l'exception des souches indiquées par * qui nous ont été fournies par Olivier Pruvost, Cirad, La Réunion.

- : aucun signal exploitable n'a été obtenu

Annexe 3. Résultats des tests de vraisemblance des analyses de pressions de sélection effectuées avec le modèle *branch-site* du programme CODEML et appliquées aux haplotypes des orthologues de XCV1933 et identification des sites probablement soumis à sélection positive.

Haplotype testé	modèles <i>Branch-site</i>	Valeur de la vraisemblance	LRT ^a	$p(\chi^2)$ ^b	codon sous pression de sélection
H19	M2	-11917,60	145,01	2,13E-33	-
	M2nul	-11990,11			
H17	M2	-11999,36	-4265,63	-	-
	M2nul	-9866,55			
H26	M2	-9646,31	3889,02	0	-
	M2nul	-11590,82			
H29	M2	-12018,29	-115,370	-	-
	M2nul	-11960,60			
H11	M2	-11192,75	1593,48	0	-
	M2nul	-11989,49			
H7	M2	-11967,94	-19,48	-	-
	M2nul	-11958,20			
H2	M2	-10798,89	2268,18	0	-
	M2nul	-11932,98			
H4	M2	-11917,33	111,81	3,92E-26	-
	M2nul	-11973,24			
H24	M2	-11930,78	147,68	5,54E-34	S 578 ($p=0,980^*$)
	M2nul	-12004,62			
H20	M2	-11651,73	364,69	2,67E-81	-
	M2nul	-11834,08			
H28	M2	-11847,50	43,48	4,26E-11	-
	M2nul	-11869,24			
H15	M2	-11785,74	94,62	2,30E-22	-
	M2nul	-11833,05			
H1	M2	-11832,33	-83,89	-	-
	M2nul	-11790,38			
H18	M2	-11804,39	74,30	6,72E-18	-
	M2nul	-11841,54			
H27	M2	-11841,01	-78,49	-	-
	M2nul	-11801,76			
H30	M2	-11783,27	160,55	8,60E-37	-
	M2nul	-11863,54			
H21	M2	-11836,91	-119,57	-	-
	M2nul	-11777,13			
H25	M2	-11753,62	148,45	3,78E-34	-
	M2nul	-11827,85			
H22	M2	-11913,95	-160,02	-	-
	M2nul	-11833,93			
H5	M2	-11782,65	173,19	1,49E-39	-
	M2nul	-11869,25			
H23	M2	-11806,52	33,17	8,41E-09	-
	M2nul	-11823,11			
H13	M2	-11856,58	-5,09	-	-
	M2nul	-11854,05			
H8	M2	-11647,72	421,15	1,37E-93	-
	M2nul	-11858,29			
H9	M2	-11851,92	10,20	0,001	-

Suite Annexe 3

Haplotype testé	modèles Branch-site	Valeur de la vraisemblance	LRT ^a	$p(\chi^2)$ ^b	codon sous pression de sélection
H3	M2nul	-11857,02			
	M2	-11893,12	31,88	1,64E-08	-
	M2nul	-11909,06			
H14	M2	-11964,21	-413,24	-	-
	M2nul	-11757,58			
H16	M2	-11787,47	332,61	2,59E-74	-
	M2nul	-11953,78			
H16	M2	-11846,77	141,00	1,61E-32	-
	M2nul	-11917,27			
H10	M2	-11888,42	0,49	0,48	-
	M2nul	-11888,67			
H12	M2	-11728,45	363,82	4,15E-81	-
	M2nul	-11910,36			

^a : likelihood ratio test. Les LRTs du modèle branch-site ont été réalisés en testant l'hypothèse de neutralité (M2nul), LRT= 2 x (lnM2 – ln M2nul). Le nombre de degré de liberté du test est égal à 1.

^b : probabilité associée à la loi de Khideux.

* indique que la probabilité p que le codon soit soumis à une pression de sélection est significative p>0.95.

Annexe 4. Résultats des tests de vraisemblance des analyses de pressions de sélection effectuées avec le modèle *branch-site* du programme CODEML et appliquées aux haplotypes des orthologues de la région N-terminale du gène XCV1940 et identification des sites probablement soumis à sélection positive.

Haplotype testé	modèles <i>Branch-site</i>	Valeur de la vraisemblance	LRT ^a	$p(\chi^2)$ ^b	codon sous pression de sélection
H5	M2	-6425,23	31,63	1,86E-08	-
	M2nul	-6441,04			
H11	M2	-6391,99	145,07	2,08E-33	-
	M2nul	-6464,52			
H2	M2	-5950,44	880,91	1,38E-193	-
	M2nul	-6390,90			
H16	M2	-6132,90	432,01	5,94E-96	-
	M2nul	-6348,91			
H7	M2	-6449,49	-710,52	-	-
	M2nul	-6094,23			
H22	M2	-6403,19	-31,99	-	-
	M2nul	-6387,19			
H6	M2	-6371,77	-182,22	-	-
	M2nul	-6280,66			
H3	M2	-6325,81	115,72	5,45E-27	-
	M2nul	-6383,67			
H9	M2	-6295,40	89,90	2,50E-21	-
	M2nul	-6340,35			
H23	M2	-6391,84	-171,01	-	-
	M2nul	-6306,33			
H8	M2	-6024,91	-74,84	-	-
	M2nul	-5987,48			
H10	M2	-6335,53	13,19	2,80E-4	-
	M2nul	-6342,13			
H4	M2	-6305,21	-24,95	-	-
	M2nul	-6292,73			
H13	M2	-6188,17	253,76	3,93E-57	N 239 ($p=0,960^*$)
	M2nul	-6315,05			
H12	M2	-6418,90	-152,31	-	-
	M2nul	-6342,74			
H21	M2	-6349,72	-71,14	-	-
	M2nul	-6314,15			
H15	M2	-6072,57	661,02	8,94E-146	-
	M2nul	-6403,09			
H19	M2	-6433,50	-10,82	-	-
	M2 nul	-6428,08			
H17	M2	-6392,07	43,61	3,99E-11	-
	M2nul	-6413,88			
H1	M2	-6345,89	178,91	8,38E-41	-
	M2nul	-6435,34			
H14	M2	-6315,91	-37,28	-	-
	M2nul	-6297,27			
H18	M2	-6342,58	201,97	7,73E-46	-
	M2nul	-6443,57			
H20	M2	-6410,01	-59,90	-	-

M2nul -6380,06

^a : likelihood ratio test. Les LRTs du modèle branch-site ont été réalisés en testant l'hypothèse de neutralité (M2nul), $LRT = 2 \times (\ln M2 - \ln M2nul)$. Le nombre de degré de liberté du test est égal à 1.

^b : probabilité associée à la loi de Khidoux.

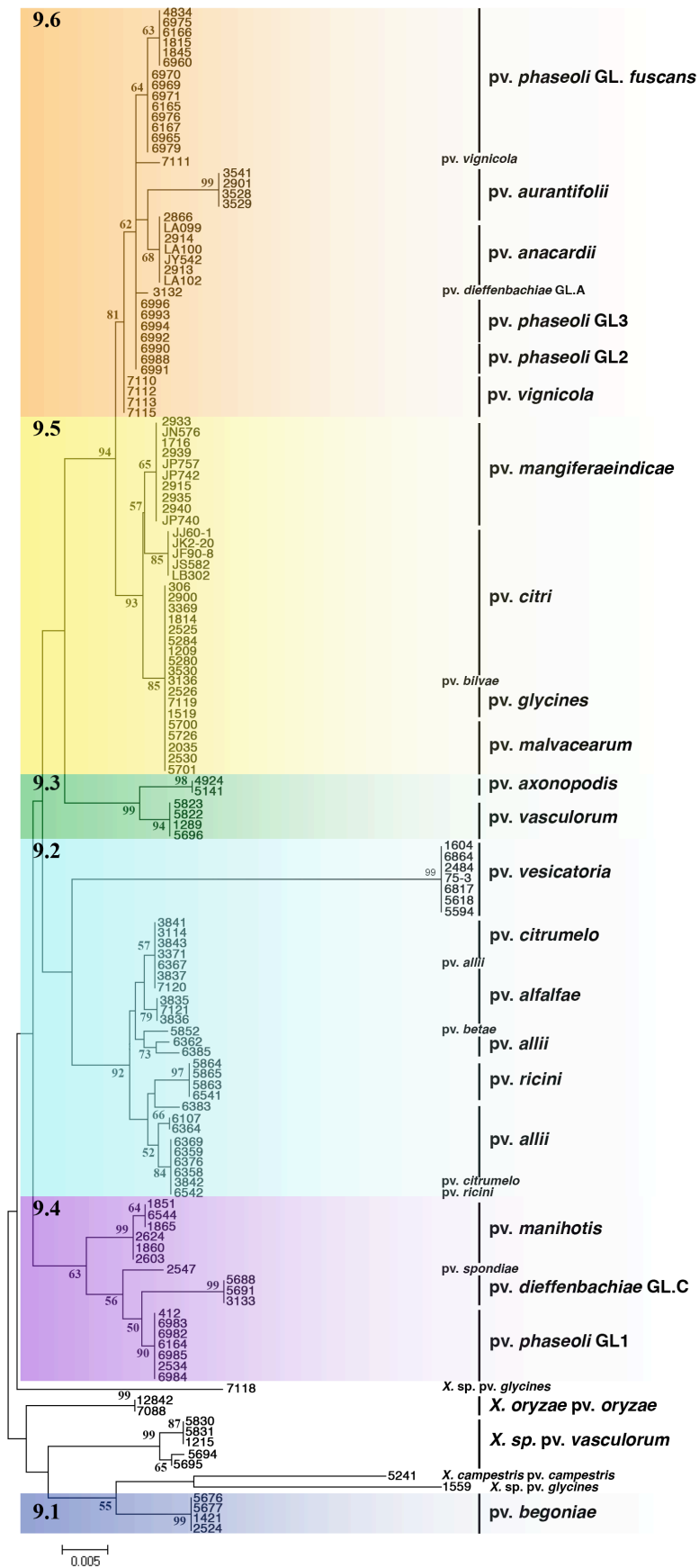
* indique que la probabilité p que le codon soit soumis à une pression de sélection est significative $p > 0.95$.

Annexe 5. Liste des souches complémentaires aux 131 souches utilisées dans l'article 2.

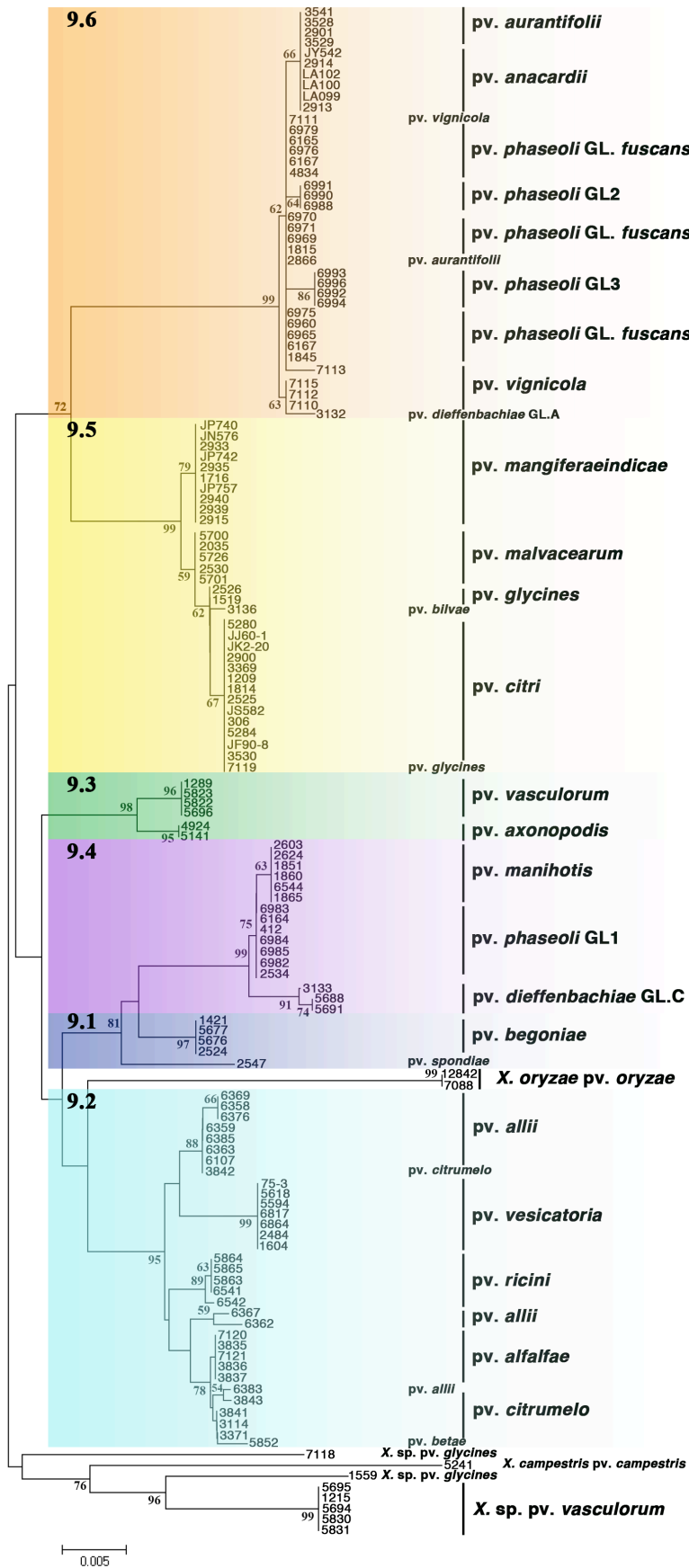
Nomenclature	Code CFBP de la souche	Hôte d'isolement	origine géographique	Années d'isolement	Group rep- PCR ^a
<i>X. axonopodis</i> pv. <i>vascolorum</i>	1290	<i>Saccharum officinarum</i>	La Réunion	1971	9.3
<i>X. vascolorum</i> pv. <i>vascolorum</i>	1216	<i>S. officinarum</i>	Kenya	1970	11
<i>X. vascolorum</i> pv. <i>holcicola</i>	2543	<i>Sorghum vulgare</i>	N ^{velles} Zélande	1969	11
<i>X. sp. glycines</i>	1559	<i>Glycine hispida</i>	Zimbabwe	1974	
	7118	<i>G. javanica</i>	Zimbabwe	1963	
<i>X. sp. vascolorum</i>	1215	<i>S. officinarum</i>	Kenya	NA	
	5694	<i>Zea Mays</i>	La Réunion	NA	
	5695	<i>Tripsacum laxum</i>	La Réunion	NA	
	5830	<i>S. officinarum</i>	Malaisie	NA	
	5831	<i>S. officinarum</i>	Afrique du Sud	NA	

^a : selon Rademaker *et al.* (2005).

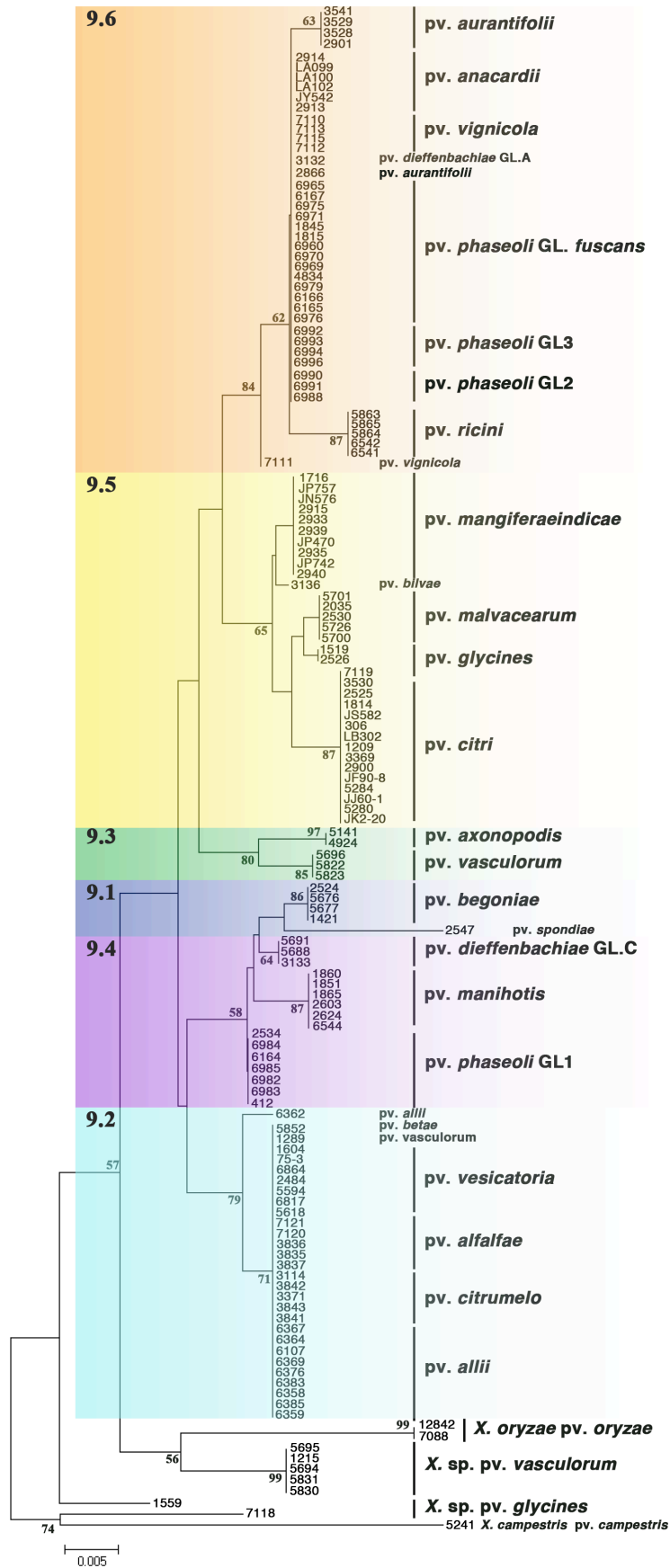
Annexe 6. Arbres phylogénétiques obtenus par Neighbor Joining avec 1000 simulations à partir des séquences partielles des gènes *atpD* (6.1), *dnaK* (6.2), *efp* (6.3), *fyuA* (6.4), *glnA* (6.5), *gyrB* (6.6) et *rpoB* (6.7) dans 138 souches de *Xanthomonas* spp. (*sensu* Vauterin *et al.*, 1995).



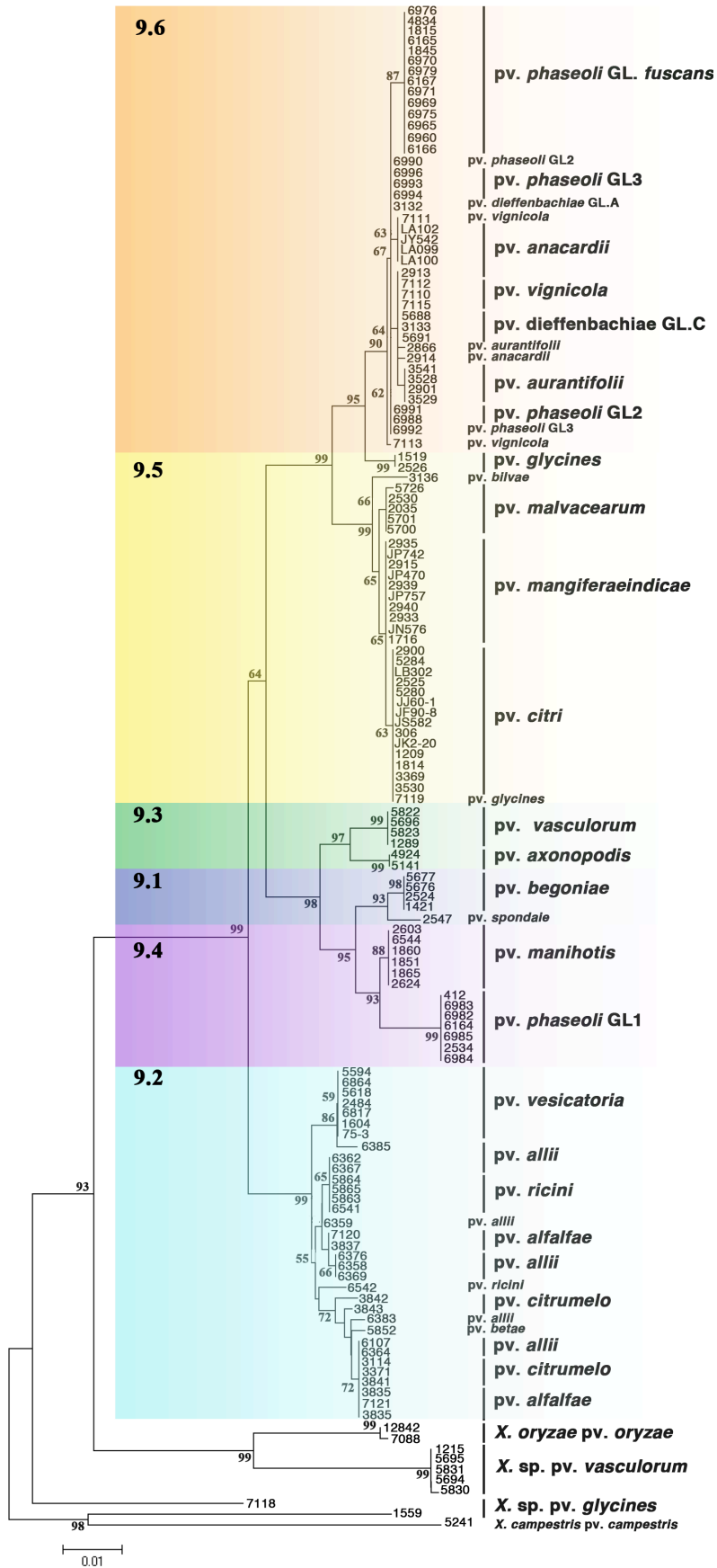
Annexe 6.1. (*atpD*)



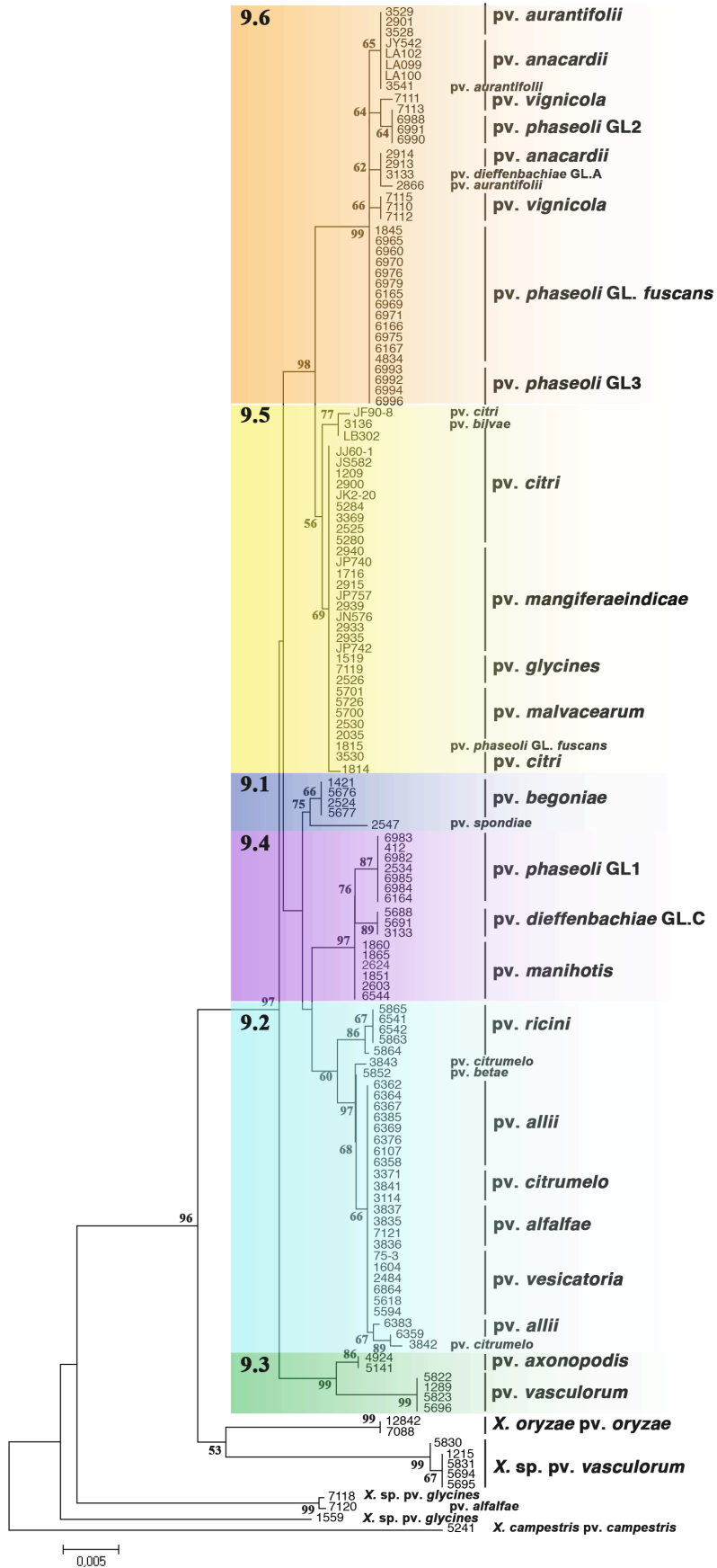
Annexe 6.2. (*dnaK*)



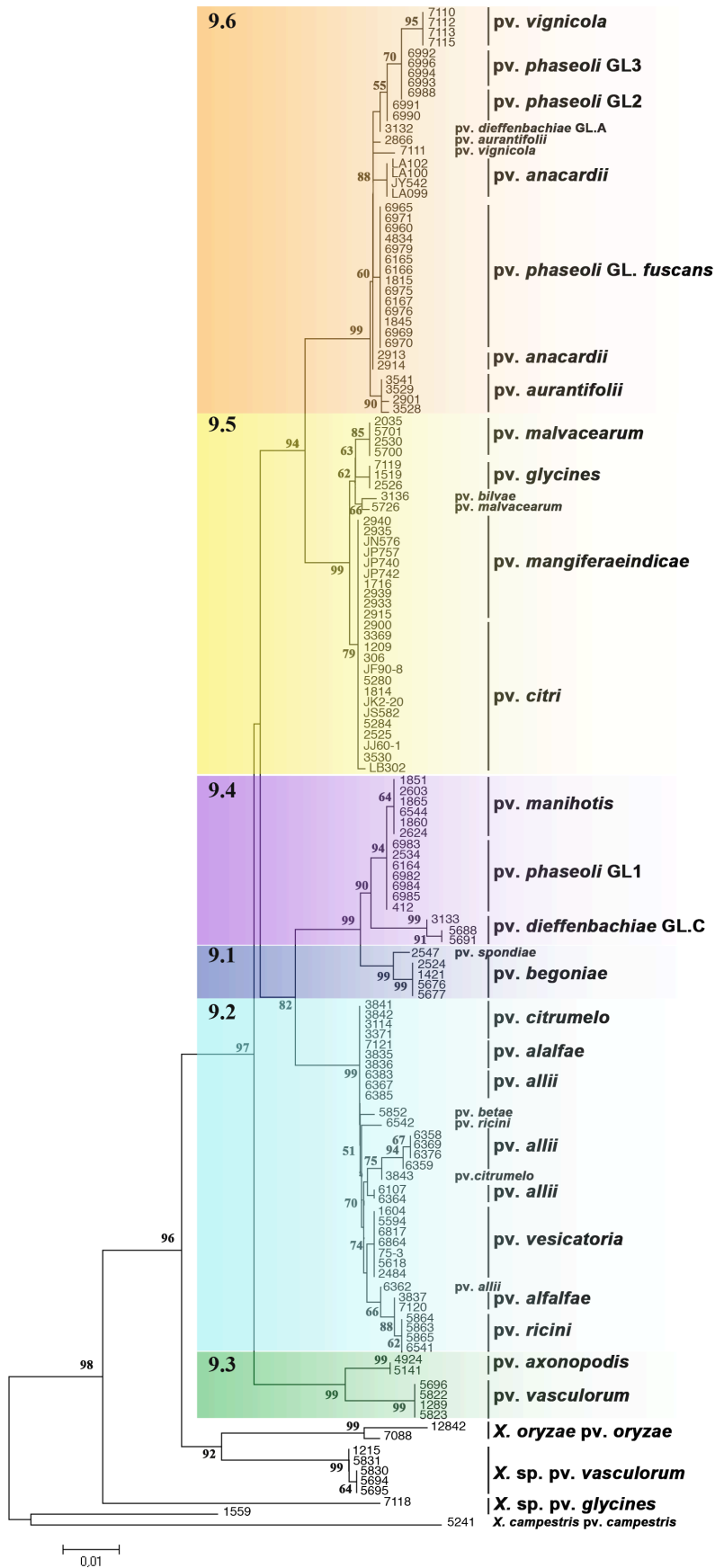
Annexe 6.3. (efp)



Annexe 6.4. (*fyuA*)



Annexe 6.5. (*glnA*)



Annexe 6.6. (*gyrB*)

Annexe 7. Diversité des pathovars en fonction des gènes de ménage étudiés. Pour plus de détails se référer à l'annexe 6.

Nom du pathovar	Gène de ménage						
	<i>atpD</i>	<i>dnaK</i>	<i>efp</i>	<i>fyuA</i>	<i>glnA</i>	<i>gyrB</i>	<i>rpoD</i>
<i>phaseoli</i> GL <i>fuscans</i>	2 groupes*				1 groupe + 1815		1 groupe + 6167 + 6165
<i>phaseoli</i> GL2				1 groupe + 6990		1 groupe + 6988	
<i>phaseoli</i> GL3				1 groupe + 6992			
<i>vignicola</i>	1 groupe + 7111	1 groupe + 7111 + 7113	1 groupe + 7111	1 groupe + 7111 + 7113	1 groupe + 7111 + 7113	1 groupe + 7111	1 groupe + 7111 + 7113
<i>anacardii</i>				1 groupe + 2913 + 2914	1 groupe + 2913 + 2914	1 groupe + 2913 + 2914	
<i>aurantifolii</i>	1 groupe + 2866	1 groupe + 2866	1 groupe + 2866	1 groupe + 2866	1 groupe + 2866	1 groupe + 2866	1 groupe + 2866
<i>citri</i>	2 groupes				1 groupe + JF90-8		2 groupes
<i>glycines</i>		1 groupe + 7119	1 groupe + 7119	1 groupe + 7119			1 groupe + 1519
<i>malvacearum</i>				1 groupe + 5726		1 groupe + 5726	1 groupe + 5700
<i>citrumelo</i>	1 groupe + 3842	1 groupe + 3842 + 3843		1 groupe + 3842 + 3843	1 groupe + 3842 + 3843	1 groupe + 3843	1 groupe + 3842
<i>alfalfae</i>	2 groupes			2 groupes	1 groupe +7120	2 groupes	1 groupe + 3835
<i>allii</i>	3 groupes + 6383 + 6367	3 groupes + 6383	1 groupe + 6362	3 groupes + 6385 + 6359 + 6383	1 groupe + 6383 + 6359	3 groupes + 6359 + 6362	3 groupes + 6367 + 6385
<i>ricini</i>	1 groupe + 6542	1 groupe + 6542		1 groupe + 6542		1 groupe + 6542	1 groupe + 6542
<i>manihotis</i> <i>dieffenbachiae</i> GLC	2 groupes					1 groupe + 3133	
<i>begoniae</i>		1 groupe + 3133					1 groupe + 2524

* un groupe est un ensemble de plus de deux souches.

Annexe 8. Types de séquences par gène et profils alléliques définis pour les 131 souches de *X. axonopodis* de la collection.

Nom de la souche	code ^a	Numéro des types de séquences (ST) par gène							ST pour les 7 loci
		<i>atpD</i>	<i>dnaK</i>	<i>efp</i>	<i>fyuA</i>	<i>glnA</i>	<i>gyrB</i>	<i>rpoD</i>	
<i>X. axonopodis</i> pv. <i>alfalfae</i>	3837	13	15	3	24	1	23	19	32
	7120	13	15	3	24	27	23	19	33
	3835	16	15	3	28	1	17	6	39
	3836	16	15	3	28	1	17	19	40
	7121	16	15	3	28	1	17	19	40
<i>X. axonopodis</i> pv. <i>allii</i>	6367	13	19	3	23	1	17	28	35
	6359	17	17	3	25	16	31	20	42
	6358	17	18	3	26	1	30	27	43
	6369	17	18	3	26	1	30	27	43
	6376	17	18	3	26	1	30	27	43
	6107	21	17	3	28	1	29	25	49
	6364	21	17	3	28	1	29	25	49
	6362	23	25	5	23	1	3	20	53
	6383	24	22	3	29	15	17	20	54
	6385	25	17	3	21	1	17	29	55
<i>X. axonopodis</i> pv. <i>anacardii</i>	2913	12	4	13	3	9	16	15	27
	2914	12	4	13	12	9	16	15	28
	JY542*	12	4	13	10	12	36	15	29
	LA099*	12	4	13	10	12	36	15	29
	LA100*	12	4	13	10	12	36	15	29
<i>X. axonopodis</i> pv. <i>aurantifolii</i>	LA102*	12	4	13	10	12	36	15	29
	3529	3	4	15	13	12	22	14	12
	3541	3	4	15	13	12	22	14	12
	2901	3	4	15	13	12	15	14	13
	3528	3	4	15	13	12	21	14	14
<i>X. axonopodis</i> pv. <i>axonopodis</i>	2866	12	5	13	32	23	14	13	30
	4924	18	31	16	15	22	25	21	45
	5141	18	31	16	15	22	25	21	45
<i>X. axonopodis</i> pv. <i>begoniae</i>	1421	5	11	6	16	14	8	4	17
	5676	5	11	6	16	14	8	4	17
	5677	5	11	6	16	14	8	4	17
	2524	5	11	6	16	14	8	10	18
<i>X. axonopodis</i> pv. <i>betae</i>	5852	19	24	3	30	19	27	29	46
<i>X. axonopodis</i> pv. <i>bilvae</i>	3136	2	32	12	8	13	20	2	11
<i>X. axonopodis</i> pv. <i>citri</i> A	306	2	7	1	1	2	2	2	3
	1209	2	7	1	1	2	2	2	3
	2525	2	7	1	1	2	2	2	3
	2900	2	7	1	1	2	2	2	3
	3369	2	7	1	1	2	2	2	3
	3530	2	7	1	1	2	2	2	3
	5280	2	7	1	1	2	2	2	3
	5284	2	7	1	1	2	2	2	3
	1814	2	7	1	1	21	2	2	4
	<i>X. axonopodis</i> pv. <i>citri</i> A [*] /A ^w	JF90-8*	29	7	1	1	11	2	39
JJ60-1*		29	7	1	1	2	2	39	62
JK2-20*		29	7	1	1	2	2	39	62
JS582*		29	7	1	1	2	2	39	62
LB302*		29	7	1	1	13	37	39	63
<i>X. axonopodis</i> pv. <i>citrumelo</i>	3114	13	16	3	28	1	17	16	31
	3371	13	16	3	28	1	17	16	31
	3841	13	16	3	28	1	17	16	31

Nom de la souche	code ^a	<i>atpD</i>	<i>dnaK</i>	<i>efp</i>	<i>fyuA</i>	<i>glnA</i>	<i>gyrB</i>	<i>rpoD</i>	ST pour les 7 gènes
	3843	13	21	3	22	18	24	16	34
	3842	17	17	3	31	17	17	20	41
<i>X. axonopodis</i> pv. <i>dieffenbachiae</i> GL.A	3132	14	1	13	2	9	18	17	36
<i>X. axonopodis</i> pv. <i>dieffenbachiae</i> GL.C	3133	15	13	9	3	24	19	18	37
	5688	15	14	9	3	24	4	22	38
	5691	15	14	9	3	24	4	22	38
<i>X. axonopodis</i> pv. <i>glycines</i>	7119	2	7	1	1	2	9	38	5
	2526	2	9	11	9	2	9	11	9
	1519	2	9	11	9	2	9	2	10
<i>X. axonopodis</i> pv. <i>malvacearum</i>	2035	2	8	4	7	2	11	9	6
	2530	2	8	4	7	2	11	9	6
	5701	2	8	4	7	2	11	9	6
	5700	2	8	4	7	2	11	23	7
	5726	2	8	4	6	2	26	9	8
<i>X. axonopodis</i> pv. <i>mangiferaeindicae</i>	1716	6	2	2	4	2	2	5	19
	2915	6	2	2	4	2	2	5	19
	2933	6	2	2	4	2	2	5	19
	2935	6	2	2	4	2	2	5	19
	2939	6	2	2	4	2	2	5	19
	2940	6	2	2	4	2	2	5	19
	JN576	6	2	2	4	2	2	5	19
	JP740	6	2	2	4	2	2	5	19
	JP742	6	2	2	4	2	2	5	19
	JP757	6	2	2	4	2	2	5	19
<i>X. axonopodis</i> pv. <i>manihotis</i>	1851	8	6	7	17	10	10	7	22
	6544	8	6	7	17	10	10	7	22
	1865	8	6	7	17	10	10	8	23
	1860	9	6	7	17	10	10	7	24
	2603	9	6	7	17	10	10	7	24
	2624	9	6	7	17	10	10	7	24
<i>X. axonopodis</i> pv. <i>phaseoli</i> GL <i>fuscans</i>	1845	7	5	13	11	5	5	6	20
	4834	7	5	13	11	5	5	6	20
	6166	7	5	13	11	5	5	6	20
	6960	7	5	13	11	5	5	6	20
	6975	7	5	13	11	5	5	6	20
	1815	7	5	13	11	2	5	6	21
	6165	22	5	13	11	5	5	13	50
	6167	22	5	13	11	5	5	26	51
	6965	22	5	13	11	5	5	6	52
	6969	22	5	13	11	5	5	6	52
	6970	22	5	13	11	5	5	6	52
	6971	22	5	13	11	5	5	6	52
	6976	22	5	13	11	5	5	6	52
	6979	22	5	13	11	5	5	6	52
<i>X. axonopodis</i> pv. <i>phaseoli</i> GL1	412	10	12	8	18	6	12	12	25
	6546	10	12	8	18	6	12	12	25
	6164	10	12	8	18	6	12	12	25
	6982	10	12	8	18	6	12	12	25
	6983	10	12	8	18	6	12	12	25
	6984	10	12	8	18	6	12	12	25
	6985	10	12	8	18	6	12	12	25
<i>X. axonopodis</i> pv. <i>phaseoli</i> GL2	6988	26	27	13	2	7	33	31	64
	6990	26	27	13	2	7	34	31	56

Nom de la souche	code ^a	<i>atpD</i>	<i>dnaK</i>	<i>efp</i>	<i>fyuA</i>	<i>glnA</i>	<i>gyrB</i>	<i>rpoD</i>	ST pour les 7 gènes
<i>X. axonopodis</i> pv. <i>phaseoli</i> GL3	6991	26	27	13	2	7	34	31	56
	6992	26	28	13	2	5	33	32	57
	6993	26	28	13	2	5	33	32	57
	6994	26	28	13	2	5	33	32	57
	6996	26	28	13	2	5	33	32	57
<i>X. axonopodis</i> pv. <i>ricini</i>	6542	17	23	14	27	4	32	30	44
	5863	20	20	14	23	4	28	24	47
	5865	20	20	14	23	4	28	24	47
	6541	20	20	14	23	4	28	24	47
	5864	20	20	14	23	20	28	24	48
<i>X. axonopodis</i> pv. <i>spondiae</i>	2547	11	26	10	19	26	13	37	26
<i>X. axonopodis</i> pv. <i>vasculorum</i>	1289	4	10	3	14	3	7	3	15
	5696	4	10	17	14	3	7	3	16
	5822	4	10	17	14	3	7	3	16
	5823	4	10	17	14	3	7	3	16
<i>X. axonopodis</i> pv. <i>vesicatoria</i>	1604	1	3	3	20	1	1	1	1
	2484	1	3	3	20	1	1	1	1
	5594	1	3	3	20	1	1	1	1
	5618	1	3	3	20	1	1	1	1
	6817	1	3	3	20	1	1	1	1
	6864	1	3	3	20	1	1	1	1
	75-3	1	3	3	20	1	1	36	2
<i>X. axonopodis</i> pv. <i>vignicola</i>	7110	27	30	13	3	8	35	33	58
	7112	27	30	13	3	8	35	33	58
	7115	27	30	13	3	8	35	33	58
	7113	27	29	13	5	7	35	35	59
	7111	28	5	18	10	25	6	34	60

^a numéro CFBP des souches, à l'exception des souches indiquées par * qui nous ont été fournies par Olivier Pruvost, Cirad, La Réunion.

Annexe 9. Liste des 41 séquences recombinantes détectées par six ou sept des sept méthodes de détection de la recombinaison implémentées dans le logiciel RDP et identification des souches de type parental et des sites de recombinaison.

Nom du pathovar	N° CFBP de la souche recombinante	Sites de recombinaison		Gènes affectés	N° CFBP des souches de type parental	Méthode n'ayant pas détecté de la recombinaison	Nb de méthodes ayant détecté de la recombinaison	
<i>citrumelo</i>	3114	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
	3371	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
	3841	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
	3842	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
	3843	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
<i>alfalfae</i>	3835	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
	3836	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
	3837	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
	7121	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
<i>betae</i>	5852	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
<i>allii</i>	6107	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
	6358	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
	6359	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
	6362	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
	6364	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
	6367	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
	6369	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
	6376	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
	6385	324	5454	<i>atpD, rpoD</i>	6976/5594	GENECONV	6	
	<i>ricini</i>	5863	1333	5331	<i>efp, rpoD</i>	6976/5594	GENECONV	6
		5864	1333	5331	<i>efp, rpoD</i>	6976/5594	GENECONV	6
5865		1333	5331	<i>efp, rpoD</i>	6976/5594	GENECONV	6	
6541		1333	5331	<i>efp, rpoD</i>	6976/5594	GENECONV	6	
6542		1333	5331	<i>efp, rpoD</i>	6976/5594	GENECONV	6	
<i>manihotis</i>	1851	4871	5343	<i>rpoD</i>	5691/2547	-	7	
	1860	4871	5343	<i>rpoD</i>	5691/2547	-	7	
	1865	4871	5343	<i>rpoD</i>	5691/2547	-	7	
	2603	4871	5343	<i>rpoD</i>	5691/2547	-	7	
	2624	4871	5343	<i>rpoD</i>	5691/2547	-	7	
	6544	4871	5343	<i>rpoD</i>	5691/2547	-	7	
<i>phaseoli</i> GL1	412	4871	5343	<i>rpoD</i>	5691/2547	-	7	
	6164	4871	5343	<i>rpoD</i>	5691/2547	-	7	
	6546	4871	5343	<i>rpoD</i>	5691/2547	-	7	
	6982	4871	5343	<i>rpoD</i>	5691/2547	-	7	
	6983	4871	5343	<i>rpoD</i>	5691/2547	-	7	
	6984	4871	5343	<i>rpoD</i>	5691/2547	-	7	
6985	4871	5343	<i>rpoD</i>	5691/2547	-	7		
<i>dieffenbachiae</i> GL.C	3133	2043	2896	<i>fyuA</i>	6546/7110	-	7	
	5688	2043	2896	<i>fyuA</i>	6546/7110	-	7	
	5691	2043	2896	<i>fyuA</i>	6546/7110	-	7	
<i>alfalfae</i>	7120	2862	3958	<i>fyuA,gyrB</i>	3837/5618	RDP	6	

Références bibliographiques

Ah-You, N., Gagnevin, L., Chiroleu, F., Jouen, E., Neto, J.R. & Pruvost, O. (2007). Pathological variations within *Xanthomonas campestris* pv. *mangiferaeindicae* support its separation into three distinct pathovars that can be distinguished by amplified fragment length polymorphism. *Phytopathology* **97**, 1568-1577.

Ah-You, N., Gagnevin, L., Grimont, P.A., Brisse, S., Nesme, X., Chiroleu, F., Bui Thi Ngoc, L., Jouen, E., Lefeuvre, P., Vernière, C. & Pruvost, O. (2009). Polyphasic characterization of xanthomonads pathogenic to members of the *Anacardiaceae* and their relatedness to species of *Xanthomonas*. *Int J Syst Evol Microbiol* **59**, 306-318.

Aini, L.Q., Hirata, H. & Tsuyumu, S. (2010). A TonB-dependent transducer is responsible for regulation of pathogenicity-related genes in *Xanthomonas axonopodis* pv. *citri*. *J Gen Plant Pathol* **76**, 132-142.

Alavi, S.M., Poussier, S. & Manceau, C. (2007). Characterization of ISXax1, a novel insertion sequence restricted to *Xanthomonas axonopodis* pv. *phaseoli* (variants *fuscans* and *nonfuscans*) and *Xanthomonas axonopodis* pv. *vesicatoria*. *Appl Environ Microbiol* **73**, 1678-1682.

Alavi, S.M., Sanjari, S., Durand, F., Brin, C., Manceau, C. & Poussier, S. (2008). Assessment of the genetic diversity of *Xanthomonas axonopodis* pv. *phaseoli* and *Xanthomonas fuscans* subsp. *fuscans* as a basis to identify putative pathogenicity genes and a type III secretion system of the SPI-1 family by multiple suppression subtractive hybridizations. *Appl Environ Microbiol* **74**, 3295-3301.

Alfano, J.R. & Collmer, A. (1997). The type III (Hrp) secretion pathway of plant pathogenic bacteria: trafficking harpins, Avr proteins, and death. *J Bacteriol* **179**, 5655-5662.

Almeida, N.F., Yan, S., Lindeberg, M., Studholme, D.J., Schneider, D.J., Condon, B., Liu, H., Viana, C.J., Warren, A., Evans, C., Kemen, E., Maclean, D., Angot, A., Martin, G.B., Jones, J.D., Collmer, A., Setubal, J.C. & Vinatzer, B.A. (2009). A draft genome sequence of *Pseudomonas syringae* pv. *tomato* T1 reveals a type III effector repertoire significantly divergent from that of *Pseudomonas syringae* pv. *tomato* DC3000. *Mol Plant Microbe Interact* **22**, 52-62.

Amano, A. (2010). Bacterial adhesins to host components in periodontitis. *Periodontol 2000* **52**, 12-37.

Amin, D.N. & Hazelbauer, G.L. (2010). The chemoreceptor dimer is the unit of conformational coupling and transmembrane signaling. *J Bacteriol* **192**, 1193-1200.

Amme, S., Rutten, T., Melzer, M., Sonsmann, G., Vissers, J.P., Schlesier, B. & Mock, H.P. (2005). A proteome approach defines protective functions of tobacco leaf trichomes. *Proteomics* **5**, 2508-2518.

- Ancillo, G., Hoegen, E. & Kombrink, E. (2003).** The promoter of the potato chitinase C gene directs expression to epidermal cells. *Planta* **217**, 566-576.
- Antúnez-Lamas, M., Cabrera-Ordóñez, E., López-Solanilla, E., Raposo, R., Trelles-Salazar, O., Rodríguez-Moreno, A. & Rodríguez-Palenzuela, P. (2009).** Role of motility and chemotaxis in the pathogenesis of *Dickeya dadantii* 3937 (ex *Erwinia chrysanthemi* 3937). *Microbiology* **155**, 434-442.
- Armitage, J.P. (1999).** Bacterial tactic responses. *Adv Microb Physiol* **41**, 229-289.
- Atkinson, S. & Williams, P. (2009).** Quorum sensing and social networking in the microbial world. *J R Soc Interface* **6**, 959-978.
- Attree, O. & Attree, I. (2001).** A second type III secretion system in *Burkholderia pseudomallei*: who is the real culprit? *Microbiology* **147**, 3197-3199.
- Bafana, B. (2008).** Trade-East Africa: Going Bananas to Fight Poverty and Hunger. IPS InterPress Service News Agency (IPS). [http:// ipsnews.net/news.asp?idnews=44210](http://ipsnews.net/news.asp?idnews=44210). Accessed October 2008.
- Bai, F., Branch, R.W., Nicolau, D.V., Pilizota, T., Steel, B.C., Maini, P.K. & Berry, R.M. (2010).** Conformational spread as a mechanism for cooperativity in the bacterial flagellar switch. *Science* **327**, 685-689.
- Beattie, G.A. & Lindow, S.E. (1995).** The secret life of foliar bacterial pathogens on leaves. *Annu Rev Phytopathol* **33**, 145-172.
- Beattie, G.A. & Lindow, S.E. (1999).** Bacterial colonization of leaves: a spectrum of strategies. *Phytopathology* **89**, 353-359.
- Beattie, G.A. (2002).** Leaf surface waxes and the process of leaf colonization by microorganisms. In *Phyllosphere microbiology*, S.E. Lindow, E.I. Hecht-Poinar and V. Elliott, eds, (St. Paul, Minnesota), APS Press, 3-26.
- Belas, R., Zhulin, I.B. & Yang, Z. (2008).** Bacterial signaling and motility: sure bets. *J Bacteriol* **190**, 1849-1856.
- Bergey, D.H., Harrison, F.C., Breed, R.S., Hammer, B.W. & Huntoon, F.M. (1923).** *Bergey's manual of determinative bacteriology*, 1st Edition, (Baltimore, Williams and Wilkins, Co), 442 pp.
- Bijlsma, J.J.E. & Groisman, E.A. (2003).** Making informed decisions: regulatory interactions between two-component systems. *Trends Microbiol* **11**, 359-366.
- Blakeman, J.P. (1972).** Effect of plant age on inhibition of *Botrytis cinerea* spores by bacteria on beetroot leaves. *Physiol Plant Pathol* **2**, 143-152.

- Blanvillain, S., Meyer, D., Boulanger, A., Lautier, M., Guynet, C., Denancé, N., Vasse, J., Lauber, E. & Arlat, M. (2007).** Plant carbohydrate scavenging through TonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. *PLoS One* **2**, e224.
- Boldog, T., Li, M. & Hazelbauer, G. (2006).** Nanodiscs separate chemoreceptor oligomeric states and reveal their signaling properties. *Proc Natl Acad Sci U. S. A.* **103**, 11509–11514.
- Boureau, T., ElMaarouf-Bouteau, H., Garnier, A., Brisset, M.N., Perino, C., Pucheu, I. & Barny, M.A. (2006).** DspA/E, a type III effector essential for *Erwinia amylovora* pathogenicity and growth in planta, induces cell death in host apple and nonhost tobacco plants. *Mol Plant Microbe Interact* **19**, 16-24.
- Bradbury, J.F. (1984).** *Xanthomonas* Dowson. 1939. In: *Bergey's Manual of Systematic bacteriology*, Vol. 1 Krieg NR, Holt JG, eds (Baltimore, Williams & Wilkins), 199-210.
- Bren, A. & Eisenbach, M. (2000).** How signals are heard during bacterial chemotaxis: protein-protein interactions in sensory signal propagation. *J Bacteriol* **182**, 6865-6873.
- Brencic, A. & Winans, S.C. (2005).** Detection of and response to signals involved in host-microbe interactions by plant-associated bacteria. *Microbiol Mol Biol Rev* **69**, 155-194.
- Brown, K. (2001).** Florida fights to stop citrus canker. *Science* **292**, 2275-2276.
- Bryant, D. & Moulton, V. (2004).** Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* **21**, 255-265.
- Bui Thi Ngoc, L., Vernière, C., Jouen, E., Ah-You, N., Lefeuvre, P., Chiroleu, F., Gagnevin, L. & Pruvost, O. (2010).** Amplified fragment length polymorphism and multilocus sequence analysis-based genotypic relatedness among pathogenic variants of *Xanthomonas citri* pv. *citri* and *Xanthomonas campestris* pv. *bilvae*. *Int J Syst Evol Microbiol* **60**, 515-525.
- Butterworth, J. & McCartney, H.A. (1991).** The dispersal of bacteria from leaf surfaces by water splash. *J Appl Bacteriol* **71**, 484-496.
- Büttner, D. & Bonas, U. (2010).** Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiol Rev* **34**, 107-133.
- Cao, H., Baldini, R.L. & Rahme, L.G. (2001).** Common mechanisms for pathogens of plants and animals. *Annu Rev Phytopathol* **39**, 259-284.
- Cafati, C.R. & Saettler, A.W. (1980).** Role of nonhost species as alternate inoculum sources of *Xanthomonas phaseoli*. *Plant Disease* **64**, 194-196.

Cangelosi, G.A., Freitag, N.E & Buckley, M.R. (2004). From outside to inside: environmental microorganisms as human pathogens. A report of the American Academy of Microbiology, Washington, DC, USA, <http://www.asm.org/ASM/files/ccLibraryFiles/FILENAME/000000001402/EnvironMicroBW>, 18.

Chan, J.W.Y.F. & Goodwin, P.H. (1999). Differentiation of *Xanthomonas campestris* pv. *phaseoli* from *Xanthomonas campestris* pv. *phaseoli* var. *fuscans* by PFGE and RFLP. *Eur J Plant Pathol* **105**, 867-878.

Chen, S.L., Hung, C-S., Xu, J., Reigstad, C.S., Magrini, V., Sabo, A., Blasiar, D., Bieri, T., Meyer, R.R., Ozersky, P., Armstrong, J.R., Fulton, R.S., Latreille, J.P., Spieth, J., Hooton, T.M., Mardis, E.R., Hultgren, S.J., Gordon, J.L. (2006). Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. *Proc Natl Acad Sci U S A* **103**, 5977-5982.

Chen, S.L., Hung, C.S., Pinkner, J.S., Walker, J.N., Cusumano, C.K., Li, Z., Bouckaert, J., Gordon, J.I. & Hultgren, S.J. (2009). Positive selection identifies an in vivo role for FimH during urinary tract infection in addition to mannose binding. *Proc Natl Acad Sci U S A* **106**, 22439-22444.

Christensen, H., Kuhnert, P., Olsen, J.E. & Bisgaard, M. (2004). Comparative phylogenies of the housekeeping genes *atpD*, *infB* and *rpoB* and the 16S rRNA gene within the *Pasteurellaceae*. *Int J Syst Evol Microbiol* **54**, 1601-1609.

Ciesiolka, L.D., Hwin, T., Gearlds, J.D., Minsavage, G.V., Saenz, R., Bravo, M., Handley, S.M., Conover, H., Zhang, J., Caporgno, J., Phengrasamy, N.B., Toms A.O., Stall, R.E. & Whalen, M.C. (1999). Regulation of expression of avirulence gene *avrRxv* and identification of a family of host interaction factors by sequence analysis of *avrBsT*. *Mol Plant Microbe Interact* **12**, 35-44.

Cohan, F.M. (2002). Sexual isolation and speciation in bacteria. *Genetica* **116**, 359-370.

Conley, M.P., Wolfe, A.J., Blair, D.F. & Berg, H.C. (1989). Both CheA and CheW are required for reconstitution of chemotactic signaling in *Escherichia coli*. *J Bacteriol* **171**, 5190-5193.

Cornelis, G.R. (2006). The type III secretion injectisome. *Nat Rev Microbiol* **4**, 811-825.

Costechareyre, D., Bertolla, F. & Nesme, X. (2009). Homologous recombination in *Agrobacterium*: potential implications for the genomic species concept in bacteria. *Mol Biol Evol* **26**, 167-176.

Cotter, S.E., Surana, N.K. & St Geme, J.W. (2005). Trimeric autotransporters: a distinct subfamily of autotransporter proteins. *Trends Microbiol.* **13**, 199-205.

Cowan, M.M. (1999). Plant products as antimicrobial agents. *Clin Microbiol Rev* **12**, 564-582.

Croft, K.P.C., Juttner, F. & Slusarenko, A.J. (1993). Volatile products of the lipoxygenase pathway evolved from *Phaseolus vulgaris* (L.) leaves inoculated with *Pseudomonas syringae* pv *phaseolicola*. *Plant Physiol* **101**, 13-24.

da Silva, A.C., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida, N.F., Alves, L.M., do Amaral, A.M., Bertolini, M.C., Camargo, L.E., Camarotte, G., Cannavan, F., Cardozo, J., Chambergo, F., Ciapina, L.P., Cicarelli, R.M., Coutinho, L.L., Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J., Ferreira, R.C., Ferro, M.I., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite, R.P., Lemos, E.G., Lemos, M.V., Locali, E.C., Machado, M.A., Madeira, A.M., Martinez- Rossi, N.M., Martins, E.C., Meidanis, J., Menck, C.F., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, H.A., Rossi, A., Sena, J.A., Silva, C., de Souza, R.F., Spinola, L.A., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. & Kitajima, J.P. (2002). Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417**, 459-463.

Darrasse, A., Bureau, C., Samson, R., Morris, C. & Jacques, M-A. (2007). Contamination of bean seeds by *Xanthomonas axonopodis* pv. *phaseoli* associated with low bacterial densities in the phyllosphere under field and greenhouse conditions. *Eur J Plant Pathol* **119**, 203-215.

Darrasse, A., Darsonval, A., Boureau, T., Brisset, M-N., Durand, K. & Jacques, M-A. (2010). Transmission of plant-pathogenic bacteria by nonhost seeds without induction of an associated defense reaction at emergence. *Appl Environ Microbiol* **76**, 6787-6796.

Darsonval, A., Darrasse, A., Meyer, D., Demarty, M., Durand, K., Bureau, C., Manceau, C. & Jacques, M-A. (2008). The Type III secretion system of *Xanthomonas fuscans* subsp. *fuscans* is involved in the phyllosphere colonization process and in transmission to seeds of susceptible beans. *Appl Environ Microbiol* **74**, 2669-2678.

Darsonval, A., Darrasse, A., Durand, K., Bureau, C., Cesbron, S. & Jacques, M-A. (2009). Adhesion and fitness in the bean phyllosphere and transmission to seed of *Xanthomonas fuscans* subsp. *fuscans*. *Mol Plant Microbe Interact* **22**, 747-757.

Dawson, W.J. (1939). On the systematic position and generic names of the Gram negative bacterial plant pathogens. *Zentralbl Bakteriol Parasitenkd Infektionskr Hyg Abt II* **100**, 177-193.

Derridj, S. (1996). Nutrients on the leaf surface. In *Aerial plant surface microbiology*, C.E. Morris, P.C. Nicot, C. Nguyen-The, eds. (New York, Plenum Press), 25-42.

Didelot, X. (2010). Based analysis of bacterial population structure. In *Bacterial population genetics in infectious disease*, D.A. Robinson, D. Falush and E.J. Feil, eds (Hoboken, New Jersey, John Wiley & sons, Inc), 37-60.

Didelot, X. & Falush, D. (2007). Inference of bacterial microevolution using multilocus sequence data. *Genetics* **175**, 1251-1266.

Diggle, S.P., Crusz, S.A. & Cámara, M. (2007). Quorum sensing. *Curr Biol* **17**, R907-R910.

Dow, J.M., Crossman, L., Findlay, K., He, Y-Q., Feng, J-X. & Tang, J-L. (2003). Biofilm dispersal in *Xanthomonas campestris* is controlled by cell-cell signaling and is required for full virulence to plants. *Proc Natl Acad Sci U S A* **100**, 10995-11000.

Dye, D.W. & Lelliot, R.A. (1974). Genus II. *Xanthomonas*. In *Bergey's Manual of Determinative Bacteriology*, R.E. Buchanan and N.E. Gibbons eds (Baltimore, Williams and Wilkins, Co).

Falke, J.J. & Hazelbauer, G.L. (2001). Transmembrane signaling in bacterial chemoreceptors. *Trends Biochem Sci* **26**, 257-265.

Falush, D. (2009). Toward the use of genomics to study microevolutionary change in bacteria. *PLoS Genet* **5**, e1000627.

Fargier, E. (2007). L'étude de la pathologie de *Xanthomonas campestris* et de la structure génétique de ses pathovars a permis l'amélioration de la détection du pathogène dans les semences de Brassicacées. Angers, Université d'Angers, pp. 236.

Fargier, E. & Manceau, C. (2007). Pathogenicity assays restrict the species *Xanthomonas campestris* into three pathovars and reveal nine races within *Xanthomonas campestris* pv. *campestris*. *Plant Pathol* **56**, 805-818.

Feil, E., Zhou, J., Smith, J.M., & Spratt, B.G. (1996). A comparison of the nucleotide sequences of the *adk* and *recA* genes of pathogenic and commensal *Neisseria* species: Evidence for extensive interspecies recombination within *adk*. *J Mol Evol* **43**, 631-640.

Feil, H., Feil, W.S., Chain, P., Larimer, F., Dibartolo, G., Copeland, A., Lykidis, A., Trong, S., Nolan, M., Goltsman, E., Thiel, J., Malfatti, S., Loper, J.E., Lapidus, A., Detter, J.C., Lands, M., Richardson, P.M., Kyrpides, N.C., Ivanova, N. & Lindow, S. (2005). Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000. *Proc Natl Acad Sci U S A* **102**, 11064-11069.

Felsenstein, J. (2005). PHYLIP (phylogeny inference package) version 3.63. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Fiala, V., Glad, C., Martin, M., Jolivet, E. & Derridj, S. (1990). Occurrence of soluble carbohydrates on the phylloplane of maize (*Zea mays* L.): variations in relation to leaf heterogeneity and position on the plant. *New Phytol* **115**, 609-615.

Flor, H.H. (1971). Current status of the gene-for-gene concept. *Annu Rev Phytopathol* **9**, 275-296.

Foussard, M., Cabantus, S., Pedelacq, J., Guiller, V., Tranier, S., Mourey, L., Birck, C. & Samama, J. (2001). The molecular puzzle of two-component signaling cascades. *Microbes Infect* **3**, 417-424.

Gagnevin, L. & Pruvost, O. (2001). Epidemiology and control of mango bacterial black spot. *Plant Disease* **85**, 928-935.

Galperin, M.Y. (2004). Bacterial signal transduction network in a genomic perspective. *Environ Microbiol* **6**, 552-567.

Galperin, M.Y. (2005). A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. *BMC Microbiol* **5**, 35.

Galperin, M.Y. & Gomelsky, M. (2005). Bacterial signal transduction modules: from genomics to biology. *ASM News* **71**, 323-333.

Gent, D.H., Al-Saadi, A., Gabriel, D.W., Louws, F.J., Ishimaru, C.A. & Schwartz, H.F. (2005). Pathogenic and genetic relatedness among *Xanthomonas axonopodis* pv. *allii* and other pathovars of *X. axonopodis*. *Phytopathology* **95**, 918-925.

Gevers, D., Vandepoele, K., Simillon, C. & Van de Peer, Y. (2004). Gene duplication and biased functional retention of paralogs in bacterial genomes. *Trends Microbiol* **12**, 148-154.

Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt B.G., Coenye, T., Feil E.J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F.L. & Swings, J. (2005). Opinion : Reevaluating prokaryotic species. *Nature Reviews Microbiology* **3**, 733-739.

Gilbertson, R.L. & Maxwell, D.P. (1992). Common bacterial blight of bean. In: Diseases of vegetables and oil seed crop, 18-39.

Gooderham, W.J. & Hancock, R.E.W. (2009). Regulation of virulence and antibiotic resistance by two-component regulatory systems in *Pseudomonas aeruginosa*. *FEMS Microbiol Rev* **33**, 279-294.

Goodwin, P.H. & Sopher C.R. (1994). Brown pigmentation of *Xanthomonas campestris* pv. *phaseoli* associated with homogentistic acid. *Can. J. Microbiol.* **40**, 8-34.

Green, I.R., Tocoli, F.E., Lee, S.H., Nihei, K. & Kubo, I. (2007). Molecular design of anti-MRSA agents based on the anacardic acid scaffold. *Bioorg Med Chem* **15**, 6236-6241.

- Gudesblat, G.E., Torres, P.S. & Vojnov, A.A. (2009).** *Xanthomonas campestris* overcomes *Arabidopsis* stomatal innate immunity through a DSF cell-to-cell signal-regulated virulence factor. *Plant Physiol* **149**, 1017-1027.
- Guidot, A., Prior, P., Schoenfeld, J., Carrère, S., Genin, S. & Boucher, C. (2007).** Genomic structure and phylogeny of the plant pathogen *Ralstonia solanacearum* inferred from gene distribution analysis. *J Bacteriol* **189**, 377-387.
- Gunasekera, T.S. & Paul, N.D. (2007).** Ecological impact of solar ultraviolet-B (UV-B:320-290nm) radiation on *Corynebacterium aquaticum* and *Xanthomonas* sp. colonization on tea phyllosphere in relation to blister blight disease incidence in the field. *Lett Appl Microb* **44**, 513-519.
- Hacker, J. & Carniel, E. (2001).** Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. *EMBO Rep* **2**, 376-381.
- Hajri, A. (2009).** Détermination des répertoires d'effecteurs de type III chez *Xanthomonas* spp. et analyse de leur rôle dans la spécificité d'hôte. Angers, Université d'Angers, pp. 182.
- Hajri, A., Brin, C., Hunault, G., Lardeux, F., Lemaire, C., Manceau, C., Boureau, T., Poussier, S. (2009).** A «Repertoire for Repertoire» hypothesis: repertoires of type three effectors are candidate determinants of host specificity in *Xanthomonas*. *PLoS One* **4**, e6632.
- Hall, T.A. (1999).** BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-98.
- Hamza, A.A., Robène-Soustrade, I., Jouen, E., Gagnevin, L., Lefeuvre, P., Chiroleu, F. & Pruvost, O. (2010).** Genetic and pathological diversity among *Xanthomonas* strains responsible for bacterial spot on tomato and pepper in the southwest indian ocean region. *Plant Disease* **94**, 993-999.
- Hanage, W.P., Fraser, C. & Spratt, B.G. (2005).** Fuzzy species among recombinogenic bacteria. *BMC Biol* **3**, 6.
- Hauben, L., Vauterin, L., Swings, J. & Moore, E.R.B. (1997).** Comparison of 16S ribosomal DNA sequences of all *Xanthomonas* species. *Int J Syst Bacteriol* **47**, 328-335.
- Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C-G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M., Shinagawa, H. (2001).** Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* **8**, 11-22.
- Hayward, A.C. (1993).** The hosts of *Xanthomonas*. In *Xanthomonas*, J.G. Swings, and E.L. Civerolo, eds (London, United Kingdom, Chapman and Hall), 1-119.

- Hazelbauer, G.L., Falke, J.J. & Parkinson, J.S. (2007).** Bacterial chemoreceptors: high-performance signaling in networked arrays. *Trends Biochem Sci* **33**, 9-19.
- He, Y-W., Ng, A.Y-J., Xu, M., Lin, K., Wang, L-H., Dong, Y-H. & Zhang, L-H. (2007).** *Xanthomonas campestris* cell-cell communication involves a putative nucleotide receptor protein Clp and a hierarchical signalling network. *Mol Microbiol* **64**, 281-292.
- He, Y-W. & Zhang, L-H. (2008).** Quorum sensing and virulence regulation in *Xanthomonas campestris*. *FEMS Microbiol Rev* **32**, 842-857.
- He, Y-W., Boon, C., Zhou, L. & Zhang, L-H. (2009).** Co-regulation of *Xanthomonas campestris* virulence by quorum sensing and a novel two-component regulatory system RavS/RavR. *Mol Microbiol* **71**, 1464-1476.
- Hildebrand D.C., Palleroni, N.G. & Schroth, M.N. (1990).** Deoxyribonucleic relatedness of 24 xanthomonad strains representing 23 *Xanthomonas campestris* pathovars and *Xanthomonas fragariae*. *J Appl Bacteriol* **68**, 263-269.
- Hirano, S.S., Clayton, M.K. & Upper, C.D. (1994).** Estimation of and temporal changes in means and variances of populations of *Pseudomonas syringae* on snap bean leaflets. *Phytopathology* **84**, 934-940.
- Hirano, S.S. & Upper, C.D. (2000).** Bacteria in the leaf ecosystem with emphasis on *Pseudomonas syringae* –a pathogen, ice nucleus, and epiphyte. *Microbiol Mol Biol Rev* **64**, 624-653.
- Holmes, D.E., Nevin, K.P. & Lovley, D.R. (2004).** Comparison of 16S rRNA, *nifD*, *recA*, *gyrB*, *rpoB* and *fusA* genes within the family *Geobacteraceae* fam. nov. *Int J Syst Evol Microbiol* **54**, 1591-1599.
- Hsiao, Y-M., Fang, M-C., Sun, P-F. & Tseng, Y-H. (2009).** Clp and RpfF up-regulate transcription of *pelA1* gene encoding the major pectate lyase in *Xanthomonas campestris* pv. *campestris*. *J Agric Food Chem* **57**, 6207-6215.
- Huang, D.L, Tang, D.J, Liao, Q., Li, X.Q., He, Y.Q., Feng, J.X., Jiang, B.L., Lu, G.T., & Tang, J.L. (2009).** The Zur of *Xanthomonas campestris* is involved in hypersensitive response and positively regulates the expression of the *hrp* cluster via *hrpX* but not *hrpG*. *Mol Plant Microbe*, 321-329.
- Huson, D.H. & Bryant, D. (2006).** Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**, 254-267.
- Jacques, M-A., Josi, K., Darrasse, A. & Samson, R. (2005).** *Xanthomonas axonopodis* pv. *phaseoli* var. *fuscans* is aggregated in stable biofilm population sizes in the phyllosphere of field-grown beans. *Appl Environ Microbiol* **71**, 2008-2015.

Joardar, V., Lindeberg, M., Jackson, R.W., Selengut, J., Dodson, R., Brinkac, L.M., Daugherty, S.C., DeBoy, R., Durkin, A.S., Giglio, M.G., Madupu, R., Nelson, W.C., Rosovitz, M. J., Sullivan, S., Haft, D.H., Creasy, T., Davidsen, T., Zafar, N., Zhou, L., Halpin, R., Holley, T., Khouri, H., Feldblyum, T., White, O., Fraser, C.M., Chatterjee, A.K., Cartinhour, S., Schneider, D.J., Mansfield, J., Collmer, A. & Buell, C.R. (2005). Whole-genome sequence analysis of *Pseudomonas syringae* pv. *phaseolicola* 1448A reveals divergence among pathovars in genes involved in virulence and transposition. *J Bacteriol* **187**, 6488-6498.

Jones, J.B., Lacy, G.H., Bouzar, H., Stall, R.E. & Schaad, N.W. (2004). Reclassification of the xanthomonads associated with bacterial spot disease of tomato and pepper. *Syst Appl Microbiol* **27**, 755-762.

Jordan, I.K., Rogozin, I.B., Wolf, Y.I. & Koonin, E.V. (2002). Microevolutionary genomics of bacteria. *Theor Popul Biol* **61**, 435-447.

Jukes, T.H. & Cantor, C.R. (1969). Evolution of protein molecules. In Mammalian protein metabolism, Munro ed (New York, Academic press), 21-132.

Kaiser, W.J. & Vakili, N.G. (1978). Insect transmission of pathogenic xanthomonads to bean and cowpea in Puerto Rico. *Phytopathology* **68**, 1057-1063.

Kamoun, S., Kamdar, H.V., Tola E. & Kado, C.I. (1992). Incompatible interactions between crucifers and *Xanthomonas campestris* involve a vascular hypersensitive response: role of the *hrpX* locus. *Mol Plant Microbe Interact* **5**, 22-33.

Kaprelyants, A.S., Suleimenov, M.K., Sorokina, A.D., Deborin, G.A., El-Registan, G.I., Stoyanovich, F.M., Lille, Y.E. & Ostrovsky, D.N. (1987). Structural-functional changes in bacterial and model membranes induced by phenolic lipids. *Biol Membr* **4**, 254-261.

Karamanoli, K., Menkissoglu-Spiroudi, U., Bosabalidis, A.M., Vokou, D. & Constantinidou, H-I.A. (2005). Bacterial colonization of the phyllosphere of nineteen plant species and antimicrobial activity of their leaf secondary metabolites against leaf associated bacteria. *Chemoecology* **15**, 59-67.

Kearney, B. & Staskawicz, B. (1990). Characterization of IS476 and its role in bacterial spot disease of tomato and pepper. *J Bacteriol* **172**, 143-148.

Kentner, D. & Sourjik, V. (2006). Spatial organization of the bacterial chemotaxis system. *Curr Opin Microbiol* **9**, 619-624.

Kline, K.A., Fälker, S., Dahlberg, S., Normark, S. & Henriques-Normark, B. (2009). Bacterial adhesins in host-microbe interactions. *Cell Host Microbe* **5**, 580-592.

Koch, K. & Ensikat, H-J. (2008). The hydrophobic coatings of plant surfaces: epicuticular wax crystals and their morphologies, crystallinity and molecular self-assembly. *Micron* **39**, 759-772.

- Koebnik, R. (2005).** TonB-dependent trans-envelope signalling: the exception or the rule? *TRENDS in Microbiology* **13**, 343-347.
- Koebnik, R. (2007).** Comparative genomics of *Xanthomonas*: the puzzle of life style. *XIII International Congress on Molecular Plant-Microbe Interactions*. July 21-27, 2007 Sorrento Italy.
- Koebnik, R., Krüger, A., Thieme, F., Urban, A. & Bonas, U. (2006).** Specific binding of the *Xanthomonas campestris* pv. *vesicatoria* AraC-type transcriptional activator HrpX to plant-inducible promoter boxes. *J Bacteriol* **188**, 7652-7660.
- Koshland, D.E. (1980).** Bacterial chemotaxis in relation to neurobiology. *Annu Rev Neurosci* **3**, 43-75.
- Krantz, S.L., Benoit, T.G. & Beasley C.W. (1999).** Phytopathogenic bacteria associated with tardigrada. *Zool Anz* **238**, 259-260.
- Kube, M., Migdoll, A.M., Müller, I., Kuhl, H., Beck, A., Reinhardt, R. & Geider, K. (2008).** The genome of *Erwinia tasmaniensis* strain Et1/99, a non-pathogenic bacterium in the genus *Erwinia*. *Environ Microbiol* **10**, 2211-2222.
- Kubo, I., Muroi, H. & Kubo, A. (1995).** Structural functions of antimicrobial long-chain alcohols and phenols. *Bioorg Med Chem* **3**, 873-880.
- Kuhner, M.K. (2008).** Coalescent genealogy samplers: windows into population history. *Trends Ecol Evol* **24**, 86-93.
- Lathe, W.C., Snel, B. & Bork, P. (2000).** Gene context conservation of a higher order than operons. *Trends Biochem Sci* **25**, 474-479.
- Lavín, J.L., Kiil, K., Resano, O., Ussery, D.W. & Oguiza, J.A. (2007).** Comparative genomic analysis of two-component regulatory proteins in *Pseudomonas syringae*. *BMC Genomics* **8**, 397.
- Lawrence, J.C & Ochman, H. (1998).** Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci U S A* **95**, 9413-9417.
- Lee, B.M., Park, Y.J., Park, D.D., Kang, H.W., Kim, J.G., Song, E.S., Park, I.C., Yoon, U.H., Hahn, J.H., Koo, B.S., Lee, G.B., Kim, H., Park, H.S., Yoon, K.O., Kim, J.H., Jung, J.H., Koh, N.H., Seo, J.S. & Seung, J.G. (2005).** The genome sequence of *Xanthomonas oryzae* pathovar *oryzae* KACC10331, the bacterial blight pathogen of rice. *Nucleic Acids Res* **33**, 577-586.
- Leveau, J.H.J. & Lindow, S.E. (2001).** Appetite of an epiphyte: quantitative monitoring of bacterial sugar consumption in the phyllosphere. *Proc Natl Acad Sci U S A* **98**, 3446-3453.

- Lim, B.L. (2010).** TonB-Dependent Receptors in Nitrogen-Fixing Nodulating Bacteria. *Microbes Environ* **25**, 67-74.
- Lindemann, J. & Upper, C. (1985).** Aerial Dispersal of Epiphytic Bacteria over Bean Plants. *Appl Environ Microbiol* **50**, 1229-1232.
- Lindow, S.E. (1996).** Role of immigration and other processes in determining epiphytic bacterial populations. In *Aerial plant surface microbiology*, C.E. Morris, P.C. Nicot and C. Nguyen-The, eds (New York, Plenum Press), 155-168.
- Lindow, S.E. (2002).** Differential survival of solitary and aggregated cells of *Pseudomonas syringae* on leaves. *Phytopathology* **92**, S97.
- Lindow, S.E. & Brandl, M.T. (2003).** Microbiology of the phyllosphere. *Appl Environ Microbiol* **69**, 1875-1883.
- Linke, D., Riess, T., Autenrieth, I.B., Lupas, A. & Kempf, V.A. (2006).** Trimeric autotransporter adhesins: variable structure, common function. *Trends Microbiol.* **14**, 264-270.
- Ma, W., Dong, F., Stavrinides, J. & Guttman, D.S. (2006).** Type III effector diversification via both pathoadaptation and horizontal transfer in response to a coevolutionary arms race. *PLoS Genet* **2**, e209.
- Madan Babu, M. (2003).** Did the loss of sigma factors initiate pseudogene accumulation in *M. leprae*? *Trends Microbiol* **11**, 59-61.
- Maddock, J.R. & Shapiro, L. (1993).** Polar location of the chemoreceptor complex in the *Escherichia coli* cell. *Science* **259**, 1717-1723.
- Mahillon, J. & Chandler, M. (1998).** Insertion sequences. *Microbiol Mol Biol Rev* **62**, 725-774.
- Mahuku, G.S., Jara, C., Henriquez, M.A., Castellanos, G. & Cuasquer, J. (2006).** Genotypic characterization of the common bean bacterial blight pathogens *Xanthomonas axonopodis* pv. *phaseoli* and *Xanthomonas axonopodis* pv. *phaseoli* var. *fuscans* by rep-PCR and PCR-RFLP of the ribosomal genes. *J Phytopathol* **154**, 35-44.
- Maiden, M.C., Bygraves, J.A, Feil, E. Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M. & Spratt B.G. (1998).** Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**, 3140-3145.
- Mkandawire, A.B., Mabagala, R.B., Guzman, P., Gepts, P. & Gilbertson R.L. (2004).** Genetic diversity and pathogenic variation of common blight bacteria (*Xanthomonas campestris* pv. *phaseoli* and *X. campestris* pv. *phaseoli* var. *fuscans*) suggests pathogen coevolution with the common bean. *Phytopathology* **94**, 593-603.

- Manson, M.D., Armitage, J.P., Hoch, J.A. & MacNab, R.M. (1998).** Bacterial locomotion and signal transduction. *J Bacteriol* **180**, 1009-1022.
- Marenda, M., Brito, B., Callard, D., Genin, S., Barberis, P., Boucher, C. & Arlat, M. (1998).** PrhA controls a novel regulatory pathway required for the specific induction of *Ralstonia solanacearum* *hrp* genes in the presence of plant cells. *Mol Microbiol* **27**, 437-453.
- Marguerettaz, M., Pieretti, I., Gayral, P., Puig, J., Brin, C., Cociancich, S., Poussier, S., Rott, P. & Royer, M. (2010).** Genomic and evolutionary features of the SPI-1 type III secretion system that is present in *Xanthomonas albilineans* but not essential for xylem colonization and symptom development of sugarcane leaf scald. *Mol Plant Microbe Interact*, in press.
- Martin, D. & Rybicki, E. (2000).** RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**, 562-563.
- Mason, T.L. & Wasserman, B.P. (1987).** Inactivation of red beet betaglucan synthase by native and oxidized phenolic compounds. *Phytochemistry* **26**, 2197-2202.
- Maude, R.B. (1996).** Seedborne diseases and their control. *Principles & Practice*. CAB International, Oxon, UK.
- Maurelli, A.T. (2007).** Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens. *FEMS Microbiol Lett* **267**, 1-8.
- Mayr, E. (1942).** Systematics and origin of species from the viewpoint of a zoologist. New York: Columbia University Press.
- Mazar, J. & Cotter, P.A. (2007).** New insight into the molecular mechanisms of two-partner secretion. *Trends Microbiol.* **15**, 508-515.
- McDonald, J.H. & Kreitman, M. (1991).** Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* **351**, 652-654.
- Melotto, M., Underwood, W., Koczan, J., Nomura, K. & He, S.Y. (2006).** Plant stomata function in innate immunity against bacterial invasion. *Cell* **126**, 969-980.
- Meng, Y., Li, Y.R., Galvani, C.D., Hao, G., Turner, J.N., Burr, T.J. & Hoch, H.C. (2005).** Upstream migration of *Xylella fastidiosa* via pilus-driven twitching motility. *J Bacteriol* **187**, 5560-5567.
- Mercier, J. & Lindow, S.E. (2000).** Role of leaf surface sugars in colonization of plants by bacterial epiphytes. *Appl Environ Microbiol* **66**, 369-374.
- Meyer, D.F. & Bogdanove, A.J. (2009).** Genomics-driven advances. In Plant pathogenic bacteria: genomics and molecular biology, Jackson, ed (Norfolk, United Kingdom, Caister Academic Press), 147-161.

Mignard, S. (2008). Phylogénie et identification des mycobactéries. Lyon, Université Claude Bernard-Lyon 1, pp. 113.

Monier, J.M. & Lindow, S.E. (2003). Differential survival of solitary and aggregated bacterial cells promotes aggregate formation on leaf surfaces. *Proc Natl Acad Sci U S A* **100**, 15977-15982.

Monier, J.M. & Lindow, S.E. (2005). Aggregates of resident bacteria facilitate survival of immigrant bacteria on leaf surfaces. *Microb Ecol* **49**, 343-352.

Montesinos, E., Bonaterra, A., Badosa, E., Francés, J., Alemany, J., Llorente, I. & Moragrega, C. (2002). Plant-microbe interactions and the new biotechnological methods of plant disease control. *Int Microbiol* **5**, 169-175.

Moreira, L.M., De Souza, R.F., Almeida Jr, N.F., Setubal, J.C., Oliveira, J.C., Furlan, L.R., Ferro, J.A. & da Silva, A.C.R. (2004). Comparative genomics analyses of citrus-associated bacteria. *Annu Rev Phytopathol* **42**, 163-184.

Moreira, L.M., Almeida, N.F., Potnis, N., Digiampietri, L.A., Adi, S.S., Bortolossi, J.C., da Silva, A.C., da Silva, A.M., de Moraes, F.E., de Oliveira, J.C., de Souza, R.F., Facincani, A.P., Ferraz, A.L., Ferro, M.I., Furlan, L.R., Gimenez, D.F., Jones, J.B., Kitajima, E.W., Laia, M.L., Leite, R.P., Nishiyama, M.Y., Neto, J.R., Nociti, L.A., Norman, D.J., Ostroski, E.H., Pereira, H.A., Staskawicz, B.J., Tezza, R.I., Ferro, J.A., Vinatzer, B.A. & Setubal, J.C. (2010). Novel insights into the genomic basis of citrus canker based on the genome sequences of two strains of *Xanthomonas fuscans* subsp. *aurantifolii*. *BMC Genomics* **11**, 238.

Morris, C.E., Monier, J-M. & Jacques, M-A. (1997). Methods for observing microbial biofilms directly on leaf surfaces and recovering them for isolation of culturable microorganisms. *Appl Environ Microbiol* **63**, 1570-1576.

Morris, C.E., Monier, J-M. & Jacques, M-A. (1998). A technique to quantify the population size and composition of the biofilm component in communities of bacteria in the phyllosphere. *Appl Environ Microbiol* **64**, 4789-4795.

Morris, C.E. & Kinkel, L.L. (2002). Fifty years of phyllosphere microbiology: significant contributions to research in related fields. In *Phyllosphere microbiology*, S.E. Lindow, E.I. Hecht-Poinar and V. Elliott, eds, (St. Paul, Minnesota), APS Press, 365-375.

Morris, C.E., Kinkel, L.L., Xiao, K., Prior, P. & Sands, D.C. (2007). Surprising niche for the plant pathogen *Pseudomonas syringae*. *Infection, Genetics and Evolution* **7**, 84-92.

Morris, C.E., Sands, D.C., Vinatzer, B.A., Glaux, C., Guilbaud, C., Buffière, A., Yan, S., Dominguez, H. & Thompson, B.M. (2008). The life history of the plant pathogen *Pseudomonas syringae* is linked to the water cycle. *ISME J* **2**, 321-334.

- Morrison, D.A. (2005).** Networks in phylogenetic analysis: new tools for population biology. *Int J Parasitol* **35**, 567-582.
- Moxon, E.R., Rainey, P.B., Nowak, M.A. & Lenski, R.E. (1994).** Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Biol* **4**, 24-33.
- Moxon, R., Bayliss, C. & Hood, D. (2006).** Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu Rev Genet* **40**, 307-333.
- Müller, C. & Riederer, M. (2005).** Plant surface properties in chemical ecology. *J Chem Ecol* **31**, 2621-2651.
- Niemann, H.H., Schubert, W.D. & Heinz, D.W. (2003).** Adhesins and invasins of pathogenic bacteria: a structural view. *Microbes Infect* **6**, 101-112.
- Niinemets, U., Loreto, F. & Reichstein, M. (2004).** Physiological and physicochemical controls on foliar volatile organic compound emissions. *Trends Plant Sci* **9**, 180-186.
- Noël, L., Thieme, F., Nennstiel, D. & Bonas, U. (2001).** cDNA-AFLP analysis unravels a genome-wide *hrpG*-regulon in the plant pathogen *Xanthomonas campestris* pv. *vesicatoria*. *Mol Microbiol* **41**, 1271-1281.
- Nuryastuti, T., van der Mei, H., Busscher, H.J., Iravati, S., Aman, A.T. & Krom, B.P. (2009).** Effect of cinnamon oil on *icaA* expression and biofilm formation by *Staphylococcus epidermidis*. *Appl Environ Microbiol* **75**, 6850-6855.
- Ochiai, H., Inoue, Y., Takeya, M., Sasaki, A. & Kaku, H. (2005).** Genome sequence of *Xanthomonas oryzae* pv. *oryzae* suggests contribution of large numbers of effector genes and insertion sequences to its race diversity. *JARQ* **39**, 275-288.
- Ochman, H., Lawrence, J.C. & Groisman, E.A. (2000).** Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299-304.
- Ochman, H. & Moran, N.A. (2001).** Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* **292**, 1096-1099.
- Oh, I., Park, A., Bae, M., Kwon, S.J., Kim, Y.S., Lee, J.E., Kang, N.Y., Lee, S., Cheong, H. & Park, O.K. (2005).** Secretome analysis reveals an *Arabidopsis* lipase involved in defense against *Alternaria brassicicola*. *Plant Cell* **17**, 2832-2847.
- Ojanen-Reuhs, T., Kalkkinen, N., Westerlund-Wikstrom, B., van Doorn, J., Haahtela, K., Nurmiaho-Lassila, E.L., Wengelnik, K., Bonas, U. & Korhonen, T.K. (1997).** Characterization of the *fimA* gene encoding bundle-forming fimbriae of the plant pathogen *Xanthomonas campestris* pv. *vesicatoria*. *J Bacteriol* **179**, 1280-1290.

- Opio, A.F., Allen, D.J. & Teri, J.M. (1996).** Pathogen variation in *Xanthomonas campestris* pv. *phaseoli*, the causal agent of common bacterial blight in *Phaseolus beans*. *Plant Pathol.* **45**, 1126-1133.
- Orr, M.R. & Smith, T.B. (1998).** Ecology and Speciation. *Trends Ecol Evol* **13**, 502-506.
- Pallen, M.J & Wren, B.W. (2007).** Bacterial pathogenomics. *Nature* **449**, 835-842.
- Palleroni, N.J. & Bradbury, J.F. (1993).** *Stenotrophomonas*, a new bacterial genus for *Xanthomonas maltophilia* (Hugh 1980) Swings *et al.*, 1983. *Int J Syst Bacteriol* **43**, 606-609.
- Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T.G., Prentice, M.B., Sebahia, M., James, K.D., Churcher, C., Mungall, K.L., Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdeño-Tárraga, A.M., Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Karlyshev, A.V., Leather, S., Moule, S., Oyston, P.C.F., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. & Barrell, B.G. (2001).** Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**, 523-527.
- Parkinson, J.S. (1976).** *cheA*, *cheB*, and *cheC* genes of *Escherichia coli* and their role in chemotaxis. *J Bacteriol* **126**, 758-770.
- Parkinson, J.S., Ames, P. & Studdert, C.A. (2005).** Collaborative signaling by bacterial chemoreceptors. *Curr Opin Microbiol* **8**, 116-121.
- Parkinson, N., Aritua, V., Heeney, J., Cowie, C., Bew, J. & Stead, D. (2007).** Phylogenetic analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. *Int J Syst Evol Microbiol* **57**, 2881-2887.
- Parkinson, N., Cowie, C., Heeney, J. & Stead, D. (2009).** Phylogenetic structure of *Xanthomonas* determined by comparison of *gyrB* sequences. *Int J Syst Evol Microbiol* **59**, 264-274.
- Peñuelas, J. & Llusà, J. (2004).** Plant VOC emissions: making use of the unavoidable. *Trends Ecol Evol* **19**, 402-404.
- Pieretti I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Cociancich, S., Couloux, A., Darrasse, A., Gouzy, C., Jacques, M-A., Lauber, E., Manceau, C., Mangenot, S., Poussier, S., Segurens, B., Szurek, B., Verdier, V., Arlat, M. & Rott. P. (2009).** The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. *BMC Genomics* **10**, 616.

- Pietrarelli, L., Balestra, G.M. & Varvaro, L. (2006).** Effects of simulated rain on *Pseudomonas syringae* pv. *tomato* populations on tomato plants. *J Plant Pathol* **88**, 245-251.
- Pollard, M., Beisson, F., Li, Y. & Ohlrogge, J.B. (2008).** Building lipid barriers: biosynthesis of cutin and suberin. *Trends Plant Sci* **13**, 236-246.
- Posada, D., Crandall, K.A. & Holmes E.C. (2002).** Recombination in evolutionary genomics. *Annu Rev Genet* **36**, 75–97.
- Poulin, R. (2007).** Host specificity. In *Evolutionary ecology of parasites*, R. Poulin, ed (New Jersey, Princeton University Press), 41-69.
- Poussier, S., Thoquet, P., Trigalet-Demery, D., Barthet, S., Meyer, D., Arlat, M. & Trigalet, A. (2003).** Host plant-dependent phenotypic reversion of *Ralstonia solanacearum* from non-pathogenic to pathogenic forms via alterations in the *phcA* gene. *Mol Microbiol* **49**, 991-1003.
- Pruvost, O., Savelon, C., Boyer, C., Chiroleu, F., Gagnevin, L. & Jacques, M-A. (2009).** Populations of *Xanthomonas citri* pv. *mangiferaeindicae* from asymptomatic mango leaves are primarily endophytic. *Microb Ecol* **58**, 170-178.
- Qian, W., Jia, Y., Ren, S-X., He, Y-Q., Feng, J-X., Lu, L-F., Sun, Q., Ying, G., Tang, DJ., Tang., H., Wu, W., Hao, P., Wang, L., Jiang, B-L., Zeng, S., Gu, W-Y., Lu, G., Rong, L., Tian, Y., Yao, Z., Fu, G., Chen, B., Fang, R., Qiang, B., Chen, Z., Zhao, G-P., Tang, J-L. & He, C. (2005).** Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome Res* **15**, 757-767.
- Qian, W., Han, Z.J. & He, C. (2008).** Two-component signal transduction systems of *Xanthomonas* spp.: a lesson from genomics. *Mol Plant Microbe Interact* **21**, 151-161.
- Rademaker, J.L.W., Hoste, B., Louws, F.J., Kersters, K., Swings, J., Vauterin, L., Vauterin, P. & de Bruijn, F.J. (2000).** Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: *Xanthomonas* as a model system. *Int J Syst Evol Microbiol* **50**, 665-677.
- Rademaker, J.L.W., Louws, F.J., Schultz, M.H., Rossbach, U., Vauterin, L., Swings, J. & de Bruijn, F.J. (2005).** A comprehensive species to strain taxonomic framework for *Xanthomonas*. *Phytopathology* **95**, 1098-1111.
- Radman, M., Taddei, F. & Matic, I. (2000).** Evolution-driving genes. *Res Microbiol* **151**, 91-95.
- Rajagopal, L., Sundari, C.S., Balasubramanian, D. & Sonti, R. (1997).** The bacterial pigment xanthomonadin offers protection against photodamage. *FEBS Lett* **415**, 125-128.

Rao, C.V. & Ordal, G.W. (2009). The molecular basis of excitation and adaptation during chemotactic sensory transduction in bacteria. *Contrib Microbiol* **16**, 33-64.

Rauha, J-P., Remes, S., Heinonen, M., Hopia, A., Kähkönen, M., Kujala, T., Pihlaja, K., Vuorela, H. & Vuorela, P. (2000). Antimicrobial effects of Finnish plant extracts containing flavonoids and other phenolic compounds. *Int J Food Microbiol* **56**, 3-12.

Ridé, M., & Ridé S. (1978). *Xanthomonas populi* (RidC) comb. nov. (syn. *Aplanobacter populi* Ride) spécifité, variabilité et absence de relation avec *Erwinia cancerogena*. *Eur J For Pathol* **8**, 310-333.

Rigano, L.A., Siciliano, F., Enrique, R., Sendín, L., Filippone, P., Torres, P.S., Qüesta, J., Dow, J.M., Castagnaro, A.P., Vojnov, A.A. & Marano, M.R. (2007). Biofilm formation, epiphytic fitness, and canker development in *Xanthomonas axonopodis* pv. *citri*. *Mol Plant Microbe Interact* **20**,1222-1230.

Robène-Soustrade, I., Laurent, P., Gagnevin, L., Jouen, E. & Pruvost, O. (2006). Specific detection of *Xanthomonas axonopodis* pv. *dieffenbachiae* in anthurium (*Anthurium andreanum*) tissues by nested PCR. *Appl Environ Microbiol* **72**, 1072-1078.

Rosenberg, N.A. & Nordborg, M. (2002). Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat Rev Genet* **3**, 380-390.

Rosenstiel, T.N., Ebbets, A.L., Khatri, W.C., Fall, R. & Monson, R.K. (2004). Induction of poplar leaf nitrate reductase: a test of extrachloroplastic control of isoprene emission rate. *Plant Biol* **6**, 12-21.

Roumagnac, P., Pruvost, O., Chiroleu, F. & Hughes, G. (2004a). Spatial and temporal analysis of bacterial blight of onion caused by *Xanthomonas axonopodis* pv. *allii*. *Phytopathology* **94**,138-146.

Roumagnac, P., Gagnevin, L., Gardan, L., Sutra, L., Manceau, C., Dickstein, E.R., Jones, J.B., Rott, P. & Pruvost, O. (2004b). Polyphasic characterization of xanthomonads isolated from onion, garlic and Welsh onion (*Allium* spp.) and their relatedness to different *Xanthomonas* species. *Int J Syst Evol Microbiol* **54**, 15-24.

Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X. & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496-2497.

Rudolph, K. (1993). Infection of the plant by *Xanthomonas*. In *Xanthomonas*, J.G. Swings, and E.L.Civerolo, eds (London, United Kingdom, Chapman and Hall), 193-264.

Ruppel, S., Krumbein, A. & Schreiner, M. (2008). Composition of the phyllospheric microbial populations on vegetable plants with different glucosinolate and carotenoid compositions. *Microb Ecol* **56**, 364-372.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A. & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945.

Ryan, R.P., Fouhy, Y., Lucey, J.F., Jiang, B-L., He, Y-Q., Feng, J-X., Tang, J-L. & Dow, J.M. (2007). Cyclic di-GMP signalling in the virulence and environmental adaptation of *Xanthomonas campestris*. *Mol Microbiol* **63**, 429-442.

Saddler, G.S. & Bradbury, J.F. (2005). Family I. *Xanthomonadaceae* fam. nov. In *Bergey's manual of systematic bacteriology* second édition, vol. 2 (The Proteobacteria), part B (The Gammaproteobacteria), G.M Garrity, D.J Brenner, N.R Krieg, J.T Staley eds (Springer, New York), 63-90.

Sarkar S.F., Gordon J.S., Martin G.B. & Guttman D.S. (2006). Comparative genomics of host-specific virulence in *Pseudomonas syringae*. *Genetics* **174**, 1041-1056.

Salzberg, S.L., Sommer, D.D., Schatz, M.C., Phillippy, A.M., Rabinowicz, P.D., Tsuge, S., Furutani, A., Ochiai, H., Delcher, A.L., Kelley, D., Madupu, R., Puiu, D., Radune, D., Shumway, M., Trapnell, C., Aparna, G., Jha, G., Pandey, A., Patil, P.B., Ishihara, H., Meyer, D.F., Szurek, B., Verdier, V., Koebnik, R., Dow, J.M., Ryan, R.P., Hirata, H., Tsuyumu, S., Won Lee, S., Seo, Y.S., Sriariyanum, M., Ronald, P.C., Sonti, R.V., Van Sluys, M.A., Leach, J.E., White, F.F. & Bogdanove, A.J. (2008). Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99^A. *BMC Genomics* **9**, 204.

Schaad, N.W., Postnikova, E., Lacy, G.H., Sechler, A., Agarkova, I., Stromberg, P., Stromberg, V.K. & Vidaver, A.K. (2005). Reclassification of *Xanthomonas campestris* pv. *citri* (ex Hasse 1915) Dye 1978 forms A, B/C/D, and E as *X. smithii* subsp. *citri* (ex Hasse) sp. nov. nom. rev. comb. nov., *X. fuscans* subsp. *aurantifolii* (ex Gabriel 1989) sp. nov. nom. rev. comb. nov., and *X. alfalfae* subsp. *citrumelo* (ex Riker and Jones) Gabriel et al., 1989 sp. nov. nom. rev. comb. nov.; *X. campestris* pv. *malvacearum* (ex smith 1901) Dye 1978 as *X. smithii* subsp. *smithii* nov. comb. nov. nom. nov.; *X. campestris* pv. *alfalfae* (ex Riker and Jones, 1935) dye 1978 as *X. alfalfae* subsp. *alfalfae* (ex Riker et al., 1935) sp. nov. nom. rev.; and "var. *fuscans*" of *X. campestris* pv. *phaseoli* (ex Smith, 1987) Dye 1978 as *X. fuscans* subsp. *fuscans* sp. nov. *Syst Appl Microbiol* **28**, 494-518.

Schaad, N.W., Postnikova, E., Lacy, G.H., Sechler, A., Agarkova, I., Stromberg, P., Stromberg, V.K. & Vidaver, A.K. (2006). Emended classification of xanthomonads pathogens on citrus. *Syst Appl Microbiol* **29**, 690-695.

Schaad, N.W., Postnikova, E., Lacy, G., Sechler, A., Agarkova, I., Stromberg, P.E., Stromberg, V.K. & Vidaver, A.K. (2007). *Xanthomonas alfalfae* sp. nov., *Xanthomonas citri* sp. nov. and *Xanthomonas fuscans* sp. nov. In list of new names and new combinations previously effectively, but not validly, published, validation list no. 115. *Int J Syst Evol Microbiol* **57**, 893-897.

- Schauer, K., Rodionov, D.A. & de Reuse, H. (2008).** New substrates for TonB-dependent transport: do we only see the 'tip of the iceberg'? *Trends Biochem Sci* **33**, 330-338.
- Schechter, L.M., Roberts, K.A., Jamir, Y., Alfano, J.R. & Collmer, A. (2004).** *Pseudomonas syringae* type III secretion system targeting signals and novel effectors studied with a Cya translocation reporter. *J Bacteriol* **186**, 543-555.
- Schönherr, J. & Baur, P. (1996).** Cuticle permeability studies: a model for estimating leaching of plant metabolites to leaf surfaces. In *Aerial plant surface microbiology*, C.E. Morris, P.C. Nicot and C. Nguyen-The, eds (New York, Plenum Press), 1-24.
- Shepherd, R.W. & Wagner, G.J. (2007).** Phylloplane proteins: emerging defenses at the aerial frontline? *Trends Plant Sci* **12**, 51-56.
- Shimodaira, H. & Hasegawa, M. (1999).** Multiple comparisons of loglikelihoods with applications to phylogenetic inference. *Mol Biol Evol* **16**, 1114-1116.
- Sikirou, R. & Wydra, K. (2004).** Persistence of *Xanthomonas axonopodis* pv. *vignicola* in weeds and crop debris and identification of *Sphenostylis stenocarpa* as a potential new host. *Eur J Plant Pathol* **110**, 939-947.
- Silverstein, K.A.T., Graham, M.A., Paape, T.D. & VandenBosch, K.A. (2005).** Genome organization of more than 300 defensin-like genes in *Arabidopsis*. *Plant Physiol* **138**, 600-610.
- Slocum, M.K. & Parkinson, J.S. (1983).** Genetics of methyl-accepting chemotaxis proteins in *Escherichia coli*: organization of the tar region. *J Bacteriol* **155**, 586-594.
- Smith, J.J., Jones, D.R., Karamura, E., Blomme, G. & Turyagyenda, F.L. (2008).** An analysis of the risk from *Xanthomonas campestris* pv. *musacearum* to banana cultivation in Eastern, Central and Southern Africa. [InfoMusa@\[http://www.promusa.org/index.php?option=com_content&task=view&id=66\]](http://www.promusa.org/index.php?option=com_content&task=view&id=66).
- Sokurenko, E.V., Chesnokova, V., Dykhuizen, D.E., Ofek, I., Wu, X-R., Krogfelt, K.A., Struve, C., Schembri, M.A. & Hasty, D.L. (1998).** Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin. *Proc Natl Acad Sci U S A* **95**, 8922-8926.
- Sokurenko, E.V., Hasty, D.L. & Dykhuizen, D.E. (1999).** Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends Microbiol* **7**, 191-195.
- Sourjik, V. (2004).** Receptor clustering and signal processing in *E. coli* chemotaxis. *Trends Microbiol.* **12**, 569-576.

- Sourjik, V. & Berg, H.C. (2000).** Localization of components of the chemotaxis machinery of *Escherichia coli* using fluorescent protein fusions. *Mol Microbiol* **37**, 740-751.
- Sourjik, V. & Berg, H.C. (2002).** Receptor sensitivity in bacterial chemotaxis. *Proc Natl Acad Sci U S A* **99**, 123-127.
- Stadler, B. & Müller, T. (2000).** Effects of aphids and moth caterpillars on epiphytic microorganisms in canopies of forest trees. *Can J For Ecol* **30**, 631-638.
- Stall, R.E., Gottwald, T.R., Koizumi, M. & Schaad, N.C. (1993).** Ecology of plant pathogenic xanthomonads. In *Xanthomonas*, J.G. Swings, and E.L. Civerolo, eds (London, United Kingdom, Chapman and Hall), 265-299.
- Starr, M.P. (1981).** The genus *Xanthomonas*. In: The prokaryotes. Starr MP, Stolp H, Truper HG, Balows A, Schlegel HG, eds. Springer-Verlag, Berlin, Germany, 742-763.
- Stasiuk, M. & Kozubek, A. (2010).** Biological activity of phenolic lipids. *Cell Mol Life Sci* **67**, 841-860.
- Steinmuller, D. & Tevini, M. (1985).** Action of ultraviolet radiation (UV-B) upon cuticular waxes of some crop plants. *Planta* **164**, 557-564.
- Stuckenbrock, E.H. & McDonald, B.A. (2008).** The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology* **46**, 75-100.
- Studholme, D.J., Kemen, E., MacLean, D., Schornack, S., Aritua, V., Thwaites, R., Grant, M., Smith, J. & Jones, J.D.G. (2010).** Genome-wide sequencing data reveals virulence factors implicated in banana *Xanthomonas* wilt. *FEMS Microbiol Lett* **310**, 182-192.
- Sutherland, J.W. (1993).** Xanthan. In *Xanthomonas*, J.G. Swings, and E.L. Civerolo, eds (London, United Kingdom, Chapman and Hall), 363-388.
- Swings, J., De Vos, P., Van den Mooter, M., & De Ley, J. (1983).** Transfer of *Pseudomonas maltophilia* Hugh 1981 to the genus *Xanthomonas* as *Xanthomonas maltophilia* (Hugh 981) comb nov. *Int J Syst Bacteriol* **33**, 409-413.
- Szurmant, H. & Ordal, G.W. (2004).** Diversity in chemotaxis mechanisms among the bacteria and archaea. *Microbiol Mol Biol Rev* **68**, 301-319.
- Takéo, T. (1992).** Green and semi-fermented teas. In *Tea: Cultivation to Consumption*, K.C. Wilson and M. Clifford, eds (London, Chapman & Hall), 413-457.
- Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007).** MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596-1599.

- Tang, X., Xiao, Y. & Zhou, J.M. (2006).** Regulation of the type III secretion system in phytopathogenic bacteria. *Mol Plant Microbe Interact.* **19**, 1159-1166.
- Teichmann, S.A. & Madan Babu, M. (2004).** Gene regulatory network growth by duplication. *Nat Genet* **36**, 492-496.
- Tenaillon, O., Taddei, F., Radmian, M. & Matic, I. (2001).** Second-order selection in bacterial evolution: selection acting on mutation and recombination rates in the course of adaptation. *Res Microbiol* **152**, 11-16.
- Tenaillon, O., Skurnik, D., Picard, B. & Denamur, E. (2010).** The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol* **8**, 207-217.
- Thébaud, P., Servant, F., Schiex, T., Kahn, D. & Gouzy J. (2000).** *JOBIM Conf Proc* ENSA, LIRM. Montpellier, 361-365.
- Thieme, F., Koebnik, R., Bekel, T., Berger, C., Boch, J., Büttner, D., Caldana, C., Gaigalat, L., Goesmann, A., Kay, S., Kirchner, O., Lanz, C., Linke, B., McHardy, A.C., Meyer, F., Mittenhuber, G., Nies, D.H., Niesbach-Klosgen, U., Patschkowski, T., Ruckert, C., Rupp, O., Schneiker, S., Schuster, S.C., Vorhölter, F.J., Weber, E., Puhler, A., Bonas, U., Bartels, D. & Kaiser, O. (2005).** Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* revealed by the complete genome sequence. *J Bacteriol* **187**, 7254-7266.
- Timmer, L.W. (1989).** Inoculum production from Asiatic citrus canker lesions and epiphytic survival of *X. campestris* pv. *citri* in Argentina. *Phytopathology* **79**, 1182.
- Timmer, L.W., Marois, J.J. & Achor, D. (1987).** Growth and survival of xanthomonads under conditions nonconductive to disease development. *Phytopathology* **77**, 1431-1445.
- Timmer, L.W., Gottwald, T.R. & Zitko, S.E. (1991).** Bacterial exudation from lesions of Asiatic citrus canker and citrus bacterial spot. *Plant Disease* **75**, 192-195.
- Triplett, L.R., Zhao, Y. & Sundin, G.W. (2006).** Genetic differences between blight-causing *Erwinia* species with differing host specificities, identified by suppression subtractive hybridization. *Appl Environ Microbiol* **72**, 7359-7364.
- Tyagi, J.S & Saini, D.K. (2004).** Did the loss of two-component systems initiate pseudogene accumulation in *Mycobacterium leprae*? *Microbiology* **150**, 4-7.
- Ulrich, L.E. & Zhulin, I.B. (2010).** The MiST2 database: a comprehensive genomics resource on microbial signal transduction. *Nucleic Acids Res* **38**, D401-407.
- van den Mooter, M. & Swings, J. (1990).** Numerical analysis of 295 phenotypic features of 266 *Xanthomonas* strains and related strains and an improved taxonomy of the genus. *Int J Syst Bacteriol* **40**, 348-369.

- van der Wolf, J.M., van der Zouwen, P.S. & Garbeva, P. (2006).** Transmission of *Xanthomonas* pv. *campestris* by insects. In Proceedings of the 11th International Conference on Pathogenic Bacteria, 99-100.
- van Elsas, J.D., Turner, S., & Bailey, M.J., (2003).** Horizontal gene transfer in the phytosphere. *New Phytol* **157**, 525-537.
- Vauterin, L., Hoste, B., Kersters, K. & Swings, J. (1995).** Reclassification of *Xanthomonas*. *Int J Syst Bacteriol* **45**, 472-489.
- Vauterin, L. & Swings, J. (1997).** Are classification and phytopathological diversity compatible in *Xanthomonas*? *Journal of Industrial Microbiology and Biotechnology* **19**, 77-82.
- Vicente, J.G., Conway, J., Roberts, S.T. & Taylor, J.D. (2001).** Identification and origin of *Xanthomonas campestris* pv. *campestris* races and related pathovars. *Phytopathology* **91**, 492-499.
- Vicente, J.G., Everett, B. & Roberts, S.T. (2006).** Identification of isolates that cause a leafspot disease of brassicas as *Xanthomonas campestris* pv. *raphani* and pathogenic and genetic comparison with related pathovars. *Phytopathology* **96**, 735-745.
- Vladimirov, N. & Sourjik, V. (2009).** Chemotaxis: how bacteria use memory. *Biological Chemistry* **390**, 1097-1104.
- Vorhölter, F.J., Schneiker, S., Goesmann, A., Krause, L., Bekel, T., Kaiser, O., Linke, B., Patschkowski, T., Rückert, C., Schmid, J., Sidhu, V.K., Sieber, V., Tauch, A., Watt, S.A., Weisshaar, B., Becker, A., Niehaus, K. & Pühler, A. (2008).** The genome of *Xanthomonas campestris* pv. *campestris* B100 and its use for the reconstruction of metabolic pathways involved in xanthan biosynthesis. *J Biotechnol* **134**, 33-45.
- Vidaver, A.K. (1993).** *Xanthomonas campestris* pv. *phaseoli* cause of common bacterial blight of bean. In *Xanthomonas*, J.G. Swings, and E.L. Civerolo, eds (London, United Kingdom, Chapman and Hall), 40-51.
- Wadhams, G.H. & Armitage, J.P. (2004).** Making sense of it all: bacterial chemotaxis. *Nat Rev Mol Cell Biol* **5**, 1024-1037.
- Wakker, J.H. (1883).** Vorläufige Mittheilungen über Hyacinthenkrankheiten. *Botanisches Centralblatt* **14**, 315-317.
- Wang, E.A., Mowry, K.L., Clegg, D.O. & Koshland, D.E. (1982).** Tandem duplication and multiple functions of a receptor gene in bacterial chemotaxis. *J Biol Chem* **257**, 4673-4676.
- Weller, D.M., & Saettler, A.W. (1980).** Evaluation of seedborne *Xanthomonas*

phaseoli and *Xanthomonas phaseoli* var. *fuscans* as primary inocula in bean blights. *Phytopathology* **70**, 148-152.

Wengelnik, K., Rossier, O. & Bonas, U. (1999). Mutations in the regulatory gene *hrpG* of *Xanthomonas campestris* pv. *vesicatoria* result in constitutive expression of all *hrp* genes. *J Bacteriol* **181**, 6828-6831.

Wertz, J.E., Goldstone, C., Gordon, D.M. & Riley, M.A. (2003). A molecular phylogeny of enteric bacteria and implications for a bacterial species concept. *J Evol Biol* **16**, 1236-1248.

Willems, A., Gillis, M., Kersters, K., Van den Broecke, L., & De Ley, J. 1987. Transfer of *Xanthomonas ampelina* Panagopoulos 1969 to a new genus *Xylophilus* gen. nov., as *Xylophilus ampelinus* (Panagopoulos 1969) comb. nov. *Int. J. Syst. Bacteriol.* **37**, 422-430.

Whipps, J.M, Hand, P., Pink, D. & Bending, G.D. (2008). Phyllosphere microbiology with special reference to diversity and plant genotype. *J Appl Microbiol* **105**, 1744-1755.

White, F.F, Potnis, N., Jones, J.B. & Koebnik, R. (2009). The type III effectors of *Xanthomonas*. *Mol Plant Pathol* **10**, 749-766.

Whitworth, D.E. & Cock, P.J.A. (2009). Evolution of prokaryotic two-component systems: insights from comparative genomics. *Amino Acids* **37**, 459-466.

Wiener, M.C. (2005). TonB-dependent outer membrane transport: going for Baroque? *Curr Opin Struct Biol* **15**, 394-400.

Wojciechowski, M.F., Lavin, M. & Sanderson, M.J. (2004). A phylogeny of legumes (*Leguminosae*) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. *Am J Bot* **91**, 1846-1862.

Yadav, R.K.P., Karamanoli, K. & Vokou, D. (2005). Bacterial colonization of the phyllosphere of mediterranean perennial species as influenced by leaf structural and chemical features. *Microb Ecol* **50**, 185-196.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13**, 555-556.

Yang, Z., Wong, W.S.W. & Nielsen, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* **22**, 1107-1118.

Yang, P., Vauterin, L., Vancanneyt, M., Swings, J. & Kersters, K. (1993). Application of fatty acid methyl esters for the taxonomic analysis of the genus *Xanthomonas*. *Syst Appl Microbiol* **16**, 47-71.

- Yao, J. & Allen, C. (2006).** Chemotaxis is required for virulence and competitive fitness of the bacterial wilt pathogen *Ralstonia solanacearum*. *J Bacteriol* **188**, 3697-3708.
- Young, J.M., Dye, D.W., Bradbury, J.F., Panagopoulos, C.G. & Robbs, C.F. (1978).** A proposed nomenclature and classification for plant pathogenic bacteria. *New Zealand Journal of Agricultural Research* **21**, 153-177.
- Young, J.M., Park, D.C., Shearman, H.M. & Fargier, E. (2008).** A multilocus sequence analysis of the genus *Xanthomonas*. *Syst Appl Microbiol* **31**, 366-377.
- Young, S.A., Guo, A., Guikema, J.A., White, F.F. & Leach, J.E. (1995).** Rice cationic peroxidase accumulates in xylem vessels during incompatible interactions with *Xanthomonas oryzae* pv *oryzae*. *Plant Physiol* **107**, 1333-1341.
- Zhang, P., Khursigara, C.M., Hartnell, L.M. & Subramaniam, S. (2007).** Direct visualization of *Escherichia coli* chemotaxis receptor arrays using cryo-electron microscopy. *Proc Natl Acad Sci U S A* **104**, 3777-3781.
- Zhang, S.S., He, Y.Q., Xu, L.M., Chen, B-W., Jiang, B-L., Liao, J., Cao, J-R., Liu, D., Huang, Y-Q., Liang, X-X., Tang, D-J., Lu, G-T. & Tang, J-L. (2008).** A putative *colR*(XC1049)-*colS*(XC1050) two-component signal transduction system in *Xanthomonas campestris* positively regulates *hrpC* and *hrpE* operons and is involved in virulence, the hypersensitive response and tolerance to various stresses. *Res Microbiol* **159**, 569-578.
- Zhang, Y., Callaway, E.M., Jones, J.B. & Wilson, M. (2009).** Visualisation of *hrp* gene expression in *Xanthomonas euvesicatoria* in the tomato phyllosphere. *Eur J Plant Pathol* **124**, 379-390.
- Zhou, H., Morgan, R.L., Guttman, D.S. & Ma, W. (2009).** Allelic variants of the *Pseudomonas syringae* type III effector HopZ1 are differentially recognized by plant resistance systems. *Mol Plant Microbe Interact* **22**, 176-189.
- Zou, L.F., Wang, X.P., Xiang, Y., Zhang, B., Li, Y.R., Xiao, Y., Wang, J., Walmsley, A.R. & Chen, G. (2006).** Elucidation of the *hrp* clusters of *Xanthomonas oryzae* pv. *oryzicola* that control the hypersensitive response in nonhost tobacco and pathogenicity in susceptible host rice. *Appl Environ Microbiol* **72**, 6212-6224.

Approches cumulées de phylogénie et d'écologie pour déterminer les bases génétiques de la spécificité d'hôte des bactéries phytopathogènes, cas des *Xanthomonas* spp.

La spécificité d'hôte est le résultat de plusieurs phases d'interaction avec l'hôte : attraction, multiplication, colonisation et transmission. Nous avons émis l'hypothèse que la spécificité d'hôte se dessine dès les étapes précoces d'attraction et d'adhésion. Pour tester cette hypothèse, nous avons sélectionné les gènes codant des senseurs du chimiotactisme (MCPs) et des régulateurs à deux composants, des transporteurs TonB-dépendants et des adhésines, et avons déterminé leur distribution dans une collection de 173 souches appartenant à différents pathovars de *Xanthomonas* spp. ayant des gammes d'hôtes différentes. Les résultats obtenus montrent que la majorité des pathovars (28/34) présente des répertoires uniques et propres de ces gènes candidats, indiquant une corrélation entre le répertoire des gènes candidats et la gamme d'hôte. Six pathovars non individualisés partagent chacun leur répertoire avec seulement un autre pathovar. Ces similarités pourraient refléter des spécificités écologiques communes. Les analyses de séquences des gènes codant les senseurs du chimiotactisme et les adhésines ont révélé la présence de signaux de sélection positive dans la divergence adaptative entre les espèces de *Xanthomonas* spp. Au sein de l'espèce *X. axonopodis*, les gènes candidats sont soumis à une pression de sélection positive ou purificatrice. La pression de sélection positive s'exerce sur des sites appartenant à des domaines fonctionnels de ces protéines. Ces résultats suggèrent que la pression de sélection purificatrice agirait sur les gènes impliqués dans la reconnaissance de molécules végétales communes alors que la pression de sélection positive agirait sur les gènes impliqués dans la reconnaissance de molécules spécifiques d'une niche particulière. Ces résultats révèlent que la spécificité d'hôte se dessine dès les étapes précoces d'attraction et d'adhésion à la plante. Les études phylogénétiques et généalogiques menées sur les souches de *X. axonopodis* confirment l'existence de sous-groupes au sein de cette espèce et indiquent une divergence très ancienne pour certains d'entre eux. Malgré un long isolement de ces sous-groupes, de nombreux échanges de matériel génétique ont été identifiés et associés à des transferts de gènes codant des facteurs de virulence. Les résultats présentés montrent l'importance de la perception de l'environnement et de l'adhésion dans les interactions entre les bactéries phytopathogènes et les leurs plantes hôtes.

Mots-clés : adaptation à l'hôte, chimiotactisme, adhésion, *Xanthomonas*, écologie, phylogénie, généalogie

Phylogeny and ecology cumulated approaches to determine the genetic basis of host specificity in plant pathogenic bacteria, case of *Xanthomonas* spp.

Host specificity combines distinct successive phases of interaction between bacteria and host plant: attraction, ability to multiply on or inside the host and transmission to new hosts. We hypothesized that determinants responsible for bacterial host specificity are expressed starting from chemotactic attraction by host tissues and adhesion. We established the distribution of 70 genes involved in chemotactic attraction, chemical environment sensing and adhesion in a large collection of xanthomonad strains. These 173 strains belong to different pathovars of *Xanthomonas* spp and display different host ranges. Most pathovars (28/34) were characterized by unique repertoires of candidate genes highlighting a correspondence between pathovar clustering and repertoires of sensors and adhesins. In contrast, six pathovars may display the same repertoire. This may reflect common ecological behaviors. To further challenge our hypothesis, we tested for molecular signatures of selective pressures on genes encoding chemotactic sensors and adhesins of xanthomonad strains. We identified strong evidence of adaptive divergence acting on most candidate genes. Within *X. axonopodis*, candidate genes were identified to be under purifying selection or positive selection. Most of sites under positive selection were located in conserved domains within proteins. These finding suggest that purifying selection may act on genes involved in recognition of common structures of plant tissues and positive selection may act on genes coding for sensors and adhesins used for colonization of specific niches. These findings provide new insights in the evolutionary importance of chemotactic attraction and adhesion in the host specificity of plant pathogenic bacteria. Thus, events leading to host specificity may occur as early as chemotactic attraction by host and adhesion to tissues for plant pathogenic xanthomonads. Phylogenetical and genealogical analyses conducted on *X. axonopodis* strains strongly supported clustering of these strains into 6 subgroups corresponding to known subgroups within this species and indicated that the divergence between some subgroups is very ancient. Furthermore, recombination events due to genetic exchanges have been found to occur recently in *X. axonopodis* strains and some were associated to transfers of virulence-associated genes. All these data support the importance of perception of the environment and adhesion in the plant-microbe interactions.

Keywords: host adaptation, chemotaxis, adhesion, *Xanthomonas*, ecology, phylogeny, genealogy