Etude du polymorphisme associé aux répétitions en tandem pour le typage de bactéries pathogènes : Pseudomonas aeruginosa et Staphylococcus aureus

Lucie Onteniente

UNIVERSITÉ D'EVRY VAL D'ESSONNE

THÈSE

Pour obtenir le grade de

DOCTEUR EN SCIENCES
DE L'UNIVERSITÉ D'EVRY VAL D'ESSONNE

Spécialité : Biologie cellulaire et moléculaire

Présentée par

Lucie ONTENIENTE

Etude du polymorphisme associé aux répétitions en tandem pour le typage de bactéries pathogènes : *Pseudomonas aeruginosa* et *Staphylococcus aureus*

Soutenue le **13 février 2004** devant la commission d'examen :

Mr Jean-Didier Cavallo	Rapporteur
Mme Véronique Vincent	Rapporteur
Mr Sylvain Brisse	Examinateur
Mr Philip Supply	Examinateur
Mr Francis Quétier	Président du jury
Mr Gilles Vergnaud	Directeur de thèse

Cette thèse a été réalisée dans le laboratoire « Génomes, Polymorphisme et Minisatellites » à l'Institut de Génétique et Microbiologie (IGM) à Orsay sous la direction de Gilles Vergnaud, que je tiens à remercier pour les moyens matériels mis à ma disposition pendant ces trois ans de travail.

Je suis très honorée d'avoir pu compter six éminents chercheurs et enseignants-chercheurs dans mon jury de thèse. Leur disponibilité et leurs remarques pertinentes m'ont permis de passer dans de bonnes conditions les dernières étapes de la thèse.

J'adresse en outre mes remerciements à tous ceux qui ont collaboré de près ou de loin à ce travail : Sylvain Brisse du laboratoire des Bactéries Pathogènes Emergentes de l'Institut Pasteur, pour sa disponibilité et les échanges fructueux sur le projet Pseudomonas, Névine El Solh et Anne Morvan du Centre National de Référence des Staphylocoques de l'Institut Pasteur pour leur expertise et leur accueil, Jean Louis Koeck du HIA Val de Grâce pour m'avoir fourni des souches de Staphylococcus aureus, Philippe Bouloc et Justine Collier de l'IGM pour leur précieux conseils en microbiologie ainsi que Christine Pourcel pour les corrections du manuscrit.

Une mention plus particulière pour celles et ceux qui ont contribué par leur présence, leur soutien, leurs humeurs (bonnes ou mauvaises), leurs conseils et leur aide (technique ou morale) à rendre agréable le séjour quotidien au laboratoire et presque supportables les moments difficiles que j'ai pu y traverser. Pêle-mêle et sans ordre de préférence : Philippe, Yolande, Sophie, Justine, Sylvie, et plus récemment Olivier. Évidemment, un grand merci à France, qui a toujours été d'un soutien sans faille et avec qui j'ai particulièrement apprécié de travailler tout au long de ces trois années de thèse.

Que soient remerciées également Marie-Claude, Marie-Christine et Murielle, les secrétaires ; ainsi que Suzie la gestionnaire du bâtiment 400.

Merci à Bernard Mignotte et Gilles Waksmann de l'école doctorale « Des Génomes aux Organismes » sans qui cette thèse n'aurait pas été possible dans ce laboratoire.

Une pensée émue pour Edmond et Francine Puvion de l'Institut de Recherche sur le Cancer à Villejuif pour leur accueil exceptionnel lors de mon premier stage en laboratoire en 1997, et les nombreux conseils qu'ils m'ont donnés tout au long de mes études universitaires.

J'associe également à ces remerciements les personnes extérieures au laboratoire dont je vais tenter de faire la liste (là encore sans ordre de préférence) : Anastasia et Adrien, Fred et Dom, Steph, Fatou et Sébastien, Jérôme et Juliette, Simon, Blandine ; mes amis de l'IGR : Laurence et Hervé, Valérie, Virginie, Benoit ; mes amis des Maternelles d'été : Manu, Djoul, Etienne, Youri et Valérie ; mes amies d'Evry : Laetitia et Valérie ; et bien sur les BioDocsciens : Latif, Frank, Christophe, Etienne, François, Marc, Nicolas L et Nicolas B, et les BioDocsciennes : Estelle, Véronique, Rosa, Carine, Virginie et Christine avec qui j'ai partagé de bons moments de débat et de franche rigolade. Je suis pleinement convaincue que le temps que j'ai consacré, parallèlement à mon travail de thèse, à la vie associative m'a apporté et m'apportera encore bien plus qu'il ne m'a coûté.

Je termine en remerciant ma famille, en particulier ma mère, mon père et Yoyo mon petit frère, de m'avoir donné tous les moyens nécessaires pour en arriver là et pour leur soutien constant. Avec une pensée pour mes grands-parents qui, je pense, auraient été fiers de leur petite-fille.

Et enfin merci à Bertrand qui a vécu une deuxième fin de thèse… Merci pour la préparation à la soutenance et la recherche de coquilles dans le manuscrit, et surtout pour sa patience, son soutien et son humour en toute circonstance.

J'allais oublier, mais je remercie aussi le Stade de France et Roland Garros pour tous ces grands moments de sport que j'y ai vécu et que j'y vivrai ; les efforts, les peines et les joies des sportifs en ces lieux me renvoient à ceux que j'ai éprouvés durant cette thèse.

Et merci à Pétillon qui chaque semaine me fait rire avec ses dessins.

Table des matières

Liste des abréviations

ADN : acide désoxyribonucléique

AFLP : amplified fragment length polymorphism

ARN : acide ribonucléique

ARNm : ARN messager

BLAST: basic local alignment search tool

CRISPR : clustered regularly interspaced short palindromic repeat

DR: direct repeat

ETR: exact tandem repeat

GISA: glycopeptide intermediate *Staphylococcus aureus*

GOLD : genomes online database

IS : insertion sequence

kb : kilobase

LPS : lipopolysaccharide

Mb : mégabase

MIRUs: mycobacterial interspersed repetitive units

MLST : multilocus sequence typing

MLVA : multilocus VNTR analysis

MMR : mismatch repair

MRSA : methicilline resistant *Staphylococcus aureus*

MSSA : methicilline sensible *Staphylococcus aureus*

MSCRAMM: microbial surface components recognizing adhesive matrix molecules

NCBI : national center for biotechnology information

ORF : open reading frame

PEG : polyéthylène glycol

pb : paire de bases

PCR : polymerase chain reaction

PIC : polymorphism information content

RAPD : random amplified polymorphic DNA

RFLP : restriction fragment length polymorphism

SNP : single nucleotide polymorphism

SPE: serial-passage experiments

SSM : slipped-strand mispairing

SSR : short sequence repeat

STAR: *Staphylococcus aureus* repeat

STR : short tandem repeat

TIGR : the institute for genomic research

TRDB : tandem repeats database

TRF : tandem repeats finder

VNTR : variable number of tandem repeats

VRSA : vancomycine resistant *Staphylococcus aureus*

WGS : whole genome shotgun

1 INTRODUCTION

1.1 Le génotypage des bactéries pathogènes

1.1.1 Pourquoi s'intéresse-t-on au génotypage des bactéries pathogènes ?

Face à une infection, l'identification de l'agent microbiologique en cause, bactérien ou viral, est le premier objectif. Cette étape est de loin la plus urgente et la plus importante à court terme puisque elle intervient directement dans la prise en charge appropriée du patient. Elle est réalisée sur site clinique, voire, si la pathologie en laisse le temps, en laboratoire d'analyse. Les hôpitaux disposent pour cette raison de services de microbiologie clinique importants. L'identification bactérienne constitue donc un champ majeur de l'activité clinique, et la majorité des espèces bactériennes pathogènes pour l'homme (quelques centaines) est maintenant bien connue.

Dans un deuxième temps cependant, il est également essentiel de rechercher l'origine de l'agent infectieux, afin si possible de « tarir » l'éventuelle source. Les sources peuvent être l'environnement médical (matériel médical et personnel soignant), l'environnement urbain (canalisations d'eau, tour de refroidissement dans le cas de la légionellose par exemple), ou même des pays aux conditions sanitaires mal contrôlées (épidémie dite du SRAS, Syndrome Respiratoire Aigu Sévère, ou en fin d'année 2002, souches de *Mycobacterium tuberculosis* très résistantes aux antibiotiques venant de Chine). Cette enquête *a posteriori* requiert le typage des souches.

Avant d'entrer dans ces notions de typage de souches, il peut être utile de revenir brièvement sur la notion même d'espèce bactérienne. Bien qu'elle puisse sembler claire à première vue, puisque l'on considère comme naturel qu'une bactérie fasse partie d'une espèce, il apparaît rapidement dès que l'on se confronte à la question d'identification bactérienne que la situation n'est pas si simple. Ceci d'ailleurs est bien reflété par les nombreuses controverses de nomenclature qui agitent les taxonomistes, et les réassignations bactériennes fréquentes. Nous tâcherons dans les quelques paragraphes qui suivent d'expliciter sans entrer dans les détails les raisons de ces difficultés.

La notion d'espèce n'est pas sans ambiguïtés même pour des espèces à reproduction sexuée pour lesquelles pourtant le critère d'interfécondité aurait pu sembler devoir être définitif et absolu. Il n'est donc pas surprenant que chez des êtres vivants qui se reproduisent par simple division cellulaire la situation soit plus délicate. Ce qui est le plus étonnant en définitive est que la notion d'espèce bactérienne a une certaine utilité pratique. Ceci est probablement lié au fait qu'une infime fraction des bactéries existant sur notre planète est pathogène pour l'homme. Ces rares bactéries ont trouvé une niche écologique. A partir d'ancêtres précurseurs,

des populations essentiellement clonales se sont développées. Dans un contexte médical, les représentants de ces populations donnent une image d'homogénéité phénotypique et par conséquent une valeur pratique à la notion d'espèce. Dans un contexte plus large, prenant en compte l'environnement, il s'avère souvent que ces bactéries pathogènes ont des voisines non pathogènes génétiquement très proches. Alors que nos connaissances des bactéries de l'environnement s'accroissent, le nombre de telles situations s'accroît également. Ceux qui voudraient définir de façon très formaliste la notion d'espèce, sur des critères de distance génétique en particulier, souhaitent ainsi imposer de rebaptiser par exemple *Yersinia pestis*, l'agent causal de la peste, en *Yersinia pseudotuberculosis*, l'espèce d'où est issue *Yersinia pestis* il y a quelques milliers d'années. Dans ce cas pourtant, la niche écologique particulière (la puce et le rat), et évidemment les conséquences en pathologie humaine, justifient de conserver la nomenclature traditionnelle. De même, chez *Brucella*, six espèces sont définies, chacune se caractérisant par un tropisme d'hôte particulier, alors que les distances génétiques entre espèces sont minimes (Moreno 2002). Il faudra donc probablement s'accommoder pendant encore longtemps de situations contradictoires, et accepter que chaque espèce bactérienne soit un cas particulier, lié à un écosystème.

Le travail qui va être présenté dans cette thèse ne porte pas sur l'identification bactérienne, mais sur le typage de souches. Pour cette raison, même si, comme nous l'avons rappelé très brièvement, les considérations sur les notions d'espèce sont très intéressantes, elles ne seront pas abordées plus avant. La question du typage de souches ne se pose que pour une population bactérienne relativement proche, ce qui est le cas dans un contexte clinique. En l'occurrence, nous avons dans ce travail abordé la question du typage pour trois agents infectieux, *Y. pestis*, *P. aeruginosa* et *S. aureus*. La première bactérie citée est l'agent causal de la peste, qui demeure associé à plusieurs milliers de cas chaque année dans le monde, les deux autres sont responsables de nombreuses infections acquises en milieu hospitalier. Comme nous l'avons rappelé, alors que la prise en charge médicale immédiate se satisfait largement de l'identification bactérienne complétée par la connaissance du profil de résistance aux antibiotiques de la bactérie incriminée, le contrôle des maladies infectieuses requiert une caractérisation plus fine des souches.

Pour une surveillance épidémiologique efficace, il est nécessaire de pouvoir identifier avec le plus de précision possible les souches bactériennes responsables d'épidémies à l'échelle planétaire ou locale. Dans le premier cas, la connaissance de l'origine de souches responsables de maladies telles que la tuberculose peut permettre de mettre en place des mesures sanitaires appropriées dans les pays concernés, voire même d'exercer une pression internationale sur ces pays pour qu'ils améliorent leur prise en charge des maladies infectieuses. Dans le second cas, le typage de souches permet de prendre des mesures très locales (mesures sanitaires dans les hôpitaux, identification de porteurs, identification de foyers d'accueil à risque, identification de systèmes de canalisations contaminés). Pour ces

deux types d'études, épidémiologie locale et épidémiologie globale, l'identification de marqueurs polymorphes est une étape préalable. De façon plus anecdotique, mais qui peut mériter d'être rappelée dans le contexte actuel, le typage des bactéries pathogènes potentiellement utilisables comme armes biologiques s'est avéré nécessaire pour identifier le plus précisément possible l'origine de la souche de *Bacillus anthracis* envoyée dans des enveloppes à l'automne 2001 aux Etats-Unis.

1.1.2 Le concept de clonalité

Les épidémies observées dans les cas de maladies infectieuses ont souvent pour origine un même agent étiologique. Généralement, l'agent étiologique responsable d'une épidémie dérive d'une cellule unique, toutes les cellules issues de cette cellule seront génétiquement identiques ou très proches. Les organismes impliqués dans une épidémie ont donc une relation de clonalité. Ils appartiennent à la même espèce et présentent des caractéristiques communes comme par exemple des facteurs de virulence, des propriétés biochimiques et des caractéristiques génomiques. Cependant on observe une diversité suffisante pour distinguer des isolats de même espèce provenant de différents lieux et prélevés à différents moments, ce qui permet de les classer en souches. On peut ainsi obtenir un typage plus approfondi qu'une simple détermination de l'espèce impliquée dans une épidémie. Ceci correspond à une démarche de sous-typage nécessaire pour reconnaître une situation d'épidémie, détecter les contaminations croisées de pathogènes responsables d'infections nosocomiales, déterminer la source de l'infection et enfin participer au développement de vaccins (Olive 1999).

Par ailleurs, pour le typage dans un contexte d'étude épidémiologique, les génotypes sont déterminés à partir de plusieurs locus suffisamment stables. Si la recombinaison est fréquente dans l'espèce considérée, on trouve les caractéristiques d'une population panmictique (association aléatoire des allèles) comme c'est le cas pour *Neisseria gonorrhoeae*. L'étude de marqueurs génétiques permet différentes applications en microbiologie (Tibayrenc 1998) :

- Le typage de souches.

- L'enrichissement des données de taxonomie, pour mieux définir les concepts d'espèce et de sous-espèce, pour vérifier la taxonomie actuelle et pour comprendre le rôle évolutif de la recombinaison.

- L'analyse des liens entre variabilité génétique et diversité biologique, pour mieux comprendre la pathogénicité, les interactions hôtes/parasite, la résistance aux antibiotiques, la perte d'efficacité des vaccins.

1.1.2.1 Différentes structures de population

Le déséquilibre de liaison correspond au réassortiment non aléatoire des allèles à différents locus, en fonction de leur position physique sur le chromosome. En absence de recombinaison, on observe un déséquilibre de liaison entre les marqueurs. Ce déséquilibre est très important dans les populations clonales. Les données MLEE montrent que beaucoup d'espèces présentent un fort déséquilibre de liaison (Smith 1993). Superficiellement, ceci suggère que la structure de nombreuses populations bactériennes est clonale, le taux de recombinaison du chromosome bactérien est donc trop faible pour permettre une distribution aléatoire des allèles. Pour vérifier ceci, des données d'études MLEE réalisées sur plusieurs populations bactériennes ont été analysées avec un test statistique afin de déterminer s'il y a effectivement des associations entre différents locus dans ces populations (Orskov 1983).

Il existe des populations dans lesquelles un déséquilibre de liaison est observé et pourtant la recombinaison est fréquente. Ceci a été observé par exemple lorsque l'échantillon étudié est constitué d'un mélange de populations dans lesquelles est observée de la recombinaison dans chaque population mais pas entre les populations. L'absence de recombinaison entre ces populations peut être due au fait qu'elles sont isolées géographiquement ou écologiquement, ou bien que des barrières biologiques empêchent tout échange génétique entre elles. Autre situation dans laquelle on conclut à une population clonale alors que ça n'est pas le cas : les populations épidémiques. Il s'agit d'une situation particulière, un génotype donné va se développer et se disséminer très vite, conduisant à observer un déséquilibre de liaison temporaire. Lorsque l'on supprime de l'analyse les types électrophorétiques récemment disséminés, on voit que la population n'a pas une structure clonale. Il peut y avoir maintien d'un déséquilibre de liaison entre locus qui ont une interaction épistatique. Enfin la dérive génétique peut aussi conduire à un déséquilibre. L'analyse de plusieurs jeux de données MLEE a conduit à conclure à l'existence de quatre types de populations :

- population panmictique (index d'association entre les marqueurs proche de 0), il n'y a pas de déséquilibre de liaison, les allèles sont distribués aléatoirement. Ex : *Neisseria gonorrhoeae*
- population épidémique : elle est panmictique mais apparaît clonale en situation d'épidémie puisque le génotype responsable de l'épidémie est très fortement surreprésenté dans la population étudiée au moment précis de l'étude. Ex : *Neisseria meningitidis*
- population clonale : fort déséquilibre de liaison, population clonale à tous les niveaux d'analyse. Ex : le genre *Salmonella*
- déséquilibre de liaison observé dans une population constituée de sous-populations dans lesquelles il y a de la recombinaison interne à chaque sous-population mais pas entres elles. Ex : *Rhizobium meliloti*

Pour conclure, les populations ne sont pas invariablement clonales, elles peuvent occuper un large spectre de structures possibles allant d'une population strictement clonale à une population panmictique. C'est pourquoi on peut parler de clones uniquement pour un jeu donné de marqueurs génétiques. La Figure 1 illustre trois structures de populations bactériennes.

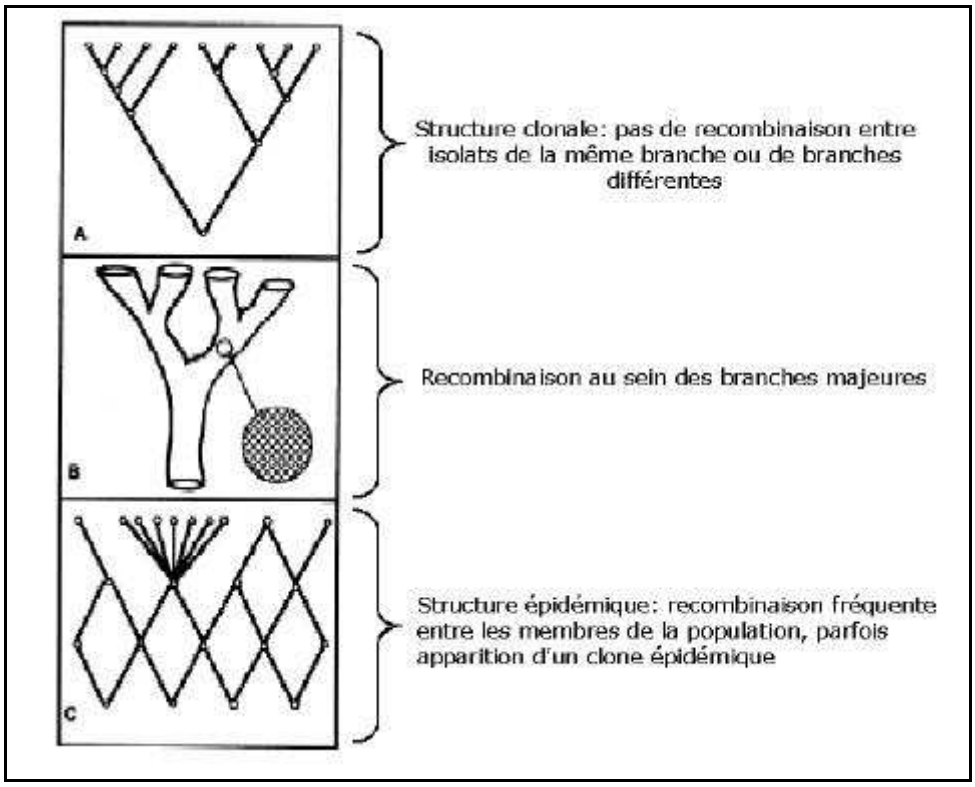


Figure 1: Structure des populations bactériennes (d'après (Smith 1993)).

Reste la question de savoir si les populations clonales sont strictement clonales. Des événements de transfert horizontal ont été mis en évidence par analyse de séquences chez *E. coli* (Smith 1991) et *Salmonella* (Smith 1990). Cependant, la fréquence de ces événements ne permet pas de supprimer le déséquilibre de liaison.

Un exemple récent d'étude de déséquilibre de liaison a été réalisé chez *Mycobacterium tuberculosis,* un des pathogènes humains les plus répandus dans le monde, afin de mieux connaître la structure de la population (Supply 2003). Pour pouvoir faire cette étude statistique, il faut disposer de marqueurs polymorphes, or la plupart des gènes de structure chez *M. tuberculosis* est peu polymorphe. Les marqueurs utilisés pour cette étude sont 12 locus VNTRs (Variable Number of Tandem Repeats) (Supply 2001). La population de souches étudiée vient d'une région où la tuberculose a une forte incidence. Les résultats montrent un fort déséquilibre de liaison entre ces marqueurs, ce qui conduit à conclure que *M. tuberculosis* a une évolution clonale.

Chez *Helicobacter pylori*, les échanges génétiques sont si fréquents que les locus sont tous à l'équilibre de liaison, *H. pylori* est donc une espèce panmictique. La recombinaison est beaucoup plus fréquente chez *H. pylori* que chez *Escherichia coli* ou *Neisseria meningitidis*. (Suerbaum 1998). Chez *N. meningitidis*, la recombinaison a été estimée à une fréquence 80 fois plus élevée que la mutation ponctuelle d'un nucléotide. L'estimation pour *E. coli* est de 10 à 50 fois supérieure et pour *S. pneumoniae*, 50 fois supérieure (Feil 1999), (Feil 2000).

Nous reviendrons dans la discussion sur la structure des populations des deux espèces qui nous intéressent dans ce travail.

1.1.2.2 Vitesse de mutation des marqueurs

La vitesse de mutation des marqueurs est à prendre en considération pour utiliser les marqueurs adaptés au type de questions posées. Comme le montre la Figure 2, plus le marqueur évolue vite, plus on observe de génotypes différents dans une population donnée. Les marqueurs qui ont une vitesse d'évolution rapide sont utiles pour des études épidémiologiques à court terme et au niveau local (un hôpital, une ville), par exemple dans des situations épidémiques. Les marqueurs qui évoluent lentement sont utiles pour des études épidémiologiques sur des mois ou des années, ainsi qu'au niveau mondial, pour les études de phylogénie.

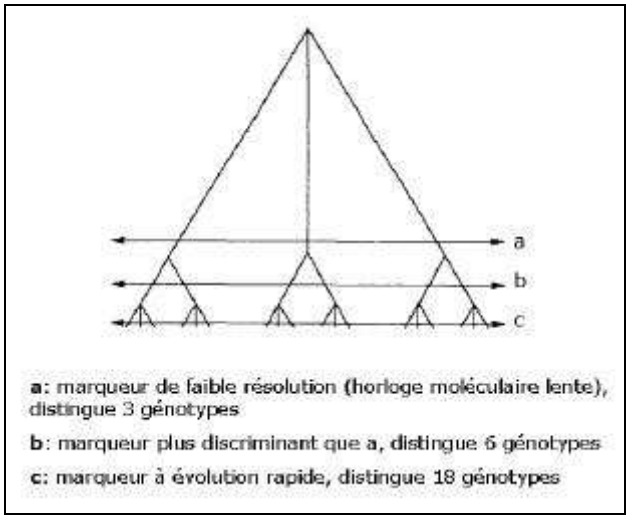


Figure 2: Arbre phylogénétique hypothétique décrivant la divergence évolutive parmi des génotypes de pathogènes (d'après (Tibayrenc 1998)).

Lorsque de nouveaux marqueurs sont décrits, le plus souvent, une étude de leur stabilité au cours du temps est réalisée. Une méthode simple est le SPE (pour Serial Passage Experiments) qui consiste à faire des dilutions en série d'une culture bactérienne pendant plusieurs semaines pour suivre les changements génotypiques et phénotypiques au cours du temps. On peut par exemple comparer les génotypes des cultures prélevées en début

d'expérience et après des centaines de générations. Ceci a été fait pour la répétition localisée dans le gène *spa* chez *S. aureus* (Frenay 1996). Lenski a également réalisé une étude sur *E. coli* après 20000 générations (Lenski 2003).

Des techniques de typage de souches ont été mises en œuvre dès les débuts de la microbiologie. Parmi les principales on peut citer le sérotypage, le typage de résistance aux phages, les tests biochimiques. La découverte du rôle de l'ADN comme support du patrimoine génétique a ouvert la voie au développement de techniques dites de génotypage. On peut noter que ces techniques vont souvent tirer parti des connaissances antérieures avec des méthodes plus simples à mettre en œuvre : la connaissance des génomes bactériens permet maintenant d'identifier l'origine d'une caractéristique biochimique, et de développer une méthode de typage par PCR (Polymerase Chain Reaction), par exemple, plus simple que le typage biochimique original. Dans les prochaines pages, ces aspects seront abordés plus en détail. Puis nous présenterons les espèces bactériennes, *Y. pestis*, et surtout *P. aeruginosa* et *S. aureus*, dont l'étude et le typage constituent l'essentiel du travail effectué dans le cadre de cette thèse.

1.2 Glossaire des principales techniques d'épidémiologie

L'identification bactérienne consiste à déterminer l'espèce de la bactérie étudiée, le typage consiste à distinguer des souches au sein d'une même espèce.

Le typage des souches peut se faire à différents niveaux. Des méthodes de phénotypage sont utilisées classiquement, comme par exemple les antibiogrammes et le sérotypage, mais pour cela, les marqueurs sérologiques doivent être présents sur toutes les souches, ce qui n'est pas systématiquement le cas. Il existe aussi des tests de phénotypage propres à chaque bactérie selon une activité enzymatique spécifique de l'espèce considérée. Les principales techniques de phénotypage seront exposées dans le paragraphe 1.2.1.

Les méthodes de génotypage, souvent associées à des sigles (RFLP, VNTR, RAPD, AFLP, PFGE etc.) permettent l'étude du polymorphisme au niveau de l'ADN. Elles seront décrites dans le paragraphe 1.2.2. Au cours de cette thèse, des marqueurs polymorphes ont été développés pour le génotypage de souches de *Y. pestis*, *P. aeruginosa* et *S. aureus*.

Ainsi, parmi les critères à prendre en considération pour le choix d'une technique de typage de souches, il faut que celles-ci soient toutes typables par la méthode choisie et que cette dernière soit suffisamment discriminante. Elle doit permettre de distinguer des souches qui ne sont pas épidémiologiquement reliées, c'est à dire par exemple de provenances géographiques différentes. Une technique sera aussi retenue sur un critère de reproductibilité. Ceci est très

important pour la comparaison des résultats entre laboratoires et pour la mise en place de bases de données qui doivent reposer sur des données solides. Le délai d'obtention des résultats et le coût sont également des critères à considérer pour évaluer le développement possible de la technique dans des laboratoires d'analyses hospitaliers et pas uniquement dans un contexte de laboratoires de recherche. Ces critères de choix d'une technique de typage seront résumés dans le paragraphe 1.2.3.

Les techniques d'épidémiologie moléculaire regroupent les techniques biochimiques et les techniques de biologie moléculaire. Il faut faire la distinction entre des approches généralistes (exemples : RAPD, isoenzyme electrophoresis) qui permettent une comparaison directe entre différentes espèces et permettent aussi de typer des souches d'une même espèce, et des approches spécialisées qui elles sont spécifiques d'une espèce donnée (exemple : le spoligotypage développé pour *M. tuberculosis*).

1.2.1 Techniques de microbiologie classiques : phénotypage

Les différentes techniques décrites ci-dessous sont utilisées pour identifier ou typer des souches en systématique bactérienne et en épidémiologie. Elles détectent des variations phénotypiques et il n'est pas toujours possible de relier une variation allélique à un gène spécifique, d'où parfois la difficulté de réaliser une analyse génétique de la structure des populations étudiées.

1.2.1.1 Les tests de résistance aux antibiotiques

Cette méthode est utilisée en routine dans les laboratoires de microbiologie, pour des raisons évidentes de prise en charge médicale. Il s'agit de déterminer le profil de résistance d'une souche vis à vis de plusieurs antibiotiques. Pour cela, des disques correspondant chacun à un antibiotique à tester sont déposés sur un étalement de bactéries. Après 24 heures, le diamètre de la zone d'inhibition autour de chaque disque imprégné d'antibiotique est mesuré et, selon sa taille, on peut déterminer par comparaison avec des données de référence si la souche est résistante ou sensible à l'antibiotique testé. L'antibiogramme de la souche analysée va permettre d'orienter le choix des antibiotiques à utiliser pour traiter le malade.

1.2.1.2 Le sérotypage

Le sérotypage des souches consiste à comparer le comportement de la surface membranaire de différentes souches par des tests d'agglutination sur lames avec des mélanges de sérums correspondant aux différents types antigéniques connus. Il y a par exemple 16 sérogroupes O chez *Pseudomonas aeruginosa*. Une difficulté importante dans le cas de *P. aeruginosa* est que

dans certaines souches, dites mucoïdes, une sorte de mucus rend la surface inaccessible et le test inefficace (Young 1974).

1.2.1.3 Le lysotypage (ou phage typing)

Cette technique permet de tester la sensibilité des souches à l'infection par des phages. C'est une technique utilisée depuis les années 1960. Elle permet de classer les souches testées en différents lysotypes (Sutter 1965). Chez *Staphylococcus aureus*, le lysotypage est une technique toujours utilisée et, récemment, de nouveaux phages ont été identifiés pour le typage (de Gialluly 2003).

1.2.1.4 MLEE (MultiLocus Enzymes Electrophoresis)

La tehnique dite MLEE a longtemps été la méthode standard d'étude de génétique de populations eucaryotes et de systématique. A partir des années 1980, des études MLEE de populations bactériennes ont été réalisées pour plusieurs espèces (*Escherichia coli*, *Bordetella* spp., *Haemophilus influenzae*, *Neisseria meningitidis*, *Legionella* spp., *Pseudomonas aeruginosa*, *Staphylococcus aureus*, etc…) (Selander 1986). Les isolats sont caractérisés par la mobilité électrophorétique d'un grand nombre d'enzymes cellulaires solubles. La charge nette et la distance de migration de la protéine sont directement liées à sa séquence en acides aminés. Les variants de mobilité d'une enzyme peuvent être associés aux allèles correspondant au locus du gène, plus facilement que pour d'autres caractères phénotypiques étudiés. L'électrophorèse permet de détecter un grand nombre de substitutions d'acide aminés (80 à 90%), ceci a été vérifié pour des enzymes de séquences nucléiques connues (Ramshaw 1979). Un certain nombre de substitutions sont sans effet sur la migration électrophorétique.

1.2.1.5 Tests biochimiques

Certains tests biochimiques sont développés à partir d'une activité enzymatique, ou d'une production de pigments spécifiques d'une espèce bactérienne. Ces tests sont essentiels en identification bactérienne médicale et peuvent permettre dans ce contexte relativement simple de faire en première intention une identification d'espèce et parfois aussi un typage succinct des souches. Ces tests sont automatisés. En effet, dans un contexte médical, comme il n'existe que quelques centaines de bactéries pathogènes pour l'homme, il est possible de constituer une base de données de profils relativement complète, et bien adaptée à l'identification bactérienne (plutôt qu'au typage proprement dit). L'utilisation de galeries API permet l'identification de l'espèce bactérienne en testant un certain nombre d'activités métaboliques, comme par exemple l'utilisation des sucres. D'autres tests sont orientés vers une espèce bactérienne en particulier. Pour *S.aureus*, on peut tester la production de staphylocoagulase.

Ce test biochimique est utilisé depuis les années 1960. Des tests rapides pour l'identification de *S. aureus* par la staphylocoagulase sont développés (Holliday 1999).

La production de pyocines, agents anti-bactériens, est utilisée comme méthode de phénotypage (pyocin typing) de *P. aeruginosa* chez qui différents « pyocin types » sont connus (Bruun 1976).

1.2.2 Les techniques de génotypage

Les techniques de génotypage (analyse de l'ADN) prennent une part croissante par rapport aux techniques de phénotypage, du fait d'une part des problèmes de « typabilité » de certaines souches et d'autre part des problèmes de reproductibilité rencontrés avec certaines techniques. L'arrivée de la PCR en 1985 (Saiki 1985) a encore accéléré cette tendance. En outre, le génotypage peut être « délocalisé » (c'est à dire se faire dans un autre laboratoire que le laboratoire de microbiologie) ce qui permet un partage efficace du travail. Cependant, comme nous le verrons dans le paragraphe 1.2.3, les techniques de génotypage ne sont pas toutes satisfaisantes pour le critère de reproductibilité.

L'accélération du développement du génotypage a aussi pour origine la disponibilité des données de séquençage des génomes complets qui ouvre la voie à l'étude plus exhaustive de marqueurs polymorphes tels que les répétitions en tandem et les SNPs (Single Nucleotide Polymorphism).

Les techniques courantes de génotypage mettent en œuvre l'électrophorèse de fragments d'ADN et l'analyse d'un profil de bandes.

1.2.2.1 PFGE (Pulse Field Gel Electrophoresis)

Avec cette technique, il n'y a pas d'amplification d'ADN par PCR. L'électrophorèse en champ pulsé de l'ADN chromosomique bactérien digéré par des enzymes de restriction est considérée comme la méthode de choix pour le typage moléculaire de nombreux pathogènes. Il s'agit d'une technique de RFLP (Restriction Fragment Length Polymorphism)-PFGE (électrophorèse en champ pulsé des fragments chromosomiques digérés).

Le champ pulsé permet de séparer des fragments d'ADN de très grande taille, de 30 à 2000kb. (Schwartz 1984). Les bactéries sont enrobées dans de l'agarose avant de subir une étape de lyse *in situ* afin de préserver l'intégrité physique de l'ADN. L'ADN est ensuite digéré par une enzyme de restriction à sites rares. Le morceau d'agarose contenant l'ADN bactérien digéré est placé en haut d'un gel d'agarose et l'électrophorèse en champ pulsé est réalisée. La polarité du courant est modifiée à intervalles réguliers ce qui permet de séparer des fragments

de grande taille. Le gel est coloré au Bromure d'Ethidium puis exposé sur une lampe UV pour visualiser les fragments d'ADN. Les photos des gels sont analysées à l'aide de logiciels. Les profils de restriction des différents isolats sont comparés entre eux pour déterminer leur proximité.

En 1995, Tenover a proposé un système de standardisation et d'interprétation des profils obtenus par champ pulsé pour déterminer la relation entre les isolats étudiés (Tenover 1995). Il faut d'abord déterminer quelle est la souche responsable de l'épidémie. Ensuite les « règles » suivantes d'interprétation des profils sont suivies :

- si deux isolats ont le même profil PFGE, on considère qu'il s'agit de la même souche.
- des souches sont considérées comme proches lorsqu'elles ont un seul événement génétique de différence, ce qui se traduit par une différence dans le profil au niveau de 2 à 3 bandes (addition ou délétion d'un site de restriction).
- des souches liées ont 4 à 6 bandes différentes, cela est dû à deux événements génétiques.
- des souches qui ont plus de 6 bandes de différence sont considérées sans lien épidémiologique.

Ces critères de Tenover ne sont pas vraiment reconnus (par Tenover lui-même), il semblerait qu'ils ne soient pas valables pour toutes les situations épidémiques.

De nombreuses études ont été menées pour déterminer quelle enzyme de restriction est la plus discriminante pour une espèce bactérienne donnée. Dans de nombreuses publications, l'enzyme utilisée pour les études PFGE de *S. aureus* est *SmaI* (Linhardt 1992) et pour *P. aeruginosa*, il s'agit de *SpeI* (Holloway 1992).

1.2.2.2 RAPD (Random Amplified Polymorphic DNA) ou AP-PCR (Arbitrarily Primed PCR)

Cette technique a été décrite pour la première fois en 1990 par Williams (Williams 1990) et simultanément par Welsh (Welsh 1990). Elle présentait l'intérêt de pouvoir amplifier par PCR différentes portions du génome sans connaître sa séquence. Le principe est basé sur l'utilisation d'amorces aléatoires courtes (9 à 10 pb) qui s'hybrident avec l'ADN chromosomique à faible température d'hybridation et vont permettre d'initier l'amplification d'un certain nombre de régions du génome. Les amorces utilisées sont déterminées de façon empirique. Le nombre et la localisation des sites d'hybridation des amorces varient d'une souche à l'autre dans une même espèce. Là aussi, les fragments amplifiés seront séparés par électrophorèse sur gel d'agarose. Les profils obtenus sont analysés et comparés. Le reproche majeur fait à cette technique est le manque de reproductibilité et de standardisation. De

nombreuses hybridations partielles ont lieu et une très légère variation des conditions de PCR (température, tampon, enzyme) conduit à des variations du profil de bandes obtenu. La reproductibilité est possible à l'intérieur d'un laboratoire, mais beaucoup moins entre laboratoires. Des variations ont été observées d'une machine PCR à l'autre (Meunier 1993).

1.2.2.3 AFLP (Amplified Fragment Length Polymorphism)

Cette technique implique trois étapes : tout d'abord la digestion de l'ADN génomique par une enzyme de restriction et la ligation d'adaptateurs aux fragments générés, puis l'amplification sélective de fragments de restriction et enfin une séparation électrophorétique des fragments amplifiés. L'amplification des fragments de restriction est réalisée à l'aide d'amorces qui s'hybrident au niveau des adaptateurs du site de restriction. On amplifie en général de 50 à 100 fragments qui seront ensuite détectés sur gels de polyacrylamide en conditions dénaturantes (Vos 1995) ou séquenceur à capillaires. Cette technique est très puissante, beaucoup plus robuste et reproductible que la précédente, et, comme cette dernière, ne requiert pas la connaissance de données de séquence génomique. Elle exige cependant de l'ADN de bonne qualité, et le respect de procédures rigoureuses.

1.2.2.4 RFLP (Restriction Fragment Length Polymorphism)

1.2.2.4.1 Typage par transfert et hybridation (Southern blot)

L'ADN génomique est digéré par une enzyme de restriction puis les fragments d'ADN sont séparés sur gel d'agarose et transférés sur membrane. Ensuite, l'hybridation est réalisée avec une sonde correspondant à un fragment du génome, un gène (par exemple *ToxA* pour *P. aeruginosa* (Grundmann 1995), ou des éléments mobiles : séquences IS). Cette approche a été largement développée au cours des années 1980, et reste encore parfois utilisée, par exemple pour le typage de polymorphisme associé à des éléments mobiles chez *Y. pestis* et *M. tuberculosis*. Le ribotypage est également une application de cette approche, dans laquelle la sonde utilisée correspond au locus de l'ADNr 16S-23S (Scieux 1992; Bingen 1994). L'ADN génomique est hydrolysé par deux enzymes de restriction au cours de réactions indépendantes. Les fragments sont ensuite séparés par électrophorèse puis transférés sur une membrane de nylon. Une sonde ribosomique 16S-23S d'*Escherichia coli* marquée est mise en contact avec la membrane pour révéler les fragments d'ADN homologues. Une à plusieurs bandes d'hybridation sont observées en fonction du nombre de sites de restriction présents dans l'opéron ribosomique des souches testées. Ces différentes étapes ont été automatisées et sont réalisées avec un « RiboPrinter » qui permet une reproductibilité satisfaisante entre laboratoires.

1.2.2.4.2 PCR-RFLP

Une PCR est réalisée en utilisant des amorces spécifiques d'un locus donné. Ensuite le produit de PCR est digéré par une enzyme de restriction. Les fragments obtenus sont séparés par électrophorèse. Les profils de bandes sont alors comparés. Par exemple chez *S. aureus*, les gènes de la coagulase, de la protéine A et de la région hypervariable proche de mecA ont été amplifiés puis digérés par *HaeII*, puis les profils comparés (Wichelhaus 2001).

L'amplification de l'ADNr 16S puis la digestion du fragment PCR et la séparation des produits de digestion sur gel constituent une variante du ribotypage, de moindre résolution. Cette méthode est appelée ARDRA (Amplified Ribosomal DNA Restriction Analysis) (Vaneechoutte 1992). Elle ne permet pas en général de distinguer des souches au sein d'une même espèce, mais permet de classer grossièrement des bactéries. L'avantage de cette approche par rapport au Southern blot est qu'elle ne nécessite que de faibles quantités d'ADN, de qualité moyenne et qu'elle est simple et rapide.

1.2.2.5 MLST (Multi Locus Sequence Typing)

Il s'agit du séquençage d'environ 500 pb de fragments internes de gènes de ménage. En général 7 gènes sont séquencés par isolat. Ensuite les séquences sont comparées aux séquences déjà rencontrées et les isolats classés en types de séquences (ST pour « sequence types »). Cette technique est utile pour des études épidémiologiques, mais peut ne pas être assez discriminante pour des analyses en situation d'épidémie. Ces locus évoluent lentement du fait de leur présence dans des gènes de ménage. Ceci explique que cette technique MLST est valable pour des études de phylogénie (dépend des espèces) plutôt que pour du typage de routine. Le MLST n'est pas valable pour des espèces d'émergence récente comme par exemple *Y. pestis* ou *M. tuberculosis*, en revanche elle a été validée pour un certain nombre d'espèces comme *H. pylori* et *N. meningitidis*. Dans cette dernière espèce, l'approche est en passe de devenir la méthode de référence. Par ailleurs, cette technique est encore coûteuse (2 fois 7 séquences à effectuer) pour des analyses de routine dans des laboratoires d'analyse hospitaliers (Maiden 1998). L'énorme avantage de cette approche est qu'elle est parfaitement reproductible quel que soit le laboratoire puisqu'il s'agit de séquencer et de classer les séquences. Il n'y a pas d'ambiguïté d'interprétation des résultats comme avec des profils de migration multibandes. Le typage MLST a été initialement développé pour l'étude de *Neisseria meningitidis* du fait de la difficulté de comparer des résultats de MLEE (MultiLocus Enzyme Electrophoresis) entre laboratoires. Il existe des bases de données en ligne pour le typage MLST (Chan 2001). Il est possible de soumettre des séquences sur le site www.mlst.net. Des données sont accessibles pour les bactéries pathogènes suivantes : *Neisseria meningitidis*, *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Campylobacter jejuni*, *Streptococcus pyogenes*, *Haemophilus influenzae*, *Burkholderia pseudomallei*,

Candida albicans, *Enterococcus faecium*, *Escherichia coli*, *Streptococcus agalactiae et Helicobacter pylori.*

Des puces à ADN ont également été développées pour mener des études MLST. Les données obtenues sont reproductibles et concordantes avec les données épidémiologiques disponibles pour les souches testées (van Leeuwen 2003). Il s'agit encore pour le moment d'approches « recherches » trop coûteuses pour un usage clinique.

1.2.2.6 SNPs (Single Nucleotide Polymorphism)

Le terme SNP désigne un changement de nucléotide à une position du génome, les insertions, délétions ou inversions ne sont pas considérées comme des SNPs (Brookes 1999). L'étude des SNPs s'est beaucoup développée avec le séquençage des génomes complets, permettant ainsi la comparaison de séquences de plusieurs génomes d'une même espèce pour rechercher les SNPs. Les SNPs identifiés par comparaison de génomes doivent être validés expérimentalement pour s'assurer qu'il ne s'agit pas d'erreurs de séquençage. Par exemple de nombreux SNPs ont été découverts chez *M. tuberculosis* par comparaison des séquences des souches H37Rv et CDC1551. Cette variabilité génétique peut avoir un rôle important pour la pathogénicité de l'espèce, par exemple lorsqu'un SNP est localisé dans un gène impliqué dans la virulence ou bien dans la réponse immunitaire de l'hôte (Fleischmann 2002).

Des SNPs ont également été étudiés chez *P. aeruginosa*. En effet, les genes exoS, exoT, exoU et exoY codant des toxines, présentent un SNP. Un test par PCR multiplexe a été développé pour étudier ces SNPs (Ajayi 2003).

La recherche de SNPs comme outil de génotypage pour des espèces très homogènes a aussi été réalisée comme par exemple pour *Bacillus anthracis,* une espèce d'émergence récente. Ainsi quelques SNPs ont été validés et constituent un outil complémentaire de typage de *B. anthracis*, en plus d'un certain nombre de VNTRs déjà décrits (Read 2002).

1.2.2.7 Amplification par PCR de séquences répétées en tandem

L'analyse MLVA ou Multi-Locus VNTR (Variable Number of Tandem Repeats) Analysis, est l'étude par PCR de plusieurs VNTRs. Il s'agit de répétitions en tandem qui présentent un polymorphisme de longueur. Le nombre de motifs répétés est variable d'une souche à l'autre. Ce type d'analyse a été développé dans ce travail pour *Pseudomonas aeruginosa* et *Staphylococcus aureus* (voir le chapitre 1.5) et pour *Yersinia pestis* (voir le chapitre 1.6).

1.2.3 Bilan des avantages/inconvénients des différentes méthodes de typage

Plusieurs paramètres sont importants pour évaluer une technique de typage (Tenover 1994). Tout d'abord la « typabilité », c'est à dire la possibilité d'obtenir un résultat pour chaque isolat analysé. Les isolats non typables ne sont pas interprétables. Ensuite, la reproductibilité, c'est à dire le fait d'obtenir le même résultat pour un isolat typé plusieurs fois. Une fois que ces deux exigences sont satisfaites, il faut aussi prendre en compte le pouvoir discriminant de la technique. Idéalement, chaque isolat non relié aux autres devrait être identifié comme unique. La facilité d'interprétation est également importante. En effet, l'épidémiologie à l'échelle mondiale nécessite de pouvoir comparer les données de typages réalisées dans des laboratoires différents. Une technique difficile à interpréter sera très difficilement mise à profit hors du laboratoire.

La faisabilité d'une technique repose sur plusieurs critères : l'accessibilité à la technique, les délais d'obtention des résultats, l'expertise technique nécessaire, le coût du matériel, les réactifs nécessaires, et enfin l'utilité de la technique pour un grand nombre de microorganismes (van Belkum 2001). Le Tableau 1 fait le bilan des avantages et inconvénients des différentes techniques de typage présentées dans le paragraphe précédent.

Les techniques de phénotypage sont dans l'ensemble moins faciles à interpréter que les techniques de génotypage, elles sont basées sur la présence ou l'absence d'activités métaboliques ou biologiques exprimées par le génome complet. Les résultats de génotypage sont le plus souvent des profils de bandes. Il faut tout de même insister sur le fait que les différentes techniques de génotypage ne fournissent pas toutes la même complexité de profil à analyser. Il y a également des différences au niveau de la reproductibilité selon la spécificité de la technique choisie. Par exemple la technique de RAPD est par définition moins stringente qu'une PCR réalisée pour amplifier un locus donné (conditions stringentes d'hybridation des amorces). La RAPD est satisfaisante en interne dans un laboratoire, mais les résutlats ne sont pas comparable entre laboratoires du fait des problèmes de reproductibilité.

D'une manière générale, les techniques de génotypage, basées sur la détection d'acides nucléiques, présentent l'avantage majeur que toutes les souches sont typables, c'est à dire même celles qui ne le sont pas par des techniques de phénotypage habituellement utilisées, celles qui sont difficilement cultivables, les non cultivables, et également les bactéries mortes.

Tableau 1: Avantages et inconvénients des différentes techniques de typage (adapté de van Belkum 2001)

	méthode de typage	"typabilité"	reproductibilité	pouvoir discriminant	facilité de mise en œuvre	facilité d'interprétation	accessibilité de la méthode	coût
PHENOTYPAGE	**Profils de résistance aux antibiotiques**	bonne	bonne	faible	excellente	excellente	excellente	faible
	Serotypage	variable	bonne	variable	bonne	bonne	variable	moyen
	Lysotypage	variable	passable	variable	faible	faible	excellente	moyen
	MLEE	excellente	excellente	bon	bonne	excellente	variable	élevé
	Tests biochimiques manuels	bonne	faible	faible	excellente	excellente	excellente	faible
	Tests biochimiques automatisés	bonne	bonne	faible	bonne	bonne	variable	moyen
GENOTYPAGE	**PFGE**	excellente	bonne	excellent	bonne	bonne	variable	élevé
	RAPD / AP-PCR	excellente	faible	faible	bonne	faible	bonne	moyen
	AFLP	excellente	bonne	excellent	bonne	passable	faible	élevé
	RFLP	excellente	variable	variable	bonne	passable	variable	moyen
	Ribotypage automatisé	excellente	excellente	bon	bonne	bonne	variable	élevé
	MLST (séquençage)	optimale	excellente	excellent	bonne	excellente	faible	élevé
	SNPs (séquençage)	excellente	excellente	faible	bonne	excellente	faible	élevé
	MLVA	excellente	excellente	excellent	excellente	excellente	bonne	faible

Pour les techniques de génotypage faisant appel à la PCR, un autre avantage est la très grande sensibilité et la spécificité de la technique permettant de travailler à partir de petites quantités de bactéries et ainsi d'éviter l'étape de mise en culture préalable pour des analyses nécessitant beaucoup de matériel, comme par exemple pour faire une analyse de type Southern blot. Cet aspect est particulièrement important pour des bactéries qui poussent très lentement comme *M. tuberculosis* (nécessite 2 à 3 semaines de culture). Un autre avantage de la PCR est la possibilité d'amplifier de l'ADN dégradé ou ancien. Cet avantage semble assez anecdotique mais permet par exemple de génotyper des souches très anciennes et ainsi de mieux connaître l'évolution des pathogènes d'intérêt (Fletcher 2003).

La seule technique permettant d'être absolument sûr que deux souches sont identiques reste le séquençage du génome complet.

Le développement d'une méthode de typage est largement guidé par des considérations pratiques. Le besoin doit être bien identifié. Dans le domaine de l'épidémiologie à échelle planétaire, il est clair qu'aucun laboratoire ou même pays ne peut espérer développer une approche mondiale sans étroite collaboration avec d'autres. Lorsque, comme c'est souvent le cas, par exemple pour la tuberculose, les foyers infectieux majeurs sont dans des pays relativement pauvres, il importe que les méthodes à mettre en œuvre soient peu coûteuses en terme de matériel et de consommables. Inversement, dans les pays les plus riches, les méthodes doivent être le plus automatisables possible en raison du coût élevé de la main d'œuvre. Enfin, les données qui en découlent doivent pouvoir être assemblées assez simplement et de façon fiable. Rares sont les méthodes qui répondent à ses critères, aucune n'est totalement satisfaisante, et actuellement ce domaine connaît une évolution très rapide.

La voie que nous explorons est celle des séquences répétées en tandem, pour un certain nombre de raisons que nous expliciterons dans le reste de ce travail.

1.3 Séquences répétées dans les génomes bactériens

Malgré leur petite taille, les génomes bactériens possèdent une grande variété de séquences répétées. Ils sont très compacts et présentent une grande densité de gènes par rapport aux génomes eucaryotes dans lesquels existe une forte proportion d'ADN extragénique (95% dans le génome humain). C'est pourquoi, à part les opérons ribosomiques connus depuis longtemps dans les génomes bactériens (souvent en plusieurs copies, 4 dans le génome de *P. aeruginosa* et 5 chez *S. aureus*), on pensait que les génomes bactériens étaient dépourvus de séquences répétées, puisque l'ADN est « cher » dans une bactérie qui doit assurer une réplication rapide de son génome, et que l'utilité de ces séquences répétées reste hypothétique. Ces dernières années cependant, différents types de séquences répétées ont été découvertes dans les génomes bactériens.

La variabilité, la complexité et la spécificité taxonomique des répétitions des génomes microbiens est similaire à celle trouvée dans les répétitions chez les plantes et les animaux. L'élément répété de base a une taille de 1 paire de bases à plusieurs kilobases. Les répétitions peuvent être dispersées dans le génome ou répétées en tandem à un seul locus. Les données de séquençage des génomes bactériens complets sont de plus en plus nombreuses et ouvrent la voie à l'étude exhaustive des répétitions dans ces génomes afin de comprendre leurs rôles.

1.3.1 Séquences répétées dispersées sur le génome

1.3.1.1 Les séquences d'insertion (IS)

Elles jouent un rôle majeur dans la plasticité des génomes procaryotes. Ces séquences, de moins de 2.5kb en général, codent des fonctions impliquées dans leur translocation dans le même génome et entre différents génomes. L'insertion d'une séquence IS dans un gène peut provoquer une inactivation de celui-ci. Ces éléments génétiques mobiles peuvent donc être à l'origine de réarrangements du génome tel que des délétions, des inversions. Ceci peut conduire à rassembler des gènes en clusters liés à une fonction spécialisée comme par exemple la virulence (Mahillon 1999). Chez *Yersinia pestis*, une espèce très riche en séquences IS, il a été montré que ces séquences participent à la plasticité du génome et qu'il existe une relation évolutive entre les souches caractéristiques des 3 grandes épidémies de peste (Antiqua, Medievalis et Orientalis) (Radnedge 2002).

Les séquences IS peuvent être utilisées comme sonde pour le typage de souches par Southern blot, comme par exemple IS256 chez *S. aureus* (Wei 1992) et aussi pour des analyses RFLP, comme par exemple les séquences IS6110 très étudiées chez *M. tuberculosis* (van Embden 1993).

1.3.1.2 Les séquences REP (Repetitive Extragenic Palindromic sequences)

Les régions intergéniques des génomes bactériens, et de la plupart des organismes, contiennent des séquences nécessaires pour le contrôle de la transcription et de la traduction. Ces séquences incluent les promoteurs et terminateurs de la transcription, les signaux de démarrage et d'arrêt de la traduction et des sites de fixations pour des protéines régulatrices.

Les séquences REP découvertes en 1984 (Stern 1984) sont aussi appelées PU pour Palindromic Units (Gilson 1984). La séquence REP a une longueur d'environ 35pb et possède une répétition inversée, elle peut être unique ou en copies multiples adjacentes. On trouve entre 500 et 1000 copies de séquences REP dans les génomes de *E. coli* et *S. typhimurium*, ce qui correspond à environ 1% de leur génome. Elles sont toujours situées dans des séquences transcrites ainsi que dans les régions intergéniques d'un opéron ou de la partie 3' non traduite du transcrit (Higgins 1982 ; Gilson 1987). Ces séquences auraient plusieurs rôles possibles : un rôle dans la terminaison de la transcription, un rôle de stabilisation du messager (les séquences REP pourraient constituer une barrière contre l'exonucléase 3'-5', et la stabilisation du messager qui en résulterait pourrait augmenter l'expression du gène) (Higgins 1988). Chez *E. coli*, les séquences PU présentent une grande similarité avec le consensus, et pourraient

constituer des sites de fixation pour des protéines. Ceci a été montré *in vitro* pour la gyrase par exemple (Yang 1988).

Les séquences PU, associées à d'autres séquences répétées, constituent les élements BIME (pour Bacterial Interspersed Mosaic Element). L'étude des BIME chez *E. coli* a conduit à définir deux familles de BIME (BIME-1 et BIME-2) selon les séquences PU qu'elles contiennent. Des expériences de retard sur gel ont montré que les différentes séquences PU ont des affinités différentes pour la gyrase, suggérant que ces deux familles de BIME sont fonctionnellement distinctes (Gilson 1991 ; Bachellier 1994).

1.3.1.3 Les séquences ERIC (Enterobacterial Repetition Intergenic Consensus)

Les séquences ERIC (Hulton 1991) sont aussi appelées IRUs pour Intergenic Repeats Units (Sharples 1990). Elles ont été décrites pour la première fois dans les génomes d'*E. coli* et de *S. typhimurium,* des bactéries à gram négatif. Ce sont des éléments de 126 pb qui possèdent dans la région centrale une répétition inversée fortement conservée. Les séquences ERIC permettent la formation de structures secondaires stables de type tige-boucles. Certaines séquences ERIC sont transcrites.

Les séquences consensus des éléments REP et ERIC sont totalement différentes (Hulton 1991). Versalovic a étudié la distribution de ces deux types de séquences répétées, ERIC et REP, dans différents génomes bactériens (Versalovic 1991). Les séquences REP et ERIC sont suffisamment conservées pour pouvoir y choisir des amorces consensus afin d'amplifier les régions situées entre les REP et les ERIC. Par PCR puis électrophorèse sur gel d'agarose, on obtient des profils de bandes. Les amorces ERIC et REP s'hybrident préférentiellement à l'ADN des bactéries à gram négatif. Les séquences REP et ERIC trouvées initialement dans des bactéries à gram négatif sont en fait conservées chez les eubactéries depuis des centaines de millions d'années. Leur répartition tout le long du génome constitue des sites de fixation pour des amorces consensus, permettant une identification rapide des espèces et souches bactériennes.

1.3.1.4 Les séquences BOX

Les séquences BOX ont été découvertes chez *Streptococcus pneumoniae* (Martin 1992). Elles sont au nombre de 25 environ dans le génome de *S. pneumoniae* et sont constituées de différentes combinaisons de 3 sous-unités : boxA (59pb), boxB (45pb) et boxC (50pb). Elles peuvent former des structures secondaires et pourraient avoir un rôle dans l'expression de ces gènes, elles seraient donc des séquences régulatrices.

Il a été montré que la sous-unité boxA est conservée dans de nombreuses espèces bactériennes, elle peut donc servir de cible pour des analyses de type rep-PCR (repetitive element PCR) afin de typer de nombreux microorganismes (Koeuth 1995). Ceci a été fait par exemple pour l'étude de souches de *Bacillus anthracis* et de *Bacillus cereus* (Kim 2002).

L'amplification des éléments BOX a montré un excellent pouvoir résolutif pour *S. pneumoniae* (van Belkum 1996). Auparavant, c'était la combinaison de 5 techniques de génotypage qui permettait de discriminer des souches de *S. pneumoniae* non reliées épidémiologiquement (Hermans 1995) ce qui n'était pas réalisable en routine pour un laboratoire d'analyses microbiologiques.

1.3.1.5 Les CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats)

Une famille de répétitions présente dans de nombreux génomes procaryotes (bactéries et archaés) a été mise en évidence (Mojica 2000). Ces éléments ont été appelés CRISPR (« clustered regularly interspaced short palindromic repeats »). Il s'agit de répétitions d'un motif bien conservé (de taille variant entre les génomes considérés : de 21 pb dans *S. typhimurium* à 37 pb dans *S. pyogenes*), espacées par des séquences non conservées, mais de longueur constante (du même ordre de taille que le motif conservé) (Jansen 2002): elles correspondent donc à une répétition en tandem ayant un motif répété comprenant une région bien conservée et une région variable (espaceurs). Dans un génome bactérien, les CRISPRs peuvent se retrouver en plusieurs localisations, avec un motif répété très conservé d'un locus à l'autre (même si le nombre de répétitions peut varier) (van Embden 2000). Dans certains cas, on trouve plusieurs types de CRISPRs, dont les motifs conservés sont différents, dans un même génome. Le nombre de répétitions peut également varier, en un même locus, d'une souche à l'autre : certains CRISPRs ont donc été exploités pour l'épidémiologie (Kamerbeek 1997). En revanche, d'une espèce bactérienne à l'autre (sauf espèces proches), les motifs conservés constituant les CRISPRs ont peu d'homologie : une exception est à noter entre *N. meningitidis* et *P. multocida*. Certaines caractéristiques de séquence sont cependant communes aux CRISPRs de toutes les espèces : le motif répété conservé contient une symétrie dyade (souvent GTT à une extrémité et AAC à l'autre extrémité). En général, une séquence flanquante de quelques centaines de pb est conservée entre tous les locus CRISPRs d'un génome donné, qui n'a pas d'homologie avec les flanquantes des CRISPRs d'autres génomes. Enfin, tous les génomes bactériens ne contiennent pas de CRISPRs, c'est le cas par exemple pour deux bactéries étudiées dans cette thèse, *P. aeruginosa* et *S. aureus*. La présence de ces CRISPRs est strictement associée à des gènes appelés *cas* (pour « CRISPR associated genes »), qui pourraient correspondre à des protéines se liant à l'ADN (Jansen 2002). Les mécanismes de création et de dispersion dans les génomes, ainsi que la fonction

biologique des CRISPRs, sont encore inconnus. Le Tableau 2 montre les caractéristiques des CRISPRs trouvés dans différents génomes bactériens (Mojica 2000).

Tableau 2 : Caractéristiques des CRISPRs trouvés dans différents génomes bactériens, (d'après (Mojica 2000))

Bactérie	Taille du motif conservé répété (a)	Taille de l'espaceur (b)	Taille de l'unité répétée en tandem (a+b)	Nombre de locus dans le génome	Nombre de répétitions en tandem / locus	Référence
T . maritima	30	39-40	69-70	8	2-40	(Nelson 1999)
A. aeolicus	29	36-38	65-67	1	6	(Mojica 2000)
E. coli	29	32-33	61-62	3	2/7/13	(Nakata 1989)
S. typhi	29	32	61	= 1	6	(Mojica 2000)
C. jejuni	36	30	66	1	5	(Mojica 2000)
Y. pestis	28	32-33	60-61	3	6/9	(Mojica 2000)
C. difficile	29	36-38	65-67	$4^{A} + 2^{B}$	5-17	(Mojica 2000)
M. tuberculosis	36	38-40	74-76	1	Variable	(Hermans 1991)
Calothrix sp.	37	35-41	72-78	>1	5	(Masepohl 1996)
Anabaena sp.	37	32-43	69-80	>1	17	(Masepohl 1996)

* A, B : CRISPRs de différents types (plus de 3 pb de différence dans le motif conservé) dans un même organisme.

Une méthode de typage du CRISPR localisé dans le génome de *M. tuberculosis,* ou DR pour Direct Repeat (Groenen 1993 ; van Embden 2000) est le « spoligotyping » (pour spacer oligonucleotide typing) (Goguet de la Salmonière 1997). Le spolygotypage permet de détecter la présence ou l'absence d'espaceurs dans ce locus. Les espaceurs de la répétition sont amplifiés par PCR et détectés par hybridation d'une membrane sur laquelle sont fixés les oligonucléotides correspondant aux différentes séquences des espaceurs connus (Kamerbeek 1997). C'est une méthode très reproductible mais lègèrement moins discriminante que la méthode de choix pour le typage de *M. tuberculosis*, la technique de RFLP au niveau des séquences IS6110 (Kremer 1999). Cependant, cette technique combinée avec une étude par PCR de locus VNTRs (Frothingham 1998) permet d'obtenir une très bonne résolution (Filliol 2000). Une étude de la distribution globale des spoligotypes de *M. tuberculosis* a été réalisée récemment. Une base de données de spoligotypes permet de rassembler des données à l'échelle internationale (Filliol 2002; Sola 2003).

Dans une autre étude récente réalisée chez *Campylobacter jejuni*, le séquençage du CRISPR unique dans son génome ainsi que les techniques de RFLP et de MLST ont été comparées. Les trois méthodes se sont montrées tout autant discriminantes. Cependant, les nombreux remaniements observés dans les séquences analysées par MLST (recombinaison 50 fois plus importante que les mutations ponctuelles), et la présence fréquente de plusieurs souches différentes dans un même prélèvement suggèrent que le typage des souches de *C. jejuni* n'est utile que pour des situations épidémiques et pas pour des études épidémiologiques plus globales (Schouls 2003).

1.3.2 Les séquences STAR (*Staphylococcus Aureus* Repeats)

On trouve ces séquences, identifiées dans le génome de *S. aureus,* en plusieurs exemplaires. Elles contiennent un élément répété en tandem (comme d'ailleurs les BOX de *S. pneumoniae*). Elles ont été identifiées par RFLP et Southern blot dans d'autres génomes du genre *Staphylococcus* (Cramton 2000). Elles sont localisées dans des régions intergéniques. On peut parler d'une famille de séquences répétées puisque la séquence consensus du motif répété est la même pour les différentes séquences STAR. Les flanquantes sont assez bien conservées également. Cette famille de répétitions en tandem sera détaillée au paragraphe 1.5.3.8.2.2.1 décrivant les répétitions en tandem chez *S. aureus*.

1.3.3 Séquences répétées à un seul locus du génome : les répétitions en tandem

1.3.3.1 Définition d'une répétition en tandem

Une répétition en tandem est une succession de motifs d'ADN répétés les uns derrière les autres, par opposition aux répétitions « dispersées » dont les unités répétées sont dispersées dans le génome. Les différentes unités formant la répétition en tandem ne sont pas nécessairement identiques entre elles : le degré d'homologie au sein d'une répétition en tandem est très variable. Les répétitions en tandem ont été en premier lieu étudiées chez les mammifères, où trois catégories ont été distinguées : les satellites, les minisatellites, et les microsatellites. Cette distinction correspond à différentes plages de taille (pour la longueur totale et la taille du motif) et a été faite de façon plus ou moins arbitraire : l'ADN « satellite » a tout d'abord été isolé par centrifugations sur gradient de densité (Britten 1968). Ensuite, des répétitions en tandem de taille inférieure, pouvant être analysés grâce à la technique de Southern Blot, ont été caractérisées (Wyman 1980) puis appelées « minisatellites » (Jeffreys 1985). Enfin, les répétitions en tandem de taille encore inférieure ont été nommées « microsatellites » lorsque l'avènement de la technique de PCR (« polymerase chain reaction ») en a fait des outils courants de génétique moléculaire. Les répétitions en tandem sont présentes, dans des proportions variables, chez tous les organismes : eucaryotes, procaryotes, et même virus. Dans ces deux derniers groupes, les mécanismes sous-jacents n'ont quasiment pas été étudiés, et la distinction entre microsatellites et minisatellites est rarement faite. D'autres termes sont couramment employés, chez les procaryotes, mais également parfois chez les eucaryotes, pour désigner les répétitions en tandem : les SSR (« simple sequence repeat ») et STR (« short tandem repeat ») désignent des répétitions en tandem « simples » : elles correspondent aux microsatellites et aux minisatellites de petite taille (quelques centaines de paires de bases). Les VNTRs (« variable number of tandem

repeats ») désignent les répétitions en tandem polymorphes, qui peuvent appartenir à la classe des microsatellites ou des minisatellites.

Dans la suite, nous ne distinguerons pas les différentes catégories de répétitions en tandem.

En raison de leur polymorphisme de longueur résultant de la variation du nombre de motifs, les répétitions en tandem ont conduit au développement des empreintes génétiques permettant l'identification humaine (Jeffreys 1985). L'intérêt des répétitions en tandem dans le domaine du typage est lié en bonne partie au fait que de nombreux allèles peuvent être observés en un locus, et que ces différents allèles peuvent être identifiés par la mesure de la taille d'un fragment d'ADN, opération très simple à réaliser. Les répétitions en tandem ont rendu possibles les études de liaison génétique et de cartographie du génome humain. On trouve des répétitions en tandem dans des séquences intergéniques et intragéniques. Les conséquences de la variation du nombre de motifs ne sont pas les mêmes selon la position de la répétition. Lorsque le minisatellite est situé dans la région 5' d'un gène, il peut réguler la transcription de celui-ci en éloignant plus ou moins le site d'initiation de la transcription. Lorsqu'il est localisé dans une phase ouverte de lecture, la protéine peut être plus ou moins longue, voire plus courte s'il y a introduction d'un codon stop. Il existe assez fréquemment une hétérogénéité des motifs constituant la répétition en tandem. Cette variation interne a été utilisée dans certains cas pour comprendre le mode de mutation de ces structures.

Les répétitions en tandem peuvent constituer des familles : des répétitions en tandem (ayant des motifs répétés similaires) peuvent se retrouver en plusieurs localisations sur les génomes bactériens, avec leurs flanquantes (Cramton 2000) ou sans leurs flanquantes (Supply 1997).

1.3.3.2 Méthode de recherche des répétitions en tandem avant le séquençage des génomes complets

Les répétitions en tandem qui ont été identifiées, essentiellement dans des génomes eucaryotes avant la généralisation du séquençage des génomes complets, l'étaient par Southern blot (on parlait alors de minisatellites) soit au hasard de la recherche de séquences polymorphes du génome humain, soit par exemple en utilisant comme sonde une séquence répétée en tandem (dans (Vergnaud 1989), les sondes utilisées sont totalement synthétiques). Le séquençage d'un nombre croissant de génomes bactériens va faciliter grandement l'identification de ce type de séquences, et va permettre leur étude exhaustive. Des outils de recherche de séquences répétées ont été développés au laboratoire et seront présentés dans la partie 2.1 du chapitre matériel et méthodes.

1.3.3.3 Utilisation en épidémiologie

Les répétitions en tandem sont souvent polymorphes et leurs allèles, de tailles différentes, sont alors facilement identifiables par PCR (à partir d'amorces spécifiques des flanquantes) suivie d'une simple migration sur gel. Ces structures, dont l'analyse est simple et rapide, ont donc une utilité en tant que marqueurs épidémiologiques. En effet, afin de comprendre le mode de dissémination des infections dans les communautés et les hôpitaux, mais également d'appréhender les changements évolutifs qui ont donné lieu à des avantages sélectifs, la distinction précise entre différents isolats d'une même espèce bactérienne est indispensable, ce qui peut être fait grâce aux répétitions en tandem polymorphes.

Quelques précautions doivent cependant être observées quant à l'utilisation des répétitions en tandem pour l'épidémiologie bactérienne (van Belkum 1999). D'une part, certaines répétitions ne sont pas neutres du point de vue évolutif (c'est-à-dire que certains allèles peuvent conférer un avantage sélectif à certains isolats). Cela pourrait donc conduire à des conclusions erronées sur la proximité des souches, qui peuvent avoir le même allèle car elles ont été confrontées au même hôte et non pas parce qu'elles sont génétiquement proches. Cela dit, de nombreuses répétitions en tandem non neutres sont utilisées de façon efficace en épidémiologie : par exemple, celle de la coagulase de *Staphylococcus aureus* (Shopsin 2000).

Par ailleurs, les locus de contingence, subissant la variation de phase, n'ont aucune valeur épidémiologique et doivent par conséquent être évités. Ces locus sont toutefois rares : dans l'ensemble, les répétitions en tandem ne subissent pas d'altérations au cours de leur manipulation en laboratoire (van Belkum 1997 ; Stothard 1998).

1.3.3.4 Exemples de répétitions en tandem étudiées dans les génomes bactériens

Les répétitions en tandem sont étudiées chez les bactéries depuis plus de dix ans. Elles ont été validées en tant que marqueurs épidémiologiques dans différentes espèces bactériennes. L'un de leurs avantages majeurs est que la mise en œuvre du typage de répétitions en tandem polymorphes est relativement facile et peu coûteuse par rapport aux autres techniques communément utilisées (voir paragraphe 1.2.3).

Les premières illustrations de la « méthodologie MLVA» chez les bactéries ont été réalisées en 1997 chez *Haemophilus influenzae* (van Belkum 1997) et *Bacillus anthracis* (Jackson 1997, Keim 2000). Ce dernier exemple est particulièrement intéressant du fait de la faible diversité au sein des souches de *B. anthracis*, une espèce d'émergence récente. Ces séquences répétées constituent une source majeure de polymorphisme dans ce génome. De même, l'étude des répétitions en tandem dans le génome de *Yersinia pestis*, une autre espèce

d'émergence récente (Achtman 1999), a été initiée à cette période (Adair 2000). Depuis, de nombreux autres marqueurs ont été développés chez *B. anthracis* et *Y. pestis*, dont un certain nombre au laboratoire, répondant ainsi à des préoccupations de la défense nationale face à une menace bioterroriste (Le Flèche 2001).

Une des bactéries pathogènes majeures en santé humaine est *Mycobacterium tuberculosis*. Le typage de répétitions en tandem est récemment apparu comme une alternative potentielle aux techniques courantes. La description des premiers VNTRs chez *M. tuberculosis* a été publiée en 1998 (Frothingham 1998). Depuis, l'étude des VNTRs chez cette espèce a fait l'objet de plusieurs publications (Supply 2000 ; Mazars 2001; Skuce 2002). Au laboratoire huit nouveaux minisatellites ont été identifiés et validés chez *M. tuberculosis* (Le Flèche 2002). Un outil d'identification en ligne a été mis en place et est accessible à la communauté. C'est pour cette espèce bactérienne que l'on trouve le plus de références concernant des études de génotypage par VNTRs et par spoligotypage. Plusieurs nomenclatures ont été employées pour désigner les répétitions en tandem chez *M. tuberculosis* dont les ETRs (Exact Tandem Repeats) et les MIRUs (Mycobacterial Interspersed Repeats Units). L'article de Le Flèche présente une synthèse des données existantes de typage de répétitions en tandem chez *M. tuberculosis*, ainsi que la présentation d'un service internet gratuit, rapide, et facile d'utilisation permettant l'identification de souches (Le Flèche 2002).

En ce qui concerne d'autres bactéries pathogènes pour l'homme, souvent tout reste à faire dans le domaine de la recherche de répétitions en tandem en tant qu'outil épidémiologique : très peu d'espèces ont été étudiées jusqu'à présent par rapport au nombre de génomes bactériens entièrement séquencés et publiés (voir Tableau 4 du paragraphe 1.4.2 sur le séquençage des génomes bactériens).

En 2001, une étude de type MLVA (Multi-Locus VNTR Analysis) a été réalisée chez *Francisella tularensis* (Farlow 2001). En 2002, la même équipe s'est intéressée à *Borrelia burgdorferi*, *Borrelia afzelii* et *Borrelia garinii* (Farlow 2002).

En 2003, les espèces suivantes ont été étudiées:

Salmonella enterica (Lindstedt 2003; Liu 2003); *Staphylococcus aureus* (Sabat 2003) et *Legionella pneumophila* étudiée au laboratoire (Pourcel 2003). Tout récemment, des études sont parues pour trois espèces du genre *Brucella* (Bricker 2003), les streptocoques de groupe B (Dore 2003) et *Neisseria spp* (Jordon 2003).

Dans la suite, je présenterai l'identification de marqueurs polymorphes pour le génotypage de *Yersinia pestis* (Le Flèche 2001) puis de *Pseudomonas aeruginosa* (Onteniente 2003) et enfin de *Staphylococcus aureus*, ainsi que les outils de bioinformatique développés au laboratoire (Denoeud 2004) pour faciliter ce type d'études.

1.3.3.5 Evolution de ces séquences répétées en tandem

Les séquences répétées en tandem sont largement distribuées à travers les génomes eucaryotes. Elles subissent parfois une instabilité génétique détectable dont on pense qu'elle est provoquée par des mécanismes de glissement, de recombinaison génétique inégale, de conversion génique (Charlesworth 1994). L'instabilité des répétitions en tandem a été observée dans des régions d'ADN non codantes (Jeffreys 1994), comme dans des gènes codant des protéines (Sharp 1987). Une conséquence de certains de ces mécanismes, en particulier le glissage à la réplication, est qu'il ne peut y avoir événement de mutation que dans une série de motifs parfaitement identiques. Cette homogénéité nécessaire est parfois appelée évolution concertée, ce qui peut être trompeur. Certaines répétitions en tandem, telle CEB1 chez l'homme, concilient parfaitement extrême instabilité méiotique et hétérogénéité interne des motifs (Buard 1994).

Les séquences répétées en tandem ont été observées dans tous les génomes procaryotes. Peu d'études sur les mécanismes d'évolution de ces séquences ont été entreprises. Les grandes répétitions sont les plus étudiées dans les génomes bactériens, en particulier chez *E. coli* (Rocha 1999). Néanmoins, le seul mécanisme connu chez les bactéries est le glissement lors de la réplication aussi appelé en anglais SSM (Slipped Strand Mispairing). Il concerne les répétitions en tandem à courts motifs (jusqu'à 10 pb), et qui sont parfaitement conservées (Levinson 1987).

Dans les génomes bactériens, il y a surtout des répétitions à motifs courts (de type microsatellites), aussi appelés SSRs (Short Sequence Repeats). Le mécanisme SSM intervient lors de la réplication et ne fait pas intervenir de cassure double brin comme c'est le cas pour des mécanismes plus complexes qui concernent surtout la catégorie des minisatellites (motifs de 10 pb et plus). Ces courtes séquences évoluent relativement vite, à une fréquence d'une insertion ou délétion pour 1000 replications (Levinson 1987). Ces mécanismes, décrits chez l'homme, sont encore peu étudiés dans les génomes bactériens, c'est pourquoi, à part le glissement lors de la réplication, on ne peut pas expliquer clairement les mécanismes impliqués dans l'expansion/réduction des répétitions. Le phénomène SSM conduit à l'augmentation ou à la réduction du nombre de copies selon le brin subissant le mésappariement, comme illustré par la Figure 3.

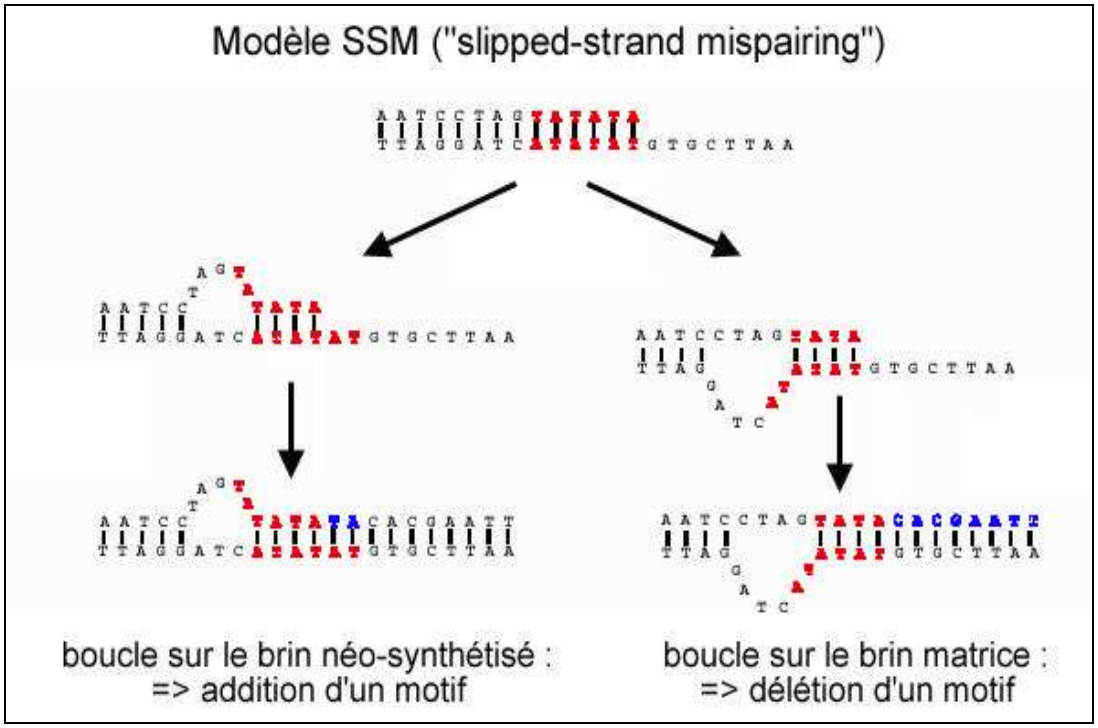


Figure 3 : Modèle de mutation des répétitions en tandem par glissement lors de la réplication.

Certaines de ces erreurs de réplication sont corrigées par l'activité de relecture de l'exonucléase et par le système de réparation des mésappariements, mais d'autres échappent à ces réparations et deviennent des mutations. Par exemple, chez la levure, les structures CTG/CAG ou CGG/CCG forment des structures secondaires qui échappent au système de réparation de l'ADN (Moore 1999). L'instabilité des répétitions en tandem à motifs courts correspondrait donc à un équilibre entre la génération d'erreurs de réplication par glissement et la correction de certaines de ces erreurs par les systèmes de réparation, comme le système de réparation des mésappariements et, à un degré moindre, l'activité de relecture de l'exonucléase (Kruglyak 1998). Des mutations dans des gènes du système de réparation des mésappariements (MMR) sont à l'origine d'une instabilité des microsatellites chez *Escherichia coli* (Levinson 1987).

Shields et McDevitt se sont intéressés plus particulièrement à la séquence répétée présente dans le gène clfA (clumping factor) chez *Staphylococcus aureus* (Shields 1995). Ils ont étudié la séquence de la souche Newman (GenBank Z18852). Le motif protéique répété est Serine-Aspartate, et il y a 308 résidus. Au niveau de l'ADN, la répétition a une taille de 18 pb, avec pour motif consensus GAY TCN GAY TCN GAY AGY (N : les 4 bases possibles, Y : T ou C). Les motifs les plus conservés par rapport au consensus sont observés au centre de la séquence. Aux extrémités, les motifs sont très différents par rapport au consensus. La comparaison de l'usage des codons dans la séquence répétée et dans le reste de la séquence du gène clfA, ainsi que dans d'autres gènes similaires, a montré un usage des codons un peu différent dans la séquence répétée par rapport aux autres séquences comparées. Ce

phénomène pourrait correspondre à un avantage sélectif plus qu'à de l'évolution concertée. Par ailleurs, le peu d'homogénéité de l'ensemble des motifs de la répétition ne va pas en faveur de l'évolution concertée. Cela peut refléter une stratégie utilisée par de nombreuses bactéries qui possèdent des répétitions instables qui leur confèrent un avantage sélectif en leur permettant par exemple des phénomènes de variation de phase, comme cela a été bien décrit chez *Neisseria meningitidis* (voir paragraphe suivant).

1.3.3.6 Rôles de ces séquences dans l'adaptation et la virulence des bactéries

Dans un certain nombre de pathogènes, des répétitions en tandem à courts motifs (ou SSRs), présentes en amont ou dans les séquences codantes de protéines de surface, sont polymorphes et permettent l'adaptation de la bactérie aux changements de conditions survenant au cours de l'infection de l'hôte. Ce phénomène est appelé variation de phase (revue : (Henderson 1999)). Les mécanismes de variation de phase sont multiples. Ils peuvent se produire pour des séquences répétées situées en amont du gène donc dans la partie régulatrice de la transcription, ou bien dans la séquence codante elle-même et, dans ce cas, c'est le produit du gène qui peut être affecté lors de la traduction par un changement de cadre de lecture.

La variation de phase peut conduire à une variation antigénique lorsque différents phénotypes de surface peuvent être exprimés. La variation de phase a été observée en particulier chez les bactéries à gram négatif, et concerne les structures de surface telles que les fimbriaes, les flagelles, les protéines de la membrane externe ainsi que les Lipopolysaccharides (LPS), ce qui permet de penser que ces mécanismes permettent à la bactérie de s'adapter à différents types de milieux, et d'échapper au système immunitaire en modifiant les protéines de surface, sans perte irréversible de patrimoine génétique.

Les locus à fort taux de mutation sont appelés locus de contingence (Moxon 1994). Par exemple, chez la bactérie *Haemophilus influenzae*, qui colonise les voies respiratoires et peut causer des pneumonies et des méningites, le rôle des répétitions en tandem dans la modulation de la virulence a été beaucoup étudié. Tout d'abord, la variabilité des répétitions a été associée expérimentalement à la modulation des gènes impliqués dans la synthèse des pili et des LPS (Weiser 1989). Une répétition de dinucléotides dans un promoteur de gènes codant des sous-unités de pili est un facteur régulateur majeur de leur expression : selon le nombre d'unités répétées, l'espacement entre les boîtes -35 et -10 est soit favorable soit défavorable à la reconnaissance de ce site par l'ARN polymérase (van Ham 1993). Les SSRs peuvent aussi moduler l'expression génique en étant à l'origine de blocages de la réplication (Krasilnikova 1998).

D'autres répétitions en tandem ont un effet au niveau de la traduction. Chez *H. influenzae*, différents gènes codant pour des enzymes de synthèse du LPS contiennent des répétitions de tétranucléotides, localisées dans les séquences codantes, ce qui est à l'origine de décalages du cadre de lecture (Weiser 1990). Un tétranucléotide, situé dans un gène homologue à une méthyltransférase de type III, est tellement instable qu'il génère même du mosaïcisme dans les cultures bactériennes (De Bolle 2000). Chez *Neisseria meningitidis*, une répétition en tandem de 7 pb dans la phase codante du gène PilQ affecte la biosynthèse des pili de façon quantitative (Tonjum 1998). Les SSRs peuvent également agir directement comme des terminateurs de transcription (Guerin 1998).

L'implication des SSRs dans la modulation de l'expression de gènes a été mise en évidence dans une grande variété d'autres bactéries [revue : (van Belkum 1999)] dont *Escherichia coli* (Foster 1994), *Neisseria meningitidis* (van der Ende 1995), ou *Mycoplasma gallisepticum* (Glew 1998), et devrait profiter du séquençage des génomes bactériens. En effet, ce séquençage a ouvert la voie à une analyse plus systématique des gènes potentiellement impliqués dans la variation de phase. Lorsque la séquence complète du génome de *H. influenzae* a été connue, un catalogue de toutes les répétitions en tandem a pu être établi (Fleischmann 1995) : la plupart de ces répétitions sont associées avec des gènes potentiellement impliqués dans la virulence (molécules d'adhésion, enzymes de synthèse du LPS…) (Hood 1996). De la même façon, le séquençage du pathogène *Helicobacter pylori* (Tomb 1997) a permis de mettre en évidence une trentaine de gènes associés à des SSRs, c'est à dire potentiellement impliqués dans la variation de phase (Saunders 1998). Ces gènes codent pour des enzymes de biosynthèse du LPS, des protéines de surface et des enzymes de restriction. Le séquençage d'une seconde souche de cette bactérie (Alm 1999) a rendu possible l'identification de SSRs polymorphes (entre les deux souches considérées) : la plupart de ces candidats se sont avérés subir une variation d'expression selon le nombre de répétitions.

Par ailleurs, des répétitions en tandem appartenant à des phases ouvertes de lecture et dont le motif est multiple de 3 génèrent des polymorphismes au niveau des protéines, ce qui peut être à l'origine d'une variation antigénique :

- Chez *Staphylococcus aureus*, des protéines de surface impliquées dans la reconnaissance des molécules d'adhésion de la matrice extracellulaire de l'hôte contiennent de nombreuses répétitions en tandem, dont le nombre de répétitions influe sur l'accessibilité du domaine actif (voir Tableau 3).

- Chez *Bacillus anthracis*, une répétition en tandem dans le gène de l'exosporium est à l'origine de variations de la longueur des filaments de la surface des spores (Sylvestre 2003).

- Chez les streptocoques du groupe A, la protéine M, protéine de surface et facteur de virulence, est soumise à une grande variabilité antigénique causée vraisemblablement par des événements de recombinaison homologue (Hollingshead 1987).

- Chez les streptocoques du groupe B, la protéine alpha C, antigène de surface, contient une répétition en tandem polymorphe qui, lorsqu'elle est délétée, permet d'échapper à la réponse immunitaire de l'hôte (Madoff 1996 ; Gravekamp 1998).

- Chez le mycoplasme *Mycoplasma hyorhinis*, les protéines de surface du système VIp confèrent aux bactéries, par leur variation de taille, une résistance contre les anticorps produits par l'hôte (porc) (Citti 1997).

Le Tableau 3 liste des répétitions en tandem associées à des gènes de fonction connue chez différentes bactéries (van Belkum 1998) : selon la taille de leurs motifs répétés, certaines appartiennent à la classe des microsatellites et d'autres à la classe des minisatellites.

Tableau 3 : Description de gènes bactériens associés à des répétitions en tandem, d'après (van Belkum 1998)

Espèce	Motif répété	Gène	Niveau de régulation du gène		Fonction du gène	Références
			transcription	Traduction / protéine		
H. influenzae	CAAT	*lic1-lic3*	-	+	biosynthèse du lipopolysaccharide	(Hood 1996) ; (Weiser 1990)
	GCAA	*yadA*	-	+	adhésine	(Hood 1996)
	GACA	*lgtC*	-	+	glycosyltransférase	(Hood 1996)
	TTGG	ND	-	+	protéines liant le fer	(Hood 1996)
	AGTC	ND	-	+	méthyltransférase	(Hood 1996)
	TTTA	ND	-	+	homologue d'une protéine de Bacillus	(Hood 1996)
	TA	*HifA/B*	+	-	synthèse des fimbriae (pili)	(van Ham 1993)
N. meningitidis	G	*lsi2*	+	-	biosynthèse du lipopolysaccharide	(Burch 1997)
	CTCTT	*opa*	-	+	protéines d'opacité	(Meyer 1990)
	A	*opa*	+	-	protéines d'opacité	(Meyer 1990)
	G	*porA*	+	-	protéine de la membrane externe	(van der Ende 1995)
S. aureus	93 pb	*fnb*	-	+	protéine se liant à la fibronectine	(Patti 1994)
	561 pb	*cna*	-	+	adhésine du collagène	(Patti 1992)
	81 pb	*coa*	-	+	coagulase	(Goh 1992); (Schwarzkopf 1994)
	24 pb	*spa*	-	+	protéine A	(Shopsin 1999)
	18 pb	*clf*	-	+	récepteur du fibrinogène	(McDevitt 1995)
Streptococcus spp.	60 pb	*pspA*	-	+	protéine de surface des pneumocoques	(Yother 1992)
	69 pb	*emm*	-	+	protéine de résistance à la phagocytose	(Bessen 1989)
	246 pb	*aC*	-	+	protéine *aC*	(Gravekamp 1997)
E. faecalis	TAGTARR	*rep1*et *rep2*	+	-	itéron: régule la réplication et le transfert des plasmides	(Heath 1995)
M. hyorhinus	36 et 39 pb	*vlp*	-	+	protéine membranaire variante	(Yogev 1991)
	A	*vlp*	+	-	protéine membranaire variante	(Yogev 1991)
M. bovis	24 pb	*vspA*	-	+	lipoprotéine de la surface membranaire	(Lysnyansky 1996)
M. fermentans	A	*P78*	-	+	lipoprotéine de transporteur ABC	(Theiss 1997)
U. urealyticum	18 pb	*MB*	-	+	antigène spécifique de MB	(Zheng 1995)
B. anthracis	12 pb	*vvrA*	-	+	homologue de la protéine de la gaine microfilaire	(Jackson 1997)
L. monocytogenes	66 pb	*prfA*	-	+	homologue à l'internaline riche en leucine	(Domann 1997)
E. coli	A et C	*lac*	+	-	ß-galactosidase	(Foster 1994); (Rosenberg 1994)
A. marginale	87 pb	*Msp1a*	-	+	protéine majeure de la surface	(Allred 1990)

1.4 Le séquençage des génomes bactériens

Depuis la publication de la première séquence complète d'un génome bactérien, celui d'*Haemophilus influenzae*, en 1995 (Fleischmann 1995), qui a démontré l'efficacité de l'approche « WGS » (whole genome shotgun) pour le séquençage de génomes complets (voir Figure 5), des progrès spectaculaires ont été faits au niveau des techniques de séquençage, des

stratégies d'assemblage et de finition, et des méthodes d'annotation. La microbiologie est certainement parmi les premiers bénéficiaires de cette évolution : à l'heure actuelle, plus de 100 génomes bactériens ont été entièrement séquencés et trois fois plus sont en cours de séquençage (au 15 novembre 2003, d'après le site GOLD : « Genome Online Database » [http://ergo.integratedgenomics.com/GOLD/] (Bernal 2001), 128 génomes bactériens étaient achevés et 391 en cours). La Figure 4 montre l'évolution du nombre de génomes procaryotes séquencés chaque année depuis 1995, ce qui témoigne des progrès effectués ces dernières années pour le séquençage systématique des génomes.

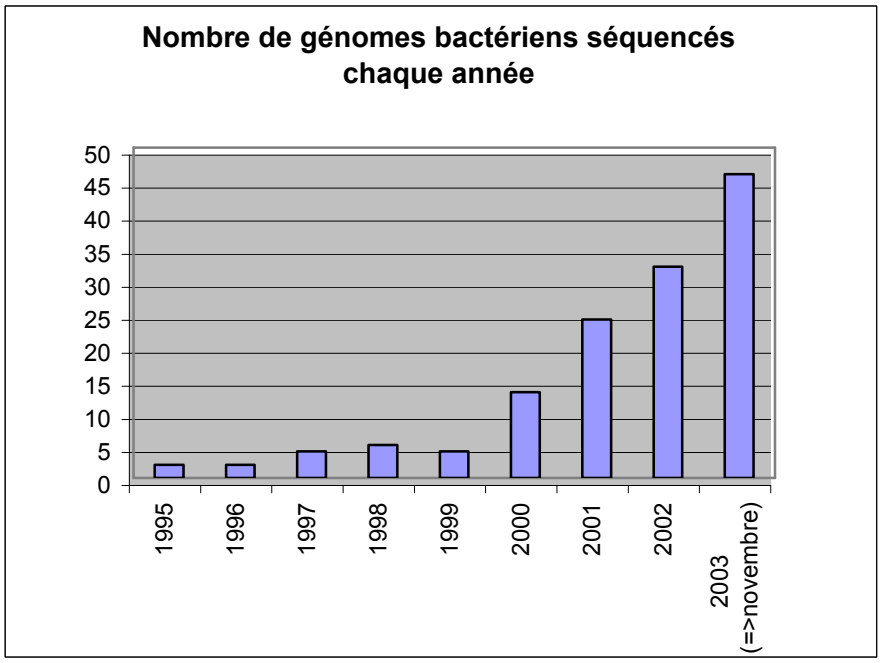


Figure 4 : Evolution du nombre de génomes bactériens séquencés chaque année depuis 1995.

En 2002 le séquençage complet d'un génome bactérien pouvait être achevé en quelques mois, avec un taux d'erreur de l'ordre de 1/100000 seulement, et au coût de 8 à 9 centimes de dollar par paire de bases (Fraser 2002), ce qui correspond, pour un génome de quelques Mégabases à un coût de quelques centaines de milliers d'euros.

1.4.1 Différentes stratégies de séquençage des génomes

Deux stratégies de séquençage des génomes sont fréquemment employées (Frangeul 1999) : la première, celle des clones ordonnés, utilise une banque de grands inserts pour construire une carte de chevauchement couvrant le génome entier ; les clones choisis sont alors séquencés un par un pour obtenir la séquence du génome entier. La deuxième stratégie, dite du « Shotgun complet », n'exige pas de carte avant le séquençage.

1.4.1.1 Stratégie des clones ordonnés

Dans cette stratégie, différentes méthodes sont employées pour construire une carte : « restriction fingerprinting » ou bien «hybridization mapping ». Le fingerprint est une méthode de comparaison de clones, basée sur la correspondance de profils de fragments de restriction caractéristiques entre différents clones. Si deux clones partagent un nombre significatif de fragments de restriction, il peut être supposé que ces deux clones se chevauchent et forment une région appelée « contig ». Cette méthode a été appliquée avec succès pour le séquençage de *Caenorhabditis elegans* et de *Mycobacterium tuberculosis*, et c'est la méthode qui a permis le séquençage du génome humain par le consortium public Human Genome Project (Lander 1987).

1.4.1.2 Whole Genome Shotgun

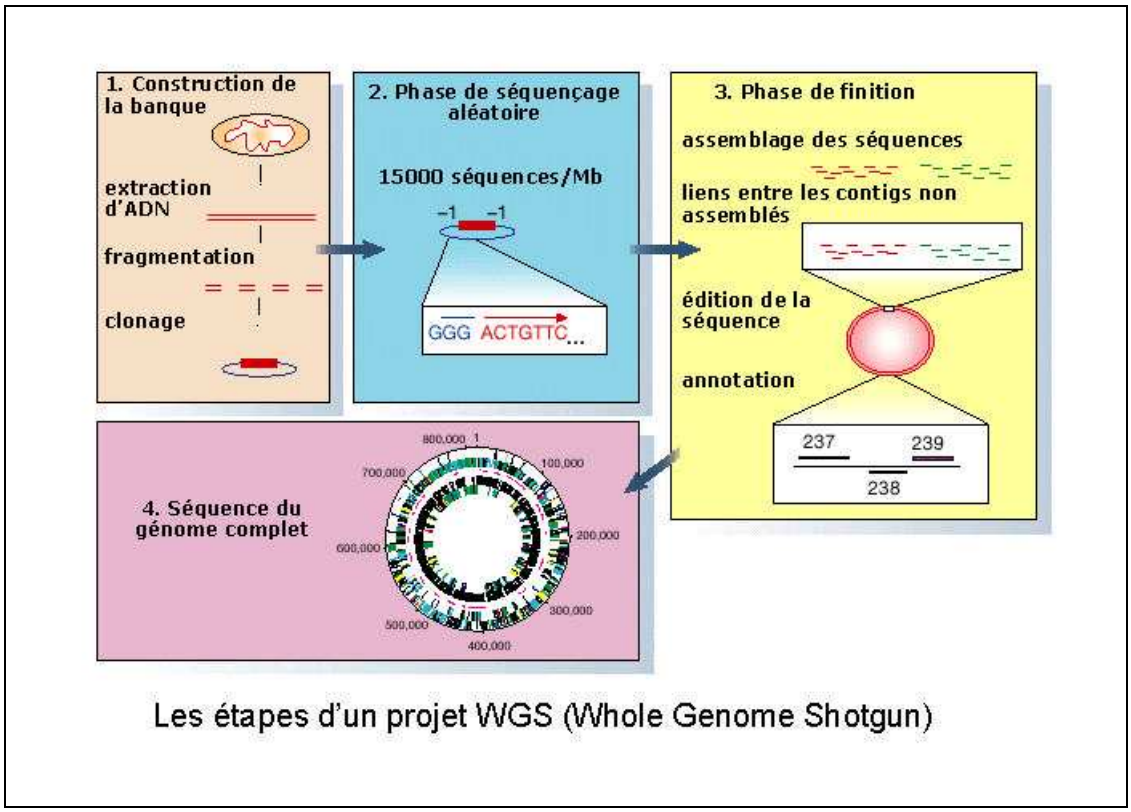


Figure 5 : Projet WGS (d'après (Fraser 2000)).

C'est actuellement la stratégie la plus largement employée pour le séquençage d'un génome microbien. Le génome est d'abord fragmenté de manière physique (nébulisation, sonication) et/ou chimique (enzyme de restriction) puis les fragments générés sont utilisés pour créer différentes banques. Un grand nombre de clones, issus de ces banques, sont séquencés

aléatoirement et assemblés. L'utilisation de fragments issus de plusieurs méthodes de fragmentation permettra d'obtenir l'ordre des contigs par rapport au génome complet, puis les séquences manquantes entre chaque contig seront déterminées par un séquençage ciblé, afin d'obtenir la séquence complète du génome (voir Figure 5). Environ 90 à 95% de la séquence d'un génome sera déterminée pendant la phase aléatoire et représentée sous la forme d'un ensemble de plusieurs centaines de contigs. La détermination de l'ordre de ces contigs et l'obtention des séquences manquantes représente la partie la plus complexe d'un projet génome complet, c'est la phase de finition (ou finishing). Cette phase utilise différentes stratégies *in silico* et expérimentales pour ordonner les contigs par rapport à la séquence finale, et sa durée dépend essentiellement de la qualité des banques utilisées lors de la phase aléatoire, du nombre de clones séquencés et de la richesse du génome en éléments répétés. La couverture minimun pour ne pas avoir de trou dans l'assemblage final est d'au moins 10X.

1.4.2 Génomes bactériens entièrement séquencés à ce jour

La Figure 6 décrit les propriétés des différentes bactéries séquencées. On peut les classer dans trois grandes catégories, correspondant aux motivations de leur séquençage :

- Les bactéries d'intérêt médical : pathogènes humains (par exemple *Yersinia pestis :* peste, deux souches ; *Mycobacterium tuberculosis* : tuberculose, 3 souches ; *Staphylococcus aureus* : infections nosocomiales et communautaires, 6 souches) et bactéries présentant un intérêt pharmaceutique (par exemple *Streptomyces coelicolor :* production d'antibiotiques).

- Les bactéries d'intérêt économique : pathogènes de végétaux (par exemple *Xylella fastidiosa* : phytopathogène), bactéries présentant un intérêt pour l'industrie agroalimentaire (*Lactococcus lactis*), bactéries utilisées pour la synthèse d'acides aminés (*Corynebacterium efficiens*) ou de solvants (*Clostridium acetobutylicum*), ou encore pour la dépollution (*Pseudomonas putida*, *Shewanella oneidensis*).

- Les bactéries d'intérêt pour la recherche en microbiologie : organismes modèles (*Escherichia coli*, *Bacillus subtilis*) ou ayant des propriétés biologiques intéressantes (résistance à des conditions de vie en milieux « extrêmes » : par exemple, *Deinococcus radiodurans* qui survit dans des milieux fortement irradiés, ou *Thermotoga maritima*, organisme thermophile…)

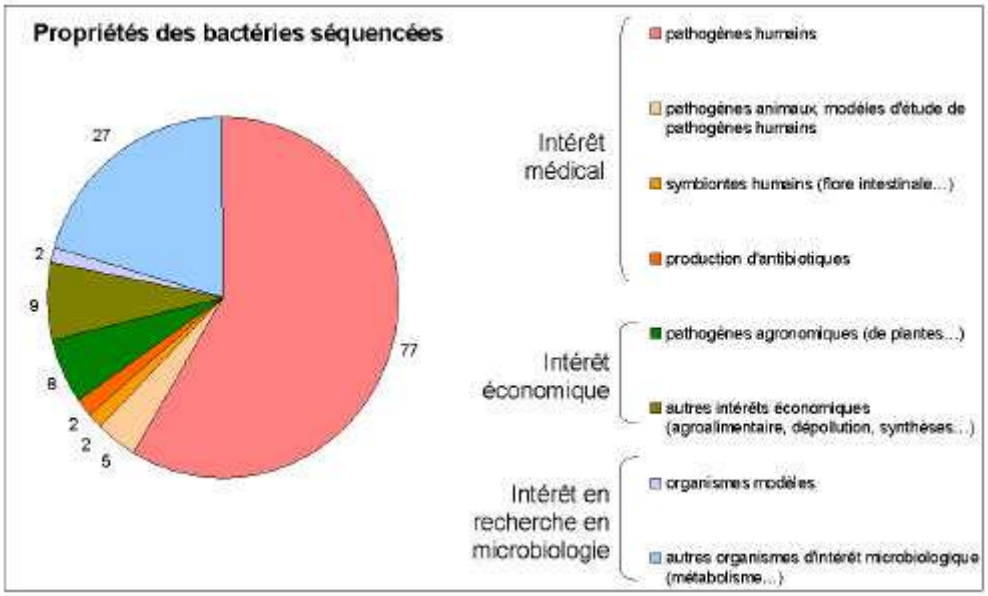


Figure 6 : Motivations du séquençage de génomes bactériens.

La grande majorité des bactéries séquencées à ce jour est d'intérêt médical : entre 1995 et 2000, les organismes séquencés étaient essentiellement des bactéries pathogènes pour l'homme auxquelles s'ajoutaient quelques organismes d'intérêt plus fondamental (organismes modèles). Depuis 2001, de plus en plus de projets de séquençage impliquent des bactéries d'intérêt économique (Nelson 2000). Ce biais en faveur des bactéries d'intérêt médical devrait donc s'atténuer dans les prochaines années.

Le règne bactérien est d'une grande hétérogénéité, du point de vue du pourcentage en GC (de 22% pour *Wigglesworthia glossinidia* à 72% pour *Streptomyces coelicolor*) comme de la taille du génome (de 580 kb pour *Mycoplasma genitalium*, pathogène intracellulaire obligatoire, à 9105 kb pour *Bradyrhizobium japonicum*) ou du nombre de gènes. La Figure 7 représente le nombre de gènes en fonction de la taille du génome pour les 128 bactéries séquencées au 15 novembre 2003 : la densité en gènes des bactéries est relativement constante et voisine d'un gène par kilobase (elle varie entre 0,49 pour *Mycobacterium leprae*, particulièrement peu dense : cette espèce présente une fraction importante d'ADN non codant et de pseudogènes -non transcrits ou non traduits- (Cole 2001), et 1,29 pour *Escherichia coli* O157:H7 EDL933). La fraction codante est généralement aux alentours de 90%. Près de la moitié des ORFs (pour « Open Reading Frames », ou phases ouvertes de lecture) de chaque espèce est de fonction inconnue. De plus, environ un quart des ORFs n'a aucune homologie avec des protéines existantes dans les bases de données de séquences. Ce pourcentage devrait diminuer avec le séquençage de plus en plus de génomes bactériens, mais il témoigne de la grande diversité biologique au sein des organismes procaryotes (Fraser 2000).

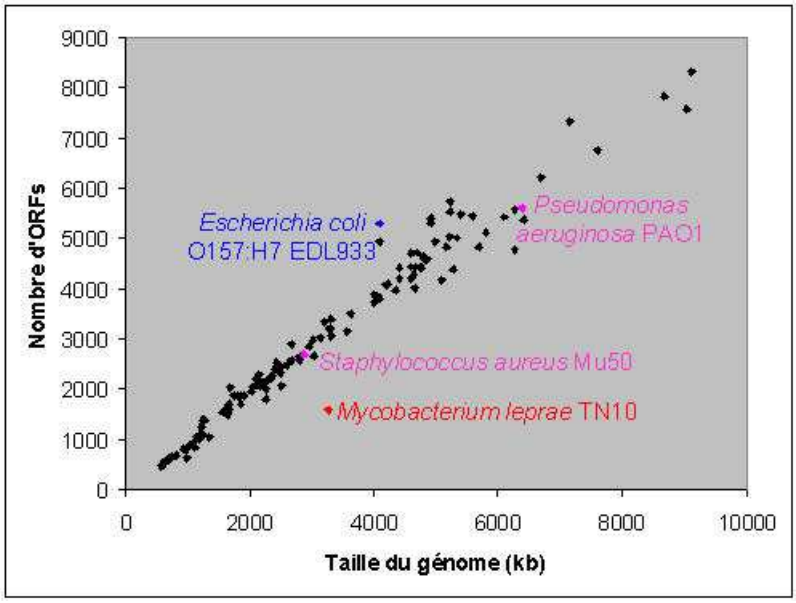


Figure 7 : Densité d'ORFs dans les génomes bactériens.

Le Tableau 4 est un récapitulatif des génomes des 66 bactéries pathogènes pour l'homme entièrement séquencées et publiées à la date du 15 novembre 2003. Parmi ces génomes, très peu ont fait l'objet d'une étude de VNTRs. Parmi ces 66 bactéries séquencées, certaines sont de même espèce, il y a au total 48 espèces bactériennes représentées dans ce tableau dont 11 ont fait l'objet d'une étude MLVA.

Tableau 4 : Génomes des bactéries pathogènes pour l'homme entièrement séquencés et publiés (à la date du 15 novembre 2003)

Espèce	Souche	Propriétés de la bactérie	Numéro(s) d'accession*	Publication*	Date de publication*	Taille du génome (kb)*	Nombre d'ORFs (phases ouvertes de lecture)*	Densité en ORFs par kb	Souches comparées pour la recherche de répétitions en tandem polymorphes[a]
Bacillus anthracis[b]	Ames	responsable de la maladie du charbon	NC_003997	Nature 423, 81-86	01/05/03	5227	5738	1,10	
***Bacillus* cereus**[b]	ATCC 14579	intoxications alimentaires	NC_004722	Nature 423, 87-91	01/05/03	5411	5477	1,01	
Bordetella parapertussis	12822 NCTC-13253	coqueluche	NC_002928	Nature Genetics 35, 32-40	10/08/03	4773	4404	0,92	
Bordetella pertussis	Tohama I NCTC-13251	coqueluche	NC_002929	Nature Genetics 35, 32-40	10/08/03	4086	3816	0,93	
Borrelia burgdorferi	B31	maladie de Lyme	NC_001318	Nature 390,580-586	11/12/97	1230	1256	1,02	
Brucella melitensis biovar suis	1330	brucellose (zoonose, infection chronique, avortement)	NC_004310, NC_004311	PNAS 99, 13148-13153	01/10/02	3310	3388	1,02	1330/16M
Brucella melitensis	16M	fièvre de Malte	AE008917, AE008918	PNAS 99, 443-448	08/01/02	3294	3197	0,97	
Campylobacter jejuni	NCTC 11168	syndrome de Guillain-Barré	AL111168	Nature 403,665-668	10/02/00	1641	1654	1,01	
Chlamydia trachomatis	serovar D	infections génitales, pulmonaires, oculaires	AE001273	Science 282,754-759	23/10/98	1042	896	0,86	
Chlamydia pneumoniae	AR39	pneumonies et bronchites	AE002161	NAR 28, 1397-1406	15/03/00	1229	1052	0,86	CWL029/ J138; CWL029/ AR39 + comparaison 3 génomes
Chlamydophila pneumoniae	J138	pneumonies et bronchites	BA000008	NAR 28, 2311-2314	15/06/00	1228	1070	0,87	
Chlamydophila pneumoniae	CWL029	pneumonies et bronchites	AE001363	Nat Genet, 21,385-389	10/04/99	1230	1052	0,86	
Chlamydophila caviae	GPIC	pathogène	NC 003361	NAR, 31, 2134-2147	15/04/03	1173	1012	0,86	
Chromobacterium violaceum	ATCC 12472	parfois pathogène pour l'homme	NC_005085	PNAS 100, 11660-11665	18/09/03	4751	4431	0,93	
Clostridium tetani	Massachusetts E88	tétanos	NC_004557	PNAS 100, 1316-21	05/02/03	2799	2640	0,94	
Clostridium perfringens	13	gangrène	BA000016, NC 003042	PNAS 99, 996-1001	22/01/02	3031	2660	0,88	
Corynebacterium diphtheriae gravis	NCTC13129	diphtérie	NC 002935	NAR 31, 6516-6523	15/11/03	2488	2320	0,93	
Coxiella burnetii	RSA 493	fièvre Q	NC 002971	PNAS 100, 5455-60	29/04/03	2100	2095	1,00	
Enterococcus faecalis	V583	infections des voies urinaires, bactériémie, endocardite	NC 004668	Science 299, 2071-4	28/03/03	3209	3337	1,04	

Espèce	Souche	Propriétés de la bactérie	Numéro(s) d'accession*	Publication*	Date de publication*	Taille du génome (kb)*	Nombre d'ORFs (phases ouvertes de lecture)*	Densité en ORFs par kb	Souches comparées pour la recherche de répétitions en tandem polymorphes[a]
Escherichia coli	O157:H7 Sakai	diarrhées, colites hémorragiques, syndrome d'urémie hémolytique	BA000007	DNA Research 8, 11-22	27/02/01	5594	5448	0,97	0157:H7 Sakai/ UPEC-CFT073; 0157:H7 Sakai/ EDL933
Escherichia coli	UPEC-CFT073	souche uropathogénique	NC 004431	PNAS 99, 17020-4	09/12/02	5231	5533	1,06	
Escherichia coli	O157:H7 EDL933	colites hémorragiques, syndrome d'urémie hémolytique	AE005174	Nature, 409, 529-533	25/01/01	4100	5283	1,29	
Fusobacterium nucleatum	ATCC 25586	pathogène dentaire	AE009951	J Bacteriol 184, 2005-2018	10/04/02	2170	2067	0,95	
Haemophilus influenzae	KW20	bronchites, otites	L42023	Science 269,496-512	28/07/95	1830	1850	1,01	
Helicobacter pylori	26695	ulcères	AE000511	Nature 388,539-547	07/08/97	1667	1590	0,95	26695/ J99
Helicobacter pylori	J99	ulcères	AE001439	Nature 397,176-180	14/01/99	1643	1495	0,91	
Leptospira interrogans serovar lai	56601	leptospirose	NC 004342, NC 004343	Nature 422, 888-93	24/04/03	4691	4728	1,01	
Listeria monocytogenes	EGD-e	pathogène alimentaire	AL591824	Science 294, 849-852	26/10/01	2944	2855	0,97	
Mycobacterium bovis[b]	AF2122/97(spoligotype9)	tuberculose	NC 002945	PNAS 100, 7877-7882	24/06/03	4345	3955	0,91	
Mycobacterium tuberculosis[b]	CDC1551	tuberculose	AE000516	J Bacteriol 184, 5479-90	02/10/01	4403	4187	0,95	H37Rv/ CDC1551
Mycobacterium tuberculosis[b]	H37Rv	tuberculose	AL123456	Nature 393,537-544	11/06/98	4411	4402	1	
Mycobacterium leprae	TN	lèpre	AL450380	Nature 409, 1007-1011	11/06/98	3268	1604	0,49	
Mycoplasma genitalium	G-37	pathogène du tractus génital	L43967	Science 270,397-403	20/10/95	580	468	0,81	
Mycoplasma penetrans	HF-2	infections urogénitales et respiratoires	NC 004432	NAR, 30, 5293-5300	01/12/02	1358	1038	0,76	
Mycoplasma pneumoniae	M129	pneumonie	U00089	NAR 24,4420-4449	15/11/96	816	677	0,83	
Neisseria meningitidis[b]	MC58 (serogroup B)	méningite	AE002098	Science 287,1809-1815	10/03/00	2272	2158	0,95	MC58/ Z2491
Neisseria meningitidis[b]	Z2491 (serogroup A)	méningite	AL 162759	Nature, 404, 502-506	30/03/00	2184	2121	0,971	
Porphyromonas gingivalis	W83	pathogène oral: infecte les gencives	NC_002950	J.Bacteriol 185, 5591-5601	09/09/03	2343	2227	0,95	

Espèce	Souche	Propriétés de la bactérie	Numéro(s) d'accession*	Publication*	Date de publication*	Taille du génome (kb)*	Nombre d'ORFs (phases ouvertes de lecture)*	Densité en ORFs par kb	Souches comparées pour la recherche de répétitions en tandem polymorphes[a]
Pseudomonas aeruginosa[b]	PAO1	infecte les poumons des individus atteints de mucoviscidose ou immunodéprimés + autres infections	AE004091	Nature 406,959-964	30/09/00	6264	5570	0,89	
Rickettsia conorii	Malish 7	fièvre éruptive méditerranéenne	AE006914	Science 293, 2093-2098	14/09/01	1268	1374	1,08	Malish 7/ Madrid E
Rickettsia prowazekii	Madrid E	typhus	AJ235269	Nature 396,133-140	12/11/98	1111	834	0,75	
Salmonella typhimurium	LT2SGSC1412	gastro-entérite, fièvre thyphoïde	NC_003197	Nature 413, 852-856	25/10/01	4857	4597	0,946	S. typhimurium/ S. enterica Typhi Ty2; S. enterica Typhi CT18/ S. enterica Typhi Ty2; S. typhimurium/ S. enterica Typhi CT18 + comparaison des 3 génomes Salmonella
Salmonella enterica	Typhi Ty2	fièvre thyphoïde	NC_004631	J. Bacteriol 185, 2330-7	21/03/03	4791	4646	0,970	
Salmonella enterica	serovar Typhi CT18	fièvre thyphoïde	AL513382	Nature, 413, 848-852	25/10/01	4809	4600	0,957	
Shigella flexneri	2a 301	dysenterie et réaction inflammatoire	NC_004337	NAR 30, 4432-4441	16/10/02	4607	4434	0,96	Salmonella typhimurium/ Shigella flexneri
Shigella flexneri	serotype 2a 2457T	dysenterie	NC_004741	Infect. Immun. 71, 2775-2786	23/04/03	4599	4706	1,02	
Staphylococcus aureus[b]	MW2	infections suppuratives cutanées, sous cutanées et des muqueuses	BA000033	Lancet 359, 1819-1827	25/05/02	2820	2632	0,93	Mu50/ MW2; Mu50/ N315
Staphylococcus aureus[b]	Mu50 (VRSA)		BA000017	The Lancet 357, 1225-1240	21/04/01	2878	2697	0,94	
Staphylococcus aureus[b]	N315 (MRSA)		BA000018	The Lancet 357, 1225-1240	21/04/01	2813	2594	0,92	
Streptococcus pyogenes	M1 GAS SF370	pharyngite, scarlatine, septicémies, syndrome de choc toxique	AE004092	PNAS, 98, 4658-4663	10/04/01	1852	1696	0,92	M1 GAS/ M3 SSI1; M1 GAS/ M3 MGAS315; M1 GAS/ M18 MGAS8232 + comparaison des 4 génomes
Streptococcus pyogenes	M3 (SSI-1)	infections suppuratives	NC 004606	Genome Res 13, 1042-55	10/06/03	1894	1861	0,98	
Streptococcus pyogenes	M3 MGAS315	infections invasives	NC_004070	PNAS 99, 10078-10083	16/07/02	1900	1865	0,98	
Streptococcus pyogenes	M18 MGAS8232	infections néonatales (maladies cardiaques)	AE009949	PNAS 99, 4648-4673	02/04/02	1895	1889	1,00	
Streptococcus mutans	UA159	caries dentaires	NC_004350	PNAS 99, 14434-14439	29/10/02	2030	1963	0,97	
Streptococcus agalactiae	NEM316	infections néonatales (septicémie, méningite, pneumonie)	AL732656	Mol Microbiol 45, 1499-513	30/09/02	2211	2118	0,96	NEM316/ 2603

Espèce	Souche	Propriétés de la bactérie	Numéro(s) d'accession*	Publication*	Date de publication*	Taille du génome (kb)*	Nombre d'ORFs (phases ouvertes de lecture)*	Densité en ORFs par kb	Souches comparées pour la recherche de répétitions en tandem polymorphes[a]
Streptococcus agalactiae	2603V/R	infections néonatales (septicémie, méningite, pneumonie)	NC_004116	PNAS 99, 12391-12396	28/08/02	2160	2175	1,01	
Streptococcus pneumoniae	TIGR4 ATCC-BAA-334	pneumonie, septicémie, méningite	AE005672	Science 293,498-506	20/07/01	2160	2094	0,97	TIGR4/ R6
Streptococcus pneumoniae	R6	pneumonie, septicémie, méningite	AE007317	J Bacteriol. 183, 5709-5717	10/10/01	2038	2043	1,00	
Treponema pallidum	Nichols	syphilis	AE000520	Science 281,375-388	17/07/98	1138	1041	0,91	
Tropheryma whipplei	TW08/27	infection chronique intestinale + autres organes	NC_004551	Lancet 361, 637-644	24/02/03	925	817	0,88	TW08/27 / Twist
Tropheryma whipplei	Twist	maladie intestinale	NC_004572	Genome Res 13, 1800-1809	01/08/03	927	808	0,87	
Ureaplasma urealyticum	serovar 3	pathogène urogénital	AF222894	Nature 407, 757-762	12/10/00	751	650	0,87	
Vibrio cholerae	Biotype El Tor, strain N16961	choléra	AE003852 AE003853	Nature 406,477-483	03/08/00	4000	3885	0,97	
Vibrio parahaemolyticus	RIMD 2210633	gastroentérite	NC 004603, NC 004605	Lancet 361, 743-9	01/03/03	5165	4832	0,94	
Yersinia pestis[b]	CO-92 (Biovar Orientalis)	peste	NC_003143, NC_003131, NC_003132, NC_003134	Nature 413,523-527	04/10/01	4653	4012	0,86	CO-92/ KIM5 P12
Yersinia pestis[b]	KIM5 P12 (Biovar Mediaevalis)	peste	NC_004088	J. Bacteriol 184, 4601-4611	29/07/02	4600	4198	0,91	

- * :D'après le site « GOLD » [http://ergo.integratedgenomics.com/GOLD/]
- **espèces bactériennes ayant fait l'objet d'une étude de VNTRs**
- [a] : résultats des comparaisons disponibles dans la base de données du laboratoire GPMS (Génomes, Polymorphisme et Minisatellites)
- [b] : espèces bactériennes étudiées au laboratoire GPMS

Des comparaisons de génomes de même espèce bactérienne ou d'espèces voisines ont été réalisées au laboratoire afin d'identifier *in silico* des répétitions en tandem polymorphes. Les résultats de ces comparaisons sont accessibles sur la page web du laboratoire http://minisatellites.u-psud.fr/comparison**.**

La Figure 8 représente l'arbre phylogénetique des procaryotes séquencés, basé sur les séquences 16S.

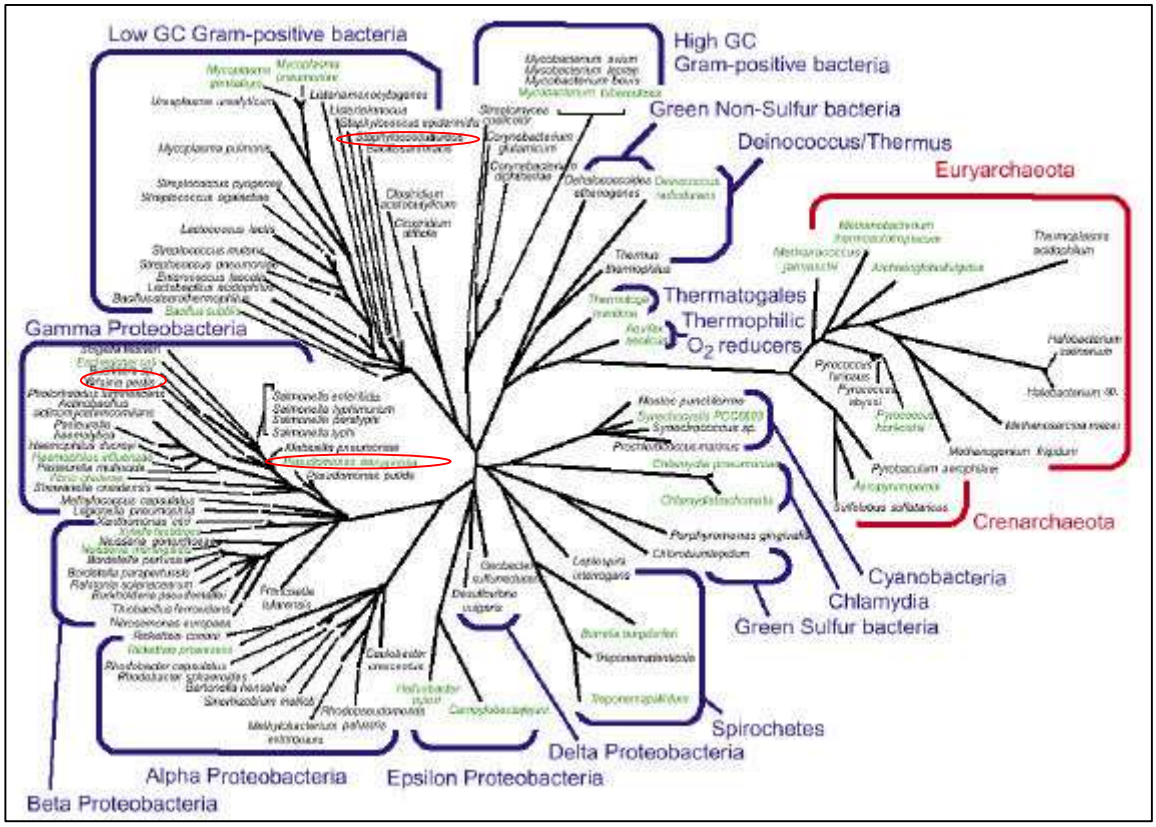


Figure 8 : Arbre phylogénétique 16S des procaryotes séquencés (d'après (Nelson 2000)).

1.4.3 Application du séquençage des génomes bactériens à l'étude de la variabilité génomique

1.4.3.1.1 Comparaison de plusieurs génomes au sein d'une même espèce

Le séquençage de génomes est l'approche la plus puissante pour identifier une variabilité génomique au sein d'espèces bactériennes, jusqu'au niveau de la souche ou même de l'isolat. Par exemple, et à la suite de l'attaque bioterroriste d'octobre 2001 qui a disséminé par voie postale une poudre de Bacille du charbon (*Bacillus anthracis*), une étude comparative menée par le TIGR (The Institute for Genomic Research : http://www.tigr.org/) sur les séquences de

deux isolats de *Bacillus anthracis* Ames a mis en évidence des locus polymorphes entre ces souches (Read 2002). De telles comparaisons de génomes sont grandement facilitées lorsqu'au moins l'un des génomes est entièrement séquencé (et non pas sous forme de fragments), ce qui est le cas pour un nombre croissant de bactéries d'intérêt médical ou économique : par exemple, 6 souches de *Staphylococcus aureus* sont actuellement disponibles. L'intérêt de la comparaison de ces 6 génomes pour la recherche de répétitions en tandem polymorphes chez *S.aureus* sera illustré dans cette thèse.

La comparaison d'isolats très proches montre que la diversité au sein d'une espèce peut varier considérablement. Alors que *Mycobacterium tuberculosis* (Sreevatsan 1997) et *Yersinia pestis* (Achtman 1999) ont une très faible diversité, il a été montré par MLST que *Neisseria meningitidis* (Feil 1999) et *Helicobacter Pylori* (Achtman 1999) présentent une grande diversité génétique. La comparaison entre deux souches d'*Helicobacter pylori* a également permis de mettre en évidence des régions variables dans leur génomes (Alm 1999). De même, la comparaison de *Listeria monocytogenes* (pathogène) et *Listeria innocua* (non pathogène) a permis d'identifier des facteurs putatifs de virulence chez Listeria (Glaser 2001). Les génomes de *Bordetella pertussis*, *Bordetella parapertussis* et *Bordetella bronchiseptica* ont été séquencés récemment et comparés. *B. pertussis* et *B. parapertussis* ont une plus grande spécificité d'hôte que *B. bronchiseptica*, elles infectent l'homme, tandis que *B. bronchiseptica* peut infecter de nombreux mammifères. *B. pertussis* et *B. parapertussis* ont dérivé de manière indépendante à partir d'ancêtres communs proches de *B. bronchiseptica*. Cette spécificité d'hôte semble être la conséquence de pertes de fonction plutôt que de gains, les différences de virulence sont probablement dues à des pertes de fonctions régulatrices (Parkhill 2003).

En utilisant des membranes à haute densité ou des lames de verre portant tous les gènes d'un génome complet (whole genome arrays), des comparaisons peuvent être effectuées expérimentalement et non plus seulement *in silico*, au sein d'un groupe de souches ou d'espèces très proches. Cette approche a été utilisée avec succès pour comparer *Mycobacterium tuberculosis* et *Mycobacterium bovis* BCG (Behr 1999), ainsi que pour différents isolats d'*Helicobacter pylori* (Salama 2000) et d'*E.* coli (Ochman 2000).

Enfin la comparaison d'espèces éloignées permet aussi d'identifier les fonctions importantes dans la pathogénicité des microorganismes. Ainsi, il a été démontré par comparaison des génomes complets d'*Haemophilus influenzae*, *Helicobacter pylori* et *E. coli*, que 70% des gènes communs aux bactéries pathogènes, mais absents d' *E. coli* sont des gènes impliqués dans l'interaction avec l'hôte (Snel 1999).

La comparaison de génomes d'espèces proches mais causant des maladies très différentes, telles que *M. leprae* et *M. tuberculosis*, ou *N. meningitidis* et *N. gonorrhoeae*, devrait permettre également d'identifier les gènes responsables de tel ou tel effet pathogène.

1.4.3.1.2 L'étude des SNPs et de leur rôle dans des espèces pathogènes

Dans le génome humain, les SNPs peuvent être associés à des sensibilités à des maladies et à des microorganismes pathogènes. Le remplacement d'un acide aminé du fait de la présence d'un SNP peut représenter un avantage adaptatif qui, s'il est fixé, peut aboutir à une nouvelle espèce (Perutz 1983).

Un certain nombre de SNPs ont été découverts chez les bactéries pathogènes. Ils peuvent participer à l'adaptation du pathogène. Par exemple des altérations localisées dans des gènes de régulation ou des gènes de structure peuvent conférer à la bactérie un avantage sélectif lors de l'infection, la propagation de l'épidémie ou bien même influer sur l'évolution à long terme de la virulence (Sokurenko 1999). Alors que beaucoup de mutations « pathoadaptatives » ont pour conséquence l'inactivation de gènes, pour les SNPs « adaptatifs », le plus souvent il en résulte une modification minime de la structure de la protéine. Par exemple il a été montré récemment que des variations alléliques des adhésines des fimbriaes chez *E. coli* et *S. thyphimurium* sont liées à des SNPs situés dans des gènes impliqués dans la spécificité d'hôte, le tropisme et peuvent servir de lien entre un mode de vie commensal et celui de pathogène (Weissman 2003).

1.5 Application du génotypage par les minisatellites à deux espèces bactériennes importantes en santé publique : *Pseudomonas aeruginosa* et *Staphylococcus aureus*

1.5.1 Les infections nosocomiales

Les infections nosocomiales sont contractées dans un établissement de soins. Une infection est considérée comme telle lorsqu'elle était absente à l'admission. Lorsque l'état infectieux du patient à l'admission est inconnu, l'infection est habituellement considérée comme nosocomiale si elle apparaît après un délai de 48 heures d'hospitalisation. Ces infections peuvent être directement liées aux soins (par exemple l'infection d'un cathéter) ou simplement survenir lors de l'hospitalisation indépendamment de tout acte médical (par exemple en période d'épidémie de grippe).

On distingue plusieurs types d'infections nosocomiales qui relèvent de modes de transmission différents :

- les infections d'origine "endogène" : le malade s'infecte avec ses propres germes, à la faveur d'un acte invasif et/ou en raison d'une fragilité particulière ;

- les infections d'origine "exogène" : il peut s'agir soit :

1) d'infections croisées, transmises d'un malade à l'autre par les mains ou les instruments de travail du personnel médical ou paramédical ;

2) d'infections provoquées par les germes du personnel porteur ;

3) d'infections liées à la contamination de l'environnement hospitalier (eau, air, matériel, alimentation...).

Quel que soit son mode de transmission, la survenue d'une infection nosocomiale est favorisée par la situation médicale du patient qui dépend de :

- son âge et sa pathologie : sont particulièrement réceptifs les personnes âgées, les immunodéprimés, les nouveaux-nés, en particulier les prématurés, les polytraumatisés et les grands brûlés ;

- certains traitements (antibiotiques qui déséquilibrent la flore des patients et sélectionnent les bactéries résistantes; traitements immunosuppresseurs) ;

- la réalisation d'actes invasifs, nécessaires au traitement du patient : sondage urinaire, cathétérisme, ventilation artificielle et intervention chirurgicale.

Les progrès médicaux permettent de prendre en charge des patients de plus en plus fragiles qui cumulent souvent de nombreux facteurs de risque. Ceci explique le caractère "inévitable" de certaines infections nosocomiales et la nécessité de prendre en compte ces facteurs de risque lors de l'interprétation des taux d'infections nosocomiales. *(Informations du ministère de la santé)*

En Europe, la prévalence des infections nosocomiales varie de 6,1% à 12,1%. Depuis 1990, plusieurs études de prévalence ont été réalisées en France, donnant des estimations de prévalence de 5,4% à 9,9%. Le taux d'incidence est deux fois plus élevé dans les grands hôpitaux universitaires. Les services les plus à risque sont dans l'ordre la réanimation (34,9%), la chirurgie (9,8%), la médecine (2,1%). Les sites anatomiques les plus fréquemment en cause sont dans l'ordre décroissant l'appareil urinaire, les poumons, le site opératoire. Les germes les plus fréquemment isolés sont *Staphylococcus aureus* (15,4%), *Pseudomonas aeruginosa* (14%), Staphylocoques coagulase négative (13,1%), *Escherichia coli (10,2%)*, et *Acinetobacter sp*. (9,7%). En réanimation, l'incidence des infections nosocomiales est beaucoup plus élevée. Elle peut atteindre ou dépasser 50% des admissions. En moyenne, 30% des malades en réanimation ont au moins une infection nosocomiale. En réanimation chirurgicale, il s'agit essentiellement d'infections urinaires, les infections des plaies opératoires représentant 8% des infections.

La létalité par infection nosocomiale dans les hôpitaux de plus de 500 lits est de 3,3%, comprenant 0,5% causant directement la mort. Les pneumopathies représentent la première cause de mortalité par infections nosocomiales, suivies des septicémies. La mortalité par pneumopathie nosocomiale est élevée (10 à 60%). La survenue d'une pneumopathie chez un malade en réanimation multiplie par 4 le risque de décès. La létalité des infections sur cathéters varie de 8 à 40% selon les études. La gravité des infections urinaires est moindre que celle des infections sur cathéters, la létalité globale étant de 0,1 %.

L'émergence épidémique de bactéries multi-résistantes en France et dans le monde est devenue depuis le début des années 90 un phénomène sur lequel des efforts de prévention particuliers ont été mis en oeuvre. Parmi ces bactéries, les staphylocoques dorés résistants à la méthicilline (MRSA) sont particulièrement préoccupants tant par leur incidence élevée que par leur fort pouvoir de dissémination interhumaine. Le risque de dissémination des souches MRSA est favorisé par les flux de malades non seulement à l'intérieur d'un hôpital, mais aussi entre les hôpitaux. De plus, cette espèce bactérienne commensale représente un exemple de résistance acquise exposant au risque de diffusion de la résistance à d'autres espèces pathogènes strictes. Les prévalences des MRSA varient de 39,6% en Italie à 0,1% au Danemark, la France se situant dans le haut de l'échelle avec 33,8%. La transmission est quasi exclusivement manuportée, le réservoir de germes étant le plus souvent un portage nasal par les malades ou le personnel soignant (Astagneau 1998) (CLIN Paris Nord).

Au cours de ma thèse, je me suis particulièrement intéressée à deux espèces bactériennes qui, comme nous venons de le voir dans ce paragraphe, posent des problèmes majeurs en santé humaine, il s'agit de *Pseudomonas aeruginosa* et de *Staphylococcus aureus*. Leurs caractéristiques et un tour d'horizon des méthodes de typage utilisées pour ces deux bactéries seront présentés dans le paragraphe suivant.

1.5.2 Pseudomonas aeruginosa

1.5.2.1 Description

Pseudomonas aeruginosa (ou bacille pyocyanique) est une bactérie à gram négatif, aérobie stricte qui appartient à la famille des Pseudomonadaceae. On trouve *P. aeruginosa* un peu partout dans l'environnement, dans le sol, l'eau, à la surface des plantes et parfois des animaux. *Pseudomonas* sécrète des pigments : la pyocyanine et la pyoverdine.

Cette bactérie, pathogène opportuniste de l'homme, est à l'origine d'infections du système urinaire, du système respiratoire (surtout chez les malades atteints de mucoviscidose) et de

septicémies, celles-ci concernant surtout les grands brûlés, les patients immunodéprimés atteints d'un cancer ou du sida.

P. aeruginosa est mobile grâce à une ciliature polaire et pousse facilement dans des milieux humides. Elle se multiplie à une température optimale de 37°C sur milieu ordinaire. Dans les hôpitaux, on trouve parfois *P. aeruginosa* dans les solutions antiseptiques. Une autre caractéristique importante est qu'elle peut vivre sous forme libre ou sous forme de biofilm. Sous cette forme « d'organisme pluricellulaire » les propriétés physiologiques de *P. aeruginosa* changent, elle devient par exemple particulièrement résistante aux antibiotiques (Costerton 1999). Des biofilms peuvent ainsi se former sur les sondes, les cathéters, les canalisations, les lavabos. Par ailleurs, *P. aeruginosa* est la première cause de décès chez les patients atteints de mucoviscidose.

1.5.2.2 La mucoviscidose

La mucoviscidose ou Cystic Fibrosis (CF) en anglais est une maladie génétique autosomale récessive. Le gène CFTR (Cystic Fibrosis Transmembrane conductance Regulator) code une protéine de la famille des transporteurs ABC (ATP- binding cassette) qui transporte de nombreux substrats ainsi que du chlore (Schwiebert 1999). Ce gène est porté par le chromosome 7. L'inactivation de CFTR par mutation provoque une absence ou une diminution des sécrétions chlorées et une augmentation de l'absorption du sodium dans l'air. On observe une production anormale de mucus broncho pulmonaire et une obstruction des poumons. En France une personne sur 30 est porteuse d'un allèle muté. Cette maladie se traduit par une insuffisance respiratoire grave et des troubles digestifs permanents. La mucoviscidose affecte la qualité du mucus, principalement dans les bronches et dans le tube digestif. Son épaississement anormal empêche son écoulement. Des lésions pulmonaires irréversibles apparaissent et conduisent à une insuffisance respiratoire chronique grave et évolutive. Aucun traitement curatif n'existe à l'heure actuelle.

Les patients atteints de mucoviscidose souffrent d'infections bactériennes chroniques (Renders 2001). Le plus souvent ils sont infectés par *Pseudomonas aeruginosa*, *Staphylococcus aureus* (Branger 1994) et *Haemophilus influenzae* mais aussi *Streptococcus pneumoniae*, *Escherichia coli*, *Burkholderia cepacia*, *Klebsiella* spp., *Proteus* spp., *Serratia* spp., *Enterobacter* spp., *Citrobacter* spp et les streptocoques de groupe A. La plupart des malades meurent des complications associées à ces infections chroniques du tissu pulmonaire.

Les souches de *P. aeruginosa* qui poussent préférentiellement dans les poumons des patients atteints de mucoviscidose sont de phénotype « mucoïde » (Oliver 2000), c'est-à-dire qu'elles sont « collantes ». Elles sont entourées d'un gel d'alginate impliqué dans la virulence et ne peuvent plus être éliminées par les cellules cilliées des muqueuses pulmonaires. De plus, ces

souches ne sont pas toujours sérotypables, ce qui rend leur surveillance particulièrement difficile.

La cause majeure de la persistance de *P. aeruginosa* dans les poumons des patients est l'apparition chez ces souches du phénotype mucoïde (Govan 1996). Ce phénomène illustre l'adaptation d'un pathogène opportuniste à une infection pulmonaire. Un lien a été proposé entre la formation d'un biofilm dans les poumons et la résistance aux antibiotiques, pour expliquer la variation phénotypique (Drenkard 2002). Le phénotype mucoïde est complexe, il ne résulte pas de l'expression d'un seul gène, mais de plusieurs gènes impliqués dans la voie de biosynthèse de l'alginate. Une étude récente d'analyse par puce à ADN de l'expression globale des gènes de souches mucoïdes a été réalisée, pour tenter de mieux caractériser l'ensemble des gènes dont l'expression est modifiée lorsque le phénotype est mucoïde (Firoved 2003).

Les travaux de Kresse et al. ont montré que les souches persistantes dans les poumons de patients CF présentent toutes une grande inversion chromosomique (Large Chromosomal Inversion ou LCI). Il s'agit d'un nouveau mécanisme d'adaptation de la bactérie. La séquence IS6100 induit le couplage d'un insertion-délétion. L'insertion dans les gènes *wbpM*, *pilB* et *mutS* conduit au phénotype habituellement observé dans les souches de patients CF, c'est à dire le défaut et l'hypermutabilité de l'antigène O et des pili de type IV. Cette adaptation de la bactérie par LCI dans une nouvelle niche écologique est un modèle pour l'étude de l'évolution des génomes bactériens (Kresse 2003). Le génotypage de *P. aeruginosa* est à développer davantage pour en améliorer la résolution afin de pouvoir caractériser et suivre toutes les souches persistantes dans les poumons.

1.5.2.3 Résistances aux antibiotiques

Les souches responsables d'infections nosocomiales sont le plus souvent multirésistantes aux antibiotiques. *P. aeruginosa* est particulièrement source de problèmes car cette espèce possède naturellement des résistances à de nombreux antibiotiques (amoxicilline, céfotaxime, tétracyclines, cotrimoxazole), et qu'elle peut acquérir de nouvelles résistances via des mutations. Ceci a pour conséquence l'apparition de souches résistantes à tous les antibiotiques utilisés (Livermore 2002). Des études épidémiologiques ont montré qu'un nombre limité de clones de *P. aeruginosa* est retrouvé régulièrement dans les hôpitaux européens. Les types antigéniques les plus souvent en cause lors d'épidémies sont représentés par le sérotype O:11 et le sérotype O:12. Ce dernier a pour origine la dissémination clonale d'une souche unique à travers toute l'Europe (Pitt 1989; Mifsud 1997). De plus en plus de souches multirésistantes de sérotype O:11 sont observées (Tassios 1998).

1.5.2.4 Méthodes d'identification de *P. aeruginosa* : phénotypage/génotypage

Les méthodes de phénotypage les plus utilisées pour *P. aeruginosa* sont le sérotypage et le lysotypage (Renders 2001). Cependant, toutes les souches ne sont pas sérotypables, et lorsqu'elles le sont, le pouvoir discriminant observé est faible, certains sérotypes étant très fréquents. *P. aeruginosa* synthétise différentes pyocines (des substances anti-bactériennes). Les différents profils de pyocines synthétisées ont été utilisés comme moyen de phénotypage des souches. Le problème majeur est que les profils de synthèse des pyocines sont instables puisqu'ils changent avec le métabolisme de la bactérie. Les conditions de culture pour effectuer les tests peuvent être contrôlées, en revanche la durée de l'antibio-thérapie utilisée pour le traitement des patients est un paramètre non contrôlable. Ceci peut conduire à l'apparition de fausses « nouvelles » souches.

Dans le cas de souches provenant de poumons de patients atteints de mucoviscidose, des conversions phénotypiques sont observées (phénotype mucoïde). Cette instabilité du phénotype entraîne des problèmes de typabilité et de reproductibilité. Des souches de génotype différent développent un phénotype commun (Romling 1994).

Les techniques de génotypage permettent de pallier le problème des variations de phénotypes non liées à de réelles modification des gènes. Les techniques de génotypage disponibles actuellement pour *P. aeruginosa* sont le RFLP et l'analyse par PFGE, le ribotypage et l'AP-PCR (Bennekov 1996) (Grundmann 1995) (Renders 1996). Plusieurs études ont été menées pour comparer les différentes techniques de typage. Par exemple en 1996 une étude a été réalisée pour comparer deux méthodes de phénotypage (lysotypage, production de pyocines) et deux de génotypage (AP-PCR,PFGE) d'une même collection de souches provenant de patients atteints de mucoviscidose. Les techniques d'AP-PCR et PFGE ont montré une meilleure résolution comparée à celles obtenue par sérotypage et par production de pyocines. Les deux techniques de génotypage ont donné des résultats concordants, bien que de résolution différente. L'AP-PCR est une technique rapide et facile à mettre en œuvre (par rapport au RFLP-PFGE) et permet de déterminer la clonalité parmi les souches de patients chroniquement infectés par *P. aeruginosa* (Renders 1996). Les techniques de génotypage sont donc préférables pour typer ces souches particulières. De très nombreuses études de comparaison des techniques de génotypage ont été réalisées.

PFGE et AP-PCR manquent de reproductibilité entre laboratoires (Dabrowski 2003), (Foissaud 1999). Cependant, le PFGE est considéré comme la méthode de choix pour le génotypage de *P. aeruginosa*, car c'est la plus discriminante qui soit actuellement disponible. Les inconvénients majeurs de cette technique sont qu'elle est couteûse et difficile à mettre en œuvre pour des analyses cliniques de routine.

Par ailleurs, il a été montré que le ribotypage automatisé avec le RiboPrinter (Qualicon) constitue une technique reproductible entre laboratoires (Brisse 2002) mais la limite majeure de cette technique chez *P. aeruginosa* est le faible pouvoir discriminant pour les souches des ribogroupes majoritaires (Brisse 2000).

Dans l'objectif de répondre aux problèmes de typabilité, de faisabilité, de reproductibilité et de résolution de l'analyse pour le typage de souches de *P. aeruginosa*, j'ai développé au cours de ma thèse une série de marqueurs de type VNTR (Variable Number of Tandem Repeats) (Onteniente 2003). Ce travail a été facilité grâce aux données de séquençage du génome complet de la souche PAO1 publié en août 2000 (Stover 2000). Jusqu'à présent, aucune étude de VNTRs n'avait été réalisée pour *P. aeruginosa*. Une seule séquence répétée en tandem a été décrite dans le passé, la répétition présente dans le gène *algP* impliqué dans la régulation de la mucoïdie chez *P. aeruginosa* (Deretic 1990). AlgP est une protéine qui possède un domaine C-terminal ressemblant à l'extrémité de l'histone H1 chez les eucaryotes. AlgP participe à l'activation transcriptionnelle du gène *algD*. Celui-ci correspond à une étape clé dans l'établissement du phénotype mucoïde chez *P. aeruginosa* (Deretic 1992). La séquence répétée en tandem dans le gène *algP* est constituée de 6 motifs de 75 pb très conservés. On observe une répétition de moindre conservation interne de 45 motifs de 12 pb qui va au delà de la répétition des motifs de 75 pb, comme si la répétition en tandem avait évolué en plusieurs étapes. Les variations du nombre de motifs ne semblent pas avoir d'effet sur le phénotype mucoïde des souches (Deretic 1990).

1.5.3 Staphylococcus aureus

1.5.3.1 Description

Le staphylocoque doré est une bactérie à gram positif de forme sphérique, non sporulante et immobile, qui forme de petites grappes. Sur la base d'une analyse du gène de l'ARN16S, le genre *Staphylococcus* est proche du groupe des Bacillus-Lactobacillus-Streptococcus (Voir Figure 8)

Chez l'homme cette bactérie est présente sur la peau et dans les fosses nasales, mais aussi dans la flore normale de l'intestin. *S. aureus* se multiplie à des températures allant de 15°C à 45°C, forme des colonies jaunes sur milieu riche et est hémolytique sur du milieu sang-agar. Ces bactéries peuvent pousser en condition d'aérobie ou bien par fermentation. Sur 19 espèces de *Staphylococcus*, deux sont des commensales de l'homme : *Staphylococcus aureus* qui nous intéresse plus particulièrement ici, et *Staphylococcus epidermidis* présente sur la peau. *S. aureus* est une bactérie pathogène, en revanche, la plupart des souches de *S.*

epidermidis ne le sont pas, et jouent un rôle protecteur de la flore normale de leur hôte. Cependant, dans l'environnement hospitalier, *S. epidermidis* peut être pathogène.

1.5.3.2 Infections liées à *S. aureus*

S. aureus peut infecter de nombreux tissus. Le plus souvent, il s'agit d'infections suppuratives, cutanées, sous-cutanées et des muqueuses : abcès divers dont des furoncles, des panaris, infections des plaies, de la sphère ORL (sinusites, otites, angines), infections oculaires, infections urinaires et rénales, infections ostéoarticulaires (ostéomyélites et arthrites), septicémies (sur cathéter, thrombophlébites, endocardites infectieuses), infections pulmonaires, infections neuro-méningées, infections intestinales, toxi-infections alimentaires. *S. aureus* exprime un certain nombre de facteurs de virulence (Kuroda 2001). :

- des protéines de surface qui favorisent la colonisation des tissus hôtes ;
- des invasines qui favorisent la diffusion de la bactérie dans les tissus (ex : leukocidine, kinases, hyaluronidase) ;
- des facteurs de surface qui inhibent la phagocytose (ex : proteine A) ;
- des facteurs biochimiques qui augmentent la survie de la bactérie dans les phagocytes (ex : catalase) ;
- des leurres immunologiques (ex : protéine A, coagulase) ;
- des toxines qui lysent les membranes des cellules eucaryotes (ex : hémolysines, leukotoxine, leukocidine) ;
- des exotoxines qui causent des dommages aux tissus hôtes et provoquent les symptômes (ex : TSST, ET).

1.5.3.3 Apparition de souches résistantes à la méthicilline en milieu hospitalier

Staphylococcus aureus pose un véritable problème de santé publique. L'adaptation de *S. aureus* à l'environnement hospitalier a été marquée par l'acquisition de résistances aux antibiotiques souvent peu de temps après leur introduction (Enright 2003). Peu après le début de l'utilisation de la pénicilline en 1944, les premières souches résistantes étaient isolées (Barber 1948). De même, l'utilisation d'une pénicilline semi-synthétique, la méthicilline, en 1959 a rapidement été suivie de l'apparition en 1961 des premières souches dites MRSA (Methicilline Resistant *Staphylococcus aureus*) observées au Royaume Uni (Jevons 1961). Ces souches MRSA ont ensuite été observées dans les autres pays européens puis au Japon, en Australie et aux Etats Unis. Ainsi depuis les années 70, les souches MRSA sont devenues la cause majeure d'infections nosocomiales à travers le monde.

Le gène *mecA* est responsable de la résistance à la méthicilline. Il code la protéine PBP2', protéine de fixation à la pénicilline. Ce gène est situé dans un élément génétique mobile, la cassette *SCCmec* (Staphylococcal Chromosomal Cassette mec) qui proviendrait d'une espèce voisine. La cassette mec contient le gène *mecA* ainsi que les gènes *ccrA* et *ccrB* (cassette chromosome recombinase genes) qui codent les recombinases nécessaires pour sa mobilité (Katayama 2000 ; Hiramatsu 2001). L'intégration de cette cassette dans le chromosome d'une souche sensible à la méthicilline convertit celle-ci en souche résistante. Les souches résistantes à la méthicilline sont la cause majeure d'infections nosocomiales et il devient de plus en plus difficile de les combattre du fait de l'apparition de résistances à toutes les classes d'antibiotiques utilisés actuellement. Différents clones MRSA ont été distingués en fonction des différents types de cassettes mec intégrées dans leur génome (Hiramatsu 2002). Quatre types de régions mec ont été décrits (types I – IV) en fonction du type de complexe du gène *ccr* (types 1–3) et de la classe du complexe *mec* (A et B) (Okuma 2002). Les premières souches MRSA isolées dans les années 1960 (clone archaïque) ont une cassette mec de type I. Les cassettes de types II et III sont typiques des clones MRSA modernes, enfin le type IV (classé en deux sous-types : IV a et IV b) est associé aux nouvelles souches MRSA émergentes dans la communauté (Daum 2002).

1.5.3.4 Emergence des souches C-MRSA acquises dans la communauté

Il a été montré récemment que la résistance à la méthicilline s'observe aussi dans des souches de la flore intestinale normale humaine. Elles ont été converties en souches MRSA par une nouvelle cassette mec (Ma 2002). Ceci indique que le problème s'étend à l'extérieur du milieu hospitalier puisque des individus sains sont colonisés par des souches MRSA. On parle de community-acquired MRSA (C-MRSA).

1.5.3.5 Evolution des souches MRSA

Deux théories s'affrontent : la première est celle d'un clone unique qui a acquis une fois le gène *mecA* et qui s'est ensuite disséminé à travers le monde (Kreiswirth 1993). Cette hypothèse a été appuyée par une étude RFLP et hybridation avec des sondes mecA et Tn554 sur des souches collectées dans le monde entier.

La seconde théorie propose que les souches MRSA aient évolué en plusieurs fois par transfert horizontal de *mecA* dans des souches MSSA phylogénétiquement distinctes.

Cette seconde théorie a été confirmée par au moins quatre types d'études épidémiologiques. Dans la première, la technique de MLEE a été mise en œuvre et les résultats ne vont pas du tout dans le sens d'une origine clonale des souches MRSA puisque le gène *mecA* a été

retrouvé dans des souches de lignées phylogénétiques différentes (Musser 1992). Dans une seconde étude, une analyse utilisant la technologie des puces à ADN a également montré que le gène *mecA* a été observé dans des souches provenant de 5 lignées non reliées, et appuie l'hypothèse du transfert horizontal, fondamental dans l'évolution des souches MRSA (Fitzgerald 2001). La troisième étude par RFLP-PFGE et ribotypage d'un millier de souches MSSA et MRSA provenant d'Amérique du nord et d'Europe, collectées depuis les années 60, a confirmé que le transfert horizontal joue un rôle important dans la dissémination du gène *mecA* dans la population de *S. aureus* (Wielders 2002). Enfin, une étude MLST va également dans ce sens (Enright 2002).

Ainsi, il est maintenant accepté que les souches MRSA modernes sont la conséquence de l'acquisition indépendante de la cassette SCC*mec* par différentes lignées de *S. aureus*.

1.5.3.6 Résistance à la vancomycine

La vancomycine était, jusqu'en 1997, l'antibiotique de dernier recours, le seul efficace contre les souches MRSA. En 1997 sont apparues des souches de résistance intermédiaire à la vancomycine (VISA : Vancomycine-Intermediately-susceptible *S.aureus*) (Hiramatsu 1997) puis à partir de 2002, des souches VRSA (Vancomycine-Resistant *S.aureus*). Des souches VISA qui sont également résistantes à la teicoplanine sont appelées GISA (pour Glycopeptide intermediate *S.aureus)* (Linares 2001). Des souches totalement résistantes aux glycopeptides (GRSA) ont été décrites récemment. Une explication pourrait être le transfert des gènes impliqués dans la résistance au glycopeptide depuis des souches de streptocoques résistantes aux glycopeptides vers des souches de *S. aureus*. En effet, ces deux espèces étaient à chaque fois présentes sur les sites où des souches GRSA ont été observées. Cependant, très peu de souches GRSA ont été décrites (8 cas aux USA en 2002) et il est difficile de prédire si ces souches vont connaître une expansion importante ou rester des évènements sporadiques (comme c'est le cas pour les souches GISA) (Johnson 2002).

1.5.3.7 Six souches de *S. aureus* entièrement séquencées

Pour tous les génomes de *S. aureus* séquencés, la méthode du WGS (whole genome shotgun) a été utilisée. Les différents projets de séquençage ont montré que les génomes du genre *Staphylococcus* sont composés d'un mélange complexe de gènes dont un certain nombre est issu de transfert horizontal. La plupart des gènes de résistance aux antibiotiques est portée par des plasmides ou des éléments génétiques mobiles (Kuroda 2001). L'analyse des génomes séquencés a aussi permis de montrer qu'environ la moitié du génome de *S. aureus* a été transmis verticalement depuis un ancêtre commun au groupe de bactéries *Bacillus/Staphylococcus*.

Le génome d'une taille de 2.81Mb de la souche N315, résistante à la méthicilline, a été entièrement séquencé. Cette souche isolée en 1982 est à l'origine d'infections nosocomiales et aussi d'infections acquises dans la communauté.

Le génome de la souche Mu50, très proche de N315, a été séquencé par le même groupe. Elle présente, en plus de la résistance à la méthicilline, une résistance à la vancomycine ; c'est donc une souche MRSA et VRSA.

MW2 est une souche MRSA très virulente, acquise dans la communauté et non dans l'environnement hospitalier comme ce fut le cas pour Mu50 et N315. Son génome a été séquencé un an après Mu50 et N315, dans le but de pouvoir comparer les 3 génomes MRSA et tenter de comprendre les bases génétiques à l'origine de cette forte virulence (Baba 2002). Le Tableau 5 décrit les 6 souches de *S. aureus* dont le génome séquencé a été utilisé pour rechercher des répétitions en tandem polymorphes par comparaison des séquences.

Tableau 5 : Six génomes *S. aureus* utilisés pour la comparaison

Souches séquencées	**N315**	**Mu50**	**MW2**	**NCTC8325**	**MRSA252**	**MSSA476**
Taille du génome	2.8Mb	2.9Mb	2.8Mb	2.9Mb	2.9Mb	2.8Mb
Résistance vis à vis de la méthicilline	**MRSA** (SCC*mec* type II)	**MRSA** (SCC*mec* type II) + VRSA	**MRSA** (SCC*mec* type IVa)	MSSA	MRSA	MSSA
Institution	NITE	NITE	NITE	Oklahoma university	Sanger Center	Sanger Center
Caractéristiques	souche hospitalière	souche hospitalière	souche hyper virulente acquise dans la communauté	vieille souche de laboratoire	souche hospitalière épidémique	souche hyper-virulente acquise dans la communauté
Publication	(Kuroda 2001)	(Kuroda 2001)	(Baba 2002)	Non publié	Non publié	Non publié

D'autres souches de *S. aureus* sont en cours de séquençage ou non publiées à la date du 15 novembre 2003 d'après le site GOLD : « Genome Online Database » [http://ergo.integratedgenomics.com/GOLD/] :

- Les souches *S. aureus* 930131 (2,56Mb), *S. aureus* ATCC 29213 (2,62Mb), ont été complètement séquencées par Integrated Genomics Inc. et non publiées.
- Une souche *S. aureus* sans aucune précision sur la souche, a été séquencée par Genomes Therapeutics et non publiée.
- Le séquençage de la souche COL (2,80Mb), souche MRSA hospitalière, n'est pas encore terminé par le TIGR.

1.5.3.8 Techniques de typage mises en œuvre pour *Staphylococcus aureus*

1.5.3.8.1 Phénotypage et génotypage

Les principales techniques de phénotypage et génotypage présentées dans le chapitre 1.2 ont été utilisées pour le typage de *S. aureus*. La littérérature est très riche en références concernant des études de typage de *S. aureus* et d'études de comparaison de techniques, nous nous limiterons à quelques unes (van Belkum 2000).

Les profils de résistance aux antibiotiques ont constitué le premier outil épidémiologique pour le typage des souches. En effet, dans un premier temps il faut identifier les souches résistantes à la méthicilline. Pour une confirmation, le test (latex agglutination) de détection de la protéine PBP2 qui confère la résistance à la méthicilline peut être mis en œuvre facilement (van Leeuwen 1999). Par ailleurs, le lysotypage a été pendant longtemps la méthode de phénotypage de référence pour *S. aureu*s (van Belkum 1993).

Pour ce qui est des techniques de génotypage, une PCR pour amplifier *mecA* permet de savoir rapidement si la souche possède le gène qui confère la résistante à la méthicilline. Ce gain de temps par rapport à un test de résistance à la méthicilline peut être très important pour commencer un traitement (Murakami 1991). Les autres techniques de génotypage évoquées dans le paragraphe 1.2.2 ont toutes été utilisées pour *S. aureus*.

Il a été montré que les techniques de génotypage pour *S. aureus* sont plus efficaces que les techniques de phénotypage (Tenover 1994). Dans cette étude, douze techniques de typage ont été comparées : antibiogramme, lysotypage (Khalifa 1989), tests biochimiques (Hebert 1988), immunodétection (Tsang 1983), typage des séquences IS257/431, FIGE (Field Inversion Gel Electrophoresis) (Goering 1992), MLEE, restriction de l'ADN plasmidique, PFGE, analyse de restriction du produit de PCR du gène de la coagulase, RFLP et ribotypage.

Parmi les techniques de détection récemment développées chez *S. aureus* on peut citer quelques exemples (van Belkum 2003) : la PCR en temps réel a été adaptée à l'analyse de souches MRSA (Grisold 2002) ; la détection combinée des souches MRSA par test d'agglutination de PBP2 et par PCR en temps réel du gène *mecA* constitue un outil de typage satisfaisant pour du diagnostic de routine (Rohrer 2001).

Des efforts sont faits aussi dans l'objectif de pouvoir analyser directement les échantillons prélevés sur les patients, et ce dans un temps réduit le plus possible , par exemple via la mise au point d'un test de détection rapide par PCR quantitative de souches MRSA à partir de

prélèvements sans passer par une étape de culture ; le résultat est obtenu en moins de 6 heures (2 à 3 jours avec la procédure habituelle) (Francois 2003).

Le typage binaire est une technique de génotypage récente développée pour l'étude de *S. aureus*. Elle a été développée à partir des études RAPD réalisées précédemment. Van Leeuwen a identifié par RAPD des régions uniques dans le génome de *S. aureus* qui pourront servir de cibles pour le typage binaire. Il s'agit de fragments qui ne sont pas communs à toutes les souches. Cette technique a été développée parce que les profils obtenus en RAPD étaient trop complexes à reproduire et à interpréter, il était souhaitable d'obtenir des résultats analysables sans ambiguïté et donc avec moins de bandes à analyser. Les 15 sondes utilisées dans l'article de Van Leeuwen (van Leeuwen 1999) ont été obtenues en sélectionnant des fragments uniques en RAPD, puis ces fragments ont été extraits sur gels et séquencés, puis testés comme sondes pour hybrider une membrane sur laquelle sont fixés les ADN hydrolysés de 14 souches qui ne sont pas reliées épidémiologiquement entre elles. La détection des fragments hybridés se fait par révélation ECL (Enhanced Chemical Luminescence). Le résultat est codé de façon binaire : 1 si la sonde est hybridée et 0 si elle ne l'est pas. On obtient un code binaire pour chaque souche, composé d'autant de chiffres que de sondes testées. Plusieurs laboratoires ont typé la même collection de souches. Les laboratoires qui ont suivi le protocole à la lettre ont obtenu des résultats comparables. Le typage binaire semble constituer selon les auteurs une méthode de typage simple et robuste permettant une bonne reproductibilité des résultats entre laboratoires différents (van Leeuwen 2002).

1.5.3.8.2 Les répétitions en tandem

Les techniques traditionnelles de phénotypage ont progressivement été remplacées par des techniques basées sur l'étude de l'ADN. La stratégie courante pour le génotypage des souches de *S. aureus* est l'amplification de séquences répétées en tandem, souvent localisées dans des séquences de protéines associées à la surface membranaire. Ces séquences de type VNTR (Variable Number of Tandem Repeat) permettent de distinguer les souches entre elles mais il faut en général étudier plusieurs locus pour discriminer les souches. Le séquençage d'un locus, s'il est suffisamment polymorphe, pourrait permettre de n'étudier qu'un locus, au lieu d'une étude MLVA pour le typage des souches. Ceci a été développé pour le gène *spa* (Proteine A), et pour *coa*, le gène codant la coagulase.

1.5.3.8.2.1 Répétitions en tandem localisées dans des séquences intragéniques

Parmi les gènes impliqués dans la virulence de *S. aureus*, un certain nombre possède des répétitions en tandem, dont certaines ont déjà été analysées lors d'études épidémiologiques. Il est particulièrement intéressant de tenter de faire un lien entre la virulence et l'expression de certain gènes qui pourrait être modulée par la variation du nombre de motifs dans une

séquence répétée en tandem. Le Tableau 6 présente les gènes possédant une répétition en tandem et dont le polymorphisme a déjà été étudié.

Tableau 6 : Répétitions en tandem intragéniques déjà étudiées chez *S. aureus*

nom du gène :	produit:	fonction:	taille du motif :	méthodes de génotypage:	références:
coa	Staphylocoagulase (exoenzyme)	coagulation des tissus hôtes	81pb	VNTR, séquençage	(Goh 1992); (Shopsin 2000)
sspA	Serine protéase, V8 protéase (exoenzyme)	protéolyse des tissus hôtes	9pb	VNTR	(Rice 2001), (Sabat 2003)
clfA, clfB	protéines de liaison au fibrinogène, riches en Ser-Asp	adhésion cellulaire aux tissus hôtes	18pb	VNTR	(McDevitt 1994), (McDevitt 1995), (Sabat 2003)
sdrC, sdrD, sdrE	protéines de liaison au fibrinogène, riches en Ser-Asp	adhésion cellulaire aux tissus hôtes	18pb	VNTR	(Josefsson 1998), (Sabat 2003)
fnbB, fnbA	protéines de liaison à la fibronectine	adhésion cellulaire aux tissus hôtes	42pb (d'après TRF[a])	VNTR	(Patti 1994)
spa	proteine A (liaison aux immunoglobulines G)	désordre immunitaire chez l'hôte	24pb	VNTR, séquençage	(Frenay 1996), (Shopsin 1999), (Sabat 2003)

[a] : d'après le logiciel Tandem Repeats Finder

La coagulase est un facteur de virulence important chez *S. aureus*. La séquence répétée en tandem a été découverte lors du séquençage du gène *coa* (Kaida 1987). Cette répétition en tandem présente un polymorphisme pour le nombre de motifs présents d'une souche à l'autre et aussi au niveau de la séquence des motifs (Goh 1992), (Hookey 1998). La répétition en tandem présente dans le gène *coa* a été séquencée pour évaluer son apport pour des études épidémiologiques en comparaison avec les résultats obtenus avec le séquençage du gène *spa* (Shopsin 1999). Des souches MRSA d'origines géographiques différentes et collectées à différentes périodes ont été analysées. Les résultats montrent que ce marqueur a un faible index de polymorphisme, donc il serait utile plutôt pour des études épidémiologiques à long terme ou en complément du séquençage du gène *spa*. Le gène *coa* n'évolue pas à la même vitesse que *spa*, il est beaucoup plus stable (Shopsin 2000). Le séquençage du locus *spa* sera présenté dans la suite de ce travail.

S. aureus exprime à sa surface des protéines qui jouent un rôle important dans la virulence (Foster 1994). Il s'agit d'une famille des protéines à motif serine-aspartate répété, les protéines sdr (**sd-r**epeats). Ces protéines de surface participent à la fixation à la matrice extracellulaire de la cellule hôte ainsi qu'à la formation de biofilms. Les protéines associées à la paroi extracellulaire ont un certain nombre de caractéristiques communes, notamment elles possèdent un peptide signal et N-terminal nécessaire pour la sécrétion ainsi qu'un signal C-

terminal conservé nécessaire pour l'attachement de la protéine à la paroi cellulaire. Ce signal de sortie de la protéine possède un motif conservé LPXTG. La Sortase est l'enzyme membranaire qui clive les protéines à motifs LPXTG entre la glycine et la thréonine. Les protéines LPXTG et la Sortase existent aussi chez beaucoup d'autres bactéries à gram positif.

S. aureus exprime 11 protéines LPXTG (proteine A (Spa), ClfA et ClfB (clumping factor), sdrC, sdrD, sdrE: proteines sdr (serine aspartate repeats), Cna (collagen binding protein), FnbpA et FnbpB (Fibronectine binding proteins), Pls (Plasmin sensitive protein) et FmtB. Parmi ces protéines de liaison à la membrane et qui ont un motif LPXTG, certaines font partie des MSCRAMM (pour Microbial Surface Components Recognizing Adhesive Matrix Molecules) (Patti 1994). Ces protéines reconnaissent des ligands dans la matrice extracellulaire de la cellule hôte. Il existe notamment des MSCRAMM chez *S. aureus* spécifiques de la fibronectine , du collagène, de la laminine, de la vitronectine, de la thrombospondine, de l'elastine, de la sialoproteine et du fibrinogène.

Dans un article récent est présentée une analyse MLVA par PCR multiplexe. Les gènes *spa*, *sdr* (un seul couple d'amorce pour amplifier les 3 locus : sdrC, sdrD sdrE) , *ClfA*, *ClfB* et s*spA* sont amplifiés dans une même réaction de PCR (Sabat 2003). Ce sont alors des profils mutlibandes qui sont analysés, comme pour d'autres techniques de génotypage.

1.5.3.8.2.2 Répétitions en tandem localisées dans des régions intergéniques

1.5.3.8.2.2.1 Les séquences STARs: (StaPhylococcal Aureus Repeat sequences)

Dans l'article sur le séquençage de Mu50 et N315 (Kuroda 2001), des séquences répétées spécifiques des *Staphylococcus* ont été observées : les séquences STAR, déjà décrites en 2000 par Cramton (Cramton 2000). Il s'agit d'une séquence d'ADN hautement variable parmi les souches de *S.aureus* et qui est trouvée en plusieurs copies dans le génome. Ces séquences répétées sont pour la plupart intergéniques. Initialement, elles ont été étudiées à deux locus précis, le premier situé entre les gènes *uvrA* et *hprK*, et le second entre les gènes *icaC* et *geh*. Ensuite, les données de Southern blot d'une part, et de séquençage des génomes complets d'autres part, ont montré la présence d'environ 70 séquences STARs dans les génomes de *S. aureus*. Ces séquences ont été détectées par Southern blot chez d'autres espèces du genre *Staphylococcus (S. haemolyticus, S. lentus, S. warneri, S. sciuri)*. La fonction de ces séquences riches en GC n'est pas connue. La présence ubiquitaire de ces séquences dans le génome suggère une importance (présente ou passée) dans la vie de la bactérie. Les auteurs ont testés par Southern blot le nombre de séquences STARs dans différentes souches *S. aureus*. Ils ont observé pour plusieurs souches qui dérivent de la souche de référence NCTC8325 que les éléments STARs sont stables sur plusieurs années en ce qui concerne leur nombre et leur position dans le génome. La plupart des séquences STARs ont le profil

suivant: B-(C)n-A. Une à six séquences C sont répétées en tandem entre les motifs A et B (Kuroda 2001) :

séquence A : 97 pb

TGACTAGAATTGAAAAAAGCTTGTTACAAGCGCATTTTCGTTCAGTCAACTACTGCCAAATATAAC
TTTGTAGAGCATTGAACATTGATTTTATGTC

séquence B :46 pb

GGGAGTGGGACAGAAATGATATTTTCGCAAAATTTATTTCGTTGTC

séquence C : 57 pb

CCCCAACTTGCACATTATTGTAAGCTGACTTTTCGTCAGCTTCTGTGTTGGGGCCCC

La séquence C a un pourcentage en GC de 51% donc plus important que celui observé pour le reste du génome de *S. aureus* (32.8% pour N315 et 32.9% pour Mu50). La diversité de séquences à un même locus parmi les souches de *S. aureus* suggère que les séquences STARs évoluent rapidement et permettent donc une étude pour l'identification et l'analyse d'isolats cliniques.

Une étude de souches MRSA par typage de séquences STARs amplifiées par PCR a été réalisée. Le pouvoir discriminant obtenu est comparable à celui observé en champ pulsé. La bonne reproductibilité des amplifications par PCR des séquences STAR en fait donc un outil de typage de *S. aureus* tout à fait satisfaisant (Quelle 2003). Nous avons étudié un certain nombre de séquences STARs pour l'étude MLVA qui a été développée au cours de cette thèse.

1.5.3.8.2.2.2 Séquence dru (**d**irect **r**epeat **u**nit)

La résistance à la méthicilline résulte de l'expression du gène *mecA*. Le gène *mecA* est localisé dans une grande région, la région mec qui possède une séquence répétée, la séquence intergénique dru. L'origine des répétitions dru chez *S.aureus* n'est pas connue (Nahvi 2001). La cassette SCC*mec* étant absente dans les souches sensibles à la méthicilline, le marqueur dru a une utilisation réservée à l'étude épidémiologique des souches MRSA. Les allèles de 6 à 11 répétitions de 40 pb ont été séquencés dans cette étude. Il semblerait que les répétitions centrales soient délétées et les premières répétitions dupliquées (Senna 2002).

1.6 Application du génotypage par les minisatellites à une espèce pathogène d'émergence récente : *Yersinia pestis*

Le genre Yersinia appartient à la famille des Enterobacteriaceae qui comprends 11 espèces dont 3 sont pathogènes pour l'homme : *Y. pestis*, *Y. pseudotuberculosis* et *Y. enterocolitica*.

Y. pestis est une bactérie à gram négatif, immobile et non sporulante. C'est une espèce d'émergence récente issue de *Y. pseudotuberculosis* (Achtman 1999).

La peste est une zoonose qui atteint en premier les rongeurs, l'homme n'a aucun rôle dans la survie à long terme de *Y. pestis :* voir (Perry 1997). La transmission entre rongeurs a lieu à travers leurs puces. La contamination peut avoir lieu par contact ou par ingestion mais ces deux voies ne permettent pas le maintien de *Y. pestis* dans des réservoirs animaux. La puce acquiert *Y. pestis* à partir d'un repas avec du sang contaminé. La bactérie se dissémine depuis la piqûre de la puce par le réseau lymphatique et les ganglions infectées forment les bubons. Ensuite elle passe dans le sang. La puce Xenopsylla cheopis est considérée comme le vecteur classique.

Trois biovars (antiqua, medievalis, orientalis) sont décrits dans l'espèce *Y. pestis*. Chacun correspondrait à une des trois grandes pandémies de peste survenues au cours de l'histoire. La première pandémie établie a eu lieu au VIème siècle, la peste Justinienne, qui démarra en Egypte. La deuxième pandémie a été décrite dans les steppes d'Asie centrale à partir de 1330. Elle a tué entre 17 et 28 millions de personnes de 1361 jusqu'à 1480. Elle a été appelée la « Grande Peste ». La troisième pandémie a pour origine la province de Yunnan, en Chine en 1855. Cette fois-ci la pandémie arriva jusqu'en Amérique et en Australie. Les 3 biovars sont distingués selon des critères biochimiques basés sur la fermentation du glycérol et la conversion du nitrate en nitrite. Antiqua a un biotype positif pour ces deux critères, Orientalis peut convertir le nitrate mais ne réalise pas la fermentation du glycérol, enfin Medievalis fermente le glyceérol mais ne peut pas convertir le nitrate. Les génomes (4,8Mb) de *Y. pestis* et *Y. pseudotuberculosis* sont très proches, celui de *Y. enterolitica* est plus éloigné des deux autres.

L'épidémie humaine démarre généralement par la forme bubonique à partir des piqûres de puces qui ont abandonné les rongeurs morts ou mourants. Il y a plusieurs forme de peste :

- la forme bubonique est la manifestation classique de la maladie. Deux à six jours après une piqûre de puce ou contamination avec du matériel infecté par des blessures ouvertes, des

ganglions grossissent et forment les bubons. Une bactériémie ou septicémie secondaire est courante chez les malades avec des bubons.

- la forme septicémique : des patients ayant le sang contaminé mais pas de bubons peuvent développer une septicémie. La forme septicémique ressemble aux septicémies causées par d'autres bactéries à gram négatif. La mortalité de cette forme est très élevée probablement parce que les antibiotiques utilisés contre les septicémies indéfinies ne sont pas efficaces contre *Y. pestis.*

- la forme pneumonique est rare mais elle est très contagieuse parce que la dissémination par voie respiratoire à partir d'un individu infecté est très rapide. Le taux de mortalité des patients qui développent une pneumonie secondaire est très élevé.

Lorsque une victime de la forme bubonique développe ensuite une pneumonie, les gouttelettes de l'air expiré deviennent contaminantes et une épidémie de la forme pneumonique se développe. En raison de la complexité du cycle biologique du parasite et du nombre et de la variété des animaux vecteurs, l'éradication de *Y. pestis* semble peu probable.

Les techniques utilisées actuellement sont le génotypage des séquences IS par PCR (Motin 2002) et depuis 2000, l'amplification par PCR de répétitions en tandem de type VNTRs (Adair 2000), (Le Flèche 2001), (Klevytska 2001).

Cette partie concernant *Y. pestis* a été volontairement moins développée que celles traitant de *P. aeruginosa* et *S. aureus*. En effet, elle correspond à un mini-projet de développement de marqueurs chez *Y. pestis* (collection de 5 souches) que j'ai mené à bien en début de thèse. Ce travail a été publié conjointement à une étude chez *B. anthracis* réalisée au laboratoire (Le Flèche 2001). Cet article est présenté dans la partie Résultats en tant que validation de l'approche MLVA et des fonctionnalités de la base de données développée au laboratoire. Cette approche a ensuite été développée plus largement chez *P. aeruginosa* et *S. aureus* dont les résultats constituent le cœur même de ce travail de thèse.

2 MATERIEL ET METHODES

2.1 Identification des répétitions en tandem

2.1.1 La base de données développée au laboratoire

Figure 9 : Page d'accueil de la base de données de répétitions en tandem**.**

Une base de données d'identification des répétitions en tandem a été développée au laboratoire par France Denoeud en utilisant le logiciel Tandem Repeats Finder (TRF) développé par Gary Benson (Benson 1999). Le TRF permet de détecter des séquences répétées en tandem même lorsque les motifs ne sont pas parfaitement conservés par rapport au consensus, et permet de trouver sans requête particulière, notamment sur la taille du motif

recherché, toutes les répétitions dans la séquence soumise, détectables par l'approche heuristique utilisée, ce qui représente un progrès considérable pour la recherche de minisatellites.

La base de données utilise le TRF pour rechercher des séquences répétées en tandem dans les génomes entièrement séquencés et accessibles à la communauté scientifique. Des requêtes peuvent être faites dans des séquences de génomes bactériens, de virus, d'archae, d'eucaryotes et enfin de plasmides et organelles. La Figure 9 montre la page d'accueil de la base.

La base de données développée au laboratoire permet à l'utilisateur de choisir un certain nombre de critères de recherche des séquences répétées (voir Figure 10) :

L : longueur du minisatellite

U : taille de l'unité répétée

N : nombre de répétition

V : conservation des motifs par rapport au motif consensus

Pos : position sur le génome (en kb)

%GC : pourcentage G+C

B : biais entre les brins

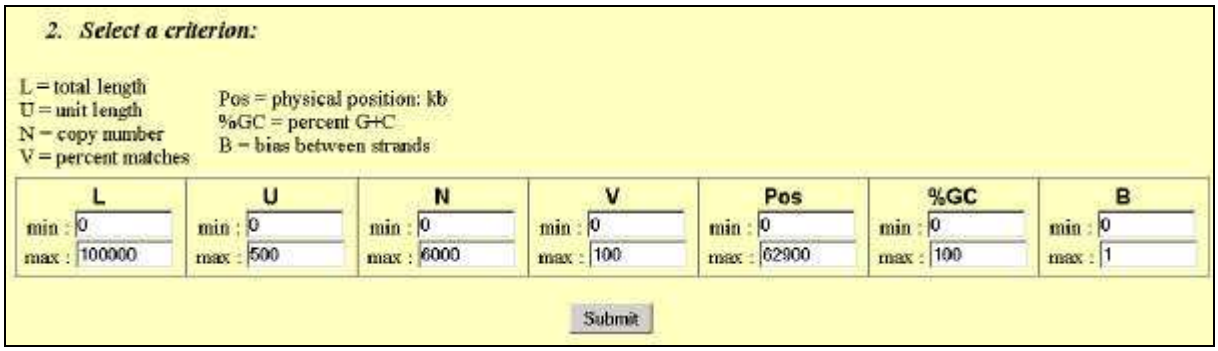


Figure 10 : Critères de recherche de répétitions en tandem dans la base de données.

La base de données fournit les résultats d'une requête sous forme de tableau avec les caractéristiques détaillées de chaque répétition en tandem ou bien sous forme d'histogramme représentant la distribution des répétitions dans le génome selon au choix : la position sur le génome, la longueur totale de la séquence répétée, la taille du motif, le nombre de répétitions ou encore le pourcentage en GC.

Le TRF fournit dans son fichier de sortie toutes les répétitions possibles identifiées dans la séquence soumise. En revanche, lorsqu'à un même locus, plusieurs tailles de motifs sont possibles, la base de données ne sélectionne qu'un seul résultat, celui correspondant à la répétition la plus longue (critère prioritaire) et ensuite selon le critère de conservation des

motifs. Pour chaque groupe de répétitions redondantes, deux valeurs sont mesurées : Lmax, la longueur maximum de la répétition (parmi les répétitions redondantes), et Mmax, le pourcentage maximum de conservation des motifs parmi les répétitions redondantes ayant une longueur totale supérieure à 80% de Lmax. Ensuite, parmi les répétitions ayant une longueur supérieure à 80% du Lmax et une conservation au moins égale à Mmax-0.1, la répétition avec le plus petit motif sera retenue (explication sur le lien suivant : http://iech5.igmors.u-psud.fr/ALIGNEMENTS/base_ms/overlapping.html).

Dans de rares cas, il faut soumettre la séquence au TRF pour voir toutes les répétitions possibles au locus étudié. En effet, il peut arriver que les variations de tailles d'allèles observées ne correspondent pas à la taille du motif proposé par la base. La base fournit, pour les génomes récemment importés, un lien vers les autres répétitions en tandem redondantes.

Les différentes fonctionnalités de la base seront décrites au fur et à mesure de leur utilisation au cours de mon travail de thèse.

2.1.2 Critères de recherche des répétitions en tandem chez *Y. pestis* et *P. aeruginosa*

Pour l'étude du génome de *Yersinia pestis*, (Parkhill 2001) nous avons recherché des répétitions en tandem uniquement dans le chromosome bactérien. Les critères de recherche sont U≥ 9pb et N≥ 7.

Ces critères ont été choisis de manière à ne pas sélectionner des motifs de moins de 9pb pour éviter des locus de type microsatellite souvent instables (Bayliss 2001) et donc sans valeur épidémiologique pour des études à plus long terme qu'une épidémie. Par ailleurs, la technique utilisée ici pour la séparation des produits PCR sur gel d'agarose imposait aussi une sélection de tailles de motifs qui puissent être résolues sur gel. Avec ces critères, on obtient 64 répétitions en tandem candidates. Dans un deuxième temps, un nouveau choix de répétitions a été fait selon les critères « U≥9pb, N= 6 et V≥80% ». Douze répétitions correspondent à ces critères, ce qui fait au total 76 locus étudiés chez *Y. pestis*. En effet, l'étude de la première série de répétitions a montré une corrélation entre le nombre d'allèles et le pourcentage de conservation des motifs, il s'agit donc pour cette espèce bactérienne d'un critère prédictif du polymorphisme des répétitions en tandem.

Pour l'étude des répétitions en tandem chez *Pseudomonas aeruginosa*, les mêmes critères ont été utilisés : U≥ 9pb et N≥ 7.

Les séquences des oligonucléotides ont été choisies dans les régions flanquantes des minisatellites avec le logiciel Primer3 (http://www.broad.mit.edu/cgi-bin/primer/primer3_www.cgi).

2.1.3 Comparaison de plusieurs génomes de même espèce : exemple de *S. aureus*

Pour les deux espèces bactériennes précédemment étudiées, nous ne disposions que d'un seul génome entièrement séquencé, ce qui a nécessité d'analyser expérimentalement un grand nombre de locus pour identifier des marqueurs polymorphes. Pour *Staphylococcus aureus* ce n'est pas le cas puisque 6 génomes complets sont disponibles (voir Tableau 5 du paragraphe 1.5.3.7) : Mu50, N315, MW2, NCTC8325, MRSA252 et MSSA476 (4 disponibles au début du projet). Actuellement 4 autres génomes sont en cours de séquençage, 3 par des entreprises privées et une par le TIGR (http://wit.integratedgenomics.com/GOLD/).

Une nouvelle fonctionnalité dans la base de données a été développée au laboratoire : la comparaison de plusieurs souches d'une même espèce bactérienne [http://minisatellites.u-psud.fr/comparison]. Cette page permet de sélectionner les répétitions en tandem polymorphes dans les souches comparées.

La Figure 11 présente la méthode utilisée pour comparer deux souches : la recherche par BLAST (Altschul 1997) des flanquantes des répétitions en tandem d'une souche dans le génome complet de l'autre souche, et réciproquement. La comparaison est faite de manière à identifier avec le TRF des répétitions en tandem d'un génome A puis de rechercher par BLAST dans le génome B les séquences homologues aux séquences des régions flanquantes des répétitions identifiées dans le premier génome. Bien entendu, lorsque dans le génome A il n'y a qu'un seul motif à un locus donné alors que plusieurs motifs ont été détectés dans le génome B, cette séquence n'est pas reconnue comme une séquence répétée.

Les raisons principales de la recherche par BLAST sont, d'une part, qu'il n'est pas possible d'utiliser la répétition elle-même pour trouver le locus homologue dans la deuxième souche avec un outil tel que BLAST : certains motifs répétés peuvent avoir des similitudes suffisantes pour générer des « matchs » entre des répétitions en tandem non apparentées. En effet, par exemple, l'hybridation d'une sonde synthétique constituée d'un motif élémentaire aléatoire répété en tandem sur des Southern Blots du génome humain (démarche qui peut s'apparenter au BLAST) identifie plusieurs locus polymorphes indépendants (Vergnaud 1989). D'autre part les BLAST doivent être effectués de façon réciproque afin d'éviter de manquer des répétitions en tandem qui n'auraient été détectées (avec les seuils utilisés lors du traitement par le TRF) que dans une des deux souches, l'autre contenant une répétition trop courte ou insuffisamment

conservée. Enfin, les positions des répétitions données par le TRF ne peuvent pas être utilisées pour calculer la taille totale de la répétition, sachant que le TRF ne trouve pas toujours le même démarrage du 1er motif répété d'un génome à l'autre.

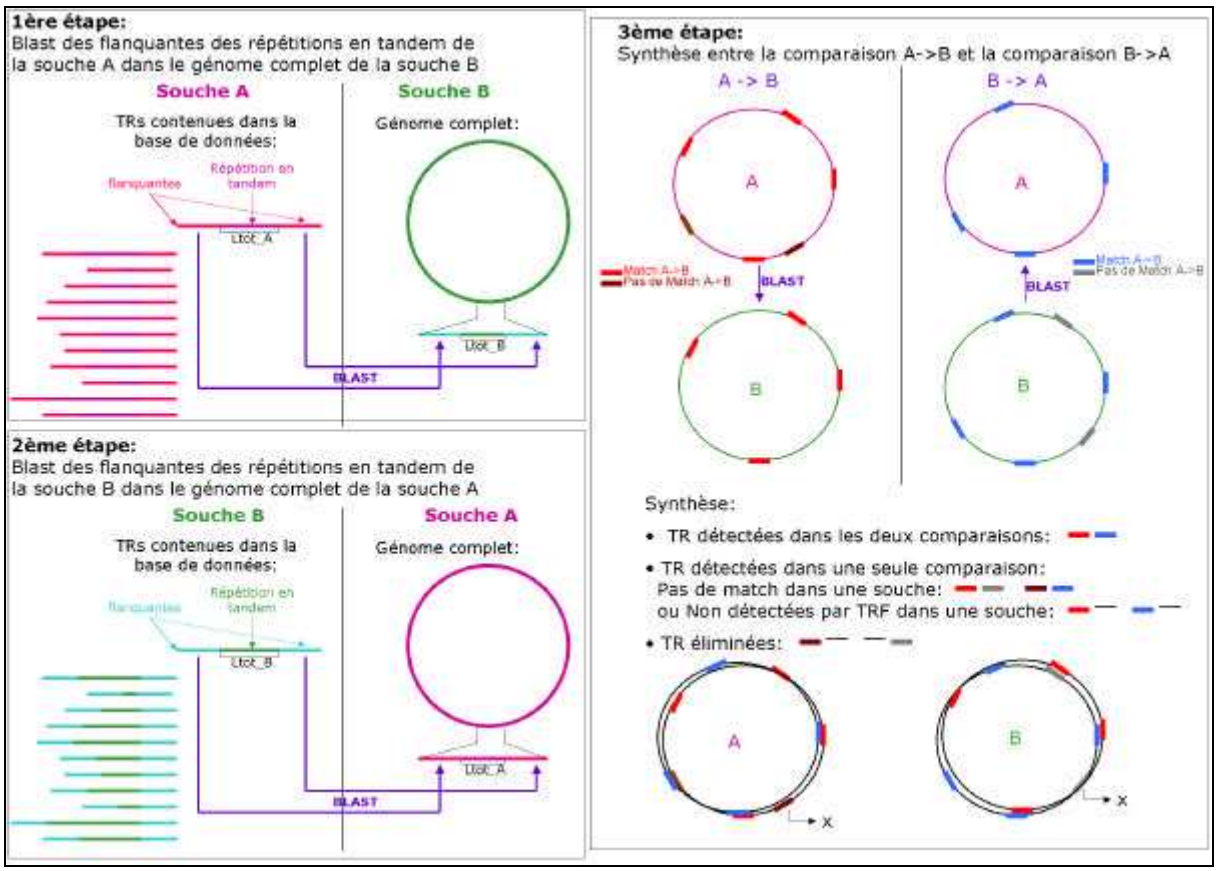


Figure 11 : Méthode de comparaison de souches bactériennes, basée sur le Blast des flanquantes des répétitions en tandem.

Lorsqu'il y a plus de 2 génomes à comparer, les comparaisons sont également faites de façon réciproque deux à deux (AB, BA, AC, CA, etc…) puis la synthèse est effectuée en utilisant comme référence les positions sur la souche A, commune à toutes les comparaisons de deux souches. Ensuite, la taille de la répétition en tandem, ainsi que le nombre d'allèles distincts sont calculés dans les différents génomes comparés. Les résultats des différentes comparaisons sont disponibles dans la base de données qui peut être interrogée via la page de comparaison de souches, accessible à l'adresse [http://minisatellites.u-psud.fr/comparison], et permet entre autre d'identifier les répétitions en tandem polymorphes, c'est à dire de longueur variable entre les souches comparées (Denoeud 2004). La Figure 12 présente la page d'accueil de la base pour les comparaisons de plus de deux génomes de même espèce.

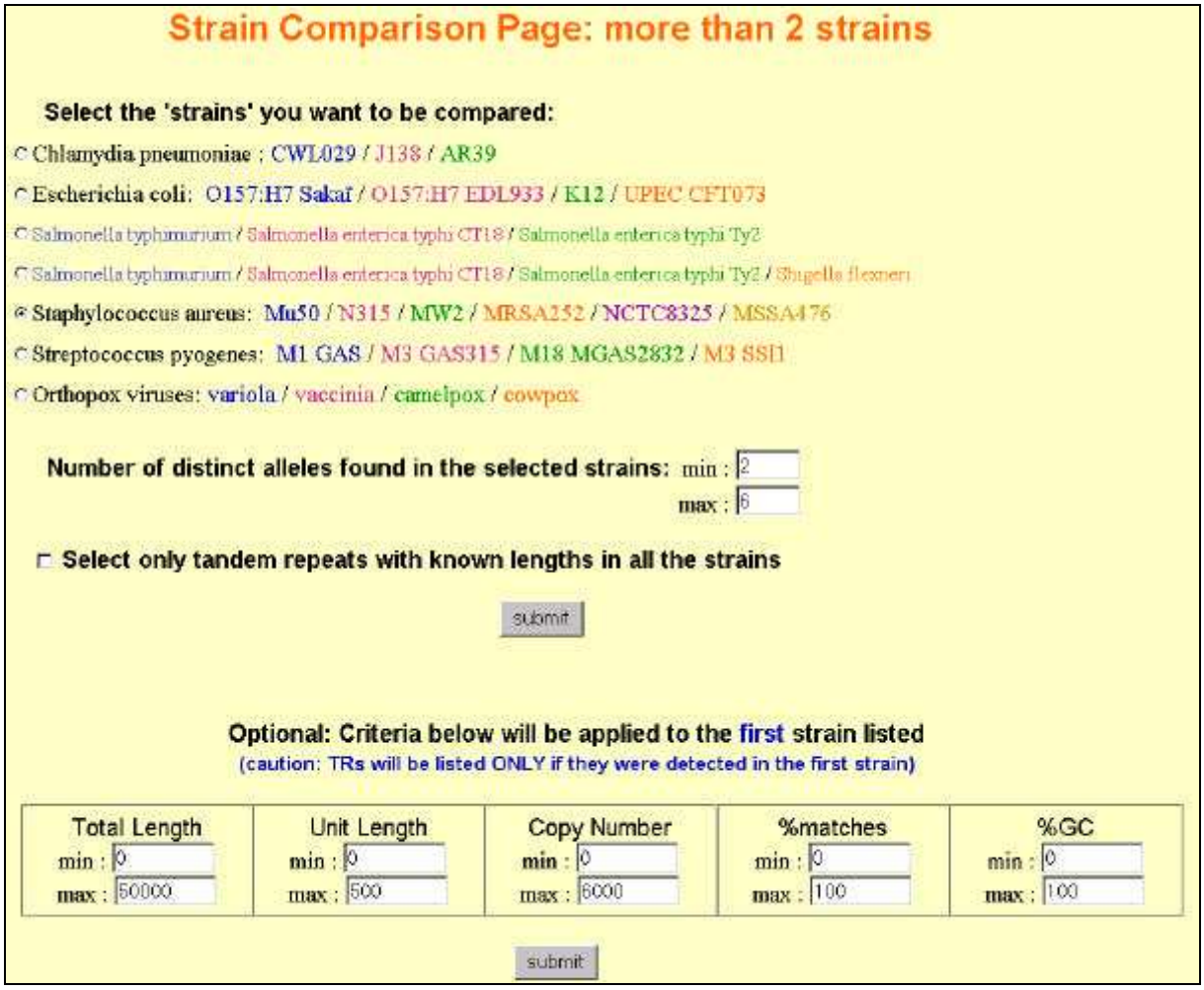


Figure 12 : Page de comparaison de plusieurs génomes de même espèce.

Un autre outil a été développé dans la base de données, il s'agit de la possibilité de faire des BLAST avec les séquences des amorces dans les génomes disponibles dans la base, et obtenir la taille du produit PCR attendu (voir Figure13).

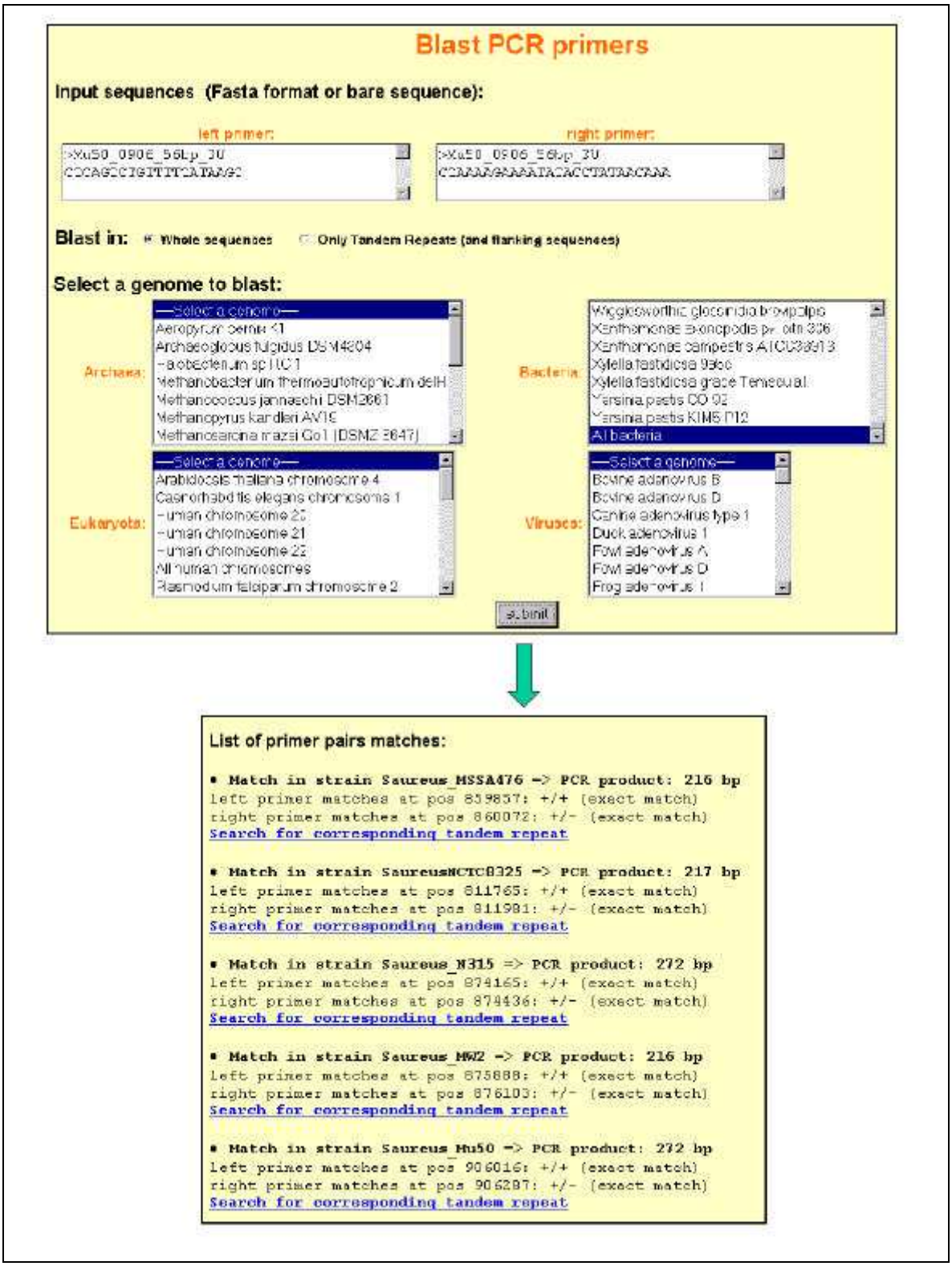


Figure 13 : Exemple de résultat de Blast des séquences des amorces.

2.2 Génotypage

La démarche globale du typage des répétitions en tandem est illustrée par la Figure 14.

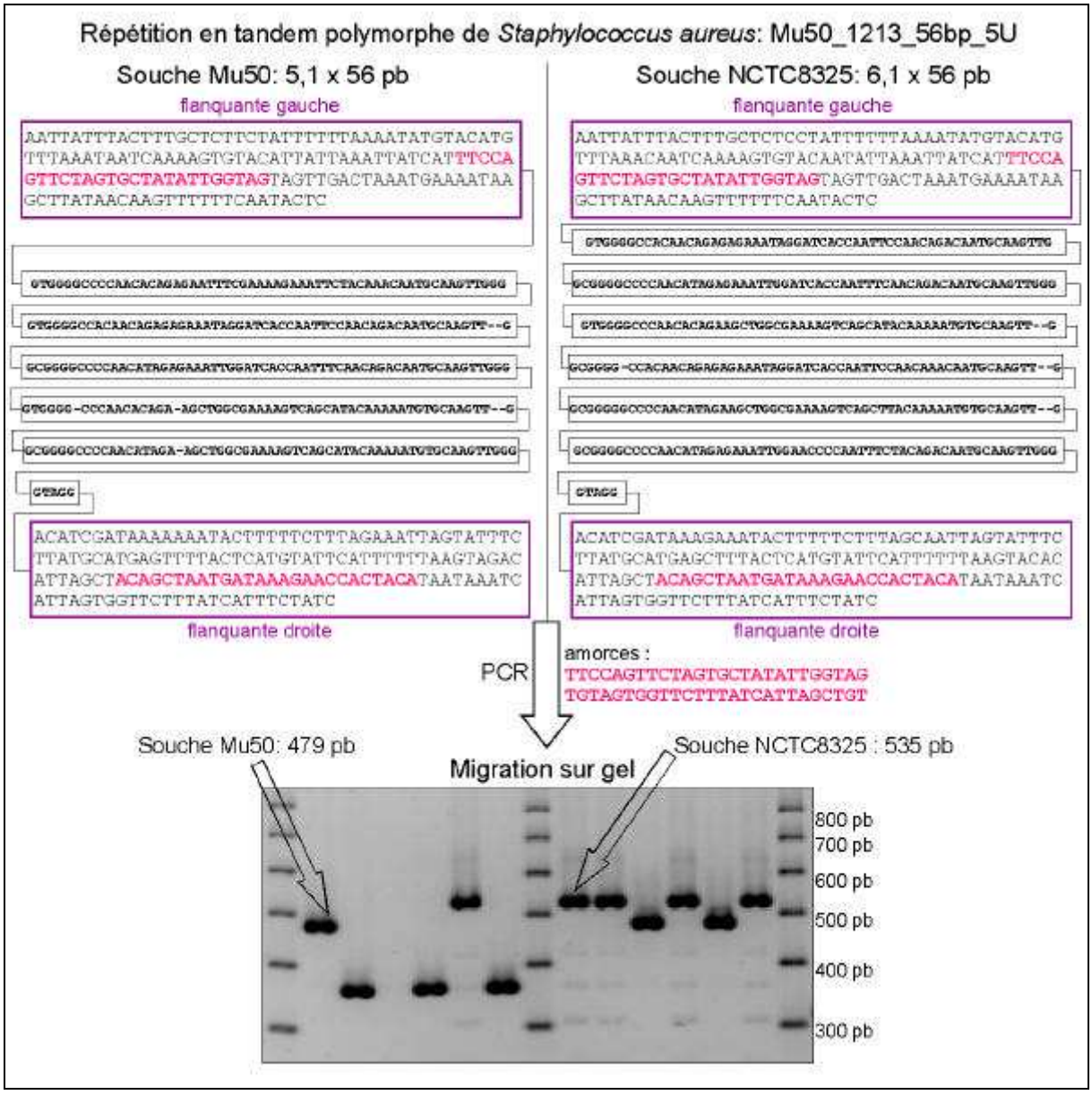


Figure 14 : Utilisation des répétitions en tandem polymorphe pour le génotypage de souches bactériennes.

2.2.1 Culture des souches et extraction d'ADN

Les 102 souches de *Pseudomonas aeruginosa* ont été fournies par le Dr Sylvain Brisse qui a étudié cette collection par ribotypage (Brisse 2000). Ces souches ont été collectées dans différents hôpitaux européens.

Les souches ont été réisolées sur boites LBA (Luria-Bertani Agar) (voir Annexe 1). Puis, une colonie de chaque souche a été mise en culture liquide en milieu LB (voir Annexe 1) à 37°C en aérobie pendant une nuit. Un millilitre de cette culture de nuit est centrifugé à 7500rpm pendant 10 minutes puis l'ADN est extrait à l'aide du kit QIAamp DNA Mini Kit selon le protocole recommandé par QIAgen (Qiagen, Hilden, Allemagne) sauf l'étape de lyse qui est plus longue, 5h à 55°C.

Pour le projet *Staphylococcus aureus*, nous disposions de 107 souches de la collection du centre National de Référence des Staphylocoques de l'Institut Pasteur Paris et de 30 souches de l'Hôpital d'Instruction des Armées du Val de Grâce. La liste des 137 souches et de leurs caractéristiques est détaillée en Annexe 2. Les souches de l'Institut Pasteur proviennent de différents hôpitaux français essentiellement, et aussi d'Espagne, de Tunisie et de Belgique. Elles ont fait l'objet d'analyse PFGE, mais les résultats ne sont pas comparables d'une série à l'autre puisque les génotypes n'ont pas été assignés selon la même nomenclature (des études indépendantes en PFGE sont difficile à comparer entre elles).

Les souches ont été cultivées à l'Institut Pasteur dans du milieu BHI (brain heart infusion). Un millilitre de culture de nuit a été prélevé pour l'extraction d'ADN. Avant d'utiliser le kit QIAamp DNA Mini kit, le culot bactérien est incubé pendant 30 minutes à 37°C avec 180µl d'un tampon de lyse (voir Annexe 1) et 20µl de lysostaphine à 1mg/ml. Une deuxième incubation a été réalisée avec 25µl de proteinase K (20mg/ml) et 200µl de tampon de lyse du kit. La préparation étant souvent très visqueuse, l'échantillon est ensuite chauffé 15 minutes à 95°C. Le protocole est ensuite suivi comme indiqué par le fournisseur. Puis, la mesure de concentration des ADN est réalisée avec le fluorimètre DyNA Quant™200 (Hoefer).

2.2.2 Amplification des répétitions en tandem par PCR

Pour le projet *P. aeruginosa,* un sous-groupe de 12 souches (voir Tableau 7) représentant 12 ribogroupes différents a été choisi pour tester les 201 répétitions en tandem.

Les répétitions en tandem identifiées comme polymorphes dans ce sous-groupe ont alors été testées sur la totalité de la collection de souches.

Pour le projet *S. aureus*, toutes les souches (137 au total) ont été testées pour les 33 répétitions en tandem polymorphes choisies parmi les 122 polymorphes dans au moins deux des six génomes comparés (Voir paragraphe 3.3.1.3). Cinq souches n'ont pas été amplifiées pour plusieurs des locus étudiés, c'est pourquoi elles ont été éliminées de l'analyse MLVA finale. Il s'agit des souches Saur025, Saur027 et Saur111 de l'Institut Pasteur et de Saur150 et Saur 154 du Val de Grâce.

Tableau 7 : Sous collection de 12 souches *P. aeruginosa*.

Souche:	RiboGroupe:	Ribotype:	Origine:
03D021	88-S-5	A	France
03D009	99-S-2	nd	France
04A036	99-S-4	O:16	France
05A400	87-S-3	O:12	France
15A178	147-S-3	O:7	Portugal
35C022	148-S-5	A6	Afrique du Sud
19A211	148-S-7	C10	Suisse
08D005	169-S-1	A3	Allemagne
03C001	169-S-4	A1	France
09A068	172-S-1	E15	Grèce
08A461	88-S-4	A8	Allemagne
01A105	88-S-6	O:4	Autriche

Protocole PCR :

Les réactions PCR sont réalisées dans un volume final de 15µl :

- 1ng d'ADN
- 1X de Tampon PCR 10X (Qbiogen)
- 1Unité de taq DNA polymerase (Qbiogen)
- 200µM de chaque dNTP
- 0.3µM de chaque primer

Le programme suivant a été utilisé pour tous les couples d'amorces :

- 96°C 5 minutes de dénaturation

suivi de 30 cycles :

- 96°C 30 sec de dénaturation
- 60°C 30 sec pour l'hybridation des primers
- 65°C 1minute pour la synthèse d'ADN

- extension finale 5 minutes à 65°C

Des mises au point ont été effectuées pour certains marqueurs en faisant des PCR avec un gradient de température de 55°Cà 65°C (12 températures testées) réalisé sur deux souches selon le programme suivant :

- 96°C 5 minutes de dénaturation

suivi de 30 cycles :

- 96°C 30 sec de dénaturation

- 12 températures testées, de 55°C à 65°C, 30 sec pour l'hybridation des amorces
- 65°C 1minute pour la synthèse d'ADN

- extension finale 70°C 5 minutes

Pour l'amplification des répétitions en tandem chez *P. aeruginosa*, les conditions de PCR spécifiques pour certains locus sont détaillées dans l'article (Onteniente 2003).

2.2.3 Séparation des produits de PCR sur gel d'agarose standard

Deux microlitres de bleu de dépôt sont ajoutés aux réactions de PCR (voir Annexe 1). Ensuite les échantillons sont déposés sur gel d'agarose à 1%, 2% ou 3% selon les tailles des produits à séparer. La migration est réalisée en TBE 0.5X (voir Annexe 1). Les conditions de migration des marqueurs chez *P. aeruginosa* sont décrites dans l'article (Onteniente 2003).

Les dépôts sont organisés de la manière suivante :

- 1 dépôt de marqueur de taille (DNA ladder BioRad 100pb ou 20pb)
- une souche de référence
- 5 échantillons à analyser

Ceci permet une analyse satisfaisante des gels à l'aide du logiciel BioNumerics (Applied Math). Le dépôt tous les 6 puits d'un marqueur de taille permet au logiciel BioNumerics de corriger correctement la courbure de l'image en cas de migration courbée des échantillons.

Ne disposant pas de la souche de référence PAO1, nous avons choisi une des 102 souches de la collection comme référence (05A400) et nous avons séquencé les allèles de cette souche pour 7 des 8 locus polymorphes.

L'Institut Pasteur nous a fourni les souches Mu50 et NTCT8325, qui ont servi d'ADN de référence.

2.2.4 Traitement des données

Les images de gels sont analysées avec le logiciel BioNumerics. Un script permet de convertir les tailles d'allèles en nombre de motifs. Elles sont alors importées dans une nouvelle base BioNumerics, pour faire une analyse « MLVA. » Les arbres sont faits avec les paramètres « categorical » et UPGMA.

Pour le calcul de l'index de polymorphisme, l'indice de Nei est utilisé :

Soit i l'indice sur les génotypes (ou classes distinguées), f_i la fréquence du génotype i, n_i le nombre de souches (individus) de génotype i, N le nombre total de souches

$f_i = n_i/N$

PIC (polymorphism information content) = $1 - \sum (f_i)^2$

Cet indice correspond à la probabilité de tomber sur deux souches de génotype différent dans une population décrite par l'échantillon testé. Ceci suppose que l'échantillon testé soit représentatif de la population globale.

2.3 Test de stabilité des répétitions en tandem polymorphes chez *P. aeruginosa*

Pour évaluer l'intérêt épidémiologique des marqueurs polymorphes identifiés chez *Pseudomonas aeruginosa*, une étude de la stabilité (à court terme) de ces marqueurs a été réalisée. Plusieurs souches ont été mises en culture et diluées quotidiennement dans un nouveau milieu pendant 3 semaines et les marqueurs polymorphes ont été étudiés après un certain nombre de générations. Ce même type d'étude a été réalisé par exemple pour des marqueurs chez *Staphylococcus aureus* (Shopsin 2000).

2.3.1 Courbe de croissance des 6 souches testées

Le temps de génération des 6 souches étudiées (12A241, 18E049, 22D032, 03D021, 04A036 et 05A400 la souche de référence) a été déterminé en faisant une courbe de croissance. A partir d'un ré-isolement sur boite, les 6 souches sont mises en culture dans du milieu LB à 37°C sous agitation, puis la densité optique à 600nm (DO600) est lue toutes les 30 minutes sur une durée totale de 10 heures.

2.3.2 Dilutions en série des cultures bactériennes

A partir des 6 cultures de nuit ayant servi à faire les courbes de croissance, des dilutions en série sont réalisées. La Densité Optique à 600nm (DO600) est lue toutes les 12 heures et à chaque fois, un ensemencement réalisé à une dilution de $1/1000^{ème}$. Un tube de milieu de culture témoin est fait à chaque dilution pour contrôler d'éventuelles contaminations. Le milieu est utilisé sans aucun antibiotique. Après les 10 premiers jours de dilutions, un étalement contrôle est réalisé sur boite LBA pour vérifier que la culture correspond bien à une culture homogène de *Pseudomonas aeruginosa.* Ceci est répété à la fin des 3 semaines. Des

aliquotes des 6 cultures initiales ont été gardés à –80°C, ainsi que tous les prélèvements quotidiens.

2.3.3 Typage des souches

Le protocole d'extraction d'ADN utilisé est celui décrit dans la partie 2.2.1. On a extrait les ADN des 6 souches correspondant au début de la série de dilution (T0) et les ADN de ces 6 souches après 23 dilutions (réalisées sur 21 jours). Les PCR sont réalisées pour les 8 minisatellites polymorphes chez *P. aeruginosa* selon les conditions décrites dans l'article (Onteniente 2003).

2.4 Séquençage d'allèles chez *P. aeruginosa* et *S. aureus*

Pour ces projets, certaines répétitions en tandem ont été séquencées dans l'objectif de trouver un locus qui, séquencé, serait aussi résolutif qu'une étude MLVA. Les produits de PCR ont été purifiés par les deux méthodes suivantes :

2.4.1.1 Précipitation au PEG (Poly Ethylène Glycol) des produits de PCR (de plus de 300pb)

Les produits de PCR sont précipités de la façon suivante (Embley 1991) :

Les PCR sont réalisées dans un volume final de 60 µl. On ajoute 0,6 volume de PEG8000 20% (p/v) NaCl2,5M. Les produits sont incubés 10 minutes à 37°C (Thermomixer comfort Eppendorf) puis centrifugés 10 minutes à 13000rpm. Le surnageant est éliminé. On ajoute ensuite 500µl d'éthanol 80% puis on centrifuge 10 minutes à 13000rpm. On vide le surnageant et on sèche les culots. Ceux-ci sont ensuite repris dans 20µl d'eau, et 2µl sont déposés sur gel d'agarose 1% pour évaluer la quantité de produit PCR purifié. On laisse évaporer le volume nécessaire pour avoir 20ng d'ADN/ 100pb à séquencer. Les produits PCR d'une taille inférieure à 300pb ne sont pas bien précipités au PEG, c'est pourquoi une autre technique est utilisée pour les petits fragments.

2.4.1.2 Traitement à l'ExoSAP-IT™ des produits de PCR (de moins de 300pb)

L'ExoSAP-IT™ contient deux enzymes thermolabiles : l'Exonucléase I qui dégrade les ADN simples brins et la phosphatase alcaline de crevette (SAP) qui hydrolyse les dNTPs libres.

Les produits PCR sont préalablement précipités à l'éthanol. On ajoute aux 60µl de réaction de PCR, 2.5µl de NaCl 5M puis 120 µl d'éthanol 100%. Les tubes sont placés deux heures à –20°C puis centrifugés à 12000rpm pendant 20 minutes à 4°C. Le surnageant est éliminé et le culot séché. Celui-ci est repris dans 5µl d'eau. On ajoute 2µl d'ExoSAP-IT (USB, Cleveland, Ohio). Puis on incube les échantillons 10 minutes à 37°C suivies de 10 minutes à 80°C pour inactiver les deux enzymes. On laisse évaporer les échantillons. Les produits de PCR purifiés sont ensuite envoyés à séquencer (MWG-Biotech).

2.4.2 Traitement des données

Les séquences sont récupérées au format .scf puis soumise au logiciel Phred qui analyse la qualité de chaque trace pour nommer les bases. Ensuite, le logiciel Phrap permet l'assemblage des différentes lectures et fournit les indices de qualité du contig obtenu.

Les séquences de même longueur sont ensuite alignées avec ClustalW et sont soumises au TRF. Un script perl a été écrit pour automatiser le codage des allèles en lettres. Un fichier dictionnaire.txt contient tous les motifs différents rencontrés pour un locus donné dans la souche de référence, avec pour chaque motif une lettre qui lui est assignée. Ensuite, on soumet le fichier contenant les séquences des allèles au script qui va assigner pour chaque motif reconnu une lettre. Les nouveaux motifs par rapport à ceux présents dans le dictionnaire initial sont signalés par une étoile (*) et il est alors possible d'enrichir le dictionnaire de ces nouveaux motifs. Ceci constitue l'étape préliminaire d'analyse des motifs des répétitions en tandem séquencés. Notre système de codage nous permet de déterminer dans la séquence à traiter le point de démarrage du premier motif.

La TRDB (Tandem Repeats Data Base, http://tandem.bu.edu/cgi-bin/trdb/trdb.exe) est la base de données de répétitions en tandem développée par Gary Benson. Une de ses fonctionnalités permet le codage des allèles en lettres. Cependant, le problème majeur vient du fait que le TRF ne démarre pas toutes les répétitions d'un locus à la même position, ce qui a pour conséquence d'observer plusieurs consensus différents (d'une souche à l'autre) pour le même locus. Le codage multiplie le nombre de motifs différents par rapport à la réalité.

Une autre fonctionnalité intéressante est la possibilité de visualiser les motifs alignés par rapport au consensus. Les mutations par rapport au consensus sont en couleur et permettent de comparer « à l'œil » assez facilement les allèles (une illustration sera présentée dans la suite).

3 RESULTATS

3.1 Développement de marqueurs polymorphes chez des espèces pathogènes d'émergence récente : *Yersinia pestis* et *Bacillus anthracis*

Ce premier article (Le Flèche 2001) est intitulé " A tandem repeats database for bacterial genomes : application to the genotyping of *Yersinia pestis* and *Bacillus anthracis* " (une base de données des répétitions en tandem des génomes bactériens : application au génotypage de *Yersinia pestis* et *Bacillus anthracis*). Le travail est orienté sur l'étude des bactéries considérées comme étant des menaces " bioterroristes " et entre dans le cadre plus large de la lutte contre le risque biologique provoqué. D'ailleurs, quelques mois après la parution de l'article, aux Etats-Unis, des enveloppes contenant des spores de *Bacillus anthracis* ont été envoyées par courrier (Jernigan 2002). L'enquête qui a suivi a eu pour but de trouver l'origine des spores utilisées : les différentes enveloppes contenaient-elles la même souche, et quelle souche? L'approche utilisée est identique à celle décrite ici, c'est à dire le typage de répétitions en tandem, et elle a fourni l'essentiel des réponses attendues en quelques jours. Le laboratoire qui a réalisé le travail aux Etats-Unis a comparé le génotype obtenu à un fichier préalablement constitué grâce à l'analyse de souches provenant du monde entier.

Résumé :

Contexte : Certaines espèces de bactéries pathogènes sont génétiquement très homogènes, ce qui rend difficile la distinction entre les souches. Ces dernières années, les répétitions en tandem se sont révélées être des marqueurs de choix pour le génotypage d'un certain nombre de pathogènes. La variabilité de ces structures semble contribuer à la flexibilité phénotypique de certaines bactéries pathogènes. La disponibilité de séquences de génomes entiers a ouvert la voie à l'évaluation systématique de la diversité des répétitions en tandem et à leur utilisation pour des études épidémiologiques.

Résultats : Cet article présente une base de données (http://minisatellites.u-psud.fr) des répétitions en tandem de génomes bactériens d'accès public, qui facilite l'identification et la sélection des répétitions en tandem. Nous illustrons son utilisation par la caractérisation de minisatellites de deux pathogènes humains importants, *Yersinia pestis* et *Bacillus anthracis*. Afin d'éviter les locus de contingence, qui sont probablement de faible valeur en tant que marqueurs épidémiologiques, et de proposer des outils de génotypage exploitables par électrophorèse sur des gels d'agarose classiques, seules les répétitions en tandem d'unité répétée d'au moins 9 pb ont été évaluées. *Yersinia pestis* contient 64 minisatellites de ce type,

dans lesquels l'unité est répétée au moins 7 fois. Un lot de 12 locus supplémentaires, contenant 6 unités répétées et ayant une forte conservation interne a également été testé. Quarante-neuf locus sont polymorphes parmi les 5 souches de Yersinia utilisées (dont 25 parmi les 3 souches de *Yersinia pestis*). *Bacillus anthracis* contient 30 structures comparables, dans lesquelles l'unité est répétée au moins 10 fois. La moitié de ces répétitions en tandem est polymorphe parmi les souches testées.

Conclusions: L'analyse des séquences des génomes bactériens actuellement disponibles montre que *Bacillus anthracis* et *Yersinia pestis* ont une densité intermédiaire en répétitions en tandem de plus de 100 pb (environ 30 par Mb) par rapport aux autres génomes bactériens analysés jusqu'à présent. Dans les deux cas, tester le polymorphisme d'une fraction seulement de ces séquences a suffi pour développer rapidement un lot de plus de 15 marqueurs informatifs, certains montrant un très fort degré de polymorphisme. Par exemple, pour le marqueur BAMS7 de *Bacillus anthracis*, l'indice de polymorphisme (PIC) atteint 0,82, avec des allèles couvrant une large plage de tailles (600 à 1950 pb), et 9 allèles sont distingués parmi le nombre restreint de souches typées dans cette étude.

L’analyse MLVA est la technique adoptée pour identifier les souches de *Y. pestis* et de *B. anthracis*. Cette technique a été validée depuis le premier article sur un plus grand nombre de souches.

Research article

A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*

Philippe Le Flèche[1,2], Yolande Hauck[2], Lucie Onteniente[2], Agnès Prieur[1,2], France Denoeud[2], Vincent Ramisse[1], Patricia Sylvestre[1], Gary Benson[3], Françoise Ramisse[1] and Gilles Vergnaud*[1,2]

Address: [1]Centre d'Etudes du Bouchet, BP3, 91710 Vert le Petit, France, [2]Génomes et Minisatellites, Institut de Génétique et Microbiologie, Bat 400, Université Paris XI, 91405 Orsay cedex, France and [3]Department of Biomathematical Sciences, Box 1023, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, USA

E-mail: Philippe Le Flèche - lefleche@igmors.u-psud.fr; Yolande Hauck - Yolande.Hauck@igmors.u-psud.fr; Lucie Onteniente - Lucie.Onteniente@igmors.u-psud.fr; Agnès Prieur - Agnes.Prieur@igmors.u-psud.fr; France Denoeud - France.Denoeud@igmors.u-psud.fr; Vincent Ramisse - Vincent.Ramisse@ceb.etca.fr; Patricia Sylvestre - psylvest@pasteur.fr; Gary Benson - benson@ecology.biomath.mssm.edu; Françoise Ramisse - f.ramisse@freesurf.fr; Gilles Vergnaud* - Gilles.Vergnaud@igmors.u-psud.fr
*Corresponding author

Published: 30 March 2001

BMC Microbiology 2001, **1**:2

Received: 19 February 2001
Accepted: 30 March 2001

This article is available from: http://www.biomedcentral.com/1471-2180/1/2

Abstract

Background: Some pathogenic bacteria are genetically very homogeneous, making strain discrimination difficult. In the last few years, tandem repeats have been increasingly recognized as markers of choice for genotyping a number of pathogens. The rapid evolution of these structures appears to contribute to the phenotypic flexibility of pathogens. The availability of whole-genome sequences has opened the way to the systematic evaluation of tandem repeats diversity and application to epidemiological studies.

Results: This report presents a database ([http://minisatellites.u-psud.fr]) of tandem repeats from publicly available bacterial genomes which facilitates the identification and selection of tandem repeats. We illustrate the use of this database by the characterization of minisatellites from two important human pathogens, *Yersinia pestis* and *Bacillus anthracis*. In order to avoid simple sequence contingency loci which may be of limited value as epidemiological markers, and to provide genotyping tools amenable to ordinary agarose gel electrophoresis, only tandem repeats with repeat units at least 9 bp long were evaluated. *Yersinia pestis* contains 64 such minisatellites in which the unit is repeated at least 7 times. An additional collection of 12 loci with at least 6 units, and a high internal conservation were also evaluated. Forty-nine are polymorphic among five *Yersinia* strains (twenty-five among three *Y. pestis* strains). *Bacillus anthracis* contains 30 comparable structures in which the unit is repeated at least 10 times. Half of these tandem repeats show polymorphism among the strains tested.

Conclusions: Analysis of the currently available bacterial genome sequences classifies *Bacillus anthracis* and *Yersinia pestis* as having an average (approximately 30 per Mb) density of tandem repeat arrays longer than 100 bp when compared to the other bacterial genomes analysed to date. In both cases, testing a fraction of these sequences for polymorphism was sufficient to quickly develop a set of more than fifteen informative markers, some of which show a very high degree of polymorphism. In one instance, the polymorphism information content index reaches 0.82 with allele length covering a wide size range (600-1950 bp), and nine alleles resolved in the small number of independent *Bacillus anthracis* strains typed here.

Background

The polymorphism associated with tandem repeats has been instrumental in mammalian genetics for the construction of genetic maps and still is the basis of DNA fingerprinting in forensic applications. Tandem repeats are usually classified among satellites (spanning megabases of DNA, associated with heterochromatin), minisatellites (repeat units in the range 6-100 bp, spanning hundreds of base-pairs) and microsatellites (repeat units in the range 1-5 bp, spanning a few tens of nucleotides).

More recently, a number of studies have supported the notion that tandem repeats reminiscent of mini and microsatellites are likely to be a highly significant source of very informative markers for the identification of pathogenic bacteria even when these pathogens are recently emerged, highly monomorphic species [1–5]. This probably reflects the important contribution of tandem repeats to the adaptation of the pathogen to its host. Tandem repeats appear to contribute to phenotypic variation in bacteria in at least two ways. Tandem repeats located within the regulatory region of a gene can constitute an on/off switch of gene expression at the transcriptional level [6,7]. Similarly, tandem repeats within coding regions with repeat units length not a multiple of three can induce a reversible premature end of translation when a mutation changes the number of repeats (reviewed in [8–10]). In other instances, the repeated unit length is a multiple of three, and the tandem repeat contributes to a coding region. In such cases, variations in the number of copies modify the gene product itself [11].

Mutation mechanisms of micro and minisatellites have been studied in some detail in eukaryotes, essentially human and yeast (reviewed in [12]). In brief, the data obtained so far suggest that microsatellites mutate by replication slippage processes; mutation rates depend upon the efficiency of mismatch repair mechanisms and an internal heterogeneity within the array strongly stabilizes the tandem repeat. In contrast, minisatellites mutate predominantly as the result of the repair of a double strand break initiated within, or very close to, the tandem repeat. In eukaryotes at least, these events can be of replicative origin [13], or can be genetically controlled, and specifically induced, during meiosis, at double strand breaks hot-spots. Minisatellite mutation rate in eukaryotes appears to be insensitive to mismatch repair efficiency, and internal heterogeneity is compatible with a high mutation rate [12, 14].

In bacteria, loci containing a tandem repeat from the microsatellite class (repeat unit sizes of 1-8 bp) have been called simple sequence contingency loci [8]. Altered number of repeats allows for reversible on and off states of expression for the corresponding gene. The mutation rate of a tetranucleotide (microsatellite) tract in *Haemophilus influenzae* is higher than 10^{-4} and contributes to the adaptation of the pathogen to its hosts as the infection progresses [15]. In such an extreme situation, the microsatellite is of limited value for strain identification, epidemiological and phylogenetic studies. The tandem repeat array is composed of perfect copies of the elementary unit, and different alleles are observed in a single culture. In contrast, the phylogenetic identity of minisatellite alleles of identical size can usually be further checked by DNA sequencing, since the repeated units are often not perfect [16]. The pattern of variants along the array provides an additional level of allele identification and phylogenetic information. In addition, tandem repeats with longer repeat unit length can be relatively easily typed in the size range of a few hundred base-pairs using ordinary horizontal gel electrophoresis.

In this report, we will first describe the use of a tandem repeats database for bacterial genomes ([http://minisatellites.u-psud.fr]) and briefly compare the general characteristics of tandem repeats in a number of bacterial genomes for which the sequence has been determined and made publicly available. We will then show how this tool can easily be applied to the rapid characterization of new highly polymorphic markers in two pathogens, *Y. pestis* and *B. anthracis*.

Both *Y. pestis* (causative agent of plague) and *B. anthracis* (causative agent of anthrax) are recently emerged clones of respectively Y. pseudotuberculosis [17] and *B. cereus* [18]. In the case of *Y. pestis*, a high resolution typing tool based on RFLP (Restriction Fragment Length Polymorphism) analysis of IS100 locations has already been developed [17]. However this technology is more demanding than PCR typing, which justifies the development of such an assay. In the case of *B. anthracis*, polymorphisms were initially identified essentially using AFLP (Amplified Fragment Length Polymorphism) typing [19]. Subsequent analyses demonstrated that the most informative fragments in AFLP patterns resulted from tandem repeat array length variations (five minisatellite loci were characterized in this way [2]).

Results and discussion

Use of the tandem repeats database

To date, 36 bacterial genome sequences from 32 species have been released in the public domain and are included in the database (Figure 1A; the nine archaebacteria genomes sequenced to date are presented in an other page, which can be accessed from [http://minisatellites.u-psud.fr/]). As many other sequencing projects are under way ([http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/bact.html] ; [http://www.tigr.org/tdb/mdb/

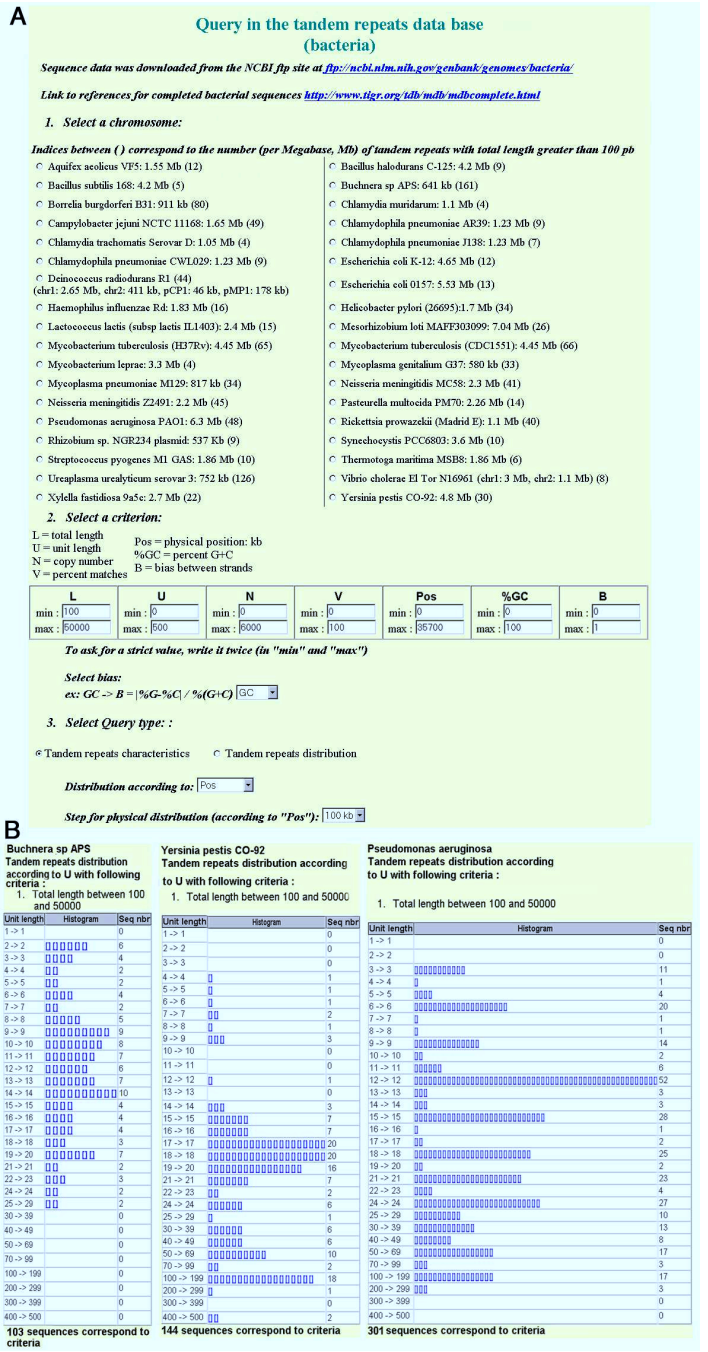


Figure 1

Querying the tandem repeats database 1A: bacterial tandem repeats main page Bacteria species are listed in alphabetical order. The name of the strain used for sequencing is indicated after the species name and before the genome size (expressed in megabase). The rightmost figure indicates the density (per Mb) of tandem repeat arrays longer than 100 bp. The search for tandem repeats can be restricted according to a combination of criteria, including total array length (L), repeat unit length (U), number of repeats (N), internal conservation of the repeats (V), position (expressed in kilobase) on the genome (Pos), GC content of the array (%GC), strand bias (B). Three different biases can be evaluated, GC bias, AT bias and Purine-Pyrimidine bias. The bias reflects strand asymmetry of the repeat sequence. The search output can either present a list of characteristics of the tandem repeats fulfilling criteria, ordered according to their position on the genome, or classify the tandem repeats according to a selected structural parameter. 1B: examples of queries in three genomes All tandem repeat arrays spanning more than 100 base-pairs are classified according to repeat unit length. The query was run on *Buchnera sp.* (left panel), *Yersinia pestis* (middle panel) and *Pseudomonas aeruginosa* (right panel).

mdbinprogress.html] ; [http://www.sanger.ac.uk/Projects/Microbes/]), the database will be regularly updated. The collection of tandem repeats present in a given genome can be queried according to a combination of criteria, total tandem repeat array length (L), repeat unit length (U), number of repeats (N), percentage of conservation of the repeats along the array (V), position on the genome (Pos), average GC percent of the repeats (%GC), strand bias in nucleotide composition (B) (these values have been precomputed using the Tandem Repeats Finder software described in [20]). The results shown on Figure 1B use the "Tandem Repeats Distribution according to repeat unit length" option (Figure 1A). Three genomes were searched for tandem repeat arrays longer than 100 base-pairs (L ≥ 100). The genomes selected illustrate three different behaviors. On the right panel, *Pseudomonas aeruginosa* shows a very striking bias towards minisatellites with a motif length multiple of three. On the left and middle panels of Figure 1B, B*uchnera sp* and Y. *pestis*, show no such bias. The overall density of tandem repeat arrays longer than 100 base-pairs varies in the different genomes. *Buchnera sp.* contains 103 such loci, for a total genome size of 641 kb, which corresponds to a density per megabase of 161. *Pseudomonas aeruginosa*, with a total genome length of 6.3 Mb, has a density of 48. *Y. pestis* has an intermediate value of 30. Figure 2 summarizes the values observed in the 32 species. Ten non pathogenic species are presented in the upper part, 22 pathogenic species on the lower part. The species are ordered from top to bottom according to increasing genome size. The dark bars indicate for each genome the density per megabase of tandem repeat arrays longer than 100 bp. The clear bars reflect the excess of tandem repeats with unit length a multiple of three. A wide range of situations is observed, with a remarkable excess of tandem repeats multiples of three in *Mycobacterium tuberculosis* and *Pseudomonas aeruginosa*, presumably reflecting a significant contribution of tandem repeats to coding regions in these two bacteria.

As a quick illustration of the use of this database to facilitate the development of genotyping tools for bacterial genomes, we have evaluated the polymorphism associated with tandem repeats from *Y. pestis* on one hand and *B. anthracis* on the other (in this second instance, the genome sequence has not been completed yet and does not appear on the publicly accessible Tandem Repeats Database page, Figure 1A).

Application to Y. pestis

Figure 3A presents the result of a query run on *Y. pestis*, to identify tandem repeats with repeat units longer than 9 base-pairs repeated at least 7 times in the strain which has been sequenced (CO-92 biovar Orientalis).Sixty-four tandem repeats fulfill these criteria (an additional group

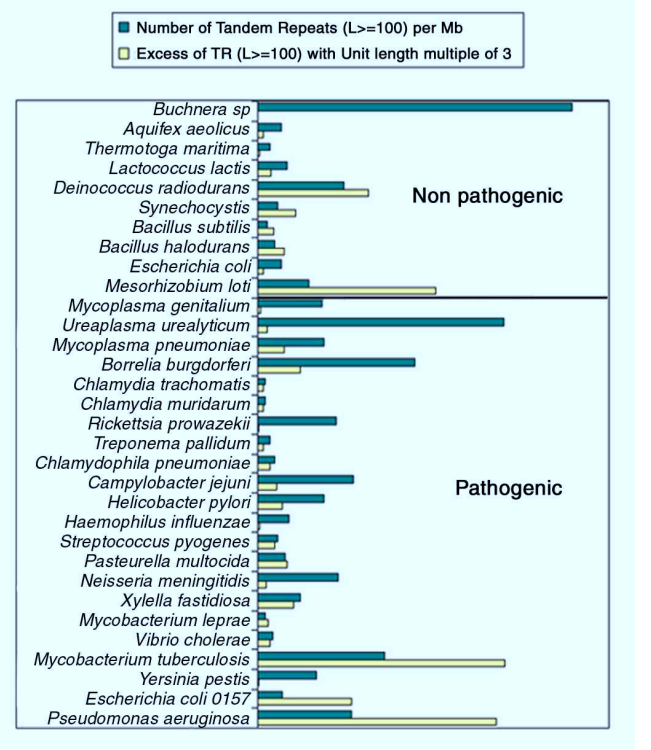


Figure 2
Relative frequency of tandem repeats within bacterial genomes The ten non-pathogen species are listed on top. Within each category, species are ordered according to genome size (smallest genome on top). The density of tandem repeat arrays longer than 100 bp is plotted for each species (dark bars). The clear bars reflect the excess (χ^2 values) of tandem repeats with a repeat unit length multiple of three.

of forty-nine have 6 copies of the motif; the twelve loci with the highest internal conservation were also included in this study). The output includes links to individual alignment files, as produced by the Tandem Repeat Finder software [20]. The alignment file also includes 200 base-pairs of flanking sequence from each side of the tandem repeat, from which primers can be selected for PCR amplification. Figure 3B shows an annotated extract of one alignment file. The positions of the primers selected for subsequent PCR amplification are underlined. Three *Y. pestis* (representing the Antiqua, Medievalis, and Orientalis biovars [17]) and two *Y. pseudotuberculosis* strains were used for the initial identification of minisatellites sufficiently polymorphic to be of interest for further studies. Table 1 summarizes the PCR conditions used for each polymorphic locus and the results obtained. A total of 76 tandem repeats were tested. PCR amplification failed in 6 cases. Twenty one loci are monomorphic in the five *Yersinia* strains typed here. Forty-nine of the loci are polymorphic (Table 1). Twenty-five of these are polymorphic among the *Y. pestis* strains.

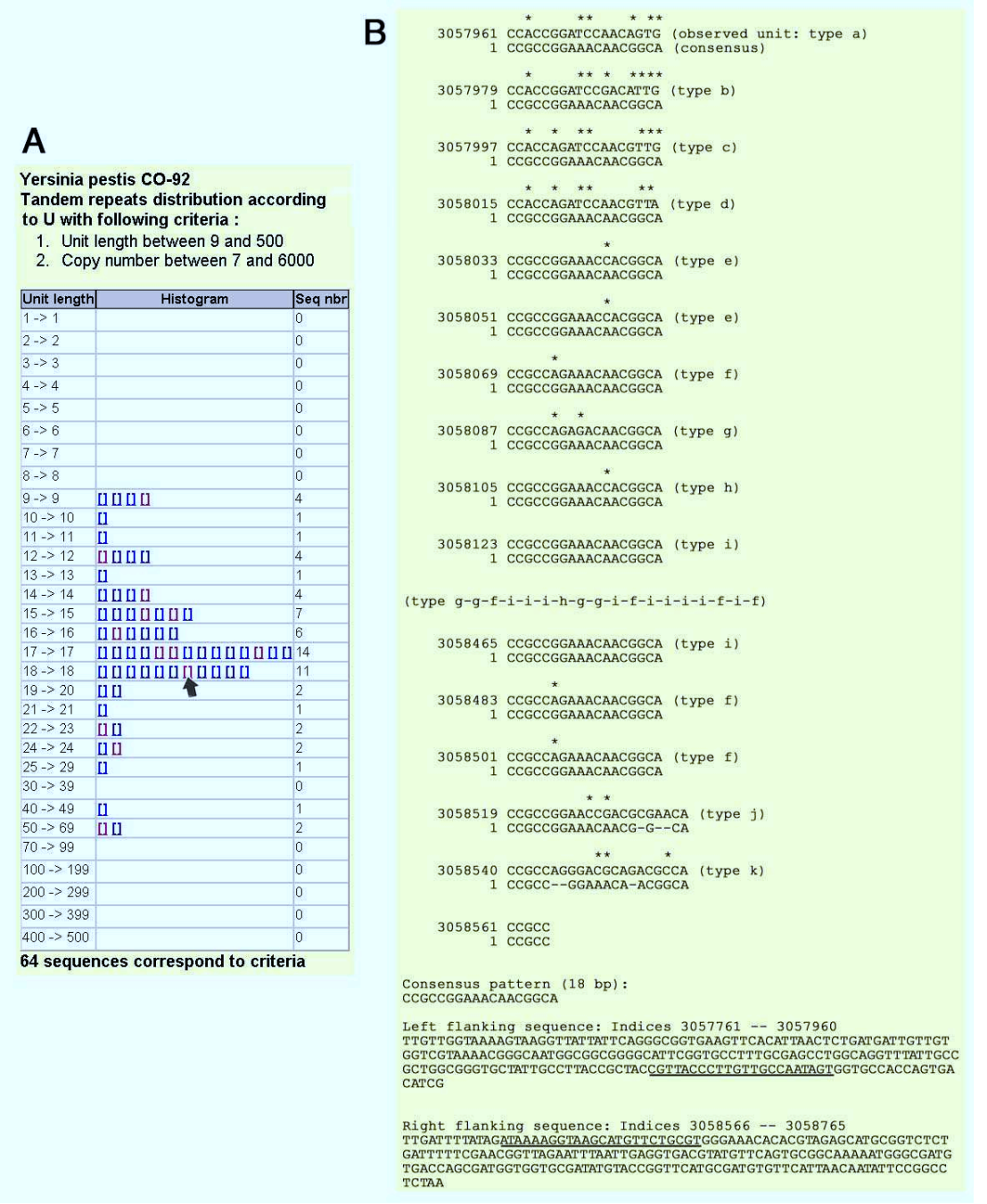


Figure 3
Selection procedure of minisatellites for *Y. pestis* 3A: Sixty-four tandem repeats have at least 7 units longer than 9 base-pairs. Panel A presents the distribution of these 64 loci according to repeat unit length. Each rectangle is an hyperlink to an alignment file. The rectangle indicated by the arrow is linked to the file illustrated in panel B. 3B: This is an annotated alignment file. The file corresponds to Yp3057ms09 (Table 1 and Figure 4; Yp : *Yersinia pestis*; 3057 : position on the genome, expressed in kilobases; MS09 : MiniSatellite index). The consensus pattern of 18 base-pairs is aligned to each motif. Annotations of the file are inserted within brackets. Although this minisatellite is very polymorphic, eleven different motifs (labeled a-k) are observed in the sequenced allele. The first four and last two copies are most diverged and rare. Four types of motifs (f, g, h, i) constitute most of the array. For convenience, 18 motifs have been removed from the alignment file and replaced by their letter code. The last two copies are 21 base-pair long instead of 18. The end of the alignment file (panel B, bottom) provides sequence data flanking the tandem repeat array. The positions of the primers chosen for PCR amplification of this locus (Table 1) are shown underlined.

Seven present a different allele in each of the five *Yersinia* strains, thirteen have a different allele in each of the three *Y. pestis* strains. Gel images for the 25 loci polymorphic among *Y. pestis* are shown in Figure 4. As can be seen, the repeat unit size and the overall length of the PCR products are such that tandem repeats differing by a single repeat unit can be distinguished by simple agarose gel electrophoresis.

Application to B. anthracis

Given the relatively low overall size of most bacterial tandem repeats, tandem repeat search can be run even on unfinished sequences. Tandem Repeats Finder was applied to *B. anthracis* sequence obtained from The Institute for Genomic Research through the website at [http://www.tigr.org] . The sequence was recovered as approximately 1000 contigs, for a total amount of slightly more than 5 Mb. Thirty tandem repeats have at least 10 copies of a repeat unit longer than 9 base-pairs. Fourteen of them are polymorphic among the 31 *B. anthracis* strains typed here (Table 2). Twenty-seven different genotypes are identified. Polymorphism information content (PIC) indexes based on the 27 genotypes vary from 0.07 to 0.82. Nine PIC values are above 0.5. Eight alleles are identified for CEB-Bams30, in a size range 270-900 base-pairs (Figure 5). In this case, the resolution of the largest alleles would probably be improved by using an automated DNA sequencer, and more alleles might be resolved. There are clear gaps in the size range coverage shown in Figure 5, and it is likely that the typing of additional strains would uncover new alleles. The genotyping data obtained was used to construct a phylogenetic tree based upon the Neighbor-Joining method ([http://www.infobiogen.fr]). In order to be able to correlate the tree obtained here with earlier studies [2], 5 minisatellites and one microsatellite reported previously were also typed. Figure 6 presents the data obtained and the resulting tree, using the nomenclature previously proposed [2]. Six *Bacillus cereus* strains have also been included and used as an outgroup in the analysis. Occasionally *B. cereus* strains will not amplify (scored as 0 in Figure 6) or will give weak amplification signals (Figure 5, last six lanes on the right). The proposed tree is in good agreement with earlier results. In particular, the A and B clusters are well defined. We have apparently no representatives for the A1b and A3a group, whereas strains 9533 and 9502 to 9505 appear to define a new branch. The correspondence between allele numbering and allele size is indicated in Table 3.

Correlations between polymorphism and structural characteristics of minisatellites

We have looked for correlations between on one hand the number of alleles and polymorphism of the minisatellites, and on the other, simple structural characteristics of the tandem repeats in the sequenced strain : motif size, number of motifs, total length, conservation of the motifs along the array (percent identity), GC content, strand bias. In the case of *B. anthracis*, a highly significant correlation (0.01 level) is observed between polymorphism and both total length and GC content. This is not true for *Y. pestis* in which a strong correlation is seen between the number of alleles and the conservation of the motifs (Figure 7).

Conclusions

We limited here our investigation of tandem repeats to minisatellites, i.e. repeat units longer than 9 base-pairs, so as to avoid simple sequence contingency loci [8] of limited epidemiological value, and to facilitate the typing of alleles with agarose gel electrophoresis. However, simple sequence contingency loci are also represented in the database and are of great interest for molecular pathogenicity studies [6–8]. The use of the tandem repeats database was demonstrated here on two of the most genetically homogeneous human pathogenes, *Y. pestis* and *B. anthracis*. There is consequently a possibility that a common database format for identification and epidemiological analyses of pathogens amenable to minisatellite typing be developed. As more data becomes available on polymorphism associated with tandem repeats, it will be added to the database presented here in order to avoid duplication of work and nomenclature.

Bacterial species differ very significantly in the density of tandem repeats within their genome, and also in their use of tandem repeats. Some species have a very strong excess of tandem repeats with repeat units length which are multiple of three, the most striking examples being *M. tuberculosis* and *P. aeruginosa*. Polymorphism in such tandem repeats is likely to modulate the protein structure rather than gene activity. In *M. tuberculosis*, all tandem repeats with total length (L) higher than 100 bp and 9 or 15 base-pairs long units are located with ORFs [21]. An important proportion of these tandem repeats correspond to the so-called PE and PPE multigene families [21].

In the two species studied here, tandem repeat polymorphism is strongly correlated with one or more of the sequenced allele characteristics, as illustrated in Figure 7. In *Yersinia pestis* a strong correlation is observed between number of alleles observed and homogeneity of the tandem array. In *Bacillus anthracis*, the strongest correlations are with total array length and GC content. It appears that the correlations are not the same in the two species, so that at present at least, the polymorphism associated with a tandem repeat cannot be inferred from its primary sequence. In particular, and in contrast to what is known for microsatellites (1-5 bp repeat units),

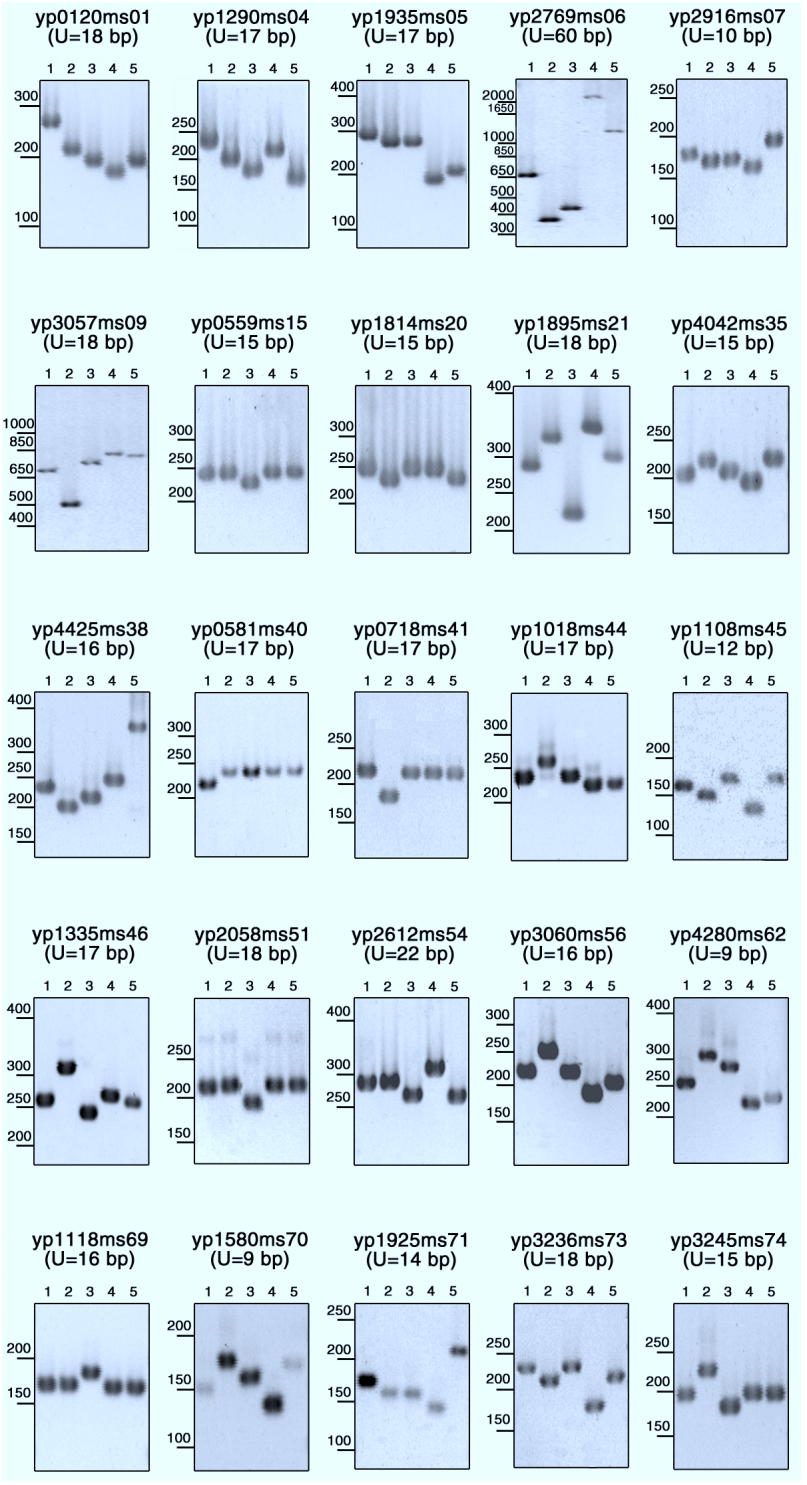


Figure 4

Images of PCR amplification of the twenty-five minisatellites polymorphic in the *Y. pestis* strains DNA from three reference *Y. pestis* strains representing each of the main biovars, antiqua (lane 1), medievalis (lane 2) and orientalis (lane 3) and two *Y. pseudotuberculosis* strains (lanes 4 and 5) have been PCR amplified and an aliquot of the products has been run on 2% horizontal agarose gels as described. The length of the minisatellite motifs (U) and the size range is indicated on each panel. Yp2916ms07 has one of the shortest (10 bp) unit. Four alleles are clearly distinguished between the 150 and 200 bp marker fragments.

some of the minisatellites are highly polymorphic in spite of a poor internal homogeneity of the sequenced allele, as is also the case for minisatellites in the human genome [12]. However, more systematic allele sequencing will be required to demonstrate that polymorphism is not associated with a subclass of alleles showing a higher internal homogeneity. Similarly, allele sequencing will be required to formally establish that the allele size variations observed are indeed (as is likely) the consequence of variations in the number of repeats.

Five among the *B. anthracis* markers described here (Ceb-Bams1, 3, 7, 13 and 30) are highly polymorphic with PIC values (or Nei's index) above 0.7. In this respect, it is important to observe that the length of the allele observed for Ceb-Bams1 in the Ames strain is not of the size expected from the sequence data (Table 2). This may result either from a high mutation rate at Ceb-Bams1 or from a sequencing error. The expected allele size corresponds to allele 4 (Table 3), which is unlikely for the Ames strain because Ceb-Bams1 allele 4 is observed only in cluster B strains (Figure 6) and Ames is well apart of cluster B [2]. A similar situation is observed for Ceb-Bams28, for which the expected product does not correspond to any existing allele in the collection of strains typed. In this case however, the locus is moderately polymorphic, with a PIC value of 0.26 and only three alleles observed (Table 2), so that a sequencing error is the most likely interpretation. This issue could be easily solved by typing with Ceb-Bams1 and Ceb-Bams28 the very strain which has been used for the sequencing project.

It is interesting to observe that, although the magnitude of allele size difference has not been taken into account when building the distance matrix, the resulting phylogenetic tree proposed in Figure 6 tends to group together strains with alleles of similar size at these most variable loci. This is reminiscent of observations made in *H. influenzae* [1] and suggest that mutation events are predominantly small size changes. Here again, more detailed studies involving full allele sequencing should now help understand the succession of events producing a population of alleles.

Materials and methods

Bacterial genomes DNA sequences

Finished sequences in the public domain were recovered by ftp from the NCBI or the Sanger center sites ([http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/bact.html] ; [http://www.sanger.ac.uk/Projects/Microbes/]). Preliminary sequence data for *B. anthracis* was obtained from The Institute for Genomic Research through the website at [http://www.tigr.org] .

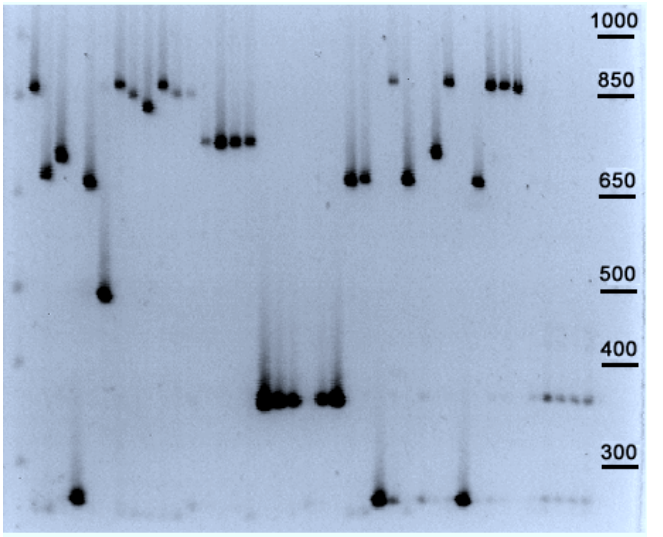


Figure 5
PCR amplification of *B. anthracis* minisatellite CEB-Bams30 DNA from *B. anthracis* and *B. cereus* (six rightmost lanes) was amplified using primers for CEB-Bams30 (Table 2). The PCR products were run on a 40 cm long 2% ordinary agarose gel.

DNA preparation

All strains used here are part of the collection maintained by the Centre d'Etudes du Bouchet (CEB). They originate either from the CIP (Collection Institut Pasteur, [http://www.pasteur.fr/]) or from AFSSA (Agence Française de Sécurité Sanitaire des Aliments, [http://www.afssa.fr/] , Dr Josée Vaissaire). DNA from each isolate was obtained by large-batch procedures or by the simplified procedure as described in [2]. In addition, 15 µg of DNA from the *B. anthracis* Ames strain were kindly provided by Dr Mats Forsman, FOA, Sweden.

Minisatellite PCR amplification and genotyping

PCR reactions were performed in 15 µl containing 1 ng of DNA, 1x Long Range Reaction Buffer 3 (Roche-Boehringer), 1 unit of Taq DNA polymerase, 200 µM of each dNTP, 0.3 µM of each flanking primer. The Taq DNA polymerase was either prepared essentially as described in [22] or purchased from Qbiogen or Roche-Boehringer. The 1x LongRange Buffer 3 is 1.75 mM $MgCl_2$, 50 mM Tris-HCl pH9.2, 16 mM $(NH_4)2SO_4$.

PCR reactions were run on a Perkin-Elmer 9600 or a MJResearch PTC200 thermocycler. An initial denaturation at 96°C for five minutes was followed by 34 cycles of denaturation at 96°C for 20 seconds, annealing at 60°C for 30 seconds, elongation at 65°C for 1 minute, followed by a final extension step of 5 minutes at 65°C. In few cases, other annealing temperatures and/or elongation times were used (see tables 1 and 2). Five microliters of

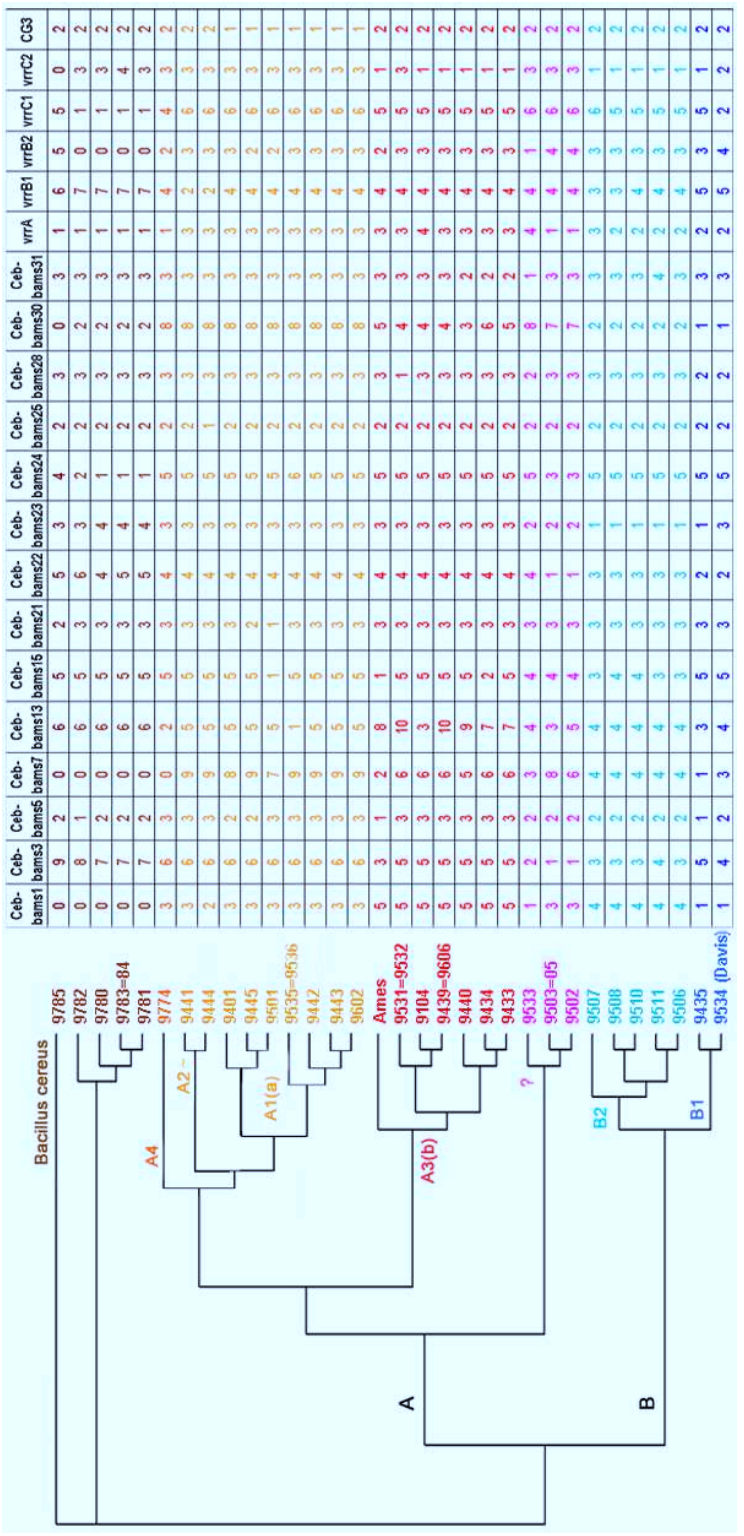


Figure 6

Bacillus anthracis **phylogenetic tree** The genotype of each strain for the polymorphic minisatellites is given (size estimates for each allele are given in Table 3). "0" indicates a failure of the PCR amplification. This is most often associated with *B. cereus* strains, and probably reflects in these cases sequence divergence in the flanking sequence. The phylogenetic tree was produced using the Neighbor-Joining method as available on-line at [http://www.infobiogen.fr.]

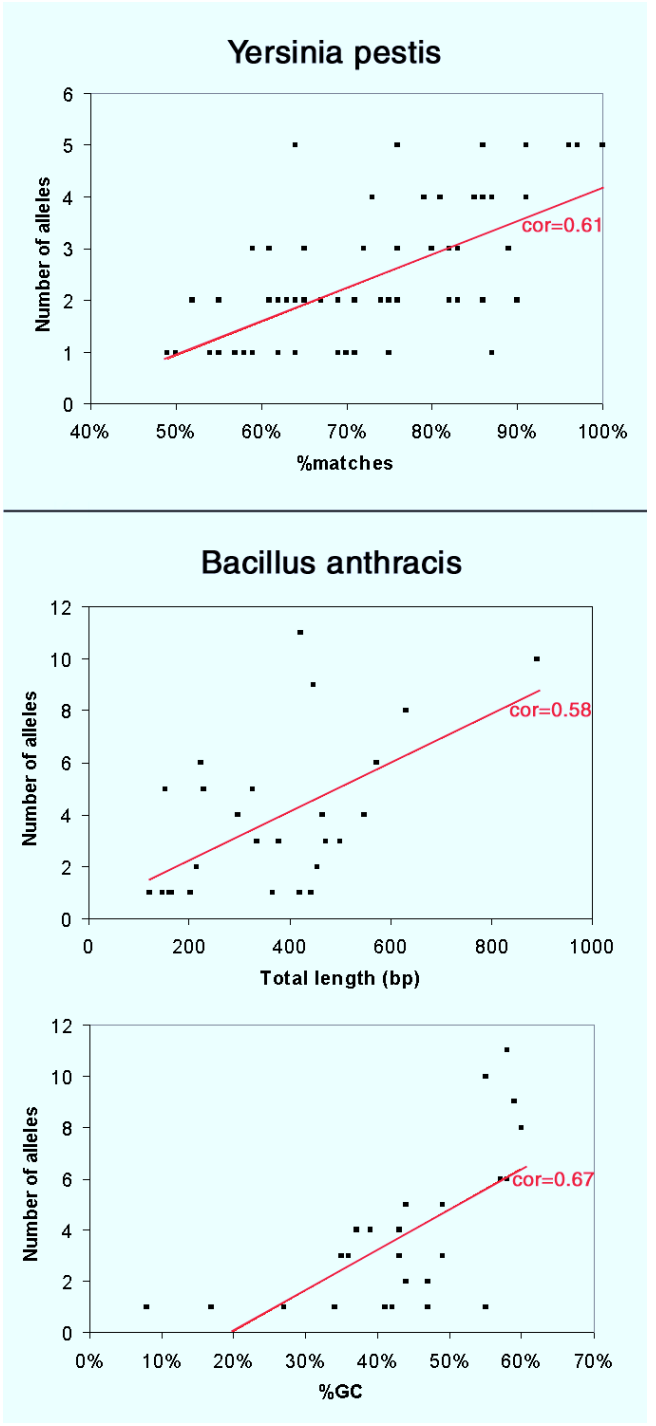


Figure 7
Significant correlation between number of alleles and minisatellites structural characteristics The number of alleles is plotted as a function of Total length and %GC for *Bacillus anthracis*, and %matches for *Yersinia pestis* (the correlations are highly significant at the 0.01 level). Number of alleles for each locus is the total number detected (i.e. *Bacillus anthracis* and *B. cereus*; *Yersinia pestis* and *Y. pseudotuberculosis*).

the PCR products where run on standard 1% or 2% agarose gel (Qbiogen) in 0.5 x TBE buffer at a voltage of 10 V/cm as indicated in Tables 1 and 2. Gel length of 10 to 40 cm were used according to PCR product size and motif length. Gels were stained with ethidium bromide and visualized under UV light. Allele sizes were estimated using as size markers the 1 kb ladder plus (Gibco-BRL which also includes a 100 bp ladder between 100 bp and 500 bp, plus 650, 850 and 1000 bp bands) or the 50 bp ladder (Euromedex) which provides a 50 bp ladder between 50 and 300 bp and a 100 bp ladder from 300 bp to 1000 bp.

Data analysis

Tandem Repeats Finder analysis:

Sequences were processed using the Tandem Repeats Finder software ([http://c3.biomath.mssm.edu/trf.html]). The output was processed to eliminate duplicates before being imported in a database (running under Access2000, Microsoft Corp.) as described previously [12]. The *B. anthracis* preliminary sequence data file uses FASTA type of headers (i.e. >sequenceId) to separate the independent contigs. The headers were replaced by runs of 10 Ns before running Tandem Repeats Finder.

Blast queries against the M. tuberculosis genome:

The identifications of the open reading frames containing a given tandem repeat from M. tuberculosis were done by running a BLAST search on the dedicated web page at [http://www.sanger.ac.uk/Projects/M_tuberculosis/blast_server.shtml] .

Estimation of the excess of tandem repeats with motif length multiple of three:

A χ^2 test was calculated for the difference between the observed number of tandem repeats with motif length multiple of 3 and the expected number of tandem repeats with motif length multiple of 3 (expected value in the absence of bias being the total number of tandem repeats divided by 3). The χ^2 values vary from 0.01 to 253.5. There is a significant excess ($\chi^2 > 3.841$) for all species but 6 (*Buchnera sp*, *T. maritima*, *H. influenzae*, *M. genitalium*, *R. prowazekii*, *Y. pestis*).

Polymorphism index:

Polymorphism Information Index (PIC) or Nei's diversity index is calculated as $1 - \Sigma (\text{allele frequency})^2$ based upon the unique genotypes.

Phylogenetic reconstruction:

A phenetic approach, based on a distance matrix was used. Distance matrix between strains was obtained by counting the number of differences between the corresponding genotypes. Then, Neighbor Joining cluster

analysis was performed with Phylip [23] accessed at [http://www.infobiogen.fr/] . An outgroup was arbitrary chose among *Bacillus cereus* strains (9785) and input order of species was randomised.

Data (genotypes, distance matrix, phylogenetic tree) are available at [http://minisatellites.u-psud.fr/ASPSamp/Phylogenie/data.htm]

Correlation analysis

Correlations were calculated with the statistical program SPSS: Pearson correlation, and non-parametric correlations (Kendall's tau and Spearman's rho) show similar results.

Table 1: Description of *Yersinia* polymorphic markers

Marker	U	N	% GC	V	Primer sequences	PCR	Expected product length (bp)	Estimated size range (bp)	Number of Alleles in *Y.pestis*	Total number of alleles
Markers polymorphic in *Yersinia pestis* strains										
yp0120ms01	18	8	34	86	**L:** CTAAGCACAATTGTTATGCTGAACC **R:** TACTGAATCTGCTTCATTGTTCAAA		228	180 - 280	3	4
yp1290ms04	17	8	27	96	**L:** CGCTGTTGAAGTTTTAGTGTAAGAA **R:** AAATGTAACCTGCCAAACGTG		230	160 - 240	3	5
yp1935ms05	17	11	36	87	**L:** CCTCAGTTCATTGTGTAAAATCTCA **R:** GTATTAGCGAGATCACAGATGAGC		291	190 - 300	2	4
yp2769ms06	60	9	48	64	**L:** AATTTTGCTCCCCAAATAGCAT **R:** TTTTCCCCATTAGCGAAATAAGTA	90 s	606	370 - 2500	3	5
yp2916ms07	10	9	44	85	**L:** ATACCGCTACGATCAGCCTCTAT **R:** ATTTAATATTGATTTTGGGACTTGC		184	150 - 200	2	4
yp3057ms09	18	33	65	91	**L:** CGTTACCCTTGTTGCCAATAGT **R:** ACGCAGAACATGCTTACCTTTTAT	90 s	682	500 - 820	3	5
yp0559ms15	15	10	30	62	**L:** TTGACCAAGTGTAAAAGGCATAAAT **R:** AAACTATCGCCAGCCATTTTAGTA		237	225 - 250	2	2
yp1814ms20	15	9	47	74	**L:** ACAACCTCAGTTTGCCCTTG **R:** GTAAAGAGCGCAATGATCGTACT		253	230 - 250	2	2
yp1895ms21	18	9	51	76	**L:** GCTTAAAGCAGATTGATACTCACG **R:** CTGCATGTTCACCGGTTCAG		278	220 - 350	3	5
yp4042ms35	15	8	41	59	**L:** CTGTTACCGGTCAAAGTGGATATT **R:** AGGCTCTCCTTATCATTATTTGGTC		204	195 - 225	2	3
yp4425ms38	16	8	41	86	**L:** GTGAGGTATAGCTAAACGGTGATGT **R:** CGCCGTAGATTATTTGTCACTTTAT		233	200 - 380	3	5
yp0581ms40	17	7	28	76	**L:** GCAATCATTCACCTAACCATATCTC **R:** GTGCAATAGGCGTTGTTGTGTA		214	220 - 240	2	2
yp0718ms41	17	7	41	75	**L:** GAAGAAAGCCAGCTAATCTGATG **R:** TAATGAATAGCAACGACAACCAATA		217	180 - 220	2	2
yp1018ms44	17	7	38	61	**L:** CAATTCCAACAGCTATTAATGCAA **R:** GAATTTTCATAACACGTTCTTCCTG		233	220 - 260	2	3
yp1108ms45	12	7	65	79	**L:** GCATCGGAGACTGGGTAAAC **R:** TTTCTGAGGATTTATCGGTGTGAT		161	120 - 170	3	4
yp1335ms46	17	7	33	73	**L:** CAGGTTTTACGTTATTTTCTGAAGG **R:** CAGCATGAAGTATGACGGGTATATTA		252	230 - 310	3	4
yp2058ms51	18	7	37	65	**L:** GGTTTTTACCGATATAAATCCTGAG **R:** GACCAAGAAGTTAAGTTGCTTATCG		207	190 - 210	2	2
yp2612ms54	22	7	28	82	**L:** GTCCACCATTTTCATACTGTCACTT **R:** GCTCTTTGTTCGATTTTATTGAATG		281	250 - 300	2	3
yp3060ms56	16	7	21	81	**L:** AACCGACTGACTCACTTTATATTGG **R:** TTCTTTTCCATTACTCAGCCTGTT		220	180 - 250	2	4
yp4280ms62	9	7	33	60	**L:** TTTAGTCTTGATTAAGCTGCGTTTT **R:** ACGGAAGACAACCTTATTATTGATG		240	220 - 310	3	5
yp1118ms69	16	6	39	82	**L:** GACGTTGCAACTGCAAAAATAAG **R:** ACTTGTTGTGAAGACCATCACTCT		179	165 - 180	2	2
yp1580ms70	9	6	32	97	**L:** AAACCAACGGTTCATATTGAATAAA **R:** CTTCTTCCGCTATTTTCCTACAGA		146	140 - 170	3	5
yp1925ms71	14	6	45	91	**L:** GCTACTCGAATATGAGTTAGCCAAA **R:** ATTGCCATATTGGATGCTAAAATAA		171	145 - 210	2	4
yp3236ms73	18	6	40	89	**L:** AATACCCTGTGGGTGATAATGAAC **R:** ATCGATTTAGGTACCACCAATTCA		225	175 - 230	2	3
yp3245ms74	15	6	44	83	**L:** CCCCGACTTATATCAAGCACTG		195	180 - 225	3	3

Table 1: Description of *Yersinia* polymorphic markers

					R: AACTGACGATCTTTTTCACTGAGTT					
Markers polymorphic in 5 *Yersinia* strains (monomorphic in *pestis*)										
yp0802ms02	**18**	**12**	**49**	**86**	**L:** CTGACACAAAACGAGAGCCTATTT **R:** AGCGTGAGTGGGCTATCAATAC	53°C 1 min	**314**	**240 - 315**	**1**	**2**
yp2925ms08	**15**	**12**	**39**	**63**	**L:** AGCCTTTTTGTTGATTATCAGTCAT **R:** CGATAATAATACTGAATTACCGGATG		**270**	**270 - 290**	**1**	**2**
yp4411ms10	**14**	**8**	**32**	**69**	**L:** ATCATGCTTTTGCCTCAATATAATC **R:** GAAACGCAGTCCCTGTTGTAG		**191**	**190 - 210**	**1**	**2**
yp0813ms16	**17**	**8**	**39**	**64**	**L:** GTTGGTTATCCGACAGTCTTCAATA **R:** GCAATTCGTTATGGCTTAGTAAAAA		**235**	**230 - 270**	**1**	**2**
yp1269ms18	**27**	**9**	**54**	**55**	**L:** GCAAAAGCTGAAGCAGATAAAATAG **R:** AAACCACCAAACAAATCATCAAC		**303**	**220 - 250**	**1**	**2**
yp2196ms22	**20**	**8**	**12**	**55**	**L:** AAACCAACAAGAAAACTGTAACCAC **R:** CATTCACCATTGATGTCCTTAGAC	90 s	**265**	**270 - 1500**	**1**	**2**
yp2324ms24	**19**	**8**	**34**	**65**	**L:** TTCACCGGGTTACCTTAATTACATA **R:** CTACCTTGCTGTCAACACTCGAC		**255**	**215 - 255**	**1**	**2**
yp2331ms25	**17**	**9**	**36**	**76**	**L:** AACGCGTTAATAAAACAATAAAGTG **R:** CAATATCCTTTTACTCAGCCGATG		**181**	**190 - 230**	**1**	**3**
yp2679ms27	**16**	**8**	**20**	**76**	**L:** ATGATTACTGGCAAGAGCACTATGT **R:** AACAAGATCACCTGGTCGTTAAA		**217**	**200 - 220**	**1**	**2**
yp2908ms28	**18**	**8**	**40**	**69**	**L:** GCAGAAATAATCTTCAGGAGAAACA **R:** AGATCGTCGTTAGTCCATGTCAG		**242**	**190 - 290**	**1**	**2**
yp2958ms29	**16**	**8**	**23**	**61**	**L:** AAAATAGTCTGTGTTTCAGCAAAGC **R:** CCTTAAAAACCCTAAGTGGGTAAAA		**215**	**215 - 245**	**1**	**2**
yp3225ms30	**54**	**11**	**51**	**52**	**L:** CAATAATACCATCGTGCGTGATAC **R:** TATTAATGGTGGTGTTAGTCGCTGT		**683**	**680 - 900**	**1**	**2**
yp3532ms31	**14**	**8**	**30**	**67**	**L:** GTTATTTATTTTTGCCCCAACTTGT **R:** TTAGCCTGTTTGTTCTTCAAATAGC		**217**	**215- 245**	**1**	**2**
yp3787ms32	**18**	**8**	**49**	**65**	**L:** CGATAACGTTAATGCCATCAGTAG **R:** GCGCCGGTAAGTTTTGTTTATTA		**218**	**190 - 240**	**1**	**3**
yp3795ms33	**15**	**8**	**43**	**67**	**L:** CCCTTCTTTTTATGCTTGAAGATACT **R:** GTTGAACCACAGGCTGTTGAG		**210**	**210 - 225**	**1**	**2**
yp4371ms37	**18**	**8**	**35**	**82**	**L:** TACTTAGGCATTGTCTCTTCACTCC **R:** CTGAAATTATCAGTAGGTGGTTCGT		**235**	**235 - 255**	**1**	**2**
yp0999ms43	**17**	**7**	**38**	**80**	**L:** ATTCCACCACCAACAATTATCAC **R:** GGTATTCGTATTGAAGATGACATTG		**211**	**220 - 300**	**1**	**3**
yp1962ms50	**18**	**7**	**34**	**71**	**L:** TACCGAGGTATTCCCTGGTCTAAT **R:** AGTTGACTCCCAGTCACTTTTCC		**225**	**225 - 240**	**1**	**2**
yp3734ms59	**16**	**7**	**36**	**69**	**L:** ATTATCATGACCCTTCCAGTGCTAT **R:** CATCAAAATGCCAGGAGAGTAAC		**218**	**200 - 220**	**1**	**2**
yp4338ms63	**17**	**7**	**38**	**72**	**L:** ATTAACGATTTCTTGTCGCTCAGT **R:** AATCAGTAACGGCATGTGTCAGTA		**194**	**190 - 275**	**1**	**3**
yp0549ms66	**18**	**6**	**41**	**83**	**L:** TAAAAGCGTCAACAAAGTAGGTCAT **R:** GTTCCTGTTGTTGAAAATGCTG		**212**	**200 - 220**	**1**	**2**
yp0782ms67	**18**	**6**	**40**	**90**	**L:** TTCCAGGCTAAAGATATTGACTTTG **R:** CTCGGCTTGTTCTACGTTTAATG		**248**	**250 - 270**	**1**	**2**
yp1053ms68	**18**	**6**	**32**	**82**	**L:** CCGTTATCTGGTGAAAGTGAACAG **R:** GTCCGGTAGCCTGATTGTTTATT		**182**	**175 - 205**	**1**	**3**
yp3634ms75	**15**	**6**	**36**	**80**	**L:** ATGTGAGCTTGATTGCTGAGTAGT **R:** TCATATTTTAGGTGTTTTGCCTTTG		**210**	**180 - 210**	**1**	**3**

Some structural characteristics of the tandem repeats are presented : U (unit length), N (number of repeats), %GC, V (% of conservation). PCR and electrophoresis conditions are as described in the material and methods section : annealing temperature is 60°C, elongation time is 60 seconds and gels are 2% agarose except when indicated otherwise. Total number of alleles means number of alleles in 3 *Y. pestis* and 2 *Y. pseudotuberculosis* strains.

Table 2: Description of *Bacillus anthracis* polymorphic markers

Marker	U	N	% GC	V	Primer sequences	PCR	Expected product length in bp (observed)	Estimated size range (bp)	Number of alleles in B. *anthracis*	Total number of alleles	PIC index
Ceb-Bams 1	21	16	44	88	**L:** GTTGAGCATGAGAGGTACCTTGTCCTTTTT **R:** AGTTCAAGCGCCAGAAGGTTATGAGTTATC		485 (520)	410-520	5	5	0.72
Ceb-Bams 3	15	25	59	73	**L:** GCAGCAACAGAAAACTTCTCTCCAATAACA **R:** TCCTCCCTGAGAACTGCTATCACCTTTAAC	1%	544	480-860	6	9	0.75
Ceb-Bams 5	39	10	49	92	**L:** GCAGGAAGAACAAAAGAAACTAGAAGAGCA **R:**ATTATTAGCAGGGGCCTCTCCTGCATTACC	53°C 60s	307	305-385	3	3	0.56
Ceb-Bams 7	18	49	55	69	**L:** GAATATTCGTGCCACCTAACAAAACAGAAA **R:** TGTCAGATCTAGTTGGCCCTACTTTTCCTC	65°C 1%	1017	600-1950	9	9	0.82
Ceb-Bams 13	9	70	60	79	**L:** AATTGAGAAATTGCTGTACCAAACT **R:** CTAGTGCATTTGACCCTAATCTTGT	120s 1%	814	330-850	8	11	0.79
Ceb-Bams 15	18	12	57	77	**L:** GTATTTCCCCCAGATACAGTAATCC **R:** GTGTACATGTTGATTCATGCTGTTT		409	410-610	5	5	0.59
Ceb-Bams 21	45	11	43	75	**L:** TGTAGTGCCAGATTTGTCTTCTGTA **R:** CAAATTTTGAGATGGGAGTTTTACT		676	540-680	3	3	0.14
Ceb-Bams 22	36	15	39	81	**L:** ATCAAAAATTCTTGGCAGACTGA **R:** ACCGTTAATTCACGTTTAGCAGA		735	590-950	4	6	0.51
Ceb-Bams 23	42	11	37	85	**L:** CGGTCTGTCTCTATTATTCAGTGGT **R:** CCTGTTGCTCCTAGTGATTTCTTAC		653	570-820	3	4	0.49
Ceb-Bams 24	42	11	44	80	**L:** CTTCTACTTCCGTACTTGAAATTGG **R:** CGTCACGTACCATTTAATGTTGTTA		630	335-670	3	6	0.2
Ceb-Bams 25	15	14	45	60	**L:** CCGAATACGTAAGAAATAAATCCAC **R:** TGAAAGATCTTGAAAAACAAGCATT		391	375-390	2	2	0.07
Ceb-Bams 28	24	14	36	70	**L:** CTCTGTTGTAACAAAATTTCCGTCT **R:** TATTAAACCAGGCGTTACTTACAGC		493 (400)	300-400	3	3	0.26
Ceb-Bams 30	27	16	58	78	**L:** AGCTAATCACCTACAACACCTGGTA **R:** CAGAAAATATTGGACCTACCTTCC	120s 1%		200-890	11	11	0.77
Ceb-Bams 31	9	64	58	57	**L:** GCTGTATTTATCGAGCTTCAAAATCT **R:** GGAGTACTGTTTGTTGAATGTTGTTT	 1%	772	300-850	4	4	0.32

Some structural characteristics of the tandem repeats are presented : U (unit length), N (number of repeats), %GC, V (% of conservation). PCR and electrophoresis conditions are as described in the material and methods section : annealing temperature is 60°C, elongation time is 60 seconds and gels are 2% agarose except when indicated otherwise. The expected product length is deduced from the sequencing data corresponding to the Ames strain. When the Ames strains typing does not fit with the expected value, the observed value is indicated between (). Only one side of the Ceb-Bams30 minisatellite can be identified in the available Ames sequence. The other side was identified in the course of the independent, partial sequencing of *B. anthracis* strains (Vergnaud and col., unpublished data). Total number of alleles includes alleles observed in the *B. cereus* strains. Polymorphism Information Index (PIC) or Nei's diversity index is calculated as $1 - \Sigma (\text{allele frequency})^2$.

Table 3: Correspondence between *B. anthracis* allele sizes and allele numbering

allele nb marker name	1	2	3	4	5	6	7	8	9	10
Ceb-Bams1	~ 410	~ 430	~ 450	~ 480	~ 520					
Ceb-Bams3	484	514	544	559	574	589	704	734	857	
Ceb-Bams5	307	346	385							
Ceb-Bams7	603	1017	1305	1503	1557	1647	1809	1899	1953	
Ceb-Bams13	328	382	454	481	490	652	742	787	814	850
Ceb-Bams15	409	535	571	589	607					
Ceb-Bams21	541	631	676							
Ceb-Bams22	591	627	699	735	~ 900	~ 950				
Ceb-Bams23	569	611	653	821						
Ceb-Bams24	336	420	462	504	630	672				
Ceb-Bams25	376	391								
Ceb-Bams28	~ 300	~ 375	~ 400							
Ceb-Bams30	266	375	500	660	695	730	760	850 to		

Table 3: Correspondence between *B. anthracis* allele sizes and allele numbering

							900
Ceb-Bams31	304	700	772	853			
vrrA	289	301	313	325	337		
vrrB1	184	193	220	229	256	~ 280	~ 290
vrrB2	~ 135	153	162	171	~ 180		
vrrC1	400	502	520	538	583	613	685
vrrC2	532	568	607	660			
CG3	153	158					

Alleles have been numbered in increasing size order. When the allele size (in base-pairs) observed in the Ames strain was in agreement with the size expected according to Ames sequence data, the values indicated in the table assume that alleles differ in size by a multiple of the motif length. These likely values will have to be confirmed by more accurate size estimation tools and allele sequencing. When the allele size in Ames is not as expected (Ceb-Bams1 and Ceb-Bams28), the estimated values are preceded by a ~. The Vrr and CG3 allele sizes were described in [2]; new alleles are indicated by a ~.

Acknowledgements

Minisatellite investigations in the laboratory are supported by grants from Délégation Générale de l'Armement (DGA/DSA/STTC and DGA/DSA/SP-Nuc). Preliminary sequence data for *B. anthracis* was obtained from The Institute for Genomic Research through the website at [http://www.tigr.org] . Sequencing of *B. anthracis* was accomplished with support from Office of Naval Research, Department of Energy, and National Institute of Allergy and Infectious diseases. We wish to thank the referees for the significant improvements they have suggested.

References

1. van Belkum A, Scherer S, van Leeuwen W, Willemse D, van Alphen L, Verbrugh H: **Variable number of tandem repeats in clinical strains of Haemophilus influenzae.** *Infect Immun* 1997, **65**:5017-27
2. Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, Okinaka R, Jackson PJ, Hugh-Jones ME: **Multiple-Locus Variable-Number Tandem Repeat Analysis Reveals Genetic Relationships within Bacillus anthracis.** *J Bacteriol* 2000, **182**:2928-2936
3. Frothingham R, Meeker-O'Connell WA: **Genetic diversity in the Mycobacterium tuberculosis complex based on variable numbers of tandem DNA repeats.** *Microbiology* 1998, **144**:1189-96
4. Supply P, Mazars E, Lesjean S, Vincent V, Gicquel B, Locht C: **Variable human minisatellite-like regions in the Mycobacterium tuberculosis genome.** *Mol Microbiol* 2000, **36**:762-71
5. Adair DM, Worsham PL, Hill KK, Klevytska AM, Jackson PJ, Friedlander AM, Keim P: **Diversity in a variable-number tandem repeat from Yersinia pestis.** *J Clin Microbiol* 2000, **38**:1516-9
6. van Ham SM, van Alphen L, Mooi FR, van Putten JP: **Phase variation of H. influenzae fimbriae: transcriptional control of two divergent genes through a variable combined promoter region.** *Cell* 1993, **73**:1187-96
7. Weiser JN, Love JM, Moxon ER: **The molecular mechanism of phase variation of H. influenzae lipopolysaccharide.** *Cell* 1989, **59**:657-65
8. Bayliss CD, Field D, Moxon ER: **The simple sequence contingency loci of Haemophilus influenzae and Neisseria meningitidis.** *J Clin Invest* 2001, **107**:657-66
9. Henderson IR, Owen P, Nataro JP: **Molecular switches - the ON and OFF of bacterial phase variation.** *Mol Microbiol* 1999, **33**:919-32
10. Wang G, Ge Z, Rasko DA, Taylor DE: **Lewis antigens in Helicobacter pylori: biosynthesis and phase variation.** *Mol Microbiol* 2000, **36**:1187-96
11. Wilton JL, Scarman AL, Walker MJ, Djordjevic SP: **Reiterated repeat region variability in the ciliary adhesin gene of Mycoplasma hyopneumoniae.** *Microbiology* 1998, **144**:1931-43
12. Vergnaud G, Denoeud F: **Minisatellites: Mutability and Genome Architecture.** *Genome Res* 2000, **10**:899-907
13. Kokoska RJ, Stefanovic L, Tran HT, Resnick MA, Gordenin DA, Petes TD: **Destabilization of yeast micro- and minisatellite DNA sequences by mutations affecting a nuclease involved in Okazaki fragment processing (rad27) and DNA polymerase delta (pol3-t).** *Mol Cell Biol* 1998, **18**:2779-88
14. Debrauwère H, Buard J, Tessier J, Aubert D, Vergnaud G, Nicolas A: **Meiotic instability of human minisatellite CEB1 in yeast requires DNA double-strand breaks.** *Nat Genet* 1999, **23**:367-71
15. De Bolle X, Bayliss CD, Field D, van de Ven T, Saunders NJ, Hood DW, Moxon ER: **The length of a tetranucleotide repeat tract in Haemophilus influenzae determines the phase variation rate of a gene with homology to type III DNA methyltransferases.** *Mol Microbiol* 2000, **35**:211-22
16. van Belkum A, Scherer S, van Alphen L, Verbrugh H: **Short-sequence DNA repeats in prokaryotic genomes.** *Microbiol Mol Biol Rev* 1998, **62**:275-93
17. Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A, Carniel E: **Yersinia pestis, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis [published erratum appears in Proc Natl Acad Sci U S A 2000 Jul 5;97(14):8192].** *Proc Natl Acad Sci U S A* 1999, **96**:14043-8
18. Helgason E, Okstad OA, Caugant DA, Johansen HA, Fouet A, Mock M, Hegna I, Kolsto : **Bacillus anthracis, Bacillus cereus, and Bacillus thuringiensis - one species on the basis of genetic evidence.** *Appl Environ Microbiol* 2000, **66**:2627-30
19. Keim P, Kalif A, Schupp J, Hill K, Travis SE, Richmond K, Adair DM, Hugh-Jones M, Kuske CR, Jackson P: **Molecular evolution and diversity in Bacillus anthracis as detected by amplified fragment length polymorphism markers.** *J Bacteriol* 1997, **179**:818-24
20. Benson G: **Tandem repeats finder: a program to analyze DNA sequences.** *Nucleic Acids Res* 1999, **27**:573-80
21. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Barrell BG, *et al*: **Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.** *Nature* 1998, **393**:537-44
22. Engelke DR, Krikos A, Bruck ME, Ginsburg D: **Purification of Thermus aquaticus DNA polymerase expressed in Escherichia coli.** *Anal. Biochem.* 1990, **191**:396-400
23. Felsenstein J: **PHYLIP - Phylogeny Inference Package (Version 3.2).** *Cladistics* 1989, **5**:164-166

3.2 Développement de marqueurs polymorphes chez *Pseudomonas aeruginosa*

3.2.1 Etude MLVA

3.2.1.1 Caractéristiques des répétitions en tandem chez *P. aeruginosa*

Le séquençage du génome complet de la souche PAO1 publié en septembre 2000 (Stover 2000) a ouvert la voie à une étude plus exhaustive des répétitions en tandem chez *P. aeruginosa*. Dans l'article décrit dans le paragraphe précédent (Le Flèche 2001), nous avions montré que le génome de la souche PAO1 comportait un excès de répétitions en tandem dont le motif avait une taille multiple de 3 donc potentiellement localisées dans des régions codantes.

Le génome de *P. aeruginosa* est très riche en répétitions en tandem, il y en a 3225 au total, et la densité de répétitions d'une longueur de plus de 100 pb par Mb est de 48. Nous avons limité l'étude des répétitions à celles ayant :

- un motif d'une taille minimum de 9 paires de bases, d'une part pour éviter les locus de type microsatellites qui peuvent être des locus de contingence (Bayliss 2001), donc potentiellement instables et par conséquent de moindre intérêt pour une étude épidémiologique, et d'autre part, en raison des limites de la technique de séparation des produits de PCR sur gel d'agarose (méthode par ailleurs la moins coûteuse et la plus facilement accessible à tous les laboratoires) ;

- un nombre de répétitions minimum de 7, en raison de l'abondance de répétitions en tandem dans le génome de *P. aeruginosa*. En effet, nous avions constaté pour les répétitions tandem du génome de *B. anthracis* que la longueur totale de la répétition semblait être corrélée au polymorphisme.

Deux cent une répétitions en tandem satisfont la requête « U≥ 9pb et N≥ 7 ». La Figure 15 présente la distribution des 201 répétitions en fonction de la taille du motif. On voit très clairement la forte proportion de motifs d'une taille multiple de 3pb.

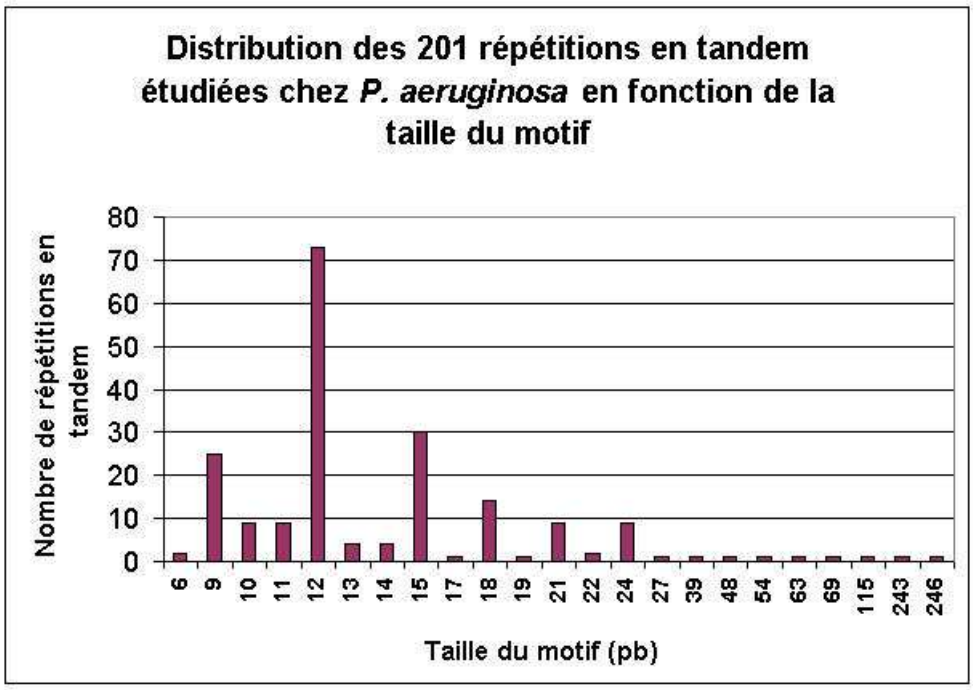


Figure 15

Nous avons comparé la distribution des 3325 répétitions en tandem du génome de PAO1 et des 201 répétitions testées pour cette étude MLVA, pour les critères de conservation du motif et de longueur totale de la répétition (ces deux critères ont été corrélés au polymorphisme chez certaines espèces bactériennes précédemment étudiées au laboratoire (Le Flèche 2001)).

La Figure 16 illustre la distribution des 3325 répétitions et des 201 répétitions en fonction de la conservation des motifs. La distribution des 201 répétitions en tandem étudiées ici présente un décalage de la conservation des motifs par rapport à la population totale, dans le sens d'une plus faible conservation.

La Figure 17 illustre la distribution des 3325 répétitions et des 201 répétitions en fonction de la longueur totale de la répétition. La majorité des répétitions chez *P. aeruginosa* est de petite taille (souvent avec seulement 2 motifs répétés). Les 201 répétitions ne constituent pas un échantillon représentatif des 3225 répétitions totales étant donné le biais introduit en faisant une requête sur le nombre de copies minimum de 7, sélectionnant ainsi les plus grandes répétitions. La Figure 17 montre le décalage de distribution des tailles dans les deux populations de répétitions en tandem considérées.

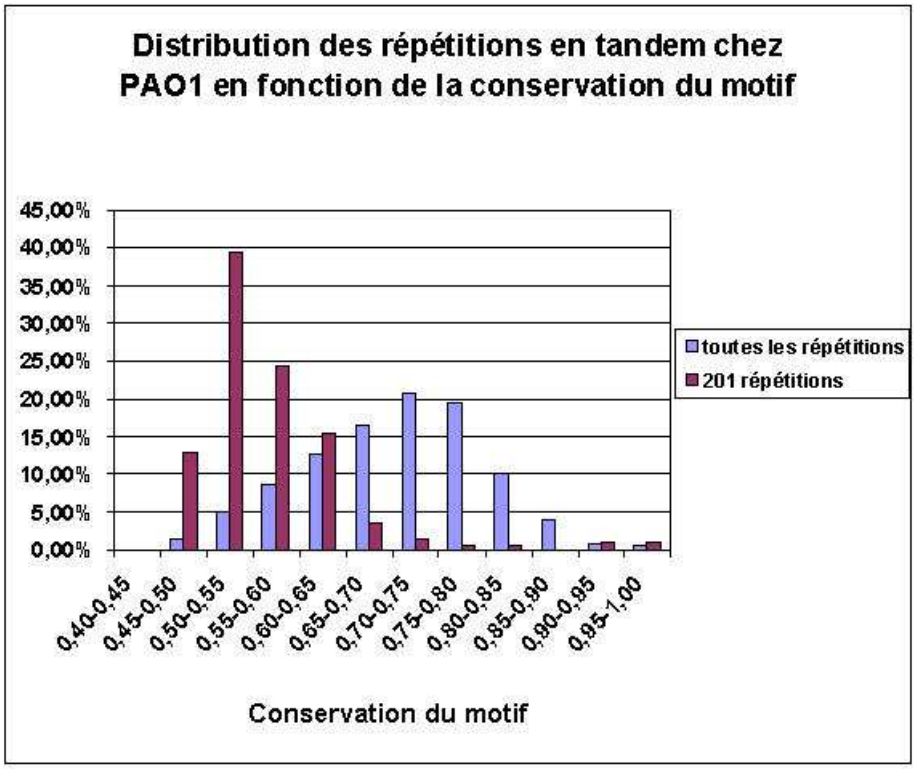


Figure 16

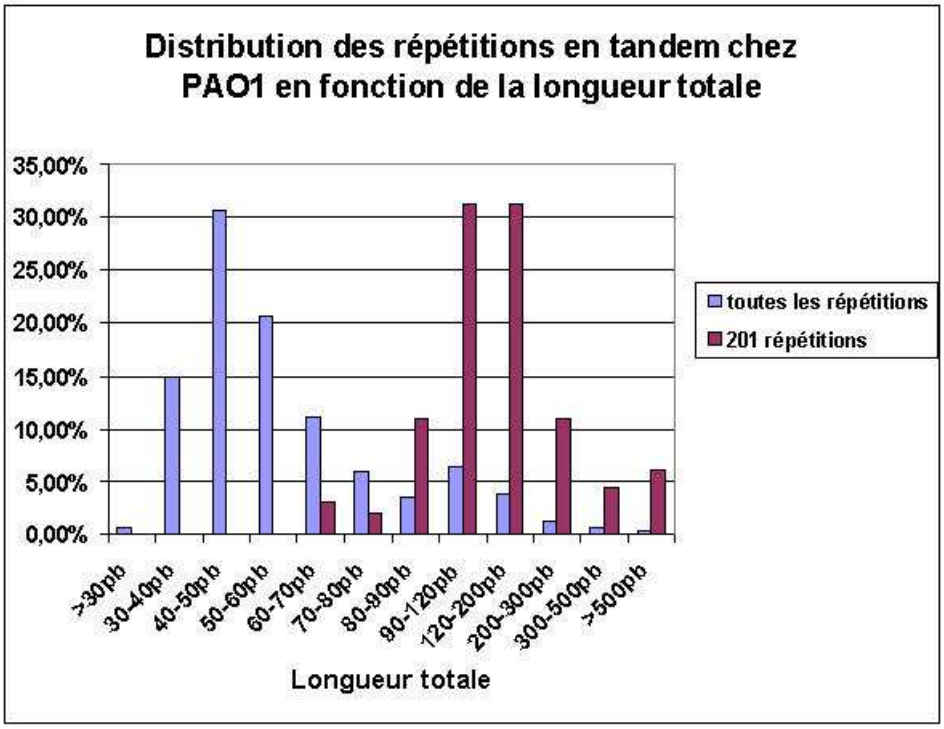


Figure 17

Nous avons ensuite mesuré expérimentalement dans une population de souches d'origine clinique le polymorphisme associé aux 201 répétitions en tandem candidates. La collection de 102 souches a été analysée par ribotypage dans une étude antérieure (Brisse 2000). Dans un premier temps nous avons sélectionné une sous-collection de 12 souches parmi les 102 totales, pour tester le polymorphisme des 201 répétitions. Les souches choisies pour cette étude préliminaire sont toutes de ribogroupe différent.

3.2.1.2 Résultats de l'étude MLVA

Vingt-trois parmi les 201 répétitions en tandem testées n'ont pu être amplifiées par PCR, malgré des tentatives de mise au point de PCR à gradient de température (voir Annexe 3). Cent soixante dix locus sont monomorphes pour la sous-collection des 12 souches. Les séquences des amorces utilisées ainsi que les conditions de PCR sont disponibles en Annexe 3 et sont également consultables dans la base de données du laboratoire.

Enfin, pour 8 minisatellites possédant au moins deux allèles, nous avons étendu l'analyse à toute la collection de souches (Onteniente 2003).

La Figure 18 représente la distribution des 201 répétitions le long du génome de PAO1 et indique la position des 8 répétitions polymorphes (A) ainsi que la distribution des répétitions monomorphes et polymorphes selon la taille du motif (B).

Un autre type d'arbre proposé par le logiciel BioNumerics positionne les souches de façon à minimiser la distance totale de l'ensemble des branches. L'analyse permet également de créer des génotypes « hypothétiques » dont l'existence réduit la taille totale. Cette représentation utilise le critère « categorical » c'est à dire que chaque marqueur est considéré comme un caractère indépendant et le même poids est donné à chaque allèle. La Figure 19 illustre la représentation « minimum spanning tree » des 87 souches *P. aeruginosa* étudiées par l'approche MLVA en utilisant les 7 marqueurs.

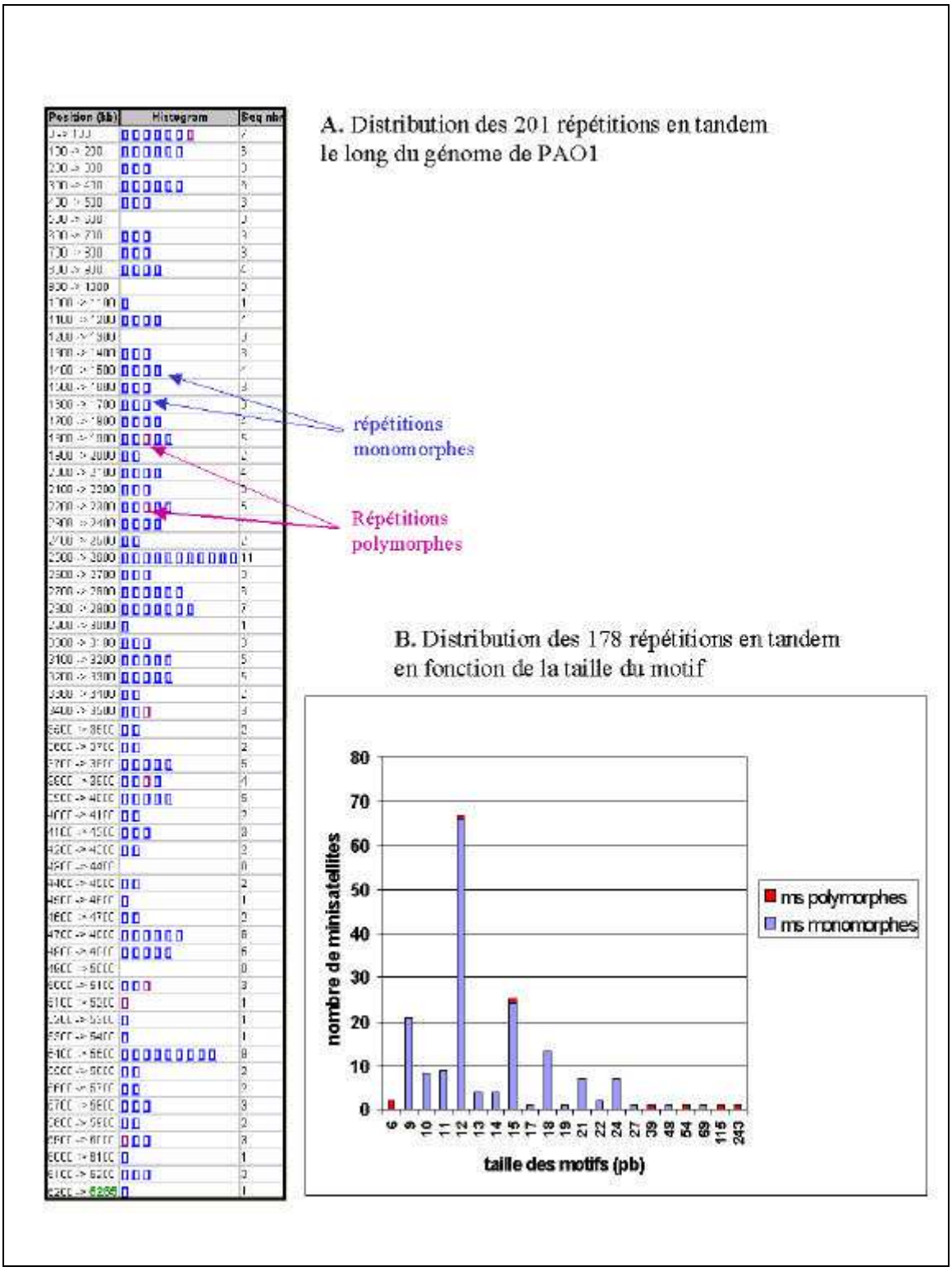


Figure 18 : Distribution des 201 répétitions en tandem sur le génome PAO1 et distribution des répétitions polymorphes/monomorphes

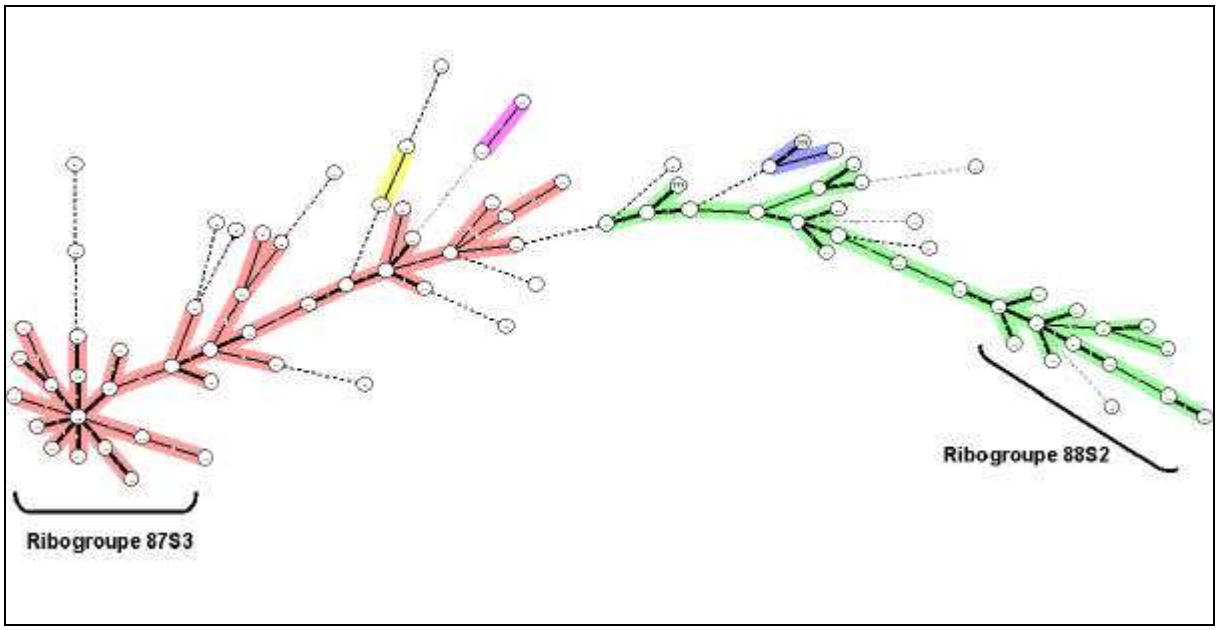


Figure 19 : Représentation « minimun spanning tree »
de l'arbre MLVA des 90 souches *P. aeruginosa*.

L'éloignement des deux ribogroupes majoritaires est illustré par cette représentation des résultats de l'analyse MLVA. Si les sérotypes avaient été représentés sur cette figure, nous aurions observé également une répartition très claire des souches de sérotype O11 d'un côté et O12 de l'autre, se superposant aux deux ribogroupes majoritaires respectivement 87S-3 et 88S-2.

Nous avons vérifié la robustesse du typage MLVA par typage en aveugle de 10 souches de la collection déjà analysées. Ce type de contrôle montre qu'il sera possible de standardiser les analyses entre laboratoires. La standardisation était essentiellement possible, jusque là, pour la technique de MLST, dont les résultats sont comparables sans ambiguïté quel que soit le laboratoire ayant réalisé les analyses (en cas de désaccord, les chromatogrammes de séquençage peuvent être échangés pour évaluation). L'analyse MLVA est reproductible et d'une grande robustesse, elle se développera probablement dans les années à venir soit dans les laboratoires hospitaliers d'analyses biologiques, soit dans des laboratoires dédiés au typage (ce pourrait être par exemple en France le rôle des CNR, Centre Nationaux de Référence, ou de sociétés privées). La page web prototype développée au laboratoire permet à tout autre laboratoire de soumettre ses analyses afin de les comparer avec celles que nous avons réalisées et d'identifier les souches les plus proches d'une souche nouvelle.

Ce deuxième article (Onteniente 2003), intitulé " Evaluation of the Polymorphisms Associated with Tandem Repeats for *Pseudomonas aeruginosa* Strain Typing " (Evaluation du polymorphisme associé aux répétitions en tandem pour le typage de souches de *Pseudomonas aeruginosa*) a constitué la première étude MLVA réalisée chez *P. aeruginosa*.

Jusqu'à présent ce sont les techniques classiques de génotypage qui étaient utilisées pour le typage de cette bactérie, comme par exemple l'électrophorèse en champ pulsé, la technique « RAPD », le ribotypage, le typage des séquences IS (voir paragraphe 1.2.1.1.3).

Résumé de l'article:

Nous présentons le développement d'un outil de typage de *Pseudomonas aeruginosa*, l'analyse MLVA, (Multiple-Locus VNTR (Variable Number of Tandem Repeats) Analysis). Nous avons d'abord évalué le polymorphisme de 201 répétitions en tandem, sélectionnées parmi plus de 3000 présentes dans la souche PAO1, avec une collection test de 12 souches cliniques de génotypes distincts. Sept locus VNTR, facilement interprétables avec la technologie utilisée ici, ont été identifiés et ont servi au génotypage de 89 souches cliniques qui avaient été classées, dans une étude préalable, dans 46 ribotypes dont 2 très répandus. Soixante et onze génotypes MLVA différents ont été observés. A deux exceptions près, les souches de même ribotype ont été regroupées ensemble après l'analyse des données MLVA. Les 27 isolats appartenant au ribotype le plus fréquent ont été séparés en 14 génotypes MLVA, et les 18 souches du deuxième ribotype le plus fréquent ont quant à elles été séparées en 15 génotypes MLVA. L'analyse par électrophorèse en champ pulsé d'un sous-groupe de 17 souches du ribotype majoritaire avec l'enzyme *Spe*I permet de distinguer 7 types, identiques au nombre de génotypes MLVA observés dans ce sous-groupe. Nos données montrent que le typage MLVA de *P. aeruginosa* basé sur un jeu de 7 locus a un fort pouvoir discriminant. Du fait de sa grande reproductibilité et de sa facilité de transfert entre laboratoires, le typage MLVA représente un outil de surveillance épidémiologique de *P. aeruginosa* très prometteur. Par ailleurs, un service gratuit en ligne d'identification de souches par genotypage a été développé au laboratoire. En pratique cependant, les données MLVA peuvent être très simplement conservées et analysées localement à partir d'un tableau. Les lectures des résultats peuvent être manuelles (comparaison des tailles d'allèles avec une ou deux souches de référence) et ne requièrent pas les outils plus sophistiqués et coûteux tels que BioNumerics, qui ne se justifient que pour de grands projets en phase de développement et validation.

JOURNAL OF CLINICAL MICROBIOLOGY, Nov. 2003, p. 4991–4997 Vol. 41, No. 11
0095-1137/03/$08.00+0 DOI: 10.1128/JCM.41.11.4991–4997.2003

Evaluation of the Polymorphisms Associated with Tandem Repeats for *Pseudomonas aeruginosa* Strain Typing

Lucie Onteniente,[1] Sylvain Brisse,[2]† Panayotis T. Tassios,[3] and Gilles Vergnaud[1,4]*

Institut de Génétique et Microbiologie, Université Paris Sud, 91405 Orsay cedex,[1] and Centre d'Etudes du Bouchet, BP3, 91710 Vert le Petit,[4] France; Eijkman-Winkler Institute, Utrecht University, 3584 CX Utrecht, The Netherlands[2]; and Department of Microbiology, Medical School, National and Kapodestrian University of Athens, M. Asias 75, 115 27 Athens, Greece[3]

Received 24 April 2003/Returned for modification 23 June 2003/Accepted 19 August 2003

We report on the development of a scheme for the typing of *Pseudomonas aeruginosa*, multiple-locus variable number of tandem repeat (VNTR) analysis (MLVA). We first evaluated the polymorphisms of 201 tandem repeat loci selected from more than 3,000 such sequences present in strain PAO1 with a test collection of 12 genotypically distinct clinical strains. Seven VNTR loci which can be easily scored with the technology used here were identified and used to genotype a collection of 89 clinical isolates that had previously been classified into 46 ribotypes, including 2 widespread ribotypes. Seventy-one different MLVA genotypes could be distinguished. With only two exceptions, strains with identical ribotypes were grouped together upon cluster analysis of the MLVA data. The 27 isolates with the most frequent ribotype were divided into 14 MLVA types, and the 18 isolates with the second most frequent ribotype were divided into 15 MLVA types. Analysis of a subset of 17 strains belonging to the major ribotype by pulsed-field gel electrophoresis with the enzyme *Spe*I distinguished seven types, identical to the number of MLVA types in this subset. Our data show that MLVA typing of *P. aeruginosa* based on the first set of loci has a high discriminatory power. Because MLVA is highly reproducible and easily portable among laboratories, it represents a very promising tool for the molecular surveillance of *P. aeruginosa*. A free, online strain identification service based on the genotyping data produced herein has been developed.

Pseudomonas aeruginosa is a ubiquitous gram-negative bacterium and a common opportunistic pathogen in hospitals. It is the most important cause of lung colonization in patients with cystic fibrosis and has high rates of multidrug resistance (for a review, see reference 14). Accurate typing and characterization of isolates are essential to understanding the epidemiology of this pathogen. Serotyping is traditionally used for strain typing, but it cannot be easily applied to the mucoid strains found in a significant proportion of cystic fibrosis patients. The development of DNA-based typing methods has circumvented this difficulty, as well as the problem of the limited discriminatory power of serotyping. To be useful, a molecular surveillance tool should be highly discriminatory and reproducible. It should also be easy to standardize. Furthermore, the resulting data should be able to be easily stored, retrieved, and compared by use of databases that can be shared between laboratories. Finally, the overall cost (including that of labor) of the methodology should be as low as possible (23). The methods available at this time for the genotyping of *P. aeruginosa* include pulsed-field gel electrophoresis (PFGE), arbitrarily primed PCR, and ribotyping (1, 10, 19). The first two methods suffer from a lack of interlaboratory reproducibility (5, 9); and furthermore, the present "gold standard" with the greatest discriminatory power, PFGE, is generally too costly and labor-intensive for routine clinical strain typing. On the other hand, automated ribotyping with the RiboPrinter (Qualicon, Wilmington, Del.) showed very high interlaboratory reproducibility (3) but suffered from a lack of discriminatory power when clinical *P. aeruginosa* strains were investigated (4).

Genetic markers called minisatellites or variable number of tandem repeats (VNTRs) were initially developed as the basis for fingerprinting the DNA of humans (11). More recently, minisatellites have been shown to exist in bacterial genomes as well, and the availability of whole-genome sequence data has opened the way to the systematic evaluation of tandem repeat polymorphisms (13, 24). When applicable, this method has been shown to fulfill most of the criteria required for an "ideal" typing system, including ease of use, speed, high discriminatory power, and reproducibility. The number of repeat units at each locus is usually estimated by measuring the sizes of the PCR products amplified with locus-specific primers flanking the repeat region. One kind of VNTR typing assay, often called multiple-locus VNTR analysis (MLVA), is based on a set of polymorphic tandem repeat loci. Approximately 20 loci are used for MLVA with *Bacillus anthracis* (13), *Yersinia pestis* (13), or the *Mycobacterium tuberculosis* complex (12).

In order to develop an MLVA scheme, one needs to identify polymorphic minisatellite loci, which must then be individually checked for variations of the repeat number among strains. The genome of *P. aeruginosa* strain PAO1 (20) is relatively rich in tandem repeats (13). We report on the identification of seven polymorphic loci and validation of their applicability in the MLVA scheme. Evaluation of these seven polymorphic

* Corresponding author. Mailing address: Institut de Génétique et Microbiologie, Université Paris Sud, 91405 Orsay cedex, France. Phone: 33 1 69 15 62 08. Fax: 33 1 69 15 66 78. E-mail: gilles.vergnaud@igmors.u-psud.fr.

† Present address: Biodiversity of Emerging Bacterial Pathogens Unit, Institut Pasteur, 75724 Paris Cedex 15, France.

loci allows a high degree of discrimination among strains with a high degree of reproducibility and easy scoring.

MATERIALS AND METHODS

Tandem repeat locus identification. The sequence data (20) for *P. aeruginosa* PAO1 were obtained from the World Wide Web (http://www.genome.pseudomonas.com) and were processed by using the Tandem Repeats Finder (TRF) software (http://c3.biomath.mssm.edu/trf.html). The output was then imported into a database accessible via the Internet (http://minisatellites.u-psud.fr), as described previously (13, 24). Tandem repeat loci are designated by using the nomenclature described previously (12); for instance, ms173-5186_243bp_14U is the tandem repeat locus named ms173, which appears at position 5186 kb in the PAO1 genome with a 243-bp repeat unit and 14 units in PAO1 (Table 1).

Isolates and DNA preparation. A total of 102 isolates were included in the present investigation as representatives of (i) all 53 ribogroups previously identified (one isolate per ribogroup) (4) and (ii) the geographic distribution of the two most frequent ribogroups found (87-S-3 [30 additional isolates] and 88-S-2 [19 additional isolates]) (4). The 203 isolates previously studied were clinical isolates collected from 20 European hospitals between 1997 and 1999 and had been analyzed by automated ribotyping (4). Fifty-three ribogroups (a ribogroup being defined as a set of strains showing the same ribotype; i.e., not a single band difference was detected among their profiles obtained by automated ribotyping with *Pvu*II) were initially identified among 203 isolates. Four ribogroups comprised approximately half of the isolates, with the two most frequent ones being ribotype 87-S-3, with 43 isolates, and ribotype 88-S-2, with 28 isolates.

The polymorphisms of candidate VNTR loci were initially evaluated by using a subset of 12 isolates from different ribogroups (see Fig. 2).

Isolates were grown at 37°C in Luria-Bertani broth. One milliliter of an overnight culture was centrifuged at 5,000 × *g* for 10 min. A QIAamp DNA mini kit (Qiagen, Hilden, Germany) was used for DNA extraction, as recommended by the manufacturer, with an extended lysis step (5 h at 55°C).

Minisatellite PCR amplification and genotyping. The PCR mixtures (15 μl) contained 1 ng of DNA, 1× *Taq* Reaction Buffer (Qbiogen, Illkirch, France), 1 U of *Taq* DNA polymerase (Qbiogen), 200 μM each deoxynucleoside triphosphate, and 0.3 μM each flanking primer (for primer sequences, see Table 1). PCRs were performed in an MJ Research PTC200 thermocycler. Initial denaturation at 96°C for 5 min was followed by 30 cycles of denaturation at 96°C for 20 s, annealing at 60°C for 30 s, and elongation at 65°C for 90 s. The final extension step was 5 min at 65°C. Different annealing temperatures and/or elongation times were used for ms173-5186_243bp and ms194-5915_12bp (for ms173, the annealing temperature was 64°C and the extension time was 5 min at 70°C; for ms194, the annealing temperature was 65°C and the extension time was 1 min at 70°C). Five microliters of each of the PCR products was run on standard 1% (ms173), 2% (ms142, ms172, ms194) or 3% (ms010, ms061, ms077, ms127) agarose gels (Qbiogen or ICN, Aurora, Ohio) in 0.5× TBE (Tris-borate-EDTA) buffer at a voltage of 10 V/cm. Gel runs (bromophenol blue position) of 20 cm (ms127, ms142), 30 cm (ms010, ms061, ms077, ms172), or 40 cm (ms173) were used according to the PCR product size and motif length. Gels were stained with ethidium bromide, visualized under UV light, and photographed (Vilber Lourmat, Marne la Vallée, France). The size markers used were a 100- or 20-bp ladder (Bio-Rad, Hercules, Calif.) or a 1-kb ladder plus (Gibco-BRL, Cergy Pontoise, France). Gel images were analyzed with the Bionumerics software package (version 3.0; Applied Maths, Sint-Martens-Latem, Belgium).

PFGE typing and analysis. DNA for PFGE was prepared as described previously (21) and digested with *Spe*I (New England Biolabs, Beverly, Mass.). Electrophoresis in a CHEF DRIII apparatus (Bio-Rad, Milano, Italy) was at 6 V/cm and 14°C for 21 h, with switching times linearly ramped from 5 to 23 s. The gels were stained with ethidium bromide and visualized under UV illumination with the E.A.S.Y. Win32 system (Herolab, Wiesloch, Germany).

Data analysis. Band size estimates were exported from the Bionumerics software and converted to numbers of units. The resulting data were imported back into Bionumerics software for use for clustering analysis with the categorical coefficient and Ward clustering parameter. Use of the categorical coefficient implies that the character states are considered unordered. The same weight is given to a large or a small number of differences in the number of repeats at any locus. The website used for identification (http://bacterial-genotyping.igmors.u-psud.fr/) was developed by using the BNserver application (version 3.0; Applied Maths).

RESULTS

Screening for informative and reliable VNTR loci in the *P. aeruginosa* genome. The TRF software (2) identified more than 3,000 tandem repeats within the *P. aeruginosa* PAO1 genome. Because no general rule has been described to predict tandem repeat polymorphisms directly from the sequence of a single allele (7, 13), present approaches rely upon systematic testing, especially for species in which only one genome has been sequenced (12). Since distinct alleles of VNTR loci with longer repeat units are generally easier to score on agarose gels, we decided to evaluate loci with repeat units more than 9 bp long. A total of 201 such tandem repeats with at least seven units each in the PAO1 sequence were identified with TRF software and were evaluated. Of these, 23 gave weak amplification signals with the set of primers used and were not considered further. One hundred seventy loci within the screening strain collection were monomorphic and were considered of very limited value for *P. aeruginosa* strain typing. Finally, only eight loci were polymorphic, defined here as showing at least two alleles among the screening strain collection (Table 1). Figure 1 shows examples of the PCR products obtained for the different loci. The number of alleles per locus ranged from 2 to 16. The two most polymorphic loci, at positions 0098 kb (ms010) and 1844 kb (ms061), varied by multiples of 6-bp units instead of the 12-bp repeat unit initially suggested by the database. The locus at position 5915 (ms194) varied by multiples of 12-bp units. Although the alleles at this locus were highly polymorphic, they could not be easily and reproducibly scored with the technology used here (ordinary agarose gel electrophoresis), given the range of allele sizes (600 to 700 bp), and this locus was therefore not characterized further.

MLVA of clinical *P. aeruginosa* isolates. MLVA analysis with the seven loci selected was then performed with the extended representative collection of 102 strains (Fig. 1 illustrates the setup of the MLVA assay for the seven loci). One or two loci of 13 strains (12%) could not be amplified (total number of missing PCR products, 17; data not shown). The PCR products could therefore be scored for all seven loci in 89 strains. The quality of the data produced was evaluated by retyping 10 coded samples. The correct MLVA genotype (Fig. 2) could be assigned to each coded sample. Figure 2 shows the results of clustering analysis for the 89 strains. PAO1 was included, based on the genome sequence data available and the estimated repeat numbers indicated in Table 1. MLVA discriminated 72 genotypes. This was higher than the number of ribotypes (46 in this set of strains) and serotypes (12 among the 68 typeable strains). Isolates with different ribotypes were all distinguished by MLVA, with only two exceptions: isolate 18A218 (ribotype 171-S-2) and isolate 18A403 (ribotype 172-S-8) from Spain (the ribotype patterns were very similar) (4). All isolates with identical ribotypes were clustered together, with only two exceptions: isolate 09A318 (ribotype 87-S-3) and isolate 12A241 (88-S-2). However, although strain 09A318 clustered in a distinct branch, at five of seven loci its MLVA genotype showed alleles that are generally encountered in the other strains of ribotype 87-S-3. On the contrary, strain 12A241 also had a discrepant serotype (serotype O1) compared to those of all other strains of ribotype 88-S-2 (serotype O11), so one can suspect a strain mix-up in this case. Ribotypes 87-S-3

TABLE 1. Characteristics of tandem repeat loci

Locus name[a]	Associated open reading frame	Motif length (bp)	No. of units in PAO1	% G+C content	% Conservation	Primer sequence[b]	Expected PCR product length in PAO1 (bp)	Estimated size range (bp)	No. of alleles	Polymorphism index
ms010-0098_6bp[c]	PA0081	6	11	65	100	L: GCAGGAACGCTTGCAGCAGGT R: CTTCGCCGACCCAGGGATCA	167	143–233	16	0.91
ms061-1844_6bp[a]	*pscP*	6	12	65	98	L: CTTGCCGTGCTACCGATCC R: CCCCCATGCCAGTTGC	127	85–139	10	0.87
ms077-2263_39bp[c]	*pcoA*	39	5	57	93	L: GCGTCATGGTCTGCATGTC R: TATACCCTCTTCGCCCAGTC	442	349–520	7	0.61
ms127-3496_15bp	PA3115	15	8	72	52	L: CTCGGAGTCTCTGCCAACTC R: GGCAGGACAGGATCTCGAC	210	210–225	2	0.45
ms142-3873_115bp	PA3463	115	7	66	94	L: AGCAGTGCCAGTTGATGTTG R: GTGGGGCGAAGGAGTGAG	890	200–775	7	0.68
ms172-5083_54bp	PA4541	54	12	64	72	L: GGATTCTCTCGCACGAGGT R: TACGTGACCTGACGTTGGTG	789	573–843	8	0.75
ms173-5186_243bp	PA4625	243	14	61	81	L: CTGCAGTTCGCGCAAGTC R: ATTTCAGCCAGCGTTACCAA	3,503	1,073–4,718	10	0.82
ms194-5915_12bp[c]	*algP*	12	45	70	64	L: CCTTAGGAGGCGCTGGTC R: AGCTGCTGGCAAGGCTCT	690	600–700	8	0.78

[a] Loci are listed according to their positions in the PAO1 genome. The proposed reference name includes the size of the repeat unit.
[b] L, left; R, right.
[c] The observed length variations do not fit with the repeat unit proposed in the minisatellite database but, rather, suggest a smaller (ms010, ms061, ms194) or a larger (ms077) unit for tandem repeat variation. This is due to the emergence of a new tandem repeat unit within a larger tandem repeat. The new unit sequence is derived from the preexisting repeat but has a different length. This was checked in particular for ms077 by sequencing the different alleles (data not shown; the data are available on request).

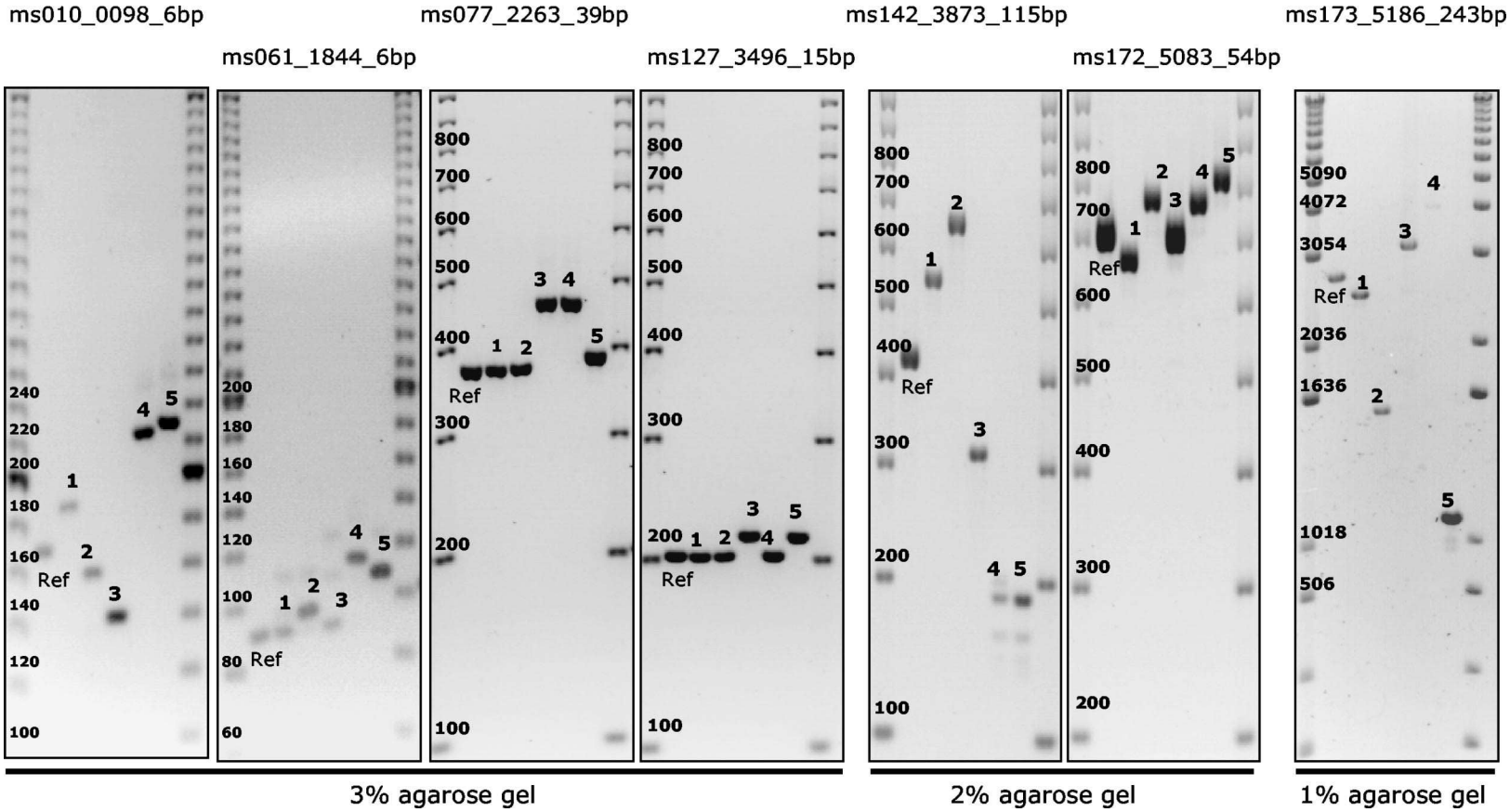


FIG. 1. MLVA setup on agarose gels. The usual setup for the running of MLVA on agarose gels is shown. Six DNA samples (including one reference DNA control lane on the left [with isolate 05A400; see Fig. 2]) are flanked by size markers (a 20-bp, 100-bp, or 1-kb ladder, according to the locus being typed). The experiment whose results are shown was part of the reproducibility test. Five blind-coded samples are numbered from 1 to 5 (strains 03D021, 04A036, 11C010, 22D032, and 18E049, respectively). The sizes (in base pairs) (indicated on the left of each gel) can be deduced by visual inspection of the patterns observed by taking into account the MLVA type for the reference strain (Ref) used here (isolate 05A400, genotype 12-6-3.5-8-3-11-11).

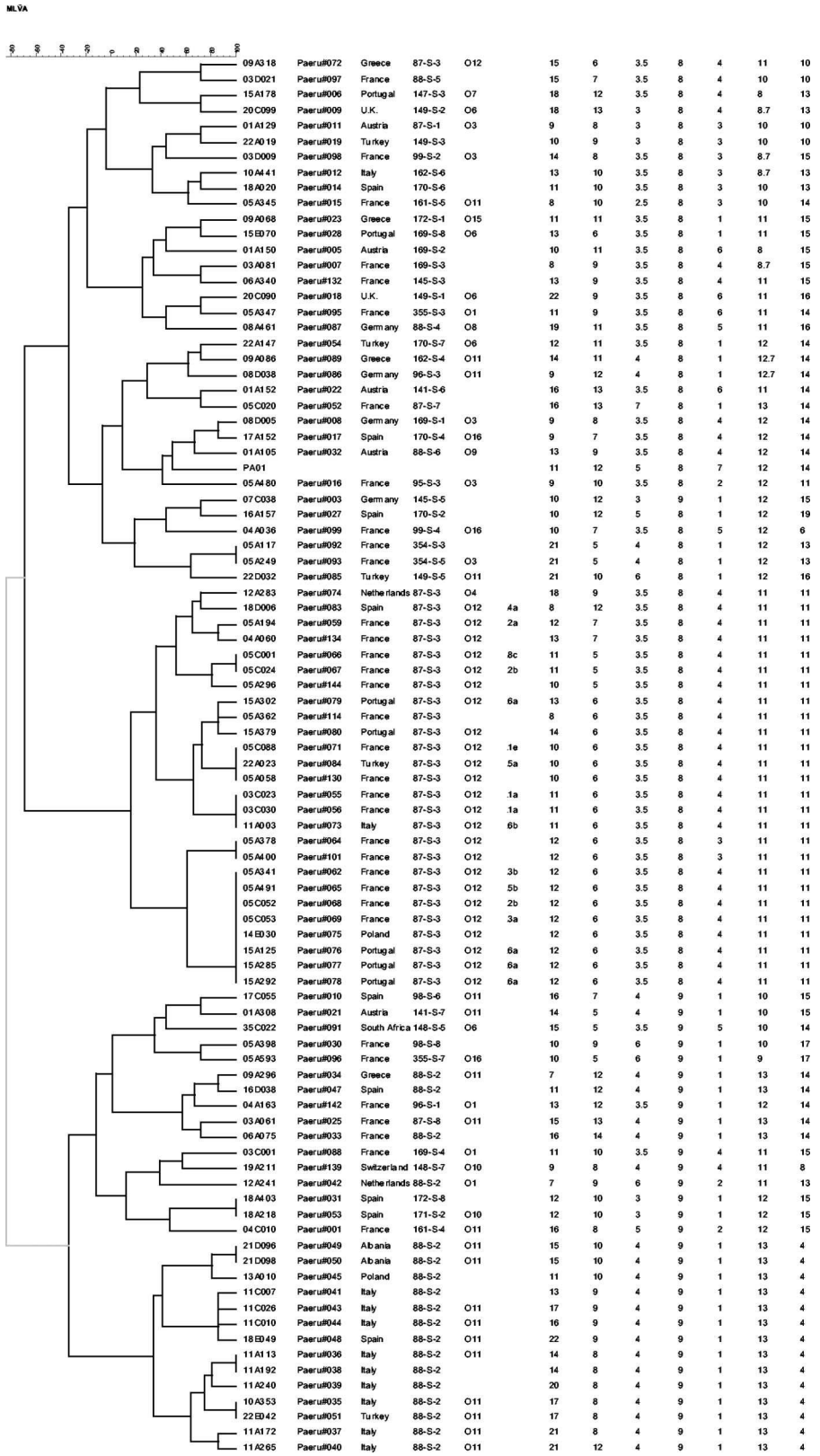


FIG. 2. Dendrogram deduced from the clustering analysis of the 90 strains (including strain PAO1). The first column on the left identifies the isolates (the screening set comprised strains 03D021, 15A178, 03D009, 09A068, 08A461, 08D005, 01A105, 04A036, 05A400, 35C022, 03C001, 19A211). The second column indicates the DNA batch, and the third column indicates the country of origin of each isolate. The fourth column indicates the ribotype, as reported previously (4). The fifth column contains the serogroup for isolates which could be typed by serotyping. The sixth column indicates the PFGE type and subtype of isolates tested by PFGE, while the last seven columns indicate the MLVA types (repeat copy number) at loci ms010, ms061, ms077, ms127, ms142, ms172, and ms173, respectively. The number of units is deduced from the sizes of the PCR products and was formally checked by sequencing only for the ms077 locus (data not shown).

(27 isolates) and 88-S-2 (18 isolates) were subdivided into 14 and 15 MLVA types, respectively. The subtyping results were in good agreement with the geographic origins of the strains, as isolates with the same MLVA genotype most often originated from the same center (indicated by the two first numbers of the isolate code in Fig. 2).

In order to compare the discriminatory power of MLVA with that of the present gold standard, 17 ribotype 87-S-3 isolates were also typed by PFGE. PFGE resulted in seven distinct types, which were further distinguished into 12 subtypes (indicated in Fig. 2 by Arabic numbers and lowercase letters), according to published criteria (22); among the isolates in this set, seven types were also distinguished by MLVA.

The results of a comparison of the MLVA genotypes with the serotyping data were in agreement with the single origin of serotype O12 postulated earlier (18) as well as with the greater genotypic diversity of serotype O11 isolates (21).

The MLVA genotypes of the strains studied here were loaded onto our web server. They can be accessed and compared with the MLVA genotypes of new, unknown strains. Identification queries can be run from the strain identification page (http://bacterial-genotyping.igmors.u-psud.fr/), as described elsewhere (12).

DISCUSSION

Although the *P. aeruginosa* genome is relatively rich in tandem repeats, only 8 of the 201 tandem repeats tested by the protocol described here proved to be polymorphic. This is in contrast to the polymorphism found in other bacterial species, including some with low levels of genetic polymorphism at housekeeping loci, such as *B. anthracis*, *Y. pestis*, and *M. tuberculosis* (12, 13). It is very unlikely that the low levels of diversity of VNTR loci found in the present sample are due to the screening strain collection chosen, since these 12 strains could all be distinguished by ribotyping and had originated from seven different countries. Our results show that minisatellite loci are more stable in *P. aeruginosa* than in other species with lower overall population genetic diversity. This suggests distinct evolutionary mechanisms for tandem repeats, such as various levels of slipped strand mispairing and repair during replication (15). The present investigation evaluated all tandem repeats with a repeat unit of 9 bp or more and at least 7 units. Additional polymorphic markers could be identified by using different queries. For instance, querying of the *P. aeruginosa* tandem repeat database (http://minisatellites.u-psud.fr) for tandem repeats with an internal conservation of at least 90% identifies 50 such loci, none of which was investigated here. It is likely that more polymorphic loci may be obtained if loci with unit lengths smaller than 9 bp (the threshold used in our screening procedure) are chosen, since in the present study the smallest units were the most polymorphic (Table 1).

One of the eight polymorphic markers identified, ms194 (position 5915), was not used in the final analysis because of the difficulty of scoring the alleles by simple agarose gel electrophoresis. Preliminary sequencing results showed that the level of sequence diversity at this locus is very high and that alleles containing an identical number of repeat units can have distinct sequences (data not shown). This tandem repeat is located within *algP*, a gene implicated in the regulation of mucoidy in *P. aeruginosa*, and its associated polymorphism was first described by Deretic and Konyecsni (8). The use of another approach, such as polyacrylamide gel or capillary electrophoresis, which offer higher degrees of resolution, would probably solve this issue, albeit at a higher overall cost.

One (and sometimes two) of the seven loci of 12% of the strains typed failed to be amplified, despite multiple amplification attempts. This may reflect the fact either that the corresponding locus is missing or that sequence divergence results in mispriming. Further investigations will be required (including tests with new primer pairs) before this lack of amplification can be used as additional data. In the course of this preliminary investigation, the corresponding strains were not included in the final analysis (Fig. 2).

The collection of isolates tested here was originally assembled with two distinct but complementary aims: first, to be representative of European clinical *P. aeruginosa* isolates, and second, to include geographically diverse strains from among the two most frequent ribotypes in order to check if MLVA could subdivide these two groups of isolates. The validity of MLVA for clustering analysis and evaluation of the phylogenetic relationships among strains is not yet formally established. In a recent work, Le Flèche et al. (12) empirically selected clustering parameters to analyze strains from the *M. tuberculosis* complex. This was made possible by the extensive knowledge of the evolutionary relationships and the underlying epidemiology independently generated with other markers to distinguish *M. tuberculosis* complex strains. The same parameters have been applied in the present study. The clustering proposed here (Fig. 2) shows similarities with the clustering reported previously, which was based on ribotyping: isolates not distinguished by ribotyping are also generally clustered by MLVA (Fig. 2). This suggests that MLVA does retain some amount of phylogenetic information which can be used to trace the evolutionary histories and relationships of genotypes, as has also been found for other organisms (12). Exceptions to this general rule are easily explained by the fact that cluster analysis is based on only a few informative characters, and therefore, a difference at a single locus can alter the positioning of strains in the dendrogram. Conversely, alleles of different origins may be of identical size (homoplasy), which will also alter the clustering analysis. Allele sequencing, at least with a representative strain collection, may eventually help correct some of these inconsistencies. In any case, for more distantly related genotypes, the relationships depicted by clustering analysis may become increasingly less meaningful, especially in light of the relatively frequent occurrence of horizontal transfer among *P. aeruginosa* strains (6, 16, 17), which can obscure the evidence of a common ancestral lineage among strains.

ACKNOWLEDGMENTS

Work on human bacterial pathogen identification is supported by grants from the Délégation Générale de l'Armement to L.O. and G.V. PFGE typing was performed in the framework of the Genetic Epidemiology Network for Europe (GENE) project (contract QLK2-2000-01404) of the Fifth Framework Program of the European Union.

P.T.T. is grateful to Georgia Diamantopoulou for excellent technical assistance with PFGE. S.B. thanks Jan Verhoef (Utrecht University) for continuous support.

REFERENCES

1. **Bennekov, T., H. Colding, B. Ojeniyi, M. W. Bentzon, and N. Hoiby.** 1996. Comparison of ribotyping and genome fingerprinting of *Pseudomonas aeruginosa* isolates from cystic fibrosis patients. J. Clin. Microbiol. **34:**202–204.
2. **Benson, G.** 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. **27:**573–580.
3. **Brisse, S., V. Fussing, B. Ridwan, J. Verhoef, and R. J. Willems.** 2002. Automated ribotyping of vancomycin-resistant *Enterococcus faecium* isolates. J. Clin. Microbiol. **40:**1977–1984.
4. **Brisse, S., D. Milatovic, A. C. Fluit, K. Kusters, A. Toelstra, J. Verhoef, and F. J. Schmitz.** 2000. Molecular surveillance of European quinolone-resistant clinical isolates of *Pseudomonas aeruginosa* and *Acinetobacter* spp. using automated ribotyping. J. Clin. Microbiol. **38:**3636–3645.
5. **Dabrowski, W., U. Czekajlo-Kolodziej, D. Medrala, and S. Giedrys-Kalemba.** 2003. Optimisation of AP-PCR fingerprinting discriminatory power for clinical isolates of Pseudomonas aeruginosa. FEMS Microbiol. Lett. **218:**51–57.
6. **Denamur, E., B. Picard, G. Decoux, J. B. Denis, and J. Elion.** 1993. The absence of correlation between allozyme and rrn RFLP analysis indicates a high gene flow rate within human clinical *Pseudomonas aeruginosa* isolates. FEMS Microbiol. Lett. **110:**275–280.
7. **Denoeud, F., G. Vergnaud, and G. Benson.** 2003. Predicting human minisatellite polymorphism. Genome Res. **13:**856–867.
8. **Deretic, V., and W. M. Konyecsni.** 1990. A procaryotic regulatory factor with a histone H1-like carboxy-terminal domain: clonal variation of repeats within *algP*, a gene involved in regulation of mucoidy in *Pseudomonas aeruginosa.* J. Bacteriol. **172:**5544–5554.
9. **Foissaud, V., J. M. Puyhardy, J. C. Chapalain, H. Salord, J. J. Depina, M. Morillon, P. Nicolas, and J. D. Perrier-Gros-Claude.** 1999. Inter-laboratory reproducibility of pulsed-field electrophoresis for the study of 12 types of *Pseudomonas aeruginosa.* Pathol Biol (Paris) **47:**1053–1059. (In French.)
10. **Grundmann, H., C. Schneider, D. Hartung, F. D. Daschner, and T. L. Pitt.** 1995. Discriminatory power of three DNA-based typing techniques for *Pseudomonas aeruginosa.* J. Clin. Microbiol. **33:**528–534.
11. **Jeffreys, A. J., V. Wilson, and S. L. Thein.** 1985. Individual-specific 'fingerprints' of human DNA. Nature **316:**76–79.
12. **Le Flèche, P., M. Fabre, F. Denoeud, J. L. Koeck, and G. Vergnaud.** 2002. High resolution, on-line identification of strains from the *Mycobacterium tuberculosis* complex based on tandem repeat typing. BMC Microbiol. **2:**37.
13. **Le Flèche, P., Y. Hauck, L. Onteniente, A. Prieur, F. Denoeud, V. Ramisse, P. Sylvestre, G. Benson, F. Ramisse, and G. Vergnaud.** 2001. A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis.* BMC Microbiol. **1:**2.
14. **Livermore, D. M.** 2002. Multiple mechanisms of antimicrobial resistance in *Pseudomonas aeruginosa*: our worst nightmare? Clin. Infect. Dis. **34:**634–640.
15. **Lopes, J., H. Debrauwere, J. Buard, and A. Nicolas.** 2002. Instability of the human minisatellite CEB1 in rad27Delta and dna2–1 replication-deficient yeast cells. EMBO J. **21:**3201–3211.
16. **Picard, B., E. Denamur, A. Barakat, J. Elion, and P. Goullet.** 1994. Genetic heterogeneity of *Pseudomonas aeruginosa* clinical isolates revealed by esterase electrophoretic polymorphism and restriction fragment length polymorphism of the ribosomal RNA gene region. J. Med. Microbiol. **40:**313–322.
17. **Pirnay, J. P., D. De Vos, C. Cochez, F. Bilocq, A. Vanderkelen, M. Zizi, B. Ghysels, and P. Cornelis.** 2002. *Pseudomonas aeruginosa* displays an epidemic population structure. Environ. Microbiol. **4:**898–911.
18. **Pitt, T. L., D. M. Livermore, D. Pitcher, A. C. Vatopoulos, and N. J. Legakis.** 1989. Multiresistant serotype O 12 *Pseudomonas aeruginosa*: evidence for a common strain in Europe. Epidemiol. Infect. **103:**565–576.
19. **Renders, N., Y. Romling, H. Verbrugh, and A. van Belkum.** 1996. Comparative typing of *Pseudomonas aeruginosa* by random amplification of polymorphic DNA or pulsed-field gel electrophoresis of DNA macrorestriction fragments. J. Clin. Microbiol. **34:**3190–3195.
20. **Stover, C. K., X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrener, M. J. Hickey, F. S. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrock-Wadman, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu, I. T. Paulsen, J. Reizer, M. H. Saier, R. E. Hancock, S. Lory, and M. V. Olson.** 2000. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. Nature **406:**959–964.
21. **Tassios, P. T., V. Gennimata, A. N. Maniatis, C. Fock, N. J. Legakis, and the Greek** *Pseudomonas aeruginosa* Study Group. 1998. Emergence of multidrug resistance in ubiquitous and dominant *Pseudomonas aeruginosa* serogroup O:11. J. Clin. Microbiol. **36:**897–901.
22. **Tenover, F. C., R. D. Arbeit, R. V. Goering, P. A. Mickelsen, B. E. Murray, D. H. Persing, and B. Swaminathan.** 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J. Clin. Microbiol. **33:**2233–2239.
23. **van Belkum, A., M. Struelens, A. de Visser, H. Verbrugh, and M. Tibayrenc.** 2001. Role of genomic typing in taxonomy, evolutionary genetics, and microbial epidemiology. Clin. Microbiol. Rev. **14:**547–560.
24. **Vergnaud, G., and F. Denoeud.** 2000. Minisatellites: mutability and genome architecture. Genome Res. **10:**899–907.

3.2.1.3 Caractéristiques des répétitions en tandem polymorphes

3.2.1.3.1 Longueur des motifs répétés

Pour la moitié des 8 locus polymorphes, les différences de taille entre les allèles observés expérimentalement ne correspondent pas à la taille du motif sélectionné dans la base de données (critères de choix de la base : longueur totale de la répétition puis conservation du motif). En soumettant ces séquences au logiciel TRF, nous avons constaté qu'il y avait plusieurs répétitions possibles pour ces locus, comme le résume le Tableau 8.

Tableau 8 : Caractéristiques des répétitions redondantes pour 4 locus chez *P. aeruginosa*

Répétitions en tandem ::		U : taille du motif (pb)	N : nombre de répétitions dans PAO1	V : conservation par rapport au consensus (%)	% G+C
ms010_0098	base de données	12	14	67	71
	observation expérimentale	**6**	**11**	**100**	**65**
ms061_1844	base de données	12	16	64	73
	observation expérimentale	**6**	**12**	**98**	**65**
ms077_2263	base de données	15	18	50	63
	observation expérimentale	**39**	**4**	**93**	**57**
ms194_5915	base de données	75	8	88	71
	observation expérimentale	**12**	**45**	**64**	**70**

Nous avons constaté pour trois de ces locus (ms010, ms061 et ms077) que les variations de taille observées correspondaient au motif répété le mieux conservé. Or la conservation moyenne des motifs dans les 201 répétitions en tandem est faible (56%) et elle ne reflète pas la conservation moyenne de 70% observée pour les 3225 répétitions totales du génome PAO1 à partir des données de séquençage. Nous n'avions pas fait de requête selon le critère de conservation des motifs.

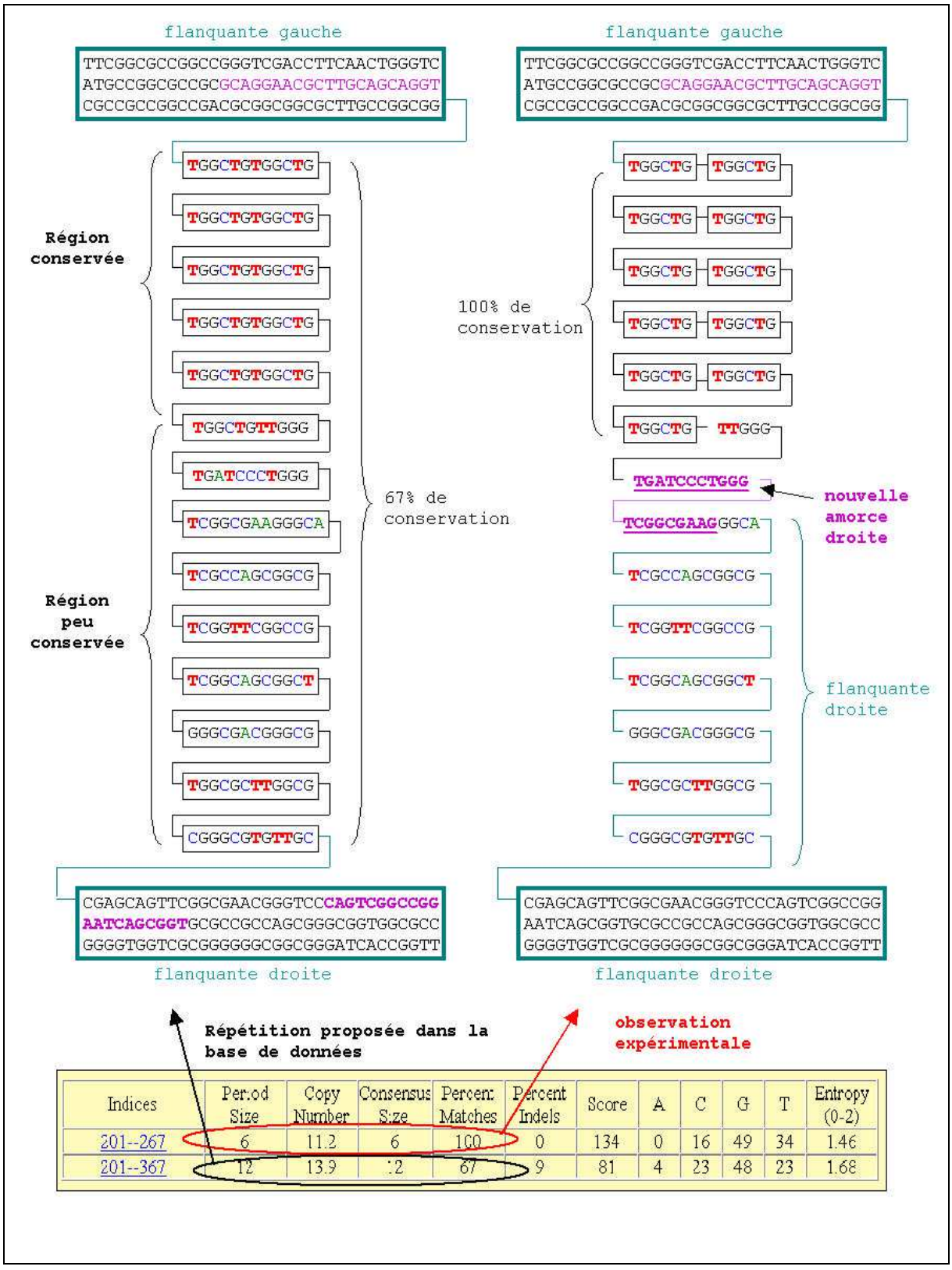


Figure 20 : Exemple de répétitions redondantes détectées par le TRF au locus ms010.

La Figure 20 illustre deux représentations de la répétition en tandem détectée par « TRF » dans la séquence de ms010, avec les deux tailles de motifs possibles, et des conservations très

différentes. En regardant plus en détail les séquences répétées proposées par le logiciel TRF, nous avons constaté par exemple pour ms010 que la répétition de 14 fois 12 pb est constituée d'une première série de motifs de 6 pb parfaitement conservés puis, dans la deuxième moitié, d'une série de motifs moins bien conservés (Figure 20). Un microsatellite semble donc avoir émergé à l'intérieur d'une répétition en tandem plus ancienne. Nous avons sélectionné une autre amorce droite dans la partie peu conservée de la séquence pour obtenir des produits de PCR relativement petits et ainsi pouvoir observer sur gel d'agarose les variations de taille de 6 pb.

En ce qui concerne ms194, deux tailles de motif sont détectées par le logiciel TRF, l'une de 75 pb et l'autre de 12 pb, moins bien conservée. Les différences de taille des allèles observées sur gel ne correspondaient pas à des tailles multiples de 75 pb, mais à une différence beaucoup plus petite. Pour confirmer cette observation et écarter l'hypothèse d'anomalies de migration, nous avons séquencé quelques allèles de ms194 (voir paragraphe 3.2.4).

3.2.1.3.2 Polymorphisme

Le nombre de répétitions en tandem polymorphes est faible, avec seulement 8 répétitions polymorphes pour la collection de souches testées (soit 4.5% des 178 répétitions en tandem).

Une étude récente des critères prédictifs du polymorphisme des répétitions en tandem chez l'homme a montré une corrélation entre le pourcentage en GC et le polymorphisme (Denoeud 2003). Par ailleurs, l'étude des répétitions en tandem dans le génome de *Yersinia pestis* a montré une corrélation entre la conservation des motifs et le polymorphisme. Chez *Bacillus anthracis*, les deux critères prédictifs du polymorphisme observés sont la longueur totale de la répétition et le pourcentage en bases G et C (GC%) (Le Flèche 2001).

Il n'apparaît pas de corrélation entre le nombre d'allèles observés et la longueur totale de la répétition, ou le pourcentage en GC. Ces deux paramètres ne semblent pas être des critères prédictifs du polymorphisme des répétitions en tandem chez *P. aeruginosa*.

En revanche, la Figure 21 montre *a posteriori* que le critère de conservation de la répétition aurait pu être un critère (imparfaitement) prédictif du polymorphisme des répétitions en tandem de *P. aeruginosa*.

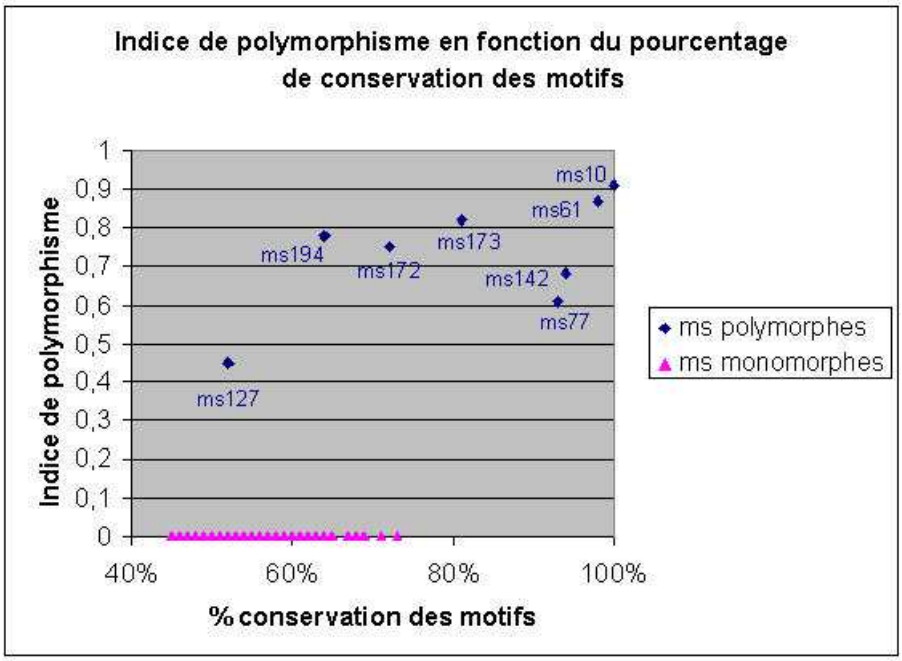


Figure 21 : Indice de polymorphisme en fonction du pourcentage de conservation des motifs chez *P. aeruginosa*

3.2.2 Stabilité des 8 répétitions en tandem polymorphes chez *P. aeruginosa*

La stabilité des répétitions en tandem polymorphes identifiées chez *P. aeruginosa* a été testée au cours de la croissance bactérienne sur une période de 3 semaines. En effet, il est important de savoir si les marqueurs proposés dans cette étude ont une valeur épidémiologique au moins à court terme, ou s'ils présentent une instabilité, en particulier en ce qui concerne les petits motifs de 6 pb qui pourraient être des locus de contingence.

Pour cela, six souches de *P. aeruginosa* ont été mise en culture liquide (LB) et diluées quotidiennement. Des mesures de densité optique à 600nm ont été prises à chaque nouvelle dilution pour pouvoir calculer le nombre total de générations qu'ont subi les cultures après 3 semaines. Les cultures atteignent la phase stationnaire au bout de 500 minutes environ, et restent en phase stationnaire pendant plusieurs heures jusqu'à la dilution suivante. Pendant la phase stationnaire, des remaniements chromosomiques peuvent se produire. Le nombre de générations n'est donc pas le seul critère à considérer. Des courbes de croissance des différentes souches ont également été réalisées pour vérifier si celles-ci présentent des différences de temps de génération. Les 6 souches ont des courbes de croissance très semblables (temps de génération de 30 minutes environ). Le nombre de générations a donc été calculé à partir des données d'une seule des six souches.

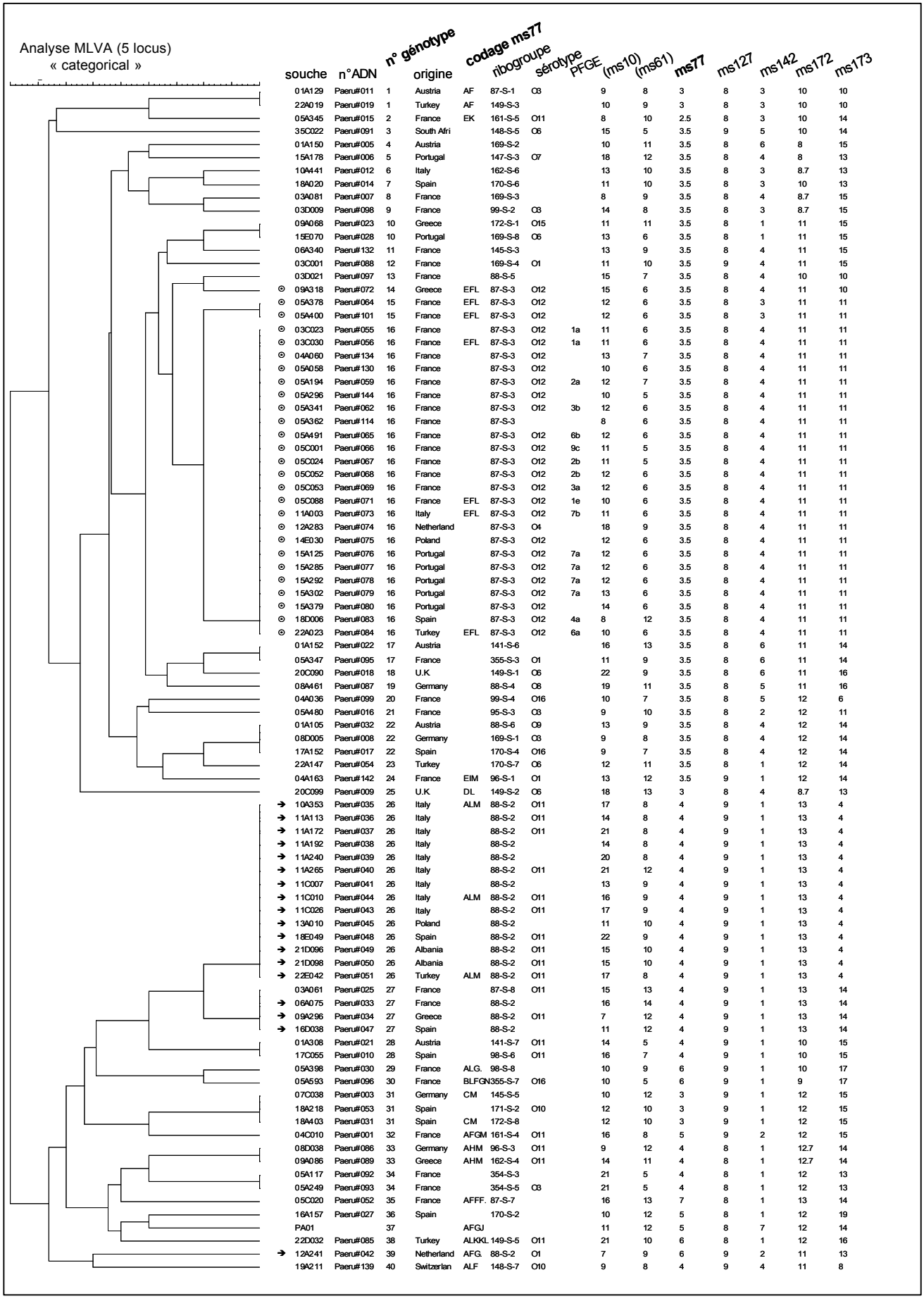
Des PCR ont été réalisées sur les échantillons prélevés au début de l'expérience et après les 3 semaines de dilutions. Les tailles des allèles observés pour les 8 répétitions en tandem de l'étude MLVA sont identiques avant et après les 3 semaines de culture. On peut simplement conclure à une stabilité de ces marqueurs à court terme. Pour pouvoir utiliser ces marqueurs pour des études épidémiologiques plus globales, il faudrait faire l'expérience sur un plus grand nombre de générations et séquencer les allèles.

Le nombre de générations est de 200 environ, ce qui est peu. Les cellules ont passé l'essentiel de la durée de l'expérience en phase stationnaire. Pour tester un plus grand nombre de générations, il serait préférable d'utiliser un chemostat dans lequel le milieu est renouvelé de façon régulière et permet d'entretenir sur un grand nombre de générations une culture bactérienne.

3.2.3 Séquençage de deux répétitions en tandem : ms77 et ms194

3.2.3.1 Séquençage de la répétition ms77

La Figure 2 de l'article sur le typage par polymorphisme de répétitions en tandem appliqué à *P. aeruginosa* illustre à quel point différentes répétitions en tandem peuvent avoir des comportements différents. Les deux marqueurs ms10 et ms61 à motif de 6 paires de bases s'apparentent plus à des microsatellites qu'à des minisatellites. Leur degré de polymorphisme est très élevé, et on observe que des souches ayant des allèles de même taille peuvent être placées en des positions très éloignées de l'arbre, et réciproquement. Ces marqueurs sont très utiles en épidémiologie locale (étude d'une épidémie) mais leur valeur phylogénétique pour des études plus globales est sans doute limitée. La Figure 22 présente une analyse des données faite en ne tenant pas compte de ms10 et ms61. Seulement 40 génotypes sont distingués.



souche	n°ADN	n° génotype	origine		codage ms77 ribogroupe	sérotype	PFGE	(ms10)	(ms61)	ms77	ms127	ms142	ms172	ms173
01A129	Paeru#011	1	Austria	AF	87-S-1	O3		9	8	3	8	3	10	10
22A019	Paeru#019	1	Turkey	AF	149-S-3			10	9	3	8	3	10	10
05A345	Paeru#015	2	France	EK	161-S-5	O11		8	10	2.5	8	3	10	14
35C022	Paeru#091	3	South Afri		148-S-5	O6		15	5	3.5	9	5	10	14
01A150	Paeru#005	4	Austria		169-S-2			10	11	3.5	8	6	8	15
15A178	Paeru#006	5	Portugal		147-S-3	O7		18	12	3.5	8	4	8	13
10A441	Paeru#012	6	Italy		162-S-6			13	10	3.5	8	3	8.7	13
18A020	Paeru#014	7	Spain		170-S-6			11	10	3.5	8	3	10	13
03A081	Paeru#007	8	France		169-S-3			8	9	3.5	8	4	8.7	15
03D009	Paeru#098	9	France		99-S-2	O3		14	8	3.5	8	3	8.7	15
09A068	Paeru#023	10	Greece		172-S-1	O15		11	11	3.5	8	1	11	15
15E070	Paeru#028	10	Portugal		169-S-8	O6		13	6	3.5	8	1	11	15
06A340	Paeru#132	11	France		145-S-3			13	9	3.5	8	4	11	15
03C001	Paeru#088	12	France		169-S-4	O1		11	10	3.5	9	4	11	15
03D021	Paeru#097	13	France		88-S-5			15	7	3.5	8	4	10	10
09A318	Paeru#072	14	Greece	EFL	87-S-3	O12		15	6	3.5	8	4	11	10
05A378	Paeru#064	15	France	EFL	87-S-3	O12		12	6	3.5	8	3	11	11
05A400	Paeru#101	15	France	EFL	87-S-3	O12		12	6	3.5	8	3	11	11
03C023	Paeru#055	16	France		87-S-3	O12	1a	11	6	3.5	8	4	11	11
03C030	Paeru#056	16	France	EFL	87-S-3	O12	1a	11	6	3.5	8	4	11	11
04A060	Paeru#134	16	France		87-S-3	O12		13	7	3.5	8	4	11	11
05A058	Paeru#130	16	France		87-S-3	O12		10	6	3.5	8	4	11	11
05A194	Paeru#059	16	France		87-S-3	O12	2a	12	7	3.5	8	4	11	11
05A296	Paeru#144	16	France		87-S-3	O12		10	5	3.5	8	4	11	11
05A341	Paeru#062	16	France		87-S-3	O12	3b	12	6	3.5	8	4	11	11
05A362	Paeru#114	16	France		87-S-3			8	6	3.5	8	4	11	11
05A491	Paeru#065	16	France		87-S-3	O12	6b	12	6	3.5	8	4	11	11
05C001	Paeru#066	16	France		87-S-3	O12	9c	11	5	3.5	8	4	11	11
05C024	Paeru#067	16	France		87-S-3	O12	2b	11	5	3.5	8	4	11	11
05C052	Paeru#068	16	France		87-S-3	O12	2b	12	6	3.5	8	4	11	11
05C053	Paeru#069	16	France		87-S-3	O12	3a	12	6	3.5	8	4	11	11
05C088	Paeru#071	16	France	EFL	87-S-3	O12	1e	10	6	3.5	8	4	11	11
11A003	Paeru#073	16	Italy	EFL	87-S-3	O12	7b	11	6	3.5	8	4	11	11
12A283	Paeru#074	16	Netherland		87-S-3	O4		18	9	3.5	8	4	11	11
14E030	Paeru#075	16	Poland		87-S-3	O12		12	6	3.5	8	4	11	11
15A125	Paeru#076	16	Portugal		87-S-3	O12	7a	12	6	3.5	8	4	11	11
15A285	Paeru#077	16	Portugal		87-S-3	O12	7a	12	6	3.5	8	4	11	11
15A292	Paeru#078	16	Portugal		87-S-3	O12	7a	12	6	3.5	8	4	11	11
15A302	Paeru#079	16	Portugal		87-S-3	O12	7a	13	6	3.5	8	4	11	11
15A379	Paeru#080	16	Portugal		87-S-3	O12		14	6	3.5	8	4	11	11
18D006	Paeru#083	16	Spain		87-S-3	O12	4a	8	12	3.5	8	4	11	11
22A023	Paeru#084	16	Turkey	EFL	87-S-3	O12	6a	10	6	3.5	8	4	11	11
01A152	Paeru#022	17	Austria		141-S-6			16	13	3.5	8	6	11	14
05A347	Paeru#095	17	France		355-S-3	O1		11	9	3.5	8	6	11	14
20C090	Paeru#018	18	U.K		149-S-1	O6		22	9	3.5	8	6	11	16
08A461	Paeru#087	19	Germany		88-S-4	O8		19	11	3.5	8	5	11	16
04A036	Paeru#099	20	France		99-S-4	O16		10	7	3.5	8	5	12	6
05A480	Paeru#016	21	France		95-S-3	O3		9	10	3.5	8	2	12	11
01A105	Paeru#032	22	Austria		88-S-6	O9		13	9	3.5	8	4	12	14
08D005	Paeru#008	22	Germany		169-S-1	O3		9	8	3.5	8	4	12	14
17A152	Paeru#017	22	Spain		170-S-4	O16		9	7	3.5	8	4	12	14
22A147	Paeru#054	23	Turkey		170-S-7	O6		12	11	3.5	8	1	12	14
04A163	Paeru#142	24	France	EIM	96-S-1	O1		13	12	3.5	9	1	12	14
20C099	Paeru#009	25	U.K	DL	149-S-2	O6		18	13	3	8	4	8.7	13
10A353	Paeru#035	26	Italy	ALM	88-S-2	O11		17	8	4	9	1	13	4
11A113	Paeru#036	26	Italy		88-S-2	O11		14	8	4	9	1	13	4
11A172	Paeru#037	26	Italy		88-S-2	O11		21	8	4	9	1	13	4
11A192	Paeru#038	26	Italy		88-S-2			14	8	4	9	1	13	4
11A240	Paeru#039	26	Italy		88-S-2			20	8	4	9	1	13	4
11A265	Paeru#040	26	Italy		88-S-2	O11		21	12	4	9	1	13	4
11C007	Paeru#041	26	Italy		88-S-2			13	9	4	9	1	13	4
11C010	Paeru#044	26	Italy	ALM	88-S-2	O11		16	9	4	9	1	13	4
11C026	Paeru#043	26	Italy		88-S-2	O11		17	9	4	9	1	13	4
13A010	Paeru#045	26	Poland		88-S-2			11	10	4	9	1	13	4
18E049	Paeru#048	26	Spain		88-S-2	O11		22	9	4	9	1	13	4
21D096	Paeru#049	26	Albania		88-S-2	O11		15	10	4	9	1	13	4
21D098	Paeru#050	26	Albania		88-S-2	O11		15	10	4	9	1	13	4
22E042	Paeru#051	26	Turkey	ALM	88-S-2	O11		17	8	4	9	1	13	4
03A061	Paeru#025	27	France		87-S-8	O11		15	13	4	9	1	13	14
06A075	Paeru#033	27	France		88-S-2			16	14	4	9	1	13	14
09A296	Paeru#034	27	Greece		88-S-2	O11		7	12	4	9	1	13	14
16D038	Paeru#047	27	Spain		88-S-2			11	12	4	9	1	13	14
01A308	Paeru#021	28	Austria		141-S-7	O11		14	5	4	9	1	10	15
17C055	Paeru#010	28	Spain		98-S-6	O11		16	7	4	9	1	10	15
05A398	Paeru#030	29	France	ALG.	98-S-8			10	9	6	9	1	10	17
05A593	Paeru#096	30	France	BLFGN	355-S-7	O16		10	5	6	9	1	9	17
07C038	Paeru#003	31	Germany	CM	145-S-5			10	12	3	9	1	12	15
18A218	Paeru#053	31	Spain		171-S-2	O10		12	10	3	9	1	12	15
18A403	Paeru#031	31	Spain	CM	172-S-8			12	10	3	9	1	12	15
04C010	Paeru#001	32	France	AFGM	161-S-4	O11		16	8	5	9	2	12	15
08D038	Paeru#086	33	Germany	AHM	96-S-3	O11		9	12	4	8	1	12.7	14
09A086	Paeru#089	33	Greece	AHM	162-S-4	O11		14	11	4	8	1	12.7	14
05A117	Paeru#092	34	France		354-S-3			21	5	4	8	1	12	13
05A249	Paeru#093	34	France		354-S-5	O3		21	5	4	8	1	12	13
05C020	Paeru#052	35	France	AFFF.	87-S-7			16	13	7	8	1	13	14
16A157	Paeru#027	36	Spain		170-S-2			10	12	5	8	1	12	19
PA01		37		AFGJ				11	12	5	8	7	12	14
22D032	Paeru#085	38	Turkey	ALKKL	149-S-5	O11		21	10	6	8	1	12	16
12A241	Paeru#042	39	Netherland	AFG.	88-S-2	O1		7	9	6	9	2	11	13
19A211	Paeru#139	40	Switzerlan	ALF	148-S-7	O10		9	8	4	9	4	11	8

Figure 22 : Arbre MLVA (ms77, ms127, ms142, ms172, ms173) des 90 souches *P. aeruginosa*

En comparaison avec ms10 et ms61, ms127 est un marqueur qui sépare très clairement les deux groupes majeurs qui ont été identifiés, le groupe contenant le ribotype 88-S-2 d'une part (allèle à 9 motifs) et le groupe contenant le ribotype 87-S-3 d'autre part (allèle à 8 motifs).

Les autres marqueurs ont un comportement intermédiaire, et montrent quelques anomalies apparentes. Par exemple pour ms077 un allèle à six unités est observé dans 4 souches, positionnées en différentes parties de l'arbre par l'analyse MLVA. Ce genre d'observation peut refléter des événements d'homoplasie (deux allèles paraissent identiques sans être de même origine phylogénétique, par coïncidence fortuite de taille en l'occurrence) ou bien trahir l'existence d'évènements plus complexes tels que du transfert horizontal entre différentes souches (ce qui réduirait fortement la valeur phylogénétique de l'approche MLVA, mais sans diminuer pour autant la valeur « identification de souche »).

Afin d'aborder ces points, j'ai réalisé le séquençage de certains allèles ms077. Ce locus qui correspond au gène *pcoA* présente l'avantage de varier dans une plage de taille bien adaptée au séquençage. La séquence nucléique de vingt-sept allèles différents a été établie : l'allèle codé 7 unités dans l'article, les 4 allèles codés 6, 2 allèles codés 5, 6 allèles codés 4, 8 allèles codés 3,5, 5 allèles codés 3, et un allèle codé 2,5. Le séquençage permet en réalité de distinguer 15 allèles (au lieu de 7 par la seule mesure de taille). Des allèles de taille égale mais placés à différents endroits de l'arbre sont effectivement différents ; l'égalité de taille est une coïncidence (homoplasie). Dans le Tableau 9 de codage des allèles, le dernier motif, invariant, est omis, il y a donc une lettre de moins que de motifs annoncés.

Tableau 9 : Codage des allèles ms77

allèles ms77	codage des motifs
05C020_520bp_7U	AFFFGM
22D032_481bp_6U	ALKKL
12A241_481bp_6U	AFGMN
05A398_481bp_6U	ALGMN
05A593_481bp_6U	BLFGN
04C010_442bp_5U	AFGM
PAO1_77_442bp_5U	**AFGJ**
19A211_403bp_4U	ALF
10A353_403bp_4U	ALM
22E042_403bp_4U	ALM
11C010_403bp_4U	ALM
08D038_403bp_4U	AHM
09A086_403bp_4U	AHM

allèles ms77	codage des motifs
04A163_388bp_3.5U	EIM
05A400_388bp_3.5U	EFL
03C030_388bp_3.5U	EFL
05A378_388bp_3.5U	EFL
11A003_388bp_3.5U	EFL
22A023_388bp_3.5U	EFL
09A318_388bp_3.5U	EFL
05C088_388bp_3.5U	EFL
01A129_364bp_3U	AF
22A019_364bp_3U	AF
18A403_364bp_3U	CM
07C038_364bp_3U	CM
20C099_364bp_3U	DL
05A345_349bp_2.5U	EK

La Figure 23 présente le dictionnaire des motifs du minisatellite ms77.

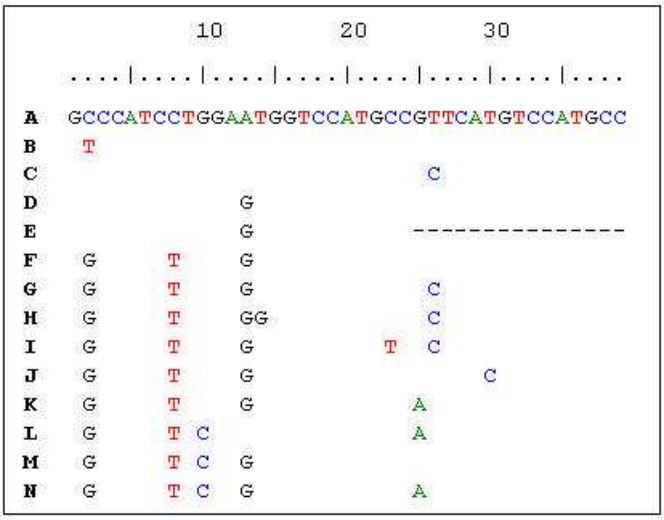


Figure 23 : Motifs pour le codage de ms77

On observe un motif exceptionnel, le motif E, à qui il manque 15 paires de bases. Ce motif est le premier de certains allèles tels que EFL, EIM et EK, qui sont présents dans la partie gauche du « minimum spanning tree » présenté Figure 24 et qui sont notés avec une taille intermédiaire (1,5 ou 2,5).

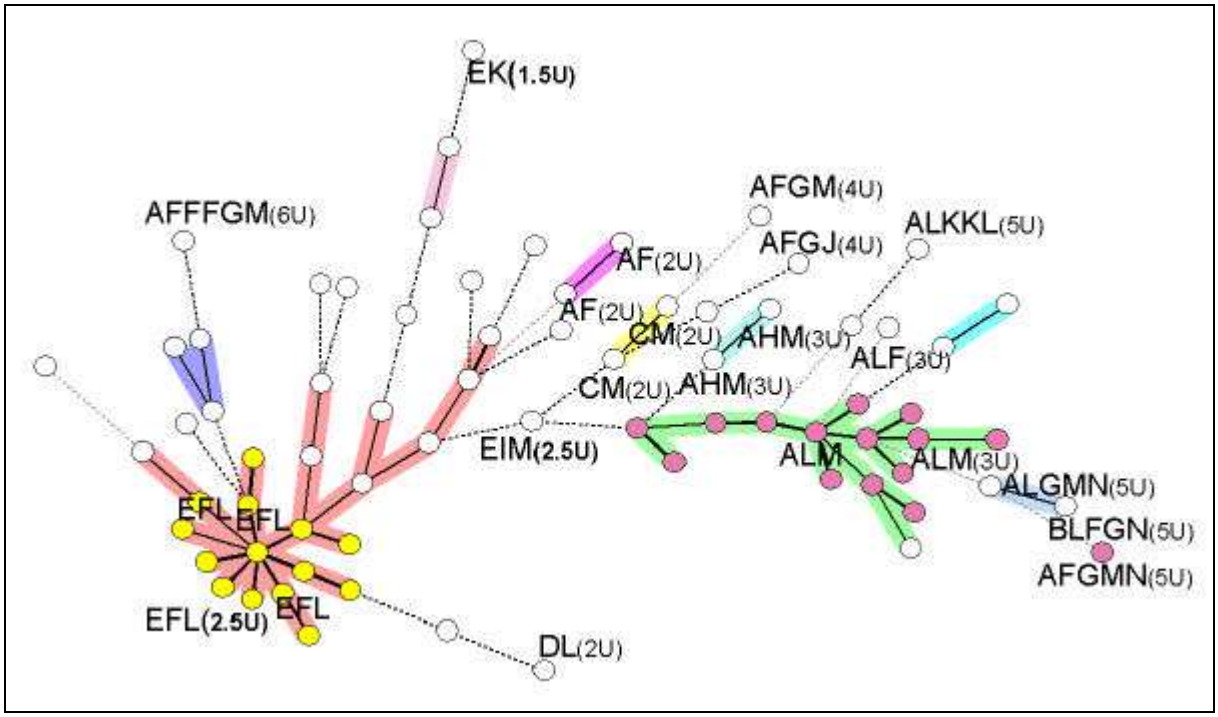


Figure 24 : Représentation « minimum spanning tree » de l'analyse MLVA des souches *P. aeruginosa* et du codage de l'allèle ms77

Certains motifs, rares, peuvent se déduire d'un autre par une mutation ponctuelle, c'est le cas de H, I, (issus de G) et J (issu de F), de B (issu de A). D'autres peuvent résulter d'un événement de recombinaison. Par exemple, il est tentant de suggérer que D, rencontré dans l'allèle DL de la souche 20C099, résulte de la perte d'un motif dans la séquence EF. D commence comme E et termine comme F, et sa position dans l'arbre de la Figure 24 suggère cette phylogénie. De même l'allèle AFGM de la souche 04C010 peut conduire à l'allèle CM des souches voisines 18A403 et 07C038 par un événement de délétion, AFG->C, C commence comme A et termine comme G. Cette étude limitée montre que, dans certains cas au moins, il sera possible de compléter le typage MLVA de mesure de tailles par le séquençage d'allèles. Dans le cas de *P. aeruginosa*, outre ms077, les locus ms142, ms172 et ms194 sont de bons candidats pour constituer une association MLVA-MLST (mesure de tailles, combinée si nécessaire au séquençage des allèles). On constate au niveau de la séquence protéique que la variation du nombre de motifs n'introduit pas de codon stop au niveau de la séquence répétée. La protéine pcoA sera plus ou moins longue selon les souches. La Figure 25 représente l'alignement des séquences protéiques au niveau de la répétition en tandem ms77 localisée dans le gène *pcoA*.

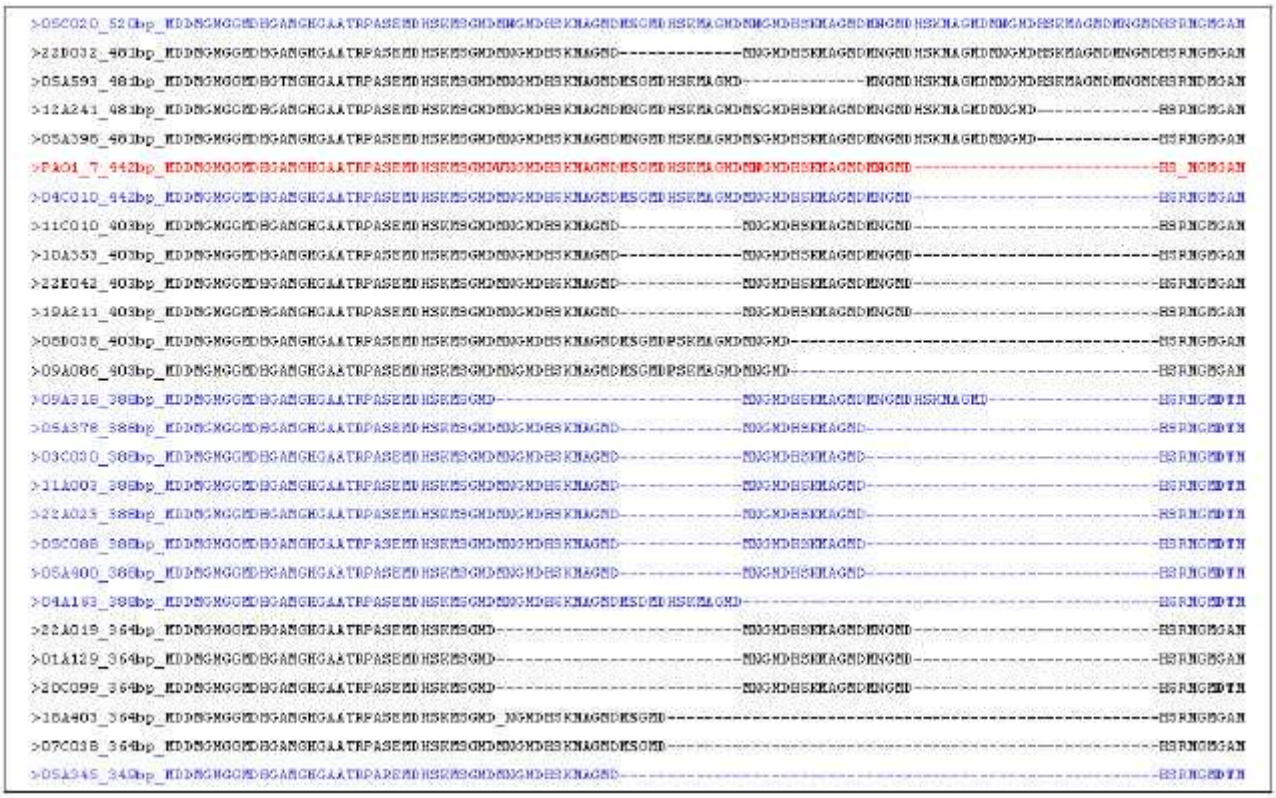


Figure 25 : Alignement des séquences protéiques ms77
de différentes souches *P. aeruginosa*

3.2.3.2 Séquençage de la répétition ms194

Cette répétition en tandem a déjà été décrite dans la littérature (Deretic 1990). Elle est constituée de répétitions de motifs de 75pb bien conservés, ou de motifs de 12pb moins conservés. Les migrations des produits de PCR sur gel d'agarose laissaient penser que la variation de taille entre les allèles était de 12pb et non de 75pb. Les allèles étant de grandes tailles (600 à 700pb) et le motif de petite taille, il était difficile d'assigner les tailles observées sur gel avec certitude. J'ai donc séquencé quelques allèles pour déterminer leur taille exacte.

L'allèle du génome PAO1 a une taille de 690pb d'après la séquence publiée. La gamme de taille des allèles va de 603 pb à 702 pb. En regardant les séquences dans le détail, on voit une succession de 5 motifs de 12 pb et d'un motif de 15 pb (ou 12 pb + 3 pb) se répéter plusieurs fois, c'est à dire plusieurs gros motifs de 75 pb. Un codage est possible en prenant par exemple un dictionnaire de motifs de 12pb et un autre de 15pb, ce qui nécessite un découpage « manuel » de la séquence en motifs des deux tailles. Le codage a été fait pour quelques allèles ms194. Les différents motifs de 12 pb et 15 pb sont présentés dans la Figure 26.

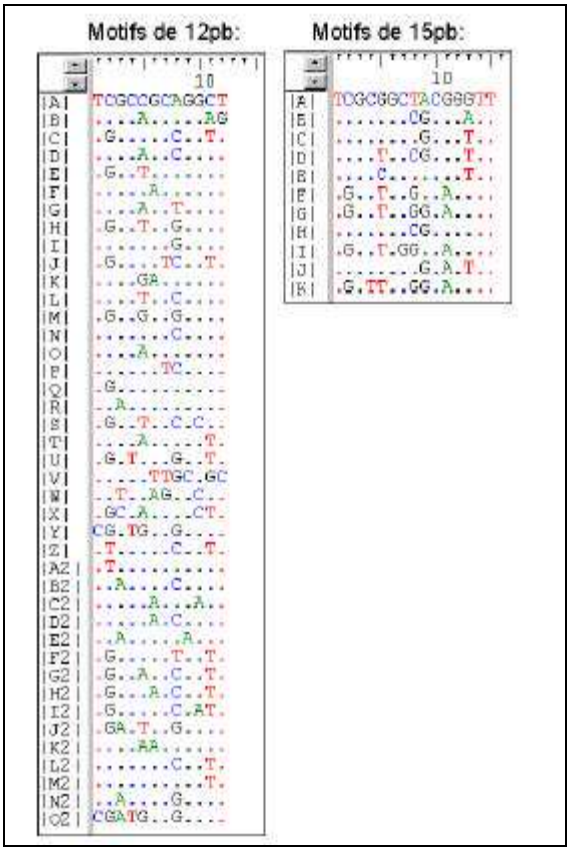


Figure 26 : Motifs de 12 pb et 15 pb pour le codage de ms77.

En noir sont indiqués les motifs de 12pb et en rouge les motifs de 15pb :

>ms194_PAO_45U_690pb_**ABYKAACADEBFCAGHCIJKLMCACFNHCICOPHDACODQERSTU**

>Pa90_194L_45U_690pb_**ABYGAACADEHFCAGHCIJKNMCAI$_2$FNJ$_2$CICOPHDACODQERSTU**

>Pa88_194L_43U_666pb_**ABYGAACADEHFCAGHCIJKNMCAI$_2$FNJ$_2$CICOPHDACO--ERSTU**

>ref_ms194_46U_702pb_**ABYFAACZA$_2$DHCICFNHCICFB$_2$HCICFNHCICC$_2$D$_2$HCACODQEE$_2$STU**

>Pa21_194L_38U_603pb_**ABO$_2$GAAF$_2$A$_2$DEHFCFNHCICFNHCICFNHDAG$_2$DQERSTU**

>Pa85_194L_42U_654pb_**ABYGAAH$_2$ADEHFNHCICFNHCICFNHCICFNHDAG$_2$DQEE$_2$STU**

>Pa89_194L_44U_678pb_**ABYGAACA$_2$DEHFCK$_2$NHCICK$_2$NHCACK$_2$NHCIL$_2$ONHDACDQEE$_2$STU**

>Pa96_194L_44U_678pb_**M$_2$BYIAACA$_2$DEHFCFNHJICFNHDACFNHCN$_2$L$_2$ONHDACDQERSTU**

Les deux allèles « 45U » se distinguent par cinq mutations ponctuelles, qui transforment un G en K, un H en B, un C en I_2 et un H en J_2 . Ce locus montre bien comment différentes évolutions de répétitions en tandem peuvent se conjuguer, avec ici amplification en un même locus de modules de 12, 15 ou 75 paires de bases. Certains allèles comptent, par exemple, 6 motifs de 12bp au lieu de 5 entre 2 motifs de 15 pb (cf la souche de référence 05A400 qui a un allèle de 702 pb).

3.2.4 Conclusions

Ce travail constitue la première étude MLVA réalisée chez *P. aeruginosa*, et a montré un pouvoir discriminant supérieur à celui observé par ribotypage de la même collection de souches, et équivalent à celui observé par analyse en champ pulsé pour les souches d'un des ribogroupes majoritaires. Par rapport au grand nombre de répétitions en tandem étudiées chez *P. aeruginosa*, peu se sont montrées polymorphes. L'échantillon des 201 répétitions testées n'était pas représentatif de la population générale des répétitions en tandem dans ce génome du fait des choix qui ont été faits sur les caractéristiques des répétitions à étudier. Le critère de conservation du motif s'est avéré être assez satisfaisant. Une autre étude des répétitions en tandem pourrait être envisagée chez *P. aeruginosa* pour identifier d'autres marqueurs polymorphes, en choisissant une requête uniquement sur le critère de conservation. Ainsi, avec une sélection des répétitions à plus de 90% de conservation interne, on obtient 51 répétitions candidates (dont une seule déjà testée et polymorphe, ms142). Le séquençage d'allèles du locus ms077 a permis de montrer que des allèles de même taille retrouvés dans différentes branches de l'arbre MLVA ont une histoire évolutive différente. Le séquençage de ms77 a aussi montré un motif d'une taille différente de tous les autres motifs de la répétition, qui explique les tailles intermédiaires observées sur gel d'agarose (2,5U et 3,5U). Pour ms194, le codage n'est pas simple, et donc difficile à généraliser pour toutes les souches.

3.3 Utilisation de la comparaison de génomes pour l'identification de répétitions en tandem polymorphes

3.3.1 Etude MLVA chez *Staphylococcus aureus*

3.3.1.1 Résultats des comparaisons de génomes

La comparaison de plusieurs génomes d'une même espèce bactérienne permet d'identifier directement des répétitions en tandem polymorphes.

Au début du projet de génotypage de *S. aureus*, les génomes des deux souches Mu50 et N315 ont pu être comparés. Ces deux souches sont des souches MRSA hospitalières, Mu50 étant en outre résistante à la vancomycine (VRSA). Ce sont les deux souches *S. aureus* les plus proches parmi toutes celles séquencées (99.95% d'homologie entre les flanquantes des répétitions de ces deux génomes). Onze répétitions en tandem polymorphes ont été identifiées dans ces deux souches. A titre de comparaison, 115 répétitions en tandem sont polymorphes

entre les souches de *Y. pestis* CO92 et KIM5 P12 qui présentent 99,66% d'homologie entre leurs flanquantes. Cette étude préliminaire sur Mu50 et N315 a permis de faire les premiers essais de marqueurs. La comparaison a été étendue à 4 génomes entièrement séquençés (Mu50, N315, NCTC8325, et MSSA476), puis 6 (Mu50, N315, NCTC8325, MW2, MSSA476 et MRSA252). Le bilan de cette comparaison est illustré par le Tableau 10. Dans le génome de Mu50, 828 répétitions en tandem ont été détectées par le TRF. Cent vingt deux d'entre elles ont plus d'un allèle parmi les 6 souches comparées, donc un tri important des répétitions est réalisé.

Tableau 10 : Bilan des comparaisons des 6 génomes *S. aureus.*

génomes comparés:	% homologie entre les flanquantes:	nombres de répétitions en tandem polymorphes:
Mu50/N315	99.95%	11
Mu50/MW2	98.68%	58
Mu50/NCTC8325	98.75%	57
Mu50/MRSA252	98.68%	63
Mu50/MSSA476	98.71%	54
Les 6 génomes		122 polymorphes dans au moins 2 des 6 génomes

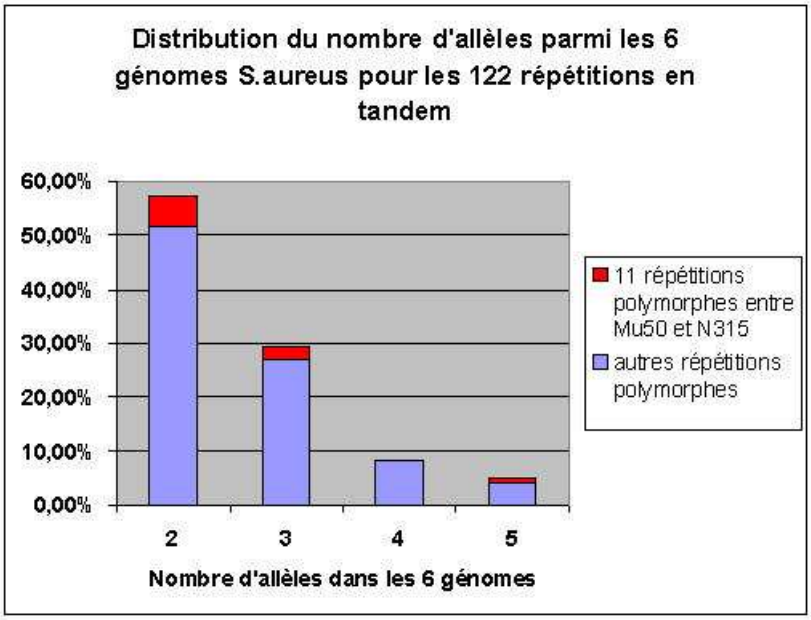


Figure 27 : Distribution du nombre d'allèles parmi les 6 génomes *S. aureus* comparés.

Curieusement, les 11 répétitions polymorphes dans les deux génomes les plus proches génétiquement ne font le plus souvent pas partie des répétitions les plus polymorphes parmi les 6 génomes comparés. La Figure 27 montre la distribution du nombre d'allèles parmi les 6 génomes *S. aureus* pour les 122 répétitions et les 11 répétitions polymorphes entre Mu50 et N315. Les 11 minisatellites polymorphes entre les souches Mu50 et N315 le sont pour la plupart uniquement dans ces deux souches et pas dans toutes les autres comme on aurait pu s'y attendre, puisqu'elles sont plus éloignées génétiquement de Mu50 que ne l'est N315. Sur 11 répétitions polymorphes entre Mu50 et N315, 7 le sont uniquement dans ces deux souches. L'allèle observé dans les autres souches pour ces 7 locus est dans 4 cas identique à l'allèle de N315, dans un cas identique à Mu50 et dans 2 cas la répétition n'a pas été détectée dans les autres souches.

La liste des 122 répétitions en tandem, telle qu'elle est fournie par la base de données, est présentée en Annexe 4. La base de données indique le nombre d'allèles observés parmi les souches comparées, puis les caractéristiques suivantes pour chaque génome : localisation de la répétition, taille (L) de la répétition, taille (U) du motif, nombre (N) de motifs et enfin détection ou non par TRF de la répétition dans le génome considéré.

Par ailleurs, nous avons regardé la localisation intra ou intergénique des 122 répétitions, les produits des gènes dans lesquels sont localisées ces répétitions, la taille du motif protéique et enfin sont indiquées les 33 répétitions qui ont été étudiées au cours de cette thèse. La Figure 28 montre la répartition des 122 répétitions selon leur localisation intra ou intergénique et la proportion des motifs multiples de 3 dans ces deux groupes de répétitions. La définition de région intragénique considérée ici est la partie codante du gène ainsi que son promoteur.

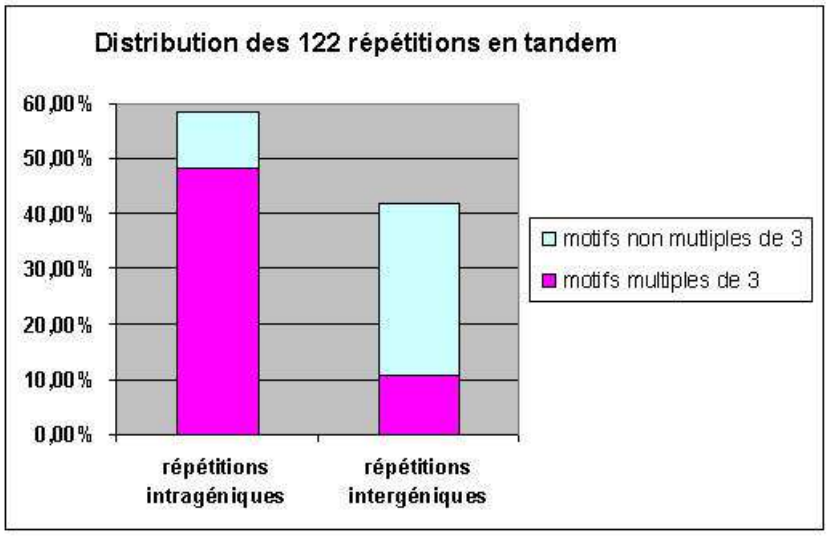


Figure 28 : Distribution des 122 répétitions selon leur localisation intra ou intergénique.

Cette figure illustre deux points évoqués précédemment. Le premier est que 60% des répétitions en tandem polymorphes de *S. aureus* sont localisées dans des séquences codantes. Les 40 % restants sont dans les régions intergéniques, alors que par ailleurs celles-ci ne représentent que 15% de l'ensemble du génome de *S. aureus*.. Le second point est que 90 % des répétitions en tandem intragéniques sont à motif multiple de 3 paires de bases. Dans les régions intergéniques, la proportion de motifs multiples de 3 est d'environ un tiers comme on peut s'y attendre. La pression de sélection est plus faible dans les régions intergéniques.

3.3.1.2 Les séquences STARs

Une autre observation intéressante parmi les 122 répétitions en tandem polymorphes est la présence de 12 répétitions correspondant à des éléments STARs (décrits dans le paragraphe 1.3.2). Je me suis intéressée plus en détail à ces séquences. Dans l'article de Cramton et al (Cramton 2000), une analyse de type Southern blot suggère la présence d'un nombre d'éléments STAR qui varie (13 à 21) en fonction de la souche testée, mais les auteurs supposent qu'il y en a probablement plus, ce qui est confirmé (environ 70) dans l'article sur le séquençage de N315 et Mu50 (Kuroda 2001). La structure d'une séquence STAR est schématiquement décrite par la formule B-(C)n-A.

J'ai recherché par BLAST (Altschul 1997) dans la base de données de répétitions en tandem, les séquences STARs dans les 6 génomes de *S. aureus*, à partir de la séquence consensus du motif central C (57pb :CCCCAACTTGCACATTATTGTAAGCTGACTTTTCGTCAGCTTCTGTGTTGGGGCCCC), qui est répété dans les éléments STARs. Certaines séquences STARs n'ont pas été identifiées par BLAST, les motifs centraux C étant probablement trop divergés par rapport au consensus pour être identifiés. Par ailleurs, on dispose des comparaisons de génomes qui permettent l'identification des répétitions dont les flanquantes sont homologues dans les autres génomes. On peut donc savoir quelles sont les séquences STARs « homologues » dans les génomes comparés par rapport à celles détectées dans Mu50.

Sur les 13 séquences STARs identifiées par BLAST chez Mu50, une seule n'est pas polymorphe parmi les 6 génomes. Cette séquence STAR n'a été retrouvée que dans MRSA252, et elle y est de même taille que dans Mu50. On obtient donc au total 12 séquences STARs polymorphes dans la comparaison des 6 souches. Les résultats de la recherche de séquences STARs dans la base de données de répétitions en tandem sont résumés dans le Tableau 11.

Les répétitions identifiées par TRF ont des motifs de 54 à 59pb (ou aussi 113bp). En effet, selon la conservation des séquences, le TRF ne trouve pas toujours la même taille de consensus, ni exactement le même démarrage pour le 1er motif.

Voici le bilan des séquences STARs identifiées par BLAST dans la base des répétitions en tandem :

Mu50 : 13 séquences STARs

N315 : 13 séquences STARs

MRSA252 : 11 séquences STARs

NCTC8325 : 14 séquences STARs

MW2 : 17 séquences STARs

MSSA476 : 19 séquences STARs

Sur les 12 séquences STARs identifiées dans Mu50 par la comparaison des 6 génomes, 7 ont été étudiées ici.

Tableau 11 : Séquences STARs identifiées dans la base de données de répétitions en tandem.

position (kb)	souche	taille unité (pb)	nombre de copies	position (kb)	souche	taille unité (pb)	nombre de copies	position (kb)	souche	taille unité (pb)	nombre de copies	position (kb)	souche	taille unité (pb)	nombre de copies	position (kb)	souche	taille unité (pb)	nombre de copies	position (kb)	souche	taille unité (pb)	nombre de copies
																115	NCTC8325	58	2,6				
																				140	MSSA476	58	2,6
								141	MW2	58	2,6												
																				165	MSSA476	58	2
																				273	MSSA476	58	2,9
								274	MW2	58	2,9												
311*	Mu50	55	3,1	311	N315	55	3,1	290	MW2	56	4,1	309	MRSA252	56	3,1	258	NCTC8325	54	3	288	MSSA476	56	3,1
																730	NCTC8325	56	3,6				
								782	MW2	56	3,6												
842*	Mu50	55	2,5	818	N315	55	2,5									762	NCTC8325	55	2,5				
																				780	MSSA476	56	3,6
847*	Mu50	59	6	823	N315	59	6	819	MW2	58	3,1	862	MRSA252	58	3,1	768	NCTC8325	58	7	817	MSSA476	58	3,1
855*	Mu50	56	5,2	830	N315	56	5,2													824	MSSA476	56	2
906*	Mu50	56	2,3	874	N315	56	2,3																
												907	MRSA252	56	2,3								
												911	MRSA252	112	2,59								
								924	MW2	58	2,6	956	MRSA252	56	2,3	860	NCTC8325	56	4	908	MSSA476	58	2,6
1213*	Mu50	56	5,1	1137	N315	56	5,1	1138	MW2	56	5					1074	NCTC8325	113	3	1167	MSSA476	56	5
								1041	MW2	58	3,3	1082	MRSA252	58	4,3					1070	MSSA476	58	3,3
1219	Mu50	55	3	1142	N315	55	3																
								1226	MW2	56	4,5												
																				1254	MSSA476	56	2

position (kb)	souche	taille unité (pb)	nombre de copies	position (kb)	souche	taille unité (pb)	nombre de copies	position (kb)	souche	taille unité (pb)	nombre de copies	position (kb)	souche	taille unité (pb)	nombre de copies	position (kb)	souche	taille unité (pb)	nombre de copies	position (kb)	souche	taille unité (pb)	nombre de copies
1425*	**Mu50**	58	4,1	1348	**N315**	58	4,1									1286	**NCTC8325**	58	4	1379	**MSSA476**	58	3,63
								1440	**MW2**	58	3,1					1375	**NCTC8325**	58	3,1	1469	**MSSA476**	58	3,1
								1483	**MW2**	58	3												
																				1509	**MSSA476**	58	2
1729	**Mu50**	56	5,1	1652	**N315**	56	5,1	1698	**MW2**	59	3	1763	**MRSA252**	54	5,1					1678	**MSSA476**	59	3
																				1764	**MSSA476**	57	3
																1776	**NCTC8325**	59	4,5				
								1784	**MW2**	57	3												
								1949	**MW2**	60	2,61	2026	**MRSA252**	58	2,2	1890	**NCTC8325**	**117**	3,1				
								1992	**MW2**	59	3									1971	**MSSA476**	59	3
2028*	**Mu50**	59	2,3	1950	**N315**	59	2,3					2068	**MRSA252**	56	6	1976	**NCTC8325**	56	3				
2039*	**Mu50**	56	2,3	1961	**N315**	56	2,3									1987	**NCTC8325**	58	2,3				
				2054	**N315**	58	2,8																
2131	**Mu50**	58	2,8									2173	**MRSA252**	55	2,5								
								2337	**MW2**	56	2,1									2316	**MSSA476**	56	2,1
2561*	**Mu50**	56	2,4	2490	**N315**	56	2,4																
								2776	**MW2**	56	3,1	2855	**MRSA252**	56	2,1	2779	**NCTC8325**	56	3,1	2756	**MSSA476**	56	3,1

séquences STARs polymorphes obtenues par comparaison avec Mu50 (* : celles étudiées dans la thèse)

séquences STARs polymorphes pour au moins 2 des 5 autres souches

séquences STARs sans homologue dans la base de données

séquence STAR de même taille dans Mu50 et MRSA

3.3.1.3 Résultats de l'étude MLVA

Une partie des 122 répétitions en tandem polymorphes a été étudiée. Des choix ont été faits en fonction du nombre d'allèles observés parmi 4 génomes séquencés (au début du projet), en fonction de la taille du motif (pour faciliter les assignations de taille des produits PCR après migration sur gel d'agarose), et en fonction de la possibilité de trouver des amorces qui s'hybrident parfaitement avec les flanquantes des répétitions dans les 4 génomes, pour limiter les éventuels problèmes d'amplification. Deux autres génomes ont été ajoutés à la comparaison en cours de projet. Le Tableau 12 présente la liste des 33 répétitions en tandem étudiées au cours de la thèse.

Tableau 12 : Liste des 33 répétitions en tandem testées par PCR.

nombre d'allèles dans les 6 souches	taille du motif (pb)	position Mu50	gène à cette position dans Mu50	produit protéique	position du gène	taille motif / 3	motif protéique répété	répétitions en tandem étudiées
5	**9 ou 18**	1098012--1098170	sspA	protéase V8, glutamyl endopeptidase	1099031..1098003	3 ou 6	PNNPDN	**Mu50_1098_18bp_9U**
5	**159**	1866118--1866562	SAV1738	protéine hypothétique	1865490..1866848	53	ALKAQQAAIKEEASANNLSDTSQEAQEIQEAKREAQAEADKSVAVSNEESKAS	**Mu50_1866_159bp_3U**
5	**18**	636061--636666	sdrD	proteine sdr de liaison au fibrinogène	632692..636849	6	SD	**Mu50_0636_18bp_33,3U**
5	**67 ou 133**	704561--704796	intergénique			22,33 ou 44,33		**Mu50_0704_67bp_4U**
5	**18**	888858--889722	clfA	clumping factor A	887186..889993	6	SD	**Mu50_0888_18bp_48U**
4	**20 ou 64**	1291998--1292219	intergénique			6,67 ou 21,33		**Mu50_1291_64bp_3,5U**
4	**56**	1729388--1729679	**intergénique**			18,67		**séquence STAR Mu50_1729_56bp_5U**
4	**24**	2351355--2351474	SAV2208	protéine hypothétique	2351393..2351563	8		**Mu50_2351_24bp_5U**
4	**42**	2642053--2642330	fnb	homologue à la protéine de liaison à la fibronectine	2641824..2644940	14	PETPTPPTPEVPSE	**Mu50_2642_42bp_7U**
4	**18**	631615--632142	sdrC	protéine sdr de liaison au fibrinogène	629464..632325	6	SD	**Mu50_0631_18bp_30,3U**
4	**43**	965164--965428	intergénique			14,33		**Mu50_0965_43bp_6U**
3	**9 ou 18**	1105143--1105186	atl	autolysine	1103624..1106455	3 ou 6		**Mu50_1105_18bp_2,5U**
3	**63**	1132682--1133067	SAV1078	protéine hypothétique	1132622..1133071	21	LQLLVVRGFYACARRMYPSFT	**Mu50_1132_63bp_6,1U**
3	**134**	1194184--1194530	intergénique			44,67		**Mu50_1194_67bp_7U**

nombre d'allèles dans les 6 souches	taille du motif (pb)	position Mu50	gène à cette position dans Mu50	produit protéique	position du gène	taille motif / 3	motif protéique répété	répétitions en tandem étudiées
3	24	122905--123156	spa	précurseur de la protéine A de liaison aux IgG	122614..123966	8	PGKEDNNK	Mu50_0122_24bp_10U
3	174	123159--123840	spa	précurseur de la protéine A de liaison aux IgG	122614..123966	58	QQNAFYEILNMPNLNEEQRNGFIQSLKDDPSQSANLLSEAKKLNESQAPKADNKFNKE	Mu50_0123_174bp_3,9U
3	58	1425109--1425340	intergénique			19,33		séquence STAR Mu50_1425_58bp_4U
3	56	2039328--2039458	intergénique			18,67		séquence STAR Mu50_2039_56bp_3U
3	9	2294935--2295172	fmtB(mrp)	FmtB protein	2287935..2295380	3		Mu50_2294_9bp_29U
3	42	2638502--2638675	fnbB	homologue à la protéine de liaison à la fibronectine	2638258..2641143	14	PEVPSEPETPVPPT	Mu50_2638_42bp_4,1U
3	81	266128--266583	coa	précurseur de la staphylocoagulase	264640..266616	27	KKPSKTNAYNVTTHANGQVSYGARPTQ	Mu50_0266_81bp_5,6U
3	18	2781740--2782399	clfB	clumping factor B	2781518..2784151	6	SD	Mu50_2781_18bp_36,7U
3	56	311490--311657	intergénique			18,67		séquence STAR Mu50_0311_55bp_3U
3	40	43142--43471	intergénique			13,33		mec HVR region (ou dru)
3	18	640048--640484	sdrE	protéine sdr de liaison au fibrinogène	637243..640668	6	SD	Mu50_0640_18bp_24,3U
2	56	1213418--1213706	intergénique			18,67		séquence STAR Mu50_1213_56bp_5U
2	231	1516384--1517097	ebhA	protéine hypothétique similaire à l'adhésine emb de streptocoque	1514410..1534551	77	KEKQALKDRINQILQQGHNGINNAMTKEEIEQAKAQLAQALKEIKDLVKAKENAKQDVDKQVQALIDEIDQNPNLTD	Mu50_1516_231bp_3U
2	90	1994271--1994532	tRNA-Gly			30		Mu50_1994_90bp_3U
2	100	2221867--2222183	intergénique			33,33		Mu50_2221_100bp_3,2U
2	15 ou 30	2547600--2547676	sbi	protéine de liaison aux IgG	2546792..2548078	5 ou 10	PKVEA	Mu50_2547_15bp_5U
2	55	842266--842402	intergénique			18,33		séquence STAR Mu50_0842_55bp_3U
2	24	899533--899596	SAV0825	protéine hypothétique conservée	899565..900182	8		Mu50_0899_24bp_2,7U
2	56	906124--906248	intergénique			18,67		séquence STAR Mu50_0906_56bp_3U

* D'après le site « gib », http://gib.genes.nig.ac.jp

Les répétitions sont nommées selon la nomenclature suivante :

Nom du génome_position de la répétition sur le génome (kb)_taille du motif_ nombre de motifs

Nous avons étudié quatre catégories de marqueurs :

- Des marqueurs à localisation intragénique dont certains ont déjà montré leur utilité pour des études épidémiologiques (gènes *spa* et *coa*). Ces marqueurs permettent d'établir un lien avec les données de la littérature.

- Des séquences STARs, qui le plus souvent sont localisées dans des régions intergéniques.

- Des répétitions présentes dans des séquences de protéines hypothétiques.

- Et enfin des répétitions localisées dans des régions intergéniques.

Bilan du typage :

Sur 107 souches fournies par l'institut Pasteur (Dr Névine El Sohl) et 30 souches fournies par le Val de Grâce (Dr Jean-Louis Koeck), 5 souches (respectivement 3 et 2) ont été éliminées de l'analyse du fait de problèmes d'amplification par PCR rencontrés pour la plupart des marqueurs testés.

Parmi les 33 répétitions en tandem testées pour le génotypage des 137 souches de *S. aureus*, nous avons retenu 14 répétitions pour l'analyse MLVA. Elles sont décrites dans le Tableau 13. Ces marqueurs ont été amplifiés par PCR sans difficulté, la bonne résolution des allèles sur gel d'agarose ainsi que la facilité d'analyse des gels par le logiciel BioNumerics ont conduit à sélectionner ce sous-groupe de 14 répétitions.

Un certain nombre de marqueurs a été éliminé du fait de la difficulté d'assignation de tailles des allèles lorsque le motif est petit et les allèles grands. D'autres ont été analysés avec des amorces choisies lorsque 4 des 6 génomes *S. aureus* étaient disponibles et des mésappariements des amorces avec les nouveaux génomes séquençés pourraient expliquer certains problèmes d'amplification. Les différentes raisons sont détaillées dans le Tableau 14.

Parmi les 14 répétitions en tandem pour l'analyse MLVA, nous avons retenu deux marqueurs étudiés ces dernières années, *spa* (Shopsin 1999) et *coa* (Shopsin 2000), qui possèdent des répétitions utiles pour des études épidémiologiques sur du long terme (*coa*) et pour des études épidémiologiques à court terme et en situation d'épidémie (*spa*). Le séquençage des allèles du locus *spa* sera décrit dans le paragraphe suivant.

Tableau 13 : Caractéristiques des 14 répétitions en tandem de l'analyse MLVA.

répétitions en tandem étudiées	ORF associée	% de conservation (Mu50)	% GC (Mu50)	typage	gamme de tailles des allèles (pb)	U: nombre de motifs	nombre d'allèles (collection de souches + génomes séquencés)	PIC (index de polymorphisme)	amorce gauche (5'=>3')	amorce droite (5'=>3')
Mu50_0122_24bp_10U	spa	88	45	VNTR + analyses de séquences	224 - 464	3U - 13U	**11**	**0,68**	AGCAGTAGTGCCGTTTGCTT	AAGACGATCCTTCAGTGAGCA
Mu50_0266_81bp_6U	coa	91	41	VNTR	468 - 873	4U - 9U	**5**	**0,35**	TTGGATATGAAGCGAGACCA	CTTCCGATTGTTCGATGCTT
Mu50_0311_55bp_3U*	intergénique	94	45	VNTR	162 - 405	1U-5,5U	**7**	**0,43**	AGGGTTAGAGCCCGAGACAT	CACGGGATTGGAACAGAAAT
Mu50_0906_56bp_3U	intergénique	92	54	VNTR	216 - 328	1U - 3U	**3**	**0,46**	CCCAGCCTGTTTTCATAAGC	CCAAAAGAAAATACACCTATAACAAA
Mu50_1213_56bp_5U	intergénique	77	47	VNTR	255 - 591	1U - 7U	**6**	**0,71**	TTCCAGTTCTAGTGCTATATTGGTAG	TGTAGTGGTTCTTTATCATTAGCTGT
Mu50_1425_58bp_4U	intergénique	65	47	VNTR	357 - 763	1U - 8U	**6**	**0,49**	GGTTTGACAAAGCTAAAGTGAAGT	AAACGTATTATTTCATTGAGCAGAA
Mu50_1729_56bp_5U	intergénique	71	48	VNTR	207 - 496	1U - 6U	**6**	**0,50**	GCATAGGGAGTGGGACAGAA	TCAACGTCGAAAATGACGAA
Mu50_2039_56bp_3U	intergénique	93	48	VNTR	170 - 338	3U - 6U	**4**	**0,49**	TTCGTTCTACCCCAACTTGC	GAGCCTGGGTCATAAATTCAA
Mu50_0704_67bp_4U	intergénique	71	31	VNTR	246 - 779	2U - 10U	**7**	**0,53**	CGCGCGTGAATCTCTTTTAT	AGTCCCATATCGTGCGTTAAA
Mu50_1132_63bp_6U	SAV1078	93	42	VNTR + analyses de séquences	217- 783	1U - 9U	**8**	**0,64**	CGTGCATAATGGCTTACGAA	AAGCAGCAGAAAAAGCTAAAGAA
Mu50_1194_67bp_7U*	intergénique	84	34	VNTR	256 - 591	3U - 8U	**7**	**0,50**	AGTGCAAGCGGAAATTGAAG	ATCGTGAAAAAGCCCAAAAA
Mu50_1291_64bp_4U*	intergénique	87	34	VNTR	177 - 473	1U-5,5U	**6**	**0,29**	GGGGGAAATTCTAAGCAACC	CGAAATTTTCCACGTCGATT
Mu50_1866_159bp_3U	SAV1738	91	37	VNTR	289 - 766	1U - 4U	**4**	**0,55**	CTGTTTTGCAGCGTTTGCTA	GCAACTTGAAGAAACGGTTG
Mu50_2547_15bp_5U	sbi	63	36	VNTR	257 - 722	5U - 36U	**6**	**0,21**	AAAGATGCTGAAAAGAAAGTGG	TGATCAATCGCACCTTTGTA

* : tailles intermédiaires observées

Tableau 14: Liste des 19 répétitions en tandem éliminées de l'analyse MLVA.

nombre d'allèles dans les 6 souches :	répétitions en tandem étudiées :	ORF associée:	motif d'abandon du locus:
4	Mu50_0631_18bp_30,3U	sdrC	Pas d'amorces « monolocus »
5	Mu50_0636_18bp_33,3U	sdrD	Pas d'amorces « monolocus »
3	Mu50_0640_18bp_24,3U	sdrE	Pas d'amorces « monolocus »
5	Mu50_0888_18bp_48U	clfA	Echecs d'amplification
5	Mu50_1098_18bp_9U	sspA	Echecs d'amplification
3	Mu50_2781_18bp_36,7U	clfB	Echecs d'amplification
2	séquence STAR Mu50_0842_55bp_3U	intergénique	Echecs d'amplification
3	mec région HVR (ou dru)	intergénique	Problème de PCR
3	Mu50_0123_174bp_3,9U	spa	Tailles intermédiaires à séquencer
2	Mu50_0899_24bp_2,7U	SAV0825	Monomorphe dans la collection de souches (2 allèles dans les génomes séquençés)
4	Mu50_0965_43bp_6U	intergénique	Echecs d'amplification
3	Mu50_1105_18bp_2,5U	atl	Echecs d'amplification
2	Mu50_1516_231bp_3U	ebhA	Echecs d'amplification
2	Mu50_1994_90bp_3U	tRNA-Gly	Echecs d'amplification
2	Mu50_2221_100bp_3,2U	intergénique	Echecs d'amplification
3	Mu50_2294_9bp_29U	fmtB(mrp)	Echecs d'amplification
4	Mu50_2351_24bp_5U	SAV2208	Problème d'analyse BioNumerics
3	Mu50_2638_42bp_4,1U	fnbB	Echecs d'amplification
4	Mu50_2642_42bp_7U	fnb	Délétions dans certains allèles

Les souches analysées pour cette étude MLVA ont été caractérisées lors d'études préalables par PFGE, mais les génotypes obtenus ne sont pas comparables d'une étude à l'autre, la nomenclature utilisée pour les génotypes n'étant pas la même. Il est donc difficile de comparer les résultats entre les différentes séries de souches, en revanche, à l'intérieur d'une série, on pourra comparer le pouvoir discriminant de l'analyse MLVA par rapport à celui du PFGE (ceci illustre une faiblesse de l'approche PFGE : la difficulté de comparer des résultats d'un gel à l'autre).

La plupart des souches sont résistantes à la méthicilline. La série de 15 souches GISA (Glycopeptide Intermediaire *Staphylococcus aureus*) provient de différents hôpitaux français. La collection du Val de Grâce comporte aussi quelques souches GISA.

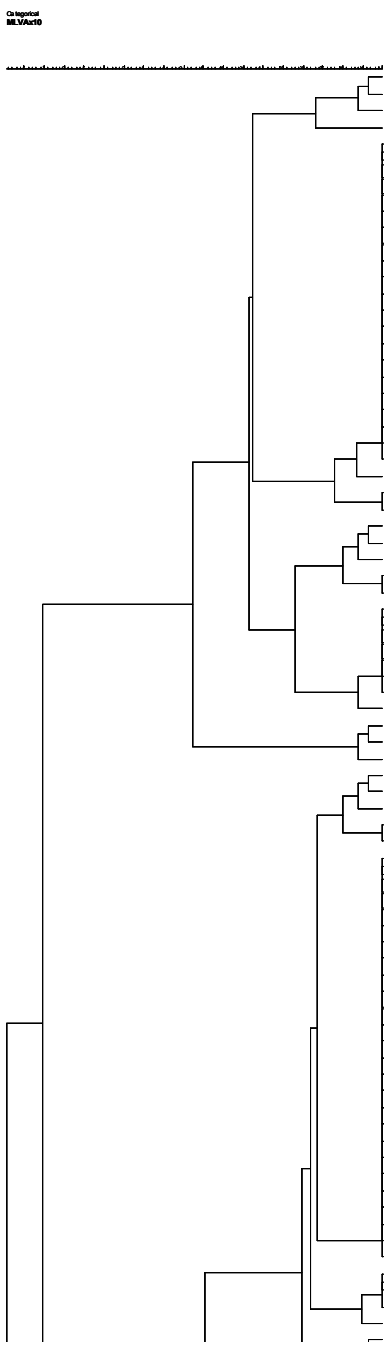
Pour l'analyse des génotypes, les tailles des allèles ont été exportées de la base de données Bionumerics et converties en nombre de motifs (U). Ensuite, ces données en nombre de motifs ont été importées dans BioNumerics et le « clustering » ou analyse de similarité réalisé en utilisant les coefficients « categorical » et « Ward ». Le coefficient categorical considère les caractères (ici chaque VNTR) comme indépendants, et le même poids est donné aux différentes tailles d'allèles, c'est-à-dire qu'un allèle de 9 motifs ne sera pas considéré comme phylogénétiquement plus proche d'un allèle de 10 motifs que d'un allèle de 3 motifs. Pour les 6 séquences STARs de l'analyse MLVA, nous avons considéré qu'une absence (confirmée par 2 tentatives) de produit d'amplification par PCR correspondait à une absence du locus dans la souche étudiée, et ceci a été exploité comme une donnée.

Nous avons utilisé deux types de représentation des arbres : arbre « classique » et arbre « minimum spanning tree ». Au total, les 137 souches se répartissent en 68 génotypes MLVA différents. Trois génotypes majoritaires sont observés :

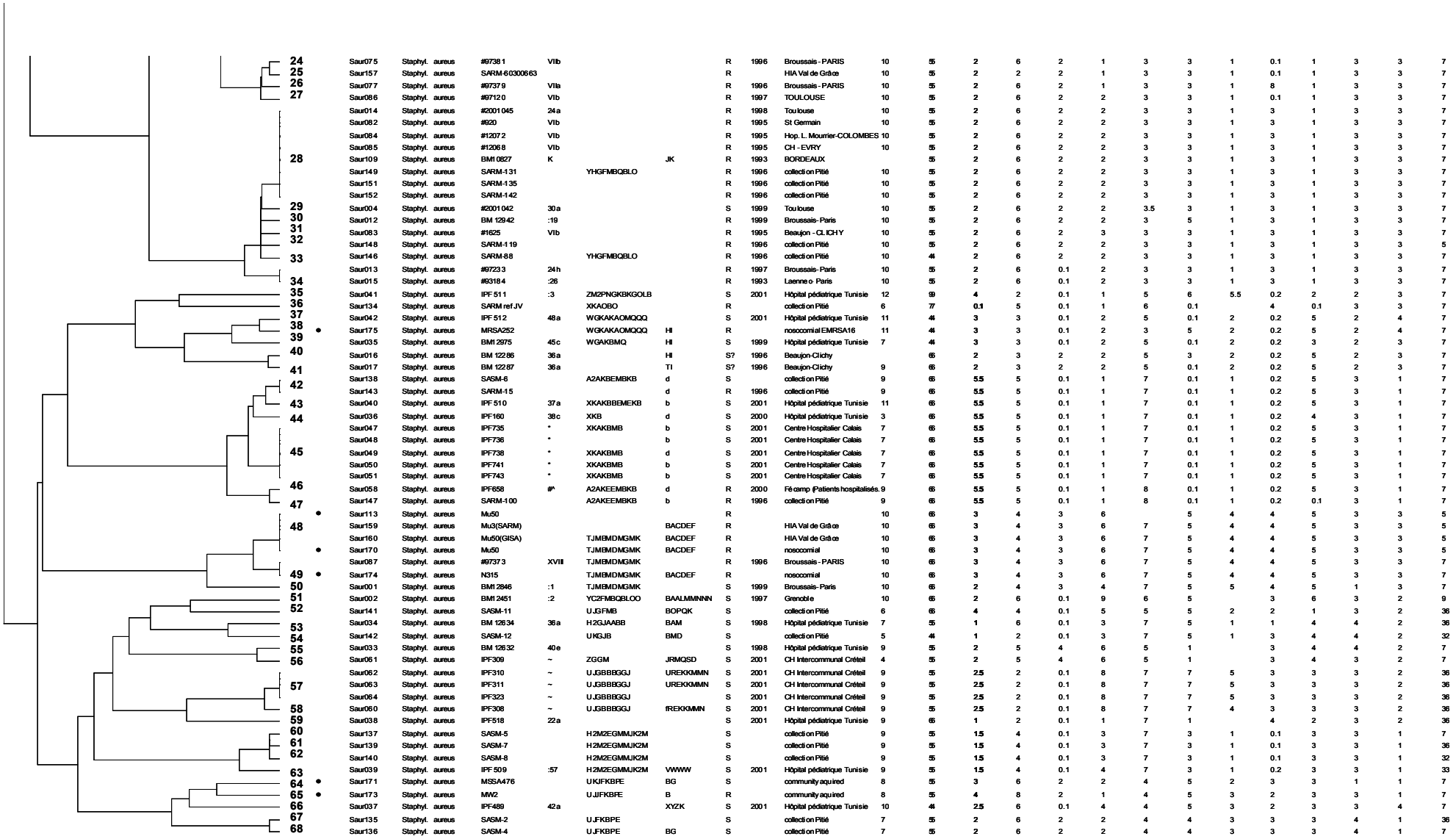
- 20 souches de génotype 11-5-2-6-2-2-3-4-1-3-1-4-3-7 (génotype n° 5 dans l'arbre)
- 25 souches de génotype 10-5-2-6-2-1-3-3-1-3-1-3-3-7 (génotype n° 21 dans l'arbre)
- 8 souches de génotype 10-5-2-6-2-2-3-3-1-3-1-3-3-7 (génotype n° 28 dans l'arbre)

On constate parmi ces génotypes MLVA que plusieurs génotypes PFGE sont observés et permettent de discriminer ces souches.

Les souches de référence sont signalées par un point, et les souches GISA par une étoile.



n° génotype MLVA		n° ADN			souche	PFGE	codage *spa*	codage 1132	Methicilline	année	origine	ms0122 (*spa*)	ms0266 (*coa*)	ms0311	ms0704	ms0906	ms1132	ms1194	ms1213	ms1291	ms1425	ms1729	ms1866	ms2039	ms2547
1	★	Saur018	Staphyl.	aureus	IPF 556	I-1	YFGFMBLO		R	1999	Broussais - Paris	8	55	2	6	0.1	2	3	6	1	3	1	4	4	7
2	★	Saur029	Staphyl.	aureus	BM 10629	I - 5	YHFGFMBQBLO		R	1993	Bordeaux	11	55	2	6	0.1	2	3	6	1	3	1	4	4	7
3	★	Saur032	Staphyl.	aureus	#96145	I - 17	YHFGFMBQBLO		R	1996	Blois	11	55	2	6	0.1	2	3	4	1	3	1	4	4	7
4	★	Saur023	Staphyl.	aureus	HM10	XXIV			R	2000	Henri Mondor-Créteil	11	55	2	6	0.1	2	3.5	6	1	3	0.1	4	3	7
		Saur010	Staphyl.	aureus	BM 3364	:13			R	1981	Broussais- Paris	11	55	2	6	2	2	3	6	1	3	1	4	3	7
	★	Saur022	Staphyl.	aureus	BM 10828	I-1			R	1993	Bordeaux	11	55	2	6	2	2	3	6	1	3	1	4	3	7
	★	Saur028	Staphyl.	aureus	Harmony 9	I - 5	YHFGFMBQBLO	JK	R	1990	Finlande	11	55	2	6	2	2	3	6	1	3	1	4	3	7
	★	Saur031	Staphyl.	aureus	#97130	I - 18			R	1997	Toulouse	11	55	2	6	2	2	3	6	1	3	1	4	3	7
		Saur089	Staphyl.	aureus	BM9 586	A			R	1987	Broussais - PARIS	11	55	2	6	2	2	3	6	1	3	1	4	3	7
		Saur090	Staphyl.	aureus	BM1 2184	A			R	1987	Broussais - PARIS	11	55	2	6	2	2	3	6	1	3	1	4	3	7
		Saur091	Staphyl.	aureus	BM9 591	A			R	1987	Broussais - PARIS	11	55	2	6	2	2	3	6	1	3	1	4	3	7
		Saur092	Staphyl.	aureus	BM1 2188	A			R	1987	Broussais - PARIS	11	55	2	6	2	2	3	6	1	3	1	4	3	7
		Saur093	Staphyl.	aureus	BM1 0761	A 4			R	1993	TOULOUSE	11	55	2	6	2	2	3	6	1	3	1	4	3	7
		Saur094	Staphyl.	aureus	BM1 0759	A 4			R	1993	TOULOUSE	11	55	2	6	2	2	3	6	1	3	1	4	3	7
5		Saur096	Staphyl.	aureus	BM9 343	A 7			R	1987	TOULOUSE	11	55	2	6	2	2	3	6	1	3	1	4	3	7
		Saur097	Staphyl.	aureus	BM1 0872	A 8			R	1992	Aalst - BELGIQUE	11	55	2	6	2	2	3	6	1	3	1	4	3	7
		Saur098	Staphyl.	aureus	BM1 0888	A 8			R	1993	Aalst - BELGIQUE	11	55	2	6	2	2	3	6	1	3	1	4	3	7
		Saur100	Staphyl.	aureus	BM1 0914	A 9			R	1991	Broussais - PARIS	11	55	2	6	2	2	3	6	1	3	1	4	3	7
		Saur101	Staphyl.	aureus	BM1 0896	A 9			R	1994	Ghent -BELGIQUE	11	55	2	6	2	2	3	6	1	3	1	4	3	7
		Saur102	Staphyl.	aureus	BM1 0138	A 10	YHFGFMBQBLO	JK	R	1989	Barcelone - ESPAGNE	11	55	2	6	2	2	3	6	1	3	1	4	3	7
		Saur103	Staphyl.	aureus	BM1 0130	A 10			R	1989	Barcelone - ESPAGNE	11	55	2	6	2	2	3	6	1	3	1	4	3	7
		Saur105	Staphyl.	aureus	BM1 2152	A 12			R	1989	Barcelone - ESPAGNE	11	55	2	6	2	2	3	6	1	3	1	4	3	7
		Saur106	Staphyl.	aureus	BM1 0877	G			R	1992	Aalst - BELGIQUE	11	55	2	6	2	2	3	6	1	3	1	4	3	7
		Saur107	Staphyl.	aureus	BM1 2156	G			R	1989	Barcelone - ESPAGNE	11	55	2	6	2	2	3	6	1	3	1	4	3	7
6		Saur071	Staphyl.	aureus	IPF 65	#&	YHFGFMBQBLO		R	2000	Rotschild- Paris	11	55	2	6	2	2	3	6	1	3	1	4	2	7
	●	Saur112	Staphyl.	aureus	Rn450 (8325)				R			11	55	2	10	2	2	3	6	1	3	1	4	3	7
7	●	Saur172	Staphyl.	aureus	NCTC8325		YHGGFMBQBLO	JK	S		old laboratory strain	11	55	2	10	2	2	3	6	1	3	1	4	3	7
8	★	Saur019	Staphyl.	aureus	IPF 555	I-1	YFGFMBLO		R	1999	Broussais - Paris	8	55	2	6	2	2	3	6	1	0.1	1	4	3	7
9	★	Saur026	Staphyl.	aureus	Harmony 26	I - 1	YHFGFMBQBQBLO		R	1989	Espagne	13	55	2	6	2	2	3	6	1	0.1	1	4	3	7
10	★	Saur021	Staphyl.	aureus	IPF 562	I-1	YFGFMBLO		R	1999	Broussais - Paris	8	55	2	6	2	2	3	6	1	0.1	1	4	2	7
11	★	Saur020	Staphyl.	aureus	IPF 557	I-1			R	1999	Broussais - Paris	8	55	2	6	2	2	3	6	1	3	1	4	3	7
	★	Saur163	Staphyl.	aureus	GISA-904062						HIA Percy	8	55	2	6	2	2	3	6	1	3	1	4	3	7
	★	Saur024	Staphyl.	aureus	BM1 2612	I - 1	YFGFMBQBLO		R	1998	Villiers St Denis	10	55	2	6	2	2	3	6	1	3	1	4	3	7
	★	Saur030	Staphyl.	aureus	HM9	XX			R	2000	Henri Mondor-Créteil	10	55	2	6	2	2	3	6	1	3	1	4	3	7
		Saur088	Staphyl.	aureus	BM9 290	A			R	1987	Hôtel Dieu - PARIS	10	55	2	6	2	2	3	6	1	3	1	4	3	7
12		Saur155	Staphyl.	aureus	SARM-4030210				R		HIA Val de Grâce	10	55	2	6	2	2	3	6	1	3	1	4	3	7
	★	Saur161	Staphyl.	aureus	GISA-Xavier16-10-3						HIA Val de Grâce	10	55	2	6	2	2	3	6	1	3	1	4	3	7
	★	Saur162	Staphyl.	aureus	GISA-1106024		YFGFMBQBLO	JK			HIA Val de Grâce	10	55	2	6	2	2	3	6	1	3	1	4	3	7
13		Saur110	Staphyl.	aureus	BM1 0883	Q			R	1993	Ghent - Belgique	10	55	2	6	2	2	3	6	1	3	1	4	4	7
14		Saur003	Staphyl.	aureus	BM 10675	:29	YHGFO	J	S?	1993	Broussais- Paris	5	55	2	2	2	1	3	6	1	2	1	4	2	7
15		Saur005	Staphyl.	aureus	BM 9520	30b	YHGFMBQBLO	J	S	1988	Broussais- Paris	10	55	2	2	2	1	3	6	1	2	1	4	2	7
16		Saur011	Staphyl.	aureus	BM 10215	:18	YHGFMBQBLO	J	S	1990	Broussais- Paris	10	55	2	2	2	1	3	6	1	2	1	4	3	7
17		Saur009	Staphyl.	aureus	BM 12174	17b			R	1996	Grenoble	10	55	2	4	2	1	3	3	1	3	1	3	3	7
18		Saur080	Staphyl.	aureus	#162	VIId			R	1995	Beaujon - CLICHY	10	55	2	4	2	1	3	3	1	3	1	3	1	7
19		Saur054	Staphyl.	aureus	IPF 648	$			R	2000	Centre Hospitalier Abbeville	10	55	2	6	2	1	3	3	1	3	1	3	1	7
20		Saur156	Staphyl.	aureus	SARM-60300661				R		HIA Val de Grâce	10	55	2	2	2	1	3	3	1	3	1	3	3	7
		Saur158	Staphyl.	aureus	SARM-5040072				R		HIA Val de Grâce	10	55	2	2	2	1	3	3	1	3	1	3	3	7
		Saur007	Staphyl.	aureus	BM 12828	16a	YHGFMBQBLO		R	1999	Broussais- Paris	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur008	Staphyl.	aureus	BM 12830	16b			R	1999	Broussais- Paris	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur043	Staphyl.	aureus	IPF 641	B			R	2000	Centre Hospitalier Nimes	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur044	Staphyl.	aureus	IPF 642	A1			R	2000	Centre Hospitalier Nimes	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur045	Staphyl.	aureus	IPF 643	C			R	2000	Centre Hospitalier Nimes	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur046	Staphyl.	aureus	IPF 644	A2			R	2000	Centre Hospitalier Nimes	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur052	Staphyl.	aureus	IPF 646	$	YHGFMBQBLO		R	2000	Centre Hospitalier Abbeville	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur053	Staphyl.	aureus	IPF 647	#$		e	R	2000	Centre Hospitalier Abbeville	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur059	Staphyl.	aureus	IPF 667	##^			R	2000	Fécamp (Patients hospitalisés.	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur065	Staphyl.	aureus	IPF 92	&			R	2000	Rotschild- Paris	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur066	Staphyl.	aureus	IPF 54	&			R	2000	Rotschild- Paris	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur067	Staphyl.	aureus	IPF 55	&			R	2000	Rotschild- Paris	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur068	Staphyl.	aureus	IPF 56	&			R	2000	Rotschild- Paris	10	55	2	6	2	1	3	3	1	3	1	3	3	7
21		Saur069	Staphyl.	aureus	IPF 57	&			R	2000	Rotschild- Paris	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur070	Staphyl.	aureus	IPF 66	&			R	2000	Rotschild- Paris	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur072	Staphyl.	aureus	#166	Ia			R	1995	Beaujon - CLICHY	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur073	Staphyl.	aureus	#165	Ia			R	1995	Beaujon - CLICHY	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur074	Staphyl.	aureus	#97386	Ib			R	1996	Broussais - PARIS	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur076	Staphyl.	aureus	#97383	VIIa			R	1996	Broussais - PARIS	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur078	Staphyl.	aureus	#97117	VIId	YHGFMBQBLO	e	R	1997	TOULOUSE	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur079	Staphyl.	aureus	#95035	VIId			R	1995	CH - NIMES	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur081	Staphyl.	aureus	#96164	VIId			R	1996	CH BLOIS	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur144	Staphyl.	aureus	SARM-29				R	1996	collection Pitié	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur145	Staphyl.	aureus	SARM-58				R	1996	collection Pitié	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur153	Staphyl.	aureus	SARM-192				R	1996	collection Pitié	10	55	2	6	2	1	3	3	1	3	1	3	3	7
		Saur055	Staphyl.	aureus	IPF 654	^			R	1999	Fécamp (Patients hospitalisés.	10	55	2	6	2	1	3	3	1	3	1	2	3	7
22		Saur056	Staphyl.	aureus	IPF 657	^			R	1999	Fécamp (Patients hospitalisés.	10	55	2	6	2	1	3	3	1	3	1	2	3	7
		Saur057	Staphyl.	aureus	IPF 659	^			R	2000	Fécamp (Patients hospitalisés.	10	55	2	6	2	1	3	3	1	3	1	2	3	7
23		Saur006	Staphyl.	aureus	#99135	:31	YHGFMBQBLO		S	1999	St Louis-Paris	10	55	2	6	2	1	3	3	1	3	1	4	3	7
		Saur075	Staphyl.	aureus	#97381	VIIb			R	1996	Broussais - PARIS	10	55	2	6	2	1	3	3	1	0.1	1	3	3	7



24	Saur075	Staphyl.	aureus	#97381	VIIb				R	1996	Broussais - PARIS	10	55	2	6	2	1	3	3	1	0.1	1	3	3	7
25	Saur157	Staphyl.	aureus	SARM-60300663					R		HIA Val de Grâce	10	55	2	2	2	1	3	3	1	0.1	1	3	3	7
26	Saur077	Staphyl.	aureus	#97379	VIIa				R	1996	Broussais - PARIS	10	55	2	6	2	1	3	3	1	8	1	3	3	7
27	Saur086	Staphyl.	aureus	#97120	VIb				R	1997	TOULOUSE	10	55	2	6	2	2	3	3	1	0.1	1	3	3	7
	Saur014	Staphyl.	aureus	#2001045	24a				R	1998	Toulouse	10	55	2	6	2	2	3	3	1	3	1	3	3	7
	Saur082	Staphyl.	aureus	#920	VIb				R	1995	St Germain	10	55	2	6	2	2	3	3	1	3	1	3	3	7
	Saur084	Staphyl.	aureus	#12072	VIb				R	1995	Hop. L. Mourrier-COLOMBES	10	55	2	6	2	2	3	3	1	3	1	3	3	7
	Saur085	Staphyl.	aureus	#12068	VIb				R	1995	CH - EVRY	10	55	2	6	2	2	3	3	1	3	1	3	3	7
28	Saur109	Staphyl.	aureus	BMI 0827	K			JK	R	1993	BORDEAUX		55	2	6	2	2	3	3	1	3	1	3	3	7
	Saur149	Staphyl.	aureus	SARM-131		YHGFMBQBLO			R	1996	collection Pitié	10	55	2	6	2	2	3	3	1	3	1	3	3	7
	Saur151	Staphyl.	aureus	SARM-135					R	1996	collection Pitié	10	55	2	6	2	2	3	3	1	3	1	3	3	7
	Saur152	Staphyl.	aureus	SARM-142					R	1996	collection Pitié	10	55	2	6	2	2	3	3	1	3	1	3	3	7
29	Saur004	Staphyl.	aureus	#2001042	30a				S	1999	Toulouse	10	55	2	6	2	2	3.5	3	1	3	1	3	3	7
30	Saur012	Staphyl.	aureus	BM 12942	:19				R	1999	Broussais- Paris	10	55	2	6	2	2	3	5	1	3	1	3	3	7
31	Saur083	Staphyl.	aureus	#1625	VIb				R	1995	Beaujon - CLICHY	10	55	2	6	2	3	3	3	1	3	1	3	3	7
32	Saur148	Staphyl.	aureus	SARM-119					R	1996	collection Pitié	10	55	2	6	2	2	3	3	1	3	1	3	3	5
33	Saur146	Staphyl.	aureus	SARM-88		YHGFMBQBLO			R	1996	collection Pitié	10	44	2	6	2	2	3	3	1	3	1	3	3	7
	Saur013	Staphyl.	aureus	#97233	24h				R	1997	Broussais- Paris	10	55	2	6	0.1	2	3	3	1	3	1	3	3	7
34	Saur015	Staphyl.	aureus	#93184	:26				R	1993	Laennec- Paris	10	55	2	6	0.1	2	3	3	1	3	1	3	3	7
35	Saur041	Staphyl.	aureus	IPF 511	:3	ZM2PNGKBKGOLB			S	2001	Hôpital pédiatrique Tunisie	12	99	4	2	0.1	1	5	6	5.5	0.2	2	2	3	7
36	Saur134	Staphyl.	aureus	SARM ref JV		XKAOBO			R		collection Pitié	6	77	0.1	5	0.1	1	6	0.1		4	0.1	3	3	7
37	Saur042	Staphyl.	aureus	IPF 512	48a	WGKAKAOMQQQ			S	2001	Hôpital pédiatrique Tunisie	11	44	3	3	0.1	2	5	0.1	2	0.2	5	2	4	7
38	Saur175	Staphyl.	aureus	MRSA252		WGKAKAOMQQQ		HI	R		nosocomial EMRSA16	11	44	3	3	0.1	2	3	5	2	0.2	5	2	4	7
39	Saur035	Staphyl.	aureus	BMI 2975	45c	WGAKBMQ		HI	S	1999	Hôpital pédiatrique Tunisie	7	44	3	3	0.1	2	5	0.1	2	0.2	3	2	3	7
40	Saur016	Staphyl.	aureus	BM 12286	36a			HI	S?	1996	Beaujon-Clichy		66	2	3	2	2	5	3	2	0.2	5	2	3	7
41	Saur017	Staphyl.	aureus	BM 12287	36a			TI	S?	1996	Beaujon-Clichy	9	66	2	3	2	2	5	0.1	2	0.2	5	2	3	7
42	Saur138	Staphyl.	aureus	SASM-6		A2AKBEMBKB	d		S		collection Pitié	9	66	5.5	5	0.1	1	7	0.1	1	0.2	5	3	1	7
	Saur143	Staphyl.	aureus	SARM-15			d		R	1996	collection Pitié	9	66	5.5	5	0.1	1	7	0.1	1	0.2	5	3	1	7
43	Saur040	Staphyl.	aureus	IPF 510	37a	XKAKBBEMEKB	b		S	2001	Hôpital pédiatrique Tunisie	11	66	5.5	5	0.1	1	7	0.1	1	0.2	5	3	1	7
44	Saur036	Staphyl.	aureus	IPF160	38c	XKB	d		S	2000	Hôpital pédiatrique Tunisie	3	66	5.5	5	0.1	1	7	0.1	1	0.2	4	3	1	7
	Saur047	Staphyl.	aureus	IPF735	*	XKAKBMB	b		S	2001	Centre Hospitalier Calais	7	66	5.5	5	0.1	1	7	0.1	1	0.2	5	3	1	7
	Saur048	Staphyl.	aureus	IPF736	*		b		S	2001	Centre Hospitalier Calais	7	66	5.5	5	0.1	1	7	0.1	1	0.2	5	3	1	7
45	Saur049	Staphyl.	aureus	IPF738	*	XKAKBMB	d		S	2001	Centre Hospitalier Calais	7	66	5.5	5	0.1	1	7	0.1	1	0.2	5	3	1	7
	Saur050	Staphyl.	aureus	IPF741	*	XKAKBMB	b		S	2001	Centre Hospitalier Calais	7	66	5.5	5	0.1	1	7	0.1	1	0.2	5	3	1	7
	Saur051	Staphyl.	aureus	IPF743	*	XKAKBMB	b		S	2001	Centre Hospitalier Calais	7	66	5.5	5	0.1	1	7	0.1	1	0.2	5	3	1	7
46	Saur058	Staphyl.	aureus	IPF658	#^	A2AKEEMBKB	d		R	2000	Fê camp (Patients hospitalisés.	9	66	5.5	5	0.1	1	8	0.1	1	0.2	5	3	1	7
47	Saur147	Staphyl.	aureus	SARM-100		A2AKEEMBKB	b		R	1996	collection Pitié	9	66	5.5	5	0.1	1	8	0.1	1	0.2	0.1	3	1	7
•	Saur113	Staphyl.	aureus	Mu50					R			10	66	3	4	3	6		5	4	4	5	3	3	5
48	Saur159	Staphyl.	aureus	Mu3(SARM)				BACDEF	R		HIA Val de Grâce	10	66	3	4	3	6	7	5	4	4	5	3	3	5
	Saur160	Staphyl.	aureus	Mu50(GISA)		TJMBMDMGMK		BACDEF	R		HIA Val de Grâce	10	66	3	4	3	6	7	5	4	4	5	3	3	5
•	Saur170	Staphyl.	aureus	Mu50		TJMBMDMGMK		BACDEF	R		nosocomial	10	66	3	4	3	6	7	5	4	4	5	3	3	5
	Saur087	Staphyl.	aureus	#97373	XVIII	TJMBMDMGMK			R	1996	Broussais - PARIS	10	66	3	4	3	6	7	5	4	4	5	3	3	7
49 •	Saur174	Staphyl.	aureus	N315		TJMBMDMGMK		BACDEF	R		nosocomial	10	66	3	4	3	6	7	5	4	4	5	3	3	7
50	Saur001	Staphyl.	aureus	BMI 2846	:1	TJMBMDMGMK			S	1999	Broussais- Paris	10	66	2	4	3	4	7	5	5	4	5	1	3	7
51	Saur002	Staphyl.	aureus	BMI 2451	:2	YC2FMBQBLOO		BAALMMNNN	S	1997	Grenoble	10	66	2	6	0.1	9	6	5		3	6	3	2	9
52	Saur141	Staphyl.	aureus	SASM-11		UJGFMB		BOPQK	S		collection Pitié	6	66	4	4	0.1	5	5	5	2	2	1	3	2	36
	Saur034	Staphyl.	aureus	BM 12634	36a	H2GJAABB		BAM	S	1998	Hôpital pédiatrique Tunisie	7	55	1	6	0.1	3	7	5	1	1	4	4	2	36
53	Saur142	Staphyl.	aureus	SASM-12		UKGJB		BMD	S		collection Pitié	5	44	1	2	0.1	3	7	5	1	3	4	4	2	32
54	Saur033	Staphyl.	aureus	BM 12632	40e				S	1998	Hôpital pédiatrique Tunisie	9	55	2	5	4	6	5	1		3	4	4	2	7
55	Saur061	Staphyl.	aureus	IPF309	~	ZGGM		JRMQSD	S	2001	CH Intercommunal Créteil	4	55	2	5	4	6	5	1		3	4	3	2	7
56	Saur062	Staphyl.	aureus	IPF310	~	UJGBBBGGJ		UREKKMMN	S	2001	CH Intercommunal Créteil	9	55	2.5	2	0.1	8	7	7	5	3	3	3	2	36
57	Saur063	Staphyl.	aureus	IPF311	~	UJGBBBGGJ		UREKKMMN	S	2001	CH Intercommunal Créteil	9	55	2.5	2	0.1	8	7	7	5	3	3	3	2	36
	Saur064	Staphyl.	aureus	IPF323	~	UJGBBBGGJ			S	2001	CH Intercommunal Créteil	9	55	2.5	2	0.1	8	7	7	5	3	3	3	2	36
58	Saur060	Staphyl.	aureus	IPF308	~	UJGBBBGGJ		fREKKMMN	S	2001	CH Intercommunal Créteil	9	55	2.5	2	0.1	8	7	7	4	3	3	3	2	36
59	Saur038	Staphyl.	aureus	IPF518	22a				S	2001	Hôpital pédiatrique Tunisie	9	66	1	2	0.1	1	7	1		4	2	3	2	36
60	Saur137	Staphyl.	aureus	SASM-5		H2M2EGMMJK2M			S		collection Pitié	9	55	1.5	4	0.1	3	7	3	1	0.1	3	3	1	7
61	Saur139	Staphyl.	aureus	SASM-7		H2M2EGMMJK2M			S		collection Pitié	9	55	1.5	4	0.1	3	7	3	1	0.1	3	3	1	36
62	Saur140	Staphyl.	aureus	SASM-8		H2M2EGMMJK2M			S		collection Pitié	9	55	1.5	4	0.1	3	7	3	1	0.1	3	3	1	32
63	Saur039	Staphyl.	aureus	IPF509	:57	H2M2EGMMJK2M		VWWW	S	2001	Hôpital pédiatrique Tunisie	9	55	1.5	4	0.1	4	7	3	1	0.2	3	3	1	33
64 •	Saur171	Staphyl.	aureus	MSSA476		UKJFKBPE		BG	S		community aquired	8	55	3	6	2	2	4	5	2	3	3	1	1	7
65 •	Saur173	Staphyl.	aureus	MW2		UJJFKBPE		B	R		community aquired	8	55	4	8	2	1	4	5	3	2	3	3	1	7
66	Saur037	Staphyl.	aureus	IPF489	42a			XYZK	S	2001	Hôpital pédiatrique Tunisie	10	44	2.5	6	0.1	4	4	5	3	2	3	3	4	7
67	Saur135	Staphyl.	aureus	SASM-2		UJFKBPE			S		collection Pitié	7	55	2	6	2	2	4	4	3	3	3	4	1	36
68	Saur136	Staphyl.	aureus	SASM-4		UJFKBPE		BG	S		collection Pitié	7	55	2	6	2	2	4	4	3	3	3	4	1	7

Figure 29 : Arbre MLVA 140 souches *S. aureus.*

On peut observer deux grands groupes de souches de même génotype MLVA. En comparant avec les génotypes de champ pulsé, on constate que plusieurs génotypes PFGE sont présents à l'intérieur de ces deux grands génotypes MLVA. Le champ pulsé apparaît ici plus résolutif que l'analyse MLVA ; cependant, les difficultés de reproductibilité et de comparaison des génotypes nous incitent à une certaine prudence face à la solidité des données PFGE. Même si le MLVA est moins résolutif ici, la robustesse des données et la reproductibilité en font un outil de génotypage plus satisfaisant.

Si l'on observe la répartition des souches selon leur sensibilité à la méthicilline, on constate que dans quelques rares cas des souches de même génotype MLVA sont sensibles ou résistantes à la méthicilline. Sinon, le complexe principal est constitué de façon homogène de souches MRSA. Des allèles de ms1132 de taille égale sont trouvés dans les différents complexes. Pour ms0266 (coa) qui évolue lentement, ainsi que pour ms0122 (spa) qui évolue plus rapidement, les allèles de même taille sont assez bien regroupés dans les différents complexes. Le plus souvent, les souches provenant d'un même hôpital sont groupées ensemble et ont parfois des génotypes uniques. Par exemple les souches « Calais » sont de même génotype MLVA et celui-ci est unique dans l'arbre. En champ pulsé, un seul génotype a également été déterminé pour les souches Calais.

3.3.2 Séquençage des locus spa et Mu50_1132

Dans un certain nombre de cas, il semble que le typage MLVA puisse être complété par, le cas échéant, le séquençage d'allèles de répétitions en tandem. Nous en avons fourni une première illustration pour *P. aeruginosa*, *S. aureus* en est un autre exemple. Nous allons l'illustrer avec 2 locus.

Le développement du séquençage de quelques marqueurs polymorphes permettant d'obtenir une résolution d'analyse équivalente à celle obtenue par analyse MLVA, réduirait sensiblement le nombre de marqueurs à analyser. En effet, le séquençage est une méthode de typage rapide à mettre en œuvre, qui permet de comparer sans ambiguïté les résultats de typages entres laboratoires, permettant ainsi la mise en commun des données concernant des milliers de souches analysées à travers le monde.

Deux locus ont été séquencés dans ce travail : *spa* et ms1132. Nous avons choisi la répétition située dans le gène *spa* codant la protéine A, pour pouvoir faire un lien entre la littérature concernant le typage des souches de *S. aureus* et nos données. Le choix de ms1132 s'explique par le fait qu'il s'agit de l'un des marqueurs les plus polymorphes parmi les 14 VNTRs de l'analyse MLVA. Pour *spa*, 11 allèles ont été observés (de 3 à 13 motifs) dans la collection de souches étudiées, et 8 allèles (de 1 à 9 motifs) pour ms1132. Les indices de polymorphisme de *spa* et de ms1132 sont respectivement de 0,68 et de 0,64.

Dans les deux exemples illustrés ici, le codage des allèles est avant tout purement descriptif, c'est un autre mode de représentation des allèles. Chaque motif est codé par une lettre. Il n'y a pas de proximité phylogénétique entre deux motifs proches dans le codage, par exemple « a » n'est pas forcément plus proche de « b » par sa séquence.

3.3.2.1 *Spa*

La répétition en tandem située dans le gène de la protéine A (spa) a déjà été étudiée par Shopsin *et al*. Cette répétition a été séquencée pour plusieurs souches et les allèles codés par des lettres. Il y a au total 37 motifs différents dans le codage de Shopsin *et al.,* (Shopsin 1999). Chaque lettre a été assignée à un motif sans prendre en considération le nombre de mutations entre motifs. La liste des motifs est présentée en Annexe 4.

Le codage dépend dans une certaine mesure de la définition du point de départ de la répétition en tandem. Dans le cas présent, un point de départ décalé de quelques paires de bases aurait permis de réduire légèrement cette liste, cependant la convention « Shopsin » constitue désormais une référence qu'il a semblé préférable de conserver. Certaines tailles d'allèles vues dans notre étude n'ont pas été observées dans l'étude de Shopsin (3U, 6U, 12U et 13U).

Pour pouvoir évaluer la cohérence de l'arbre obtenu avec l'analyse MLVA présentée ici, nous avons en parallèle amplifié et séquencé la répétition située dans le gène *spa* pour 51 souches de la collection. Au total nous disposons donc de 57 allèles spa codés, en ajoutant les allèles des 6 génomes complets.

Tableau 15 : Nombre de codages spa différents observés.

allèles *spa* séquencés (taille en nombre de motifs)	Nombre de codages différents
1 allèle 13U	1
1 allèle 12U	1
9 allèles 11U	4
16 allèles 10U	4
11 allèles 9U	3
5 allèles 8U	3
8 allèles 7U	4
2 allèles 6U	2
2 allèles 5U	2
1 allèle 4U	1
1 allèle 3U	1

Le Tableau 16 présente le résultat du codage des 57 allèles spa.

Tableau 16 : Codage des allèles spa séquencés.

souches séquencées:	codage spa:	u: nombre de motifs	souches séquencées:	codage spa:	u: nombre de motifs
>Saur026-0122	YHFGFMBQBQBLO	13U	>Saur138-0122	A2AKBEMBKB	9U
>Saur041-0122	Z**M2**PNGKBKGOLB	12U	>Saur147-0122	A2AKEEMBKB	
>MRSA252_0122_11U	**WGKAKAOMQQQ**	11U	>Saur058-0122	A2AKEEMBKB	
>Saur042-0122	WGKAKAOMQQQ		>Saur060-0122	UJGBBBGGJ	
>NCTC8325_0122_11U	**YHGGFMBQBLO**		>Saur063-0122	UJGBBBGGJ	
>Saur029-0122	YHFGFMBQBLO		>Saur064-0122	UJGBBBGGJ	
>Saur032-0122	YHFGFMBQBLO		>Saur062-0122	UJGBBBGGJ	
>Saur071-0122	YHFGFMBQBLO		>Saur139-0122	H2**M2**EGMMJK2M	
>Saur028-0122	YHFGFMBQBLO		>Saur039-0122	H2**M2**EGMMJK2M	
>Saur102-0122	YHFGFMBQBLO		>Saur140-0122	H2**M2**EGMMJK2M	
>Saur040-0122	XKAKBBEMEKB		>Saur137-0122	H2**M2**EGMMJK2M	
>Saur160-0122	TJMBMDMGMK	10U	**>MSSA_0098_8U**	**UKJFKBPE**	8U
>N315_0122_10U	**TJMBMDMGMK**		**>MW2_0099_8U**	**UJJFKBPE**	
>Mu50_0122_10U	**TJMBMDMGMK**		>Saur018-0122	YFGFMBLO	
>Saur087-0122	TJMBMDMGMK		>Saur019-0122	YFGFMBLO	
>Saur001-0122	TJMBMDMGMK		>Saur021-0122	YFGFMBLO	
>Saur024-0122	YFGFMBQBLO		>Saur034-0122	H2GJAABB	7U
>Saur162-0122	YFGFMBQBLO		>Saur035-0122	WGAKBMQ	
>Saur007-0122	YHGFMBQBLO		>Saur047-0122	XKAKBMB	
>Saur052-0122	YHGFMBQBLO		>Saur049-0122	XKAKBMB	
>Saur005-0122	YHGFMBQBLO		>Saur050-0122	XKAKBMB	
>Saur146-0122	YHGFMBQBLO		>Saur051-0122	XKAKBMB	
>Saur149-0122	YHGFMBQBLO		>Saur136-0122	UJFKBPE	
>Saur011-0122	YHGFMBQBLO		>Saur135-0122	UJFKBPE	
>Saur078-0122	YHGFMBQBLO		>Saur134-0122	XKAOBO	6U
>Saur006-0122	YHGFMBQBLO		>Saur141-0122	UJGFMB	
>Saur002-0122	YC2FMBQBLOO		>Saur142-0122	UKGJB	5U
			>Saur003-0122	YHGFO	
			>Saur061-0122	ZGGM	4U
			>Saur036-0122	XKB	3U

Nous avons séquencé un nouveau motif, M2, qui comporte une délétion de 3pb. Ce motif a été observé dans 5 souches, dont 4 sont très proches dans l'arbre obtenu par l'analyse MLVA. On constate que, parmi des allèles de même taille, plusieurs séquences différentes sont observées. L'intérêt majeur du séquençage d'allèles est d'affiner la précision du typage lorsque plusieurs allèles sont de même taille. On peut ainsi créer des sous-groupes parmi ces allèles proches sur le simple critère de taille de la répétition. Par exemple pour les allèles à 11 motifs, 4 types de séquences ont été observées. Le Tableau 15 récapitule le nombre de codages différents observés pour chaque taille d'allèles.

Les 57 souches séquencées se répartissent en 26 codages différents. Par typage MLVA, ces souches se répartissent en 45 génotypes différents, comme le montre la Figure 30. Dans cet exemple, on voit que le seul séquençage du locus *spa* ne permet pas d'avoir la même résolution des souches qu'avec l'analyse MLVA de 14 locus. Par exemple l'allèle YHGFMBQBLO est retrouvé dans 8 souches, réparties par ailleurs en 6 génotypes par l'analyse MLVA.

On observe aussi différents groupes d'allèles avec des codages très différents. Ceux qui sont très ressemblants du point de vue du codage mais de longueur différente sont proches dans l'arbre. Par exemple XKAKBMB (7 motifs) est plus proche de XKB (3 motifs) que de ZGGM plus proche en taille (4 motifs). Cette proximité des souches XKAKBMB (Saur047, Saur049, Saur050, Saur051), et XKB (Saur036) obtenue par typage MLVA correspond probablement à une réelle proximité biologique des souches. Les souches Saur051 et Saur036 sont identiques pour 13 des 14 locus, elles diffèrent uniquement pour le locus *spa*.

D'autres types de codages particuliers sont retrouvés groupés ensemble dans l'arbre MLVA, il s'agit par exemple des allèles commençant par UJGB, YHGF, H2M2E et UJF.

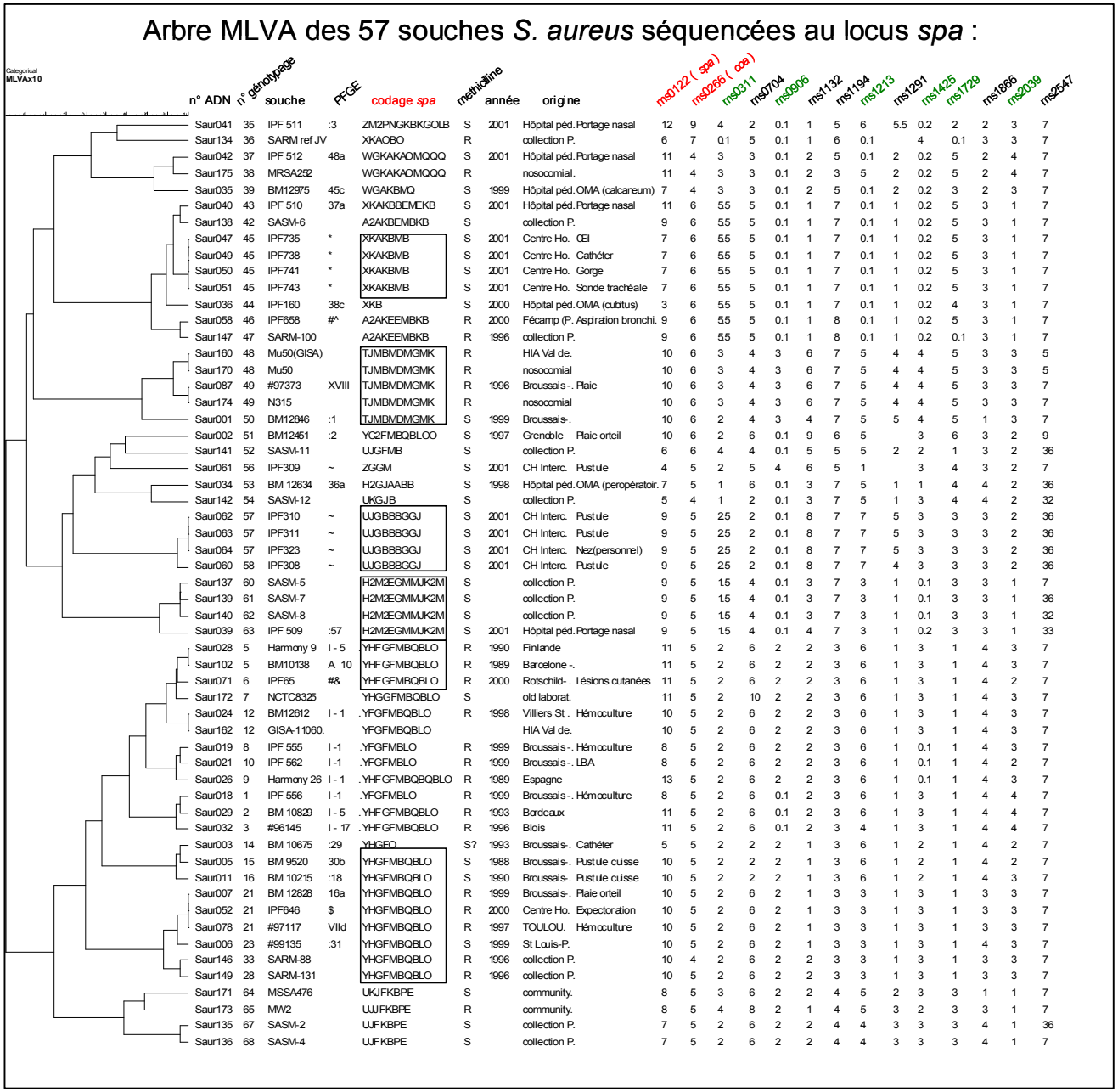

Arbre MLVA des 57 souches S. aureus séquencées au locus spa :
Categorical
MLVAx10
n° ADN n° genotypage souche PFGE codage spa methicilline année origine ms0122 (spa) ms0266 (spa) ms0311 ms0704 ms0906 ms1132 ms1194 ms1213 ms1291 ms1425 ms1729 ms1866 ms2039 ms2547
Saur041 35 IPF 511 :3 ZM2PNGKBKGOLB S 2001 Hôpital péd. Portage nasal 12 9 4 2 0.1 1 5 6 5.5 0.2 2 2 3 7
Saur134 36 SARM ref JV XKAOBO R collection P. 6 7 0.1 5 0.1 1 6 0.1 4 0.1 3 3 7
Saur042 37 IPF 512 48a WGKAKAOMQQQ S 2001 Hôpital péd. Portage nasal 11 4 3 3 0.1 2 5 0.1 2 0.2 5 2 4 7
Saur175 38 MRSA252 WGKAKAOMQQQ R nosocomial. 11 4 3 3 0.1 2 3 5 2 0.2 5 2 4 7
Saur035 39 BM12975 45c WGAKBMQ S 1999 Hôpital péd. OMA (calcaneum) 7 4 3 3 0.1 2 5 0.1 2 0.2 3 2 3 7
Saur040 43 IPF 510 37a XKAKBBEMEKB S 2001 Hôpital péd. Portage nasal 11 6 5.5 5 0.1 1 7 0.1 1 0.2 5 3 1 7
Saur138 42 SASM-6 A2AKBEMBKB S collection P. 9 6 5.5 5 0.1 1 7 0.1 1 0.2 5 3 1 7
Saur047 45 IPF735 * XKAKBMB S 2001 Centre Ho. Œil 7 6 5.5 5 0.1 1 7 0.1 1 0.2 5 3 1 7
Saur049 45 IPF738 * XKAKBMB S 2001 Centre Ho. Cathéter 7 6 5.5 5 0.1 1 7 0.1 1 0.2 5 3 1 7
Saur050 45 IPF741 * XKAKBMB S 2001 Centre Ho. Gorge 7 6 5.5 5 0.1 1 7 0.1 1 0.2 5 3 1 7
Saur051 45 IPF743 * XKAKBMB S 2001 Centre Ho. Sonde trachéale 7 6 5.5 5 0.1 1 7 0.1 1 0.2 5 3 1 7
Saur036 44 IPF160 38c XKB S 2000 Hôpital péd. OMA (cubitus) 3 6 5.5 5 0.1 1 7 0.1 1 0.2 4 3 1 7
Saur058 46 IPF658 #^ A2AKEEMBKB R 2000 Fécamp (P. Aspiration bronchi. 9 6 5.5 5 0.1 1 8 0.1 1 0.2 5 3 1 7
Saur147 47 SARM-100 A2AKEEMBKB R 1996 collection P. 9 6 5.5 5 0.1 1 8 0.1 1 0.2 0.1 3 1 7
Saur160 48 Mu50(GISA) TJMBMDMGMK R HIA Val de. 10 6 3 4 3 6 7 5 4 4 5 3 3 5
Saur170 48 Mu50 TJMBMDMGMK R nosocomial 10 6 3 4 3 6 7 5 4 4 5 3 3 5
Saur087 49 #97373 XVIII TJMBMDMGMK R 1996 Broussais -. Plaie 10 6 3 4 3 6 7 5 4 4 5 3 3 7
Saur174 49 N315 TJMBMDMGMK R nosocomial 10 6 3 4 3 6 7 5 4 4 5 3 3 7
Saur001 50 BM12846 :1 TJMBMDMGMK S 1999 Broussais -. 10 6 2 4 3 4 7 5 5 4 5 1 3 7
Saur002 51 BM12451 :2 YC2FMBQBLOO S 1997 Grenoble Plaie orteil 10 6 2 6 0.1 9 6 5 3 6 3 2 9
Saur141 52 SASM-11 UJGFMB S collection P. 6 6 4 4 0.1 5 5 5 2 2 1 3 2 36
Saur061 56 IPF309 ~ ZGGM S 2001 CH Interc. Pustule 4 5 2 5 4 6 5 1 3 4 3 2 7
Saur034 53 BM 12634 36a H2GJAABB S 1998 Hôpital péd. OMA (peropératoir. 7 5 1 6 0.1 3 7 5 1 1 4 4 2 36
Saur142 54 SASM-12 UKGJB S collection P. 5 4 1 2 0.1 3 7 5 1 3 4 4 2 32
Saur062 57 IPF310 ~ UJGBBBGGJ S 2001 CH Interc. Pustule 9 5 2.5 2 0.1 8 7 7 5 3 3 3 2 36
Saur063 57 IPF311 ~ UJGBBBGGJ S 2001 CH Interc. Pustule 9 5 2.5 2 0.1 8 7 7 5 3 3 3 2 36
Saur064 57 IPF323 ~ UJGBBBGGJ S 2001 CH Interc. Nez(personnel) 9 5 2.5 2 0.1 8 7 7 5 3 3 3 2 36
Saur060 58 IPF308 ~ UJGBBBGGJ S 2001 CH Interc. Pustule 9 5 2.5 2 0.1 8 7 7 4 3 3 3 2 36
Saur137 60 SASM-5 H2M2EGMMJK2M S collection P. 9 5 1.5 4 0.1 3 7 3 1 0.1 3 3 1 7
Saur139 61 SASM-7 H2M2EGMMJK2M S collection P. 9 5 1.5 4 0.1 3 7 3 1 0.1 3 3 1 36
Saur140 62 SASM-8 H2M2EGMMJK2M S collection P. 9 5 1.5 4 0.1 3 7 3 1 0.1 3 3 1 32
Saur039 63 IPF 509 :57 H2M2EGMMJK2M S 2001 Hôpital péd. Portage nasal 9 5 1.5 4 0.1 4 7 3 1 0.2 3 3 1 33
Saur028 5 Harmony 9 I - 5 YHFGFMBQBLO R 1990 Finlande 11 5 2 6 2 2 3 6 1 3 1 4 3 7
Saur102 5 BM10138 A 10 YHFGFMBQBLO R 1989 Barcelone -. 11 5 2 6 2 2 3 6 1 3 1 4 3 7
Saur071 6 IPF65 #& YHFGFMBQBLO R 2000 Rotschild -. Lésions cutanées 11 5 2 6 2 2 3 6 1 3 1 4 2 7
Saur172 7 NCTC8325 YHGGFMBQBLO S old laborat. 11 5 2 10 2 2 3 6 1 3 1 4 3 7
Saur024 12 BM12612 I - 1 .YFGFMBQBLO R 1998 Villiers St . Hémoculture 10 5 2 6 2 2 3 6 1 3 1 4 3 7
Saur162 12 GISA-11060. YFGFMBQBLO HIA Val de. 10 5 2 6 2 2 3 6 1 3 1 4 3 7
Saur019 8 IPF 555 I -1 .YFGFMBLO R 1999 Broussais -. Hémoculture 8 5 2 6 2 2 3 6 1 0.1 1 4 3 7
Saur021 10 IPF 562 I -1 .YFGFMBLO R 1999 Broussais -. LBA 8 5 2 6 2 2 3 6 1 0.1 1 4 2 7
Saur026 9 Harmony 26 I - 1 .YHFGFMBQBQBLO R 1989 Espagne 13 5 2 6 2 2 3 6 1 0.1 1 4 3 7
Saur018 1 IPF 556 I -1 .YFGFMBLO R 1999 Broussais -. Hémoculture 8 5 2 6 0.1 2 3 6 1 3 1 4 4 7
Saur029 2 BM 10829 I - 5 .YHFGFMBQBLO R 1993 Bordeaux 11 5 2 6 0.1 2 3 6 1 3 1 4 4 7
Saur032 3 #96145 I - 17 .YHFGFMBQBLO R 1996 Blois 11 5 2 6 0.1 2 3 4 1 3 1 4 4 7
Saur003 14 BM 10675 :29 YHGFO S? 1993 Broussais -. Cathéter 5 5 2 2 2 1 3 6 1 2 1 4 2 7
Saur005 15 BM 9520 30b YHGFMBQBLO S 1988 Broussais -. Pustule cuisse 10 5 2 2 2 1 3 6 1 2 1 4 2 7
Saur011 16 BM 10215 :18 YHGFMBQBLO S 1990 Broussais -. Pustule cuisse 10 5 2 2 2 1 3 6 1 2 1 4 3 7
Saur007 21 BM 12828 16a YHGFMBQBLO R 1999 Broussais -. Plaie orteil 10 5 2 6 2 1 3 3 1 3 1 3 3 7
Saur052 21 IPF646 $ YHGFMBQBLO R 2000 Centre Ho. Expectoration 10 5 2 6 2 1 3 3 1 3 1 3 3 7
Saur078 21 #97117 VIId YHGFMBQBLO R 1997 TOULOU. Hémoculture 10 5 2 6 2 1 3 3 1 3 1 3 3 7
Saur006 23 #99135 :31 YHGFMBQBLO S 1999 St Louis -P. 10 5 2 6 2 1 3 3 1 3 1 4 3 7
Saur146 33 SARM-88 YHGFMBQBLO R 1996 collection P. 10 4 2 6 2 2 3 3 1 3 1 3 3 7
Saur149 28 SARM-131 YHGFMBQBLO R 1996 collection P. 10 5 2 6 2 2 3 3 1 3 1 3 3 7
Saur171 64 MSSA476 UKJFKBPE S community. 8 5 3 6 2 2 4 5 2 3 3 1 1 7
Saur173 65 MW2 UJJFKBPE R community. 8 5 4 8 2 1 4 5 3 2 3 3 1 7
Saur135 67 SASM-2 UJFKBPE S collection P. 7 5 2 6 2 2 4 4 3 3 3 4 1 36
Saur136 68 SASM-4 UJFKBPE S collection P. 7 5 2 6 2 2 4 4 3 3 3 4 1 7

Figure 30

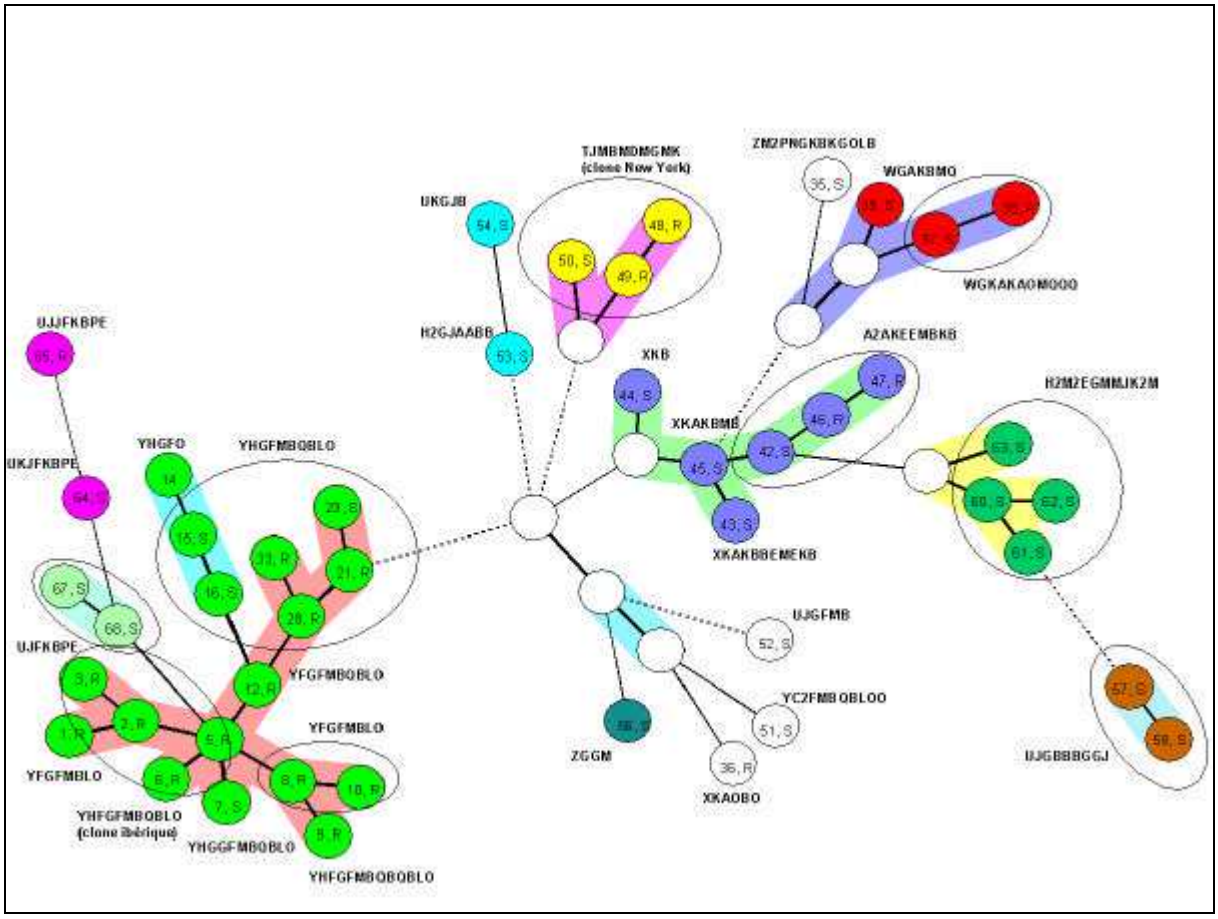


Figure 31

Dans cette représentation « minimum spanning tree » (voir Figure 31), les numéros correspondent aux numéros des génotypes de l'arbre MLVA des 140 souches *S. aureus*. Chaque couleur correspond à une grande branche de l'arbre. R et S correspondent à la résistance ou à la sensibilité à la méthicilline. En blanc, il s'agit des génotypes très éloignés des autres. Les cercles sans numéro de génotype correspondent aux souches hypothétiques qui seraient intermédiaires d'un génotype à l'autre. D'une manière générale, pour que cette réprésentation « minimum spanning tree » soit intéressante, il faut avoir une représentation la plus exhaustive possible des individus de la population (c'est-à-dire un échantillonage global), afin de limiter le nombre de souches hypothétiques. Globalement, si on se refère à l'arbre, les souches MRSA sont assez bien regroupées dans le complexe central (en vert clair), elles présentent pour la majorité un allèle de type YHFG. Les souches MSSA sont réparties dans de nombreux complexes. Le clone ibérique est le clone MRSA le plus répandu actuellement. On retrouve un certain nombre de souches dans l'étude présentée ici qui ont le même codage spa que ce clone : YHFGFMBQBLO. Dans l'étude réalisée par Crisostomo (Crisostomo 2001), toutes les souches MRSA étudiées présentent un codage commençant par YHFGF ou YHGFM, en revanche, parmi les souches MSSA, une grande diversité de codages a été observée (dont des allèles commençant eux aussi par YHGFM ou YHFGF). Un clone MSSA ancien (1963) présente une duplication d'un motif G : YHGGFMBQBLO, ce codage est

observé dans notre étude pour la souche de référence NCTC8325 qui est de type MSSA. Des allèles avec un codage de type TJMBM…, UJFK…ou WGKA… sont aussi retrouvés dans notre étude. Certaines souches MRSA présentent un allèle spa de type UJFK ou WGKA.

3.3.2.2 Mu50_1132

La répétition en tandem ms1132 est localisée dans une séquence codante (protéine hypothétique). Elle a un motif de 63 pb. Trente six allèles représentant les 8 tailles observées sur gel ont été séquencés. Avec les séquences des 6 souches de référence, nous disposons de 42 séquences. Le codage a été réalisé comme pour celui de la protéine A, selon l'ordre d'apparition des motifs dans les séquences, sans classement des motifs selon leur proximité. La Figure 32 présente l'alignement des 34 motifs de 63pb observés dans les séquences des différents allèles ms1132.

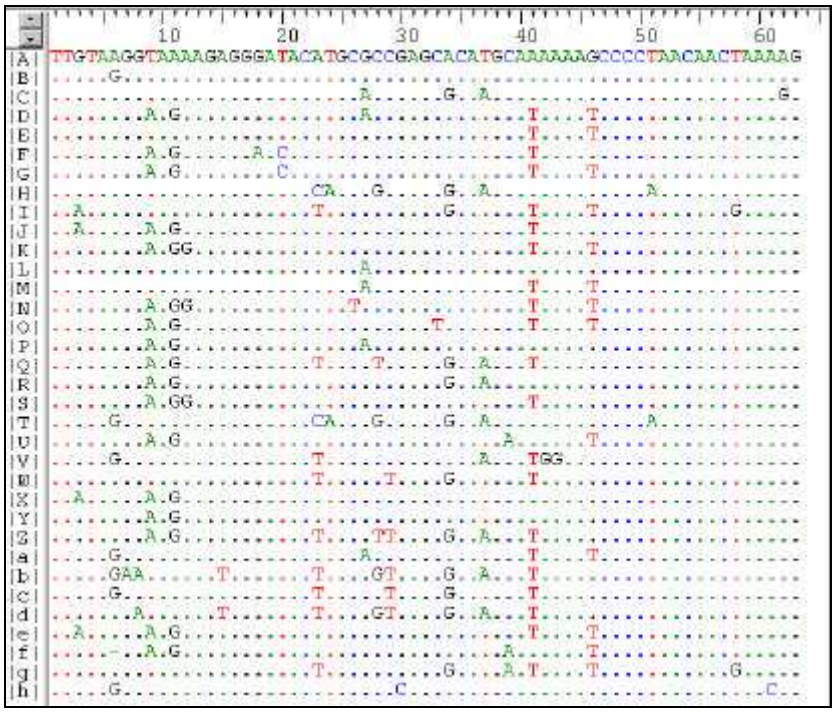


Figure 32 : Alignement des motifs pour le codage des allèles Mu50_1132

Les allèles codés sont présentés dans le Tableau 17.

Tableau 17 : Codage ms1132 des allèles séquencés.

allèles ms1132:	codage des motifs:	allèles ms1132:	codage des motifs:
Saur002_9U_720bp	BAALMMNNN	Saur028_2U_280bp	JK
Saur062_8U_658bp	UREKKMMN	Saur102_2U_280bp	JK
Saur063_8U_658bp	UREKKMMN	Saur109_2U_280bp	JK
Saur060_8U_658bp	fREKKMMN	Saur162_2U_280bp	JK
Mu50_6U_532bp	**BACDEF**	**MW2_1U_217bp**	**B**
N315_6U_532bp	**BACDEF**	Saur003_1U_217bp	J
Saur159_6U_532bp	BACDEF	Saur011_1U_217bp	J
Saur160_6U_532bp	BACDEF	Saur005_1U_217bp	J
Saur061_6U_532bp	JRMQSD	Saur040_1U_217bp	b
Saur141_5U_469bp	BOPQK	Saur047_1U_217bp	b
Saur039_4U_406bp	VWWW	Saur048_1U_217bp	b
Saur037_4U_406bp	XYZK	Saur050_1U_217bp	b
Saur142_3U_343bp	BMD	Saur051_1U_217bp	b
Saur034_3U_343bp	BAM	Saur147_1U_217bp	b
MSSA_2U_280bp	**BG**	Saur049_1U_217bp	d
Saur136_2U_280bp	BG	Saur058_1U_217bp	d
MRSA_2U_280bp	**HI**	Saur036_1U_217bp	d
Saur016_2U_280bp	HI	Saur138_1U_217bp	d
Saur035_2U_280bp	HI	Saur143_1U_217bp	d
Saur017_2U_280bp	TI	Saur053_1U_217bp	e
8325_2U_280bp	**JK**	Saur078_1U_217bp	e

Le Tableau 18 indique le nombre de codages différents observés pour chaque taille d'allèle séquençée.

Tableau 18 : Nombre de codages ms1132 observés.

allèles ms1132 séquencés (taille en nombre de motifs)	nombre de codages différents
1 allèle 9U	1
3 allèles 8U	2
5 allèles 6U	3
1 allèle 5U	1
2 allèles 4U	2
2 allèles 3U	2
11 allèles 2U	4
17 allèles 1U	5

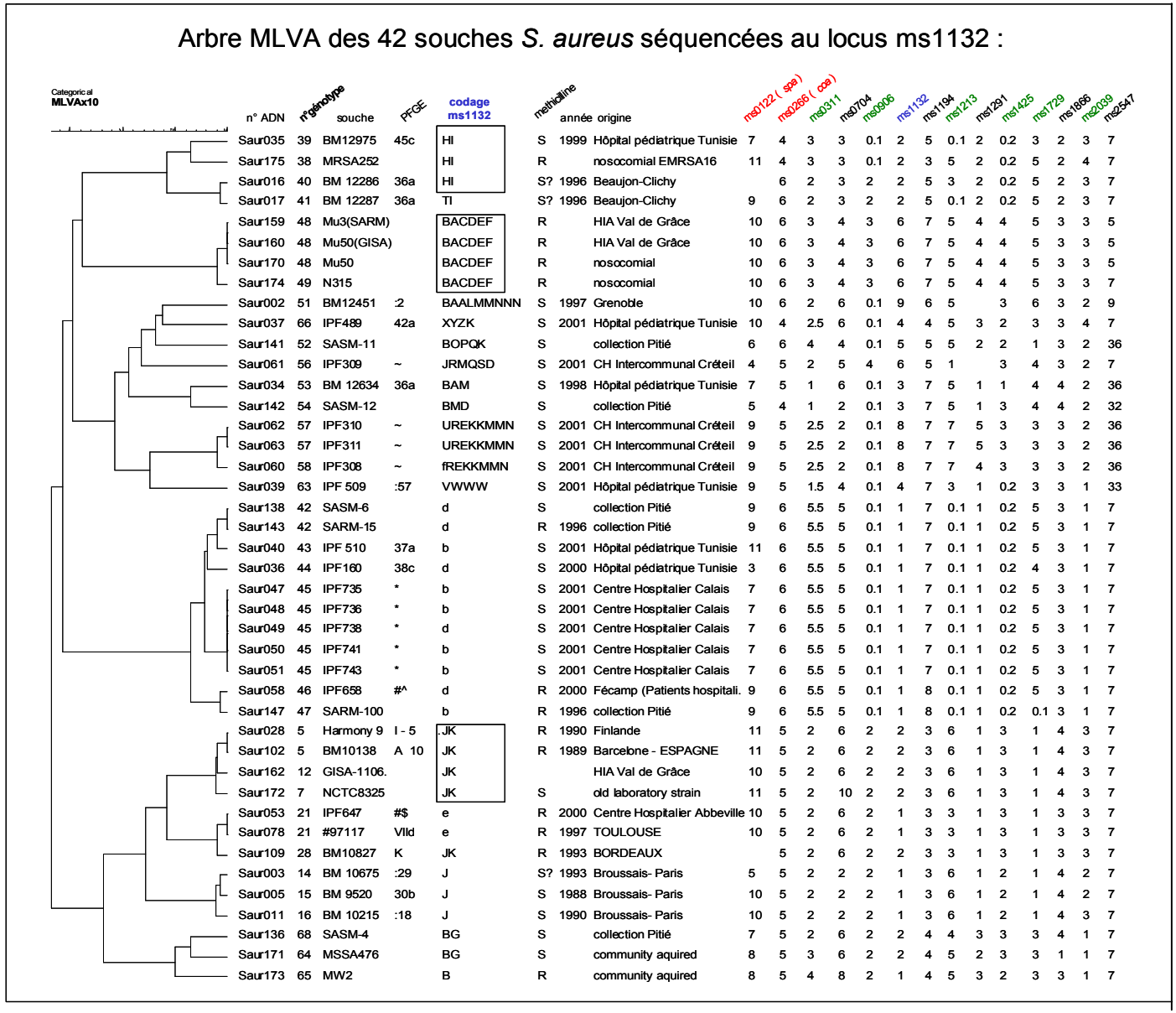

Arbre MLVA des 42 souches S. aureus séquencées au locus ms1132 :
Categorical
MLVAx10
n° ADN
n°génotype
souche
PFGE
codage ms1132
methicilline
année
origine
ms0122 (spa)
ms0266 (coa)
ms0311
ms0704
ms0906
ms1132
ms1194
ms1213
ms1291
ms1425
ms1729
ms1866
ms2039
ms2547
Saur035 39 BM12975 45c HI S 1999 Hôpital pédiatrique Tunisie 7 4 3 3 0.1 2 5 0.1 2 0.2 3 2 3 7
Saur175 38 MRSA252 HI R nosocomial EMRSA16 11 4 3 3 0.1 2 3 5 2 0.2 5 2 4 7
Saur016 40 BM 12286 36a HI S? 1996 Beaujon-Clichy 6 2 3 2 2 5 3 2 0.2 5 2 3 7
Saur017 41 BM 12287 36a TI S? 1996 Beaujon-Clichy 9 6 2 3 2 2 5 0.1 2 0.2 5 2 3 7
Saur159 48 Mu3(SARM) BACDEF R HIA Val de Grâce 10 6 3 4 3 6 7 5 4 4 5 3 3 5
Saur160 48 Mu50(GISA) BACDEF R HIA Val de Grâce 10 6 3 4 3 6 7 5 4 4 5 3 3 5
Saur170 48 Mu50 BACDEF R nosocomial 10 6 3 4 3 6 7 5 4 4 5 3 3 5
Saur174 49 N315 BACDEF R nosocomial 10 6 3 4 3 6 7 5 4 4 5 3 3 7
Saur002 51 BM12451 :2 BAALMMNNN S 1997 Grenoble 10 6 2 6 0.1 9 6 5 3 6 3 2 9
Saur037 66 IPF489 42a XYZK S 2001 Hôpital pédiatrique Tunisie 10 4 2.5 6 0.1 4 4 5 3 2 3 3 4 7
Saur141 52 SASM-11 BOPQK S collection Pitié 6 6 4 4 0.1 5 5 5 2 2 1 3 2 36
Saur061 56 IPF309 ~ JRMQSD S 2001 CH Intercommunal Créteil 4 5 2 5 4 6 5 1 3 4 3 2 7
Saur034 53 BM 12634 36a BAM S 1998 Hôpital pédiatrique Tunisie 7 5 1 6 0.1 3 7 5 1 1 4 4 2 36
Saur142 54 SASM-12 BMD S collection Pitié 5 4 1 2 0.1 3 7 5 1 3 4 4 2 32
Saur062 57 IPF310 ~ UREKKMMN S 2001 CH Intercommunal Créteil 9 5 2.5 2 0.1 8 7 7 5 3 3 3 2 36
Saur063 57 IPF311 ~ UREKKMMN S 2001 CH Intercommunal Créteil 9 5 2.5 2 0.1 8 7 7 5 3 3 3 2 36
Saur060 58 IPF308 ~ fREKKMMN S 2001 CH Intercommunal Créteil 9 5 2.5 2 0.1 8 7 7 4 3 3 3 2 36
Saur039 63 IPF 509 :57 VWWW S 2001 Hôpital pédiatrique Tunisie 9 5 1.5 4 0.1 4 7 3 1 0.2 3 3 1 33
Saur138 42 SASM-6 d S collection Pitié 9 6 5.5 5 0.1 1 7 0.1 1 0.2 5 3 1 7
Saur143 42 SARM-15 d R 1996 collection Pitié 9 6 5.5 5 0.1 1 7 0.1 1 0.2 5 3 1 7
Saur040 43 IPF 510 37a b S 2001 Hôpital pédiatrique Tunisie 11 6 5.5 5 0.1 1 7 0.1 1 0.2 5 3 1 7
Saur036 44 IPF160 38c d S 2000 Hôpital pédiatrique Tunisie 3 6 5.5 5 0.1 1 7 0.1 1 0.2 4 3 1 7
Saur047 45 IPF735 * b S 2001 Centre Hospitalier Calais 7 6 5.5 5 0.1 1 7 0.1 1 0.2 5 3 1 7
Saur048 45 IPF736 * b S 2001 Centre Hospitalier Calais 7 6 5.5 5 0.1 1 7 0.1 1 0.2 5 3 1 7
Saur049 45 IPF738 * d S 2001 Centre Hospitalier Calais 7 6 5.5 5 0.1 1 7 0.1 1 0.2 5 3 1 7
Saur050 45 IPF741 * b S 2001 Centre Hospitalier Calais 7 6 5.5 5 0.1 1 7 0.1 1 0.2 5 3 1 7
Saur051 45 IPF743 * b S 2001 Centre Hospitalier Calais 7 6 5.5 5 0.1 1 7 0.1 1 0.2 5 3 1 7
Saur058 46 IPF658 #^ d R 2000 Fécamp (Patients hospitali. 9 6 5.5 5 0.1 1 8 0.1 1 0.2 5 3 1 7
Saur147 47 SARM-100 b R 1996 collection Pitié 9 6 5.5 5 0.1 1 8 0.1 1 0.2 0.1 3 1 7
Saur028 5 Harmony 9 I - 5 JK R 1990 Finlande 11 5 2 6 2 2 3 6 1 3 1 4 3 7
Saur102 5 BM10138 A 10 JK R 1989 Barcelone - ESPAGNE 11 5 2 6 2 2 3 6 1 3 1 4 3 7
Saur162 12 GISA-1106. JK HIA Val de Grâce 10 5 2 6 2 2 3 6 1 3 1 4 3 7
Saur172 7 NCTC8325 JK S old laboratory strain 11 5 2 10 2 2 3 6 1 3 1 4 3 7
Saur053 21 IPF647 #$ e R 2000 Centre Hospitalier Abbeville 10 5 2 6 2 1 3 3 1 3 1 3 3 7
Saur078 21 #97117 VIId e R 1997 TOULOUSE 10 5 2 6 2 1 3 3 1 3 1 3 3 7
Saur109 28 BM10827 K JK R 1993 BORDEAUX 5 2 6 2 2 3 3 1 3 1 3 3 7
Saur003 14 BM 10675 :29 J S? 1993 Broussais- Paris 5 5 2 2 2 1 3 6 1 2 1 4 2 7
Saur005 15 BM 9520 30b J S 1988 Broussais- Paris 10 5 2 2 2 1 3 6 1 2 1 4 2 7
Saur011 16 BM 10215 :18 J S 1990 Broussais- Paris 10 5 2 2 2 1 3 6 1 2 1 4 3 7
Saur136 68 SASM-4 BG S collection Pitié 7 5 2 6 2 2 4 4 3 3 3 4 1 7
Saur171 64 MSSA476 BG S community aquired 8 5 3 6 2 2 4 5 2 3 3 1 1 7
Saur173 65 MW2 B R community aquired 8 5 4 8 2 1 4 5 3 2 3 3 1 7

Figure 33

On observe une certaine diversité même dans les allèles à 1 motif. Les 42 séquences ms1132 se répartissent en 20 codages différents. Par typage MLVA, ces souches se répartissent en 32 génotypes (voir Figure 33 arbre MLVA). Là aussi, le séquençage d'un seul locus n'atteint pas le pouvoir discriminant de l'analyse MLVA. On peut noter que la répartition des allèles ms1132 dans l'arbre est en accord avec les données MLVA. Globalement, les allèles ne sont pas groupés de façon aléatoire, par exemple tous les allèles ayant J comme premier (ou unique) motif sont regroupés dans la même partie de l'arbre. Une exception cependant, les allèles commençant par un motif B sont dispersés dans l'arbre.

Les souches « Calais », de génotype identique en champ pulsé, avec l'analyse MLVA et par séquençage *spa*, sont divisées en 2 codages ms1132 : allèle b ou allèle d. Lorsque l'on regarde la séquence de ces deux motifs (voir Figure 32) on voit qu'ils sont proches puisqu'ils diffèrent pour seulement 2 nucléotides sur 63. De même, H et T se distinguent par une seule mutation.

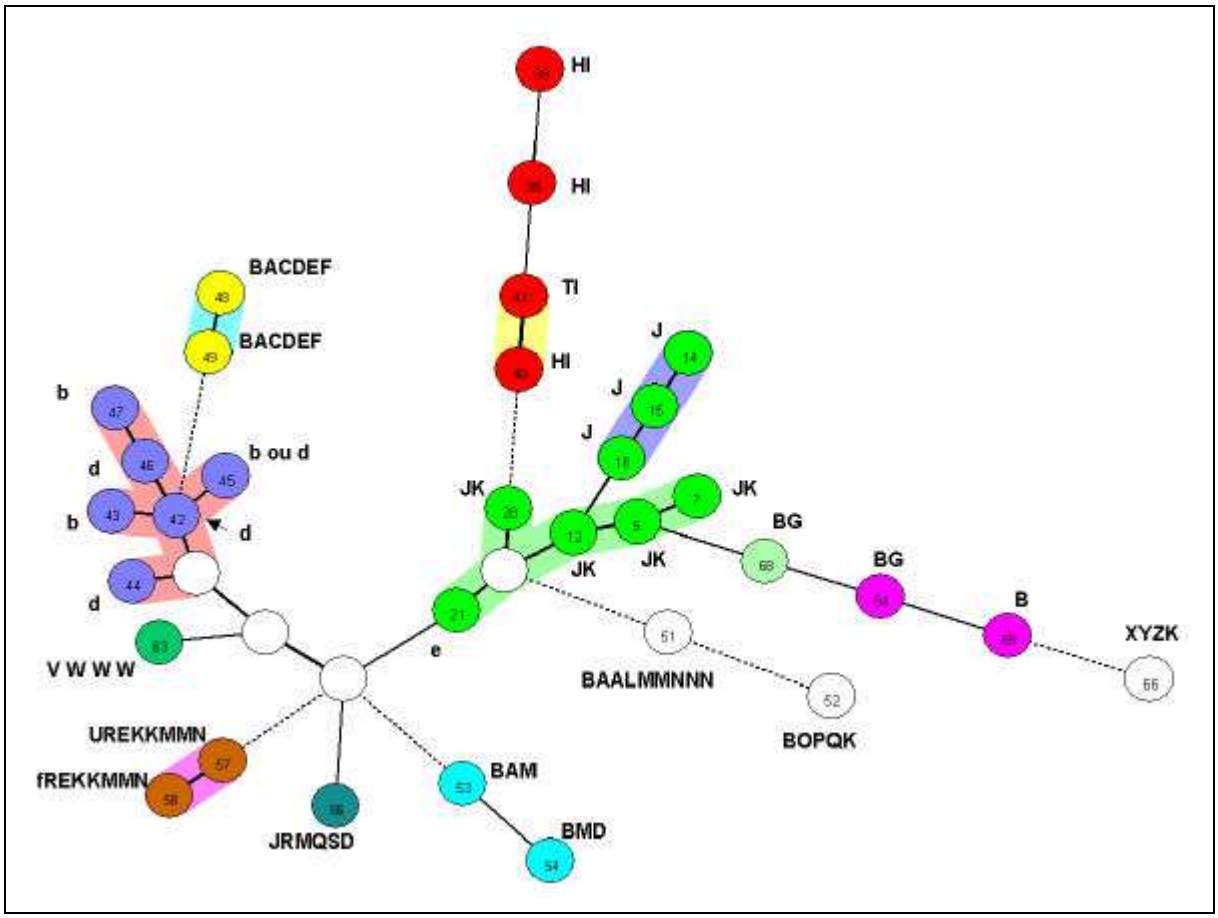


Figure 34 : Représentation « minimum spanning tree » de l'arbre MLVA des 42 souches séquencées au locus ms1132

De cette représentation « minimum spanning tree », il apparaît clairement que les souches commençant par B sont réparties largement. La souche à allèle BAM n'est pas particulièrement proche de celle à allèle BAALMMNNN contrairement à ce que nous avons observé pour le codage spa, où des allèles de taille différente, mais ressemblants en ce qui concerne leur codage, sont proches dans l'arbre. Les souches avec un allèle à 1 motif, b ou d sont toutes regroupées ensembles dans l'un des complexes principaux. Parmi ces souches, on trouve des souches MRSA et quelques MSSA. Le souches qui étaient regroupées par l'analyse MLVA ont dans l'ensemble des allèles avec un codage très identique ou proche.

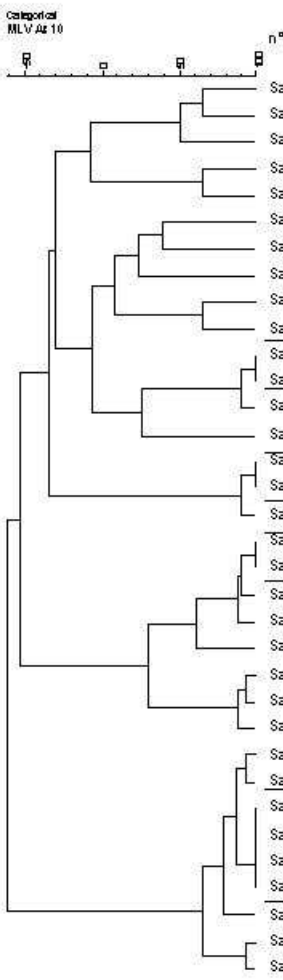
3.3.2.3 Comparaison de la résolution des typages par séquençage spa/ms1132 et par l'analyse MLVA (14 locus) :

Environ un tiers de la collection de souches a été séquencé pour ces deux locus, il était donc intéressant de voir en combien de génotypes le séquençage de *spa* et ms1132 permet de classer ces souches par rapport à l'étude MLVA avec 14 VNTRs.

Avec l'analyse MLVA, les 34 souches se répartissent en 28 génotypes différents. Le codage des allèles des deux locus séquencés a presque la même résolution puisque 27 génotypes sont observés. La Figure 35 représente l'arbre MLVA des 34 souches.

Globalement, on observe une certaine cohérence entre les génotypes obtenus par typage MLVA et les codages spa/ms1132. Des souches proches pour le codage spa sont proches pour le codage ms1132 (congruence des analyses). Ces marqueurs semblent liés. Leurs positions sur le génome sont très éloignées (spa en position 122kb et ms1132 en position 1132kb). Ceci reflète la clonalité observée chez *S.aureus* (Feil 2003), il y a peu de recombinaison, donc les allèles ne sont pas aléatoirement distribués. L'observation de deux types d'allèles ms1132, b et d (qui diffèrent pour 2 nucléotides) pour des souches identiques en codage spa, PFGE et MLVA, est une illustration de l'évolution des *S. aureus* par mutation ponctuelle plus que par recombinaison où l'on verrait un réassortiment des allèles (Smith 1993).

Arbre MLVA (14 locus) des 34 souches *S. aureus* séquencées (*spa* et ms1132):



n° ADN	n° génotype	souche	PFGE	codage *spa*	codage ms1132	méticilline	année	origine	ms0122 (spa)	Ms0266 (coa)	ms0311	ms0704	ms0806	ms1132	ms1194	ms1213	ms1291	ms1425	ms1729	ms1866	ms2039	ms2547
Saur136	68	SASM-4		UJFKBPE	BG	S		collection Pitié	7	5	2	6	2	2	4	4	3	3	3	4	1	7
Saur171	64	MSSA476		UKJFKBPE	BG	S		community aquired	8	5	3	6	2	2	4	5	2	3	3	1	1	7
Saur173	65	MW2		UJJFKBPE	B	R		community aquired	8	5	4	8	2	1	4	5	3	2	3	3	1	7
Saur035	39	BM12975	45c	WGAKBMQ	HI	S	1999	Hôpital pédiatrique Tunisie	7	4	3	3	0.1	2	5	0.1	2	0.2	3	2	3	7
Saur175	38	MRSA252		WGKAKAOMQQQ	HI	R		nosocomial EMRSA16	11	4	3	3	0.1	2	3	5	2	0.2	5	2	4	7
Saur002	51	BM12451	2	YC2FMBQBLOO	BAALMMNNNN	S	1997	Grenoble	10	6	2	6	0.1	9	6	5		3	6	3	2	9
Saur141	52	SASM-11		UJGFMB	BOPQK	S		collection Pitié	6	6	4	4	0.1	5	5	5	2	2	1	3	2	36
Saur061	56	IPF309	~	ZGGM	JRMQSD	S	2001	CH Intercommunal Créteil	4	5	2	5	4	6	5	1		3	4	3	2	7
Saur034	53	BM12634	36a	H2GJAABB	BAM	S	1998	Hôpital pédiatrique Tunisie	7	5	1	6	0.1	3	7	5	1	1	4	4	2	36
Saur142	54	SASM-12		UKGJB	BMD	S		collection Pitié	5	4	1	2	0.1	3	7	5	1	3	4	4	2	32
Saur062	57	IPF310	~	UJGBBBGGJ	UREKKMMN	S	2001	CH Intercommunal Créteil	9	5	2.5	2	0.1	8	7	7	5	3	3	3	2	36
Saur063	57	IPF311	~	UJGBBBGGJ	UREKKMMN	S	2001	CH Intercommunal Créteil	9	5	2.5	2	0.1	8	7	7	5	3	3	3	2	36
Saur060	58	IPF308	~	UJGBBBGGJ	fREKKMMN	S	2001	CH Intercommunal Créteil	9	5	2.5	2	0.1	8	7	7	4	3	3	3	2	36
Saur039	63	IPF509	:57	H2M2EGMMJK2M	WWWW	S	2001	Hôpital pédiatrique Tunisie	9	5	1.5	4	0.1	4	7	3	1	0.2	3	3	1	33
Saur160	48	Mu50(GISA)		TJMBMDMGMK	BACDEF	R		HIA Val de Grâce	10	6	3	4	3	6	7	5	4	4	5	3	3	5
Saur170	48	Mu50		TJMBMDMGMK	BACDEF	R		nosocomial	10	6	3	4	3	6	7	5	4	4	5	3	3	5
Saur174	49	N315		TJMBMDMGMK	BACDEF	R		nosocomial	10	6	3	4	3	6	7	5	4	4	5	3	3	7
Saur028	5	Harmony 9	I - 5	.YHFGFMBQBLO	JK	R	1990	Finlande	11	5	2	6	2	2	3	6	1	3	1	4	3	7
Saur102	5	BM10138	A 10	YHFGFMBQBLO	JK	R	1989	Barcelone - ESPAGNE	11	5	2	6	2	2	3	6	1	3	1	4	3	7
Saur162	12	GISA-1106.		YFGFMBQBLO	JK			HIA Val de Grâce	10	5	2	6	2	2	3	6	1	3	1	4	3	7
Saur172	7	NCTC8325		YHGGFMBQBLO	JK	S		old laboratory strain	11	5	2	10	2	2	3	6	1	3	1	4	3	7
Saur078	21	#97117	VIId	YHGFMBQBLO	e	R	1997	TOULOUSE	10	5	2	6	2	1	3	3	1	3	1	3	3	7
Saur003	14	BM10675	29	YHGFO	J	S	? 1993	Broussais- Paris	5	5	2	2	2	1	3	6	1	2	1	4	2	7
Saur005	15	BM9520	30b	YHGFMBQBLO	J	S	1988	Broussais- Paris	10	5	2	2	2	1	3	6	1	2	1	4	2	7
Saur011	16	BM10215	:18	YHGFMBQBLO	J	S	1990	Broussais- Paris	10	5	2	2	2	1	3	6	1	2	1	4	3	7
Saur040	43	IPF510	37a	XKAKBBEMEKB	b	S	2001	Hôpital pédiatrique Tunisie	11	6	5.5	5	0.1	1	7	0.1	1	0.2	5	3	1	7
Saur138	42	SASM-6		A2AKBEMBKB	d	S		collection Pitié	9	6	5.5	5	0.1	1	7	0.1	1	0.2	5	3	1	7
Saur047	45	IPF735	*	XKAKBMB	b	S	2001	Centre Hospitalier Calais	7	6	5.5	5	0.1	1	7	0.1	1	0.2	5	3	1	7
Saur049	45	IPF738	*	XKAKBMB	d	S	2001	Centre Hospitalier Calais	7	6	5.5	5	0.1	1	7	0.1	1	0.2	5	3	1	7
Saur050	45	IPF741	*	XKAKBMB	b	S	2001	Centre Hospitalier Calais	7	6	5.5	5	0.1	1	7	0.1	1	0.2	5	3	1	7
Saur051	45	IPF743	*	XKAKBMB	b	S	2001	Centre Hospitalier Calais	7	6	5.5	5	0.1	1	7	0.1	1	0.2	5	3	1	7
Saur036	44	IPF160	38c	XKB	d	S	2000	Hôpital pédiatrique Tunisie	3	6	5.5	5	0.1	1	7	0.1	1	0.2	4	3	1	7
Saur058	46	IPF658	#^	A2AKEEMBKB	d	R	2000	Fécamp (Patients hospitali.	9	6	5.5	5	0.1	1	8	0.1	1	0.2	5	3	1	7
Saur147	47	SARM-100		A2AKEEMBKB	b	R	1996	collection Pitié	9	6	5.5	5	0.1	1	8	0.1	1	0.2	0.1	3	1	7

Figure 35

La représentation « minimum spanning tree » des 34 souches nous montre que les souches choisies pour le séquençage sont très diverses du point de vue des génotypes MLVA.

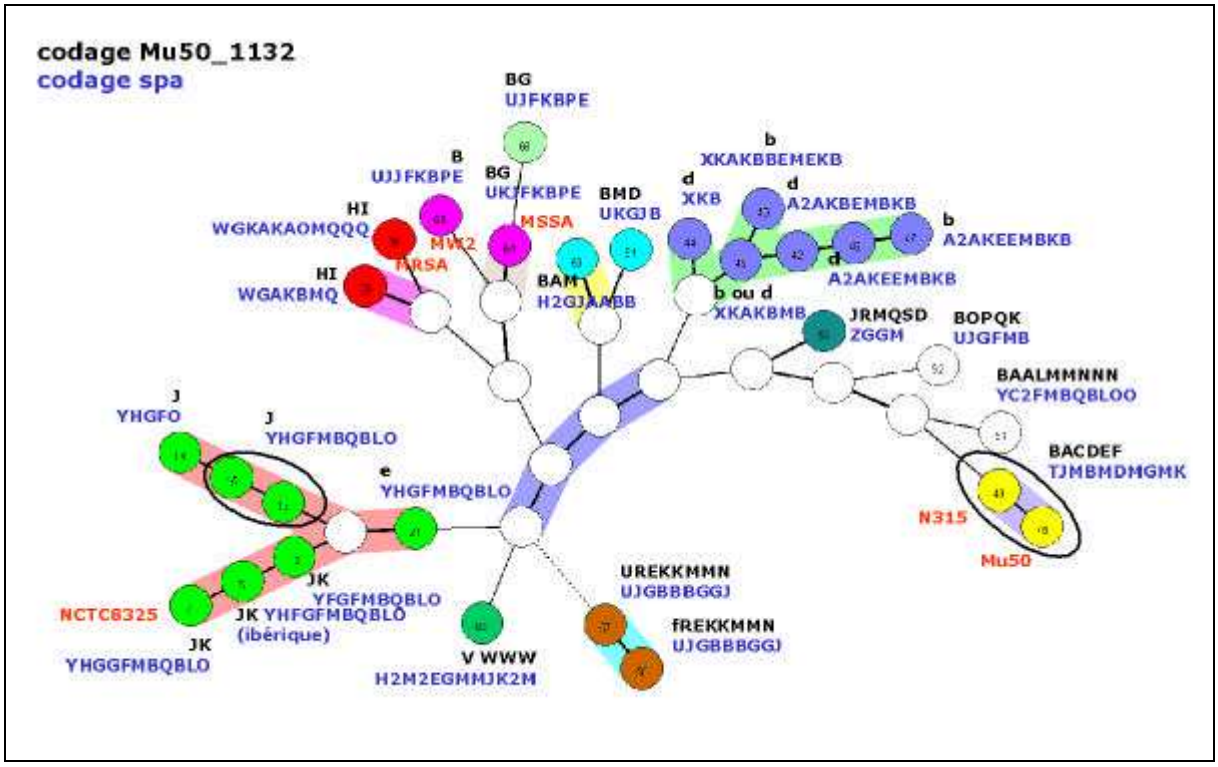


Figure 36 : Représentation « minimum spanning tree » de l'analyse MLVA des 34 souches *S. aureus* séquencées spa/ms1132.

Les souches qui ont un génotype proche par l'analyse MLVA ont aussi des codages d'allèles très proches, il y a une certaine cohérence entre les données de séquençage et celle de l'analyse de 14 VNTRs.

3.3.3 Conclusions de l'étude sur *S. aureus*

Le développement des marqueurs polymorphes pour l'analyse MLVA chez *S. aureus* nous a permis de valider douze nouveaux marqueurs de type VNTRs. Ces douze marqueurs, combinés à *spa* et *coa*, deux marqueurs déjà utilisés pour le génotypage de *S. aureus*, ont été utilisés pour l'analyse MLVA de 137 souches *S. aureus* MRSA, MSSA et GISA provenant pour la majorité d'hôpitaux français. Les souches sont réparties en 68 génotypes différents. Cette résolution semble moins bonne que celle observée par l'analyse de ces souches en PFGE, puisque l'on observe deux grands groupes de souches de même génotype MLVA qui peuvent être réparties en plusieurs génotypes PFGE. Cependant, la robustesse de l'analyse MLVA tant au niveau de la mise en œuvre technique que de l'analyse des résultats, en font un outil puissant d'analyse des souches de *S. aureus*.

Une étude de séquençage d'allèles a été réalisée pour 2 des 14 locus de l'analyse MLVA, *spa* et ms1132. Les allèles séquencés ont été codés par des lettres pour faciliter la comparaison de la composition en motifs des allèles de taille identique, et aussi pour observer des ressemblances de motif entre des allèles de tailles très différentes.

Individuellement, ces marqueurs sont moins résolutifs que l'analyse MLVA, en revanche, ces deux marqueurs combinés atteignent une résolution quasi identique à celle du MLVA.

4 DISCUSSION ET PERSPECTIVES

4.1 Intérêts du typage des répétitions en tandem pour les bactéries pathogènes

Comme nous l'avons illustré dans ce travail, le génotypage par MLVA d'une espèce bactérienne dont le génome complet est accessible est relativement facile à mettre en œuvre. Cela est encore plus vrai lorsque deux souches différentes ont été séquencées. L'absence d'un deuxième génome a été un handicap important pour le projet *P. aeruginosa*, et lorsque le deuxième génome en cours de séquençage aura été publié, il sera sans doute intéressant de réexaminer les données. Les nombreux outils bioinformatiques développés au laboratoire dans le cadre du travail de thèse de France Denoeud (Université Paris Sud, 2003) ont contribué grandement à faciliter les différents projets réalisés au cours de cette thèse.

Par rapport aux techniques de génotypage évoquées dans l'introduction, il apparaît très clairement que l'amplification par PCR de répétitions en tandem est l'outil de génotypage le plus simple et l'un des plus puissants pour la discrimination des souches dans les espèces bactériennes étudiées ici. Cette méthode présente plusieurs avantages:

- Pas ou peu de problème de « typabilité », sauf dans le cas d'espèces dont les membres présentent des pertes de matériel chromosomique importantes ou une grande divergence, et chez qui le choix d'amorces fonctionnant chez toutes les souches est difficile (par exemple pour *N. meningitidis*). Les réarrangements chromosomiques, sans perte de matériel, n'empêchent pas le typage MLVA (par exemple *Y. pestis*).

- Grande sensibilité et spécificité de la technique.

- Bonne reproductibilité intra et inter -laboratoires.

- Pas de profils multibandes à analyser donc moins d'ambiguïté pour l'analyse et pour la comparaison de résultats entre laboratoires (illustré par le développement d'une page de comparaison en ligne des résultats de typage de répétitions en tandem).

- Utilisation de matériel courant de laboratoire.

- Peu coûteuse en termes de consommables et de degré de qualification requis et donc utilisable dans des pays où les infections bactériennes sont importantes, et dont les structures d'analyse ne sont pas forcément aussi développées que dans les pays riches.

- Développable pour d'autres espèces bactériennes. La liste des génomes de bactéries pathogènes pour l'homme, entièrement séquencés et qui n'ont pas encore fait l'objet d'une étude MLVA est longue.

- Délais d'obtention des résultats fortement réduit par rapport aux techniques de phénotypage, on passe de 24 heures minimum (pour des bactéries qui poussent vite) à moins de 6 heures pour un génotypage de type MLVA automatisé.

Le génotypage par MLVA a un fort pouvoir discriminant chez *P. aeruginosa* par rapport au ribotypage, plus faible chez *S. aureus* par rapport à l'analyse PFGE, où l'on observe dans certains cas plusieurs génotypes PFGE pour un seul génotype MLVA. Cependant l'avantage majeur de l'analyse MLVA est que les résultats sont interprétables sans ambiguïté. Le socle de données pour la classification des souches est donc relativement solide pour comparer par la suite d'autres souches avec les données existantes, en développant par exemple des bases de données accessibles à la communauté.

En définitive, cette efficacité du typage de bactéries pathogènes par les répétitions en tandem est somme toute assez étonnante. Un nombre croissant de bactéries d'importance en santé humaine ou animale se trouve se prêter à cette approche. La liste s'allonge rapidement, elle compte déjà le complexe tuberculosis, *Bacillus anthracis*, *Brucella*, *Yersinia pestis*, *Salmonella*, *E. coli O157*, *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Legionella, Neisseria meningitidis*. Il peut être utile de rappeler par ailleurs que les bactéries pathogènes pour l'homme sont très rares, au regard de la diversité bactérienne dans l'environnement. Ces exceptions sont celles qui ont pu s'adapter à l'homme mais leur pathogénicité même met en danger ce biotope, et du même coup leur propre existence. Il est ainsi tentant de spéculer que les pathogènes les plus dangereux sont condamnés par l'évolution. De ce fait, les espèces correspondantes sont jeunes, à structure le plus souvent clonale, génétiquement homogènes. Le fait que les répétitions en tandem soient, en dépit de la forte homogénéité génétique par ailleurs, des sites à polymorphisme élevé, est un argument assez fort suggérant que ces structures contribuent à l'adaptation du pathogène à son hôte. On pourrait illustrer ce point pour chacune des bactéries citées en montrant qu'un certain nombre de répétitions en tandem est associé à des gènes de surface.

De ce point de vue, *Pseudomonas aeruginosa* et *Legionella pneumophila* sont dans une certaine mesure des contre-exemples qui « illustrent la règle ». Ces deux espèces sont très anciennes, et génétiquement très diverses, mais certains représentants, rares, ont des conséquences en santé humaine. *Legionella*, ou au moins certaines souches, s'adapte à des biotopes nouveaux, créés récemment par l'homme : canalisations d'eau chaude, tours de refroidissement, conduites de climatisation. La colonisation de ces biotopes nouveaux va être réalisée par les bactéries les plus versatiles, les plus aptes à s'adapter à un nouveau milieu. La rencontre éventuelle entre ces populations et l'homme qui est un autre biotope à coloniser peut alors conduire à l'émergence d'une nouvelle bactérie pathogène. En résumé, l'apparition de nouveaux biotopes en raison de modifications écologiques (liées au développement de l'agriculture, de l'élevage, de l'urbanisation) favorise ainsi mécaniquement le développement de bactéries particulièrement versatiles, susceptibles, si l'occasion d'un contact se présente, de coloniser un être humain. Le plus souvent, cette colonisation restera anecdotique, parce qu'il n'y aura pas contagion. Parfois, l'homme, ou l'animal, deviendra un biotope à part entière. La bactérie perdra alors la capacité à survivre dans son milieu d'origine. C'est le cas de

Burkholderia mallei, récemment issue de *B. pseudomallei*, de biotope largement restreint aux chevaux, et probablement en voie de disparition. C'est le cas également de *M. tuberculosis*, adaptée à l'homme.

Ces clones, qui réussissent plus vite que d'autres à coloniser des milieux nouveaux, dépendent pour cela de leur capacité d'adaptation génétique. Les changements liés aux répétitions en tandem constituent très certainement une composante de cette capacité. Ces hypothèses et spéculations expliqueraient ainsi, *a posteriori* comme souvent, la raison pour laquelle les bactéries pathogènes, le seraient devenues parce que génétiquement versatiles par nature, se prêtent aussi bien au typage MLVA.

4.2 Rôle fonctionnel de certaines répétitions en tandem

Les répétitions en tandem ont la particularité de pouvoir provoquer des variations de l'expression de gènes, ou des variations de la protéine elle-même, de façon réversible. Parmi l'ensemble des gènes d'une bactérie, une poignée seulement de gènes est affectée par ces mécanismes. La recherche de telles situations a débuté immédiatement après le séquençage en 1995 d'*Haemophilus influenzae,* premier génome bactérien entièrement séquencé (Hood 1996). Neuf locus comportant des unités de 4 pb ont été identifiés. Les séquences homologues dans d'autres génomes bactériens correspondent à des gènes impliqués dans la biosynthèse du LPS (lipopolysaccharide), la fixation du fer, ou encore l'adhésion. Une recherche similaire d'exemples de « variation de phase » liée à la mutation (réversible) de répétitions en tandem a été entreprise pour *N. meningitidis* (Martin 2003). Des preuves expérimentales d'une régulation par variation de phase ont été observées pour 14 gènes. Il s'agit là de la recherche de répétitions en tandem instables, et dont l'instabilité provoque, de façon réversible, l'expression ou l'arrêt de l'expression du gène associé. Ce ne sont donc pas les répétitions en tandem auxquelles nous nous sommes intéressés ici à des fins épidémiologiques. La variation de répétitions en tandem dans la région codante, tout en préservant la phase, est également relativement rare. Nous l'avons vu pour *P. aeruginosa* et *S. aureus*, seuls quelques gènes sont concernés. Parfois les variations de tailles sont remarquables. Par exemple il existe dans *M. tuberculosis* un gène qui contient 3 répétitions en tandem différentes (Le Flèche 2002). La taille de ce gène peut varier de plusieurs kilobases entre différentes souches du complexe tuberculosis.

Il est alors tentant d'utiliser le polymorphisme des séquences répétées en tandem codantes comme révélateur de gènes associés à la pathogénicité. Bien entendu par la suite, l'exploration du rôle fonctionnel de ces « candidats » doit être faite dans des modèles expérimentaux, et dépendra de l'existence d'une part d'outils de génétique permettant de créer

des souches mutées et d'autre part de modèles animaux. Le travail de typage ne constitue donc que l'étape la plus facile de ces approches.

4.3 Recherche de critères prédictifs du polymorphisme des répétitions en tandem

L'analyse des caractéristiques des répétitions en tandem dans les différents génomes bactériens a pour objectif de trouver des critères prédictifs du polymorphisme, à partir de la séquence d'un seul génome.

Lorsque l'on étudie une espèce bactérienne dont un seul génome est entièrement séquencé, il n'est pas possible d'utiliser la méthode de comparaison de génomes pour rechercher des répétitions polymorphes. Il serait donc intéressant de pouvoir sélectionner des répétitions en tandem selon des critères prédictifs de leur polymorphisme à partir des données de séquence d'un seul allèle, comme cela a été montré pour les minisatellites humains (Denoeud 2003). Chez l'homme, les critère prédictifs du polymorphisme des répétitions en tandem sont essentiellement le pourcentage en GC et le critère HistoryR.

Chez les bactéries, nous n'avons pas trouvé de critère prédictif du polymorphisme qui s'appliquerait à tous les génomes bactériens. Des tendances sont observées, comme par exemple le critère de conservation des motifs par rapport au consensus, et la longueur totale de la répétition. Le critère HistoryR développé par Gary Benson (indice de 0 à 1), correspond à un calcul de « coût » pour revenir par contractions successives au motif initial, avec le moins d'événements possibles. Lorsque la reconstruction est « simple », le critère HistoryR est élevé, soit proche de 1, et constitue un assez bon critère prédictif du polymorphisme. Malheureusement, le critère History R n'a pas été retenu comme un bon critère prédictif du polymorphisme chez les bactéries, parce qu'il est souvent égal à 1 en raison de la petite taille des répétitions, et donc peu informatif. Un travail spécifique devrait être réalisé pour adapter son mode de calcul. Dans différentes espèces bactériennes étudiées au laboratoire, nous avons regardé les caractéristiques des séquences des répétitions en tandem pour trouver des corrélations avec le polymorphisme. Une corrélation entre la conservation des motifs et le polymorphisme des répétitions en tandem avait déjà été observée pour *Y. pestis*. Pour *B. anthracis*, il semblerait que les critères corrélés au polymorphisme soient plutôt le pourcentage en GC et la longueur totale de la répétition. Nous avons observé ici que pour *S. aureus* et *P. aeruginosa,* le pourcentage de conservation des motifs serait un bon critère prédictif du polymorphisme associé aux répétitions.

4.4 Etude de population dans les espèces bactériennes étudiées au cours de cette thèse

L'étude de structure des populations chez *P. aeruginosa* a été entreprise récemment. Ces études sont pour l'instant très préliminaires et il est encore difficile de statuer sur la structure des populations de *P. aeruginosa*. Les auteurs ont analysé plusieurs données de typage : des séquences de trois protéines membranaires (oprI, oprL et oprD), des données AFLP, des sérotypages et enfin des typages de la pyoverdine (Pirnay 2002). Les souches étudiées étaient des souches cliniques et des souches environnementales, collectées partout dans le monde. La conclusion de ce travail est que *P. aeruginosa* aurait une structure de population de type épidémique, comparable à celle observée chez *N. meningitidis*, c'est à dire une structure clonale superficielle avec de fréquentes recombinaisons, et parfois l'apparition d'une population épidémique à partir d'un clone qui a « réussi », comme par exemple le clone de sérotype O12 (Pitt 1989)

Nous avons constaté expérimentalement que le sérotype O12, clone très largement répandu en Europe, de ribogroupe 87S-3, peut être réparti en plusieurs génotypes MLVA, essentiellement grâce aux deux marqueurs les plus polymorphes, ms10 et ms61 qui présentent un petit motif répété de 6pb. Ces marqueurs pourraient appartenir aux locus dit de contingence, du fait de la taille du motif et du très grand nombre d'allèles observés, et ne seraient peut être pas utilisables pour des études plus globales de suivi du clone O12 à travers l'Europe. Ces marqueurs permettent de révéler une situation épidémique (par rapport aux autres marqueurs qui semblent beaucoup plus stables).

L'évolution des souches peut être étudiée par SPE (Serial Passage Experiments), c'est-à-dire par des dilutions en série d'une culture bactérienne, afin de tester les variations phénotypiques et génotypiques après des centaines de générations. Nous avons testé la stabilité des répétitions en tandem (à court terme), par des dilutions successives de cultures de *P. aeruginosa* pendant au moins trois semaines. Le milieu de culture était du milieu LB, c'est à dire un milieu riche standard, sans antibiotique. Il pourrait être également intéressant de tester l'effet de différentes conditions de culture (antibiotiques, pH, milieu de culture etc…) sur le polymorphisme des répétitions décrites dans l'article, pour voir s'il existe une régulation par les minisatellites de l'expression de certains gènes selon les conditions expérimentales.

En ce qui concerne les souches de patients atteints de mucoviscidose, il serait intéressant de tester l'analyse MLVA directement à partir du prélèvement dans les poumons. Il faudra très certainement faire des mises au point pour la lyse des bactéries de phénotype mucoïde afin de pouvoir réaliser les PCR. Ceci permettrait de voir rapidement si le patient est infecté par une seule souche ou par plusieurs. En parallèle un ré-isolement sur boite sera certainement

nécessaire afin, dans un second temps et en cas de population hétérogène, d'analyser les clones individuels et de déterminer leur génotype.

Dans l'article « How clonal is *Staphylococcus aureus* », les auteurs ont analysé les données MLST de 334 isolats (isolats de la communauté, isolats hospitaliers et isolats de portage nasal de personnes saines) (Feil 2003). Du fait que l'analyse MLST est basée sur des gènes qui évoluent lentement, elle fournit des données qui permettent des études épidémiologiques globales et donc des études de populations (Maiden 1998). Les auteurs ont fait cette analyse pour savoir si la recombinaison a une contribution plus importante que les mutations ponctuelles dans l'évolution du génome de *S. aureus*. Ils ont fait des comparaisons des séquences MLST pour identifier les mutations entre les différents allèles. Deux souches très proches présenteront de nombreux locus identiques, et si certains locus diffèrent, ce sera pour un petit nombre de nucléotides. Cependant, si les clones se diversifient par recombinaison essentiellement, les allèles qui diffèrent entre souches très proches vont être différents pour un très grand nombre de positions, appuyant le fait que ces allèles ont été importés depuis des lignées non reliées. La conclusion de cette étude des données MLST chez *S. aureus* est que les mutations ponctuelles ont une fréquence quinze fois plus importante que la recombinaison. Cependant, la recombinaison contribue quand même à l'évolution de l'espèce à plus long terme.

Nous avons constaté, dans l'étude réalisée sur 137 souches *S. aureus* par séquençage des allèles à deux locus spa et ms1132, que les différents allèles observés à ces deux locus semblent liés, malgré la distance (environ 1Mb) qui les sépare sur le chromosome de *S. aureus.* Cette observation va bien dans le sens d'une faible recombinaison chez *S. aureus*, mais il faudrait bien sûr confirmer ces observations par des tests statistiques.

4.5 Quelle méthode pour reconstruire l'histoire évolutive des répétitions en tandem à partir de la séquence ?

Un deuxième niveau d'analyse des VNTRs a été exploré dans ce travail, il s'agit du séquençage d'allèles. En effet, nous avons constaté que des allèles de taille identique à un locus donné (et donc considérés comme proches) peuvent avoir des séquences différentes. Il s'agit du phénomène d'homoplasie. Il n'existe pas actuellement de méthodes d'analyse et de comparaison des répétitions en tandem. Les méthodes classiques d'alignement sont peu efficaces du fait du polymorphisme de longueur des séquences à aligner. Par ailleurs, les mécanismes d'évolution des répétitions en tandem (autres que les motifs courts parfaitement conservés qui évoluent par glissement à la réplication), sont mal connus et peu étudiés chez

les bactéries. Il existe un certain nombre d'études mais elles ont porté sur de très longues séquences répétées dispersées dans les génomes bactériens (Rocha 1999), or la plupart des répétitions en tandem situées à un seul locus dans le génome sont des séquences courtes.

Nous avons tenté de faire un codage des allèles selon la séquence des motifs qui se succèdent dans la répétition. Il est parfois possible de déduire « à l'œil » qu'un allèle est issu d'un autre par simple duplication d'un motif par exemple. Le plus souvent, il est difficile d'expliquer simplement comment un allèle est passé d'une taille à une autre. Le développement d'outils d'analyse des motifs permettra de mieux comprendre l'évolution de ce type de séquences répétées.

L'analyse des allèles *spa* très ressemblants pour essayer de voir comment passer simplement d'une taille d'allèle à la suivante est facilitée par le codage d'allèle et aussi par un outil de visualisation des motifs, le TRView développé par Gary Benson et accessible sur sa base de données de répétitions en tandem, le TRDB. Ces deux moyens de visualisation des motifs peuvent aider à repérer des duplications ou délétions de motifs conduisant d'un allèle à un autre. Mais tout cela reste une analyse « manuelle », plus ou moins complexe selon les répétitions étudiées. Un exemple d'allèles spa visualisés avec le TRView est illustré par la Figure 37.

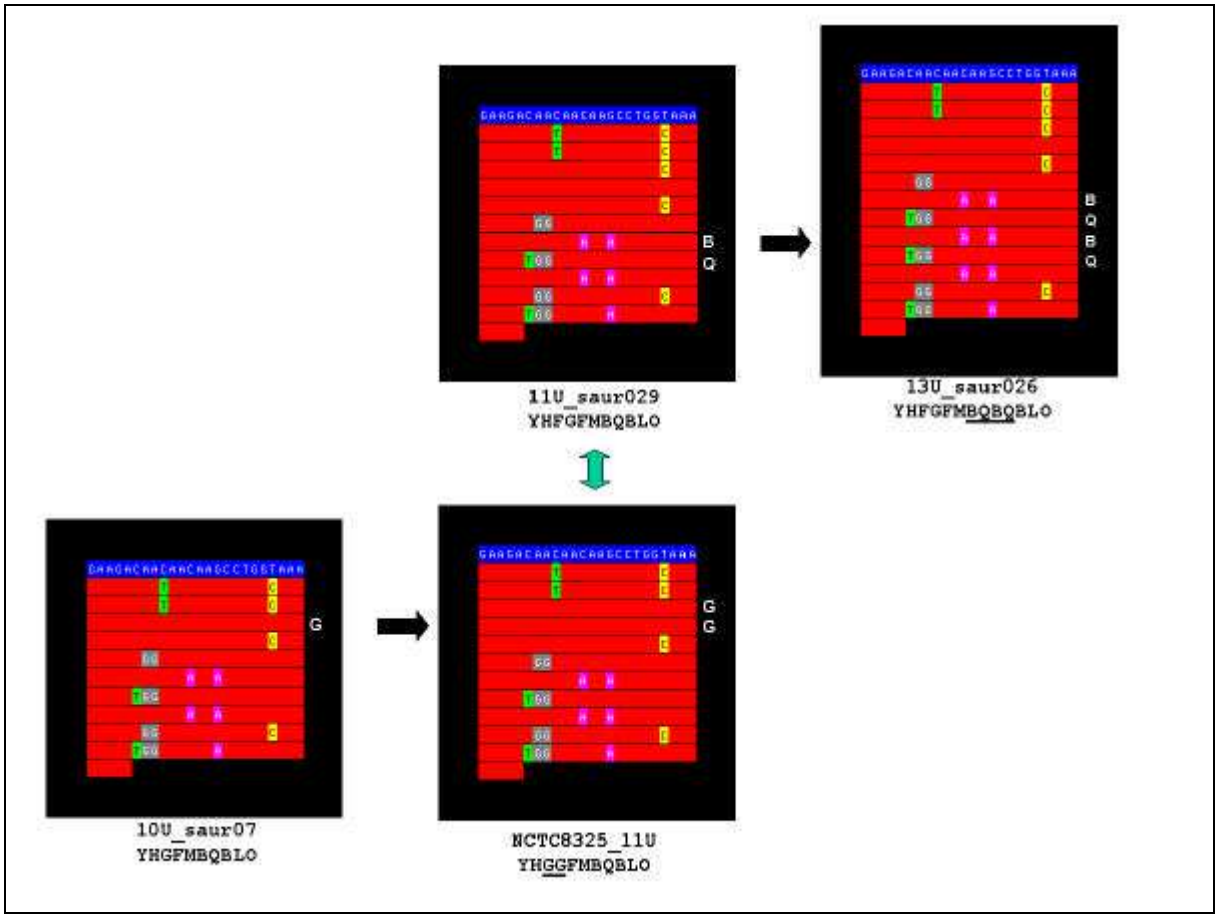


Figure 37 : Visualisation d'allèles *spa* à l'aide du TRView

Cette figure illustre comment passer d'un allèle à 10 motifs à un allèle à 11 motifs par une simple duplication. De même pour passer d'un allèle à 11 motifs à un allèle à 13 motifs.

Il serait souhaitable pour une meilleure compréhension des mécanismes d'évolution des répétitions en tandem chez les bactéries de disposer d'outils d'analyse des séquences des motifs (y compris lorsque les allèles sont de tailles très différentes). Par exemple faire le codage d'allèle en classant les motifs selon leur ressemblance et pas uniquement selon la position dans le minisatellite. C'est à dire que le 1er motif codé « a » puis le 2ème codé « b » soient ensuite codés « a » et « b » si ce sont bien les motifs les plus proches phylogénetiquement pour donner un autre sens au codage.

4.6 Développements futurs

4.6.1 Etudier le lien entre génotype et pathogénicité

Il faudrait développer l'étude des gènes impliqués dans la virulence et qui possèdent une répétition en tandem polymorphe par exemple, pour essayer d'associer un phénotype à un génotype. Ce type d'étude a déjà été effectué en ce qui concerne la variation de phase. Ceci est possible par l'étude du niveau de transcription des gènes impliqués dans la pathogénicité ayant une répétition en tandem dans la région promotrice, ou l'étude de la traduction lorsque la répétition est dans la région codante (et éventuellement étude de l'activité de la protéine).

On peut aussi faire de l'inactivation des gènes possédant une répétition en tandem dans des ORF, pour tenter d'observer un phénotype. Puis de comparer le phénotype de la souche inactivée avec ceux des différentes souches présentant un nombre de répétitions différent. Chez *P. aeruginosa*, l'inactivation de gènes semble tout à fait réalisable (Windgassen 2000).

Il va être de plus en plus nécessaire d'essayer de trouver un lien entre les génotypes observés et la pathogénicité des souches, ceci est très important pour le diagnostic clinique. En effet, des populations bactériennes différentes peuvent avoir un potentiel pathogène différent. Il a été montré récemment le lien immédiat entre un facteur de virulence chez *S. aureus* et la sévérité de la maladie. Des souches de *S. aureus* qui possèdent le gène codant la leukocidine Panton-valentine, prédisposent certains groupes de patients (enfants et jeunes adultes) à une forme fatale de pneumonie hémorragique (Gillet 2002). Ceci constitue une nouvelle forme de diagnostique moléculaire qui associe l'identification de la bactérie pathogène en cause et la détection de gènes de virulence.

Une autre étude très intéressante a consisté à tester la présence de 33 facteurs de virulence dans des souches de *S. aureus* (29 facteurs testés par PCR et 4 par test phénotypique)

(Peacock 2002). La majorité des maladies provoquées par *S. aureus* ne peuvent s'expliquer par l'action d'un seul facteur de virulence (sauf par exemple pour le syndrome de choc toxique provoqué par la toxine TSST1, Toxic Shock Syndrome Toxin 1). D'une manière générale, il semblerait plutôt qu'il s'agisse de l'action combinée de plusieurs facteurs de virulence qui permette d'expliquer le développement et la sévérité de certaines maladies. Le « fonds » génétique du patient infecté est également important pour tenter de relier une forme de maladie à des facteurs de prédisposition. D'une manière générale, le but est d'essayer de relier une forme d'infection donnée à un « profil » de facteurs de virulence. Un groupe de 7 facteurs de virulence (dont certains possédant une répétition en tandem ont été étudiés au cours de cette thèse) sont communs aux souches invasives. Il s'agit de fnbA, cna, sdrE, sej, eta, hlg et ica. Leur effet semble cumulatif. Il semblerait aussi que les variants alléliques d'un locus polymorphe puissent avoir des contributions différentes dans le développement de la maladie. Il serait intéressant de pouvoir relier certains allèles à une forme d'infection. Le transfert horizontal de gènes de virulence se produit chez *S. aureus*, et des mutations dans ces gènes peuvent avoir un effet sur la gravité de l'infection. Une autre étude a été réalisée chez les streptocoques de groupe A, avec la même démarche tentant de relier certains facteurs de virulence aux différentes manifestations de l'infection (Vlaminckx 2003).

Actuellement, de plus en plus d'études de génomique et de protéomique sont réalisées sur des pathogènes. Par exemple pour *P. aeruginosa*, d'autres génomes ont été séquençés et comparés à celui de PAO1. Deux souches persistantes dans des poumons de patients CF (Cystic Fibrosis) et une souche provenant du milieu aquatique ont été séquencées. Le génome de PAO1 constitue le « core » génome caractéristique de *P. aeruginosa*. Les 3 autres souches possèdent 10% de génome en plus dont la moitié sont des séquences nouvelles (Spencer 2003). Par ailleurs, une analyse par puces à ADN à permis d'aborder l'expression globale du génome d'une souche mucoïde, pour essayer de trouver des facteurs de virulence spécifiques du phénotype étudié. En effet, la conversion en phénotype mucoïde n'est pas due à l'expression d'un seul gène, c'est pourquoi il était nécessaire de rechercher d'autres gènes outre ceux impliqués dans la biosynthèse de l'alginate (Firoved 2003).

4.6.2 Etendre les études MLVA à d'autres bactéries pathogènes

Les pathogènes émergents ou ré-émergents représentent une préoccupation permanente pour la santé humaine. Les virus et les bactéries pathogènes sont responsables de la mort de 14 millions de personnes chaque année. Au cours des trente dernières années, quelques bactéries pathogènes ont été découvertes, par exemple *Legionella pneumophila*, *Campylobacter jejuni*, *Borrelia burgdorferi*, *Helicobacter pylori*. L'émergence de nouveaux pathogènes est souvent liée à des changements écologiques. Les facteurs de risque liés à un pathogène sont dus au

type de pathogène, à la voie de contamination et à sa spécificité d'hôte. La plupart des pathogènes infectent plusieurs hôtes, et cette spécificité d'hôte est assez mal connue (Woolhouse 2002). Les méthodes de typage des bactéries pathogènes sont donc indispensables au suivi épidémiologique des souches pathogènes à l'échelle mondiale comme à l'échelle locale, pour tenter de mettre en place des mesures sanitaires adaptées aux différentes situations. Pour cet objectif ambitieux, le typage des répétitions en tandem constitue une des solutions à étendre à d'autres espèces bactéries pathogènes.

5 BIBLIOGRAPHIE

Achtman, M., Azuma, T., Berg, D. E., Ito, Y., Morelli, G., Pan, Z. J., Suerbaum, S., Thompson, S. A., van der Ende, A. et van Doorn, L. J. 1999. Recombination and clonal groupings within Helicobacter pylori from different geographical regions. *Mol Microbiol* **32:** 459-70.

Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A. et Carniel, E. 1999. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci U S A* **96:** 14043-8.

Adair, D. M., Worsham, P. L., Hill, K. K., Klevytska, A. M., Jackson, P. J., Friedlander, A. M. et Keim, P. 2000. Diversity in a variable-number tandem repeat from *Yersinia pestis*. *J Clin Microbiol* **38:** 1516-9.

Ajayi, T., Allmond, L. R., Sawa, T. et Wiener-Kronish, J. P. 2003. Single-nucleotide-polymorphism mapping of the Pseudomonas aeruginosa type III secretion toxins for development of a diagnostic multiplex PCR system. *J Clin Microbiol* **41:** 3526-31.

Allred, D. R., McGuire, T. C., Palmer, G. H., Leib, S. R., Harkins, T. M., McElwain, T. F. et Barbet, A. F. 1990. Molecular basis for surface antigen size polymorphisms and conservation of a neutralization-sensitive epitope in Anaplasma marginale. *Proc Natl Acad Sci U S A* **87:** 3220-4.

Alm, R. A. *et al.* 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori. *Nature* **397:** 176-80.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. et Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389-402.

Astagneau, P. 1998. Epidemiology of nosocomial infections. *Rev Prat* **48:** 1525-9.

Baba, T. *et al.* 2002. Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet* **359:** 1819-27.

Bachellier, S., Saurin, W., Perrin, D., Hofnung, M. et Gilson, E. 1994. Structural and functional diversity among bacterial interspersed mosaic elements (BIMEs). *Mol Microbiol* **12:** 61-70.

Barber, M. et Rozwadnowska-Dowzenko, M. 1948. Infection by penicillin-resistant staphylococci. *Lancet* **ii**: 641-644.

Bayliss, C. D., Field, D. et Moxon, E. R. 2001. The simple sequence contingency loci of *Haemophilus influenzae* and *Neisseria meningitidis*. *J Clin Invest* **107:** 657-666.

Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane, S. et Small, P. M. 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284:** 1520-3.

Bennekov, T., Colding, H., Ojeniyi, B., Bentzon, M. W. et Hoiby, N. 1996. Comparison of ribotyping and genome fingerprinting of *Pseudomonas aeruginosa* isolates from cystic fibrosis patients. *J Clin Microbiol* **34:** 202-4.

Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27:** 573-580.

Bernal, A., Ear, U. et Kyrpides, N. 2001. Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. *Nucleic Acids Res* **29:** 126-7.

Bessen, D., Jones, K. F. et Fischetti, V. A. 1989. Evidence for two distinct classes of streptococcal M protein and their relationship to rheumatic fever. *J Exp Med* **169:** 269-83.

Bingen, E. H., Denamur, E. et Elion, J. 1994. Use of ribotyping in epidemiological surveillance of nosocomial outbreaks. *Clin Microbiol Rev* **7:** 311-27.

Branger, C. *et al.* 1994. Epidemiology of Staphylococcus aureus in patients with cystic fibrosis. *Epidemiol Infect* **112:** 489-500.

Bricker, B. J., Ewalt, D. R. et Halling, S. M. 2003. Brucella 'Hoof-Prints': strain typing by multi-locus analysis of variable number tandem repeats (VNTRs). *BMC Microbiol* **3:** 15.

Brisse, S., Fussing, V., Ridwan, B., Verhoef, J. et Willems, R. J. 2002. Automated ribotyping of vancomycin-resistant Enterococcus faecium isolates. *J Clin Microbiol* **40:** 1977-84.

Brisse, S., Milatovic, D., Fluit, A. C., Kusters, K., Toelstra, A., Verhoef, J. et Schmitz, F. J. 2000. Molecular surveillance of European quinolone-resistant clinical isolates of *Pseudomonas aeruginosa* and *Acinetobacter* spp. using automated ribotyping. *J Clin Microbiol* **38:** 3636-45.

Britten, R. J. & Kohne, D. E. 1968. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* **161:** 529-40.

Brookes, A. J. 1999. The essence of SNPs. *Gene* **234:** 177-86.

Bruun, F. N., McGarrity, G. J., Blakemore, W. S. et Coriell, L. L. 1976. Epidemiology of Pseudomonas aeruginosa infections: determination by pyocin typing. *J Clin Microbiol* **3:** 264-71.

Buard, J. & Vergnaud, G. 1994. Complex recombination events at the hypermutable minisatellite CEB1 (D2S90). *EMBO J.* **13:** 3203-3210.

Burch, C. L., Danaher, R. J. et Stein, D. C. 1997. Antigenic variation in Neisseria gonorrhoeae: production of multiple lipooligosaccharides. *J Bacteriol* **179:** 982-6.

Chan, M. S., Maiden, M. C. et Spratt, B. G. 2001. Database-driven multi locus sequence typing (MLST) of bacterial pathogens. *Bioinformatics* **17:** 1077-83.

Charlesworth, B., Sniegowski, P. et Stephan, W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371:** 215-220.

Citti, C. & Rosengarten, R. 1997. Mycoplasma genetic variation and its implication for pathogenesis. *Wien Klin Wochenschr* **109:** 562-8.

Cole, S. T., Supply, P. et Honore, N. 2001. Repetitive sequences in Mycobacterium leprae and their impact on genome plasticity. *Lepr Rev* **72:** 449-61.

Costerton, J. W., Stewart, P. S. et Greenberg, E. P. 1999. Bacterial biofilms: a common cause of persistent infections. *Science* **284:** 1318-22.

Cramton, S. E., Schnell, N. F., Gotz, F. et Bruckner, R. 2000. Identification of a new repetitive element in Staphylococcus aureus. *Infect Immun* **68:** 2344-8.

Crisostomo, M. I., Westh, H., Tomasz, A., Chung, M., Oliveira, D. C. et de Lencastre, H. 2001. The evolution of methicillin resistance in Staphylococcus aureus: similarity of genetic backgrounds in historically early methicillin-susceptible and -resistant isolates and contemporary epidemic clones. *Proc Natl Acad Sci U S A* **98:** 9865-70.

Dabrowski, W., Czekajlo-Kolodziej, U., Medrala, D. et Giedrys-Kalemba, S. 2003. Optimisation of AP-PCR fingerprinting discriminatory power for clinical isolates of Pseudomonas aeruginosa. *FEMS Microbiol Lett* **218:** 51-7.

Daum, R. S., Ito, T., Hiramatsu, K., Hussain, F., Mongkolrattanothai, K., Jamklang, M. et Boyle-Vavra, S. 2002. A novel methicillin-resistance cassette in community-acquired methicillin-resistant Staphylococcus aureus isolates of diverse genetic backgrounds. *J Infect Dis* **186:** 1344-7.

De Bolle, X., Bayliss, C. D., Field, D., van de Ven, T., Saunders, N. J., Hood, D. W. et Moxon, E. R. 2000. The length of a tetranucleotide repeat tract in *Haemophilus influenzae* determines the phase variation rate of a gene with homology to type III DNA methyltransferases. *Mol Microbiol* **35:** 211-22.

de Gialluly, C., Loulergue, J., Bruant, G., Mereghetti, L., Massuard, S., van der Mee, N., Audurier, A. et Quentin, R. 2003. Identification of new phages to type Staphylococcus aureus strains and comparison with a genotypic method. *J Hosp Infect* **55:** 61-7.

Denoeud, F. & Vergnaud, G. 2004. Identification of polymorphic tandem repeats by direct comparison of genome sequence from different bacterial strains : a Web-based ressource. *BMC Bioinformatics* **5:** 4.

Denoeud, F., Vergnaud, G. et Benson, G. 2003. Predicting Human Minisatellite Polymorphism. *Genome Res* **13:** 856-867.

Deretic, V., Hibler, N. S. et Holt, S. C. 1992. Immunocytochemical analysis of AlgP (Hp1), a histonelike element participating in control of mucoidy in Pseudomonas aeruginosa. *J Bacteriol* **174:** 824-31.

Deretic, V. & Konyecsni, W. M. 1990. A procaryotic regulatory factor with a histone H1-like carboxy-terminal domain: clonal variation of repeats within *algP*, a gene involved in regulation of mucoidy in *Pseudomonas aeruginosa*. *J Bacteriol* **172:** 5544-54.

Domann, E., Zechel, S., Lingnau, A., Hain, T., Darji, A., Nichterlein, T., Wehland, J. et Chakraborty, T. 1997. Identification and characterization of a novel PrfA-regulated gene in Listeria monocytogenes whose product, IrpA, is highly homologous to internalin proteins, which contain leucine-rich repeats. *Infect Immun* **65:** 101-9.

Dore, N., Bennett, D., Kaliszer, M., Cafferkey, M. et Smyth, C. J. 2003. Molecular epidemiology of group B streptococci in Ireland: associations between serotype, invasive

status and presence of genes encoding putative virulence factors. *Epidemiol Infect* **131:** 823-33.

Drenkard, E. & Ausubel, F. M. 2002. Pseudomonas biofilm formation and antibiotic resistance are linked to phenotypic variation. *Nature* **416:** 740-3.

Embley, T. M. 1991. The linear PCR reaction: a simple and robust method for sequencing amplified rRNA genes. *Lett Appl Microbiol* **13:** 171-4.

Enright, M. C. 2003. The evolution of a resistant pathogen--the case of MRSA. *Curr Opin Pharmacol* **3:** 474-9.

Enright, M. C., Robinson, D. A., Randle, G., Feil, E. J., Grundmann, H. et Spratt, B. G. 2002. The evolutionary history of methicillin-resistant Staphylococcus aureus (MRSA). *Proc Natl Acad Sci U S A* **99:** 7687-92.

Farlow, J., Postic, D., Smith, K. L., Jay, Z., Baranton, G. et Keim, P. 2002. Strain typing of Borrelia burgdorferi, Borrelia afzelii, and Borrelia garinii by using multiple-locus variable-number tandem repeat analysis. *J Clin Microbiol* **40:** 4612-8.

Farlow, J., Smith, K. L., Wong, J., Abrams, M., Lytle, M. et Keim, P. 2001. Francisella tularensis strain typing using multiple-locus, variable-number tandem repeat analysis. *J Clin Microbiol* **39:** 3186-92.

Feil, E. J., J. E. Cooper, H. Grundmann, D. A. Robinson, M. C. Enright, T. Berendt, S. J. Peacock, J. M. Smith, M. Murphy, B. G. Spratt, C. E. Moore and N. P. Day. 2003. How clonal is Staphylococcus aureus? *J Bacteriol* **185**: 3307-16.

Feil, E. J., Enright, M. C. et Spratt, B. G. 2000. Estimating the relative contributions of mutation and recombination to clonal diversification: a comparison between Neisseria meningitidis and Streptococcus pneumoniae. *Res Microbiol* **151:** 465-9.

Feil, E. J., Maiden, M. C., Achtman, M. et Spratt, B. G. 1999. The relative contributions of recombination and mutation to the divergence of clones of Neisseria meningitidis. *Mol Biol Evol* **16:** 1496-502.

Filliol, I. *et al.* 2002. Global distribution of Mycobacterium tuberculosis spoligotypes. *Emerg Infect Dis* **8:** 1347-9.

Filliol, I., Ferdinand, S., Negroni, L., Sola, C. et Rastogi, N. 2000. Molecular typing of Mycobacterium tuberculosis based on variable number of tandem DNA repeats used alone and in association with spoligotyping. *J Clin Microbiol* **38:** 2520-4.

Firoved, A. M. & Deretic, V. 2003. Microarray analysis of global gene expression in mucoid Pseudomonas aeruginosa. *J Bacteriol* **185:** 1071-81.

Fitzgerald, J. R., Sturdevant, D. E., Mackie, S. M., Gill, S. R. et Musser, J. M. 2001. Evolutionary genomics of Staphylococcus aureus: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proc Natl Acad Sci U S A* **98:** 8821-6.

Fleischmann, R. D. *et al.* 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* **269:** 496-512.

Fleischmann, R. D. *et al.* 2002. Whole-Genome Comparison of *Mycobacterium tuberculosis* Clinical and Laboratory Strains. *J Bacteriol* **184:** 5479-5490.

Fletcher, H. A., Donoghue, H. D., Taylor, G. M., van der Zanden, A. G. et Spigelman, M. 2003. Molecular analysis of Mycobacterium tuberculosis DNA from a family of 18th century Hungarians. *Microbiology* **149:** 143-51.

Foissaud, V., Puyhardy, J. M., Chapalain, J. C., Salord, H., Depina, J. J., Morillon, M., Nicolas, P. et Perrier-Gros-Claude, J. D. 1999. [Inter-laboratory reproducibility of pulsed-field electrophoresis for the study of 12 types of Pseudomonas aeruginosa]. *Pathol Biol (Paris)* **47:** 1053-9.

Foster, P. L. & Trimarchi, J. M. 1994. Adaptive reversion of a frameshift mutation in *Escherichia coli* by simple base deletions in homopolymeric runs. *Science* **265:** 407-409.

Foster, T. J. & McDevitt, D. 1994. Surface-associated proteins of Staphylococcus aureus: their possible roles in virulence. *FEMS Microbiol Lett* **118:** 199-205.

Francois, P., Pittet, D., Bento, M., Pepey, B., Vaudaux, P., Lew, D. et Schrenzel, J. 2003. Rapid detection of methicillin-resistant Staphylococcus aureus directly from sterile or nonsterile clinical samples by a new molecular assay. *J Clin Microbiol* **41:** 254-60.

Frangeul, L., Nelson, K. E., Buchrieser, C., Danchin, A., Glaser, P. et Kunst, F. 1999. Cloning and assembly strategies in microbial genome projects. *Microbiology* **145 (Pt 10):** 2625-34.

Fraser, C. M., Eisen, J. A., Nelson, K. E., Paulsen, I. T. et Salzberg, S. L. 2002. The value of complete microbial genome sequencing (you get what you pay for). *J Bacteriol* **184:** 6403-5; discusion 6405.

Fraser, C. M., Eisen, J. A. et Salzberg, S. L. 2000. Microbial genome sequencing. *Nature* **406:** 799-803.

Frenay, H. M., Bunschoten, A. E., Schouls, L. M., van Leeuwen, W. J., Vandenbroucke-Grauls, C. M., Verhoef, J. et Mooi, F. R. 1996. Molecular typing of methicillin-resistant Staphylococcus aureus on the basis of protein A gene polymorphism. *Eur J Clin Microbiol Infect Dis* **15:** 60-4.

Frothingham, R. & Meeker-O'Connell, W. A. 1998. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology* **144:** 1189-1196.

Gillet, Y. *et al.* 2002. Association between Staphylococcus aureus strains carrying gene for Panton-Valentine leukocidin and highly lethal necrotising pneumonia in young immunocompetent patients. *Lancet* **359:** 753-9.

Gilson, E., Clement, J. M., Brutlag, D. et Hofnung, M. 1984. A family of dispersed repetitive extragenic palindromic DNA sequences in E. coli. *Embo J* **3:** 1417-21.

Gilson, E., Perrin, D., Saurin, W. et Hofnung, M. 1987. Species specificity of bacterial palindromic units. *J Mol Evol* **25:** 371-3.

Gilson, E., Saurin, W., Perrin, D., Bachellier, S. et Hofnung, M. 1991. Palindromic units are part of a new bacterial interspersed mosaic element (BIME). *Nucleic Acids Res* **19:** 1375-83.

Glaser, P. *et al.* 2001. Comparative genomics of Listeria species. *Science* **294:** 849-52.

Glew, M. D., Baseggio, N., Markham, P. F., Browning, G. F. et Walker, I. D. 1998. Expression of the pMGA genes of Mycoplasma gallisepticum is controlled by variation in the GAA trinucleotide repeat lengths within the 5' noncoding regions. *Infect Immun* **66:** 5833-41.

Goering, R. V. & Winters, M. A. 1992. Rapid method for epidemiological evaluation of gram-positive cocci by field inversion gel electrophoresis. *J Clin Microbiol* **30:** 577-80.

Goguet de la Salmoniere, Y. O., Li, H. M., Torrea, G., Bunschoten, A., van Embden, J. et Gicquel, B. 1997. Evaluation of spoligotyping in a study of the transmission of Mycobacterium tuberculosis. *J Clin Microbiol* **35:** 2210-4.

Goh, S. H., Byrne, S. K., Zhang, J. L. et Chow, A. W. 1992. Molecular typing of Staphylococcus aureus on the basis of coagulase gene polymorphisms. *J Clin Microbiol* **30:** 1642-5.

Govan, J. R. & Deretic, V. 1996. Microbial pathogenesis in cystic fibrosis: mucoid Pseudomonas aeruginosa and Burkholderia cepacia. *Microbiol Rev* **60:** 539-74.

Gravekamp, C., Kasper, D. L. et Madoff, L. C. 1997. Immunization with a single-repeat alpha C protein may prevent escape of lower repeat mutants of group B Streptococcus. *Adv Exp Med Biol* **418:** 855-7.

Gravekamp, C., Rosner, B. et Madoff, L. C. 1998. Deletion of repeats in the alpha C protein enhances the pathogenicity of group B streptococci in immune mice. *Infect Immun* **66:** 4347-54.

Grisold, A. J., Leitner, E., Muhlbauer, G., Marth, E. et Kessler, H. H. 2002. Detection of methicillin-resistant Staphylococcus aureus and simultaneous confirmation by automated nucleic acid extraction and real-time PCR. *J Clin Microbiol* **40:** 2392-7.

Groenen, P. M., Bunschoten, A. E., van Soolingen, D. et van Embden, J. D. 1993. Nature of DNA polymorphism in the direct repeat cluster of Mycobacterium tuberculosis; application for strain differentiation by a novel typing method. *Mol Microbiol* **10:** 1057-65.

Grundmann, H., Schneider, C., Hartung, D., Daschner, F. D. et Pitt, T. L. 1995. Discriminatory power of three DNA-based typing techniques for *Pseudomonas aeruginosa*. *J Clin Microbiol* **33:** 528-34.

Guerin, M., Robichon, N., Geiselmann, J. et Rahmouni, A. R. 1998. A simple polypyrimidine repeat acts as an artificial Rho-dependent terminator in vivo and in vitro. *Nucleic Acids Res* **26:** 4895-900.

Heath, D. G., An, F. Y., Weaver, K. E. et Clewell, D. B. 1995. Phase variation of Enterococcus faecalis pAD1 conjugation functions relates to changes in iteron sequence region. *J Bacteriol* **177:** 5453-9.

Hebert, G. A., Cooksey, R. C., Clark, N. C., Hill, B. C., Jarvis, W. R. et Thornsberry, C. 1988. Biotyping coagulase-negative staphylococci. *J Clin Microbiol* **26:** 1950-6.

Henderson, I. R., Owen, P. et Nataro, J. P. 1999. Molecular switches--the ON and OFF of bacterial phase variation. *Mol Microbiol* **33:** 919-32.

Hermans, P. W., Sluijter, M., Hoogenboezem, T., Heersma, H., van Belkum, A. et de Groot, R. 1995. Comparative study of five different DNA fingerprint techniques for molecular typing of Streptococcus pneumoniae strains. *J Clin Microbiol* **33:** 1606-12.

Hermans, P. W., van Soolingen, D., Bik, E. M., de Haas, P. E., Dale, J. W. et van Embden, J. D. 1991. Insertion element IS987 from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. *Infect Immun* **59:** 2695-2705.

Higgins, C. F., Ames, G. F., Barnes, W. M., Clement, J. M. et Hofnung, M. 1982. A novel intercistronic regulatory element of prokaryotic operons. *Nature* **298:** 760-2.

Higgins, C. F., McLaren, R. S. et Newbury, S. F. 1988. Repetitive extragenic palindromic sequences, mRNA stability and gene expression: evolution by gene conversion? A review. *Gene* **72:** 3-14.

Hiramatsu, K., Cui, L., Kuroda, M. et Ito, T. 2001. The emergence and evolution of methicillin-resistant Staphylococcus aureus. *Trends Microbiol* **9:** 486-93.

Hiramatsu, K., Hanaki, H., Ino, T., Yabuta, K., Oguri, T. et Tenover, F. C. 1997. Methicillin-resistant Staphylococcus aureus clinical strain with reduced vancomycin susceptibility. *J Antimicrob Chemother* **40:** 135-6.

Hiramatsu, K., Katayama, Y., Yuzawa, H. et Ito, T. 2002. Molecular genetics of methicillin-resistant Staphylococcus aureus. *Int J Med Microbiol* **292:** 67-74.

Holliday, M. G., Ford, M., Perry, J. D. et Gould, F. K. 1999. Rapid identification of Staphylococcus aureus by using fluorescent staphylocoagulase assays. *J Clin Microbiol* **37:** 1190-2.

Hollingshead, S. K., Fischetti, V. A. et Scott, J. R. 1987. Size variation in group A streptococcal M protein is generated by homologous recombination between intragenic repeats. *Mol Gen Genet* **207:** 196-203.

Holloway, B. W., Escuadra, M. D., Morgan, A. F., Saffery, R. et Krishnapillai, V. 1992. The new approaches to whole genome analysis of bacteria. *FEMS Microbiol Lett* **79:** 101-5.

Hood, D. W., Deadman, M. E., Jennings, M. P., Bisercic, M., Fleischmann, R. D., Venter, J. C. et Moxon, E. R. 1996. DNA repeats identify novel virulence genes in Haemophilus influenzae. *Proc Natl Acad Sci U S A* **93:** 11121-5.

Hookey, J. V., Richardson, J. F. et Cookson, B. D. 1998. Molecular typing of Staphylococcus aureus based on PCR restriction fragment length polymorphism and DNA sequence analysis of the coagulase gene. *J Clin Microbiol* **36:** 1083-9.

Hulton, C. S., Higgins, C. F. et Sharp, P. M. 1991. ERIC sequences: a novel family of repetitive elements in the genomes of Escherichia coli, Salmonella typhimurium and other enterobacteria. *Mol Microbiol* **5:** 825-34.

Jackson, P. J. *et al.* 1997. Characterization of the variable-number tandem repeats in vrrA from different Bacillus anthracis isolates. *Appl Environ Microbiol* **63:** 1400-5.

Jansen, R., Embden, J. D., Gaastra, W. et Schouls, L. M. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43:** 1565-75.

Jeffreys, A. J., Tamaki, K., MacLeod, A., Monckton, D. G., Neil, D. L. et Armour, J. A. L. 1994. Complex gene conversion events in germline mutation at human minisatellites. *Nat. Genet.* **6:** 136-145.

Jeffreys, A. J., Wilson, V. et Thein, S. L. 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature* **314:** 67-73.

Jeffreys, A. J., Wilson, V. et Thein, S. L. 1985. Individual-specific 'fingerprints' of human DNA. *Nature* **316:** 76-79.

Jernigan, D. B. *et al.* 2002. Investigation of bioterrorism-related anthrax, United States, 2001: epidemiologic findings. *Emerg Infect Dis* **8:** 1019-28.

Jevons, M. P. 1961. Celbenin-resistant staphylococci. *Br. Med. J.* **i:** 124-125.

Johnson, A. P. & Woodford, N. 2002. Glycopeptide-resistant Staphylococcus aureus. *J Antimicrob Chemother* **50:** 621-3.

Jordon, P., Snyder, L. A. et Saunders, N. J. 2003. Diversity in coding tandem repeats in related Neisseria spp. *BMC Microbiol* **3:** 23.

Josefsson, E., McCrea, K. W., Ni Eidhin, D., O'Connell, D., Cox, J., Hook, M. et Foster, T. J. 1998. Three new members of the serine-aspartate repeat protein multigene family of Staphylococcus aureus. *Microbiology* **144 (Pt 12):** 3387-95.

Kaida, S., Miyata, T., Yoshizawa, Y., Kawabata, S., Morita, T., Igarashi, H. et Iwanaga, S. 1987. Nucleotide sequence of the staphylocoagulase gene: its unique COOH-terminal 8 tandem repeats. *J Biochem (Tokyo)* **102:** 1177-86.

Kamerbeek, J. *et al.* 1997. Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. *J Clin Microbiol* **35:** 907-14.

Katayama, Y., Ito, T. et Hiramatsu, K. 2000. A new class of genetic element, staphylococcus cassette chromosome mec, encodes methicillin resistance in Staphylococcus aureus. *Antimicrob Agents Chemother* **44:** 1549-55.

Keim, P., Price, L. B., Klevytska, A. M., Smith, K. L., Schupp, J. M., Okinaka, R., Jackson, P. J. et Hugh-Jones, M. E. 2000. Multiple-Locus Variable-Number Tandem Repeat Analysis Reveals Genetic Relationships within *Bacillus anthracis*. *J Bacteriol* **182:** 2928-2936.

Khalifa, K. I., Heiba, A. A. et Hancock, G. 1989. Nontypeable bacteriophage patterns of methicillin-resistant Staphylococcus aureus involved in a hospital outbreak. *J Clin Microbiol* **27:** 2249-51.

Kim, W., Hong, Y. P., Yoo, J. H., Lee, W. B., Choi, C. S. et Chung, S. I. 2002. Genetic relationships of Bacillus anthracis and closely related species based on variable-number tandem repeat analysis and BOX-PCR genomic fingerprinting. *FEMS Microbiol Lett* **207:** 21-7.

Klevytska, A. M., Price, L. B., Schupp, J. M., Worsham, P. L., Wong, J. et Keim, P. 2001. Identification and characterization of variable-number tandem repeats in the *Yersinia pestis* genome. *J Clin Microbiol* **39:** 3179-85.

Koeuth, T., Versalovic, J. et Lupski, J. R. 1995. Differential subsequence conservation of interspersed repetitive Streptococcus pneumoniae BOX elements in diverse bacteria. *Genome Res* **5:** 408-18.

Krasilnikova, M. M., Samadashwily, G. M., Krasilnikov, A. S. et Mirkin, S. M. 1998. Transcription through a simple DNA repeat blocks replication elongation. *Embo J* **17:** 5095-102.

Kreiswirth, B., Kornblum, J., Arbeit, R. D., Eisner, W., Maslow, J. N., McGeer, A., Low, D. E. et Novick, R. P. 1993. Evidence for a clonal origin of methicillin resistance in Staphylococcus aureus. *Science* **259:** 227-30.

Kremer, K. *et al.* 1999. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. *J Clin Microbiol* **37:** 2607-2618.

Kresse, A. U., Dinesh, S. D., Larbig, K. et Romling, U. 2003. Impact of large chromosomal inversions on the adaptation and evolution of Pseudomonas aeruginosa chronically colonizing cystic fibrosis lungs. *Mol Microbiol* **47:** 145-58.

Kruglyak, S., Durrett, R. T., Schug, M. D. et Aquadro, C. F. 1998. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc Natl Acad Sci U S A* **95:** 10774-8.

Kuroda, M. *et al.* 2001. Whole genome sequencing of meticillin-resistant Staphylococcus aureus. *Lancet* **357:** 1225-40.

Lander, E. S. & Green, P. 1987. Construction of multi-locus genetic linkage maps in humans. *Proc. Natl. Acad. Sci. USA* **84:** 2363-2367.

Le Flèche, P., Fabre, M., Denoeud, F., Koeck, J. L. et Vergnaud, G. 2002. High resolution, on-line identification of strains from the *Mycobacterium tuberculosis* complex based on tandem repeat typing. *BMC Microbiol* **2:** 37.

Le Flèche, P., Hauck, Y., Onteniente, L., Prieur, A., Denoeud, F., Ramisse, V., Sylvestre, P., Benson, G., Ramisse, F. et Vergnaud, G. 2001. A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*. *BMC Microbiol* **1:** 2.

Lenski, R. E., Winkworth, C. L. et Riley, M. A. 2003. Rates of DNA sequence evolution in experimental populations of Escherichia coli during 20,000 generations. *J Mol Evol* **56:** 498-508.

Levinson, G. & Gutman, G. A. 1987. High frequency of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res.* **15:** 5323-5338.

Linares, J. 2001. The VISA/GISA problem: therapeutic implications. *Clin Microbiol Infect* **7 Suppl 4:** 8-15.

Lindstedt, B. A., Heir, E., Gjernes, E. et Kapperud, G. 2003. DNA fingerprinting of Salmonella enterica subsp. enterica serovar typhimurium with emphasis on phage type DT104 based on variable number of tandem repeat loci. *J Clin Microbiol* **41:** 1469-79.

Linhardt, F., Ziebuhr, W., Meyer, P., Witte, W. et Hacker, J. 1992. Pulsed-field gel electrophoresis of genomic restriction fragments as a tool for the epidemiological analysis of Staphylococcus aureus and coagulase-negative staphylococci. *FEMS Microbiol Lett* **74:** 181-5.

Liu, Y., Lee, M. A., Ooi, E. E., Mavis, Y., Tan, A. L. et Quek, H. H. 2003. Molecular typing of Salmonella enterica serovar typhi isolates from various countries in Asia by a multiplex PCR assay on variable-number tandem repeats. *J Clin Microbiol* **41:** 4388-94.

Livermore, D. M. 2002. Multiple mechanisms of antimicrobial resistance in *Pseudomonas aeruginosa*: our worst nightmare? *Clin Infect Dis* **34:** 634-40.

Lysnyansky, I., Rosengarten, R. et Yogev, D. 1996. Phenotypic switching of variable surface lipoproteins in Mycoplasma bovis involves high-frequency chromosomal rearrangements. *J Bacteriol* **178:** 5395-401.

Ma, X. X., Ito, T., Tiensasitorn, C., Jamklang, M., Chongtrakool, P., Boyle-Vavra, S., Daum, R. S. et Hiramatsu, K. 2002. Novel type of staphylococcal cassette chromosome mec identified in community-acquired methicillin-resistant Staphylococcus aureus strains. *Antimicrob Agents Chemother* **46:** 1147-52.

Madoff, L. C., Michel, J. L., Gong, E. W., Kling, D. E. et Kasper, D. L. 1996. Group B streptococci escape host immunity by deletion of tandem repeat elements of the alpha C protein. *Proc Natl Acad Sci U S A* **93:** 4131-6.

Mahillon, J., Leonard, C. et Chandler, M. 1999. IS elements as constituents of bacterial genomes. *Res Microbiol* **150:** 675-87.

Maiden, M. C. *et al.* 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:** 3140-5.

Martin, B. *et al.* 1992. A highly conserved repeated DNA element located in the chromosome of Streptococcus pneumoniae. *Nucleic Acids Res* **20:** 3479-83.

Martin, P., van de Ven, T., Mouchel, N., Jeffries, A. C., Hood, D. W. et Moxon, E. R. 2003. Experimentally revised repertoire of putative contingency loci in Neisseria meningitidis strain MC58: evidence for a novel mechanism of phase variation. *Mol Microbiol* **50:** 245-57.

Masepohl, B., Gorlitz, K. et Bohme, H. 1996. Long tandemly repeated repetitive (LTRR) sequences in the filamentous cyanobacterium Anabaena sp. PCC 7120. *Biochim Biophys Acta* **1307:** 26-30.

Mazars, E., Lesjean, S., Banuls, A. L., Gilbert, M., Vincent, V. V., Gicquel, B., Tibayrenc, M., Locht, C. et Supply, P. 2001. High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proc Natl Acad Sci U S A* **98:** 1901-1906.

McDevitt, D. & Foster, T. J. 1995. Variation in the size of the repeat region of the fibrinogen receptor (clumping factor) of Staphylococcus aureus strains. *Microbiology* **141 (Pt 4):** 937-43.

McDevitt, D., Francois, P., Vaudaux, P. et Foster, T. J. 1994. Molecular characterization of the clumping factor (fibrinogen receptor) of Staphylococcus aureus. *Mol Microbiol* **11:** 237-48.

Meunier, J. R. & Grimont, P. A. 1993. Factors affecting reproducibility of random amplified polymorphic DNA fingerprinting. *Res Microbiol* **144:** 373-9.

Meyer, T. F., Gibbs, C. P. et Haas, R. 1990. Variation and control of protein expression in Neisseria. *Annu Rev Microbiol* **44:** 451-77.

Mifsud, A. J., Watine, J., Picard, B., Charet, J. C., Solignac-Bourrel, C. et Pitt, T. L. 1997. Epidemiologically related and unrelated strains of *Pseudomonas aeruginosa* serotype O12 cannot be distinguished by phenotypic and genotypic typing. *J Hosp Infect* **36:** 105-16.

Mojica, F. J., Diez-Villasenor, C., Soria, E. et Juez, G. 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol* **36:** 244-6.

Moore, H., Greenwell, P. W., Liu, C. P., Arnheim, N. et Petes, T. D. 1999. Triplet repeats form secondary structures that escape DNA repair in yeast. *Proc Natl Acad Sci U S A* **96:** 1504-9.

Moreno, E., Cloeckaert, A. et Moriyon, I. 2002. *Brucella* evolution and taxonomy. *Vet Microbiol* **90:** 209-27.

Motin, V. L. *et al.* 2002. Genetic variability of Yersinia pestis isolates as predicted by PCR-based IS100 genotyping and analysis of structural genes encoding glycerol-3-phosphate dehydrogenase (glpD). *J Bacteriol* **184:** 1019-27.

Moxon, E. R., Rainey, P. B., Nowak, M. A. et Lenski, R. E. 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Biol* **4:** 24-33.

Murakami, K., Minamide, W., Wada, K., Nakamura, E., Teraoka, H. et Watanabe, S. 1991. Identification of methicillin-resistant strains of staphylococci by polymerase chain reaction. *J Clin Microbiol* **29:** 2240-4.

Musser, J. M. & Kapur, V. 1992. Clonal analysis of methicillin-resistant Staphylococcus aureus strains from intercontinental sources: association of the mec gene with divergent phylogenetic lineages implies dissemination by horizontal transfer and recombination. *J Clin Microbiol* **30:** 2058-63.

Nahvi, M. D., Fitzgibbon, J. E., John, J. F. et Dubin, D. T. 2001. Sequence analysis of dru regions from methicillin-resistant Staphylococcus aureus and coagulase-negative staphylococcal isolates. *Microb Drug Resist* **7:** 1-12.

Nakata, A., Amemura, M. et Makino, K. 1989. Unusual nucleotide arrangement with repeated sequences in the Escherichia coli K-12 chromosome. *J Bacteriol* **171:** 3553-6.

Nelson, K. E. *et al.* 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima. *Nature* **399:** 323-9.

Nelson, K. E., Paulsen, I. T., Heidelberg, J. F. et Fraser, C. M. 2000. Status of genome projects for nonpathogenic bacteria and archaea. *Nat Biotechnol* **18:** 1049-54.

Ochman, H. & Jones, I. B. 2000. Evolutionary dynamics of full genome content in Escherichia coli. *Embo J* **19:** 6637-43.

Okuma, K. *et al.* 2002. Dissemination of new methicillin-resistant Staphylococcus aureus clones in the community. *J Clin Microbiol* **40:** 4289-94.

Olive, D. M. & Bean, P. 1999. Principles and applications of methods for DNA-based typing of microbial organisms. *J Clin Microbiol* **37:** 1661-9.

Oliver, A., Canton, R., Campo, P., Baquero, F. et Blazquez, J. 2000. High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection. *Science* **288:** 1251-4.

Onteniente, L., Brisse, S., Tassios, P. T. et Vergnaud, G. 2003. Evaluation of the polymorphisms associated with tandem repeats for Pseudomonas aeruginosa strain typing. *J Clin Microbiol* **41:** 4991-7.

Orskov, F. & Orskov, I. 1983. From the national institutes of health. Summary of a workshop on the clone concept in the epidemiology, taxonomy, and evolution of the enterobacteriaceae and other bacteria. *J Infect Dis* **148:** 346-57.

Parkhill, J. *et al.* 2003. Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. *Nat Genet* **35:** 32-40.

Parkhill, J. *et al.* 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413:** 523-7.

Patti, J. M., Allen, B. L., McGavin, M. J. et Hook, M. 1994. MSCRAMM-mediated adherence of microorganisms to host tissues. *Annu Rev Microbiol* **48:** 585-617.

Patti, J. M., Bremell, T., Krajewska-Pietrasik, D., Abdelnour, A., Tarkowski, A., Ryden, C. et Hook, M. 1994. The Staphylococcus aureus collagen adhesin is a virulence determinant in experimental septic arthritis. *Infect Immun* **62:** 152-61.

Patti, J. M., Jonsson, H., Guss, B., Switalski, L. M., Wiberg, K., Lindberg, M. et Hook, M. 1992. Molecular characterization and expression of a gene encoding a Staphylococcus aureus collagen adhesin. *J Biol Chem* **267:** 4766-72.

Peacock, S. J., Moore, C. E., Justice, A., Kantzanou, M., Story, L., Mackie, K., O'Neill, G. et Day, N. P. 2002. Virulent combinations of adhesin and toxin genes in natural populations of Staphylococcus aureus. *Infect Immun* **70:** 4987-96.

Perry, R. D. & Fetherston, J. D. 1997. Yersinia pestis--etiologic agent of plague. *Clin Microbiol Rev* **10:** 35-66.

Perutz, M. F. 1983. Species adaptation in a protein molecule. *Mol Biol Evol* **1:** 1-28.

Pirnay, J. P., De Vos, D., Cochez, C., Bilocq, F., Vanderkelen, A., Zizi, M., Ghysels, B. et Cornelis, P. 2002. *Pseudomonas aeruginosa* displays an epidemic population structure. *Environ Microbiol* **4:** 898-911.

Pitt, T. L., Livermore, D. M., Pitcher, D., Vatopoulos, A. C. et Legakis, N. J. 1989. Multiresistant serotype O 12 *Pseudomonas aeruginosa*: evidence for a common strain in Europe. *Epidemiol Infect* **103:** 565-76.

Pourcel, C., Vidgop, Y., Ramisse, F., Vergnaud, G. et Tram, C. 2003. Characterization of a Tandem Repeat Polymorphism in *Legionella pneumophila* and Its Use for Genotyping. *J Clin Microbiol* **41:** 1819-1826.

Quelle, L. S., Corso, A., Galas, M. et Sordelli, D. O. 2003. STAR gene restriction profile analysis in epidemiological typing of methicillin-resistant Staphylococcus aureus: description of the new method and comparison with other polymerase chain reaction (PCR)-based methods. *Diagn Microbiol Infect Dis* **47:** 455-64.

Radnedge, L., Agron, P. G., Worsham, P. L. et Andersen, G. L. 2002. Genome plasticity in *Yersinia pestis*. *Microbiology* **148:** 1687-98.

Ramshaw, J. A., Coyne, J. A. et Lewontin, R. C. 1979. The sensitivity of gel electrophoresis as a detector of genetic variation. *Genetics* **93:** 1019-37.

Read, T. D. *et al.* 2002. Comparative genome sequencing for discovery of novel polymorphisms in Bacillus anthracis. *Science* **296:** 2028-33.

Renders, N., Romling, Y., Verbrugh, H. et van Belkum, A. 1996. Comparative typing of *Pseudomonas aeruginosa* by random amplification of polymorphic DNA or pulsed-field gel electrophoresis of DNA macrorestriction fragments. *J Clin Microbiol* **34:** 3190-5.

Renders, N., van Belkum, A., Barth, A., Goessens, W., Mouton, J. et Verbrugh, H. 1996. Typing of Pseudomonas aeruginosa strains from patients with cystic fibrosis: phenotyping versus genotyping. *Clin Microbiol Infect* **1:** 261-265.

Renders, N., Verbrugh, H. et Van Belkum, A. 2001. Dynamics of bacterial colonisation in the respiratory tract of patients with cystic fibrosis. *Infect Genet Evol* **1:** 29-39.

Rice, K., Peralta, R., Bast, D., de Azavedo, J. et McGavin, M. J. 2001. Description of staphylococcus serine protease (ssp) operon in Staphylococcus aureus and nonpolar inactivation of sspA-encoded serine protease. *Infect Immun* **69:** 159-69.

Rocha, E. P., Danchin, A. et Viari, A. 1999. Functional and evolutionary roles of long repeats in prokaryotes. *Res Microbiol* **150:** 725-33.

Rohrer, S., Tschierske, M., Zbinden, R. et Berger-Bachi, B. 2001. Improved methods for detection of methicillin-resistant Staphylococcus aureus. *Eur J Clin Microbiol Infect Dis* **20:** 267-70.

Romling, U., Fiedler, B., Bosshammer, J., Grothues, D., Greipel, J., von der Hardt, H. et Tummler, B. 1994. Epidemiology of chronic Pseudomonas aeruginosa infections in cystic fibrosis. *J Infect Dis* **170:** 1616-21.

Rosenberg, S. M., Longerich, S., Gee, P. et Harris, R. S. 1994. Adaptive mutation by deletions in small mononucleotide repeats. *Science* **265:** 405-407.

Sabat, A., Krzyszton-Russjan, J., Strzalka, W., Filipek, R., Kosowska, K., Hryniewicz, W., Travis, J. et Potempa, J. 2003. New method for typing Staphylococcus aureus strains: multiple-locus variable-number tandem repeat analysis of polymorphism and genetic relationships of clinical isolates. *J Clin Microbiol* **41:** 1801-4.

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. et Arnheim, N. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230:** 1350-4.

Salama, N., Guillemin, K., McDaniel, T. K., Sherlock, G., Tompkins, L. et Falkow, S. 2000. A whole-genome microarray reveals genetic diversity among Helicobacter pylori strains. *Proc Natl Acad Sci U S A* **97:** 14668-73.

Saunders, N. J., Peden, J. F., Hood, D. W. et Moxon, E. R. 1998. Simple sequence repeats in the Helicobacter pylori genome. *Mol Microbiol* **27:** 1091-8.

Schouls, L. M., Reulen, S., Duim, B., Wagenaar, J. A., Willems, R. J., Dingle, K. E., Colles, F. M. et Van Embden, J. D. 2003. Comparative genotyping of Campylobacter jejuni by amplified fragment length polymorphism, multilocus sequence typing, and short repeat sequencing: strain diversity, host range, and recombination. *J Clin Microbiol* **41:** 15-26.

Schwartz, D. C. & Cantor, C. R. 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37:** 67-75.

Schwarzkopf, A. & Karch, H. 1994. Genetic variation in Staphylococcus aureus coagulase genes: potential and limits for use as epidemiological marker. *J Clin Microbiol* **32:** 2407-12.

Schwiebert, E. M., Benos, D. J., Egan, M. E., Stutts, M. J. et Guggino, W. B. 1999. CFTR is a conductance regulator as well as a chloride channel. *Physiol Rev* **79:** S145-66.

Scieux, C., Grimont, F., Regnault, B. et Grimont, P. A. 1992. DNA fingerprinting of Chlamydia trachomatis by use of ribosomal RNA, oligonucleotide and randomly cloned DNA probes. *Res Microbiol* **143:** 755-65.

Selander, R. K., Caugant, D. A., Ochman, H., Musser, J. M., Gilmour, M. N. et Whittam, T. S. 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol* **51:** 873-84.

Senna, J. P., Pinto, C. A., Carvalho, L. P. et Santos, D. S. 2002. Comparison of pulsed-field gel electrophoresis and PCR analysis of polymorphisms on the mec hypervariable region for typing methicillin-resistant Staphylococcus aureus. *J Clin Microbiol* **40:** 2254-6

Sharp, P. M. & Li, W. H. 1987. Ubiquitin genes as a paradigm of concerted evolution of tandem repeats. *J Mol Evol* **25:** 58-64.

Sharples, G. J. & Lloyd, R. G. 1990. A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. *Nucleic Acids Res* **18:** 6503-8.

Shields, D. C., McDevitt, D. et Foster, T. J. 1995. Evidence against concerted evolution in a tandem array in the clumping factor gene of Staphylococcus aureus. *Mol Biol Evol* **12:** 963-5.

Shopsin, B., Gomez, M., Montgomery, S. O., Smith, D. H., Waddington, M., Dodge, D. E., Bost, D. A., Riehman, M., Naidich, S. et Kreiswirth, B. N. 1999. Evaluation of protein A gene polymorphic region DNA sequencing for typing of *Staphylococcus aureus* strains. *J Clin Microbiol* **37:** 3556-63.

Shopsin, B., Gomez, M., Waddington, M., Riehman, M. et Kreiswirth, B. N. 2000. Use of coagulase gene (coa) repeat region nucleotide sequences for typing of methicillin-resistant Staphylococcus aureus strains. *J Clin Microbiol* **38:** 3453-6.

Skuce, R. A., McCorry, T. P., McCarroll, J. F., Roring, S. M., Scott, A. N., Brittain, D., Hughes, S. L., Hewinson, R. G. et Neill, S. D. 2002. Discrimination of *Mycobacterium tuberculosis* complex bacteria using novel VNTR-PCR targets. *Microbiology* **148:** 519-528.

Smith, J. M., Dowson, C. G. et Spratt, B. G. 1991. Localized sex in bacteria. *Nature* **349:** 29-31.

Smith, J. M., Smith, N. H., O'Rourke, M. et Spratt, B. G. 1993. How clonal are bacteria? *Proc Natl Acad Sci U S A* **90:** 4384-8.

Smith, N. H., Beltran, P. et Selander, R. K. 1990. Recombination of Salmonella phase 1 flagellin genes generates new serovars. *J Bacteriol* **172:** 2209-16.

Snel, B., Bork, P. et Huynen, M. A. 1999. Genome phylogeny based on gene content. *Nat Genet* **21:** 108-10.

Sokurenko, E. V., Hasty, D. L. et Dykhuizen, D. E. 1999. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends Microbiol* **7:** 191-5.

Sola, C., Filliol, I., Legrand, E., Lesjean, S., Locht, C., Supply, P. et Rastogi, N. 2003. Genotyping of the Mycobacterium tuberculosis complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. *Infect Genet Evol* **3:** 125-33.

Spencer, D. H., Kas, A., Smith, E. E., Raymond, C. K., Sims, E. H., Hastings, M., Burns, J. L., Kaul, R. et Olson, M. V. 2003. Whole-genome sequence variation among multiple isolates of Pseudomonas aeruginosa. *J Bacteriol* **185:** 1316-25.

Sreevatsan, S., Pan, X., Stockbauer, K. E., Connell, N. D., Kreiswirth, B. N., Whittam, T. S. et Musser, J. M. 1997. Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci U S A* **94:** 9869-74.

Stern, M. J., Ames, G. F., Smith, N. H., Robinson, E. C. et Higgins, C. F. 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell* **37:** 1015-26.

Stothard, D. R., Van Der Pol, B., Smith, N. J. et Jones, R. B. 1998. Effect of serial passage in tissue culture on sequence of omp1 from Chlamydia trachomatis clinical isolates. *J Clin Microbiol* **36:** 3686-8.

Stover, C. K. *et al.* 2000. Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* **406:** 959-64.

Suerbaum, S., Smith, J. M., Bapumia, K., Morelli, G., Smith, N. H., Kunstmann, E., Dyrek, I. et Achtman, M. 1998. Free recombination within Helicobacter pylori. *Proc Natl Acad Sci U S A* **95:** 12619-24.

Supply, P., Lesjean, S., Savine, E., Kremer, K., van Soolingen, D. et Locht, C. 2001. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. *J Clin Microbiol* **39:** 3563-3571.

Supply, P., Magdalena, J., Himpens, S. et Locht, C. 1997. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. *Mol Microbiol* **26:** 991-1003.

Supply, P., Mazars, E., Lesjean, S., Vincent, V., Gicquel, B. et Locht, C. 2000. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Mol Microbiol* **36:** 762-771.

Supply, P., Warren, R. M., Banuls, A. L., Lesjean, S., Van Der Spuy, G. D., Lewis, L. A., Tibayrenc, M., Van Helden, P. D. et Locht, C. 2003. Linkage disequilibrium between minisatellite loci supports clonal evolution of Mycobacterium tuberculosis in a high tuberculosis incidence area. *Mol Microbiol* **47:** 529-38.

Sutter, V. L., Hurst, V. et Fennell, J. 1965. A Standardized System for Phage Typing Pseudomonas Aeruginosa. *Health Lab Sci* **30:** 7-16.

Sylvestre, P., Couture-Tosi, E. et Mock, M. 2003. Polymorphism in the collagen-like region of the *Bacillus anthracis* BclA protein leads to variation in exosporium filament length. *J Bacteriol* **185:** 1555-63.

Tassios, P. T., Gennimata, V., Maniatis, A. N., Fock, C. et Legakis, N. J. 1998. Emergence of multidrug resistance in ubiquitous and dominant *Pseudomonas aeruginosa* serogroup O:11. The Greek *Pseudomonas Aeruginosa* Study Group. *J Clin Microbiol* **36:** 897-901.

Tenover, F. C. *et al.* 1994. Comparison of traditional and molecular methods of typing isolates of Staphylococcus aureus. *J Clin Microbiol* **32:** 407-15.

Tenover, F. C., Arbeit, R. D., Goering, R. V., Mickelsen, P. A., Murray, B. E., Persing, D. H. et Swaminathan, B. 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol* **33:** 2233-9.

Theiss, P. & Wise, K. S. 1997. Localized frameshift mutation generates selective, high-frequency phase variation of a surface lipoprotein encoded by a mycoplasma ABC transporter operon. *J Bacteriol* **179:** 4013-22.

Tibayrenc, M. 1998. Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. *Int J Parasitol* **28:** 85-104.

Tomb, J. F. *et al.* 1997. The complete genome sequence of the gastric pathogen Helicobacter pylori. *Nature* **388:** 539-47.

Tonjum, T., Caugant, D. A., Dunham, S. A. et Koomey, M. 1998. Structure and function of repetitive sequence elements associated with a highly polymorphic domain of the Neisseria meningitidis PilQ protein. *Mol Microbiol* **29:** 111-24.

Tsang, V. C., Peralta, J. M. et Simons, A. R. 1983. Enzyme-linked immunoelectrotransfer blot techniques (EITB) for studying the specificities of antigens and antibodies separated by gel electrophoresis. *Methods Enzymol* **92:** 377-91.

van Belkum, A. 1999. The role of short sequence repeats in epidemiologic typing. *Curr Opin Microbiol* **2:** 306-11.

van Belkum, A. 1999. Short sequence repeats in microbial pathogenesis and evolution. *Cell Mol Life Sci* **56:** 729-34.

van Belkum, A. 2000. Molecular epidemiology of methicillin-resistant Staphylococcus aureus strains: state of affairs and tomorrow' s possibilities. *Microb Drug Resist* **6:** 173-88.

van Belkum, A. 2003. Molecular diagnostics in medical microbiology: yesterday, today and tomorrow. *Curr Opin Pharmacol* **3:** 497-501.

van Belkum, A., Bax, R., Peerbooms, P., Goessens, W. H., van Leeuwen, N. et Quint, W. G. 1993. Comparison of phage typing and DNA fingerprinting by polymerase chain reaction for discrimination of methicillin-resistant Staphylococcus aureus strains. *J Clin Microbiol* **31:** 798-803.

van Belkum, A., Scherer, S., van Alphen, L. et Verbrugh, H. 1998. Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol Rev* **62:** 275-93.

van Belkum, A., Scherer, S., van Leeuwen, W., Willemse, D., van Alphen, L. et Verbrugh, H. 1997. Variable number of tandem repeats in clinical strains of *Haemophilus influenzae*. *Infect Immun* **65:** 5017-27.

van Belkum, A., Sluijuter, M., de Groot, R., Verbrugh, H. et Hermans, P. W. 1996. Novel BOX repeat PCR assay for high-resolution typing of Streptococcus pneumoniae strains. *J Clin Microbiol* **34:** 1176-9.

van Belkum, A., Struelens, M., de Visser, A., Verbrugh, H. et Tibayrenc, M. 2001. Role of genomic typing in taxonomy, evolutionary genetics, and microbial epidemiology. *Clin Microbiol Rev* **14:** 547-60.

van der Ende, A., Hopman, C. T., Zaat, S., Essink, B. B., Berkhout, B. et Dankert, J. 1995. Variable expression of class 1 outer membrane protein in Neisseria meningitidis is caused by variation in the spacing between the -10 and -35 regions of the promoter. *J Bacteriol* **177:** 2475-80.

van Embden, J. D. *et al.* 1993. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* **31:** 406-9.

van Embden, J. D., van Gorkom, T., Kremer, K., Jansen, R., van Der Zeijst, B. A. et Schouls, L. M. 2000. Genetic variation and evolutionary origin of the direct repeat locus of Mycobacterium tuberculosis complex bacteria. *J Bacteriol* **182:** 2393-401.

van Ham, S. M., van Alphen, L., Mooi, F. R. et van Putten, J. P. 1993. Phase variation of H. influenzae fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. *Cell* **73:** 1187-96.

van Leeuwen, W., Verbrugh, H., van der Velden, J., van Leeuwen, N., Heck, M. et van Belkum, A. 1999. Validation of binary typing for Staphylococcus aureus strains. *J Clin Microbiol* **37:** 664-74.

van Leeuwen, W. B., Jay, C., Snijders, S., Durin, N., Lacroix, B., Verbrugh, H. A., Enright, M. C., Troesch, A. et van Belkum, A. 2003. Multilocus sequence typing of Staphylococcus aureus with DNA array technology. *J Clin Microbiol* **41:** 3323-6.

van Leeuwen, W. B. *et al.* 2002. Intercenter reproducibility of binary typing for Staphylococcus aureus. *J Microbiol Methods* **51:** 19-28.

van Leeuwen, W. B., van Pelt, C., Luijendijk, A., Verbrugh, H. A. et Goessens, W. H. 1999. Rapid detection of methicillin resistance in Staphylococcus aureus isolates by the MRSA-screen latex agglutination test. *J Clin Microbiol* **37:** 3029-30.

Vaneechoutte, M., Rossau, R., De Vos, P., Gillis, M., Janssens, D., Paepe, N., De Rouck, A., Fiers, T., Claeys, G. et Kersters, K. 1992. Rapid identification of bacteria of the Comamonadaceae with amplified ribosomal DNA-restriction analysis (ARDRA). *FEMS Microbiol Lett* **72:** 227-33.

Vergnaud, G. 1989. Polymers of random short oligonucleotides detect polymorphic loci in the human genome. *Nucleic Acids Res.* **17:** 7623-7630.

Versalovic, J., Koeuth, T. et Lupski, J. R. 1991. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Res* **19:** 6823-31.

Vlaminckx, B. J., Mascini, E. M., Schellekens, J., Schouls, L. M., Paauw, A., Fluit, A. C., Novak, R., Verhoef, J. et Schmitz, F. J. 2003. Site-specific manifestations of invasive group a streptococcal disease: type distribution and corresponding patterns of virulence determinants. *J Clin Microbiol* **41:** 4941-9.

Vos, P. *et al.* 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* **23:** 4407-14.

Wei, M. Q., Udo, E. E. et Grubb, W. B. 1992. Typing of methicillin-resistant Staphylococcus aureus with IS256. *FEMS Microbiol Lett* **78:** 175-80.

Weiser, J. N., Love, J. M. et Moxon, E. R. 1989. The molecular mechanism of phase variation of H. influenzae lipopolysaccharide. *Cell* **59:** 657-65.

Weiser, J. N., Williams, A. et Moxon, E. R. 1990. Phase-variable lipopolysaccharide structures enhance the invasive capacity of Haemophilus influenzae. *Infect Immun* **58:** 3455-7.

Weissman, S. J., Moseley, S. L., Dykhuizen, D. E. et Sokurenko, E. V. 2003. Enterobacterial adhesins and the case for studying SNPs in bacteria. *Trends Microbiol* **11:** 115-7.

Welsh, J. & McClelland, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res* **18:** 7213-8.

Wichelhaus, T. A., Hunfeld, K. P., Boddinghaus, B., Kraiczy, P., Schafer, V. et Brade, V. 2001. Rapid molecular typing of methicillin-resistant Staphylococcus aureus by PCR-RFLP. *Infect Control Hosp Epidemiol* **22:** 294-8.

Wielders, C. L., Fluit, A. C., Brisse, S., Verhoef, J. et Schmitz, F. J. 2002. mecA gene is widely disseminated in Staphylococcus aureus population. *J Clin Microbiol* **40:** 3970-5.

Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. et Tingey, S. V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18:** 6531-6535.

Windgassen, M., Urban, A. et Jaeger, K. E. 2000. Rapid gene inactivation in Pseudomonas aeruginosa. *FEMS Microbiol Lett* **193:** 201-5.

Woolhouse, M. E. 2002. Population biology of emerging and re-emerging pathogens. *Trends Microbiol* **10:** S3-7

Wyman, A. R. & White, R. 1980. A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. USA* **77:** 6754-6758.

Yang, Y. & Ames, G. F. 1988. DNA gyrase binds to the family of prokaryotic repetitive extragenic palindromic sequences. *Proc Natl Acad Sci U S A* **85:** 8850-4.

Yogev, D., Rosengarten, R., Watson-McKown, R. et Wise, K. S. 1991. Molecular basis of Mycoplasma surface antigenic variation: a novel set of divergent genes undergo spontaneous mutation of periodic coding regions and 5' regulatory sequences. *Embo J* **10:** 4069-79.

Yother, J. & Briles, D. E. 1992. Structural properties and evolutionary relationships of PspA, a surface protein of Streptococcus pneumoniae, as revealed by sequence analysis. *J Bacteriol* **174:** 601-9.

Young, V. M. & Moody, M. R. 1974. Serotyping of Pseudomonas aeruginosa. *J Infect Dis* **130 Suppl:** S47-52.

Zheng, X., Teng, L. J., Watson, H. L., Glass, J. I., Blanchard, A. et Cassell, G. H. 1995. Small repeating units within the Ureaplasma urealyticum MB antigen gene encode serovar specificity and are associated with antigen size variation. *Infect Immun* **63:** 891-8.

ANNEXES

ANNEXE 1 : Liste des solutions utilisées

Milieu gélosé LBA (Luria Bertani)+ Agar :

Bactotryptone	1%
Extrait de levure	0,5%
NaCl	0,171M
Agar	0,7%
Ajuster le pH à 7,5	

Milieu LB (Luria Bertani) :

Bactotryptone	1%
Extrait de levure	0,5%
NaCl	0,171M
Ajuster le pH à 7,5	

Tampon de lyse :

Tris HCl ph 8,5	50mM
EDTA	5mM
Triton 100X	1%

« Bleu de dépôt » :

Bleu de xylène cyanol (Sigma)	0,1mg/ml
Glycerol	50%
TE	0,5X

TE 1X :

Tris HCl pH 8,0	0,01mM
EDTA	0,001mM

TBE 0,5 X :

Tris HCl	0,045M
Acide borique	0,045M
EDTA	0,001M
Ajuster à pH 8,3	

ANNEXE 2 : 137 souches *Staphylococcus aureus* étudiées

Souches Pasteur :

n° ADN	n° souche	Génotype PFGE	Année	Hôpital - Ville	Marqueurs de Résistance aux antibiotiques	Références
1	BM12846	1	1999	Broussais- Paris	Chl, Pt,PIB,PIIA,Pef	
2	BM12451	2	1997	Grenoble	Pc,Sm,Km,Tm,Gm,Em,Pt,PIB,PIIA,Lc,Su,Tmp	
3	BM 10675	29	1993	Broussais- Paris	Pc,**Ox**,Sm,Km,Nm,Tm,Gm,PIIA,Rf	
4	2001042	30a	1999	Toulouse	Pc,PIIA,Lc,Pef	
5	BM 9520	30b	1988	Broussais- Paris	Pc,Sm,Spe,Km,Nm,Tm,Gm,Pt,PIB,PIIA,Lc	
6	99135	31	1999	St Louis-Paris	Pc,Km,Tm,Gm,Tc,Em,Spi,Cl,Pt,PIB,PIIA,Lc,Fa	
7	BM 12828	16a	1999	Broussais- Paris	Pc,**Ox**,Spe,Km,Nm,Tm,Em,Spi,Cl,Pt,PIB,PIIA,Lc,Tp,Pef	
8	BM 12830	16b	1999	Broussais- Paris	Pc,**Ox**,Spe,Km,Nm,Tm,Gm,Em,Spi,Cl,Pt,PIB,PIIA,Lc,Pef,Rf	
9	BM 12174	17b	1996	Grenoble	Pc,**Ox**,Spe,Km,Nm,Tm,Gm,Em,Spi,Cl,Pt,PIB,PIIA,Lc,Tp,Pef,Fm	
10	BM 3364	13	1981	Broussais- Paris	Pc,**Ox**,Sm,Spe,Km,Nm,Tm,Gm,Tc,Mn,Em,Spi,Cl,Pt,PIB,PIIA,Lc	
11	BM 10215	18	1990	Broussais- Paris	Pc,Sm,Km,Nm,Em,PIIA,Lc,Pef	
12	BM 12942	19	1999	Broussais- Paris	Pc,**Ox**,Sm,Km,Nm,Em,Spi,Cl,Pt,PIB,PIIA,Lc,Su,Tp,Fa	
13	97233	24h	1997	Broussais- Paris	Pc,**Ox**,Km,Nm,Tm,Gm,Em,Spi,Cl,Pt,PIB,PIIA,Lc,Tp,Pef	
14	2001045	24a	1998	Toulouse	Pc,**Ox**,Km,Nm,Tm,PIIA,Pef	
15	93184	26	1993	Laennec- Paris	Pc,**Ox**,Km,Nm,Tm,Pt,PIB,PIIA,Pef,Rf	
16	BM 12286	36a	1996	Beaujon-Clichy	Pc,**Ox**,Sm,Km,Nm,Pt,PIB,PIIA	
17	BM 12287	36a	1996	Beaujon-Clichy	Pc,**Ox**,Sm,Spe,Km,Nm,Pt,PIB,PIIA	
18	IPF 556	I -1	1999	Broussais - Paris	Pc,**Ox**,Sm,Km,Tm,Gm,Tc,Mn,Em,Lc,Pt,Su,Pef,Fa,Rf,Fm	
19	IPF 555	I -1	1999	Broussais - Paris	Pc,**Ox**,Sm,Km,Tm,Gm,Tc,Mn,Em,Lc,Pt,Su,Pef,Fa,Rf,	

n° ADN	n° souche	Génotype PFGE	Année	Hôpital - Ville	Marqueurs de Résistance aux antibiotiques	Références
20	IPF 557	I -1	1999	Broussais - Paris	Pc,**Ox**,Sm,Km,Tm,Gm,Tc,Mn,Em,Lc,Pt,Su,Pef,Fa,Rf,Fm	
21	IPF 562	I -1	1999	Broussais - Paris	Pc,**Ox**,Sm,Km,Tm,Gm,Tc,Mn,Em,Lc,Pt,Su,Pef,Fa,Rf,Fm	
22	BM 10828	I -1	1993	Bordeaux	Pc,**Ox**,Sm,Spe,Km,Nm,Tm,Gm,Tc,Pef,Rf,Fa,Fm,Em,Spi,Lc,Cl	
23	HM10	XXIV	2000	Henri Mondor-Créteil	Pc,**Ox**,Sm,Km,Tm,Gm,Rf, fm,Em,Spi,Lc,Cl,PIB	
24	BM12612	I - 1	1998	Villiers St Denis	Pc,**Ox**,Sm,Spe,Km,Nm,Tm,Gm,Tc,Pef,Rf,Fm,Em,Spi,Lc,Cl	
25	IPF 216b	XVII	2000-2001	Bordeaux	Pc,**Ox**,Spe,Nm,Su,	
26	Harmony 26	I - 1	1989	Espagne	Pc,**Ox**,Sm,Spe,Km,Tm,Gm,Tc,Su,Pef,Fa,Em,Spi,Lc,Cl	
27	IPF 340	XVIII	2001	St Etienne	Pc,**Ox**,Km,Tm,Gm,Nm,Rf	
28	Harmony 9	I - 5	1990	Finlande	Pc,**Ox**,Sm,Spe,Km,Tm,Gm,Tc,Pef,Rf,Em,Spi,Lc,Cl	
29	BM 10829	I - 5	1993	Bordeaux	Pc,**Ox**,Sm,Spe,Km,Nm,Tm,Gm,Tc,Pef,Rf,Fa,Fm,Em,Spi,Lc,Cl,PIB	
30	HM 9	XX	2000	Henri Mondor-Créteil	Pc,**Ox**,Sm,Spe,Km,Nm,Tm,Gm,Tc,Mn,Su,Pef,Rf,Fm,Em,Spi,Lc,Cl,PIB	
31	97130	I - 18	1997	Toulouse	Pc,**Ox**,Sm,Spe,Km,Nm,Tm,Gm,Tc,Pef,Rf,Fa,Fm,Em	
32	96145	I - 17	1996	Blois	Pc,**Ox**,Sm,Spe,Km,Nm,Tm,Gm,Tc,Pef,Rf,Fm,Em,Spi,Lc,Cl	
33	BM 12632	40e	1998	Hôpital pédiatrique Tunisie	Lc	
34	BM 12634	36a	1998	Hôpital pédiatrique Tunisie	Pc,Sm,Km,Em	
35	BM12975	45c	1999	Hôpital pédiatrique Tunisie	Pc,Cm	
36	IPF160	38c	2000	Hôpital pédiatrique Tunisie	Pc	
37	IPF489	42a	2001	Hôpital pédiatrique Tunisie	Pc,Sm,Km,Tc,Fa	
38	IPF518	22a	2001	Hôpital pédiatrique Tunisie	Pc	
39	IPF 509	57	2001	Hôpital pédiatrique Tunisie	Pc,Tc	
40	IPF 510	37a	2001	Hôpital pédiatrique Tunisie	Pc	
41	IPF 511	3	2001	Hôpital pédiatrique Tunisie	Pc	
42	IPF 512	48a	2001	Hôpital pédiatrique Tunisie	Pc	
43	IPF641	B	2000	Centre Hospitalier Nimes	Pc,**Ox**,Tm,Em,Cl,Pt,Pef,Fm	

n° ADN	n° souche	Génotype PFGE	Année	Hôpital - Ville	Marqueurs de Résistance aux antibiotiques	Références
44	IPF642	A1	2000	Centre Hospitalier Nimes	Pc,**Ox**,Tm,Em,Cl,Pef,Fm	
45	IPF643	C	2000	Centre Hospitalier Nimes	Pc,**Ox**,Tm,Em,Cl,Pef,Fm	
46	IPF644	A2	2000	Centre Hospitalier Nimes	Pc,**Ox**,Tm,Em,Cl,Pef,Fm	
47	IPF735	*	juin-01	Centre Hospitalier Calais	Pc,Em,(Spi)	
48	IPF736	*	juin-01	Centre Hospitalier Calais	Pc,Em,(Spi)	
49	IPF738	*	juin-01	Centre Hospitalier Calais	Pc,(Spi)	
50	IPF741	*	juin-01	Centre Hospitalier Calais	Pc,Em,(Spi)	
51	IPF743	*	juin-01	Centre Hospitalier Calais	Pc,Em,(Spi)	
52	IPF646	$	Fev.2000	Centre Hospitalier Abbeville	Pc,**Ox**,Tm,Km,Em,Lc,Pef,Rf	
53	IPF647	#$	Janv.2000	Centre Hospitalier Abbeville	Pc,**Ox**,Tm,Km,Em,Lc,Pef	
54	IPF648	$	Janv.2000	Centre Hospitalier Abbeville	Pc,**Ox**,Tm,Km,Em,Lc,Pef	
55	IPF654	^	Juil.1999	Fécamp (clinique privée)	Pc,**Ox**,Km,Tm,Em,Lc,Pef,Fm	
56	IPF657	^	Dec.1999	Fécamp (clinique privée)	Pc,**Ox**,Km,Tm,Em,Lc,Pef,Fm	
57	IPF659	^	Janv.2000	Fécamp (clinique privée)	Pc,**Ox**,Km,Tm,Em,Lc,Pef	
58	IPF658	#^	Fev.2000	Fécamp (clinique privée)	Pc,**Ox**,Km,Tm,Em	
59	IPF667	##^	Mar.2000	Fécamp (clinique privée)	Pc,**Ox**,Km,Tm,Em,Lc,Pef	
60	IPF308	~	mai-01	CH Intercommunal Créteil	Pc	
61	IPF309	~	juin-01	CH Intercommunal Créteil	Pc	
62	IPF310	~	mai-01	CH Intercommunal Créteil	Pc	
63	IPF311	~	Juil.2001	CH Intercommunal Créteil	Pc	
64	IPF323	~	Oct.2001	CH Intercommunal Créteil	Pc	
65	IPF92	&	Juil.2000	Rotschild- Paris	Pc,**Ox**,Sm,Spe,Km,Nm,Tm,MLSc,Pf,Fm	
66	IPF54	&	Aout2000	Rotschild- Paris	Pc,**Ox**,Sm,Spe,Km,Nm,Tm,MLSc,SgA,Pt,Pf,Fm	
67	IPF55	&	Sept.2000	Rotschild- Paris	Pc,**Ox**,Sm,Spe,Km,Nm,Tm,MLSc,Pf,Fm	
68	IPF56	&	Sept.2000	Rotschild- Paris	Pc,**Ox**,Sm,Spe,Km,Nm,Tm,MLSc,SgA,Pt,Pf,Fm	
69	IPF57	&	2000	Rotschild- Paris	Pc,**Ox**,Spe,Km,Nm,Tm,MLSc,SgA,Pt,Pef,Fm	
70	IPF66	&	Sept.2000	Rotschild- Paris	Pc,**Ox**,Sm,Spe,Km,Nm,Tm,MLSc,SgA,Pt,Pf,Fm	
71	IPF65	#&	Aout2000	Rotschild- Paris	Pc,**Ox**,Sm,Spe,Km,Nm,Tm,Gm,Tc,Mn,MLSc,SgA,Pt,Rf,Tp,Fa,Pf,Fm,TSU	

n° ADN	n° souche	Génotype PFGE	Année	Hôpital - Ville	Marqueurs de Résistance aux antibiotiques	Références
72	166	Ia	1995	Beaujon - CLICHY	Pc,**Ox**,Nm,Sp,MLSc,Pf	Galdbart JO, J Clin Microbiol. 2000 Jan;38(1):185-90
73	165	Ia	1995	Beaujon - CLICHY	Pc,**Ox**,Nm,Sp,MLSc,Pf	Galdbart JO, J Clin Microbiol. 2000 Jan;38(1):185-90
74	97386	Ib	1996	Broussais - PARIS	Pc,**Ox**,Gm,Nm,Sp,MLSc,Pf,Fm	Galdbart JO, J Clin Microbiol. 2000 Jan;38(1):185-90
75	97381	VIIb	1996	Broussais - PARIS	Pc,**Ox**,Gm,Nm,Sp,MLSC,Pf,Fm	Galdbart JO, J Clin Microbiol. 2000 Jan;38(1):185-90
76	97383	VIIa	1996	Broussais - PARIS	Pc,**Ox**,Nm,Sp,MLSc,Pf,Fm	Galdbart JO, J Clin Microbiol. 2000 Jan;38(1):185-90
77	97379	VIIa	1996	Broussais - PARIS	Pc,**Ox**, Nm,Sp,MLSc,SgA,Pf,Fm,Tp	Galdbart JO, J Clin Microbiol. 2000 Jan;38(1):185-90
78	97117	VIId	1997	TOULOUSE	Pc,**Ox**,Nm,Sm,Sp,MLSc,SgA,Pf	Galdbart JO, J Clin Microbiol. 2000 Jan;38(1):185-90
79	95035	VIId	1995	CH - NIMES	Pc,**Ox**,Nm,Sm,Sp,MLSc,Pf,Fm	Galdbart JO, J Clin Microbiol. 2000 Jan;38(1):185-90
80	162	VIId	1995	Beaujon - CLICHY	Pc,**Ox**,NM,Sm,Sp,MLSc,Pf	Galdbart JO, J Clin Microbiol. 2000 Jan;38(1):185-90
81	96164	VIId	1996	CH BLOIS	Pc,**Ox**,Nm,Sp,MLSc,Pf,Fm	Galdbart JO, J Clin Microbiol. 2000 Jan;38(1):185-90
82	920	VIb	1995	St Germain	Pc,**Ox**,Nm,Sm,Sp,Lc,SgA,Pf	Galdbart JO, J Clin Microbiol. 2000 Jan;38(1):185-90
83	1625	VIb	1995	Beaujon - CLICHY	Pc,**Ox**,Nm,Sm,Sp,MLSc,Pf,Rf	Galdbart JO, J Clin Microbiol. 2000 Jan;38(1):185-90
84	12072	VIb	1995	Hop. L. Mourrier-COLOMBES	Pc,**Ox**,Nm,Sm,Sp,Pf	Galdbart JO, J Clin Microbiol. 2000 Jan;38(1):185-90
85	12068	VIb	1995	CH - EVRY	Pc,**Ox**,Nm,Sm,Sp,Pf	Galdbart JO, J Clin Microbiol. 2000 Jan;38(1):185-90
86	97120	VIb	1997	TOULOUSE	Pc,**Ox**,Nm,Sm,Pf,Rf,Fa	Galdbart JO, J Clin Microbiol. 2000 Jan;38(1):185-90
87	97373	XVIII	1996	Broussais - PARIS	Pc,**Ox**,Nm,MLSc,Tc,Mn	Galdbart JO, J Clin Microbiol. 2000 Jan;38(1):185-90
88	BM9290	A	1987	Hôtel Dieu - PARIS	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm,Rf	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.

n° ADN	n° souche	Génotype PFGE	Année	Hôpital - Ville	Marqueurs de Résistance aux antibiotiques	Références
89	BM9586	A	Jan.1987	Broussais - PARIS	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm,Rf,Fm	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.
90	BM12184	A	Avr.1997	Broussais - PARIS	Pc,**Ox** ,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm,Rf,Fm	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.
91	BM9591	A	Juil.1987	Broussais - PARIS	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm,Rf	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.
92	BM12188	A	Dec.1987	Broussais - PARIS	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm,Rf	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.
93	BM10761	A 4	mai-93	TOULOUSE	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm,Rf,Nm,SgA,Tp,Fa,Fm	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.
94	BM10759	A 4	mai-93	TOULOUSE	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm,Rf,Nm,SgA,Tp,Fa,Fm	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.
96	BM9343	A 7	1987	TOULOUSE	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm,Rf,Nm,SgA,Tp,Fa,Fm	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.
97	BM10872	A 8	Dec.1992	Aalst - BELGIQUE	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.
98	BM10888	A 8	Aout 1993	Aalst - BELGIQUE	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm,Nm	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.
100	BM10914	A 9	Ma1991	Broussais - PARIS	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm,Rf,Nm,SgA,Tp,Fm	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.
101	BM10896	A 9	Dec.1994	Ghent -BELGIQUE	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm,Nm	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.
102	BM10138	A 10	Oct.1989	Barcelone - ESPAGNE	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm,Rf,Nm	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.
103	BM10130	A 10	Oct.1989	Barcelone - ESPAGNE	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm,Rf,Nm	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.
105	BM12152	A 12	Oct.1989	Barcelone - ESPAGNE	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm,Rf,Nm	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.
106	BM10877	G	Fev.1992	Aalst - BELGIQUE	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.
107	BM12156	G	Oct.1989	Barcelone - ESPAGNE	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm,Rf,Nm	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.
109	BM10827	K	Nov.1993	BORDEAUX	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm,Rf,Nm,Fa,Fm	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.

n° ADN	n° souche	Génotype PFGE	Année	Hôpital - Ville	Marqueurs de Résistance aux antibiotiques	Références
110	BM10883	Q	Fev.1993	Ghent - Belgique	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.
111	BM10886	R	avr-93	Ghent - Belgique	Pc,**Ox**,Pf,Tc,Mn,MLSc,Sp,Sm,Km,Tm,Gm,Nm	Morvan A. J Clin Microbiol. 1997 Jun;35(6):1415-23.

Marqueurs de résistance aux antibiotiques

Pc : Pénicilline
Ox :Oxacilline (= méthicilline)
Spe :Spectinomycine
Sm :Streptomycine
Km :Kanamycine
Nm :Néomycine
Tm :Tobramycine
C :Chloramphénicol
Tc :Tétracycline
Mn :Minocycline
Em :Erythromycine
Sp :Spiramycine
Lc :Lincomycine
Pt :Pristinamycine
PIB :Pristinamycine IB
PIIA :Pristinamycine IIA
Su :Sulfamides
Tp :Triméthoprime
Pef :Pefloxacine
Fa :Acide fusidique
Fm :Fosfomycine
MLSc: Macrolides-Lincosamides-Streptogramines constitutif

Souches Val de Grâce :

n° ADN	n° souche	Année	Hôpital
134	SARM réf JV		collection Pitié
135	SASM-2		collection Pitié
136	SASM-4		collection Pitié
137	SASM-5		collection Pitié
138	SASM-6		collection Pitié
139	SASM-7		collection Pitié
140	SASM-8		collection Pitié
141	SASM-11		collection Pitié
142	SASM-12		collection Pitié
143	SARM-15	1996	collection Pitié
144	SARM-29	1996	collection Pitié
145	SARM-58	1996	collection Pitié
146	SARM-88	1996	collection Pitié
147	SARM-100	1996	collection Pitié
148	SARM-119	1996	collection Pitié
149	SARM-131	1996	collection Pitié
150	SARM-132	1996	collection Pitié
151	SARM-135	1996	collection Pitié
152	SARM-142	1996	collection Pitié
153	SARM-192	1996	collection Pitié
154	SARM-5170201		HIA Val de Grâce
155	SARM-4030210		HIA Val de Grâce
156	SARM-60300661		HIA Val de Grâce
157	SARM-60300663		HIA Val de Grâce
158	SARM-5040072		HIA Val de Grâce
159	Mu3(SARM)		HIA Val de Grâce
160	Mu50(GISA)		HIA Val de Grâce
161	GISA-Xavier16-10-3		HIA Val de Grâce
162	GISA-1106024		HIA Val de Grâce
163	GISA-904062		HIA Percy

SARM: *S.aureus* résistante à la méthicilline
SASM: *S.aureus* sensible à la méthicilline
GISA: Glycopeptide Intermédiaire *S.aureus*

6 souches de Référence séquencées :

Nom de la souche	caractéristiques	Références
Mu50	souche hospitalière	Kuroda M, Lancet. 2001 Apr 21;357(9264):1225-40
MSSA476	souche hyper-virulente acquise dans la communauté	non publié
NCTC8325	vieille souche de laboratoire	non publié
MW2	souche hyper virulente acquise dans la communauté	Baba T, Lancet. 2002 May 25;359(9320):1819-27.
N315	souche hospitalière	Kuroda M, Lancet. 2001 Apr 21;357(9264):1225-40
MRSA252	souche hospitalière épidémique	non publié

ANNEXE 3 :

Liste des 23 répétitions en tandem de *Pseudomonas aeruginosa* abandonnées lors de l'étude MLVA, pour lesquelles des problèmes d'amplification par PCR ont été rencontrés :

Nom de la répétition	Position dans le génome PAO1 (en kb)	L: taille du motif souche PAO1 (pb)	N: nombre de motifs souche PAO1
ms006	44	**63**	56
ms019	308	**9**	12
ms023	388	**9**	31
ms025	417	**12**	30
ms031	750	**15**	7
ms041	1165	**15**	7
ms047	1466	**9**	15
ms057	1727	**9**	11
ms068	2036	**246**	22
ms074	2247	**10**	8
ms084	2512	**18**	9
ms088	2550	**24**	21
ms102	2768	**21**	119
ms137	3750	**12**	12
ms138	3791	**15**	7
ms140	3858	**21**	11
ms144	3925	**12**	11
ms154	4432	**15**	8
ms156	4590	**15**	8
ms160	4730	**12**	24
ms162	4732	**24**	7
ms170	5043	**12**	10
ms189	5700	**12**	33

Liste des 170 répétitions en tandem monomorphes chez *P. aeruginosa* : amorces utilisées et conditions de PCR :

Nom de la répétition	Position dans le génome PAO1 (en kb)	L: taille du motif souche PAO1 (pb)	N: nombre de motifs souche PAO1	Taille du produit PCR PAO1 (pb)	Motif consensus	Amorce gauche	Amorce droite	T° hybridation des amorces (°C)
ms001	3989	**12**	14	263	AGCCGGCGATGA	ATGACCAGGATCGGCTTCAG	GACACGGCTTTCGTCCTGAT	60
ms002	239	**9**	18	272	CCGCGCCGG	GAGCGTCCTCGACGATACCT	GAAGGTCAGGGTTTCCATCTC	60
ms003	1370	**15**	15	336	GGCGGCAGCAGCGAC	GAAGAACAGGATCGCCAGTAAC	CCTGATCAACCTGCCGATCT	59,3
ms004	22	**15**	5	166	CGCCGACGGCAAGTT	ATGACGAGAAGACCTTCACCAG	CTTGGTCTTCGGATCGATGTA	60
ms005	30	**21**	6	239	GCGGCAGCCTTCGTAGTGCAG	ACAAGCTGCAGCGGGAAT	CCAGGTGAACTTCATCGACTT	60
ms007	76	**24**	11	357	GCCAGCAGGCCGACGCCGAGCAGC	AGCTGGTCGGCGTAGCAC	GATGCAGTTCCTTCCCTACG	60
ms008	82	**12**	11	227	GCCGGCCAGGCC	ATCGCTGATCTGCTTCTGCT	ACTACCCGTCCTGCTACGAG	60
ms009	85	**9**	9	248	CGGCGATGG	GGAAGCCCAGCAGACGTAG	CCTGTACGCCCGAGACCT	60
ms011	114	**12**	9	228	CGGCGGCGACCA	CCCTTGACGATCTGGAACAT	CCCTAGCTGACCCGAGTACC	60
ms012	116	**48**	10	573	GTCGGCAACAACGAGACCATCAGCATCGGCGCGACCGCACCGAGAAC	GAGCAGTTGTTCATCCATGC	GTTGAGCGTGAAGCTCTTGC	60
ms013	155	**13**	9	230	TCGCCGGCGGCGG	ACCTCGACGAGCTGTTCCT	GAGAGGCCAGCAGCATCC	60
ms014	156	**10**	16	336	TCGGCGCTGC	ACTACATCGCGCTGAACCTG	ACAGCAGCACGAACCAGAC	60
ms015	161	**14**	7	227	CGGGCCGGCTGGC	GTGCTGTCCTCGCTGCTG	ATGAGCGAAGAACCCACTGT	60
ms016	196	**12**	5	216	CAGCGCCAGCAA	GTGCAGCAGGGTGAAGTTCT	GCTTTCGATCCATGATTTCG	60
ms017	243	**12**	9	232	TGCTGCTCGGCG	GCAACCGAGGACAACAATAA	GTAGACCGCGCTCCAGAACT	60
ms018	269	**15**	14	296	GCCCGCGCTGGAGCT	GTCGATGGACGAAATCCTTC	AAAGAAACGATCGGCTTCAA	51
ms020	311	**12**	11	263	TGCTGCTGGCGC	CATTCCGATGATCGTCCTCT	AATGACCGACGAACTGGATG	60

Nom de la répétition	Position dans le génome PAO1 (en kb)	L: taille du motif souche PAO1 (pb)	N: nombre de motifs souche PAO1	Taille du produit PCR PAO1 (pb)	Motif consensus	Amorce gauche	Amorce droite	T° hybridation des amorces (°C)
ms021	330	**11**	8	250	CGGATATAAAT	GTCATGGAGCACCTCGTGT	CCTACCTGGGGCTCTACCTG	60
ms022	376	**15**	8	274	GCCGAGCAGCGGCAG	GAAGCCGCTGAGGATCAC	GCCTGTACAACGGCCTCTAC	60
ms024	388	**24**	8	299	GGCCAGGGCCAGCACCAGCAGCCG	CTGAGCCACAGCCAGAGC	CCGCGCCGTATTCTCACC	60
ms026	435	**12**	8	200	GCCACGGCCTCG	AGGCTGGCGCAGGGGATG	AGCAACGCGCGAAGATCG	60
ms027	461	**11**	8	200	CGGCGGCCCGG	CTCGGTGCATTGCTCGAC	ATGTGCTGGGCATAGAGGAA	60
ms028	622	**15**	8	244	CAGCCCCAGCCAGCC	TCAAAAAGTGAGGCCGTACA	CGCCTGGAACCTGTACAACTA	60
ms029	622	**12**	10	253	CAGCAGCGCCGG	GTGTACTGGGCGAAACTCG	GCAGCTACCGTTTCGTCAC	60
ms030	658	**12**	8	203	CGGCGGCCTGCT	AGATGCTCTGGCCCGCTAC	CAGATAGCCGAGCAACTGAA	51
ms032	758	**18**	8	308	CTGGAGGCCGGCACCCGG	ACTACGACCTGCCGATCAAG	CAGGACATTGGGTCTGCTG	63,9
ms033	775	**10**	10	220	AGCAGCGCCG	AGGACGGCGAGGAAGACT	CCCTTACCGAGCTACACCTG	60
ms034	810	**18**	8	254	CGACGGCGCTCTGCACCAGGT	CGAGGTACAGGCCTTCGAC	GGAGGAGCTCCATGAAACAG	60
ms035	819	**15**	7	245	GCACGTCGGCCAGCG	CCAGCACTTGCTCGATCACT	CGTCACCTGGAGGCTGTACT	60
ms036	825	**12**	7	237	GGCGGCGCCGAG	CTGATCCACCTGGCCAAC	TCACCCTGGATCCACACC	60
ms037	880	**15**	8	242	GCTCGGCCTGCCGCT	CCTCGGTCGCACGCAACT	ATCAGGCCACCTCCTGGAC	60
ms038	1055	**24**	7	326	GCCAAGAAGAGCGCCGAGGACGAG	CGATGAAGCCAAGAAAGCTG	GGTCGTATCCGAAAGCAACT	60
ms039	1140	**10**	7	207	TGGCCGGCGC	GACCCTTGGCGAGGACTT	GTCTTGCCGTGCAGGCTCT	60
ms040	1160	**15**	9	302	CGCCGAACACGGCGG	GCTTGAGGGACTGGCTCATA	CATGCAGTACGGCGTGAT	60
ms042	1196	**11**	10	220	CGCTCGGCCGCG	CTTTTCCGGTAGCAAGGTGT	CTTGAGCGGGTTGATCGT	60
ms043	1349	**21**	7	275	GCGCTGGCCGGCACGCTCGGC	GCTGACATCGGTGCATGA	CGACCAGCTTCAGGTAATCC	57
ms044	1379	**13**	7	212	GCCGCGGCGGAC	AGTTGGCGAACCAGGGCAAG	CAGGCCGAGCAGGTCGAG	64,7
ms045	1428	**9**	8	185	CAGGACATG	CTCACCAAGTACGAGCACCA	TGCGCTTACTCCTGGGTACT	60
ms046	1457	**12**	9	205	GGCATCGCCAGC	TGAGCGGTTCGTCACTCGAC	CATCCGGCTGGCCGTTCC	60

Nom de la répétition	Position dans le génome PAO1 (en kb)	L: taille du motif souche PAO1 (pb)	N: nombre de motifs souche PAO1	Taille du produit PCR PAO1 (pb)	Motif consensus	Amorce gauche	Amorce droite	T° hybridation des amorces (°C)
ms048	1494	**15**	7	207	TCGGCCAGCTCGATG	GATGTTGTCGTGGCTGAACA	CAGCGGATCAACCTCTACG	60
ms049	1574	**10**	8	198	GCGCCGCCGG	GACATGCGTCGCTACCTGTT	AGGCGAGCAGGAAGATCAG	60
ms050	1579	**21**	8	306	GCTCGGCCTGCTGCTGACCCT	ACTGAGCCTGCTGATGCTG	ACGAGGAACCCGATGAGCTA	60
ms051	1593	**18**	8	234	GCCGCTGCTCGACGACCC	GTGCTCGAGGAGCAGGTC	GAACAGCAGGCACTCGAAG	65
ms052	1624	**12**	22	366	CAGGCCAGCAGC	GGCCAGGCGCAGGTACAG	ATCGGCGAGATCATCTTCCT	65
ms053	1637	**24**	7	262	GGCGATGCCCACGCCGAACATCCC	GTGGGAAGAAACGCACGAT	GAGGAAATCGCCTTCCTGAT	60
ms054	1688	**9**	14	246	GATGCAGTC	AGCCCTTTCGGGTAGTTCAT	ACAAGGACACCCTGATCGTC	60
ms055	1702	**15**	8	227	CGGCCAGGGGCAGGT	CTGTAGGCCTCGACCAGCTT	CTGACCGTCACGCAGATG	60
ms056	1706	**12**	14	271	CGAGCAGGCTGG	GTAGCTGTTGGCCTGGAAGC	CTACAAGCGCCTGGTCAAG	60
ms058	1737	**9**	7	170	GCCGGCGCC	AGCTGACGCTGGAGAAGAAC	CGGTTCGACTTGGACCAG	60
ms059	1812	**24**	8	313	GGCCGAAGGCGAGCGCCAGCGCCA	GTCCTCGACACCGCCTGT	GGCCCTGGCCAATTGCTG	60
ms060	1826	**12**	9	236	GCCAGGCGCCAGC	GAGGCAGAGCGACAGCAG	CGGGATGAAGTTGTCCGATA	60
ms062	1859	**12**	11	223	GGTGCTGGCGCC	ATCCCGAGCGACTCGAAC	CTGCGCCACATAGTCGTAAG	60
ms063	1878	**9**	12	242	GCTGGTGGC	GGCAATGGCTTCTTCTATGG	ACAGCAGGGTCGAACACAG	60
ms064	1919	**12**	8	206	GCCAAGCTGCTG	CTTCATCCCCAACCTGCTC	CGAACCAGTAGACGATCTTGC	60
ms065	1998	**12**	8	218	GCAGCCGACCGC	GCCCATGAACACCACTTTCT	GCCTGCTGGTGGAACTGG	60
ms066	2014	**12**	20	356	AGGCGCCGAGGC	GGCGAACATCAGCAGCAT	GTACGCACCTGGCTGAAAG	60
ms067	2026	**18**	7	246	GGCGCGCTGGTGCGACCT	CCAGGGCTTCTACGAACG	GCCTCGTCCACCAGTACC	60
ms069	2080	**12**	8	209	GTCGCAGAGGCC	GACCAGCACCAGCGGAGT	AAGACGCCATTGGAGAAGC	60
ms070	2149	**12**	7	202	CGAGGCGCTCGG	ATCAAGTCGCGCTTCGTC	GCACTTCGACGCTGACCT	60
ms071	2166	**15**	40	696	CAGGTCGGCGTCGGC	AATCCTTGCAACCCTGCAT	CGGAGTATCCCGATGAAGAC	60
ms072	2197	**12**	12	223	CCGGCGGCTTCG	GGTAGGTGCATTTGGCACTC	ACTCGATCCTCGACCTGGAC	60
ms073	2245	**12**	7	186	GAGGGCCGGCTG	GACGCCTCGGCGATTCTC	CAGCGGGTAGATGCCACT	61,3
ms075	2248	**14**	9	251	CAGCAGGCGCGCAGC	GACGCTGAGGATCACGATG	TCCTGGCCATCTACTTCCAG	60

Nom de la répétition	Position dans le génome PAO1 (en kb)	L: taille du motif souche PAO1 (pb)	N: nombre de motifs souche PAO1	Taille du produit PCR PAO1 (pb)	Motif consensus	Amorce gauche	Amorce droite	T° hybridation des amorces (°C)
ms076	2248	**12**	19	340	GCGCCGCCGAGC	GGGACGGAAAGGTAAAGTCC	CCTGGCGATCTGGAGCAT	60
ms078	2303	**14**	8	236	GGCCGGCGCGCTGCT	ACGGTGGAACTGGTGGTG	AGCAACAGGGCGAAGCTC	51
ms079	2309	**12**	19	375	CGGCTGCCTGCG	AGCTTGCCGACTATGACCTG	GTAGTCCACCGCGCAGAC	60
ms080	2371	**12**	10	211	CTGCAGCGCGGC	ATGAGCGAACCGATGTCC	GGTCGCTCTGCTCCAGTTC	60
ms081	2374	**19**	7	281	CGGCGGCCAGCAGGTAGC	GCCATCAGCGACCAGTTG	CTCACCTTCAGCCTGGTGAC	60
ms082	2475	**13**	10	254	GCCACCGCGGGGC	CCGGAGAAACCGAAGGAAG	GGTCAGGTAGGCATCGAGGT	60
ms083	2492	**18**	7	282	CCGGCCTGCGCCTGGCG	GGAGCATCCGATGAAGAAAG	CGGTCGAGGTAGTCGATGTC	60
ms085	2533	**11**	12	248	CGGCGGCGCGG	GATACTGGATCGCCAGCAG	ACTGAGCCTGGGGCTGTT	60
ms086	2544	**12**	7	188	GCGCCAGGCCGC	CGAACAGATGGTGGAGTACG	CGATCCCCAGTACGACGAAT	60
ms087	2546	**10**	8	187	CGCGGCTCGC	GGATATCGGCGAAGTCGAG	ACCGCATCCTGCTGCTCT	60
ms089	2552	**27**	12	428	CCGAGCAGCCAGCGCCAGCCGGCAG	CGACGATCACCAGCAATACC	TGCTGATCTTCTTCGTCTGC	61,3
ms090	2553	**12**	13	269	GCCGCCAGCGCT	GAATGCCGACGCAGTCCT	CTCCGTGCCCTTGCTCTAC	60
ms091	2559	**9**	8	181	GGCGACGCG	GGCCACGTAGCCGAGTTC	GAGCATCACCAAGAACGACA	51
ms092	2566	**12**	11	235	CGCCATCGACGC	CGTCATCCTGGACAACGTG	GACCTGCTGGATGGTGTAGG	60
ms093	2587	**10**	9	182	CGCCGCGCAG	CATGCACGGATTGTTCCTC	ATGGCTTCCAGCCAGTCC	59,3
ms094	2597	**12**	7	199	GCCTGCCCGGCC	GTGGTGGTCCTCGAATGC	GGTGTACTGGAGATCGAGCAG	60
ms095	2625	**24**	15	470	GGCGGCAACCTGACCATCAAGGCC	ACCGTCACCCTGGAGAAAG	CGCTTCTCCTCAGTTGACCT	60
ms096	2630	**15**	8	223	GCGCCAGCGCCATGC	CCACCCCTGCAACTGATT	CAAGGTCCAGCAGGACAACT	60
ms097	2635	**21**	7	273	GCTGGGCGAGGATCGCCGCGA	CGAAGGACTCGACAGGAGAA	GGTCTCGAGGGGATGCTC	60
ms098	2727	**12**	8	192	CAGCGCCGGGCG	GACTGGATCGCATCGTTGA	GGTACCCTCCGAGGTGCT	60
ms099	2735	**12**	16	313	CGGCGGCCTGCG	GTCCGAGTTGGCGGTAGTAG	AGCGGCGTCGAGGTACTG	60
ms100	2755	**12**	7	192	CGAGTCCGGCGC	TGGTGGAGTTCCATCCTTTC	GGAACTCGTAGCTGGGATGA	60
ms101	2758	**12**	9	238	CGGCAGCACCGC	ACCTGACCCAGGTGACCA	TCGTAGCTCAGCCAACTGC	60

Nom de la répétition	Position dans le génome PAO1 (en kb)	L: taille du motif souche PAO1 (pb)	N: nombre de motifs souche PAO1	Taille du produit PCR PAO1 (pb)	Motif consensus	Amorce gauche	Amorce droite	T° hybridation des amorces (°C)
ms103	2790	**11**	21	381	CTGCGGCGCGC	GATGGTTACCGCGATGTTCT	AGGTGCCGTCGGACTCGT	60
ms104	2818	**12**	9	232	TGCTGGCCGCGC	CCTGGTTGCCGGTGTACTA	ATCAACGCCACCCACAGT	60
ms105	2848	**9**	8	241	CGATCAGCG	ATCATGATCGGCCGGAAG	ACCCGCTGACCATCCTCT	60
ms106	2866	**21**	8	311	GCAGGCCCAGCACGGCGGCCA	GGTCCAGGTCTTGGGTCTG	GAGCCATCGGCGGAACTG	60
ms107	2866	**22**	13	409	GCGCCAGCGCCAGCGCCGCCA	AGAGCCAGGCGTTGATCC	CTGCTGCATGGCTGGTTC	60
ms108	2886	**12**	7	207	GCGGCGAGCACC	ACATCTTGTTCAGCGTCGAG	GCGACTTCAAGGACGTCAAC	60
ms109	2887	**15**	7	234	GCGGCCTCGGCCTG	ATTGCGTCCGAGGACCTT	GCGAAGATGCTGGTGGAG	60
ms110	2890	**12**	11	233	CCCTGCTCGCCG	GTAGGGTTTGCCGAGCAG	CGCGATTGAGAGGATGGAG	60
ms111	2979	**15**	8	214	GCGCCGCTCCCGCCA	GAACAACCCGCACATGAAC	GGCGTTTGAAAGGCAAGC	60
ms112	3011	**15**	8	232	CCATGCCAGCAGCG	CGAGATTGAGCAGAGCGAAA	CTGCTCTGGTCTCTGCACCT	60
ms113	3011	**12**	7	198	GCAGCGCCAGCA	CGGGTAATTGCGCTTCTG	CAGACCTGGACCGGCGTA	60
ms114	3055	**12**	26	434	CGCGCTGCTCGG	CTCAACGCCCTCTCCTTCTT	AAATAGAGGTTGCGCAGCAG	60
ms115	3143	**12**	7	223	CCCGCCGCCTGC	GGAAACCGTCTCATGACCTAA	GAGCGTGTCGAGGTAGGC	60
ms116	3170	**12**	21	384	GCAGCAGGAACA	AGGCCATCATGCTCAAGCTA	GCTGCTGGTCCAGTTGCT	60
ms117	3189	**15**	7	207	GCCCGCGTCGACCAG	GCCGGCAGCCTGCAGAAC	GCGGAGTTCGCAGTAGGAGA	60
ms118	3192	**12**	9	219	GCCGCCTGCGGC	CGAAACGCTGCTCGATCT	ATCTTGGCGTCCAGGTAGC	60
ms119	3194	**9**	12	216	CGCTCGCCG	CAGGTTCTTCGCCGCAAC	AGGCGCTGGACCTGGAGT	60
ms120	3210	**15**	40	698	CCTGCTCGCCTGCGC	GCGAAACCCTGCTCCTCTAC	CAACAACTGGCCGAGTCC	60
ms121	3212	**18**	9	306	CGGCCTGCTAGGCGGCGC	CTCTTCGCCGGTCTCCAG	CAGCGGGCTGCTGTTGAG	60
ms122	3268	**12**	11	238	CCGCCAGCCGGT	CACAAGGGCCTGTTCGAC	GTTCGTCGAGGATCAGCAG	60
ms123	3270	**12**	20	361	CTGCGTCGGCGC	GAGCACATCGTCTGGCTGAT	CGAGGAACAGCAGCAGGT	60
ms124	3292	**12**	7	196	CCGCGGCGAGCA	ATACCGAAGCCGATCGAAGT	GGTTGTTCCTGGTGATCCTC	60
ms125	3341	**9**	59	639	GCCAGCCAGC	ACAATGCCGGCAGTAGCA	CCTGGAGCTACGGGTTGAT	60
ms126	3342	**17**	15	376	GCGCCAGCTCAGGCGCGCC	GTCCTGGAACATTCGCCACT	CTTCCGGCTGGAGGAGGT	60

Nom de la répétition	Position dans le génome PAO1 (en kb)	L: taille du motif souche PAO1 (pb)	N: nombre de motifs souche PAO1	Taille du produit PCR PAO1 (pb)	Motif consensus	Amorce gauche	Amorce droite	T° hybridation des amorces (°C)
ms128	3496	**21**	11	358	CGGCAGGCGCCGGAGGCTTCG	GCTGCTCTGCCGCCTGTA	AGTTGGCCAAGCTGCAAG	60
ms129	3497	**10**	8	224	GCGGGCGCTT	GAAACGGTGCGATAGGTGTC	CCCTACCTGAACTTCCTGGTC	60
ms130	3569	**15**	9	272	GCCGCCGGCTTGCGG	GCGACCATTAGACACCGATT	AGCAGGGCACTCCGGTTG	60
ms131	3589	**12**	8	214	CCTCCGCGACGG	TGAATTCTACAGGGTGGAACG	GAGCAGGGCGACGAGTCC	60
ms132	3609	**12**	10	239	CGCCGCGCCAGC	GGCCAACTGGTGAACTGG	AGGAACAGGCTGTGGAAGC	60
ms133	3675	**12**	13	243	GGCCGGCGGCTC	ATAGGCGAATGGCGACAG	CTGGAGCTGGCTCACTACG	60
ms134	3701	**12**	10	245	CTGGGCGCGGCG	GGTGCAGAGCTTCCTGCT	GTGCAGGGGGACTGAACG	60
ms135	3717	**13**	11	228	CGCGCGACCCCGC	CCTCTACGCCAACGAACAAC	CTGAACAGGCCGGGGATA	60
ms136	3723	**12**	12	291	CGCCCGAGCAGA	ATCGCCACCGTCAGGTAG	TGGTACTGATGGCAGTTTCG	60
ms139	3844	**11**	8	191	GCCGCCAGCGC	ACCCCGTTGATGACCTACTG	AATCGGATTGATGCAGGGTA	60
ms141	3864	**9**	10	215	CTCGGCGGC	CGGCTTCCTGATCTACCTGA	CTACCCGCCGTAGAGCTTC	60
ms143	3910	**12**	10	307	GCCGCCGCCGAG	AAGAAGGCCGAGAAACAGTTG	CTTCGCAACGCTCGACCT	60
ms145	3956	**9**	8	177	CGGCGGCAG	TCCGCCGTCGAAGTCAAT	GAAAGCTCCACCGCGTAG	60
ms146	3975	**11**	10	215	GTGCCCGCCGC	CTACAACCAGCCATTGCAGA	CCAGCTCGGCTGGTAGAC	60
ms147	4013	**12**	8	221	CGGCCTGCTGAT	TTCCTCGTCGAAGTGGTGAT	CCAGGTAGGTCATCAGCTTG	60
ms148	4036	**18**	7	243	TTCGACCTGCTCGGCCTG	GAACCGAAGCCTGAAGATTG	TCCACCAGATAGCGGATGTT	60
ms149	4106	**12**	11	232	CGGCCTGCTCGC	CTGGACGAAGCCGAAGTC	GTACCTCGAGGCGTTGCAG	60
ms150	4157	**12**	7	212	GCCGCTGGCGCC	GTAGGCCTTGTAGGCGATCA	GGCAGGAAATGTTGCAGAAC	60
ms151	4199	**9**	18	297	CTGCGCCTG	ACCCTGGTCAACGTCAGC	GTCCGCTGTCGTCCTGTT	60
ms152	4201	**9**	8	218	CCGCCAGTC	TAATCGACCCCAGCTTATCC	GGATGGTTATGGCCAGTGC	60
ms153	4221	**22**	9	316	GCCTGCGGCGCGCCTCCACGACCA	CGCATTCATGAAATGGTAGG	GCTGCTGGTATTCGACAATG	60
ms155	4495	**69**	11	876	GCCGACGGCGCCCGCTACCACGGCGGCTTCAGCAGGGCCTGCTGCACGGCCAGGGCCAGCTGGACGGC	CAGGTCTCTCTCGCTCTGCT	CCCAGAGTCCCTGCTCAC	60

Nom de la répétition	Position dans le génome PAO1 (en kb)	L: taille du motif souche PAO1 (pb)	N: nombre de motifs souche PAO1	Taille du produit PCR PAO1 (pb)	Motif consensus	Amorce gauche	Amorce droite	T° hybridation des amorces (°C)
ms157	4620	**9**	13	237	CGACTGCAT	GACAAGGACACCCTGATCGT	GCGTTCATAGCCCATCTTGT	60
ms158	4688	**12**	8	176	CGCCCTGGCCGA	CCATCCACAACGTGCATC	AGACGATCGCCTGGTTCAC	60
ms159	4728	**11**	19	347	GCCGGCGCAGC	GAACAGCCCGAGGATCAC	GTTGGTCGAGGCCTGGTT	60
ms161	4730	**9**	17	247	CGGCGGCGC	GTCACGAGGCTTGCTCCAG	AGCGCCCAGCTCTGGTAT	60
ms163	4740	**18**	13	359	GGCCAGCAGGTCGAGCCA	ACGGAAGACAGCTCGAACG	GACCTCAGATGCTTGCCCTA	60
ms164	4746	**12**	8	211	CTGCTGCGGGGC	TCGAAGCTGTCGTAGTCCTG	GATCGCTACACCACCGAGAT	60
ms165	4800	**12**	10	219	AGGCGCGGCGCC	CCGGTGATCAGGCTGAAATA	CGTCTCTACCTGCGTCGTTA	60
ms166	4808	**12**	13	281	CGCGGCAGCGCG	ATCGAGGAAATCCCGAACAC	GACCTGTTCGACCAGTACCAG	60
ms167	4811	**12**	9	198	TCGGCCTGGCCC	GCCCTCTGCCAGTTCCTT	CTTCAGGCCATGCAGCAC	60
ms168	4879	**21**	8	263	CGCCTGCTGCGCGCTGCGCA	GACGTGTTCATCCTGCTCAA	GCTGACCCTGACCTCGAA	60
ms169	4898	**9**	10	199	GCGCCCGCC	GAATCCGCCATGACTCGTAT	AATTCTTGATCACCGCCTTG	53,3
ms171	5045	**9**	10	204	CGCCGAACC	GACGACGTCCTGGTGATGTA	CTTGTGGAAGGCGCAACC	60
ms174	5282	**15**	9	240	CGGCGCCGACTTCCG	GAAGGCTTCAACCGACAGAA	CGCAACGGGAAATACGTACC	60
ms175	5394	**12**	7	246	GCCGAGCGCCTC	CCCGCGCAGGGCTTCCTC	ACCACCCTGCGCCTGCTC	60
ms176	5411	**12**	8	226	GCCCTGCCCTG	GCCTGTACACCCTGATGGTC	CCGAAGGTTGCATCAGTTG	60
ms177	5440	**15**	9	226	CGGCTCGATCCGCGG	CATGTTCGGCTTCTTCGAG	CCGCTGGCGCTGATATAGA	60
ms178	5444	**9**	16	225	GCTGCCGCT	AACGGTGTACTCCTGCTGCT	ATTGGCCAGGTGGCAGGTA	60
ms179	5460	**15**	9	272	CGACGCCAGCGGCCG	AGAAGGAGCAATGAGGACCA	CCGATGTACAGGGTCAGGTC	60
ms180	5463	**14**	7	226	CGGCGCGCTGCTGG	AGATCGACACCGACCTGCT	GAGGGGACAAGTCCATCTCA	60
ms181	5468	**18**	17	511	CGCTGCTCGGCGCCTGG	CAACTGCACGGCCTGTATC	GCGACAGTCTGGTTCAGGAT	60
ms182	5485	**24**	7	314	GTCGCGCACGCAGCGCTGCGCAG	GTTGTTCCGGGGTCAGTTC	GTCGCAGCGGGAAGTCTC	60
ms183	5489	**18**	7	251	CTGCTCGCCGCCCTCGCC	ACTTCGCAACGACAAGGAAC	CATGGTGCCGACGAAGGT	60
ms184	5489	**12**	10	241	CTCGCCGTGGCC	CACCTTCGTCGGCACCAT	GAAACCGAGCGCGTACAG	60
ms185	5584	**12**	8	206	GCCGACCACGCC	ATCTTTTCCAGGACGAACGA	CTGGGGCACCACTTCCTT	60

Nom de la répétition	Position dans le génome PAO1 (en kb)	L: taille du motif souche PAO1 (pb)	N: nombre de motifs souche PAO1	Taille du produit PCR PAO1 (pb)	Motif consensus	Amorce gauche	Amorce droite	T° hybridation des amorces (°C)
ms186	5588	**12**	13	262	CGCCTGCAGCGG	GGTATAGGCCAGTTGGGTCA	GCGCAAGTTCAACGACTTCT	53,3
ms187	5602	**12**	11	264	CGGCCAGCGCGG	CTATCCGATCTCCTGCATCC	ATCTCCCGTTGCAGGTTG	60
ms188	5635	**10**	8	189	GCCAGGTCGC	CCGAACTGCTCGCACTGG	CGCCGACGAGGACGAACT	60
ms190	5758	**18**	27	588	GCGCCGACCAGCCGCCAG	CAGGCCATAAGGCACGAAG	CAGGGTGGGGTATTCGACTC	60
ms191	5791	**15**	10	275	CGCCGGCGGCTGGT	GCCTGGACCTGGCGAAGAT	CAGCGGCTTGCCGGCTTC	60
ms192	5836	**12**	21	399	GGCGGCGCCATC	CTCTCCTGGTAATCGCGTTC	ATCAACGATGCTTGTGCTTG	60
ms193	5873	**12**	10	273	AGGCCGGCAGCG	CTCGCTGGTCATCTCGAAG	GCCTACCTGATCCGTTTCAT	60
ms195	5960	**11**	12	249	CGGCCTGGCCGG	ATCCGCGGTCTCAAGGAC	AGCCGCTCGGCAATTGAG	61,3
ms196	5990	**18**	7	260	CCTGGTCAGCGCCCGCGA	CAACCACAGTTCCGGAAAGA	CGCTGGTAGGGTCTTTCTTC	53,3
ms197	6020	**9**	36	441	CGCAGCCGC	ACCAGGCGTCGATATTGC	CCGGAGTCGAGTACCACAAG	60
ms198	6113	**18**	14	371	CCGCGCGCGCTGCGCAGG	CGTTCTCGAGCAACACCTG	CACCTACGTTGCCGAGTTG	60
ms199	6130	**15**	7	185	CGCCGAGCAGCCAGC	CTATCGGCCTGCTGGAAG	CGTACCCGGCTCCTTTATCT	60
ms200	6155	**15**	8	208	GCCGCGCTGATGCT	CGTATTGCGCTTCAACACCT	GAGGCAGGCGAGGATCAG	60
ms201	6225	**9**	13	240	GCTGGGCGG	TTGATGTCGCTGTCGTTGAG	GAAACCCAAGCCCAAACC	60

ANNEXE 4:

122 répétitions en tandem polymorphes dans au moins deux des six génomes *S. aureus* comparés (tableau fourni par la page de comparaison de la base de données du laboratoire)

nombre d'allèles dans les 6 souches	Mu50 Position	L: Longueur totale du minisatellite (pb)	U : taille du motif (pb)	N : nombre de copies	séquence détectée par le TRF	N315 Position	L:	U:	N:	TRF	Position MW2	L:	U:	N:	TRF	Position MRSA252	L:	U:	N:	TRF	Position NCTC8325	L:	U:	N:	TRF	Position MSSA476	L:	U:	N:	TRF
5	1098012--1098170	**159**	18	9	**oui**	1021633--1021791	**159**	18	9	**oui**	1023399--1023512	**114**	9	9,66	**oui**	1064018--1064221	**204**	9	19,7	**oui**	959425--959565	**141**	9	13	**oui**	1052122--1052253	**132**	9	11,2	**oui**
5	1866118--1866562	**445**	159	3	**oui**	1788113--1788557	**445**	159	3	**oui**	1825797--1826247	**451**	159	3,24	**oui**	1897230--1897521	**292**	159	2,1	**oui**	1760294--1760897	**604**	159	4	**oui**	1805450--1805582	**133**	?	?	non
5	2511193--2511367	**175**	61	2,9	**oui**	2440643--2440817	**175**	61	2,9	**oui**	?	**53**	?	?	non	2543631--2545649	**2019**	?	?	non	?	**261**	?	?	non	?	**114**	?	?	non
5	636061--636666	**606**	18	33,3	**oui**	611789--612394	**606**	18	33	**oui**	602378--602869	**492**	18	27,3	**oui**	622352--622777	**426**	18	22,8	**oui**	554474--554971	**498**	18	27	**oui**	601015--601560	**546**	18	30,3	**oui**
5	704561--704796	**236**	67	3,5	**oui**	680287--680522	**236**	67	3,5	**oui**	675398--675899	**502**	133	3,78	**oui**	687428--687596	**169**	?	?	non	623744--624378	**635**	133	5	**oui**	674030--674398	**369**	133	2,78	**oui**
5	888858--889722	**865**	18	48	**oui**	850133--851159	**1027**	18	61	**oui**	858671--859571	**901**	18	52,5	**oui**	?	**?**	?	?	non	794658--795498	**841**	18	50	**oui**	842718--843564	**847**	18	49,5	**oui**
4	1000868--1000965	**98**	?	?	non	?	**?**	?	?	non	924472--924622	**151**	58	2,6	**oui**	956430--956560	**131**	56	2,3	**oui**	860386--860611	**226**	56	4	**oui**	908441--908591	**151**	58	2,6	**oui**
4	1291998--1292219	**222**	64	4	**oui**	1215585--1215806	**222**	64	4	**oui**	1215471--1215628	**158**	64	3,04	**oui**	1257576--1257668	**93**	20	2,5	**oui**	1151427--1151456	**30**	20	2,5	**oui**	?	**?**	20	2,5	**oui**
4	1729388--1729679	**292**	56	5,1	**oui**	1652885--1653176	**292**	56	5,1	**oui**	1698498--1698675	**178**	59	3	**oui**	1763381--1763661	**281**	54	5,1	**oui**	1631653--1631711	**59**	?	?	non	1678150--1678327	**178**	59	3	**oui**
4	1757525--1757639	**115**	?	?	non	?	**115**	?	?	non	1726650--1726890	**241**	126	1,9	**oui**	?	**172**	?	?	non	?	**46**	?	?	non	1706302--1706542	**241**	126	1,9	**oui**
4	2152442--2152714	**273**	6	48	**oui**	2075919--2076191	**273**	6	48	**oui**	2104534--2104818	**285**	39	7,84	**oui**	2180593--2180829	**237**	39	7,02	**oui**	2090361--2090645	**285**	39	8	**oui**	2083776--2083943	**168**	39	4,84	**oui**
4	2351355--2351474	**120**	24	5	**oui**	2281762--2281881	**120**	24	5	**oui**	2300573--2300643	**71**	24	2,8	**oui**	2383917--2384011	**95**	24	3,7	**oui**	2293971--2294048	**78**	24	3,2	**oui**	2279678--2279748	**71**	24	2,8	**oui**
4	2561664--2561795	**132**	56	2,4	**oui**	2490953--2491084	**132**	56	2,4	**oui**	2504456--2504567	**112**	?	?	non	2596190--2596265	**76**	?	?	non	2503207--2503262	**56**	?	?	non	2483624--2483735	**112**	?	?	non
4	2642053--2642330	**278**	42	7	**oui**	2570943--2571220	**278**	42	7	**oui**	2580884--2581077	**194**	42	4,61	**oui**	2661066--2661217	**152**	?	?	non	2582156--2582361	**206**	42	4,7	**oui**	2560243--2560436	**194**	42	4,61	**oui**
4	631615--632142	**528**	18	30,3	**oui**	607343--607870	**528**	18	30	**oui**	597926--598459	**534**	18	29,3	**oui**	?	**?**	?	?	non	549902--550555	**654**	18	36	**oui**	596557--597096	**540**	18	29,3	**oui**
4	965164--965428	**265**	43	6	**oui**	888833--889097	**265**	43	6	**oui**	890502--890552	**51**	?	?	non	922321--922500	**180**	43	4,18	**oui**	826375--826467	**93**	42	2,2	**oui**	874471--874521	**51**	?	?	non
3	1086975--1087029	**55**	22	2,5	**oui**	1010596--1010650	**55**	22	2,5	**oui**	1010526--1010580	**55**	22	2,5	**oui**	1051143--1053033	**1891**	?	?	non	946612--948442	**1831**	22	2,5	**oui**	?	**?**	?	?	non
3	1105143--1105186	**44**	18	2,4	**oui**	1028774--1028817	**44**	18	2,4	**oui**	1030495--1030538	**44**	18	2,4	**oui**	1071205--1071257	**53**	9	5,2	**oui**	966548--966591	**44**	18	2,4	**oui**	1059236--1059261	**26**	?	?	non

nombre d'allèles dans les 6 souches	Mu50 Position	L: Longueur totale du minisatellite (pb)	U : taille du motif (pb)	N : nombre de copies	séquence détectée par le TRF	N315 Position	L:	U:	N:	TRF	Position MW2	L:	U:	N:	TRF	Position MRSA252	L:	U:	N:	TRF	Position NCTC8325	L:	U:	N:	TRF	Position MSSA476	L:	U:	N:	TRF
3	1116357--1116488	132	?	?	non	?	?	?	?	non	1041734--1041923	190	58	3,3	oui	1082254--1082498	245	58	4,3	oui	?	?	?	?	non	1070457--1070646	190	58	3,3	oui
3	1132682--1133067	386	63	6,1	oui	1056313--1056698	386	63	6,1	oui	?	71	9	9,11	oui	1098554--1098687	134	63	2,1	oui	994051--994184	134	63	2,1	oui	1086698--1086831	134	63	2,1	oui
3	1194184--1194530	347	134	2,7	oui	1117818--1118164	347	134	2,7	oui	1119272--1119421	150	?	?	non	1160092--1160176	85	?	?	non	1055411--1055496	86	?	?	non	1148058--1148207	150	?	?	non
3	1219058--1219224	167	55	3	oui	1142646--1142812	167	55	3	oui	1142369--1142602	234	?	?	non	?	?	?	?	non	1078520--1078640	121	?	?	non	1171155--1171275	121	?	?	non
3	122905--123156	252	24	10,5	oui	122917--123168	252	24	11	oui	99820--100023	204	24	8,5	oui	124119--124394	276	24	11,5	oui	73719--73994	276	24	11	oui	98526--98729	204	24	8,5	oui
3	123159--123840	682	174	3,9	oui	123171--123852	682	174	3,9	oui	100026--100881	856	174	4,9	oui	124397--125252	856	174	4,9	oui	74021--74852	832	174	4,9	oui	98732--99587	856	174	4,9	oui
3	1425109--1425340	232	58	4,1	oui	1348687--1348918	232	58	4,1	oui	1350816--1350933	118	?	?	non	?	?	?	?	non	1286639--1286814	176	58	4	oui	1379312--1379487	176	58	3,63	oui
3	1695062--1695135	74	?	?	non	?	74	?	?	non	1664175--1664305	131	57	2,3	oui	?	16	?	?	non	?	73	?	?	non	?	74	?	?	non
3	1985673--1985771	99	?	?	non	?	?	?	?	non	1949784--1949940	157	60	2,61	oui	?	?	58	2,2	oui	1890158--1890518	361	117	3,1	oui	1929118--1929274	157	60	2,61	oui
3	2039328--2039458	131	56	2,3	oui	1961441--1961571	131	56	2,3	oui	2002095--2002111	17	?	?	non	2078902--2079089	188	?	?	non	1987680--1987811	132	58	2,3	oui	1981342--1981358	17	?	?	non
3	2185031--2185605	575	256	2,2	oui	2108506--2109080	575	256	2,2	oui	2136921--2137239	319	?	?	non	?	?	?	?	non	2122742--2122804	63	?	?	non	?	?	?	?	non
3	2294935--2295172	238	9	29	oui	2225122--2225359	238	9	29	oui	2245083--2245380	298	9	33,1	oui	2324595--2324892	298	9	33,3	oui	2232991--2233279	289	18	7,6	oui	2224188--2224485	298	9	33,1	oui
3	2546242--2546492	251	123	3	oui	2475692--2475751	60	?	?	non	2489123--2489250	128	69	2,4	oui	2580875--2581002	128	69	2,4	oui	?	?	?	?	non	2468291--2468418	128	69	2,4	oui
3	2611979--2612851	873	?	?	non	?	873	?	?	non	2554557--2555866	1310	437	3	oui	?	1982	?	?	non	?	873	?	?	non	2533849--2535158	1310	437	3	oui
3	2630754--2632463	1710	384	4,5	oui	2560028--2561353	1326	384	3,5	oui	?	?	?	?	non	?	?	?	?	non	2571285--2574147	2863	384	7,5	oui	?	?	?	?	non
3	2638502--2638675	174	42	5	oui	2567392--2567565	174	42	5	oui	2577387--2577518	132	42	3,3	oui	2661081--2661212	132	42	4,8	oui	2578667--2578810	144	42	3,3	oui	2556704--5556877	174	42	4,3	oui
3	2642422--2642721	300	114	2,6	oui	2571312--2571611	300	114	2,6	oui	2581169--2581468	300	114	2,6	oui	?	186	?	?	non	2578822--2579180	359	114	3,1	oui	2560528--2560827	300	114	2,6	oui
3	266128--266583	456	81	5,6	oui	266141--266596	456	81	5,6	oui	244215--244589	375	81	4,6	oui	259889--260182	294	81	3,7	oui	213234--213608	375	81	5	oui	242915--243289	375	81	4,6	oui
3	2781740--2782399	660	18	42	oui	2717342--2718001	660	18	42	oui	2720626--2721375	750	18	47,9	oui	?	?	?	?	non	2723734--2724393	660	18	42	oui	2699972--2700715	744	18	47,6	oui
3	2818769--2819116	348	42	17	oui	2754371--2754718	348	42	17	oui	2757747--2758106	360	18	18,8	oui	2839324--2839445	122	42	2,9	oui	2760766--2761113	348	42	17	oui	2737087--2737446	360	18	18,8	oui
3	2837314--2837372	59	?	?	non	?	?	?	?	non	2776867--2777037	171	56	3,1	oui	2855668--2855783	116	56	2,1	oui	2779312--2779482	171	56	3,1	oui	2756207--2756377	171	56	3,1	oui
3	311490--311657	168	55	3,1	oui	311503--311670	168	55	3,1	oui	290110--290334	225	56	4,1	oui	309197--309366	170	56	3,1	oui	258535--258646	112	54	3	oui	288810--288979	170	56	3,1	oui

nombre d'allèles dans les 6 souches	Mu50 Position	L: Longueur totale du minisatellite (pb)	U : taille du motif (pb)	N : nombre de copies	séquence détectée par le TRF	N315 Position	L:	U:	N:	TRF	Position MW2	L:	U:	N:	TRF	Position MRSA252	L:	U:	N:	TRF	Position NCTC8325	L:	U:	N:	TRF	Position MSSA476	L:	U:	N:	TRF
3	43142--43471	330	40	8,3	oui	43137--43506	370	40	9,3	oui	37788--38077	290	40	7,3	oui	43105--43394	290	40	7,3	oui	?	?	?	?	non	?	?	?	?	non
3	465115--465161	47	?	?	non	?	?	?	?	non	427477--427525	49	8	6,1	oui	?	37	?	?	non	?	37	?	?	non	426122--426186	65	8	8,1	oui
3	529716--529869	154	?	?	non	?	154	?	?	non	?	147	?	?	non	513361--513642	282	133	2,1	oui	?	147	?	?	non	?	147	?	?	non
3	535447--536092	646	277	2,3	oui	511215--511860	646	277	2,3	oui	496440--497086	647	277	2,33	oui	519278--521105	1828	18	2	oui	453788--454412	625	257	2	oui	495102--495748	647	277	2,33	oui
3	550991--551087	97	21	4,6	oui	526760--526814	55	21	2,6	oui	511986--512019	34	?	?	non	536007--536039	33	?	?	non	469312--469345	34	?	?	non	510648--510681	34	?	?	non
3	632181--632224	44	21	2,1	oui	607909--607952	44	21	2,1	oui	598270--598541	272	21	2,1	oui	?	?	?	?	non	550570--550637	68	21	2,1	oui	?	?	?	?	non
3	640048--640484	437	18	24,3	oui	615776--616212	437	18	24	oui	606251--606687	437	18	24,3	oui	?	413	18	22,4	oui	554474--554970	497	?	?	non	604942--605378	437	18	24,3	oui
3	683273--683427	155	?	?	non	?	155	?	?	non	?	275	?	?	non	665553--665843	291	72	4,1	oui	?	273	?	?	non	?	275	?	?	non
3	684180--684299	120	?	?	non	?	?	?	?	non	656481--656659	179	61	2,9	oui	?	?	?	?	non	604760--605006	247	61	2,7	oui	655173--655291	119	62	1,9	oui
3	748508--748564	57	?	?	non	?	?	?	?	non	719614--719947	334	69	4,8	oui	?	?	?	?	non	?	?	?	?	non	718112--718375	264	69	4,68	oui
3	847717--848067	351	59	6	oui	823499--823849	351	59	6	oui	819191--819361	171	58	3,1	oui	862219--862392	174	58	3,1	oui	768247--768646	400	58	7	oui	817619--817789	171	58	3,1	oui
3	855096--855387	292	56	5,2	oui	830878--831169	292	56	5,2	oui	826400--826409	10	?	?	non	?	?	?	?	non	775684--775692	9	?	?	non	824828--824893	66	56	2	oui
2	101929--101971	43	20	2,1	oui	101941--101983	43	20	2,1	oui	80524--80566	43	20	2,1	oui	108840--108908	69	20	2,1	oui	53573--53615	43	20	2,1	oui	79231--79273	43	20	2,1	oui
2	1105876--1105912	37	?	?	non	?	?	?	?	non	1031228--1031288	61	24	4,37	oui	1071946--1072004	59	21	2,7	oui	967281--967341	61	24	4	oui	1059951--1060011	61	24	4,37	oui
2	1107863--1107903	41	18	2,2	oui	1031494--1031534	41	18	2,2	oui	1033239--1033279	41	18	2,3	oui	1073953--1074002	50	?	?	non	969292--969332	41	18	2,2	oui	1061962--1062002	41	18	2,3	oui
2	1139138--1139258	121	?	?	non	?	?	?	?	non	?	?	?	?	non	1104758--1104879	122	14	8,4	oui	1000232--1000293	62	17	3,4	oui	?	?	?	?	non
2	1175797--1175833	37	?	?	non	?	?	?	?	non	?	?	?	?	non	1141665--1141708	44	7	6,6	oui	?	?	?	?	non	?	?	?	?	non
2	1182603--1182659	57	9	6	oui	1106237--1106293	57	9	6	oui	1107690--1107746	57	9	6,3	oui	1148478--1148555	78	18	5	oui	1043829--1043885	57	9	6,3	oui	1136476--1136532	57	9	6,3	oui
2	1183288--1183311	24	?	?	non	?	?	?	?	non	?	?	?	?	non	1149183--1149218	36	18	2	oui	?	?	?	?	non	?	?	?	?	non
2	1183549--1183618	70	15	4,3	oui	1107183--1107252	70	15	4,3	oui	1108636--1108705	70	15	4,3	oui	1149435--1149525	91	15	5,7	oui	1044776--1044845	70	15	4,3	oui	1137422--1137491	70	15	4,3	oui
2	1213418--1213706	289	56	5,1	oui	1137053--1137341	289	56	5,1	oui	1138300--1138587	288	56	5	oui	?	?	?	?	non	1074391--1074735	345	113	3	oui	1167086--1167373	288	56	5	oui
2	1326853--1326899	47	?	?	non	?	?	?	?	non	?	?	?	?	non	?	?	?	?	non	1186083--1186154	72	25	2,9	oui	?	?	?	?	non

nombre d'allèles dans les 6 souches	Mu50 Position	L: Longueur totale du minisatellite (pb)	U : taille du motif (pb)	N : nombre de copies	séquence détectée par le TRF	N315 Position	L:	U:	N:	TRF	Position MW2	L:	U:	N:	TRF	Position MRSA252	L:	U:	N:	TRF	Position NCTC8325	L:	U:	N:	TRF	Position MSSA476	L:	U:	N:	TRF
2	1385013--1385033	21	?	?	non	?	?	?	?	non	1308759--1308790	32	11	2,9	oui	?	?	?	?	non	?	?	?	?	non	1337254--1337285	32	11	2,9	oui
2	149437--149526	90	18	8	oui	149449--149538	90	18	8	oui	126936--127025	90	18	8,27	oui	150846--150875	30	18	4,7	oui	100907--100994	88	18	8	oui	125642--125731	90	18	8,27	oui
2	1513544--1513661	118	?	?	non	?	?	?	?	non	1440617--1440792	176	58	3,1	oui	?	?	?	?	non	1375167--1375342	176	58	3,1	oui	1469172--1469347	176	58	3,1	oui
2	1516384--1517097	714	231	3	oui	1439879--1440592	714	231	3	oui	1443514--1443996	483	231	2,09	oui	1504861--1505574	714	231	3,22	oui	1378065--1378778	714	231	3	oui	1472069--1472551	483	231	2,09	oui
2	1649052--1649072	21	?	?	non	?	?	?	?	non	?	?	?	?	non	1683319--1683360	42	21	2	oui	?	?	?	?	non	?	?	?	?	non
2	1654721--1654766	46	?	?	non	?	?	?	?	non	1623788--1623878	91	45	2,02	oui	?	?	?	?	non	?	?	?	?	non	?	?	?	?	non
2	167280--167319	40	18	2,1	oui	167292--167331	40	18	2,1	oui	145525--145564	40	18	2,1	oui	168564--168603	40	18	2,1	oui	119397--119436	40	18	2,1	oui	144231--144263	33	?	?	non
2	1673654--1673689	36	18	2	oui	1597150--1597185	36	18	2	oui	1642765--1642800	36	18	2,1	oui	1707940--1707957	18	?	?	non	1576290--1576307	18	?	?	non	1622475--1622510	36	18	2,1	oui
2	1756265--1756417	153	?	?	non	?	?	?	?	non	1725259--1725542	284	131	2,2	oui	?	?	?	?	non	?	?	?	?	non	1704911--1705194	284	131	2,2	oui
2	1853651--1853790	140	12	12	oui	1775646--1775785	140	12	12	oui	1813311--1813462	152	12	13,8	oui	?	?	?	?	non	1747808--1747959	152	12	14	oui	1792964--1793115	152	12	13,8	oui
2	1865545--1865762	218	?	?	non	?	?	?	?	non	?	?	?	?	non	1896672--1896880	209	9	26,4	oui	?	?	?	?	non	?	?	?	?	non
2	1873108--1873168	61	?	?	non	?	?	?	?	non	?	?	?	?	non	1904067--1904169	103	42	2,5	oui	?	?	?	?	non	?	?	?	?	non
2	1877455--1877488	34	15	2,4	oui	1799450--1799483	34	15	2,4	oui	1837138--1837172	35	16	3,6	oui	1908454--1915095	6642??	24	4,5	oui	1771789--1771822	34	15	2,4	oui	1816473--1816507	35	16	3,6	oui
2	1886389--1886434	46	12	3,8	oui	1808384--1808548	165	?	?	non	1846208--1846372	165	?	?	non	1923642--1923806	165	?	?	non	1781029--1781193	165	?	?	non	1825544--1825708	165	?	?	non
2	1893467--1893523	57	?	?	non	?	?	?	?	non	?	?	?	?	non	1930849--1930903	55	14	4	oui	1788227--1788300	74	17	6	oui	?	?	?	?	non
2	1949758--1949836	79	?	?	non	?	?	?	?	non	?	?	?	?	non	?	?	?	?	non	1861659--1861725	67	20	3,2	oui	?	?	?	?	non
2	1953870--1953917	48	24	2	oui	1875986--1876033	48	24	2	oui	?	?	?	?	non	1994971--1995042	72	24	2	oui	?	?	?	?	non	?	?	?	?	non
2	1953969--1954020	52	24	2	oui	1876085--1876136	52	24	2	oui	?	?	?	?	non	1995070--1995145	76	24	2,9	oui	?	?	?	?	non	?	?	?	?	non
2	1956568--1956589	22	?	?	non	?	?	?	?	non	?	?	?	?	non	1997693--1997729	37	15	2,5	oui	?	?	?	?	non	?	?	?	?	non
2	1977349--1977436	88	46	1,9	oui	1899463--1899550	88	46	1,9	oui	1941499--1941540	42	?	?	non	2018571--2018612	42	?	?	non	?	?	?	?	non	1920833--1920874	42	?	?	non
2	1994271--1994532	262	90	2,9	oui	1916384--1916645	262	90	2,9	oui	1956920--1957181	262	90	2,9	oui	2033984--2034157	174	90	1,9	oui	1899022--1899283	262	90	2,9	oui	1936254--1936427	174	90	1,9	oui
2	2002370--2002389	20	?	?	non	?	?	?	?	non	?	?	?	?	non	2041984--2042035	52	16	3,25	oui	?	?	?	?	non	?	?	?	?	non

nombre d'allèles dans les 6 souches	Mu50 Position	L: Longueur totale du minisatellite (pb)	U : taille du motif (pb)	N : nombre de copies	séquence détectée par le TRF	N315 Position	L:	U:	N:	TRF	Position MW2	L:	U:	N:	TRF	Position MRSA252	L:	U:	N:	TRF	Position NCTC8325	L:	U:	N:	TRF	Position MSSA476	L:	U:	N:	TRF
2	2028521--2028651	131	59	2,3	oui	1950634--1950764	131	59	2,3	oui	1991229--1991359	131	?	?	non	2068164--2068468	305	56	6	oui	1976877--1977004	128	56	3	oui	1970476--1970606	131	?	?	non
2	2029936--2030051	116	?	?	non	?	?	?	?	non	1992644--1992818	175	59	3	oui	?	?	?	?	non	?	?	?	?	non	1971891--1972065	175	59	3	oui
2	20320--20353	34	15	2,3	oui	20319--20352	34	15	2,3	oui	20313--20346	34	15	2,3	oui	20301--20334	34	15	2,3	oui	20309--20348	40	6	7	oui	20313--20346	34	15	2,3	oui
2	2051404--2051428	25	?	?	non	?	?	?	?	non	?	?	?	?	non	2091034--2091069	36	11	3,3	oui	?	?	?	?	non	?	?	?	?	non
2	2081629--2082475	847	327	3	oui	2003743--2004589	847	327	3	oui	2044650--2045496	847	315	4,98	oui	?	?	?	?	non	2030349--2031201	853	315	5	oui	2023896--2024742	847	315	4,98	oui
2	2089256--2089300	45	?	?	non	2012180--2012272	93	18	5,2	oui	?	?	?	?	non	?	?	?	?	non	?	?	?	?	non	?	?	?	?	non
2	2168990--2169070	81	16	5,3	oui	2092465--2092545	81	?	?	non	2121093--2121173	81	?	?	non	2197120--2197206	87	?	?	non	2106915--2106995	81	?	?	non	2100218--2100298	81	?	?	non
2	2197107--2197142	36	18	1,9	oui	2120582--2120617	36	18	1,9	oui	2148817--2148852	36	18	1,9	oui	2225236--2225271	36	18	1,9	oui	2135312--2135336	25	18	2	oui	2127942--2127977	36	18	1,9	oui
2	2221867--2222183	317	100	4	oui	2145343--2145659	317	100	4	oui	2172073--2172190	118	16	4,7	oui	2249472--2249587	116	28	2,3	oui	2158559--2158875	317	100	4	oui	2151178--2151295	118	16	4,7	oui
2	2289406--2289469	64	23	2,7	oui	2219593--2219656	64	23	2,7	oui	2239554--2239617	64	23	2,7	oui	2319072--2319129	58	23	2,7	oui	2227462--2227525	64	?	?	non	2218659--2218722	64	23	2,7	oui
2	2325183--2325236	54	12	4,3	oui	2255590--2255643	54	12	4,3	oui	2275635--2275688	54	12	4,3	oui	2356984--2357037	54	?	?	non	2265087--2267095	2009	?	?	non	2254740--2254793	54	12	4,3	oui
2	2421817--2421975	159	9	18	oui	2352786--2352944	159	9	18	oui	2363993--2364157	165	9	16,9	oui	2451527--2451691	165	9	18,3	oui	2363139--2363297	159	9	16	oui	2343098--2343262	165	9	16,9	oui
2	2426421--2426489	69	9	8,1	oui	2357390--2357458	69	9	8,1	oui	2368602--2368670	69	9	8,1	oui	2456140--2456199	60	9	4,1	oui	2367742--2367810	69	9	8,1	oui	2347707--2347775	69	9	8,1	oui
2	2458089--2458260	172	60	2,9	oui	2389058--2389229	172	60	2,9	oui	?	?	60	4,8	oui	2488212--2488383	172	60	2,9	oui	2400032--2400383	352	60	6	oui	?	?	60	4,8	oui
2	2467281--2467306	26	?	?	non	?	?	?	?	non	?	?	?	?	non	2497391--2497443	53	27	2,03	oui	2409402--2409455	54	9	6	oui	?	?	?	?	non
2	2495005--2495040	36	15	2,4	oui	2425974--2426009	36	15	2,4	oui	2437791--2437826	36	15	2,4	oui	?	?	?	?	non	2437997--2438047	51	15	3,4	oui	2416896--2416931	36	15	2,4	oui
2	2495094--2495218	125	9	16	oui	2426063--2426187	125	9	16	oui	2437880--2438004	125	9	16,6	oui	?	?	?	?	non	2438086--2438225	140	9	17	oui	2416985--2417109	125	9	16,6	oui
2	2546034--2546073	40	10	4,2	oui	2475484--2475523	40	10	4,2	oui	2488914--2488953	40	10	5,4	oui	2580665--2580705	41	?	?	non	2487615--2487709	95	10	4,2	oui	2468082--2468121	40	10	5,4	oui
2	2547600--2547676	77	15	5,1	oui	2476859--2476965	107	30	3	oui	2490358--2490464	107	30	3,36	oui	2582110--2582216	107	15	6,73	oui	2489113--2489219	107	15	7	oui	2469526--2469632	107	30	3,36	oui
2	2569627--2569663	37	18	2	oui	2498916--2498938	23	18	2	oui	2512390--2512426	37	18	2	oui	2604097--2604133	37	18	2	oui	2511096--2511132	37	18	2	oui	2491558--2491594	37	18	2	oui
2	2640837--2640874	38	18	2,1	oui	2569727--2569764	38	18	2,1	oui	2579677--2579705	29	18	2,1	oui	?	?	?	?	non	2580948--2580976	29	?	?	non	2559036--2559064	29	18	2,1	oui
2	2644556--2644605	50	24	2,1	oui	2573446--2573495	50	24	2,1	oui	2583303--2583367	65	24	2,1	oui	2663335--2663399	65	?	?	non	2584584--2584648	65	24	2,1	oui	2562662--2562726	65	24	2,1	oui

nombre d'allèles dans les 6 souches	Mu50 Position	L: Longueur totale du minisatellite (pb)	U : taille du motif (pb)	N : nombre de copies	séquence détectée par le TRF	N315 Position	L:	U:	N:	TRF	Position MW2	L:	U:	N:	TRF	Position MRSA252	L:	U:	N:	TRF	Position NCTC8325	L:	U:	N:	TRF	Position MSSA476	L:	U:	N:	TRF
2	2644617--2644699	**83**	9	9	**oui**	2573507--2573589	**83**	9	9	**oui**	2583364--2583461	**98**	9	12,2	**oui**	2663396--2663493	**98**	9	12,2	**oui**	2584645--2584742	**98**	9	9	**oui**	2562723--2562820	**98**	9	12,2	**oui**
2	2654323--2654372	**50**	?	?	non	?	**?**	?	?	non	?	**?**	?	?	non	2673118--2673167	**50**	18	2,8	**oui**	2594368--2594410	**43**	18	2,4	**oui**	?	**?**	?	?	non
2	267130--267160	**31**	?	?	non	?	**?**	?	?	non	245136--245182	**47**	16	2,9	**oui**	?	**?**	26	3,4	**oui**	214156--214202	**47**	16	2,9	**oui**	243836--243882	**47**	16	2,9	**oui**
2	2806967--2807000	**34**	12	2,8	**oui**	2742569--2742602	**34**	12	2,8	**oui**	2745944--2745977	**34**	12	3,1	**oui**	2827546--2827555	**10**	?	?	non	2748963--2748996	**34**	12	3,1	**oui**	2725284--2725317	**34**	12	3,1	**oui**
2	2877111--2877148	**38**	18	2,1	**oui**	2812712--2812749	**38**	18	2,1	**oui**	2819551--2819570	**20**	?	?	non	2901703--2901722	**20**	?	?	non	2820994--2821013	**20**	?	?	non	2798891--2798910	**20**	?	?	non
2	344399--344444	**46**	?	?	non	?	**?**	?	?	non	323320--323365	**46**	17	2,7	**oui**	?	**?**	?	?	non	293261--293326	**66**	21	3,1	**oui**	321965--322010	**46**	17	2,7	**oui**
2	348324--348354	**31**	?	?	non	?	**?**	?	?	non	?	**?**	?	?	non	344418--344462	**45**	14	5,57	**oui**	297232--297262	**31**	20	2,2	**oui**	?	**?**	?	?	non
2	474268--474313	**46**	?	?	non	?	**?**	?	?	non	439066--439104	**39**	17	2,2	**oui**	?	**?**	?	?	non	?	**?**	?	?	non	437727--437765	**39**	17	2,2	**oui**
2	502396--502454	**59**	?	?	non	?	**?**	?	?	non	?	**?**	?	?	non	?	**?**	?	?	non	420460--420602	**143**	49	2,9	**oui**	?	**?**	?	?	non
2	509671--509872	**202**	?	?	non	485129--485531	**403**	201	2	**oui**	470529--470931	**403**	201	2	**oui**	?	**?**	201	2	**oui**	427819--428221	**403**	201	2	**oui**	469191--469593	**403**	201	2	**oui**
2	751815--751992	**178**	6	30	**oui**	727541--727718	**178**	6	30	**oui**	723198--723375	**178**	6	29,7	**oui**	766466--766637	**172**	12	23,8	**oui**	671478--671655	**178**	6	30	**oui**	721626--721803	**178**	6	29,7	**oui**
2	800299--800522	**224**	?	?	non	?	**?**	?	?	non	?	**?**	?	?	non	814942--814985	**44**	18	2,4	**oui**	?	**?**	?	?	non	?	**?**	?	?	non
2	801381--801460	**80**	?	?	non	777107--777242	**136**	56	2,4	**oui**	?	**?**	56	2,4	**oui**	?	**?**	?	?	non	?	**?**	?	?	non	?	**?**	56	2,4	**oui**
2	842266--842402	**137**	55	2,5	**oui**	818048--818184	**137**	55	2,5	**oui**	813797--813877	**81**	?	?	non	856819--856902	**84**	?	?	non	762798--762933	**136**	55	2,5	**oui**	812225--812305	**81**	?	?	non
2	899533--899596	**64**	24	2,7	**oui**	860970--861033	**64**	24	2,7	**oui**	869406--869469	**64**	24	2,8	**oui**	900911--900974	**64**	?	?	non	805283--805346	**64**	24	2,7	**oui**	853399--853438	**40**	?	?	non
2	906124--906248	**125**	56	2,3	**oui**	874273--874397	**125**	56	2,3	**oui**	875996--876064	**69**	?	?	non	?	**?**	?	?	non	811873--811942	**70**	?	?	non	859965--860033	**69**	?	?	non

Caractéristiques des 122 répétitions en tandem de *Staphylococcus aureus* :

nombre d'allèles dans les 6 souches	taille du motif (pb)	position Mu50	gène à cette position dans Mu50	produit protéique	position du gène	taille motif / 3	motif protéique répété	répétitions en tandem étudiées
5	**9 ou 18**	1098012--1098170	sspA	V8 protéase, glutamyl endopeptidase	1099031..1098003	3 ou 6	PNNPDN	**Mu50_1098_18bp_9U**
5	**159**	1866118--1866562	SAV1738	hypothetical protein similar to smooth muscle caldesmon	1865490..1866848	53	ALKAQQAAIKEEASANNLSDTSQEAQEIQEAKREAQAEADKSVAVSNEESKAS	**Mu50_1866_159bp_3U**
5	**61**	2511193--2511367	intergénique			20,33		
5	**18**	636061--636666	sdrD	Ser-Asp rich fibrinogen-binding, bone sialoprotein-binding protein	632692..636849	6	SD	**Mu50_0636_18bp_33,3U**
5	**67 ou 133**	704561--704796	intergénique			22,33 ou 44,33		**Mu50_0704_67bp_4U**
5	**18**	888858--889722	fnb	fibrinogen-binding protein	887186..889993	6	SD	**Mu50_0888_18bp_48U**
4	**58**	1000868--1000965	intergénique			19,33		
4	**20 ou 64**	1291998--1292219	intergénique			6,67 ou 21,33		**Mu50_1291_64bp_3,5U**
4	**56**	1729388--1729679	**intergénique**			18,67		**séquence STAR Mu50_1729_56bp_5U**
4	**126**	1757525--1757639	intergénique			42		
4	**6 ou 39**	2152442--2152714	SAV2032	hypothetical protein similar to SdrH	2151760..2153007	2 ou 13	PK ou PN ou PD	
4	**24**	2351355--2351474	SAV2208	hypothetical protein	2351393..2351563	8		**Mu50_2351_24bp_5U**
4	**56**	2561664--2561795	**intergénique**			18,67		**séquence STAR**
4	**42**	2642053--2642330	fnb	fibronectin-binding protein homolog	2641824..2644940	14	PETPTPPTPEVPSE	**Mu50_2642_42bp_7U**
4	**18**	631615--632142	sdrC	Ser-Asp rich fibrinogen-binding, bone sialoprotein-binding protein	629464..632325	6	SD	**Mu50_0631_18bp_30,3U**
4	**43**	965164--965428	intergénique			14,33		**Mu50_0965_43bp_6U**
3	**22**	1086975--1087029	intergénique			7,33		

nombre d'allèles dans les 6 souches	taille du motif (pb)	position Mu50	gène à cette position dans Mu50	produit protéique	position du gène	taille motif / 3	motif protéique répété	répétitions en tandem étudiées
3	**9 ou 18**	1105143--1105186	atl	autolysin	1103624..1106455	3 ou 6		**Mu50_1105_18bp_2,5U**
3	**58**	1116357--1116488	intergénique			19,33		
3	**63**	1132682--1133067	SAV1078	hypothetical protein	1132622..1133071	21	LQLLVVRGFYACARRMYPSFT	**Mu50_1132_63bp_6,1U**
3	**134**	1194184--1194530	intergénique			44,67		**Mu50_1194_67bp_7U**
3	**55**	1219058--1219224	**SAV1165**	hypothetical protein	1219137..1219268	18,33	pas de répétition d'acides aminés	**séquence STAR**
3	**24**	122905--123156	spa	Immunoglobulin G binding protein A precursor	122614..123966	8	PGKEDNNK	**Mu50_0122_24bp_10U**
3	**174**	123159--123840	spa	Immunoglobulin G binding protein A precursor	122614..123966	58	QQNAFYEILNMPNLNEEQRNGFIQSLKDDPSQSANLLSEAKKLNESQAPKADNKFNKE	**Mu50_0123_174bp_3,9U**
3	**58**	1425109--1425340	**intergénique**			19,33		**séquence STAR Mu50_1425_58bp_4U**
3	**57**	1695062--1695135	SAV1584	hypothetical protein	1695032..1695190	19		
3	**60**	1985673--1985771	intergénique			20		
3	**56**	2039328--2039458	**intergénique**			18,67		**séquence STAR Mu50_2039_56bp_3U**
3	**256**	2185031--2185605	rRNA-5S	SAVrRNA11	2185545..2185659	85,33		
3	**9**	2294935--2295172	fmtB(mrp)	FmtB protein	2287935..2295380	3		**Mu50_2294_9bp_29U**
3	**69**	2546242--2546492	intergénique			23		
3	**437**	2611979--2612851	SAV2475	hypothetical protein	2611978..2612397	145,67		
3	**384**	2630754--2632463	SAV2496	hypothetical protein similar to accumulation-associated protein	2630406..2631983	128	KNPLTGEIISKGESKEEITKDPINELTEYGPETITPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTR	
3	**42**	2638502--2638675	fnbB	fibronectin-binding protein homolog	2638258..2641143	14	PEVPSEPETPVPPT	**Mu50_2638_42bp_4,1U**

nombre d'allèles dans les 6 souches	taille du motif (pb)	position Mu50	gène à cette position dans Mu50	produit protéique	position du gène	taille motif / 3	motif protéique répété	répétitions en tandem étudiées
3	**114**	2642422--2642721	fnb	fibronectin-binding protein homolog	2641824..2644940	38		
3	**81**	266128--266583	coa	staphylocoagulase precursor	264640..266616	27	KKPSKTNAYNVTTHANGQVSYGARPTQ	**Mu50_0266_81bp_5,6U**
3	**18**	2781740--2782399	clfB	Clumping factor B	2781518..2784151	6	SD	**Mu50_2781_18bp_36,7U**
3	**18 ou 42**	2818769--2819116	SAV2654	hypothetical protein similar to streptococcal hemagglutinin protein	2818476..2825291	6 ou 14	riche en S	
3	**56**	2837314--2837372	SAV2670	hypothetical protein	2837278..2837388	18,67	pas de répétition d'acides aminés	
3	**56**	311490--311657	**intergénique**			18,67		**séquence STAR Mu50_0311_55bp_3U**
3	**40**	43142--43471	**intergénique**			13,33		**mec HVR region (ou dru)**
3	**8**	465115--465161	intergénique			2,67		
3	**133**	529716--529869	intergénique			44,33		
3	**277**	535447--0536092	rRNA-5S	SAVrRNA03	535393..535507	92,33		
3	**21**	550991--551087	intergénique			7		
3	**21**	632181--632224	sdrC	Ser-Asp rich fibrinogen-binding, bone sialoprotein-binding protein	629464..632325	7	SD	
3	**18**	640048--640484	sdrE	Ser-Asp rich fibrinogen-binding, bone sialoprotein-binding protein	637243..640668	6	SD	**Mu50_0640_18bp_24,3U**
3	**72**	683273--683427	intergénique			24		
3	**61**	684180--684299	intergénique			20,33		
3	**69**	748508--748564	intergénique			23		
3	**59**	847717--848067	**intergénique**			19,67		**séquence STAR**
3	**56**	855096--855387	**intergénique**			18,67		**séquence STAR**
2	**20**	101929--101971	SAV0094	hypothetical protein	101923..102405	6,67	pas de répétition d'acides aminés	

nombre d'allèles dans les 6 souches	taille du motif (pb)	position Mu50	gène à cette position dans Mu50	produit protéique	position du gène	taille motif / 3	motif protéique répété	répétitions en tandem étudiées
2	**24**	1105876--1105912	atl	autolysin	1103624..1106455	8		
2	**18**	1107863--1107903	SAV1056	conserved hypothetical protein	1107769..1108986	6		
2	**14**	1139138--1139258	SAV1085	conserved hypothetical protein	1139228..1139461	4,67		
2	**7**	1175797--1175833	SAV1121	hypothetical protein similar to glycerophosphoryl diester phosphodiesterase	1175802..1176728	2,33	pas de répétition d'acides aminés	
2	**9**	1182603--1182659	SAV1129	conserved hypothetical protein	1181001..1182938	3		
2	**18**	1183288--1183311	SAV1130	cell surface protein	1183141..1184193	6		
2	**15**	1183549--1183618	SAV1130	cell surface protein	1183141..1184193	5		
2	**56**	1213418--1213706	**intergénique**			18,67		**séquence STAR Mu50_1213_56bp_5U**
2	**25**	1326853--1326899	intergénique			8,33		
2	**11**	1385013--1385033	intergénique			3,67		
2	**18**	149437--149526	148984..149709	hypothetical protein	148984..149709	6		
2	**58**	1513544--1513661	intergénique			19,33		
2	**231**	1516384--1517097	ebhA	hypothetical protein similar to streptococcal adhesin emb	1514410..1534551	77	KEKQALKDRINQILQQGHNGINNAMTKEEIEQAKAQLAQALKEIKDLVKAKENAKQDVDKQVQALIDEIDQNPNLTD	**Mu50_1516_231bp_3U**
2	**51**	1649052--1649072	intergénique			17		
2	**45**	1654721--1654766	SAV1540	hypothetical protein	1654602..1654901	15	pas de répétition d'acides aminés	
2	**18**	167280--167319	intergénique			6		
2	**18**	1673654--1673689	dnaG	DNA primase	1673192..1675009	6		
2	**131**	1756265--1756417	intergénique			43,67		

nombre d'allèles dans les 6 souches	taille du motif (pb)	position Mu50	gène à cette position dans Mu50	produit protéique	position du gène	taille motif / 3	motif protéique répété	répétitions en tandem étudiées
2	**12**	1853651--1853790	SAV1731	hypothetical protein	1853262..1855937	4		
2	**9**	1865545--1865762	SAV1738	hypothetical protein similar to smooth muscle caldesmon	1865490..1866848	3		
2	**42**	1873108--1873168	SAV1742	phenylalanyl-tRNA synthetase homolog beta subunit	1872669..1873265	14		
2	**15**	1877455--1877488	intergénique			5		
2	**12**	1886389--1886434	SAV1755	hypothetical protein	1885515..1886459	4		
2	**14**	1893467--1893523	intergénique			4,67		
2	**20**	1949758--1949836	intergénique			6,67		
2	**24**	1953870--1953917	intergénique			8		
2	**24**	1953969--1954020	intergénique			8		
2	**15**	1956568--1956589	yent1	Pathogenicity island SaPIn3	1956394..1956795	5		
2	**46**	1977349--1977436	intergénique			15,33		
2	**90**	1994271--1994532	tRNA-Gly			30		**Mu50_1994_90bp_3U**
2	**16**	2002370--2002389	intergénique			5,33		
2	**59**	2028521--2028651	**intergénique**			19,67		**séquence STAR**
2	**59**	2029936--2030051	intergénique			19,67		
2	**15**	20320--20353	rplI	50S ribosomal protein L9	20293..20739	5		
2	**11**	2051404--2051428	SAV1907	conserved hypothetical protein	2051419..2051721	3,67	pas de répétition d'acides aminés	
2	**315 ou 327**	2081629--2082475	truncated-mapW	truncated map-w protein	2081196..2082626	105 ou 109		
2	**15**	2089256--2089300	intergénique			5		

nombre d'allèles dans les 6 souches	taille du motif (pb)	position Mu50	gène à cette position dans Mu50	produit protéique	position du gène	taille motif / 3	motif protéique répété	répétitions en tandem étudiées
2	**16**	2168990--2169070	intergénique			5,33		
2	**18**	2197107--2197142	rsbU	sigmaB regulation protein	2196323..2197324	6		
2	**100**	2221867--2222183	intergénique			33,33		**Mu50_2221_100bp_3,2U**
2	**23**	2289406--2289469	fmtB(mrp)	protéine FmtB	2287935..2295380	7,67		
2	**12**	2325183--2325236	asp23	**alkaline shock protein 23**	2325166..2325675	4		
2	**9**	2421817--2421975	ssaA	homologue à la protéine SsaA (secretory antigen precursor)	2421604..2422407	3	NNY	
2	**9**	2426421--2426489	SAV2304	protéine hypothétique similaire à SsaA	2426325..2426825	3	NNY	
2	**60**	2458089--2458260	SAV2332	protéine hypothétique similaire au régulateur transcriptionnel LysR	2457210..2458094	20	pas de répétition d'acides aminés	
2	**9 ou 27**	2467281--2467306	SAV2342	protéine hypothétique conservée	2467258..2467878	3 ou 9		
2	**15**	2495005--2495040	SAV2368	protéine hypothétique similaire à la protéine TpgX	2494687..2495301	5		
2	**9**	2495094--2495218	SAV2368	protéine hypothétique similaire à la protéine TpgX	2494687..2495301	3		
2	**10**	2546034--2546073	intergénique			3,33		
2	**15 ou 30**	2547600--2547676	sbi	protéine de liaison aux IgG	2546792..2548078	5 ou 10	PKVEA ?	**Mu50_2547_15bp_5U**
2	**18**	2569627--2569663	SAV2438	homologue au transporteur ABC permease amino acid	2568739..2570148	6		
2	**18**	2640837--2640874	fnbB	homoloque à la protéine de fixation à la fibronectine	2638258..2641143	6		
2	**24**	2644556--2644605	fnb	homoloque à la protéine de fixation à la fibronectine	2641824..2644940	8		
2	**9**	2644617--2644699	fnb	homoloque à la protéine de fixation à la fibronectine	2641824..2644940	3		
2	**18**	2654323--2654372	SAV2512	protéine hypothétique similaire à au transporteur glucarate	2654001..2655278	6		
2	**16 ou 26**	267130--267160	intergénique			5,33 ou 8,67		

nombre d'allèles dans les 6 souches	taille du motif (pb)	position Mu50	gène à cette position dans Mu50	produit protéique	position du gène	taille motif / 3	motif protéique répété	répétitions en tandem étudiées
2	**12**	2806967--2807000	SAV2646	protéine hypothétique conservée	2805671..2807578	4		
2	**18**	2877111--2877148	intergénique			6		
2	**21**	344399--344444	SAV0300	protéine hypothétique conservée	344169..344630	7		
2	**14**	348324--348354	SAV0305	protéine hypothétique similaire au transporteur NirC	347438..348262	4,67		
2	**17**	474268--474313	set14	Ilôt de pathogénicité SaPIn2 exotoxin 14	474574..475257	5,67		
2	**49**	502396--502454	intergénique			16,33		
2	**201**	509671--509872	SAV0465	protéine hypothétique	509460..510263	67		
2	**6 ou 12**	751815--751992	SAV0677	protéine hypothétique conservée	751712..752107	2 ou 4	KD	
2	**18**	800299--800522	SAV0724	protéine hypothétique similaire à l'histidinol-phosphate aminotransferase	799323..800381	6		
2	**56**	801381--801460	intergénique			18,67		
2	**55**	842266--842402	**intergénique**			18,33		**séquence STAR Mu50_0842_55bp_3U**
2	**24**	899533--899596	SAV0825	protéine hypothétique conservée	899565..900182	8		**Mu50_0899_24bp_2,7U**
2	**56**	906124--906248	**intergénique**			18,67		**séquence STAR Mu50_0906_56bp_3U**

Codage Shopsin des motifs de la répétition en tandem localisée dans le gène *spa*

AAAGAAGACAACAACAAGCCTGGC	|F| : consensus
AAAGAAGACAACAAAAAACCTGGC	|A|
AAAGAAGACAACAAAAAACCTGGT	|B|
AAAGAAGACAACAAAAAGCCTGGC	|C|
AAAGAAGACAACAACAAACCTGGC	|D|
AAAGAAGACAACAACAAACCTGGT	|E|
AAAGAAGACAACAACAAGCCTGGC	|F|
AAAGAAGACAACAACAAGCCTGGT	|G|
AAAGAAGACAATAACAAGCCTGGC	|H|
AAAGAAGACGGCAACAAACCTGGC	|J|
AAAGAAGACGGCAACAAACCTGGT	|K|
AAAGAAGACGGCAACAAGCCTGGC	|L|
AAAGAAGACGGCAACAAGCCTGGT	|M|
AAAGAAGATGGCAACAAACCTGGC	|N|
AAAGAAGATGGCAACAAACCTGGT	|O|
AAAGAAGATGGCAACAAGCCTGGC	|P|
AAAGAAGATGGCAACAAGCCTGGT	|Q|
AAAGAAGATGGTAACAAACCTGGC	|R|
GAGGAAGACAACAAAAAACCTGGC	|S|
GAGGAAGACAACAAAAAACCTGGT	|T|
GAGGAAGACAACAACAAACCTGGT	|U|
GAGGAAGACAACAACAAGCCTAGC	|V|
GAGGAAGACAACAACAAGCCTGGC	|W|
GAGGAAGACAACAACAAGCCTGGT	|X|
GAGGAAGACAATAACAAGCCTGGC	|Y|
GAGGAAGACAATAACAAGCCTGGT	|Z|
GAGGAAGACGGCAACAAACCTGGT	|A2|
AAAGAAGACAACAAAAAAGCCTGGT	|B2|
AAAGAAGACAATAACAAGCCTGGT	|C2|
GAGGAAGACAATAACAAACCTGGT	|D2|
AAAGAAGACAGCAACAAGCCTGGC	|E2|
GAGGAAGACAATAACAAGCCTAGT	|F2|
AAAGAAGACGGCAAAAAACCTGGC	|G2|
GAGGAAGACAACAACAAACCTGGC	|H2|
AAAGAAGACAACAAAAAAGCCTAGC	|I2|
AAAGAAGACAACAAAAAAGCCTAGC	|J2|
AAAGAAGATGGCAACAAGCCTAGT	|K2|
AAAGAAGACAACAA___GCCTGGT	|M2| (nouveau)

Résumé

Les répétitions en tandem sont constituées de successions de motifs d'ADN. Ces structures présentes dans tous les organismes, procaryotes comme eucaryotes, ont des applications dans de nombreux domaines. Depuis quelques années seulement, les répétitions en tandem sont étudiées chez les bactéries. Le polymorphisme associé à ces séquences peut être utilisé pour le génotypage de bactéries pathogènes, permettant une identification précise au niveau de la souche. Le polymorphisme des séquences répétées est de deux types : polymorphisme de longueur et mutations internes aux motifs. Les génomes des deux bactéries pathogènes responsables d'infections nosocomiales, *Staphylococcus aureus* et *Pseudomonas aeruginosa,* ont été étudiés dans le but d'identifier des séquences répétées polymorphes. Un ensemble de marqueurs polymorphes a été validé expérimentalement pour ces deux espèces permettant un typage dit MLVA (pour « Multiple Locus VNTR Analysis »). Le travail plus classique de typage par la taille de la répétition a été complété par un travail de séquençage de certains allèles. Les résultats obtenus montrent comment le typage « MLVA » complété si nécessaire par le séquençage d'allèles, pourraient constituer de nouvelles méthodes peu coûteuses participant au contrôle des infections bactériennes.

Abstract

Tandem repeats are constituted by the succession of DNA units. These structures, present in all organisms, prokaryotes as well as eukaryotes, have many fields of application. Since a few years, and owing to a large extend to the release of whole genome sequence data, these sequences are being studied in bacteria. Polymorphism associated with repeated sequences is useful for the genotyping of pathogenic bacteria in the aim to identify bacteria at the strain level. Two levels of polymorphism can be associated with tandem repeats: length polymorphism (counting the number of repeat units) and internal variants in units (interpreting the internal organisation of the array). Two bacterial species implicated in hospital aquired infections, *Pseudomonas aeruginosa* and *Staphylococcus aureus,* were studied in order to validate sets of polymorphic tandem repeats witch discriminate strains. A collection of polymorphic markers were developped for the two bacterial species as a MLVA scheme typing. In addition, sequence data was produced for some loci, in order to see to which extend strain similarity suggested by MLVA does correlate with the often more complex evolutionary history revealed by tandem repeat sequence analysis. The results obtained further illustrate the potential of tandem repeat analysis for the monitoring of a growing number of infectious diseases.