De l'usage du polymorphisme de répétitions en tandem pour l'étude des populations bactériennes : mise au point et validation d'un système de génotypage automatisé utilisant la technique de MLVA

Daniel Sobral

THÈSE DE DOCTORAT DE L'UNIVERSITE PARIS SUD

Spécialité: Génomique
Ecole Doctorale: Gènes, Génomes, Cellules

Présentée et soutenue publiquement par:
Daniel SOBRAL

Le 02 mai 2012

Pour obtenir le grade de:
DOCTEUR de l'UNIVERSITÉ PARIS SUD

De l'usage du polymorphisme de répétitions en tandem pour l'étude des populations bactériennes

Mise au point et validation d'un système de génotypage automatisé utilisant la technique de MLVA

Devant le jury composé de:

Rapporteurs :

Monsieur Frédéric Laurent	Hôpital de la Croix Rousse, Lyon
Madame Nathalie van der Mee-Marquet	Hôpital Trousseau, Tours

Examinateurs :

Monsieur Pierre Capy	Université Paris Sud, Gif sur Yvette
Monsieur Jean-Louis Gaillard	Hôpital Raymond Poincaré, Garches
Monsieur Thierry Wirth	Muséum National d'Histoire Naturelle, Paris

Encadrants :

Madame Fabienne Loisy-Hamon	Ceeram, La Chapelle sur Erdre
Monsieur Gilles Vergnaud	Direction Générale de l'Armement, Directeur de thèse

Sommaire

Remerciements

Cette thèse est le fruit du travail de recherche mené en collaboration avec l'équipe GPMS (Gènes, Polymorphismes et MiniSatellites) de l'IGM (Institut de Génétique et Microbiologie) et la société de biotechnologies Ceeram (Centre Européen d'Expertise et de Recherche sur les Agents Microbiens). Aussi, j'aimerais remercier toutes les personnes qui, de près ou de loin, m'ont guidé et soutenu durant ces trois dernières années.

Je tenais tout d'abord à remercier Monsieur Gilles Vergnaud, directeur de l'équipe GPMS, pour la confiance qu'il m'a témoignée en m'accueillant dans son équipe de recherche et en acceptant de diriger cette thèse. Pour sa disponibilité, son oeil critique, son pragmatisme, sa rigueur intellectuelle, je lui suis particulièrement reconnaissant.

Mes vifs remerciements vont à Madame Christine Pourcel pour ses précieux conseils, ses critiques constructives, son suivi quotidien et sa rigueur scientifique.

Je souhaite particulièrement remercier Monsieur Benoit Lebeau et Madame Fabienne Loisy-Hamon d'avoir cru en mon projet et pris le risque de le porter financièrement. Ils m'ont permis de m'épanouir professionnellement en m'accordant leur confiance dans la gestion de ce projet ambitieux.

J'exprime toute ma gratitude envers les membres du jury pour avoir consacré leur temps à ma thèse. Je cite en particulier Madame le Docteur Nathalie van der Mee-Marquet et Monsieur le Docteur Frédéric Laurent qui m'ont fait l'honneur d'accepter d'être rapporteurs de ma thèse.

Je remercie mes parrains de thèse Messieurs les Professeurs Jean-Louis Gaillard et Christophe Sola pour leur suivi et leurs précieuses indications.

Mes plus chaleureux remerciements s'adressent également, à toutes les personnes de l'équipe GPMS, avec qui j'ai partagé un café, un repas ou une discussion pendant mes années de thèse : Rim Bouchouicha, Christiane Essoh, Yolande Hauck et Philippe Le Flèche. Je témoigne ici de ma sincère reconnaissance et amitié envers Yann Blouin avec qui j'ai partagé d'excellents moments. Je les remercie tous d'avoir assuré une ambiance particulièrement favorable pour mener à bien mon travail.

L'occasion m'est donnée ici d'exprimer ma gratitude envers les membres, anciens et actuels, de l'équipe Ceeram : Franck Chatigny, Isabelle Davieau, Axelle Delage, Angélique Fourrier, Isabelle Grasland, Sandrine Hattet et Géraldine Leturnier. J'adresse un remerciement particulier à Denis Bidot qui, bien plus qu'un collègue, est devenu un ami. Je les remercie tous pour avoir contribué à la bonne humeur de l'open-space malgré les contraintes liées à l'activité d'une biotech en pleine croissance.

Je dédie ce travail à mes parents Olimpio et Marie-Hélène, à ma sœur Sonia, à ma grand-mère Henriette et à mes amis pour leur soutien. Je remercie particulièrement mon amie, Marion, pour sa patience et pour tout le réconfort qu'elle a su m'apporter sans relâche.

Liste des abréviations

ADN : Acide DésoxyriboNucléqiue
ARN : Acide RiboNucléique
AFLP : Amplified Fragment Length Polymorphism
ATP : Adenosine TriPhosphate
BIME : Bacterial Interspaced Mosaic Elements
CC : Complexe Clonal
CIFRE : Convention Industrielle de Formation par la Recherche
CRISPR : Clustered Regularly Interspaced Short Palindromic Repeats.
ERIC : Enterobacterial Repetition Intergenic Consensus
ESGEM : European Study Group on Epidemiological Markers
GPMS : Génomes, Polymorphismes et MiniSatellites
GSMM : General Stepwise Mutation Model
HACCP : Hard Analysis Critical Control Point
HIA : Hôpital d'Instruction des Armées
IS : Insertion Sequence
kb : kilobase
MIRU : Mycobacterial Interspersed Repetitive Units
MITE : Miniature Inverted repeat Transposable Elements
MLEE : Multi Locus Enzyme Electrophoresis
MLST : Multi Loci Sequence Typing
MLVA : Multiple Loci VNTR Analysis
MST : Minimum Spanning Tree
NEMIS : Neisseria Miniature Insertion Sequences
NGS : Next Generation Sequencing
ORF : Open Reading Frame
pb : paire de bases
PCR : Polymerase Chain Reaction
PFGE : Pulsed Field Gel Electrophoresis
PSPE : Parallel Serial Passage Experiment
QTL : Quantitative Trait Loci
RAPD : Random Amplified Length Polymorphism
REP : Repetitive Extragenic Palindromic
RFLP : Restriction Fragment Length Polymorphism
RIVM : RijkInstitut voor Volksgezondheid en Milieu
SBT : Sequence Based Typing
SNP : Single Nucleotide Polymorphism
SSM : Slipped-Strand Mispairing
SSMM : Single Stepwise Mutation Model
ST : Sequence Type
STR : Short Tandem Repeat
T_{ADN} : Transposon
TIGR : The Institute for Genomic Research
TRF : Tandem Repeats Finder
UPGMA : Unweighted Pair Group Method with Arithmetic Mean
VACC : VNTR Analysis Clonal Complex
VNTR : Variable Number of Tandem Repeats

Liste des tables et figures

Avant-propos

En 1942, Ernst Mayr définissait une espèce de la façon suivante : « *Les espèces sont des groupes de populations naturelles, effectivement ou potentiellement interfécondes, qui sont génétiquement isolées d'autres groupes similaires* » [1]. Cette notion n'est pas applicable pour les organismes asexués. Un problème conceptuel s'impose alors aux taxonomistes : celui de l'espèce bactérienne. Identifiés comme « *animalcules des infusions* » par Antoni van Leeuwenhoek lors de leur découverte au 17ème siècle, les micro-organismes ont longtemps été considérés comme les représentants d'une même espèce, douée de pléomorphisme *i.e.* la capacité à prendre une grande variété de formes. Dans son *Traité de systématique bactérienne* paru en 1961, le pasteurien André-Romain Prévot avance l'idée que l'espèce bactérienne n'a pas de sens en systématique mais que l'établissement d'une classification bactérienne permet de structurer notre compréhension du monde bactérien. Il définit ainsi l'espèce bactérienne comme « *une mosaïque d'enzymes et d'antigènes* ». Les Canadiens Sorin Sonea et Maurice Panisset désignent alors l'ensemble des bactéries comme un superorganisme global sexué [2]. Cependant, cette vision du monde bactérien est en pratique peu utile pour décrire et rendre compte de la réalité de la diversité bactérienne et des particularités métaboliques, physiologiques, écologiques, pathogéniques des microorganismes. Le monde bactérien n'est en effet pas un continuum, et il est utile d'y introduire une notion d'espèce. Frederik Cohan résume ainsi l'espèce bactérienne : « *Bacterial species exist… bacterial diversity is organized into discrete phenotypic and genetic clusters, which are separated by large phenotypic and genetic gaps, and these clusters are recognized as species* » [3].

L'avènement des théories fondées par les microbiologistes du 19ème siècle tels que Louis Pasteur et Robert Koch a contribué à l'établissement d'une classification procaryote. Dans cette classification, la notion d'espèce est largement définie par la pathogénicité du microorganisme. En 1957, le microbiologiste Peter Sneath applique la taxonomie numérique à la systématique bactérienne en axant son analyse sur plus de cent caractères d'ordre morphologique, biochimique, cultural ou structurel [4]. La classification de référence publiée dans le *Bergey's Manual of Systematic Bacteriology* repose sur cette approche [5]. L'avènement des méthodes moléculaires a contribué à l'augmentation de la robustesse de cette classification. L'application des méthodes d'hybridation moléculaire a permis de proposer une définition génomique de l'espèce bactérienne. Celle-ci est définie de manière opérationnelle comme l'ensemble des isolats caractérisés par un taux d'hybridation ADN/ADN supérieur à 70%. Cette caractérisation est toujours reconnue par *l'International Committee on Systematics of Prokaryotes*. Le développement du séquençage de l'ADN par Sanger, Maxam et Gilbert, et les études de similarité de séquence de l'ADNr 16S ont ensuite conduit à la proposition du seuil de 97% pour définir une espèce bactérienne [6]. La classification bactérienne est une science évolutive dont les termes et les codes sont modulés par les perpétuelles découvertes scientifiques. L'identification bactérienne et la notion d'espèce chez les procaryotes ne constituant pas le cœur du sujet de cette thèse, je m'intéresserai ici davantage à l'évaluation de la diversité au sein de plusieurs espèces bactériennes. Néanmoins, les concepts abordés sont similaires et sont confrontés aux mêmes

problématiques : il s'agit d'explorer la diversité, de rechercher des déterminants génétiques ou phénotypiques de différenciation, de classer les individus bactériens. Les génomes bactériens sont des structures mosaïques, flexibles, dynamiques, et soumises à des flux génétiques perpétuels. Chaque génome bactérien résulte d'une histoire évolutive différente et peut être différencié des membres de son espèce. Durant ces dernières années, l'essor de la génomique a permis de révéler l'existence de nombreuses sources de polymorphisme pour distinguer les membres d'une même espèce : les réarrangements chromosomiques, les mutations ponctuelles, ou encore les séquences répétées. Ces objets génétiques polymorphes sont le support de nouveaux outils conçus pour évaluer la diversité bactérienne : c'est le génotypage bactérien. Celui-ci permet d'appréhender la diversité d'une population bactérienne en différenciant avec précision les isolats bactériens. Ces techniques ont révolutionné la microbiologie dans nombre de ses facettes comme la taxonomie et la phylogénie, établissant ainsi les bases d'une meilleure connaissance du monde bactérien. Le typage constitue également un puissant outil pour l'épidémiologie des maladies infectieuses, science qui caractérise le pathogène impliqué dans le déclenchement d'une maladie, ses facteurs de transmission et les mécanismes de virulence déployés. L'isolat incriminé se voit attribuer une véritable carte d'identité microbiologique permettant à l'épidémiologiste de le tracer précisément et d'identifier son foyer. La découverte de la source de la contamination permettra ainsi d'appréhender et de définir la mise en œuvre de méthodes de lutte adaptées contre la dissémination du pathogène et l'expansion de la maladie.

Les bactéries sont des organismes ubiquitaires, pléiotropes et représenteraient, aujourd'hui, une biomasse totale de plus de 10^{30} cellules [7]. La connaissance de la diversité d'une espèce bactérienne ne peut se limiter à l'analyse de quelques isolats issus d'un même écosystème puisqu'il s'agit d'identifier des discontinuités de distribution de variations au sein de grandes populations de microorganismes. Le développement d'outils rapides et haut-débit permettant de constituer un panorama global de la diversité génétique d'une population bactérienne doit donc être envisagé. C'est dans cette quête d'une méthodologie accessible, facile à mettre en œuvre et permettant le typage systématique de tout isolat bactérien que s'inscrivent mes travaux. Au cours de cette thèse, je me suis penché sur l'étude des séquences répétées en tandem polymorphes ou VNTR (*Variable Number of Tamdem Repeats*). Ces structures génétiques complexes et dynamiques sont d'excellents marqueurs de discrimination utilisés pour génotyper les bactéries. L'étude du polymorphisme d'une collection de VNTR est le cœur de la méthodologie de génotypage bactérien par MLVA (*Multiple Loci VNTR Analysis*). Durant cette thèse, trois espèces bactériennes pathogènes ont été choisies comme modèles expérimentaux : *Legionella pneumophila*, *Pseudomonas aeruginosa* et *Staphylococcus aureus*. Ces espèces bactériennes sont différentes à bien des égards : mécanismes de virulence, spectres d'hôtes, réservoir environnemental, histoire évolutive ou encore structure de population. L'objet de mes travaux a été, en premier lieu, de trouver un procédé technologique permettant au MLVA de répondre aux besoins de génotypage haut-débit. Puis, il était nécessaire de confirmer les VNTRs en tant

qu'outils épidémiologiques pour ces pathogènes. Enfin, j'ai dû évaluer la pertinence de ces loci en tant que marqueurs informatifs de structuration des populations bactériennes.

Les trois dernières années ont vu l'avènement des nouvelles techniques de séquençage, autrement nommées NGS (*Next Generation Sequencing*). Ces procédés permettent la fourniture rapide et massive de données génomiques, une manne d'information extraordinaire pour les biologistes. L'essor parallèle de la science d'étude de ces données, la bioinformatique, contribue au développement de nouveaux outils et à l'ouverture de nouvelles perspectives de recherche. Ainsi, en génomique, les concepts et techniques sont en constante mutation. Après ces trois années, un premier bilan s'impose et plusieurs interrogations surviennent. L'avancée technologique du MLVA à laquelle mes travaux de thèse ont participé sera-elle le garant de sa généralisation ? Peut-on imaginer désormais de séquencer de manière systématique et en temps réel les génomes des isolats récoltés ? Les VNTRs vont-ils devenir des marqueurs désuets ? D'autres particularités de ces régions génomiques peuvent-elles être exploitées ?

Dans le chapitre introductif, je décrirai les principales composantes de l'épidémiologie des maladies infectieuses ainsi que l'apport du typage bactérien dans la connaissance de la diversité bactérienne. Une description des VNTRs en tant que structures génomiques accompagnée d'un bilan relatant leurs principales forces et faiblesses comme marqueurs épidémiologiques finalisera l'introduction. Puis, les travaux entrepris au cours de cette thèse seront illustrés par quatre articles, chacun précédé d'un résumé relatant les tenants et aboutissants des travaux entrepris. Enfin, plusieurs éléments seront discutés à la lumière des résultats obtenus et permettront de répondre aux précédentes interrogations.

Introduction

I) Epidémiologie des maladies infectieuses

1. Préambule

1.1. Notions d'épidémiologie

Par définition, l'épidémiologie est la science qui évalue l'apparition, les déterminants et la distribution d'une maladie dans une population [8]. Par extension, l'épidémiologie des maladies infectieuses est la science qui étudie la dynamique des maladies infectieuses et les agents responsables de leur transmission (bactéries, virus, parasites, champignons, …). L'épidémiologie est intimement liée aux statistiques. Les outils utilisés servent pour mesurer des fréquences, taux de mortalité ou taux de morbidité, et permettent de qualifier la typologie d'une maladie selon sa fréquence temporelle et géographique. La terminologie épidémiologique recense ainsi quatre grandes familles de maladies : la sporadie, l'épidémie, la pandémie et l'endémie [8]. Une maladie sporadique se déclare occasionnellement à intervalles de temps irréguliers. Une maladie endémique est spécifique d'une zone géographique donnée et maintenue à un taux bas et constant avec une incidence régulière. Une épidémie est la survenue brutale d'une maladie sur un espace restreint alors que la pandémie est une épidémie distribuée globalement.

Le cycle d'une maladie infectieuse est représenté par une chaîne composée de cinq maillons : l'agent pathogène, la source de l'agent, la transmission à l'hôte, la sensibilité de l'hôte et la sortie de l'hôte [8]. Plusieurs facteurs interviennent dans le déroulement de ce cycle : des facteurs microbiologiques, des facteurs d'hôtes et des facteurs environnementaux. L'analyse des facteurs microbiologiques se concentre sur l'étude de l'agent infectieux, de son origine et de son réservoir. L'origine ou foyer est le lieu à partir duquel l'organisme pathogène est transmis à l'hôte soit indirectement par l'environnement soit directement par un agent intermédiaire. L'intervention au niveau du foyer de l'agent rompt par conséquent le cycle infectieux. La recherche de l'origine et du réservoir de l'agent étiologique est donc fondamentale pour le contrôle des maladies infectieuses. En cela, le Dr John Snow est considéré comme le premier épidémiologiste. Ce médecin anglais a, après investigations, identifié une pompe à eau publique comme source de l'épidémie de choléra à Londres en 1849. Ce pionnier de l'épidémiologie moderne incrimina, en 1854 lors de la deuxième épidémie de choléra, une compagnie de distribution d'eau et attribua l'épidémie à la contamination de l'eau potable. Près de 30 ans avant la découverte de l'agent étiologique du choléra, *Vibrio cholerae*, par Robert Koch, John Snow affirmait que l'agent responsable des épidémies de choléra avait la capacité de se multiplier dans l'eau et en déduisait des mesures préventives efficaces [8].

1.2. Vers le typage bactérien

L'épidémiologie des maladies infectieuses cherche à élucider les mécanismes de transmission des agents infectieux, l'existence de réservoirs et leur origine. L'épidémiologiste doit pouvoir caractériser avec précision l'isolat bactérien incriminé afin d'identifier sa source et ses vecteurs de transmission. Pour répondre à ce besoin, une approche d'épidémiologie qualitative ne se contente pas de l'identification au niveau de l'espèce du pathogène incriminé. En effet, le développement d'outils moléculaires reproductibles, standardisés et permettant une caractérisation plus fine, le typage, de la bactérie isolée est nécessaire afin de déterminer la probabilité pour deux isolats d'avoir la même origine. Par opposition à l'identification qui se contente de déterminer l'espèce selon une approche descendante, le typage bactérien permet de caractériser la variabilité génétique d'une espèce donnée et d'identifier alors les relations de similitude entre les souches qui la composent par une approche ascendante. Les bactéries mutent, recombinent et s'adaptent ; cette flexibilité génomique guidée par la dérive génétique et la sélection naturelle [9] est un élément de compréhension de la diversité génétique observée au sein d'une même espèce. Le typage épidémiologique repose sur le postulat selon lequel les isolats d'un même agent pathogène impliqués dans la même chaîne de transmission sont clonaux *i.e.* ils sont tous descendants d'une même cellule ancêtre sans apport génétique externe. Ces isolats sont donc les représentants d'un même clone défini par un type spécifique, c'est-à-dire par le résultat d'une technique de typage. La caractérisation fine du clone, génotypique et phénotypique, définira une souche.

2. Problématiques épidémiologiques

L'épidémiologie des maladies infectieuses se décline sur différentes échelles spatio-temporelles selon les problématiques identifiées. De l'investigation épidémique à la biosurveillance des agents pathogènes, l'épidémiologiste s'apparente à un détective chargé de tracer le pathogène localement ou de déterminer plus globalement son comportement et sa distribution mondiale.

2.1. Investigation des cas épidémiques

Une épidémie se reconnait à l'augmentation drastique, dans un court laps de temps, de l'incidence d'une maladie au sein d'une population sur une aire géographique définie. La détection, l'investigation et le contrôle des épidémies sont étroitement liés et ont pour finalité la santé publique et la prévention. L'investigation sur le terrain repose sur une approche multidisciplinaire (clinique, épidémiologique, environnementale et microbiologique) et une démarche méthodologique basée sur la description et le test d'hypothèses. La déclaration de l'épidémie et la transmission de l'agent infectieux résultent de l'interaction entre l'agent lui-même, l'environnement, et l'hôte. L'altération

d’un membre de cette triade ou de leurs interactions peut entraîner l’augmentation de la transmission et de l’incidence d’une maladie. Couplée aux investigations épidémiologiques (caractéristiques spatio-temporelles de l’épidémie, recensement et répartition des cas, symptomatologie, données démographiques, physiques, psychologiques et sociales des malades) coordonnées avec l’enquête environnementale (identification d’un lieu, d’un site ou d’un processus identifié à risque), l’analyse microbiologique par typage est un des piliers de l’investigation des épidémies [10]. Elle n’identifie pas les modes de transmission ni les facteurs de risque mais s’attache à identifier la source, le véhicule et le réservoir de l’agent infectieux.

La première étape de l’investigation est la mise en évidence du caractère épidémique par le signalement, à l’autorité sanitaire en charge, de phénomènes de santé jugés inhabituels [10]. La connaissance du contexte épidémiologique, clinique et microbiologique d’une aire géographique limitée permet aux professionnels de terrain de faire le constat de l’épidémie lorsqu’un seuil épidémique statistique est franchi. Cependant, dans d’autres situations, l’augmentation du nombre de cas peut être modéré ou apparaître comme compatible avec les fluctuations temporelles de la maladie. C'est d'ailleurs la capacité à signaler le plus précocement possible une épidémie, alors que les cas sont dispersés géographiquement et peu nombreux, qui rend l’intervention préventive d’autant plus efficace. Le typage bactérien peut contribuer à révéler précocement l’état épidémique. La comparaison génétique des différents isolats cliniques confirmera ou infirmera l’épidémie en révélant la similarité ou la différence des différents isolats : l’alerte épidémique sera donnée lorsque deux patients indépendants s’avèrent contaminés par le même type bactérien [10]. Dans le cas contraire, les isolats sont différents, les patients ont donc été contaminés par des sources distinctes et/ou l’agent infectieux n’a pas circulé entre différents hôtes : ces cas isolés infectés par la même espèce bactérienne s’apparentent donc à des manifestations sporadiques indépendantes. Par extension, les techniques de typage garantissent la distinction du groupe des évènements sporadiques de celui des cas épidémiques. Ainsi, les cas sporadiques de salmonellose, d’incidence constante, ont été écartés de ceux impliqués dans l’épidémie alimentaire de 1993 en France [11]. La confirmation de la nature clonale de l’agent infectieux est un argument en faveur de la mise en évidence du caractère épidémique de l’infection. Cependant, les épidémies ne sont pas toujours clonales, plusieurs génotypes différents peuvent être impliqués dans des infections concomitantes. Le typage de plusieurs isolats cliniques impliqués dans une épidémie de listériose apparemment homogène a permis de montrer que deux génotypes, issus de deux sources alimentaires distinctes, étaient responsables de deux épidémies distinctes et concomitantes [12]. Des modifications climatiques peuvent également être à l'origine d'évènements épidémiques causés par différentes souches [13].

2.2. Traçabilité des agents pathogènes

Le typage bactérien permet d'attribuer une carte d'identité microbiologique, ou empreinte, à chaque "individu" bactérien. Tous les isolats récoltés peuvent ainsi être tracés au sens propre du terme (par définition au sens de la norme ISO 8402, la traçabilité correspond à la capacité de retrouver l'historique, l'utilisation ou la localisation d'un produit ou d'une activité au moyen d'informations enregistrées). Cette traçabilité permet d'identifier, d'authentifier et de localiser. Le typage bactérien est donc un outil de traçabilité d'évènements infectieux d'origine naturelle, accidentelle, ou criminelle touchant humains, animaux d'élevage ou cultures.

a. Traçabilité en santé humaine et animale

Le bactériologiste disposera d'outils de typage pour comparer les isolats cliniques afin de mettre en évidence une épidémie, comparer les isolats cliniques aux isolats de l'environnement médical lors des contrôles effectués par le service d'hygiène hospitalière ou comparer plusieurs isolats d'un même malade issus de prélèvements de nature différente ou à plusieurs intervalles de temps.

Sur le plan du diagnostic individuel, il permet d'établir la chronicité d'une infection en différenciant récidive, *i.e.* persistance d'une infection incomplètement traitée avec éventuellement acquisition de résistances aux antibiotiques par la souche bactérienne en cause, et réinfection, *i.e.* l'infection par une nouvelle souche de la même espèce [14]. Le suivi génétique des infections chroniques est utile pour comprendre les mécanismes de persistance et adapter les traitements dont l'efficacité peut être évaluée par typage : le remplacement de la souche persistante par une autre indiquera l'efficacité du traitement sur la souche primo-colonisante mais également la résistance développée par la souche récidiviste.

Le typage assure également l'identification du foyer initial d'une infection disséminée et établit son origine : endogène (flore digestive, cutanée ou respiratoire présente à l'admission) ou exogène (acquisition en cours d'hospitalisation) [15]. Ainsi, par exemple, l'emploi d'un procédé de typage doit permettre d'identifier l'origine nosocomiale ou communautaire d'une infection et de mettre en place les procédures pour y remédier [16]. La mise en place d'une méthode de typage dans un hôpital américain a permis de réduire, en moins de quatre ans, de 23% le taux d'infections nosocomiales désignant ainsi le typage bactérien comme un réel outil préventif de santé publique [17].

Dans une situation épidémique, il permet de tracer la propagation et le suivi d'infections microbiennes en identifiant les modes et vecteurs de diffusion : interhumain (transmission manuportée) ou par contact avec un réservoir ou un véhicule contaminé (siphon d'évier, matériel médicochirurgical, solutés médicamenteux ou antiseptiques, etc.). Les sources peuvent être multiples (environnement médical, environnement urbain, produit alimentaire, etc. …), une investigation analytique préliminaire doit donc orienter les recherches.

b. Traçabilité en agro-alimentaire

Les industries agro-alimentaires ont l'obligation de contrôler et de maîtriser les différentes étapes du processus de fabrication de leurs produits. Dans le cadre de l'application de mesures de traçabilité, l'identification rapide et certaine de la source d'une contamination microbienne offre la possibilité d'un confinement rapide. La caractérisation au-delà de l'espèce de la flore microbienne contaminante apporte des informations sur les sites potentiels de la contamination et les vecteurs de sa dissémination dans l'usine de production. Ces informations sont de puissants outils d'aide à la décision implémentables dans des procédures de gestion des risques du type HACCP (*Hard Analysis Critical Control Point*). Certaines bactéries, comme *Bacillus cereus* ou *Clostridium botulinium*, sont des contaminants récurrents des chaînes de production alimentaire ; l'origine du souillage par ce type de bactéries endémiques doit être déterminée pour incriminer les matières premières. Ainsi, Scheinbach et Hong ont montré que des génotypes de *Salmonella enterica* et *Escherichia coli* étaient persistants sur de longues périodes dans les usines agro-alimentaires [18]. L'endémicité de *S. aureus* dans une usine de traitement de la volaille est à l'origine de la contamination des carcasses dont la flore en *S. aureus* a été modifiée [19]. L'évaluation de points critiques relatifs aux sites d'entrée putatifs du contaminant ainsi qu'aux process favorisant les contaminations croisées, a été validée par des méthodes de typage [20]. Les techniques de typage sont donc des outils puissants applicables à la traçabilité bactérienne en agro-alimentaire dans le cadre de l'HACCP [21].

c. Traçabilité des agents bioterroristes

La crainte du bioterrorisme, diffusion délibérée d'agents biologiques susceptibles de provoquer une maladie mortelle, pousse également à rechercher des outils performants pour identifier les germes dispersés dans l'environnement, retrouver l'origine des souches et comprendre les éventuels trafics de ces armes biologiques [22]. Heureusement de tels évènements sont rares, puisque l'on cite toujours le siège de la ville de Caffa en Crimée [22]. Colonie génoise d'importance commerciale majeure, Caffa fut assiégée en 1346 par les Tatars alors que la peste se déclarait dans leurs rangs. Décimés par la maladie et sur le point de battre en retraite, le général mongol Djanibeğ donna l'ordre de catapulter les soldats morts par-dessus les remparts de la cité. La peste éclata à Caffa et les Tatars réussirent à s'emparer de la ville. La diaspora génoise a véhiculé la peste sur tout le pourtour méditerranéen en 1347 et la deuxième épidémie de peste noire, emportant plus d'un tiers de la population de l'Europe occidentale en cent ans, est déclarée en 1348. Au 20ème siècle, l'utilisation des armes biologiques est étroitement liée à la guerre, notamment à partir de la Seconde Guerre Mondiale pendant laquelle les Japonais sont suspectés d'avoir largué des puces porteuses de l'agent de la peste sur des villes chinoises. Les armes biologiques sont davantage susceptibles d'être utilisées aujourd'hui comme menaces terroristes mises en application lors d'attentats. La secte Rajneesh a, en 1984, cultivé une souche de salmonelle et dispersé le pathogène dans les bars à salades de l'Oregon provoquant

l'intoxication alimentaire de 750 personnes [23]. Cet acte fut considéré comme le premier attentat terroriste bactériologique aux États-Unis. L'envoi d'enveloppes piégées à l'anthrax en 2001 représente le premier attentat bioterroriste ayant eu un impact planétaire majeur en dépit du nombre relativement modeste de victimes [23]. La souche utilisée, *Ames*, est issue d'une lignée rare aux Etats-Unis et était utilisée par un certain nombre de laboratoires de ce pays [24,25]. L'identification de la souche incriminée par typage a été décisive dans l'orientation de l'enquête criminelle.

d. Traçabilité industrielle

Le génotypage des bactéries d'intérêt industriel s'éloigne de la thématique épidémiologique mais représente un enjeu majeur. Certaines souches bactériennes sont utilisées comme probiotiques, d'autres comme agents de vaccination et d'autres encore sont de véritables usines à production de métabolites industriels. L'étude de la diversité et de la variabilité génétique des souches d'intérêt permet une meilleure connaissance des espèces correspondantes et peut guider dans la recherche de souches ayant des propriétés nouvelles. Le génotypage fournit des outils commodes pour la traçabilité de ces souches industrielles dans le cadre d'un contrôle qualité ou pour certifier une propriété industrielle. Ainsi, le suivi de la chaîne de production par échantillonnage raisonné de différents lots est une mesure de contrôle qualité qui permet de maintenir l'intégrité des produits et de déterminer si des changements génotypiques ont eu lieu pendant la production, le conditionnement ou le stockage de ces produits biologiques.

2.3. Biosurveillance des agents pathogènes

a. Définition

La surveillance épidémiologique comprend la collecte, l'analyse et la diffusion systématique des données sanitaires pour la planification, l'exécution et l'évaluation des programmes de santé publique [8]. Les méthodes de typage contribuent à enrichir les enquêtes de surveillance épidémiologique des maladies infectieuses en attribuant un profil génotypique et/ou phénotypique à tous les isolats cliniques et environnementaux collectés et typés. Ce recensement permet d'évaluer la distribution environnementale des différents types au sein d'une espèce bactérienne qui pourra alors être corrélée avec la distribution des isolats cliniques. La diversité des isolats cliniques est souvent réduite par rapport à celle des isolats d'origine environnementale. Les souches responsables de pathologie humaine sont généralement peu retrouvées dans la nature et ont développé des mécanismes de survie, de colonisation et de résistance à l'hôte [26]. La surveillance épidémiologique prend en charge le suivi des maladies zoonotiques et notamment les modalités de transfert de souches de l'animal à l'homme. La spécificité d'hôte de certaines souches ajoutée au typage peut permettre de déterminer l'origine d'une contamination microbienne.

b. Un outil d'aide à la décision

La surveillance des agents pathogènes est un véritable outil d'aide à la décision qui permet de révéler l'émergence de certains clones ou de les corréler avec un pouvoir pathogène particulier.

A l'instar d'une épidémie, l'émergence correspond à l'apparition soudaine, localisée et imprévue d'un nouveau clone dont le profil atypique est décelé par typage bactérien. Par exemple, la souche Lorraine de *L. pneumophila* n'avait jamais été identifiée avant 1995 et émerge depuis 2001 [27]. Initialement identifiée dans l'est de la France, elle est présente à l'échelle européenne. A l'inverse des clones épidémiques, les clones émergents sont persistants et parviennent à s'établir dans une niche écologique. La détection et la surveillance des clones émergents nécessitent de disposer de bases de données fournies, robustes, internationales afin de justifier la nouveauté du type suspecté émergent et de suivre sa dissémination. L'émergence est rendue possible par l'acquisition d'une nouvelle fonction exploitée par la bactérie pour accroître sa capacité adaptative, ou *fitness*, dans une niche particulière. La plasticité génomique bactérienne favorise l'émergence [28]. L'émergence est ainsi souvent décrite comme un signal d'alerte. En effet, les souches ou clones émergents ont potentiellement un nouveau bagage génétique influant sur la pathogénicité ou la résistance aux antibiotiques. Par exemple, la souche Lorraine responsable de 10 et 25% des cas de légionelloses diagnostiqués en France et au Royaume-Uni respectivement est qualifiée d'hyper-virulente.

Le pouvoir pathogène ou pathogénicité est l'ensemble des mécanismes conditionnant une maladie induite par un agent infectieux [8]. Cette aptitude est fonction de plusieurs paramètres principalement liés aux facteurs de l'hôte, aux conditions environnementales et à la dose, au pouvoir invasif et à la virulence de l'agent biologique. La virulence bactérienne est liée à la synthèse de macromolécules appelées facteurs de virulence interférant avec des fonctions physiologiques de l'organisme infecté, aux niveaux moléculaire, cellulaire et tissulaire. Corréler génotype et virulence s'avère donc, au premier abord, une démarche périlleuse. Cependant, les épidémiologistes observent pour la majorité des espèces bactériennes suivies un lien entre génotype et virulence. Par exemple, Jaureguy et collaborateurs ont montré que l'ensemble de facteurs de virulence d'*E. coli* était différent selon les complexes clonaux [29]. La diversité limitée des souches cliniques est un facteur important pour corréler le génotype avec une capacité à déclencher une pathologie. Le génotype est de toute évidence, dans certains cas du moins, corrélé avec la virulence mais ne peut suffire à la prédire du fait par exemple de l'hétérogénéité génétique de l'hôte.

3. Structure des populations et épidémiologie

La notion d'espèce bactérienne permet de rendre compte d'une certaine réalité biologique qui s'exprime dans une structure de population. La diversité intra-espèce est très variable selon l'espèce considérée. L'épidémiologiste doit appréhender ces particularités génétiques et les considérer lors de

ses investigations : quel est le moteur de la diversité bactérienne, comment sont structurées les populations bactériennes, en quoi les considérations de génétique des populations sont-elles essentielles à l'épidémiologie des maladies infectieuses ?

3.1. Le modèle clonal remis en question

Les populations bactériennes doivent satisfaire deux processus nécessaires pour assurer leur survie : conserver l'information génétique tout en s'adaptant à des environnements changeants. La conception clonale a longtemps permis de comprendre comment se conservait et se pérennisait l'information génétique en l'absence de reproduction sexuée [30]. Les bactéries sont des organismes asexués qui se reproduisent par scissiparité et dont l'information génétique est, *a priori*, transmise intégralement à la descendance, aux mutations *de novo* près. La découverte des mécanismes horizontaux de transfert génétique a fait vaciller la conception d'une évolution strictement clonale. Des évènements génomiques extrinsèques contribuent à la plasticité génomique bactérienne et favorisent l'adaptation rapide à de nouvelles niches écologiques. Chez de nombreuses espèces bactériennes, les génomes sont des mosaïques ayant incorporé du matériel génétique étranger par transfert horizontal. Par opposition au transfert vertical, héritage du matériel génétique par scissiparité, le transfert horizontal est caractérisé par l'intégration, dans un génome bactérien, de matériel génétique libre, provenant d'éléments mobiles (virus tempérés) ou d'autres bactéries. La transmission horizontale de l'information génétique est réalisée selon trois mécanismes connus : la transformation qui permet l'internalisation d'une molécule d'ADN libre, la transduction faisant intervenir un bactériophage pour véhiculer l'ADN transformant, et la conjugaison nécessitant le contact entre deux micro-organismes. Le transfert horizontal apparaît ainsi être un facteur majeur de l'évolution des bactéries à même de contribuer à l'adaptabilité. Le *fitness* bactérien dans une nouvelle niche écologique peut par exemple être amélioré par l'acquisition de matériel génétique issu de bactéries déjà adaptées à cet écosystème (acquisition de gènes de résistance aux antibiotiques).

3.2. Structures des populations bactériennes

Les populations bactériennes ne sont pas homogènes, plusieurs structures pouvant coexister au sein d'une espèce (Figure 1). Le généticien britannique John Maynard Smith a défini les structures de populations bactériennes fondamentales en caractérisant trois structures possibles, de strictement clonale à panmictique, selon l'importance des transferts horizontaux et la nature de leur écosystème [31]. Plusieurs méthodes déterminent la structure d'une population bactérienne, la plus utilisée consistant à évaluer l'indice de déséquilibre de liaison entre les allèles des marqueurs étudiés [32]. Cet indice estime le degré d'association non aléatoire des allèles à différents loci. En l'absence de recombinaison chromosomique, les loci étudiés ne sont pas distribués aléatoirement au sein de la population : un déséquilibre de liaison est observé entre les marqueurs suggérant alors la clonalité de

l'échantillon étudié. L'absence de recombinaison entre des populations résulte de barrières biologiques empêchant tout échange génétique ou tient à l'isolement géographique ou écologique de ces populations.

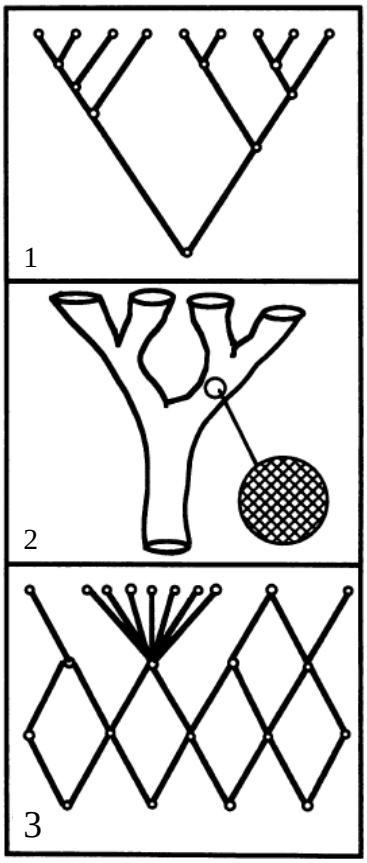


Figure 1. Modèles de structures de populations bactériennes (extrait de [31])

Population clonale, aucune recombinaison entre les individus asexués (ex. : *S. aureus*, *E. coli*)
Population à structure mixte, des évènements de recombinaison ont lieu entre les individus appartenant à la même lignée (ex. : *L. pneumophila*)
Population panmictique, représentation en réseau de la relation de parenté entre les individus sexués qui échangent du matériel génétique (ex. : *P. aeruginosa*, *H. pylori*)

a. Clonalité

L'asexualité apparente des bactéries, l'endémicité de certains complexes clonaux ainsi que la relation étroite entre génotypes et pathogénicité et l'intérêt apporté à l'étude des microorganismes pathogènes ont conduit à surestimer l'importance de la clonalité bactérienne [30]. Certaines populations bactériennes sont caractérisées par la présence d'un nombre limité de génotypes ou clones largement répandus et de nombreux génotypes rares. Dans les populations à structure clonale, ces clones prédominants appartiennent à des complexes clonaux dont la divergence radiale a été initiée à partir d'un génotype fondateur. Les bactéries clonales sont structurées exclusivement en complexes clonaux qui ont divergé par dérive génétique *i.e.* par accumulation de mutations neutres, stochastiques et héritées [33]. Les évènements de transferts horizontaux sont supposés quasi inexistants chez ce type de population bactérienne au fort déséquilibre de liaison. Cependant, des traces de transfert horizontal

ont été mises en évidence par analyse de séquences chez *E. coli* [34] et *Salmonella* [35], bactéries considérées clonales. La fréquence de ces évènements ne permet pas de supprimer le déséquilibre de liaison et ne modifie pas considérablement la structure clonale de leurs populations. Les régions génomiques d'un individu membre d'une population clonale partagent une même histoire évolutive ; les relations phylogénétiques inférées à partir de différents gènes sont donc congruentes. Cette corrélation évolutive, un des arguments de l'hypothèse clonale d'une population, est cependant à nuancer car elle n'est valable que pour les gènes de ménage peu mutables et dont la stabilité est maintenue par sélection naturelle [36].

Ces populations clonales sont donc structurées en groupes stables bien caractérisés permettant le suivi épidémiologique global. La diversité au sein de ces complexes clonaux résulte essentiellement de l'apparition de mutations *de novo*. L'épidémiologiste devra adapter son outil de typage pour que les individus de ces complexes puissent être distingués.

b. Panmixie

Certaines espèces de bactéries, telles quel *Neisseria gonorrhoeae* et *Helicobacter pylori*, évoluent principalement par transfert horizontal et recombinaison [37,38]. Les échanges génétiques au sein de ces populations, dites panmictiques, sont si fréquents que les loci sont tous à l'équilibre de liaison. La panmixie est très élevée chez *H. pylori.* Les recombinaisons sont 200 à 300 fois plus fréquentes que les mutations ponctuelles, 50% du génome d'*H. pylori* ayant pu être remplacés par recombinaison homologue durant son évolution [39]. Chaque isolat typé représente un génotype unique. L'histoire évolutive et l'établissement de relations de parenté au sein de populations panmictiques ne peuvent être déduits selon un modèle classique *i.e.* clonal [30]. Une phylogénie des génomes n'est pas concevable pour ce type de bactéries : une représentation en réseau, mettant en évidence les réticulations phylogénétiques dues aux transferts horizontaux, est préférée. Le suivi global et à long terme de ce type de pathogènes est impossible. Les déterminants qui qualifient chaque souche bactérienne mutent de manière permanente empêchant ainsi l'étude de sa distribution et répartition. Néanmoins, les vitesses de mutation demeurent raisonnables pour que l'isolat incriminé puisse être tracé localement, lors d'une investigation épidémique par exemple.

L'acquisition de nouveaux allèles par transfert horizontal peut conférer un pouvoir adaptatif et conduire à l'émergence d'un clone épidémique. Une population à structure panmictique pourra générer, en situation épidémique, des clones particuliers adaptés à leurs hôtes : on parlera de structure panmictique épidémique. Ces clones épidémiques peuvent persister dans le temps. L'observation de cette apparente clonalité caractéristique d'une population épidémique peut induire en erreur sur la structure de cette population. Le clone particulièrement adapté à un hôte profitera de son pouvoir adaptatif pour disséminer rapidement conduisant alors artéfactuellement à un déséquilibre de liaison.

c. Structure mixte

La dichotomie clonale/panmictique n'est pas la norme pour comprendre les structures de populations bactériennes. Un modèle intermédiaire correspondrait vraisemblablement à une réalité biologique plus adaptée pour la plupart des espèces bactériennes. En effet, la structure de la majorité des espèces bactériennes ne rentre pas dans les cadres rigides de la clonalité ou de la panmixie : on parle alors de structure mixte. Ainsi, les bactéries évoluent classiquement par expansion clonale entrecoupée d'évènements de recombinaison. Un déséquilibre de liaison sera alors observé dans une population constituée de plusieurs branches indépendantes au sein desquelles des évènements de recombinaison ont lieu. La divergence des différentes branches résulte d'une séparation clonale médiée par sélection naturelle ou dérive génétique (Figure 2). L'expansion clonale résulte généralement de l'adaptation d'un mutant à une nouvelle niche écologique ou à un nouvel hôte [40]. Ce mutant adapté fonde alors un complexe clonal voire, le cas échéant, un nouvel écotype *i.e.* un groupe de microorganismes génétiquement similaires adaptés à une niche écologique. Ces différents écotypes à structure clonale persistent et ne peuvent donc être confondus avec des clones épidémiques éphémères [40,41]. *L. pneumophila* est une espèce caractéristique de cette structure mixte, une dizaine de complexes clonaux s'apparentant à des écotypes étant clairement identifiés [42,43,44]. Par ailleurs, des écotypes adaptés à une nouvelle niche écologique peuvent ensuite diverger et conduire à la définition de nouvelles espèces [45]. Ainsi, le complexe *M. tuberculosis* pourrait avoir émergé de *M. canetti*, selon le modèle de l'écotype, ce qui se traduit par une apparence de réduction génétique postérieure à un goulot d'étranglement [46]. Ces complexes clonaux émergés à partir d'une population recombinante sont biologiquement différents. Edward J. Feil suggère d'affiner cette notion en remplaçant le concept de génome noyau ou *core genome*, sur lequel s'appuient certaines caractérisations génétiques de l'espèce bactérienne, par celui de *clonal core i.e.* l'ensemble des gènes communs à chaque complexe clonal [40].

L'épidémiologie des pathogènes à structure mixte est plus problématique. A l'instar des bactéries panmictiques, l'investigation locale est facilitée par la diversité des isolats. Cependant, lors de l'émergence récente d'une lignée à expansion clonale, la diversité peut être limitée : l'épidémiologiste sera alors dans une situation de clonalité. Lorsque les lignées sont établies, ses membres auront pu diverger par recombinaison : l'épidémiologiste sera alors dans une situation de panmixie. Dans ce cadre particulier mais finalement généralisable à la majorité des espèces bactériennes, l'épidémiologiste doit ajouter un élément de complexité à son enquête. Pour conclure, l'épidémiologie des maladies infectieuses est une science complexe largement influencée par la génétique des populations bactériennes et nécessitant une bonne connaissance de celles-ci avant toute investigation.

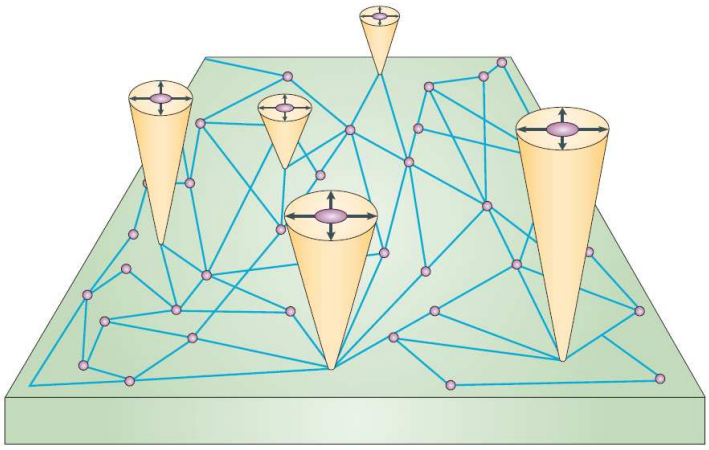


Figure 2. Emergence de complexes clonaux au sein d'une population panmictique (extrait de [40] avec la permission de Nature Publishing Group, division de Macmillan Publishers)

II) Polymorphisme chez les bactéries

Les principaux enjeux de l'épidémiologie des maladies infectieuses ont précédemment été mis en lumière et l'importance du typage bactérien comme support de l'investigation épidémiologique a été soulignée. Typer les isolats bactériens vise à pouvoir les distinguer s'il y a lieu. L'identification de sources de variabilité ou polymorphismes permet de répondre à ce besoin. Quelles sont les différentes sources de polymorphisme ? Comment sont-elles utilisées par les méthodes de typage ? Les objets polymorphes doivent-ils être adaptés à la question épidémiologique ? Je ne discuterai ici que des polymorphismes génétiques qui sont à la base des méthodes de génotypage. Faute de mieux, les premières techniques de typage tiraient parti de polymorphismes phénotypiques et permettaient la distinction d'un type défini au sein d'une espèce bactérienne. Ainsi plusieurs familles de types pouvaient être définies selon les caractères utilisés : biotypes (caractères biochimiques), sérotypes (caractères antigéniques), pathotypes (caractères de pathogénicité), zymotypes (profils enzymatiques), antibiotypes (sensibilité aux antibiotiques) ou lysotypes (sensibilité aux bactériophages). Ces méthodes de typage ne sont pratiquement plus utilisées en général pour des raisons de coûts et de manque de pouvoir discriminant. L'utilisation de techniques moléculaires permettant la production d'empreintes génétiques a augmenté considérablement la puissance des méthodes de typage bactérien.

1. Les objets génétiques polymorphes

L'essor de la génomique a permis de révéler l'existence de structures génétiques dynamiques et/ou mobiles dans les génomes procaryotes. Ces structures mutatrices représentent des sources de polymorphisme entre les individus bactériens. Ici, les principaux objets génétiques polymorphes seront

brièvement décrits. Cinq grandes familles d'éléments interviennent dans la plasticité génomique et constituent l'essentiel des sources de polymorphisme : les transferts latéraux d'éléments extrachromosomiques**,** les réarrangements chromosomiques, les mutations ponctuelles, les séquences répétées dispersées et les séquences répétées en tandem.

1.1. Eléments extrachromosomiques transférés latéralement

Tous les éléments transférés latéralement sont susceptibles de représenter une source de polymorphisme intéressante. Dans la pratique, la présence/absence de deux familles d'éléments est particulièrement recherchée : les cassettes de gènes de résistance aux antibiotiques et les plasmides. Lorsque la résistance aux antibiotiques est acquise par transfert horizontal, elle peut se propager dans la population. En considérant plusieurs antibiotiques, une population bactérienne représentera donc une véritable mosaïque dont les spectres de résistance aux antibiotiques sont très variables. De même, la présence d'un plasmide, transmis par mécanisme conjugatif, au sein d'une espèce bactérienne est un mode de discrimination utilisé mais de faible valeur. Ces plasmides sont d'ailleurs, souvent, porteurs de gènes de virulence ou de résistance aux antibiotiques.

1.2. Réarrangements chromosomiques

Les réarrangements chromosomiques sont des mutations modifiant la structure et l'organisation des génomes. Seuls les réarrangements susceptibles de créer du polymorphisme observable au sein d'une espèce bactérienne seront ici détaillés : la transposition, les délétions et les insertions.

a. Transposition

Décrits dans la théorie du gène égoïste, proposée par Richard Dawkins en 1976 [47], les éléments mobiles sont des structures génétiques ayant la faculté de produire des copies identiques de leur séquence et de les disperser dans leur génome hôte assurant ainsi leur pérennité. Les principaux éléments mobiles caractérisés sont les éléments transposables ou transposons (T_{ADN}). Plusieurs familles de T_{ADN} ont été recensées et varient autant dans leurs mécanismes transpositionnels que dans leurs structures et leurs choix des séquences cibles pour l'insertion. Les séquences d'insertion ou IS sont les transposons les plus simples et les plus largement retrouvés chez les bactéries. Les IS sont des éléments transposables de petite taille (de 750 à 2500 pb) présentant une organisation très compacte. Elles sont universellement retrouvées dans le phylum procaryote. Une IS présente à ses extrémités de courtes séquences inversées répétées (de 10 à 40 pb), séquences d'ADN actives en recombinaison, qui encadrent une phase ouverte de lecture codant pour la transposase (TnpA) de l'élément [48]. La réaction de transposition constitue le cœur de la réplication et de la propagation des séquences d'insertion. La dynamique de ces éléments mobiles résulte donc en un polymorphisme de nombre et de position d'insertion. Une particularité importante des IS est qu'elles peuvent collaborer entre elles lors

de leur transposition. Cette propriété permet de mobiliser toute séquence d'ADN comprise entre deux copies d'une même IS, en orientation directe ou inverse. La structure déplacée est appelée transposon composite, les plus connus étant Tn5 (IS50), Tn9 (IS1) et Tn10 (IS10) auxquels sont associés des gènes de résistance à un antibiotique. La collaboration entre deux copies d'une même IS peut provoquer des remaniements multiples et complexes des génomes en fonction des extrémités impliquées et de la localisation du site cible [49]. Ces transposons peuvent ainsi conduire à l'insertion et à la délétion de larges portions génomiques. La transposition est divisée en deux grandes catégories, réplicative ou conservative. Durant la transposition conservative, l'élément mobile ne se réplique pas : c'est le principe du couper-coller [49]. Ce type de transposition conduit à des évènements de délétion. Au cours de la transposition réplicative, l'élément mobile subit une étape de réplication selon le mode du copier-coller. La duplication de segments d'ADN peut résulter de ce type d'évènements [49].

b. Délétions et duplications

Les évènements de délétion et de duplication de segments d'ADN sont des éléments essentiels de la plasticité génomique. Ils constituent également une source de polymorphisme importante ; 10% des mutations neutres sont des délétions et le taux d'apparition de ces évènements est de 10^{-6} par génération [50]. Les régions de délétion sont d'ailleurs des marqueurs de polymorphisme importants notamment utilisés pour inférer des phylogénies. Les mutations par délétion et duplication sont principalement dues aux évènements de transpositions (IS et transposons composites) et à la dynamique des prophages [51]. D'autres mécanismes peuvent expliquer ces macrolésions génomiques. Lors de la réplication, un mécanisme de recombinaison inégale recA-dépendant peut conduire à la délétion ou la duplication de régions bornées par de courtes séquences répétées. Par ailleurs, des mécanismes de type recA-indépendants impliquant des erreurs réplicatives causées par un bégaiement de la polymérase, des cassures double-brin résolues par des exonucléases ou des erreurs introduites lors de l'enroulement de l'ADN ont été proposés [50].

1.3. Mutations ponctuelles

Les mutations ponctuelles désignent l'ensemble des modifications qui affectent de manière spécifique un point particulier du génome bactérien. Ces mutations conduisent alors à l'observation d'un polymorphisme nucléotidique ou SNP (*Single Nucleotide Polymorphism*). *Stricto sensu,* un SNP se rapporte au changement d'un nucléotide par un autre : on parle de transversion lorsqu'une purine est substituée en pyrimidine et inversement, et de transition lorsque la substitution maintient le type nucléotidique. Par extension, le polymorphisme nucléotidique est également caractérisé lorsque des évènements d'insertion ou délétion d'un seul nucléotide, communément nommés indels, ont lieu. La modification pourra éventuellement altérer le site de reconnaissance d'une enzyme de restriction : on parlera alors de polymorphisme de restriction, concept central des techniques de génotypage de

première génération. Selon la zone génomique au sein de laquelle se produit la mutation, on distinguera trois familles de SNPs : les SNPs intragéniques synonymes, non synonymes, et les SNPs intergéniques. Les SNPs intergéniques affectent les zones intergéniques non-codantes du génome. Ces polymorphismes n'ont, en général, pas de répercutions directes sur le phénotype bactérien et subissent, en cela, une pression sélective faible. Ces mutations neutres sont plus fréquentes que celles qui affectent directement la séquence nucléotidique d'un gène. A l'inverse, les SNPs situés au sein des régions codantes altèrent la structure du codon associé pouvant éventuellement par voie de conséquence modifier l'acide aminé codé. Cependant, le code génétique est redondant et, selon la modification du codon, les deux classes de SNPs codant, synonymes et non synonymes, ont été définis. Les SNPs non-synonymes résultent en des changements d'acides aminés ou en l'introduction de codons-stop pouvant provoquer un changement de phénotype. A l'opposé, les SNP synonymes sont caractérisés par un changement redondant de codon, le même acide aminé est donc codé n'altérant pas la séquence de la protéine.

1.4. Séquences répétées dispersées

Les séquences répétées peuvent être définies comme des zones génomiques présentant une forte similarité de séquences avec d'autres zones situées sur le même génome. Les séquences répétées dispersées sont ubiquitaires chez les procaryotes et caractérisées par une grande diversité structurelle et fonctionnelle. Leur polymorphisme résulte d'une variation du nombre d'unités répétées et de leur position dans le génome. Quatre grandes familles, dont le polymorphisme est beaucoup étudié, seront ici détaillées : les transposons, les MITEs, les REPs et les CRISPRs. Ces quatre types de structures n'ont pas d'histoire commune, de propriétés structurelles identiques ou de fonctions biologiques similaires mais sont ici regroupées en raison de leur caractère répété dispersé.

Par définition, les éléments transposables sont des éléments répétés dispersés. Ils possèdent également les caractéristiques d'objets polymorphes par réarrangement chromosomique et ont donc fait l'objet d'une description ci-avant. La translocation de ces structures leur confère une grande capacité d'envahissement génomique : certains transposons sont très grands et peuvent contenir plusieurs gènes altérant de manière significative la synthénie de l'espèce. Les transferts et réarrangements des transposons ont un impact direct sur la pathogénicité bactérienne lorsque la transposition implique par exemple des gènes de résistance aux antibiotiques ou des îlots de pathogénicité.

La deuxième famille d'éléments répétés dispersés correspond aux séquences MITEs (*Miniature Inverted repeat Transposable Elements*). Leur formation résulte d'une défaillance du système de transposition. Lors du mécanisme transpositionnel réplicatif, un défaut des systèmes de réparation de l'ADN peut mener à la fusion des extrémités du T_{ADN} conduisant à la formation de structures de type MITEs. Le lien supposé entre les MITEs et les T_{ADN} a été suggéré en premier lieu

par la mise en évidence de deux séquences inversées répétées à l'extrémité des MITEs encadrant une région, d'environ 100pb, présentant une similarité de séquences avec l'IS630 chez *Brevibacterium lactofermentum* [52], *Neisseria* [53] et *Streptococcus pneumoniae* [54]. Les MITEs sont des structures multifacettes dont plusieurs familles ont été caractérisées chez les bactéries : ERIC (*Enterobacterial Repetition Intergenic Consensus*), RUP (*Repeat Unit of Pneumococcus*), NEMIS (*Neisseria Miniature Insertion Sequences*) ou BOX. Ils sont impliqués dans de nombreuses fonctions métaboliques : régulation transcriptionnelle, modification post-transcriptionnelle, interaction avec l'hôte.

La troisième famille d'éléments répétés dispersés polymorphes est composée des séquences REP (*Repetitive Extragenic Palindromic*). Initialement identifiées chez *E. coli* [55], les séquences REP, dont le motif répété est d'environ 35 pb, sont des séquences palindromiques présentes en tant qu'unité seule et indépendante ou structurées en *clusters* nommés BIME pour *Bacterial Interspaced Mosaic Elements* [56]. La bactérie *E. coli* possède près de 1000 copies de ces éléments REP, constituant plus de 1% de son génome. Leur caractère extragénique et leur propriété à former des structures de type tige-boucle amènent à évoquer l'implication des REPs dans des mécanismes régulateurs divers [57,58].

A l'inverse des MITEs et des REPs, les répétitions formant la quatrième famille sont regroupées en *clusters* appelés CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*). L'organisation des loci CRISPR a été décrite : une séquence *leader* est suivie de répétitions directes conservées courtes ou *direct repeat* de 24-48 pb séparées par des séquences uniques nommées *spacer* de 26-72 pb. Le polymorphisme des CRISPRs repose sur la diversité de leur structure (présence/absence des différents spacers). La similarité de séquence entre les spacers et des séquences phagiques ou plasmidiques a conduit à penser que ces systèmes pourraient intervenir comme mécanisme de résistance contre l'ADN étranger. Cette hypothèse a été validée expérimentalement en testant la réponse de *Streptococcus thermophilus* à une attaque virale [59], le polymorphisme des loci CRISPR influant sur le phénotype de résistance. Les CRISPRs constitueraient ainsi un élément majeur de réponse immunitaire adaptée de la bactérie et une clé centrale de la compréhension de la coévolution phages/bactéries.

1.5. Séquences répétées en tandem

a. Les VNTRs, des structures répétées en tandem polymorphes

Par opposition aux séquences répétées dispersées, les répétitions en tandem sont, par définition, constituées par la répétition en tandem, *i.e.* adjacentes et dans un même sens, d'un motif d'ADN. Initialement décrite chez les eucaryotes, les séquences répétées en tandem forment des familles de répétitions souvent distinguées en deux grandes classes : les microsatellites et les minisatellites. Les séquences répétées, intra ou inter-géniques sont différenciées par la taille du motif répété, de 1 à 9pb pour les microsatellites et plus de 9pb pour les minisatellites. Les microsatellites

sont parfois également nommés STR (*Short Tandem Repeat*) lorsque le motif répété est inférieur à 4pb. Les raisons de cette classification structurelle reposent dans une certaine mesure sur l'hétérogénéité des mécanismes d'instabilité sous-jacents. Certaines séquences répétées en tandem présentent un polymorphisme de répétitions intra-espèce. Ces séquences polymorphes, qu'elles soient de type micro ou minisatellites, sont appelées VNTRs pour *Variable Number of Tandem Repeats*. Selon les génomes bactériens, les VNTRs peuvent représenter de 2 à 10% de l'ensemble des répétitions en tandem [60].

b. Identification du polymorphisme des VNTRs

Alors que Wyman et White recherchaient des loci utilisables en cartographie génétique, ils identifièrent un fragment chromosomique qui présentait un polymorphisme de longueur dans la population humaine [61]. Cinq ans plus tard, ce fragment a été caractérisé et a révélé la présence d'un locus minisatellite nommé MS32 [62].

Ces dernières années, la disponibilité des données de génomes complets couplée à l'utilisation d'outils de bioinformatique a permis l'identification systématique et exhaustive, *in silico*, des structures répétées en tandem. Le logiciel TRF, pour *Tandem Repeats Finder*, créé par Gary Benson en 1999 [63] assure le traitement de longues séquences nucléotidiques et la détection de structures répétées en tandem. La détection des VNTRs implique obligatoirement la comparaison des structures répétées identifiées dans plusieurs souches d'une même espèce afin de savoir si le locus considéré est polymorphe. Lorsque plusieurs génomes de différentes souches d'une même espèce ont été séquencés, cette recherche comparative peut être effectuée *in silico* grâce par exemple à l'utilisation de bases de données. L'Institut de Génétique et Microbiologie a développé la base minisatellites.u-psud (http://minisatellites.u-psud.fr/GPMS/) [64]. Cette base de données utilise le logiciel TRF pour l'analyse préalable des génomes et facilite les requêtes de répétitions en tandem dans les séquences de génomes en accès libre. Cette base propose une liste de caractéristiques des répétitions qui peuvent être interrogées par l'utilisateur (longueur de la répétition en tandem, taille de l'unité répétée, nombre de répétitions, conservation des motifs par rapport au motif consensus, position sur le génome, pourcentage de G/C). Elaborée en 2001, la base a été améliorée trois ans plus tard par l'ajout d'une nouvelle fonctionnalité : la recherche comparative [60]. Ce module permet de comparer deux à deux des génomes séquencés de la même espèce afin d'identifier les répétitions en tandem polymorphes. Le polymorphisme d'une répétition en tandem doit néanmoins toujours être vérifié expérimentalement, l'utilisation des nouvelles techniques de séquençage étant source d'erreur dans l'évaluation du nombre de copies.

2. Diversité des méthodes de génotypage

Ces sources de polymorphisme constituent le cœur des techniques de génotypage. Compte tenu de la diversité des polymorphismes exploitables, un nombre important de méthodes a été décrit. Cependant, à l'image des techniques phénotypiques, certaines techniques de génotypage développées ces dernières années sont aujourd'hui obsolètes. Cette multiplicité technologique crée de la confusion et ne contribue pas à la standardisation et l'échange de données entre laboratoires. Quelles sont ces méthodes ? Quels sont les critères de sélection ?

2.1. Profils de bandes

Avant l'ère du séquençage en masse des génomes bactériens, les techniques moléculaires de génotypage reposaient sur la génération et la comparaison de profils génétiques. Ces techniques de génotypage de première génération sont séparées en deux grandes familles. Durant les années 80, des méthodes fondées sur la restriction d'ADN ont été proposées ; la révolution apportée par la PCR en 1985 a permis, dans les années 90, le développement de méthodes basées sur l'amplification génétique.

a. Profils produits par restriction d'ADN

Ces méthodes reposent essentiellement sur l'analyse du polymorphisme des fragments générés par digestion enzymatique et détectés par hybridation de séquences spécifiques (RFLP ou *Restriction Fragment Length Polymorphism*) ou par différences à des sites de macrorestriction révélée par électrophorèse en champ pulsé (PFGE ou *Pulsed Field Gel Electrophoresis*).

La technique RFLP consiste en une digestion de l'ADN génomique total par des enzymes de restriction à nombre élevé de sites de coupure, une migration par électrophorèse des fragments d'ADN digérés suivi d'un transfert sur membrane, puis, enfin, une hybridation spécifique d'un fragment génomique ciblé par une sonde complémentaire de sa séquence [65]. La technique RFLP est en mesure de détecter à la fois un polymorphisme de restriction résultant d'une mutation ponctuelle et un polymorphisme de répétitions et/ou de distribution de la séquence ciblée. Le pouvoir discriminant de cette technique dépend principalement de la séquence ciblée par la sonde, déterminant ainsi la variabilité des fragments détectés. Deux types de cibles sont généralement détectées : IS-RFLP (ciblage des séquences d'insertion, exemple du typage IS6110 de *M. tuberculosis*) ou ribotypage (ciblage du locus de l'ADNr 16S-23S, automatisation par le RiboPrinter).

Décrite en 1984 par les généticiens Charles Cantor et David Schwartz, la technique PFGE ou *Pulsed Field Gel Electrophoresis* permet l'analyse des profils de macrorestriction de l'ADN total par électrophorèse en champ pulsé. Le résultat est un profil de restriction définissant un pulsotype caractérisant l'isolat bactérien étudié. Des essais de standardisation et d'interprétation des profils ont

été réalisés afin de déterminer la relation de parenté entre les isolats étudiés [66]. Malgré la technicité demandée et le défaut de reproductibilité, sauf si des protocoles parfaitement standardisés ont été mis en place (PulseNet), la PFGE demeure la technique de typage standard pour de nombreuses espèces bactériennes bénéficiant de son pouvoir discriminant élevé et du réseau de laboratoires référents, PulseNet, utilisant cette méthode.

b. Profils produits par amplification génétique

La seconde famille de méthodes de génotypage a exploité la puissance apportée par la PCR. Ces méthodes produisent des profils génétiques par amplification PCR au hasard (RAPD ou *Random Amplified Length Polymorphism*), amplification PCR spécifique directe (REP-PCR, BOX-PCR, ERIC-PCR) ou après restriction d'ADN (AFLP ou *Amplified Fragment Length Polymorphism*).

Proposée simultanément en 1990 par Welsh et McClelland [67] et Williams [68], la RAPD repose sur l'amplification aléatoire de différentes portions du génome permise par l'utilisation d'amorces aléatoires courtes s'hybridant à faible température et permettant, le cas échéant, l'amplification d'un fragment d'ADN. Le polymorphisme résulte du nombre et de la localisation des sites d'hybridation des amorces, variables par mutation ponctuelle, distinguant plusieurs allèles au sein de la même espèce. Le manque de reproductibilité et de standardisation ainsi que l'hybridation partielle des amorces formant des *smears* ininterprétables sur gel d'agarose, ont conduit au développement de méthodes dérivées, utilisant des amorces spécifiques. Ces méthodes s'appuient sur le polymorphisme de répétitions des séquences répétées dispersées qui constituent des marqueurs de choix. Des amorces correctement définies ciblent ces séquences répétées, dont la nature diffère selon l'espèce analysée (REP, BOX, ERIC, …), et autorisent l'amplification de fragments. Le nombre et la taille des amplicons sont fonction de la position et du nombre de séquences répétées dispersées sur le génome. Malgré les efforts de standardisation entrepris (diversilab de Biomérieux utilisant la technique de REP-PCR), ces méthodes, quoique discriminantes, sont peu utilisées principalement en raison de leur faible reproductibilité inter-laboratoire.

Méthode de référence pour la cartographie génétique chez les plantes et la localisation des loci intervenant dans l'expression des caractères quantitatifs ou QTL (*Quantitative Trait Loci*), l'AFLP a été adaptée au typage bactérien par Lin et collaborateurs [69] selon la suggestion de Vos et collaborateurs [70]. Les empreintes génétiques générées par AFLP sont reproductibles et discriminantes. L'AFLP implique la digestion enzymatique de l'ADN génomique, la ligation d'adaptateurs aux fragments générés et l'amplification sélective de ces fragments par l'utilisation d'amorces au niveau des adaptateurs du site de restriction. Anciennement méthode de référence pour le typage de nombreuses bactéries, l'AFLP demeure couteuse, fastidieuse et peu standardisée.

Contrairement aux méthodes basées sur la restriction génomique, celles faisant appel à la PCR ne sont pas versatiles *i.e.* un protocole particulier doit être adapté pour chaque espèce bactérienne.

Cependant, elles bénéficient de la sensibilité et de la spécificité de la technique de PCR qui permet le traitement de petites quantités de matériel génétique. La réaction de polymérisation en chaîne confère également l'avantage de pouvoir analyser l'ADN dégradé, issu d'une extraction brute d'ADN d'échantillons biologiques, ou d'ADN ancien, permettant des études de paléogénétique.

2.2. Identification individuelle d'allèles à plusieurs loci

Les résultats obtenus par ces précédents procédés sont analogiques et se présentent sous la forme de profils ou *patterns* qui rendent difficiles la standardisation et la normalisation de ces technologies. Au cours de la dernière décennie, la disponibilité croissante de séquences génomiques complètes de nombreuses bactéries pathogènes a permis de développer de nouvelles techniques. Au contraire des précédentes, les techniques de nouvelle génération sont de type numérique.

a. Le MLVA

Le généticien britannique Alec Jeffreys découvre, en 1984, l'hypervariabilité de séquences répétées en tandem chez l'Homme. Cette découverte inaugure l'ère de l'empreinte génétique. Initialement identifiés par RFLP, les VNTRs sont des marqueurs de choix pour discriminer les individus et définir des cartes d'identité génétique. Rapidement, cette méthodologie va compléter la technique des empreintes digitales en médecine légale. L'étude des minisatellites humains par RFLP permet pour la première fois, en 1986, de confondre le violeur et meurtrier de deux jeunes filles de Narborough dans le Leicestershire. Cette technique révolutionnaire présentait cependant une limitation majeure : la construction du profil génétique par RFLP nécessite l'accès à une grande quantité d'ADN initial de bonne qualité, prérequis difficilement satisfait lorsque l'ADN est prélevé sur une scène de crime. La standardisation des techniques de PCR rendait envisageable l'utilisation des traces d'ADN et de nouveaux marqueurs polymorphes ont été introduits pour remplacer les grandes répétitions minisatellites d'Alec Jeffreys. En 1989, D. Tautz démontre le polymorphisme des répétitions en tandem à court motif et suggère leur utilisation pour la cartographie génétique [71]. Durant les années 90, Peter Gill du Forensic Science Service, pionnier des analyses génétiques modernes, identifie différents STRs d'intérêt et développe la PCR-STR, technique d'analyse concomitante du polymorphisme des STRs par PCR multiplexe [72]. Adoptée en 1997 comme méthode de référence par le FBI, cette technique est actuellement pratiquée en routine par tous les laboratoires d'analyse mandatés.

L'équipe de Craig Venter au TIGR (*The Institute for Genomic Research*) publie, pour la première fois, en 1995, le génome séquencé d'un organisme vivant, la bactérie *Haemophilus influenzae*. L'équipe dirigée par Alex van Belkum en tire parti pour identifier les STRs polymorphes chez *H. influenzae* et propose leur utilisation comme marqueurs moléculaires [73]. Paul Keim définit le terme MLVA (*Multiple Loci VNTR Analysis*) pour désigner cette méthode d'analyse [24]. Depuis,

des protocoles de typage par MLVA ont été publiés pour plusieurs dizaines d'espèces bactériennes (parmi les études pionnières, [64,74,75,76,77,78,79]). Le génotypage par MLVA repose sur l'amplification par PCR de plusieurs VNTRs dispersés sur le génome bactérien, à l'aide d'amorces spécifiques des régions flanquantes, et sur la détermination des tailles des amplicons par électrophorèse permettant d'évaluer le nombre de répétitions à un locus donné. La longueur des unités répétées étant connue, ces tailles reflètent le nombre d'unités répétées dans les régions amplifiées. Cette méthode de détermination d'empreintes génétiques augmente considérablement, pour certaines espèces, l'efficacité du génotypage bactérien. Le résultat du typage est un code numérique incluant le nombre de motifs à chaque locus. Le processus MLVA est illustré dans la Figure 3 ci-dessous. L'intérêt et les limitations de cette méthodologie seront décrits en fin d'introduction.

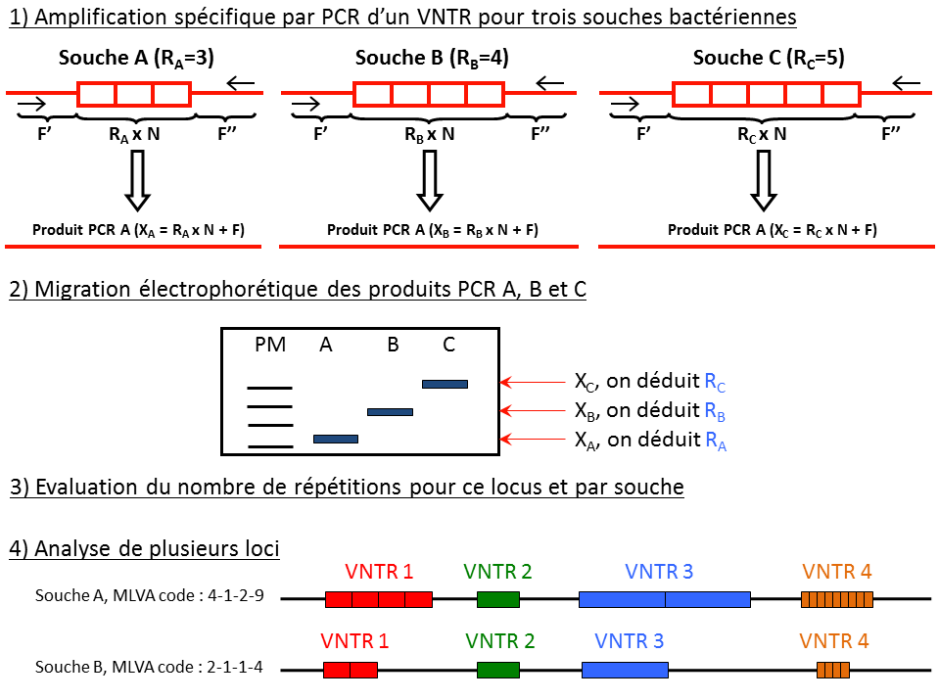


Figure 3. Processus MLVA

b. Le MLST

L'analyse MLST ou *Multi Loci Sequence Typing* repose sur le séquençage d'une portion de sept gènes de ménage [80]. Chaque allèle rencontré est nommé. La combinaison de ces différents allèles permet l'attribution d'un ST ou *Sequence Type* sous forme d'un code numérique de sept chiffres. Le MLST est actuellement une méthode centrale pour les analyses phylogénétiques bénéficiant de la disponibilité de nombreuses bases de données. Cependant, l'étude de gènes de ménage, à l'évolution lente, n'est pas adaptée à l'investigation épidémiologique et des techniques complémentaires ont émergé. Ainsi, le SBT, ou *Sequence Base Typing*, désormais préconisé comme méthode de référence pour le typage de *L. pneumophila*, est l'homologue du MLST à la différence que les gènes étudiés sont impliqués dans la virulence ou dans l'expression membranaire [81,82]. Ces

gènes davantage soumis aux processus évolutifs que sont la sélection naturelle et la dérive génétique ont des taux de mutation plus élevés et procurent un pouvoir de discrimination augmenté. Par réciprocité, le SBT a une valeur moindre pour des études phylogénétiques. Aussi bien le MLST que le SBT sont fondés sur la comparaison de séquences et *a fortiori* l'analyse de SNPs. Ces changements nucléotidiques stables transmis à la descendance peuvent être détectés de manière spécifique et individuelle par utilisation de sondes d'hybridation, de puces à ADN ou de séquençage localisé. Le bouleversement majeur est annoncé dans la décennie à venir avec la recherche exhaustive de SNPs par séquençage complet. Les NGS ont d'ores et déjà révolutionné notre façon d'appréhender l'épidémiologie bactérienne ; le récent cas épidémique causé par une souche d'*E. coli* ultra-virulente le montre. Le séquençage du génome de la souche incriminée a non seulement contribué à identifier la source de contamination mais également à comprendre son pouvoir pathogène. Cette souche résulterait d'une hybridation lui conférant la virulence d'une souche d'*E. coli* enterohémorragique toxinogène et la capacité d'adhésion aux cellules de la muqueuse intestinale d'une souche d'*E. coli* enteroaggregative [83]. En théorie, une carte d'identité génétique d'un isolat bactérien est strictement révélée par la séquence totale de son génome ; le pouvoir de discrimination d'une telle méthode est alors parfait [84]. Cependant, les génomes sont des structures souples, flexibles, modulables et des changements génétiques peuvent intervenir entre isolats issus du même clone épidémique. Outre les problématiques techniques liées à la difficulté d'associer certains contigs, l'interprétation des données de NGS doit être effectuée avec précaution pour éviter des erreurs d'analyse. Les épidémiologistes réfléchissent à des méthodologies de type MLST-plus (séquençage et comparaison de plusieurs centaines de gènes) pour bénéficier de l'apport des NGS en effectuant une fouille de données pertinente.

3. Des méthodes adaptées au besoin épidémiologique

Le positionnement d'une problématique initiale est décisif pour entreprendre avec pertinence une étude épidémiologique. L'échelle de l'étude engagée déterminera le choix de la méthode.

3.1. Epidémiologie d'intervention et épidémiologie descriptive

L'investigation épidémiologique est déclinée en plusieurs facettes. Les buts de l'enquête épidémiologique sont doubles : traçabilité et biosurveillance. Selon l'échelle d'analyse considérée, l'épidémiologie sera caractérisée comme d'intervention ou descriptive. L'épidémiologie d'intervention relève *stricto sensu* de l'enquête épidémiologique *i.e.* la définition de la relation entre isolats suspectés de faire partie de la même chaîne de transmission, la confirmation de l'état épidémique ou la validation des hypothèses sur les réservoirs, sources et modes de transmission. Des systèmes de typage comparatifs et résolutifs sans considération de référence extérieure, comme ceux

fondés sur la comparaison de profils de bandes, peuvent alors être adoptés dans ce contexte de micro-épidémiologie [85]. L'épidémiologie descriptive rentre davantage dans le cadre de la biosurveillance et du suivi épidémiologique *i.e.* l'étude de la distribution des clones ou la mise en évidence de l'émergence. Des systèmes de typage définitifs avec une nomenclature standardisée des types identifiés et stockés dans une base de données sont préférés [85]. Le développement et l'utilisation croissante de méthodes comme le MLVA ou le MLST, associés à plusieurs bases de données, s'expliquent par leur versatilité : ces deux méthodes sont à la fois comparatives et descriptives.

3.2. Vitesse d'évolution des marqueurs

Les marqueurs génétiques sur lesquels reposent les techniques de typage présentent des caractéristiques très diverses, décrites plus tôt dans ce chapitre. Certains sont stables et évoluent lentement, d'autres au contraire sont hypermutables. Parallèlement, les premiers exhibent logiquement un faible pouvoir de discrimination contrairement aux seconds. La vitesse de mutation des marqueurs est donc une composante principale à considérer et à adapter au besoin épidémiologique. L'investigation épidémiologique se conçoit sur différents niveaux de l'échelle spatio-temporelle. Cette notion de spatio-temporalité est directement corrélée avec celle de la vitesse de mutation des marqueurs et de leur pouvoir discriminant (Figure 4). Ainsi, les marqueurs à évolution rapide sont utilisés, en raison de leur discrimination élevée, pour des études épidémiologiques d'intervention, par exemple dans des situations épidémiques. *A contrario,* les marqueurs à évolution lente, comme les SNPs identifiés dans les gènes de ménage, sont pertinents pour l'épidémiologie descriptive *i.e.* définir des complexes clonaux et les suivre à l'échelle planétaire [86]. Cette dernière phrase est à nuancer car les SNPs considérés à l'échelle du génome complet deviennent collectivement un marqueur à évolution rapide, les nouvelles techniques de séquençage permettant l'analyse exhaustive de ces SNPs. Enfin, le taux de mutation des marqueurs est fonction de la diversité génétique de l'espèce étudiée. Des marqueurs pourront se montrer peu résolutifs chez une espèce et très discriminants pour une autre [87].

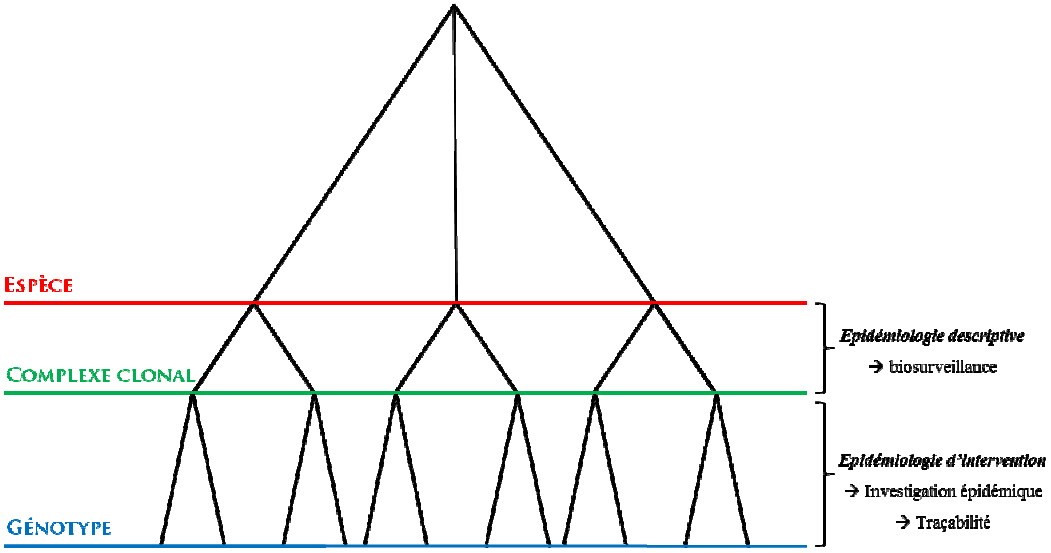


Figure 4. Vitesse d'évolution des marqueurs et niveaux de résolution

Selon le niveau de résolution voulu et la question épidémiologique, le choix du marqueur sera déterminant. L'analyse simple locus d'un gène universel très conservé comme l'ADNr 16S permettra d'identifier le genre bactérien et parfois l'espèce. Afin d'identifier les complexes clonaux au sein d'une espèce bactérienne, des marqueurs de typage d'évolution lente seront préférés (ex. : gènes de ménage du MLST). Pour affiner l'analyse génotypique, des techniques reposant sur la détection du polymorphisme de régions instables seront utilisées (ex. : profil de macrorestriction de la PFGE analysant les génomes complets).

3.3. Synthèse

L'épidémiologiste dispose donc d'une multitude de méthodes de typage qui exploitent des sources de polymorphisme différentes, présentent des critères de qualité variables et doivent être adaptées à la problématique épidémiologique. Le Tableau 1 résume la diversité des méthodes de génotypage. Les méthodes produisant des profils de bandes analysent le génome de manière exhaustive et sont donc sensibles à l'ensemble des sources de polymorphisme. A l'inverse, les procédés identifiant de manière individuelle les marqueurs ciblés sont indépendants de polymorphismes secondaires. En théorie, le séquençage génomique complet investigue toutes sortes de polymorphismes. Cependant, les techniques de séquençage les plus récentes ne sont pas encore adaptées pour la recherche des séquences répétées, éléments mobiles inclus.

		Méthodes de typage	Eléments transférés horizontalement	Réarrangements chromosomiques	Mutations ponctuelles	Séquences répétées dispersées	Séquences répétées en tandem
Profils de bandes	Restriction d'ADN	**RFLP**	✓	✓	✓✓✓	✓✓✓*	✓
		PFGE	✓✓✓	✓✓✓	✓	✓✓✓	✓✓✓
	Amplification génique	**RAPD**	✓	✓	✓✓✓	✓	✓
		REP/BOX/ERIC-PCR	✓	✓		✓✓✓	✓
		AFLP	✓	✓	✓✓✓	✓	✓
Identification individuelle		**MLVA**					✓✓✓
		MLST			✓✓✓		
		Séquençage complet		✓✓✓	✓✓✓		

Tableau 1. Sources de polymorphisme exploitées par les méthodes de génotypage

(✓✓✓ : polymorphisme spécifiquement recherché ; ✓ : polymorphisme secondaire)
*ce polymorphisme est utilisé par la méthode RFLP lorsque les séquences ciblées sont de type répétées dispersées (IS) ou présentes en multi-copies (ADNr 16S)

L'épidémiologie des maladies infectieuses est une science complexe avec son vocabulaire, sa méthodologie et ses concepts. Le typage bactérien est un des piliers de cette science, des critères d'évaluation et des méthodes de référence sont donc proposés afin de le standardiser. Le groupe d'étude sur les marqueurs épidémiologiques, l'ESGEM (*European Study Group on Epidemiological Markers*), pose les fondamentaux qui régissent le choix d'une méthode de typage. Ce comité d'experts évoque deux catégories de critères pour définir la qualité d'une méthode de typage : les critères dits de performance et ceux dits de commodité [88]. La performance d'une technique est évaluée, entre autres, par la stabilité des marqueurs utilisés, le pouvoir de discrimination du typage ainsi que sa reproductibilité. La rapidité, le coût ou la facilité d'utilisation sont des critères déterminant la commodité d'une technique de typage. Certains de ces critères ont été énumérés lors de la description des méthodes et sont rassemblés dans le Tableau 2. Il apparaît distinctement que les méthodes produisant des profils de bandes sont globalement peu commodes et peu performantes malgré la discrimination élevée de certaines, en particulier la PFGE et l'AFLP. Celles-ci demeurent donc employées pour de l'investigation d'intervention strictement comparative. Les techniques d'évaluation individuelle du polymorphisme de loci sont beaucoup plus performantes. Le MLST souffre d'un manque de discrimination et est donc davantage utilisé en épidémiologie descriptive. Le MLVA a les atouts pour faire consensus en prouvant son utilité pour les deux facettes de l'investigation épidémiologique. Le séquençage complet n'est pas encore approprié pour l'analyse épidémiologique.

Cette voie n'est empruntée que dans de rares cas exceptionnels comme celui de l'enquête menée pour l'étude de la souche ultra-virulente d'*E. coli* responsable d'une épidémie européenne, en 2011, causant 126 décès et des centaines de millions d'euros de préjudices. Des avancées technologiques importantes sont à venir, mais, actuellement, l'utilisation du séquençage complet est encore du domaine de la recherche et ne peut être proposé comme analyse de routine.

	Critères de performance					Critères de commodité			
Méthodes de typage	**Stabilité**	**Typabilité**	**Reproductibilité**	**Pouvoir de discrimination**	**Concordance épidémiologique**	**Rapidité**	**Facilité d'utilisation**	**Cout***	**Portabilité**
RFLP	★	★★★★	★	★★	★★	★★	★★	★★★	★
PFGE	★	★★★	★	★★★★	★★	★	★	★★	★
RAPD	★	★★★★	★	★★	★★	★★★	★★★	★★★	★
REP/BOX/ERIC-PCR	★★★	★★★★	★★	★★	★★	★★★	★★★	★★★	★
AFLP	★★★	★★★★	★★★	★★★	★★	★	★★	★★	★
MLVA	★★★★	★★★★	★★★★	★★★★	★★★★	★★★★	★★★★	★★★★	★★★★★
MLST	★★★★★	★★★★	★★★★	★	★★★★★	★★★	★★★★	★★	★★★★★
Séquençage complet	?	★★★★★	?	★★★★★	?	★	★	★	?

Tableau 2. Critères de validation des méthodes de génotypage

(★ : très faible ; ★★ : faible ; ★★★ : moyen ; ★★★★ : élevé ; ★★★★★ : très élevé ;
? : paramètres difficiles à estimer pour le séquençage complet, ils dépendent des technologies mises en œuvre, de la gestion des données et des marqueurs analysés pour l'investigation
*pour le coût la symbolique est inversée)

III) Les VNTRs, des outils épidémiologiques

Le MLVA est la technique qui à ce jour semble convenir au génotypage du plus grand nombre d'espèces bactériennes à un coût suffisamment réduit pour pouvoir envisager sa large utilisation dans les laboratoires de microbiologie. Toutefois, alors que son concept est validé et admis pour certaines espèces, sa généralisation n'est pas acquise. Comment l'expliquer ? Les caractéristiques biologiques des VNTRs sont-elles un frein à la propagation du MLVA ? Cependant, le MLVA s'impose ces dernières années comme méthode de référence pour plusieurs pathogènes. Quels sont alors ses atouts ?

1. Les VNTRs, des structures génomiques complexes

1.1. Instabilité des répétitions en tandem

L'instabilité des répétitions en tandem est le moteur du polymorphisme de ces loci. Celle-ci est la conséquence de mécanismes mutationnels intervenant à plusieurs étapes de modification de l'ADN : la réplication, la transcription, la recombinaison et la réparation. Nous verrons quels mécanismes ont été proposés pour expliquer leur instabilité et dévoilerons les facteurs génétiques ou environnementaux supposément impliqués. La compréhension de ces aspects est primordiale pour la prédiction du polymorphisme des VNTRs ou leur appréciation en tant qu'outils épidémiologiques.

a. Mécanismes d'instabilité : réplication, recombinaison et transcription

i) La réplication

Le modèle de glissement et mésappariement ou SSM (*Slipped-Strand Mispairing*) est le mécanisme général évoqué pour expliquer l'instabilité des répétitions en tandem courtes [89]. Ce phénomène, se produisant lors de la réplication, conduit à l'augmentation ou à la réduction du nombre de copies selon le brin subissant le mésappariement. Ce modèle suppose que lors de la polymérisation du néo-brin d'ADN, l'enzyme polymérase effectue des pauses. Durant ces pauses, le brin nouvellement synthétisé peut se désapparier et se réapparier de manière décalée, laissant une boucle non appariée. Selon le décalage produit, la fin de la réplication produit une addition ou une perte d'unité de la répétition. La formation d'une boucle sur le brin néo-synthétisé conduit à l'addition d'un motif alors que la formation d'une boucle sur le brin matrice explique la perte de motifs. Le modèle SSM est illustré sur la Figure 5. Les propriétés structurales des répétitions en tandem contribuent fortement au bégaiement de la polymérase. En effet, les répétitions en tandem sont à même de former des structures secondaires, caractérisées par des conformations de type tige/boucle, des S-ADN ou des brins d'ADN dénoués, qui sont susceptibles de provoquer l'arrêt de la polymérase. Des études *in vitro* montrent que ce phénomène est observable pour la plupart des polymérases [90].

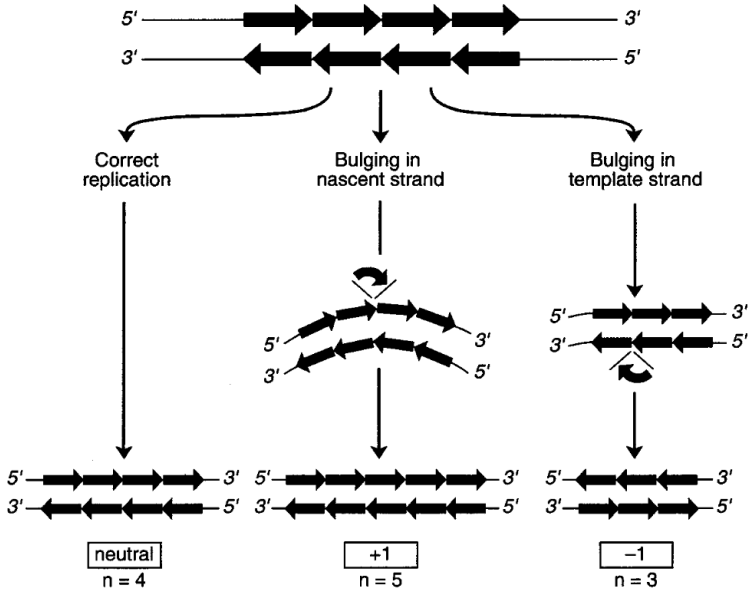


Figure 5. Modèle SSM (extrait de [91] avec la permission de American Society of Microbiology)

ii) La recombinaison

Le modèle SSM est communément accepté par la communauté scientifique pour expliquer l'instabilité des répétitions en tandem chez les procaryotes. D'autres mécanismes ont cependant été mis en avant. Si la conversion génique semble rarement impliquée dans l'évolution des répétitions en tandem, des mécanismes de conversion par recombinaison homologue ont été décrits chez les bactéries [92]. La recombinaison homologue, processus de diversification génétique, ne semble pas avoir d'impact sur l'instabilité des microsatellites mais paraît avoir un rôle prépondérant dans celle des minisatellites [93]. Très peu d'études décrivent le lien entre la recombinaison homologue et l'instabilité des répétitions en tandem chez les bactéries. La détection d'une cassure au niveau d'une répétition en tandem fera intervenir le recrutement de séquences homologues à la zone endommagée via la mise en place de la machinerie recombinatoire. Si la cassure arrive au niveau de la boucle réplicative, RecA créera un hétéroduplex entre les brins homologues. Si l'ADN est endommagé en dehors du processus de réplication, RecA assemblera l'hétéroduplex au niveau de la structure répétée en tandem au sein de laquelle l'identification de séquences homologues est facilitée : on parle de recombinaison intra-répétition. Une longueur minimale entre les séquences homologues échangées a été suggérée. Les évènements de recombinaison ont été démontrés possibles entre des séquences séparées d'au moins 25pb chez *E. coli* [94]. La fréquence de recombinaison homologue est considérablement augmentée pour les séquences de 74pb et plus [95]. De même, deux séquences séparées par plus de 1kb sur le chromosome ont sensiblement moins la capacité de recombiner [96]. Enfin, la fréquence de la recombinaison homologue est accrue par la similarité partagée par les deux séquences. Chez *E.coli*, une divergence de 10% entre les deux répétitions réduit les évènements de recombinaison d'un facteur 40 [94].

iii) La transcription

Le mécanisme transcriptionnel a été mis en évidence comme source d'instabilité des microsatellites chez de nombreux organismes modèles [97]. Cette hypothèse a été également avancée, chez *S. cerevisiae*, par la mise en exergue de l'augmentation de l'instabilité de la répétition du dinucléotide (CG) lors de l'activation de sa transcription [98]. Un modèle a été proposé pour expliquer l'influence des mécanismes transcriptionnels. Lors de la formation du complexe de transcription, la séparation puis la réassociation des brins complémentaires d'ADN est susceptible de créer des structures secondaires de type épingles à cheveux ou boucles déstabilisant les répétitions en tandem. Ce mécanisme n'affecte que les microsatellites transcrits et susceptibles d'adopter des structures secondaires. La transcription n'est donc impliquée que marginalement dans la dynamique des structures répétées en tandem.

b. Facteurs de stabilité des répétitions en tandem

i) Facteurs trans : la réparation de l'ADN

Des systèmes multiples assurent la sauvegarde et la stabilité de l'information génétique en limitant la fixation des mutations [99]. Le système de correction des mésappariements ou MMR, pour *MisMatch Repair*, participe au maintien de l'intégrité du génome chez les organismes vivants. Ce système de réparation post-réplicatif reconnaît et corrige les mésappariements résultant d'une mauvaise incorporation nucléotidique ou d'un dérapage de la polymérase ayant échappé au *proofreading* ou fonction de relecture. Ainsi, la boucle non appariée est la cible des enzymes de réparation des mésappariements. Les protéines Mut du système MMR ont été initialement identifiées chez *E. coli* [100]. La mutation d'un gène codant pour une de ces enzymes entraîne une augmentation des fréquences d'instabilité des répétitions en tandem d'un facteur 20 pour *E.coli* (gènes *mutS* ou *mutL*) [101]. La stabilité des microsatellites représente donc un enjeu majeur pour la physiologie bactérienne qui déploie les mécanismes de réparation adéquats afin de limiter les évènements de mutation. Cependant les minisatellites ne semblent pas affectés par les dysfonctionnements du système de réparation des mésappariements chez *S. cerevisiae* [102,103] et *E. coli* [104].

Les enzymes impliquées dans les systèmes de réparation par excision de nucléotide ou NER pour Nucleotide Excision Repair font partie des complexes uvrABC [105]. Cependant, le complexe uvrABC n'intervient pas dans la stabilité des répétitions en tandem via sa fonction NER. La sous-unité uvrA de uvrABC se lie aux structures en épingle à cheveux, intermédiaires structuraux du modèle SSM, et bloque le bégaiement de la polymérase [106]. Les exonucléases simple brin, enzymes catalysant l'hydrolyse séquentielle des nucléotides d'un ADN double brin dans le sens 3'→5', sont supposées participer au maintien de la structure des minisatellites par dégradation de la séquence répétée mésappariée. Feschenko et collaborateurs, après avoir étudié l'instabilité d'un minisatellite

plasmidique, ont conclu que 90% des évènements de délétion potentiels étaient inhibés par l'action conjointe de trois exonucléases simple brin chez *E. coli* [107].

ii) Facteurs cis : caractéristiques structurelles des répétitions

Les répétitions en tandem forment des structures atypiques dont le taux d'instabilité est affecté par plusieurs caractéristiques inhérentes : la longueur et le nombre d'unités répétées , la taille et la composition du motif répété , le degré de similitude entre les motifs. Chez la levure *Saccharomyces cerevisiae*, l'instabilité du microsatellite dinucléotidique GT croît avec la taille totale de la répétition [98]. Cependant, F. Denoeud et G. Vergnaud n'ont pas mis en évidence de corrélation positive significative entre le polymorphisme des répétitions en tandem et leur longueur totale chez quatre espèces bactériennes (*E. coli*, *S. aureus*, *S. typhi* et *S. pyogenes*) [60]. Une analyse fine du locus VNTR O157-10 chez *E. Coli* a montré que le taux de mutation était corrélé positivement au nombre d'unités répétées portées par ce microsatellite [104]. Cette corrélation a été généralisée pour plusieurs loci VNTR chez *E. coli* et *Y. pestis* [104,108,109]. Le glissement lors de la réplication surviendrait donc plus fréquemment si le microsatellite compte davantage d'unités répétées. L'occurrence élevée des mécanismes de type SSM explique la forte instabilité des microsatellites par rapport aux minisatellites. Par extension, la fréquence de dérapage de l'ADN polymérase dans les mononucléotides est d'autant plus élevée que la répétition est longue [110]. Chez *S. cerevisiae*, un microsatellite de 14 adénines est par exemple 400 fois plus instable que son homologue contenant 4 adénines [111]. Des expériences de stabilité ont montré, chez *E. coli*, que les microsatellites polymorphes dont le motif est composé de moins de six nucléotides sont soumis à davantage d'instabilité et arborent un taux de mutation trois à soixante-dix fois plus élevé que les grands microsatellites ou les minisatellites [104]. Une étude chez *E. coli* a montré que l'instabilité d'un microsatellite d'une longueur donnée dépend de son orientation par rapport au sens de la réplication, l'instabilité d'un poly(TG) étant deux fois plus élevée que celle d'un poly(AC) de même longueur [112]. Enfin, les répétitions imparfaites, *i.e.* dont le motif répété est faiblement conservé, se révèlent très stables. Petes et collaborateurs ont mis en exergue la stabilité des microsatellites dont les motifs répétés avaient divergé par mutation ponctuelle [113]. L'étude comparative *in silico* menée par F. Denoeud et G. Vergnaud entre quatre espèces bactériennes (*E. coli*, *S. aureus*, *S. typhi* et *S. pyogenes*) corrobore cette observation : moins le motif répété est conservé, plus la répétition en tandem s'avèrera monomorphe [60]. Cette observation avait été décrite quelques années plus tôt par Le Flèche et collaborateurs qui avaient observé que le nombre d'allèles des répétitions en tandem identifiées chez *Y. pestis* était corrélé à la conservation du motif répété [64]. L'imperfection de ces structures désavantagerait logiquement les mécanismes de recombinaison homologue et diminue les propriétés structurales qui favorisent le glissement de la polymérase.

iii) Facteurs environnementaux

Les bactéries sont constamment soumises à des stress environnementaux qui induisent des mécanismes de réponse au stress dont le facteur sigma est la clé de voute. La mise en place de ce système contribue à l'augmentation du pouvoir adaptatif des bactéries par l'augmentation du taux de mutation. Cette plasticité des génomes procaryotes est un élément essentiel de leur capacité d'adaptation. La littérature ne distingue qu'une seule étude décrivant l'influence des stress environnementaux sur l'instabilité des répétitions en tandem chez ces bactéries [114]. La température est un facteur important d'instabilité des VNTRs : trois fois plus de mutations sont observées quand la bactérie se développe à 43°C au lieu de 10°C. De manière surprenante, l'exposition à la lumière ou aux rayons UV ne change pas significativement le taux de mutation observé. Cette absence d'effet a également été observée pour les minisatellites de *L. pneumophila* [115]. La bactérie, placée en situation de carence nutritive, présente un taux de mutation trois fois plus important. La déficience des mécanismes de réparation d'ADN, et notamment du système MMR, induite lors de situations de carence a été décrite mais son effet sur la dynamique des répétitions en tandem n'est pas encore correctement caractérisé. De longues périodes de carence induisent un changement de la conformation de l'ADN superenroulé favorisant le phénomène de glissement de la polymérase [116,117].

1.2. Des loci impliqués dans le *fitness* bactérien

Les VNTRs sont des régions génomiques instables dont la mutation est médiée par plusieurs facteurs. La réponse à des stress environnementaux, par essence imprévisibles, est une des raisons de promouvoir cette forme de variabilité. Ainsi, chez *E. coli*, les gènes impliqués dans la réponse aux stress environnementaux, garants de l'intégrité génomique, sont riches en microsatellites [118]. Une hypothèse évolutive suggère que la fonction de ces microsatellites serait d'induire, chez certains individus, un phénotype mutateur en modulant l'expression de ces gènes. Ces individus subiraient alors une diminution de *fitness* dans des conditions optimales de croissance, mais seraient capables, lors de changements environnementaux, d'explorer plus rapidement l'espace des phénotypes possibles. Le *fitness* bactérien ou degré de compétition dans un environnement est un concept central en théorie de l'évolution. Mesure de la sélection naturelle, cette valeur adaptative décrit la capacité d'un génotype particulier à se reproduire et à assurer sa pérennité. La théorie de la contingence génétique a été popularisée par R. Moxon en 1994 [36]. Cette théorie avance l'existence de loci dits de contingence, régions génomiques hypermutables qui modulent l'expression génétique. Cette possibilité de *switch* génotypique contribue à la plasticité génomique, un des éléments générateurs du *fitness* bactérien. Certains VNTRs ont un rôle physiologique pour les bactéries comme loci de contingence. Certaines familles ou classes de répétitions sont-elles davantage impliquées dans le fitness bactérien ? Par quels mécanismes ces loci interviennent-ils dans la plasticité génomique ?

a. Variation de phase

La variation de phase est un évènement génétique naturel qui résulte d'altérations génétiques ou épigénétiques [119]. Cette variation cause le décalage du cadre de lecture du gène affecté. Un changement du cadre de lecture peut survenir par mutation ponctuelle ou par une indel (insertion/délétion) non multiple de trois. Une variation de phase induit donc un changement sévère de l'expression des gènes pouvant aller jusqu'à l'introduction d'un codon-stop et l'arrêt prématuré de la transcription. La variation de phase permet alors des basculements génétiques de type « marche/arrêt » fréquents, stochastiques, héritables et réversibles [119]. Ce processus d'adaptation des populations bactériennes aux changements de l'environnement module donc l'expression génétique. Cette stratégie adaptative conduit à l'apparition simultanée et à haute fréquence de plusieurs phénotypes nouveaux dans une population bactérienne, générant une diversité fonctionnelle intra-espèce. L'implication de la variation de phase, générée par une répétition en tandem, dans un système régulateur complet a été décrite pour la première fois pour le gène *mod*. Codant pour une méthyltransférase composante d'un système de restriction-modification d'ADN, ce gène comporte un locus de contingence, une répétition du tétranucléotide AGTC. Ce gène fait partie d'une cascade de régulation qui induit un changement phénotypique du pathogène lors de son changement de site de colonisation. La variation de phase du gène *mod* régule ainsi l'expression de seize gènes codant pour une grande diversité de protéines (transporteurs de fer, protéines membranaires, protéines de choc thermique). L'instabilité du locus de contingence est si élevée qu'elle génère du mosaïcisme phénotypique observable sur boîtes *in vitro* [120] ; les individus dont le gène *mod* est éteint s'avèrent plus compétitifs en conditions de stress chez *Neisseria* [121]. La variation de phase de ce gène est à l'origine d'effets pléiotropiques. En effet, le basculement en position « arrêt » du gène *mod* chez *N. gonorrhoeae* consolide le biofilm formé par cette espèce augmentant son adhésion, et *a fortiori* sa survie, sur les cellules cervicales épithéliales humaines [121].

b. Variation antigénique

La variation antigénique est un mécanisme original de survie de certaines bactéries extracellulaires. Elle traduit une induction de la diversité antigénique caractérisée par la capacité d'une bactérie à modifier de façon itérative certains antigènes majeurs exposés à sa surface au cours du processus infectieux [119]. Les exemples de variation antigénique décrivent classiquement une fonction d'échappement au système immunitaire et une fonction d'adaptation aux récepteurs endothéliaux de l'hôte. Dans le cadre des interactions hôtes/parasites où les pressions sélectives sur le parasite sont fortes, la variation phénotypique joue un rôle central dans le succès de l'infection. L'instabilité locale des répétitions en tandem, dont le motif est congru à trois, appartenant à des phases ouvertes de lecture, génère des polymorphismes protéiques caractérisés par une variation antigénique des protéines membranaires susceptibles d'être reconnues par le système immunitaire de l'hôte. En

effet, le séquençage complet des génomes a permis de montrer que la quasi-exclusivité de ces loci de contingence est associée aux gènes codant pour des protéines membranaires [122,123]. Ces protéines jouent un rôle déterminant dans la spécificité antigénique des bactéries et contribuent à leur virulence. L'objet n'est pas de lister les différents cas identifiés mais simplement d'illustrer le rôle des répétitions en tandem dans les mécanismes d'échappement à l'hôte avec quelques exemples. Chez les streptocoques du groupe B, le gène *bca* comporte un minisatellite très polymorphe. Ce gène code pour la protéine Alpha C, une protéine membranaire transporteur de polysaccharide et un puissant antigène de surface qui s'avère très peu immunogène lorsque le minisatellite porté par le gène *bca* présente peu d'unités répétées. Ce changement de conformation des épitopes permettrait à la bactérie d'échapper à la réponse immunitaire de l'hôte [124,125]. Cette hypothèse a été confirmée *in vivo* par la démonstration de la relation de proportionnalité entre la réponse immunitaire de rats soumis à la protéine Alpha C et le nombre de motifs répétés du minisatellite étudié. Ce mécanisme de survie adaptative est bien documenté et décrit dans quelques autres cas. Par exemple, chez *Mycoplasma hyorhinis*, les protéines de surface du système Vlp confèrent aux bactéries, par leur variation de taille, une résistance contre les anticorps produits par l'hôte [126].

c. Variation fonctionnelle

De manière plus anecdotique, certains exemples décrivent l'impact de l'instabilité de certains satellites sur la fonction protéique. Ainsi, chez *E. coli*, l'insertion d'un motif supplémentaire du triplet répétée TCT dans le gène *ahpC* modifie la fonction de la protéine codée : d'une péroxidase, la protéine devient une disulfide réductase par l'addition d'une phénylalanine [127]. Ce *switch* fonctionnel a été observé lors de conditions de stress, durant lesquelles la capacité à former des ponts disulfures confère un avantage sélectif.

d. Régulation de la transcription

i) Contrôle du promoteur et mise sous silence de certains gènes

Les répétitions en tandem insérées dans les régions codantes sont à l'origine de basculements génétiques de type « marche/arrêt ». Les microsatellites non codants peuvent également induire ce type de changements génétiques drastiques lorsqu'ils sont présents dans les séquences promotrices. Les promoteurs bactériens sont courts, environ 40 bases, et sont constitués de deux motifs caractéristiques appelés boîtes : une séquence en position -10 appelée boîte de Pribnow (TATAAT) et une séquence en -35 (TGTTGACA). La distance entre ces deux séquences représente un seul tour d'hélice permettant aux deux motifs d'interagir de manière concomitante avec le facteur sigma de l'ARN polymérase. Les microsatellites instables situés dans cette région intercalée induisent alors une modification de la région promotrice. Ces loci contrôlent par conséquent l'expression ou le silence de certains gènes. Chez *Neisseria* spp, *Yersinia* et *H. influenzae*, la répétition du dinucléotide TA est

localisée dans un promoteur dual qui contrôle l'expression de deux gènes codants pour les sous-unités du pilus, *hifA* et *hifB* [128]. Cette répétition est un locus de contingence qui modifie la structure promotrice entre les boîtes -35 et -10. Selon le nombre de répétitions, le site de reconnaissance de l'ARN polymérase sera intact ou altéré. Ainsi un isolat possédant 9 unités répétées sera déficient en pilus contrairement aux isolats possédant 10, 11 ou 12 répétitions.

ii) Interaction avec les facteurs de transcription

La régulation de la transcription fait intervenir plusieurs protéines appelées facteurs de transcription dits *trans*-régulateurs. Ces protéines se lient sur l'ADN au niveau de sites de fixation spécifiques, situés en amont des promoteurs, appelés éléments *cis*-régulateurs. Cette interaction entre les éléments *cis* et *trans* permet alors la régulation de l'activité de l'ARN polymérase. La conformation adoptée par l'élément *trans* conduira à l'activation ou à l'inhibition de la transcription. Certains VNTRs ont été impliqués dans la modulation des éléments *cis*, inhibant alors la fonction des facteurs *trans*-régulateurs. La transcription du gène *nadA* chez *N. meningitidis* est le modèle d'étude. Ce gène exhibe une répétition du motif tétranucléotidique TAAA situé en amont de la région promotrice -35 [129]. La transcription de ce gène est réprimée par l'élément *trans*-régulateur NadR dont le site de reconnaissance se situe à proximité du microsatellite. La fixation de NadR sur le motif *cis* est dépendante du nombre d'unités répétées du microsatellite. La dynamique de cette répétition en tandem régule ainsi l'expression de NadA, une invasine impliquée dans l'adhésion cellulaire et déterminante dans la virulence bactérienne. Chez *M. tuberculosis*, le VNTR3690 est supposé intervenir dans la régulation du gène *lpdA*. Ce gène code pour la protéine lpdA, un facteur de virulence pléiotropique du bacille tuberculeux. Cette protéine est un antigène majeur caractérisé comme un membre de la famille des flavoprotéines disulfide réductase assurant la réponse au stress oxydatif et catalysant la production d'ATP en conditions anaérobies. Le gène *lpdA* s'avère douze fois plus exprimé lorsque VNTR3690 présente quatre motifs répétés contre une unité seule [130]. Le mécanisme régulateur n'est pas élucidé mais les auteurs ont mis en évidence une séquence palindromique dans le minisatellite. Cette séquence est susceptible de former des structures de type tige-boucle, conformation classique de fixation de protéines régulatrices [131].

2. Le MLVA, une méthode parfois contestée

2.1. Concordance épidémiologique et homoplasie

L'interrogation majeure du MLVA décrit dans la majorité des revues sur le MLVA concerne la concordance épidémiologique [88,132]. La problématique du rôle biologique des VNTRs soulève de nombreuses interrogations. Bien que la majorité des mutations affectant les séquences répétées en tandem semblent neutres et n'engendrent pas de modifications phénotypiques, certaines sont associées

à d'importantes fonctions biologiques. En effet, certains VNTRs sont impliqués dans des mécanismes de variations de phase ou de modifications antigéniques. Le rôle fonctionnel de l'instabilité de ces séquences est supposé jouer un rôle dans l'adaptation de la bactérie à une nouvelle niche écologique et notamment à un nouvel hôte par l'altération de la réponse immunitaire. Dans le cadre de contextes particuliers (épidémie, invasion de l'hôte, colonisation d'une nouvelle niche écologique), un allèle particulier pourra être sélectionné positivement et se répandra au sein de la population. Deux clones identiques pourront artéfactuellement être différenciés par un marqueur soumis à sélection et dont la mutation confère un pouvoir adaptatif [132]. Ainsi, lors d'un contexte épidémique, la transmission d'un clone d'hôtes en hôtes peut être favorisée par la modulation d'un nombre d'unités répétées d'une séquence répétée en tandem situé sur un gène codant pour un facteur antigénique, le génotype de ce même clone dans son réservoir environnemental pouvant alors être différent. Cet exemple illustre la faible pertinence de ce type de VNTRs dans un contexte épidémiologique. La difficulté réside dans l'identification de ces VNTRs impliqués dans un changement phénotypique. Très peu d'études se sont penchées sur ces aspects mais certains paramètres prédictifs peuvent être évalués : présence du VNTR dans une phase ouverte de lecture codant pour une protéine de surface ou un facteur de virulence, VNTR situé dans une zone codante et dont le motif unitaire est congru à trois, VNTR situé au niveau d'un promoteur, présence d'une séquence régulatrice sur le VNTR, révélation d'une structure en tige-boucle ou épingle à cheveux au niveau d'un VNTR. Ces quelques paramètres permettent de suspecter un rôle fonctionnel mais, en aucun cas, ne le prouvent ; la réalisation d'expériences de pathogénicité *in vivo* pourrait démontrer cette corrélation entre génotype et phénotype. Des tests d'évaluation de la concordance épidémiologique, par comparaison génotypique d'isolats corrélés épidémiologiquement, permettent de garantir que le marqueur concerné est adapté aux études épidémiologiques *i.e.* le marqueur doit être polymorphe mais assez stable pour qu'il puisse être invariable dans un contexte géographique et temporel localisé. Seul ce type de validation justifie le critère de concordance épidémiologique et valide l'utilisation du VNTR.

L'apparition indépendante du même état transformé chez différentes clades est appelé convergence, ou par extension, si les clades sont proches, parallélisme. *A contrario*, la réversion est l'apparition d'un caractère ayant l'apparence de la morphologie ancestrale. Ces trois notions, convergence, parallélisme et réversion, définissent, par opposition à l'homologie, des caractères similaires non hérités d'un ancêtre commun. Ces similitudes qui ne traduisent pas les relations évolutives entre taxons ont été originellement conceptualisées en 1870 par le zoologue britannique Edwin Ray Lankester qui les a qualifiées d'homoplasie [133]. L'instabilité des répétitions en tandem et leur évolution sélective sont responsables de l'inhérente homoplasie de certains VNTRs. Ces mutations convergentes rendent inenvisageable l'utilisation de ces marqueurs pour inférer des phylogénies. L'homoplasie peut mettre en cause la concordance épidémiologique lorsque par erreur deux isolats sont considérés identiques en raison d'un marqueur convergent. Cependant, l'analyse de plusieurs loci par MLVA permet de compenser l'homoplasie individuelle de certains VNTRs [134].

Ainsi, deux isolats identiques par MLVA, récoltés dans un contexte géographique et temporel restreint, sont issus d'une même chaîne clonale. Cependant, la clonalité n'empêche pas l'homoplasie [36]. Des mutations convergentes sont ainsi susceptibles d'apparaître au sein d'un complexe clonal. Cet effet homoplasique pourrait conduire à l'assignation erronée d'une souche à un complexe clonal aboutissant à des inexactitudes lors d'études prospectives de biosurveillance. L'utilisation de plusieurs loci VNTRs évite également ce type de désagrément. L'homoplasie des VNTRs doit être rapportée mais a peu de conséquences sur leur utilisation en épidémiologie compte tenu de la méthodologie *mutiple loci* du MLVA.

2.2. Un défaut de standardisation

Le MLVA souffre d'un défaut de standardisation évident. L'exemple le plus frappant est l'absence complète de nomenclature des VNTRs. Ainsi, pour un marqueur, il est courant de le retrouver dans plusieurs publications différentes mais sous des noms distincts. Par exemple, STTR5 [135], Sal16 [75], SE-5 [136] et SENTR5 [137] sont les déclinaisons nominatives du même microsatellite. Certains auteurs déclarent l'antériorité de la découverte d'un locus et justifient le changement de nom par l'harmonisation de leur propre nomenclature. D'autres, en revanche, omettent de rappeler que le marqueur utilisé a précédemment été décrit sous un nom différent ou ne s'en aperçoivent pas. L'instauration d'une nomenclature standardisée favoriserait la démocratisation du MLVA et sa promotion en tant que méthode de référence. Une nomenclature formalisée n'entraverait nullement le développement de nouveaux protocoles. Nous proposons de définir un nom qui comprendrait les informations essentielles caractéristiques du VNTR en question. Les VNTRs sont, de manière constante, caractérisés par une taille de motif unitaire. Ensuite, ils peuvent être, pour une souche considérée comme référence, qualifiés par une position dans le génome et un nombre de motifs. Puis selon les amorces utilisées, l'amplicon pourra être de taille différent. Enfin, le VNTR est défini par un nom, qui, selon toute logique, devrait correspondre à celui donné lors de sa première description. Ainsi, une nomenclature standardisée informative serait :
X_P_Apb_Bpb_NU avec X le nom originel du VNTR, P sa position dans le génome de référence, A la taille de l'amplicon pour un couple d'amorces donné utilisé sur le génome de référence, B la taille du motif unitaire répété et N le nombre d'unités répétées sur le génome de référence.

La convention de désignation des allèles est parfois responsable de mauvaises assignations engendrant des incompatibilités entre protocoles. Comme recommandé par Gilles Vergnaud et Christine Pourcel, une souche de référence doit être utilisée pour signaler explicitement quelle convention est utilisée [134]. De plus, les résultats des analyses *in silico*, obtenues par le logiciel TRF, montrent rarement des allèles avec des nombres entiers d'unités répétées. Par exemple, le VNTR Sa0266 de *S. aureus* est indiqué à 5.6U pour la souche Mu50. Cet allèle ne s'apparente pas à un demi-motif rare car il est décliné dans toutes les souches séquencées sous cette forme tronquée. Plusieurs

convention de codage sont possibles : a) l'allèle 5.6U est maintenu, ce choix est le moins pertinent et le moins commode car les demi-motifs seront la généralité pour le VNTR étudié ; b) l'allèle est nommé 5U, ce choix est inapproprié car il risque de conduire à l'identification d'allèles nuls ; c) l'allèle est arrondi à 6U. La majorité des protocoles publiés propose un arrondi au supérieur des allèles identifiés par TRF, cependant d'autres utilisent la méthode de la troncature conduisant à un désaccord de une unité répétée selon la nomenclature adoptée.

Aucune convention ne régit les modalités de codage des allèles MLVA. Certains auteurs répertoriaient les allèles selon la taille de l'amplicon obtenu. Ce principe a rapidement été abandonné car les données différaient en fonction des amorces utilisées rendant alors complexe l'harmonisation des données. Les protocoles classiques de MLVA présentent les données sous forme de nombre de motifs répétés, issus de la conversion de la taille des amplicons. Quel que soit le protocole utilisé, les allèles seront communs et compréhensibles par l'ensemble des utilisateurs. Cependant, la conversion entre l'allèle brut (la taille de l'amplicon) et l'allèle final (nombre d'unités répétées) doit respecter certaines règles établies. Même si le modèle évolutif et la dynamique des VNTRs restent méconnus, la désignation du nombre d'unités répétées révèle une signification biologique. La valeur biologique apportée par les données MLVA est masquée lorsqu'un codage sous forme de lettres est défini [138]. Cette nomenclature alphabétique est très restrictive et peut porter à confusion ; elle n'est d'ailleurs que très peu utilisée.

3. Le MLVA, une méthode adaptée

Malgré l'interrogation sur la concordance épidémiologique et le problème de standardisation de la méthodologie, le MLVA s'impose comme méthode de référence pour le génotypage de plusieurs espèces bactériennes. En criminalistique microbienne (*microbial forensics*), la traçabilité des principaux agents de bioterrorisme, *B. anthracis*, *Y. pestis* et *F. tularensis*, est assurée par MLVA et typage SNPs limité ou génome complet. En microbiologie agro-alimentaire, le réseau PulseNet de surveillance américain des toxi-infections alimentaires propose, comme méthode alternative à la PFGE, l'utilisation de protocoles MLVA standardisés pour le génotypage de *S. enterica* et *E. coli*. Pourquoi les VNTRs sont-il des marqueurs d'intérêt ? Le MLVA s'imposera-t-il en tant que méthode de référence ?

3.1. Les VNTRs, des marqueurs de choix

Les loci VNTRs sont considérés comme stables avec un taux de mutation classiquement compris entre 10^{-4} et 10^{-6}. La stabilité des VNTRs est très variable selon les loci : elle dépend des mécanismes évolutifs et des différents facteurs de stabilité décrits précédemment. Cette stabilité peut être évaluée *in vitro* par la méthode PSPE (*Parallel Serial Passage Experiment*) [139]. La PSPE est

dérivée de la SPE (Serial Passage Experiment), méthode qui, comme son nom le laisse entendre, consiste à effectuer plusieurs passages de repiquages en série d'une culture bactérienne afin de suivre les changements génotypiques et/ou phénotypiques. Girard et collaborateurs ont repris et décliné le procédé en le répétant plusieurs fois en parallèle. Dans la pratique, la PSPE permet de générer x fois plus de générations que la SPE, x étant le nombre de répétitions expérimentales parallèles. La PSPE conjugue la robustesse du SPE et la puissance de l'expérimentation parallèle. De plus, certaines collections de référence intègrent des panels dits de stabilité, constitués par des intermédiaires de repiquage successifs d'une souche en laboratoire, qui évaluent la stabilité des VNTRs. Par ailleurs, des études *in vivo* de suivi longitudinal d'une population bactérienne dans un environnement limité permettent de suivre les changements génétiques et d'estimer la stabilité des marqueurs utilisés.

Quelle que soit la nature de l'espèce bactérienne, les protocoles de génotypage par MLVA montrent un pouvoir de discrimination élevé voire très élevé. La puissance du MLVA repose sur le choix et le nombre de VNTRs utilisés. L'utilisation d'outils efficaces pour la sélection de répétitions en tandem polymorphes et la validation sur une collection de référence permettent de définir un panel de marqueurs pertinents. La combinaison des allèles des différents VNTRs rend presque infini le nombre théorique de profils MLVA identifiables. Un protocole MLVA comprend l'analyse de 8 à 15 VNTRs, ceux-ci ayant en moyenne, et empiriquement, 4 allèles. En dehors de toute considération théorique, l'indice de discrimination, selon l'index de Simpson, vérifie les recommandations de l'ESGEM, et, de surcroît, est souvent supérieur à ceux des méthodes dites de référence.

Le génotypage par MLVA présente l'avantage de reposer uniquement sur la technique de PCR. La méthode bénéficie donc de la puissance et de la robustesse de ce procédé garantissant sa haute reproductibilité. Les VNTRs choisis comme marqueurs moléculaires sont conservés parmi l'ensemble des génomes séquencés d'une même espèce. Les VNTRs font donc partie du *core genome* ou génome commun de l'espèce considérée. L'absence d'un locus dans une souche est rarement répertoriée dans la littérature, une donnée manquante est le plus fréquemment la conséquence d'un mésappariement des amorces utilisées pour l'amplification du locus. Les régions flanquantes de la répétition en tandem sont généralement conservées, cependant certaines espèces très diverses génétiquement comme *L. pneumophila* ou *P. aeruginosa* peuvent poser des difficultés d'amplification. Le concepteur du protocole devra donc adapter les amorces quitte à les générer au niveau de zones codantes davantage conservées. Les résultats du MLVA sont produits sous forme d'un code numérique échangeable et facilement stockable dans des bases de données : le MLVA est donc, par essence, une méthode portable.

3.2. Un protocole modulaire

Le MLVA, à l'instar du MLST, examine le polymorphisme de plusieurs loci ciblés de manière individuelle. Les protocoles de typage peuvent donc être adaptés selon la question biologique posée :

on parle de protocole modulaire. Le MLVA peut se décliner selon une démarche progressive appliquée de manière hiérarchique par la constitution de panels de marqueurs. Plusieurs panels, *i.e.* sous-ensembles de VNTRs, sont élaborés selon la problématique épidémiologique. Un premier panel de loci composé essentiellement de minisatellites peu polymorphes, très stables, signatures de *clusters* phylogénétiques est pertinent pour affiner les relations de parentés entre isolats et définir des complexes clonaux afin d'établir un suivi épidémiologique en biosurveillance. Cet ensemble traduit la réalité structurelle de la population et, de ce fait, doit être concordant avec des méthodes de type MLST. Ensuite, un deuxième panel élaboré à partir de microsatellites polymorphes, plus instables, ne reflétant pas la structure de la population étudiée, affine la distinction entre les isolats et établit alors la puissance résolutive du MLVA. Ce panel, utilisé seul, n'a aucun sens phylogénétique et ne témoignera d'aucune valeur de *clustering*. Cependant, si la question est micro-épidémiologique et l'objectif la caractérisation fine et aboutie des isolats récoltés lors d'un cas épidémique, alors ce panel dit de discrimination sera choisi pour déterminer la singularité de ces isolats. En somme, le premier panel permet la caractérisation de la population analysée : c'est l'épidémiologie descriptive. Le deuxième garantit la résolution des différents individus au sein de cette population : c'est l'épidémiologie d'intervention. La combinaison des deux confère au MLVA une modularité unique en son genre. L'instabilité et la discrimination différentielles des VNTRs sont souvent discutées dans les articles décrivant des protocoles MLVA mais très peu d'auteurs regroupent les loci selon ces caractéristiques. Des essais de distinction de panels adaptés ont été accomplis pour le typage de *Brucella spp.* [140] et *M. tuberculosis* [141].

3.3. Des bases de données disponibles

L'accès à des bases de données est essentiel pour garantir la qualité d'une méthode de typage. Historiquement, la première base de données MLVA développée et rendue publique fut celle de l'Institut de Génétique et Microbiologie [64]. Mise en ligne en 2002, cette base, *The Orsay bacterial genotyping page*, était un prototype de la base actuelle MLVA bank accessible via l'URL http://minisatellites.u-psud.fr/MLVAnet/ [142]. Cette base de données comporte deux volets d'accessibilité : *public databases* et *private databases*. La version publique de la base MLVAbank permet d'accéder aux résultats de typage MLVA d'isolats de quatorze espèces bactériennes différentes ; MLVAbank est d'ailleurs la base de données MLVA la plus généraliste aujourd'hui disponible. Après avoir choisi son espèce de prédilection, la base permet de sélectionner un ensemble de VNTRs parmi ceux proposés dans une liste. Ce choix est inhérent à l'aspect modulaire des protocoles de génotypage par MLVA. L'utilisateur indique alors le génotype MLVA d'un isolat typé et la base donne les isolats les plus proches génétiquement pour les loci considérés. La version privée de la base est unique en son genre et offre à l'utilisateur une certaine confidentialité de ses données. MLVAbank fournit tous les outils nécessaires à la comparaison des génotypes et à la production de

dendrogrammes selon différents algorithmes. Sur le même modèle, plusieurs bases de données MLVA ont été développées et ont vu le jour entre 2005 et 2010. Trois grandes bases généralistes ont ainsi été créées et mises en ligne par de grandes institutions : l'H.I.A. Robert Picqué (www.mlva.eu), l'Institut Pasteur (www.pasteur.fr/recherche/genopole/PF8/mlva) et RIVM (http://www.mlva.net/default.asp). Parmi ces bases, celle du RIVM semble être la plus dynamique et la plus riche en profils MLVA (plus de 3000 pour *S. aureus* et plus de 1000 pour *S. pneumoniae* par exemple). La multiplicité de ces bases de données peut constituer un frein à l'utilisation extensive du génotypage par MLVA. La formation d'un consensus regroupant l'ensemble des données stockées apporterait sans aucun doute un avantage concurrentiel face au MLST, principale technique alternative. Peu de bases de données spécialistes, *i.e.* spécifiques d'une bactérie en particulier, existent : MLVA for *Enterococcus faecium* de l'Université d'Utrecht [143] et MIRU-VNTRplus hébergée par l'Université de Munster en Allemagne [144]. MIRU-VNTRplus, base polyphasique intégrant les données de plusieurs méthodologies de typage, est considérée comme référence pour la comparaison de profils MLVA pour *Mycobacterium tuberculosis* [144].

Positionnement du travail expérimental

L'analyse génomique a eu un impact considérable sur la façon de mener les investigations épidémiologiques et d'appréhender la structure des populations bactériennes. Ces analyses détectent le polymorphisme de certaines régions génomiques, source de variabilité au sein d'une même espèce bactérienne. L'exploitation de ces différents polymorphismes à des fins de discrimination est communément nommée typage ou, par extension, génotypage lorsque les objets polymorphes étudiés sont de nature génétique. La question pour l'épidémiologiste est de savoir parmi les nombreuses méthodes existantes laquelle utiliser. Pour cela, différents critères ont été élaborés ; la distinction des différentes méthodes selon ces paramètres a été présentée dans l'introduction. Ensuite le choix de la méthode de typage pourra être orienté selon la typologie de l'investigation épidémiologique, qu'elle soit d'intervention ou descriptive. Le MLVA, méthode tirant profit du polymorphisme de répétitions des VNTRs, combine la majorité des critères requis. De plus, cette technique semble appropriée aux différentes facettes de l'analyse épidémiologique. Néanmoins, cette méthode est souvent négligée. Comment rendre la méthodologie accessible ? L'automatisation et la standardisation du MLVA sont des éléments essentiels pour que les laboratoires chargés du suivi épidémiologique puissent changer leurs habitudes. La conception d'un outil clé en main, sous forme d'un kit de génotypage par exemple, dont les résultats sont facilement exportables, devrait être envisagée. La manipulation d'un tel outil ne doit dépendre d'aucune expertise technique particulière, un manuel d'utilisateur détaillé suffirait pour que le manipulateur puisse réaliser les différentes étapes du procédé.

Au cours de cette thèse, des protocoles de génotypage automatisés et standardisés ont été développés pour trois espèces pathogènes : *Legionella pneumophila*, *Pseudomonas aeruginosa* et *Staphylococcus aureus*. Ces trois espèces n'ont pas été choisies par hasard. Le choix de ces bactéries a principalement été orienté par la connaissance que le laboratoire GPMS avait sur leur diversité et leur structure. La plupart des VNTRs avaient préalablement été identifiés et des études épidémiologiques ont montré l'intérêt du MLVA. Un accès public dans la base de données MLVAbank permet de faire des requêtes sur les génotypes intégrés, d'étudier leur distribution et de comparer leurs caractéristiques. Sous l'égide du processus CIFRE (Convention Industrielle de Formation par la Recherche en Entreprise), les travaux entrepris et les technologies développées dans le cadre de cette thèse devaient également être en adéquation avec les activités de l'entreprise. La société Ceeram, au sein de laquelle j'ai effectué en partie ma thèse, est un laboratoire privé d'analyse génétique, spécialisé dans la détection, l'identification et la caractérisation d'agents microbiens et dont l'activité est centrée autour de trois secteurs : santé, agro-alimentaire et environnement. Le développement d'outils de génotypage pour les trois pathogènes cités s'intègre donc parfaitement aux secteurs d'activité de l'entreprise. Enfin, ces trois espèces bactériennes sont des modèles représentatifs des structures de populations décrites dans l'introduction : *L. pneumophila* présente une structure mixte, *P. aeruginosa* est fortement panmictique et *S. aureus* est une bactérie clonale.

A la lumière des différentes problématiques avancées dans l'avant-propos et du contexte défini dans l'introduction, j'ai initié ma thèse avec plusieurs interrogations. Quelles sont les contraintes

techniques limitantes du MLVA et par quels procédés techniques les contourner ? Un kit de génotypage est-il concevable ? Si oui, comment serait-il conditionné et quels seraient ses éléments de validation ? L'objectif initial était donc de mettre à disposition un outil à moindre coût permettant le génotypage de routine et de le valider sur des collections de souches pertinentes. Enfin, les études épidémiologiques menées, à la fois prospectives et rétrospectives, descriptives et d'intervention, permettent d'évaluer la capacité du MLVA à définir la structure de populations bactériennes aussi diverses que celles de *L. pneumophila*, *P. aeruginosa* et *S. aureus*.

Articles

1. Technologie développée

1.1. Etat de l'art

Pour plusieurs raisons invoquées en introduction, le MLVA possède tous les atouts nécessaires pour répondre aux différentes exigences de l'épidémiologie des maladies infectieuses. Cependant, la diffusion du MLVA est limitée par plusieurs contraintes techniques. Les VNTRs analysés doivent être amplifiés de manière indépendante. Pour des protocoles classiques qui comprennent l'analyse de 8 à 16 marqueurs, le génotypage d'une dizaine d'isolats nécessitera, de manière laborieuse, une centaine de réactions PCR. De plus, l'analyse des tailles des loci est classiquement réalisée par électrophorèse sur gel d'agarose. Il paraît difficile d'envisager d'entreprendre le typage systématique par MLVA des nouveaux isolats identifiés dans un contexte de suivi épidémiologique en utilisant le même support. L'optimisation du procédé et le développement d'appareillages dédiés doivent alors obligatoirement être envisagés. Enfin, la conversion des tailles d'amplicons en nombre d'unités répétées s'effectue par l'expertise des bandes sur gel d'agarose. La production des génotypes MLVA est donc soumise à subjectivité. En conclusion, il existe un réel besoin dans le développement d'un procédé MLVA permettant un typage rapide, au pouvoir de discrimination élevé, reproductible et à haut-débit des bactéries pathogènes dans le cadre d'une application en typage de routine. Cette demande avait été considérée et quelques tentatives de développement de protocoles automatisés ont été réalisées [74,137,145,146,147,148,149,150]. J'ai bénéficié du recul apporté par ces différents travaux pour définir la méthodologie la plus pertinente afin d'automatiser et de standardiser les protocoles MLVA.

1.2. Développement technologique

Afin de pallier aux contraintes techniques, les VNTRs identifiés sont amplifiés simultanément par PCR multiplexe afin de limiter le caractère fastidieux du MLVA. Le premier avantage de cette adaptation technique est la réduction du coût et la diminution du temps d'obtention et d'analyse des résultats. La réalisation d'une PCR multiplexe rencontrant de nombreux obstacles, notre procédé a résulté de plusieurs étapes d'optimisation dans le choix des amorces et la conception du protocole PCR. Les différents critères sont détaillés dans le brevet « Method for genotyping *Staphylococcus aureus* » (WO/2011/151586). Il est exclu d'analyser le résultat de la PCR multiplexe sous forme d'un profil de bandes, souvent peu reproductibles et dont la standardisation est hasardeuse. Le procédé a donc été conçu pour une application du typage avec analyse des fragments amplifiés par électrophorèse capillaire. Une amorce sens est donc marquée avec un fluorophore défini par un spectre d'excitation unique et permet alors d'associer un VNTR donné à une gamme de couleur lors de la visualisation des profils électrophorétiques. De plus, les amorces ont été élaborées de telle sorte que chaque VNTR soit associé à une gamme de taille connue. Ainsi, un plan de multiplexage calibré par

les deux dimensions, spectrale, assurée par le marquage des amplicons, et spatiale, garantie par le choix pertinent des amorces, permet d'identifier d'une manière unique chaque VNTR. Les amorces utilisées ont donc la triple particularité i) de garantir l'unicité d'un VNTR sur un profil électrophorétique par la caractérisation du vecteur taille/couleur, ii) de permettre leur association dans une seule réaction de PCR (la formation de structures en dimères ou *hairpin* entre les amorces a été évitée) et iii) de s'hybrider au niveau de zones conservées chez l'espèce étudiée. Le profil électrophorétique obtenu peut être analysé avec plusieurs logiciels différents disponibles en libre accès (Peak Scanner, CLC Sequence Viewer,…). Cependant, le traitement des données électrophorétiques ainsi que l'automatisation des résultats de typage ne sont possibles que par l'utilisation du logiciel Genemapper (Applied Biosystems). Ce logiciel analyse le profil électrophorétique, interprète les données brutes et attribue une taille à chaque amplicon marqué par un fluorochrome. Un utilisateur non initié peut alors obtenir simplement le code MLVA de la souche typée avec pour chaque locus un indice de qualité indiquant le degré de confiance attribué à la valeur numérique associée, reflet du nombre de motifs répétés pour le VNTR considéré. Les différentes contraintes ainsi que les actions menées pour les contourner sont schématisées sur la Figure 6 ci-dessous.

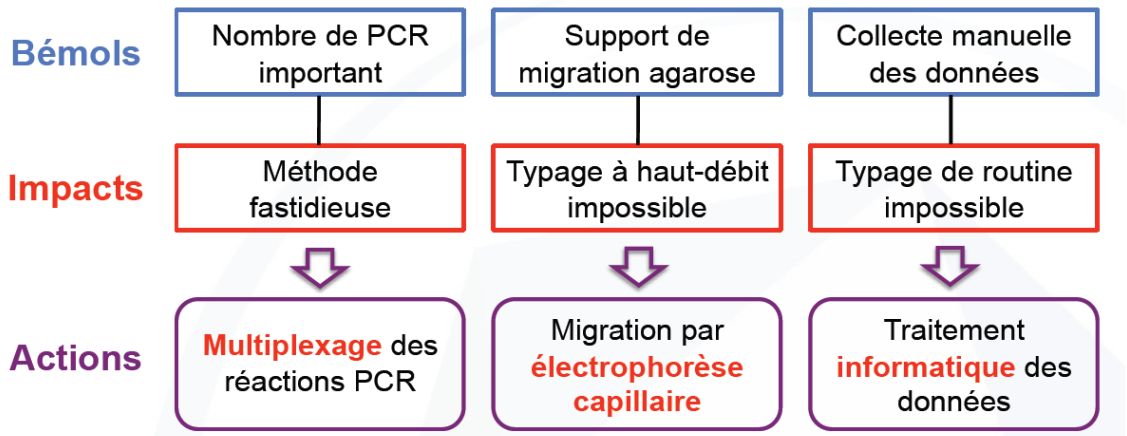


Figure 6. Automatisation et standardisation du MLVA

Cette méthodologie a été appliquée aux espèces *L. pneumophila*, *P. aeruginosa et S. aureus*, et a donné lieu à la conception de kits de génotypage respectivement typlegio@ceeramTools™, typpseudo@ceeramTools™ et typstaph@ceeramTools™. Ces kits sont des trousses complètes prêtes à l'emploi pour génotyper les isolats des espèces considérées à partir d'échantillons d'ADN extraits et purifiés. Ils contiennent les réactifs nécessaires à l'amplification des VNTRs au sein de protocoles développés et validés. Les kits comprennent un ou deux mélanges réactionnels permettant la coamplification de l'ensemble des VNTRs, un mix d'amplification ainsi que les témoins négatifs et positifs indispensables. Ces kits sont une première étape vers l'industrialisation du génotypage et sa généralisation dans les laboratoires de microbiologie.

2. Validations et applications épidémiologiques

2.1. *L. pneumophila* et écotype

a. Résumé

L. pneumophila est l'agent étiologique de la légionellose, maladie respiratoire dont la forme la plus grave se traduit par une pneumopathie aiguë parfois mortelle et de la fièvre de Pontiac. Cette bactérie hydro-tellurique thermophile présente la particularité de coloniser aussi bien les eaux naturelles que les milieux hydriques artificiels comme les circuits des tours aéroréfrigérantes ou les réseaux d'eaux chaudes sanitaires. L'extraordinaire capacité d'adaptation développée par *L. pneumophila* ainsi que la multiplicité des milieux colonisés rendent difficile l'identification de la source de la contamination ainsi que le suivi du pathogène incriminé. Un protocole de typage haut-débit permettant le suivi épidémiologique doit donc être envisagé.

Le système MLVA développé, le MLVA-12_{Orsay}, repose sur l'analyse de 12 VNTRs coamplifiés simultanément en une seule réaction de PCR. La procédure employée permet d'obtenir, à partir d'un ADN d'une souche bactérienne, un code numérique unique et reproductible en moins de quatre heures. Le MLVA-12_{Orsay} a été validé sur une collection de référence de 96 souches cliniques. Les critères de performance obéissent aux recommandations établies par l'ESGEM. Ainsi, l'indice de discrimination calculé est similaire à celui obtenu par le SBT, la méthode de référence. Le panel dit de concordance épidémiologique confirme la pertinence du MLVA-12_{Orsay} lors d'une investigation épidémique. La congruence avec le SBT et l'assignation des souches typées à différents complexes clonaux suggèrent que l'homoplasie des VNTRs est atténuée ; le MLVA est une alternative au SBT pour l'épidémiologie descriptive.

Cette procédure a été appliquée à l'étude de 217 isolats environnementaux dans le cadre du suivi de populations de *L. pneumophila* entre 2000 et 2009 dans la ville de Rennes. Nous avons montré la prépondérance d'un clone non impliqué dans des cas épidémiques rennais. En effet, ce clone, bien que responsable de plus de 10% des cas de légionellose en Europe, colonise de manière exclusive et persistante les réseaux d'eaux rennais sans être lié aux épidémies ou aux cas sporadiques détectés dans la région. Cette observation confirme l'existence d'écotypes dans la population de *L. pneumophila* dont la pathogénicité semble être modulée par des facteurs environnementaux. De plus, la mise en évidence de la colonisation exclusive sur plus de huit ans d'un même génotype est un argument pour la stabilité des VNTRs choisis. Enfin, cet article justifie l'emploi du MLVA12_{Orsay} comme outil de traçabilité et d'évaluation du risque : la source des épidémies déclarées de légionellose en 2000 et 2006, une tour aéroréfrigérante d'un centre commercial, est révélée ; la présence de souches endémiques virulentes dans les tours aéroréfrigérantes d'industriels ou d'hôpitaux est signalée afin que des mesures sanitaires soient rapidement engagées.

b. ***Article 1*** ***(publié par Applied and Environmental Microbiology)***

High-throughput typing method to identify a non-outbreak-involved *Legionella pneumophila* strain colonizing the entire water supply system in the town of Rennes, France.

Sobral D.[1,2], Le Cann P.[3], Gerard A.[3], Jarraud S.[4], Lebeau B.[2], Loisy-Hamon F.[2], Vergnaud G.[1,5], Pourcel C.[1*]

Université Paris-Sud, Institut de Génétique et Microbiologie, Orsay, F-91405, France; CNRS, Orsay, F-91405, France[1]

Ceeram (Centre Européen d'Expertise et de Recherche sur les Agents Microbiens), Allée de la Filée, 44244 La Chapelle sur Erdre, France[2]

LERES (Laboratoire d'Etude et de Recherche en Environnement et Santé), EHESP (Ecole des Hautes Etudes en Santé Publique), Avenue du Professeur Léon-Bernard, 35043 Rennes, France[3]

Centre National de Référence des Legionella, Université de Lyon, INSERM U851, Faculté de Médecine, IFR 128, Lyon, France[4]

DGA/MRIS- Mission pour la Recherche et l'Innovation Scientifique, 92221 Bagneux, France[5].

ABSTRACT

Two legionellosis outbreaks occurred in the city of Rennes, France, during the past decade, requiring in-depth monitoring of *Legionella pneumophila* in the water network and the cooling towers in the city. In order to characterize the resulting large collection of isolates, an automated low-cost typing method was developed. The multiplex capillary-based variable-number tandem repeat (VNTR) (multiple-locus VNTR analysis [MLVA]) assay requiring only one PCR amplification per isolate ensures a high level of discrimination and reduces hands-on and time requirements. In less than 2 days and using one 4-capillary apparatus, 217 environmental isolates collected between 2000 and 2009 and 5 clinical isolates obtained during outbreaks in 2000 and 2006 in Rennes were analyzed, and 15 different genotypes were identified. A large cluster of isolates with closely related genotypes and representing 77% of the population was composed exclusively of environmental isolates extracted from hot water supply systems. It was not responsible for the known Rennes epidemic cases, although strains showing a similar MLVA profile have regularly been involved in European outbreaks. The clinical isolates in Rennes had the same genotype as isolates contaminating a mall's cooling tower. This study further demonstrates that unknown environmental or genetic factors contribute to the pathogenicity of some strains. This work illustrates the potential of the high-throughput MLVA typing method to investigate the origin of legionellosis cases by allowing the systematic typing of any new isolate and inclusion of data in shared databases.

INTRODUCTION

Legionella pneumophila, an aerobic, non-spore-forming, Gram-negative bacillus, is the main etiologic agent of legionellosis, a severe respiratory disease transmitted by inhalation of aqueous aerosols. *L. pneumophila* can be found in freshwater sources, such as rivers, lakes, or ponds, where it multiplies in free-living amoebae and ciliated protozoa. The bacteria also have a widespread distribution in human-made environments (1). Most legionellosis outbreaks are linked to contaminated hot water systems or cooling towers (16). The bacteria can persist in these artificial aquatic environments through the formation of biofilms that protect Legionella populations from temperature changes and biocide treatment (25). Protozoa serve as a reservoir for legionellae, and it is thought that growth within amoebae enhances pathogenicity (6, 34).

L. pneumophila is a very diverse and heterogeneous species which is divided into three subspecies (*L. pneumophila* subsp. pneumophila, *L. pneumophila* subsp. fraseri, and *L. pneumophila* subsp. pascullei). *L. pneumophila* can be differentiated into 17 serogroups (sg), sg1 being responsible for more than 98% of the reported legionellosis cases (12). The *L. pneumophila* population is considered clonal, and the adaptation of clonal complexes (CC) to a specific ecological niche was suggested (14). Studies reported that *L. pneumophila* is capable of plasmid-mediated recombination exchange of DNA material (13) and demonstrated the occurrence of horizontal genetic transfer (9, 10, 14). Conclusions about the *L. pneumophila* population structure might be biased since the strains investigated are usually recovered from clinical cases or human-made environments. The *L. pneumophila* population is supposed to be much more diverse, and the observed clonal complexes may be a nonrandom set that represents specific lineages adapted to human habitats (32).

Due to widespread *L. pneumophila* occurrence in both natural and artificial aquatic systems, it is essential, in order to allow the implementation of prevention measures, to identify potential environmental sources of infection by comparing clinical and environmental isolates (31, 35). The investigation and characterization of large collections require low-cost typing procedures. The current typing method recommended by the European EWGLI Consortium is a multilocus sequence typing (MLST)-like protocol called sequence-based typing (SBT) (18, 30). This approach is highly portable and widely used to perform epidemiological surveys and outbreak investigation (27) owing to the existence of databases accessible over the internet. However, SBT remains tedious and time-consuming, as it requires the sequencing of both strands of seven gene fragments. For this reason, its use as a first-line assay for large-scale investigations seems precluded. Multiple-locus variable-number tandem repeat (VNTR) analysis (MLVA) is a PCR-based typing method that relies on the variability found in some tandemly repeated DNA sequences which represent sources of genetic polymorphism (37). Ten VNTRs suitable for *L. pneumophila* MLVA genotyping were previously reported (29), and a selection of 8 loci comprising Lpms01, Lpms03, Lpms13, Lpms17, Lpms19, Lpms33, Lpms34, and Lpms35 and compatible with low-resolution DNA sizing equipment (such as agarose gels) was proposed. The MLVA protocol based on these eight VNTRs and called MLVA-8_{Orsay} was useful for

both epidemiological studies and analysis of population structure (28, 29). Indeed, recent work showed that a majority of clinical strains were distributed into a limited number of CCs defined by MLVA, called VNTR analysis CC (VACC) and characterized by epidemic strains such as Paris (VACC1), Philadelphia-1 (VACC2), and Corby (VACC9) (38).

In the city of Rennes, one of the largest cities in the west of France, with a population of 200,000 inhabitants, two legionellosis outbreaks occurred in the past 10 years. The first outbreak happened in autumn 2000 and caused 22 legionellosis cases, including 4 deaths. Public health authorities adopted a plan in March 2001 to assess and prevent Legionella-associated risks. Despite the implementation of this plan, a second outbreak occurred in January 2006 and led to 2 deaths among the 8 diagnosed cases (4). In the context of water networks and cooling towers monitoring for the presence of *L. pneumophila*, isolates are regularly collected from multiple sites representative of the whole water supply system in Rennes. Here we genotyped 217 isolates collected in Rennes between 2000 and 2009, revealing that strains involved in the two outbreaks were not related to the few clones which colonize the whole water supply system in this city. For this, we automated the MLVA procedure by coamplifying 12 VNTRs (MLVA-12_{Orsay}) in a single PCR and performed fluorescent capillary electrophoresis. Compared to MLVA-8_{Orsay}, the new protocol provides increased informativity.

MATERIALS AND METHODS

Strains. A total of 320 isolates were used, including 95 isolates from the EWGLI reference collection. This collection contains clinical and environmental isolates originating from 10 countries and is composed of one epidemiologically unrelated panel of 74 isolates (panel 1), one epidemiologically related panel of 16 isolates (panel 2), and one stability panel of 5 isolates (panel 3) (17). All isolates from this collection belong to serogroup 1 (sg1) and have been characterized extensively by other genotypic and phenotypic methods. In addition, three reference strains for which the genome has been sequenced (Philadelphia-1, Paris, and Lens) were used as controls. The type strain of *L. pneumophila*, Philadelphia-1 (NCTC 11192), was obtained from the National Collection of Type Cultures, London, United Kingdom. The Lens and Paris reference strains have been deposited by the French National Reference Center for Legionella in Lyon, France, in the EWGLI collection (EUL 160 and EUL146, respectively). Five clinical isolates came from two epidemic cases that occurred in autumn 2000 and winter 2006 in Rennes. Two hundred and seventeen environmental isolates collected from human-made sources (cooling towers and water distribution systems) in Rennes between 2000 and 2009 were selected from the collection maintained by the Laboratoire d'étude et de Recherche en Environnement et Santé (LERES) in Rennes, France (see Table S1 in the supplemental material).

DNA extraction. Strains were cultured at least 2 days at 36 ± 2°C on buffered charcoal-yeast extract (BCYE) with l-cysteine. Genomic DNA was extracted using the DNeasy blood and tissue kit

(Qiagen, Courtaboeuf, France). DNA concentration was measured using an ND-1000 spectrophotometer (NanoDrop; Labtech, Palaiseau, France) and adjusted to 3 ng/µl with water.

Selection of VNTRs. The typing procedure is based on the analysis of 12 VNTRs, 8 of which were published by Pourcel and colleagues (29). Four additional VNTRs (Table 1) were selected from the four available *L. pneumophila* genomes using the Microorganisms Tandem Repeats Database (20): Lpms38 (8-bp repeat unit), Lpms39 (6-bp repeat unit), Lpms40 (6-bp repeat unit), and Lpms44 (6-bp repeat unit). They were selected among the 10 microsatellites with a 6- to 8-bp repeat unit and at least a 60% average internal conservation and showing at least two different alleles among the four sequenced genomes and were tested on panel 1 from the EWGLI collection. Lpms17 was excluded from the present MLVA scheme because of its low informativity, whereas Lpms31, which is difficult to type on agarose gel, was incorporated. As a result, the extended MLVA scheme, MLVA-12_{Orsay}, includes 12 VNTRs: eight are minisatellites (Lpms01, Lpms03, Lpms13, Lpms19, Lpms31, Lpms33, Lpms34, and Lpms35) and 4 are microsatellites (Lpms38, Lpms39, Lpms40, and Lpms44).

VNTR name	Primer name	Sequence (5’ – 3’)	Repeat size (bp)	Expected size (bp) (no. of repeats) for strain Philadelphia-1
Lpms01	Lpms01R	GCATATGACAAAGCCTTGGC	45	494 (8)
	Lpms01F	NED-TGAATTTCTCCCTCTTGCTTG		
Lpms03	Lpms03R	TGATGGTCTCAATGGTTCCG	96	942 (8)
	Lpms03F	VIC-GGACAAACAACCAATGAAGC		
Lpms13	Lpms13R	GCATCGGACTGAGCAAAGTA	24	793 (11)
	Lpms13F	NED-CTCACCAGGATGCTTTGTCG		
Lpms19	Lpms19R	TCCAGAGGCTCTGGATTATC	21	128 (4)
	Lpms19F	VIC-GAACTATCAGAAGGAGGCGA		
Lpms31	Lpms31R	ATCGCCTAATTGCCGCCTA	45	1043 (17)
	Lpms31F	6FAM-CCTCGCAAGCCTATGTGG		
Lpms33	Lpms33R	CGAGGAAATCTTCTTCAGCC	125	317 (1)
	Lpms33F	VIC-GACACCACAGCAGTTTGAAC		
Lpms34	Lpms34R	ATGCAGGATGTTTGCGCATG	125	265 (1)
	Lpms34F	6FAM-AAGGAATAAGGCGCAGCAC		
Lpms35	Lpms35R	TATCAACCTCATCATCCCTG	18	205 (3)
	Lpms35F	PET-GAATCTGAAACAGTTGAGGATG		
Lpms38	Lpms38R	GGATTGCCTTGGGCATTAAT	6	264 (3)
	Lpms38F	NED-CCTATCAACAGATGACGCTT		
Lpms39	Lpms39R	CCAACTCCTCAACGCAACAA	6	79 (6)
	Lpms39F	PET-CTTGACGAAGTAGGTGTGGG		
Lpms40	Lpms40R	TTACCCAAGCCCTTATTGCG	6	198 (4)
	Lpms40F	6FAM-TAGATCTCTTGCCGAGCTTC		
Lpms44	Lpms44R	TTATGCGAGAGTTTCATGA	6	173 (9)
	Lpms44F	NED-GCTACTGCAGCAACATCC		

Table 1. Oligonucleotide primers used and VNTRs analysed in this study

MLVA typing. Primers were designed to be able to simultaneously amplify all 12 loci, taking into account the allele size range of each locus previously evaluated by agarose gel electrophoresis (Table 2). The 12 VNTR loci were amplified in a unique multiplex PCR using the genotyping kit Typlegio ceeramTools (Ceeram, Nantes, France). Briefly, this kit includes forward primers labeled at the 5′ end with either 6-carboxyfluorescein (6-FAM), 2′-chloro-7′-phenyl-1,4-dichloro-6-carboxyfluorescein (VIC), 2′-chloro-5′-fluoro-7′,8′-fused phenyl-1,4-dichloro-6-carboxyfluorescein (NED), or PET (Applied Biosystems, Courtaboeuf, France). Reverse primers were synthesized unlabeled and tailed (Applied Biosystems, Courtaboeuf, France). The use of reverse-tailed primers, which promotes the addition of a nontemplated A to the product, efficiently avoided the so-called +A peak artifact (3). The multiplex PCR was performed in a 15-µl final volume using the Qiagen multiplex PCR kit (Qiagen, Courtaboeuf, France). The reaction mixture contained 2 µl template DNA (3 ng/µl), 7.5 µl of 2× multiplex PCR mastermix, and 5.5 µl of primer mix. The PCR was run on a Veriti thermal cycler (Applied Biosystems, Courtaboeuf, France) using the following conditions: initial denaturation cycle for 15 min at 95°C, 15 cycles of touchdown PCR (30 s at 95°C; 60 s at 82°C, with a 1.2°C drop in temperature each next cycle; 70 s at 72°C); and 15 cycles of long-range PCR (30 s at 95°C; 60 s at 64°C; 70 s at 72°C, with a 5-s increase in time each next cycle), with a final 10 min at 72°C. PCR fragments were purified using Qiagen DyeEx plates (Qiagen, Courtaboeuf, France). Then, 2 µl of purified PCR product was combined with 7.75 µl HiDi formamide and 0.25 µl GS1200LIZ (Applied Biosystems, Courtaboeuf, France). The samples were run on the ABI3130 capillary sequencer (Applied Biosystems, Courtaboeuf, France). Electrophoresis was performed using a 50-cm capillary filled with performance-optimized polymer 7 (Applied Biosystems, Courtaboeuf, France) at 60°C for 6,200 s with a running voltage of 12 kV and an injection time of 10 s at an injection voltage of 1.6 kV.

Automated binning was performed with the GeneMapper software (Applied Biosystems, Courtaboeuf, France). Each bin is characterized by the allele mean size plus or minus 10% of the unit size length. Insertion or deletion mutations sometimes arise in flanking regions or directly in tandem repeats and affect the amplicon size. Therefore, we assessed a confidence interval (CI) to overcome these sequence polymorphisms.

VNTR	Allele min (number of repeats, expected size in bp)	Allele max (number of repeats, expected size in bp)	Allele number	HGDI	Standard deviation
Lpms01	6, 404	10, 584	6	0.6501	{0.5821,0.7181}
Lpms03	7, 846	8, 942	2	0.5054	{0.4928,0.5179}
Lpms13	3, 601	17, 937	10	0.7790	{0.7152,0.8428}
Lpms19	4, 128	5, 149	2	0.2936	{0.1788,0.4084}
Lpms31	6, 548	20, 1178	12	0.8563	{0.8061,0.9066}
Lpms33	1, 317	5, 817	6	0.7020	{0.6196,0.7844}
Lpms34	1, 265	3, 515	4	0.6649	{0.6203,0.7096}
Lpms35	3, 205	32, 727	17	0.8815	{0.8416,0.9215}
Lpms38	3, 264	19, 392	4	0.2710	{0.1389,0.4031}
Lpms39	6, 79	26, 199	12	0.8301	{0.7699,0.8902}
Lpms40	4, 198	5, 204	2	0.5054	{0.4928,0.5179}
Lpms44	6, 155	18, 227	6	0.5391	{0.4512,0.6269}
			Genotype number	HGDI[a]	95% confidence interval
MLVA-12$_{Orsay}$			39	0.9534	{0.9234,0.9833}

Table 2. Range and HGDI for individual or combined VNTR loci

[a] calculated from typing results obtained with the 74 *L. pneumophila* strains included in the panel 1 of the EWGLI collection

Data analysis. Each VNTR locus was identified according to color and automatically assigned a size by the GeneMapper software (Applied Biosystems, Courtaboeuf, France). This size was then converted into an allele designation associated with a quality index. Intermediate-size alleles (which may result from intermediate-size repeat units or from small deletions in the flanking sequence) were reported as half size (0.5). The typing datum file was imported into BioNumerics version 6.5 (Applied-Maths, Sint-Martens-Latem, Belgium).

A complete strain allele string expressed as its allelic profile corresponding to the number of repeats at each VNTR was constructed in the order Lpms01, Lpms03, Lpms13, Lpms19, Lpms31, Lpms33, Lpms34, Lpms35, Lpms38, Lpms39, Lpms40, and Lpms44. The genotype of the Philadelphia-1 strain deduced from its genomic sequence NC_002942 is 8-8-11-4-17-1-1 3-3-6-4-9 and that of strain Paris NC_006368 is 7-7-10-4-9.5-4-2-17-3-13-5-6.

Repeatability of the multiplex PCR and of the size measurement by the capillary electrophoresis device was tested in two ways. The 98 isolates of the validation group were typed twice independently using the same DNA samples. In addition, the three reference strains Lens (EUL160), Paris (EUL146), and Philadelphia (NCTC11192) were typed eight times independently using the same DNA samples.

The Hunter-Gaston diversity index (HGDI) (23), an application of Simpson's index of diversity (33), was used as a polymorphism index for individual or combined VNTR loci. CI were calculated as described previously (21). The unweighted pair group method with the arithmetic mean (UPGMA) clustering method was run using the categorical coefficient (also called Hamming's distance). A cutoff value of 60% similarity was applied to define clusters. The minimum spanning tree was produced in BioNumerics, allowing the creation of missing links and scaling with member count. The logarithm scale was used when drawing branches.

RESULTS

Automated multiplex capillary-based MLVA assay development. The new optimized primers and the high-stringency PCR protocol yielded clean electrophoretic profiles without artifactual peaks, such as stutter, spike, or shoulder peaks (Fig. 1). Half-size alleles were observed for Lpms01, Lpms31, and Lpms39. In previous reports, correct assignment of unit repeats was not simple for Lpms31 because of the high frequency of these half-size alleles, and therefore, use of the higher-precision capillary electrophoresis equipment is mandatory for this locus.

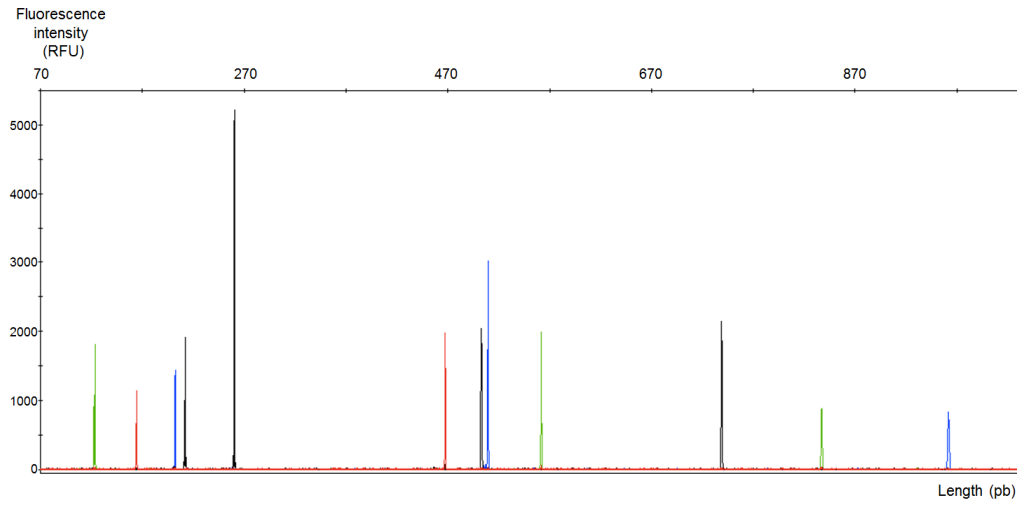


Fig. 1. Electrophoregram showing PCR amplicons of all twelve dye-colored coamplified VNTR loci separated by size and color with capillary electrophoresis.

The Study Group on Epidemiological Markers (ESGEM) refers to two categories of criteria to define the quality of a typing method: performance and convenience criteria (36). The performance criteria include marker stability, typeability (*T* value), discriminatory power combined with epidemiological concordance, and reproducibility (including repeatability). These criteria were evaluated and tested on 98 strains of the EWGLI reference collection, of which 74 are unrelated (panel 1) (17). The dendrogram in Fig. 2 displays the MLVA data and the clustering of strains, allowing an assessment of the different criteria.

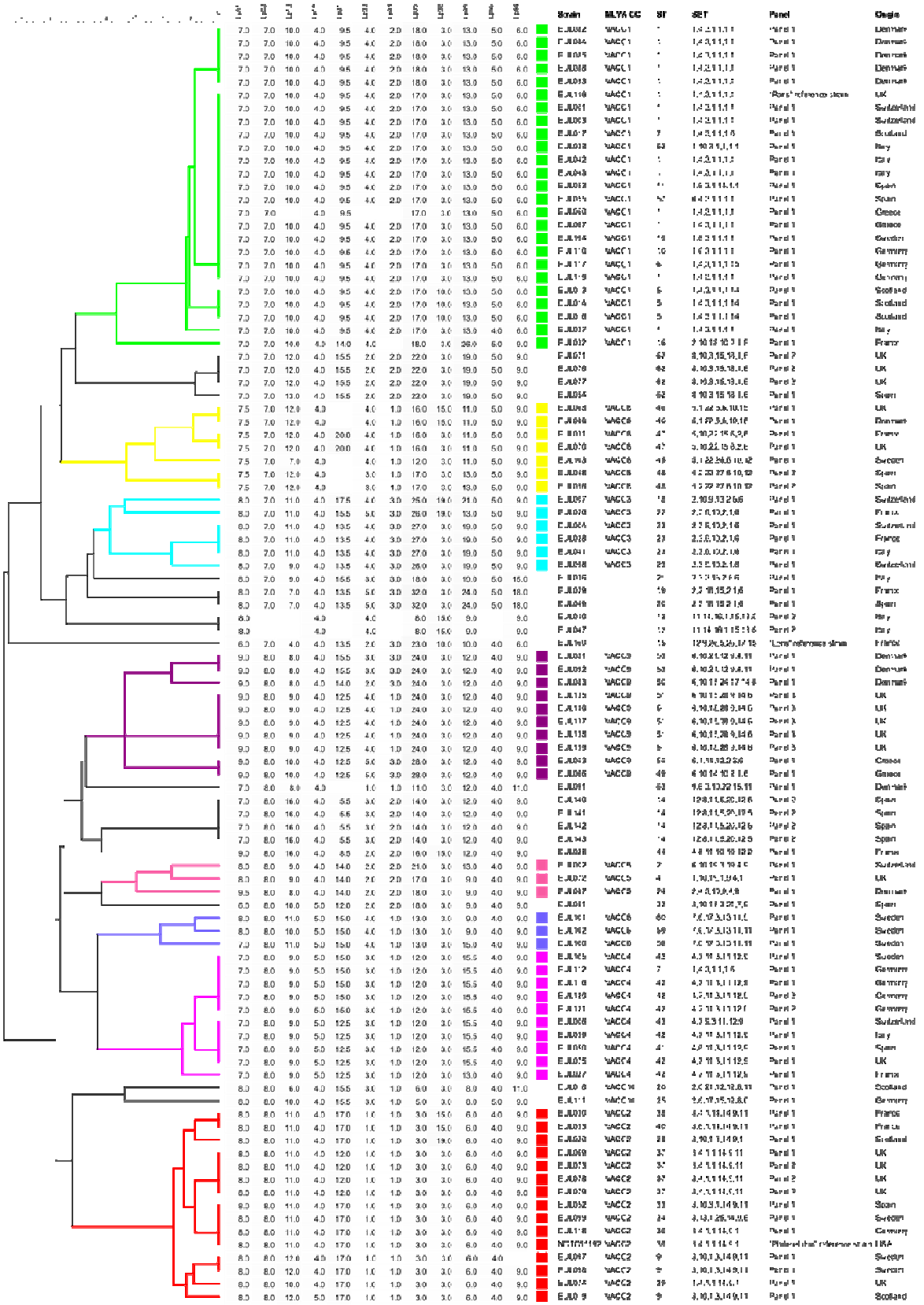


Fig. 2. Dendrogram deduced from the clustering analysis of the 98 strains from the panel 1 of the EWGLI collection using MLVA-12$_{Orsay}$.

Performance criteria. (i) Stability.The stability panel composed of 5 isolates of the Corby strain (EUL135 to EUL139) showed the same MLVA allelic profile.

(ii) *T* value.Out of 1,176 expected datum values, 21 were missing and are shown as an absent value in the dendrogram shown in Fig. 2, corresponding to a typeability, or *T* value, of 98.22%. Two isolates (EUL040 and EUL047) accounted for 10 of these missing data. Excluding these two isolates, the *T* value of MLVA-12_{Orsay} is 99.3%.

(iii) Discriminatory power (HGDI).Indices of discrimination (HGDI value) were calculated for the panel of 74 unrelated strains for each VNTR and for the two combinations of loci (MLVA-12_{Orsay} and MLVA-8_{Orsay}) (Table 2). The MLVA-12_{Orsay} index is 0.9534, a significant increase over the previous MLVA-8_{Orsay} scheme (HGDI = 0.9189). The new protocol distinguished 7 additional genotypes (39 versus 32). This improved resolution was provided primarily by Lpms38, which generated 4 new genotypes.

(iv) Epidemiological concordance.The epidemiologically related panel included 16 isolates divided into 6 groups comprising isolates with an established epidemiological link. Members of each group had the same unique MLVA code.

(v) Repeatability.The 98 isolates of the validation group were typed twice and showed the same genotypic profile. The observed sizes at each locus for the eight trials are listed in Tables S2, S3, and S4 in the supplemental material. All values are included in their respective confidence interval, showing that the described typing protocol was reproducible.

Clustering was performed, and a cutoff value of 60%, corresponding to a maximum of 5 differences out of 12 VNTRs, was applied, defining 10 clusters corresponding to MLVA clonal complexes (VACC) observed with MLVA-8_{Orsay} in a previous work (38). Nine additional genotypes were observed with the EWGLI collection (Fig. 2). The polymorphism provided by the new markers of MLVA-12_{Orsay} (Lpms31, Lpms38, Lpms39, Lpms40, and Lpms44) induced minor changes in the clusters. EUL029, EUL036, and EUL049 were not included in VACC3; EUL071, EUL076, EUL077, and EUL054 formed a group dissociated from VACC1; EUL025 and EUL051 were not included in VACC5.

Long-term epidemiological monitoring of *L. pneumophila* population in Rennes. The MLVA-12_{Orsay} assay was used to type 217 environmental isolates collected from anthropic sources (cooling towers and water distribution systems) between 2000 and 2009 in Rennes and 5 clinical isolates from two epidemic cases that occurred in autumn 2000 and winter 2006 in the same city. The 222 isolates were resolved into 14 MLVA types and distributed into 4 clusters, VACC1, VACC2, VACC6, and VACC8 (see Fig. S1 in the supplemental material). All the tested isolates of VACC1 (22/34), VACC2 (13/16), and VACC6 (128/170) were sg1. The two isolates of VACC8 were non-sg1. Surprisingly, a low diversity (HGDI = 0.5372) was observed, due largely to the high prevalence of VACC6 isolates (77% of all the isolates) in this subset and particularly of genotype 9 (87% of the VACC6 isolates and 67% of all the isolates).

Members of VACC1 (34 isolates) with 4 MLVA genotypes were frequent in cooling towers (Fig. 3). Genotype 2 represented the largest part of VACC1 (80%) and included ST1 isolates (sequence-based type [SBT] 1,4,3,1,1,1,1 [with allele numbers separated by commas]), among which was the endemic Paris strain. Genotype 2 isolates were harvested predominantly in cooling towers (24 isolates out of 31, 77%) and in hot water systems (Fig. 3). This genotype was detected profusely in hospital A's cooling tower and had an extensive colonizing period (from 2001 to 2009). Moreover, genotype 1, also present in hospital A's cooling tower, was likely to have derived from genotype 2 by a change at VNTR Lpms35. No other genotype was found in this cooling tower. Two other variants of genotype 2, each represented by a single isolate, were found in the water supplies of two gymnasiums: genotype 3, an Lpms39 variant, and genotype 4, an Lpms31 and Lpms34 variant. None of the genotypes grouped in VACC1 were reported as outbreak strains in Rennes.

VACC2 with three MLVA genotypes (12, 13, and 14) comprised all 5 clinical isolates and 11 environmental isolates, of which 9 (81%) came from cooling towers (Fig. 3). Genotype 13 is the genotype of reference strain NCTC11192 Philadelphia-1 (ST36, 3,4,1,1,14,9,1). The 5 clinical isolates were previously typed by other methods (unpublished data) and were assigned to the same pulsed-field gel electrophoresis (PFGE) type and to three different SBT sequence types (ST439, 3,4,1,28,14,11,11; ST626, 3,4,1,28,14,11,1; and ST628, 3,4,1, 1,14,13,30). ST439 isolates with MLVA genotype 12 were responsible for two legionellosis cases that occurred during the two outbreaks in 2000 and 2006. The ST626 and ST628 isolates with MLVA genotype 13 were involved in the 2000 outbreak. Epidemiological studies revealed that these isolates were associated with the two cooling towers inside mall A and likely represent a long-term contamination of the mall with the same strain. The main mall cooling tower was colonized by genotype 12 and genotype 13 isolates, whereas only genotype 13 isolates were recovered from the cooling tower of night club A, located inside the mall.

VACC6 is represented here essentially by genotype 9. Six single-locus variants are observed, with genotype 9* differing from 9 by the absence of data at Lpms13 (Fig. 3). All VACC6 isolates were harvested exclusively in hot water systems in diverse facilities, such as schools, swimming pools, spas, hotels, gymnasiums, and malls (Fig. 3), throughout the study period from 2000 to 2009. Genotype 9 was found during each annual sampling campaign. It was sampled from swimming pool A in March 2000 and October 2010 and also from gymnasium A in May 2003 and February 2005. EUL102, a Swedish isolate from the EWGLI collection, also exhibited MLVA genotype 9. Both L0794, an MLVA genotype 9 isolate, and EUL 102 were ST59 (7,6,17,3,13,11,11), which confirmed their genetic relatedness. The six single-locus variants of genotype 9 (shown as 6, 7, 8, 9*, 10, and 11 in Fig. 3) differed at loci Lpms13, Lpms35, and Lpms31 (3, 2, and 1 genotypes, respectively). One of the Lpms13 variants, genotype 8, presented the same MLVA profile as EUL101 (ST60; 7,6,17,3,13,11,9), a strain closely related to genotype 9 strain EUL102, according to SBT typing, illustrating good congruence between MLVA and SBT.

Isolates L1201 and L1202 belonged to VACC8, which also includes the Lorraine strain. They differed from Lorraine type strain EUL070 (ST47; 5,10,22,15,6,2,6) at two markers (Lpms19 and Lpms31), for which PCR amplification failed. L1201 was typed by SBT and was assigned to ST74 (5,1,22,30,6,10,6). This ST has four alleles in common with ST47 and differs by 12, 2, and 9 nucleotide mutations at *pilE*, *mip*, and *proA*, respectively.

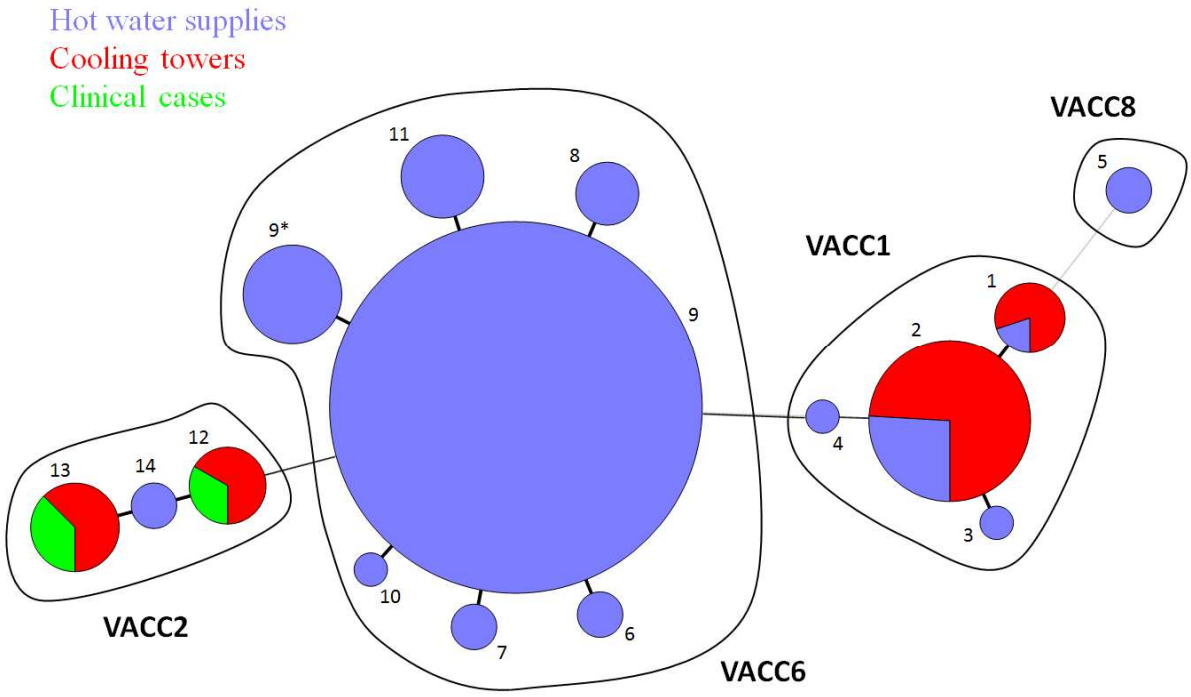


Fig. 3. Minimum spanning tree of the 222 *L. pneumophila* isolates of the Rennes study using MLVA-12_{Ceeram} (distribution according to equipment name). Each circle represents a MLVA genotype (the genotype number is indicated, genotype 9* corresponds to genotype 9 with a missing value at Lpms13). The VACC assignment is represented in bold.

DISCUSSION

Automated capillary-based MLVA system, a powerful genotyping tool. Nederbragt and collaborators described an automated method based upon the MLVA-8_{Orsay} loci. The products of 8 individual PCRs were analyzed by capillary electrophoresis runs (26). We describe here an MLVA-12_{Orsay} assay optimized for capillary electrophoresis analysis, which can be performed in a single PCR. This is, to our knowledge, the first report of a multiplex-based MLVA assay that involves the coamplification of more than nine loci. Taking advantage of the higher sizing precision provided by capillary electrophoresis, we added Lpms31 and four newly identified loci (Lpms38, Lpms39, Lpms40, and Lpms44) to the previous MLVA-8_{Orsay} selection of loci. Lpms17 was excluded because of its lack of informativity.

The clonal complex assignment was realized according to Visca and colleagues (38) in order to allow the distinction of the 10 highlighted CCs. Thus, a cutoff of 60%, which groups in the same CC any isolates that exhibit at least 7 identical markers, was applied. Due to the supplementary information brought with the added markers, some previously VACC-associated isolates were outgrouped with MLVA-12_{Orsay}.

The new VNTRs possess shorter repeat units. Lpms38, despite its low apparent discrimination value (HGDI value = 0.2710) due to the preponderance of the 3-repeat unit (3U) allele (83% of the isolates from EWGLI panel 1 have this allele), allows the differentiation of 2 extra genotypes in VACC2 and 1 extra genotype in VACC1 and VACC8. Its large allele range (from 3 to 19 repeat units) suggests that new alleles could be discovered when studying other *L. pneumophila* populations. Lpms39 has 12 alleles and a high HGDI value (0.83). Lpms40 presents only 2 alleles, with VACC1, VACC3, and VACC8 sharing the 5U allele and VACC2, VACC4, VACC5, VACC6, and VACC10 sharing allele 4U. Finally, Lpms44 offers no advantage for discrimination, at least in the population investigated here, but shows six alleles from 6U to 18U, suggesting the existence of additional alleles in *L. pneumophila*.

MLVA-12_{Orsay} shows excellent reproducibility and typeability. Five markers (Lpms01, Lpms31, Lpms35, Lpms38, and Lpms39) could not be amplified in EUL040 and EUL047. Their ST (ST12; 11,14,16,1,15,13,6) is related to that of reference strains Los Angeles-1 ATCC 33156 (11,14,16,25,7,13,F) and Dallas-1E ATCC 33216 (11,14,16,18,15,13,F) belonging to *L. pneumophila* subsp. fraseri, suggesting that the PCR primers do not perfectly match their genome. Weak amplification for these two isolates was previously reported by Pourcel and colleagues using the primers designed for MLVA-8_{Orsay} (29).

The VNTR markers proved to be stable and epidemiologically concordant. The discrimination power of MLVA-12_{Orsay} is 0.9534, compared to 0.9189 for MLVA-8_{Orsay}. Dendrogram topologies obtained for MLVA-12_{Orsay} and SBT were very congruent. The ESGEM recommends that the diversity value of a genotyping scheme should be at least 0.95. The new MLVA protocol is beyond this prerequisite, even if it still remains less discriminatory than the SBT protocol (HGDI = 0.9608). However, some studies report the difficulty in amplifying the last SBT marker, *neuA*. Removing this marker, the SBT is significantly less discriminatory (HGDI = 0.9256).

The described technical improvements are a major advance toward a routine and easy use of Legionella genotyping. The four steps of the MLVA procedure (DNA extraction, specific coamplification of the 12 VNTRs in a single multiplex PCR, fragment analysis by capillary electrophoresis, and strain code assignment) were standardized to be usable and understandable by nonexpert users. Finally, MLVA-12_{Orsay} results are delivered as an easily sharable and storable numeric code (http://mlva.u-psud.fr).

Epidemiological monitoring of *L. pneumophila* population in Rennes. We investigated strains collected as part of a large-scale epidemiological study to better understand the terms of epidemic outbreaks that occurred in Rennes and to assess the *L. pneumophila* diversity in Rennes' water supply. It represents the largest time-scale MLVA analysis of *L. pneumophila* isolates harvested in a local area. Twenty-three percent of the isolates were clustered with VACC1 (Paris lineage, 34 isolates), VACC2 (Philadelphia-1 lineage, 16 isolates) and VACC8 (Lorraine lineage, 2 isolates). All the others gather in the previously defined VACC6 cluster (Fig. 3). VACC1 Paris isolates were harvested mostly in cooling towers (70% of the VACC1 isolates; see Fig. S1 in the supplemental material), in agreement with a recent investigation in Japan suggesting that the Paris strain is particularly well adapted to cooling towers (2). This clone, assigned to ST1 and variants, has been shown to be very persistent in the environment (24) and to be responsible for sporadic cases as well as for outbreaks reported since 1981 in several countries all over the world (5). Although this CC is colonizing the Rennes' hospital cooling tower, it has not generated reported legionellosis outbreaks. All 5 clinical cases are found in VACC2 (Philadelphia-1 lineage), two with genotype 12 (ST439) and three with genotype 13 (ST626 and ST628). These STs were reported only for the Rennes outbreaks, but they are highly related to the type strain Philadelphia-1. Isolates of the VACC8 Lorraine cluster were harvested twice in a single place, the hot water system of a nursing home. The Lorraine strain itself (ST47) is a highly virulent emerging strain frequently found in Europe (19). It was responsible for 10% of diagnosed cases of Legionnaires' disease in 2009, but it is found rarely in the environment. The two Rennes VACC8 isolates were non-sg1, whereas the Lorraine strains involved in clinical cases were always sg1.

More surprisingly, we observed the large repartition of a unique MLVA clone (VACC6) recovered exclusively from hot water systems. Its distribution was not restricted to a specific neighborhood of the city, as 93% of hot water systems are colonized by VACC6 isolates (Fig. 3). This is, to our knowledge, the first report of such colonization by a single clone in anthropic habitats. The level of contamination of the network is highly variable, since it ranged from 50 to 200,000 CFU/liter (mean, 800) for VACC6 isolates. Moreover, we show that the contamination of the water network is very persistent and presumably resilient to treatments used for decontamination of the networks (50 mg/liter chloride treatment or heat shock), as described by Cooper and Hanlon (8) and Farhat et al. (15). One MLVA genotype 9 isolate from Rennes was analyzed by SBT and assigned to ST59 (7,6,17,3,13,11,11), a ST that has regularly been observed during epidemiological surveys (22, 31) and was responsible for several clinical cases in Great Britain, France, and Canada. Interestingly, the study by Reimer and collaborators showed that ST59 isolates and one-locus-variant ST670 isolates (7,6,17,3,11,11,11) were recovered in Calgary, Alberta, Canada, from water over a period of 17 years but were not involved in clinical cases in that town in the same period. Moreover, in the present study all the tested VACC6 isolates were sg1, suggesting that the clone is potentially pathogenic. In spite of the abundance of this clone in Rennes' water supplies and the pathogenicity of highly similar isolates,

it did not trigger the reported outbreaks which were linked to the contamination of cooling towers. Thus, in Rennes, isolates of VACC6, which seem particularly well adapted to water supply facilities, might be less prone to express their pathogenic power than in other niches, a potential illustration of the ecotype theory (7). However, VACC6 may have been involved in sporadic clinical cases not investigated in the present study. During the survey, the number of sporadic cases of legionellosis diagnosed in Rennes was very low (a few cases each year). Whole-genome sequencing and comparison between VACC6 isolates from Rennes and from outbreak-associated strains may shed light on this question (11).

The developed MLVA genotyping assay may represent a decision-making tool to help define targeted and priority risk facilities at which it seems important to act quickly. The present investigation suggests that a major treatment plan of Rennes' water supply is not to be advised since it might result in the recolonization of a different strain with a higher resilience and pathogenic potential.

Evolution of the strains involved in clinical cases. MLVA typing confirms the likely sources of the two outbreaks as cooling towers in a single mall in Rennes. The 5 isolates are members of a single CC with two MLVA genotypes and 3 STs. ST628 and ST629 isolates, both of MLVA genotype 13, differ at the *mip*, *proA*, and *neuA* markers at 5, 16, and 2 nucleotides, respectively. The ST 439 isolate with MLVA genotype 12 presents the same *mip* and *proA* alleles as ST626. Although these SBT loci are supposed to be prone to selection pressure, such a high divergence between alleles at a unique geographical location and during a short period of time could be explained only by horizontal transfer, given the sequence identity observed at the other loci. Thus, it is possible that ST626 diverged from ST628 through acquisition of DNA material from an ST439 isolate.

Conclusion. We discovered the massive colonization of Rennes water supplies essentially by strains of a single *L. pneumophila* clonal complex, VACC6. The presence of the major strain (genotype 9) and its variants, identified over a period of 10 years, was not associated with an outbreak of clinical cases. These observations were the result of high-resolution genotyping of more than 200 strains with a fast, efficient, automated, multiplex capillary-based MLVA method. The genotype in the form of a code which facilitates datum transfer and interlaboratory comparison was produced automatically. Such a method will allow the follow-up of a very large population of isolates, opening the way to relevant epidemiological studies as well as investigation of the *L. pneumophila* population structure.

Acknowledgments

We thank Arlette Rouxel for technical support in the isolation of strains and the Agence Régionale de Santé and the Association pour le Développement de l'Hygiène et de l'épidémiologie en Bretagne of Rennes for their financial support. We thank Norman Fry for providing DNA samples of the EWGLI collection.

D.S., F.L.-H., and B.L. are employees of Ceeram and hold stocks. Patent licensing arrangements exist with D.S., F.L.-H., B.L., and C.P.

Work by G.V. is part of the European Biodefense Laboratory Network (EBLN) supported by the European Defense Agency.

References

1. Albert-Weissenberger, C., C. Cazalet, and C. Buchrieser. 2007. *Legionella pneumophila*—a human pathogen that co-evolved with fresh water protozoa. Cell. Mol. Life Sci. 64:432–448.

2. Amemura-Maekawa, J., F. Kura, B. Chang, and H. Watanabe. 2005. *Legionella pneumophila* serogroup 1 isolates from cooling towers in Japan form a distinct genetic cluster. Microbiol. Immunol. 49:1027–1033.

3. Ballard, L. W., et al. 2002. Strategies for genotyping: effectiveness of tailing primers to increase accuracy in short tandem repeat determinations. J. Biomol. Tech. 13:20–29.

4. Campese, C., S. Jarraud, and D. Che. 2007. Legionnaire's disease: surveillance in France in 2005. Med. Mal. Infect. 37:716–721. (In French.)

5. Cazalet, C., et al. 2008. Multigenome analysis identifies a worldwide distributed epidemic *Legionella pneumophila* clone that emerged within a highly diverse species. Genome Res. 18:431–441.

6. Cirillo, J. D., et al. 1999. Intracellular growth in *Acanthamoeba castellanii* affects monocyte entry mechanisms and enhances virulence of *Legionella pneumophila*. Infect. Immun. 67:4427–4434.

7. Cohan, F. M. 2001. Bacterial species and speciation. Syst. Biol. 50:513–524.

8. Cooper, I. R., and G. W. Hanlon. 2010. Resistance of *Legionella pneumophila* serotype 1 biofilms to chlorine-based disinfection. J. Hosp. Infect. 74:152–159.

9. Corrigan, R. M., D. Rigby, P. Handley, and T. J. Foster. 2007. The role of *Staphylococcus aureus* surface protein SasG in adherence and biofilm formation. Microbiology 153:2435–2446.

10. Coscolla, M., I. Comas, and F. Gonzalez-Candelas. 2011. Quantifying nonvertical inheritance in the evolution of *Legionella pneumophila*. Mol. Biol. Evol. 28:985–1001.

11. D'Auria, G., N. Jimenez-Hernandez, F. Peris-Bondia, A. Moya, and A. Latorre. 2010. *Legionella pneumophila* pangenome reveals strain-specific virulence factors. BMC Genomics 11:181.

12. Doleans, A., et al. 2004. Clinical and environmental distributions of *Legionella* strains in France are different. J. Clin. Microbiol. 42:458–460.

13. Dreyfus, L. A., and B. H. Iglewski. 1985. Conjugation-mediated genetic exchange in *Legionella pneumophila*. J. Bacteriol. 161:80–84.

14. Edwards, M. T., N. K. Fry, and T. G. Harrison. 2008. Clonal population structure of *Legionella pneumophila* inferred from allelic profiling. Microbiology 154:852–864.

15. Farhat, M., et al. 2010. Development of a pilot-scale 1 for *Legionella* elimination in biofilm in hot water network: heat shock treatment evaluation. J. Appl. Microbiol. 108:1073–1082.

16. Fields, B. S., R. F. Benson, and R. E. Besser. 2002. *Legionella* and Legionnaires' disease: 25 years of investigation. Clin. Microbiol. Rev. 15:506–526.

17. Fry, N. K., et al. 1999. A multicenter evaluation of genotypic methods for the epidemiologic typing of *Legionella pneumophila* serogroup 1: results of a pan-European study. Clin. Microbiol. Infect. 5:462–477.

18. Gaia, V., et al. 2005. Consensus sequence-based scheme for epidemiological typing of clinical and environmental isolates of *Legionella pneumophila*. J. Clin. Microbiol. 43:2047–2052.

19. Ginevra, C., et al. 2008. Lorraine strain of *Legionella pneumophila* serogroup 1, France. Emerg. Infect. Dis. 14:673–675.

20. Grissa, I., P. Bouchon, C. Pourcel, and G. Vergnaud. 2008. On-line resources for bacterial micro-evolution studies using MLVA or CRISPR typing. Biochimie 90:660–668.

21. Grundmann, H., S. Hori, and G. Tanner. 2001. Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. J. Clin. Microbiol. 39:4190–4192.

22. Harrison, T. G., B. Afshar, N. Doshi, N. K. Fry, and J. V. Lee. 2009. Distribution of *Legionella pneumophila* serogroups, monoclonal antibody subgroups and DNA sequence types in recent clinical and environmental isolates from England and Wales (2000-2008). Eur. J. Clin. Microbiol. Infect.
Dis. 28:781–791.

23. Hunter, P. R., and M. A. Gaston. 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol. 26:2465–2466.

24. Lawrence, C., et al. 1999. Single clonal origin of a high proportion of *Legionella pneumophila* serogroup 1 isolates from patients and the environment in the area of Paris, France, over a 10-year period. J. Clin. Microbiol. 37:2652–2655.

25. Murga, R., et al. 2001. Role of biofilms in the survival of *Legionella pneumophila* in a model potable-water system. Microbiology 147:3121–3126.

26. Nederbragt, A. J., et al. 2008. Multiple-locus variable-number tandem repeat analysis of *Legionella pneumophila* using multi-colored capillary electrophoresis. J. Microbiol. Methods 73:111–117.

27. Olsen, J. S., et al. 2010. Alternative routes for dissemination of *Legionella pneumophila* causing three outbreaks in Norway. Environ. Sci. Technol. 44:8712–8717.

28. Pourcel, C., Y. Vidgop, F. Ramisse, G. Vergnaud, and C. Tram. 2003. Characterization of a tandem repeat polymorphism in *Legionella pneumophila* and its use for genotyping. J. Clin. Microbiol. 41:1819–1826.

29. Pourcel, C., et al. 2007. Identification of variable-number tandem-repeat (VNTR) sequences in *Legionella pneumophila* and development of an optimized multiple-locus VNTR analysis typing scheme. J. Clin. Microbiol. 45:1190–1199.

30. Ratzow, S., V. Gaia, J. H. Helbig, N. K. Fry, and P. C. Lück. 2007. Addition of *neuA*, the gene encoding *N*-acylneuraminate cytidylyl transferase, increases the discriminatory ability of the consensus sequence-based scheme for typing *Legionella pneumophila* serogroup 1 strains. J. Clin. Microbiol.45:1965–1968.

31. Reimer, A. R., S. Au, S. Schindle, and K. A. Bernard. 2010. *Legionella pneumophila* monoclonal antibody subgroups and DNA sequence types isolated in Canada between 1981 and 2009: laboratory component of national surveillance. Eur. J. Clin. Microbiol. Infect. Dis. 29:191–205.

32. Selander, R. K., et al. 1985. Genetic structure of populations of *Legionella pneumophila*. J. Bacteriol. 163:1021–1037.

33. Simpson, E. H. 1949. Measurement of diversity. Nature 163:688.

34. Swanson, M. S., and B. K. Hammer. 2000. *Legionella pneumophila* pathogenesis: a fateful journey from amoebae to macrophages. Annu. Rev. Microbiol. 54:567–613.

35. Tijet, N., et al. 2010. New endemic *Legionella pneumophila* serogroup I clones, Ontario, Canada. Emerg. Infect. Dis. 16:447–454.

36. van Belkum, A., et al. 2007. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin. Microbiol. Infect. 13(Suppl. 3):1–46.

37. Vergnaud, G., and C. Pourcel. 2009. Multiple locus variable number of tandem repeats analysis. Methods Mol. Biol. 551:141–158.

38. Visca, P., et al. 26 May 2011. Investigation of the *Legionella pneumophila* population structure by analysis of tandem repeat copy number and internal sequence variation. Microbiology [Epub ahead of print.] doi: 10.1099/mic.0.047258-0.

2.2. *P. aeruginosa* et suivi longitudinal

a. Résumé

P. aeruginosa, également connue sous le nom de bacille pyocyanique, est une bactérie pathogène opportuniste, particulièrement virulente lors de la contamination de sujets immunodéprimés. Ce germe ubiquitaire vit à l'état saprophyte dans l'eau et les sols humides, les réservoirs d'eau constituant alors la principale source de contamination à *P. aeruginosa*. Les pathologies induites par une infection à *P. aeruginosa* sont très diverses, allant de la conjonctivite à la septicémie. En particulier, *P. aeruginosa* est un agent pathogène majeur responsable des surinfections pulmonaires dans la mucoviscidose. La dégradation de la fonction pulmonaire est accélérée lorsque la bactérie passe sous la forme mucoïde. La population colonisatrice se rassemble alors au sein d'un biofilm, véritable rempart contre les défenses de l'hôte et l'action des antibiotiques. L'identification précoce de la souche colonisatrice permettra d'établir le foyer initial de l'infection disséminée et de tracer la propagation de la bactérie. Aussi, l'analyse systématique des isolats récoltés est un préalable pour un suivi épidémiologique complet, lui seul permettant de différencier récidive et réinfection ou de mettre en évidence une contamination inter-patients. Le génotypage est la réponse adaptée à ces desiderata. Nous avons envisagé le développement d'un procédé automatisé de génotypage par MLVA afin de permettre le typage systématique, haut-débit et en temps réel.

Le protocole développé, le MLVA-16_{Orsay}, assure l'analyse de 16 loci VNTRs coamplifiés en deux réactions de PCR multiplexe. Le développement a été mené dans le cadre d'une étude comparative avec une autre technique de typage, la PFGE, sur une collection d'isolats cliniques issus de patients atteints de mucoviscidose soignés au sein d'un hôpital parisien. Sur le plan épidémiologique les deux méthodes sont parfaitement concordantes, néanmoins le MLVA bénéficie d'une discrimination légèrement supérieure. De plus, le résultat numérique obtenu par MLVA et la mise à disposition d'une base de données riche offrent la possibilité d'étudier la distribution des isolats typés au niveau mondial. Les principaux clones épidémiques, comme PA14 ou Lesb, ne sont pas retrouvés. Néanmoins, nos observations suggèrent que les isolats typés appartiennent à un petit nombre de clones adaptés à la colonisation pulmonaire et circulant en France.

Par ailleurs, l'outil a permis de suivre longitudinalement les populations de *P. aeruginosa* colonisant les poumons de patients atteints de mucoviscidose. Les résultats montrent qu'un patient pouvait être infecté par le même génotype pendant plusieurs années et que ce génotype lui était spécifique. La colonisation est donc persistante démontrant, par là même, la stabilité des VNTRs utilisés. Les mutants observés résultent de l'instabilité d'un seul VNTR, ms61 dont le taux de mutation *in vivo* a été évalué à 10^{-4}-10^{-5} mutations par génération. Enfin, la comparaison génotypique des populations colonisatrices des différents patients indique l'absence de contamination croisée entre les patients.

b. ***Article 2 (Publié par European Journal of Clinical Microbiology & Infectious Diseases)***

A new highly discriminatory multiplex capillary-based MLVA assay as a tool for epidemiological survey of *Pseudomonas aeruginosa* in cystic fibrosis patients.

Sobral D.[1,2,3], Mariani-Kurkdjian , P[4]., Bingen, E.[4], Vu-Thien H.[5]., Hormigos, K.[1,2], Lebeau B.[3], Loisy-Hamon F.[3], Munck A.[6], Vergnaud G.[1,2,7], Pourcel C.[1,2]

Univ Paris-Sud, Institut de Génétique et Microbiologie, UMR 8621, 91405, Orsay, France[1]

CNRS, 91405, Orsay, France[2]

Ceeram (Centre Européen d'Expertise et de Recherche sur les Agents Microbiens), Allée de la Filée, 44244, La Chapelle sur Erdre, France[3]

Laboratoire de Microbiologie, Hôpital Robert Debré, Université Paris 7, Assistance Publique-Hôpitaux de Paris (APHP), 75019, Paris, France[4]

Laboratoire de Microbiologie, Hôpital Armand Trousseau, Université Paris 6, Assistance Publique-Hôpitaux de Paris (APHP), 75012 Paris, France[5],

Département de Gastroentérologie pédiatrique, Nutrition et Mucoviscidose, Université Paris 7, Hôpital Robert Debré, Assistance Publique-Hôpitaux de Paris (APHP), 75019, Paris, France[6]

DGA/MRIS- Mission pour la Recherche et l'Innovation Scientifique, 92221 Bagneux, France[7].

ABSTRACT

Multiple Locus Variable Number of Tandem Repeats (VNTR) analysis (MLVA) has been shown to provide a high level of information for epidemiological investigations and the follow-up of *Pseudomonas aeruginosa* chronic infection. In the present study an automatized MLVA assay has been developed for the analysis of 16 VNTRs in two multiplex PCRs followed by capillary electrophoresis. The result in the form of a code is directly usable for clustering analyses. This MLVA-16_{Orsay} scheme was applied to the genotyping of 83 isolates from 8 cystic fibrosis patients, demonstrating that the same genotype persisted during 8 years of chronic infection in the majority of cases. Comparison with pulse-field gel electrophoresis analysis showed that both methods were congruent, MLVA providing in some cases additional informativity. Evolution of strains during long-term infection was revealed by the presence of VNTR variants.

INTRODUCTION

Pseudomonas aeruginosa is an ubiquitous environmental species widely spread in soil and natural water, and an opportunistic human pathogen responsible for severe infections in immune-compromised patients. The bacteria infect the pulmonary tract, urinary tract, burns, wounds, and also

cause blood infections. Due to its large occurrence in hospital water systems and its capacity to persist on medical devices, *P. aeruginosa* is a leading cause of hospital-acquired pneumonia. In addition, *P. aeruginosa* is the major pathogen in cystic fibrosis (CF) lung pathology [1,2]. CF is caused by a mutation in the cftr gene affecting the salt and water balance in the lung cells. The induced physiological disorder leads to the production of a thick mucus in lungs, a conducive environment for *P. aeruginosa* colonization. The persistence of chronic *P. aeruginosa* lung infections in CF patients is due to the production of a polymer matrix by mucoid strains growing inside biofilms leading to increased tolerance to antibiotics and phagocytosis [3,4].

The population structure of *P. aeruginosa* has been investigated revealing the existence of some clonal complexes inside a panmictic population [5,6]. Worldwide successful lineages such as clones C, PA14 or Les [7-9], were reported in CF patients even if cross-infection by *P. aeruginosa* is believed to be uncommon [10]. The most probable source of colonization is environmental [11,7]. Longitudinal survey of young CF patients from primary to chronic infection revealed that both the bacterial clone and the disease stage influence the outcome of the infection [8].

To track the primary source of infection and the spread of *P. aeruginosa* between patients, several typing methods have been developed. Molecular typing methods such as pulsed-field gel electrophoresis (PFGE), restriction fragment length polymorphism (RFLP) or amplified fragment length polymorphism (AFLP) are currently used but while these techniques are robust and discriminatory, they rely on the analysis of multiband patterns which reduces their reproducibility and portability. Yet, PFGE is still considered the gold standard technique as it demonstrates the highest discriminatory power [12]. Multi locus sequence typing (MLST) is an emerging method for *P. aeruginosa* genotyping and currently the database at http://pubmlst.org/paeruginosa/ holds 1095 sequence types (ST) [13]. However it has been demonstrated that inside a single ST, PFGE shows a wide diversity which suggests that MLST informativity might not be sufficient for epidemiological studies [14]. Multi Locus Variable Number Tandem Repeat (VNTR) Analysis (MLVA) is a PCR based typing method that relies on the inherent variability found in some regions of repetitive DNA [15]. Previous studies reported the usefulness of a discriminatory MLVA protocol for *P. aeruginosa* genotyping [16-19,8]. However, in spite of its many advantages, MLVA requires a relatively high hands-on time when monoplex PCR products are run on agarose gels.

In the present study, we extended and automated our previously described MLVA scheme by amplifying 16 loci in two multiplex PCRs followed by capillary electrophoresis. The new high-throughput MLVA protocol was applied to the longitudinal survey of *P. aeruginosa* colonization in CF patients for eight years. The presence of a highly informative VNTR locus in the assay gave some insight into the strain evolution during long-term colonization.

MATERIALS AND METHODS

Patients. Eight CF patients attended the CRCM (Centre de ressources et de compétences de la mucoviscidose) in the “Robert Debré” hospital in Paris. Strains were collected from sputum specimens as part of the patients' usual care, without any additional sampling. All the patients’ data were anonymously reported, without offering any possibility to trace back the actual patients.

Microbiology. *P. aeruginosa* isolates were identified by colony morphology, characteristic pigment production, and by using the biochemical profile index procedure API NE (bioMérieux, Grenoble, France). Different isolates from a single patient sample were investigated when showing different phenotypes (mucoid/non-mucoid), or serotypes, or when more than one difference on the antimicrobial susceptibility pattern (22 antibiotics) was observed (Table S1). The studied patients were selected based on the long history of *P. aeruginosa* sampling in their airways. Eighty three isolates from the eight studied patients were genotyped manually by MLVA-15_{Orsay} [8] and using the new automated protocol. From 6 to 14 different isolates were analysed for each patient. In addition eleven isolates from a single patient, recovered over a period of three years in the “Armand Trousseau” hospital, and previously manually analysed were typed with the new automated protocol [8]. Reference strains PAO1 and PA14 were purchased from the Institut Pasteur culture collection (CRBIP). Strains C50 and SG17, belonging to clone C were generously provided by Utte Römling.

PFGE. Chromosomal DNA was prepared as previously described and digested overnight at 37 °C with 40 units of XbaI in a 250 µL reaction volume [20]. Electrophoresis was performed in a 1% agarose gel for 19 h at 14°C, using a CHEF DRIII apparatus (Bio-Rad, France) with the following parameters: 220 volts, 120° angle, 5 to 20 sec pulse time [21]. Chromosomal DNA restriction fragments were stained with ethidium bromide, and visualised by UV transillumination. Isolates showing no more than three restriction fragment differences were considered to be subtypes of a common strain [20,22].

DNA purification. DNA was purified with a DNeasy tissue kit (Qiagen, Courtaboeuf, France). The quality and the concentration of DNA were estimated by measuring UV absorbance with a ND-1000 spectrophotometer (NanoDrop, Labtech, Palaiseau, France). Diluted samples of 10 ng/µL in water were used as DNA template for PCR amplification.

Selection of VNTRs and MLVA typing. The MLVA-15_{Orsay} assay was performed as described by Vu Thien et al. [8]. The new MLVA scheme, MLVA-16_{Orsay}, comprises 16 previously described VNTRs [16,8] present in the three currently published MLVA protocols: MLVA-15_{Orsay} [8], MLVA-9_{London} [17] and MLVA-$9_{Utrecht}$ [19]. The constitution of the different MLVA schemes is described in Table 1. New primers were designed so that the allele ranges of VNTRs labeled with the same dye do not overlap. The 16 VNTR loci were amplified in two multiplex PCRs using the genotyping kit TYPPSEUDO ceeramTools® (Ceeram, La Chapelle Sur Erdre, France). Briefly, PCR reaction 1 amplifies ten VNTR loci (ms61, ms77, ms127, ms172, ms212, ms214, ms216, ms217, ms222 and ms223) and PCR reaction 2 amplifies six VNTR loci (ms142, ms207, ms209, ms211,

ms213 and ms215). The primers are listed in Table 2. The PCR reaction was run on a Veriti® Thermal Cycler (Applied Biosystems, Courtaboeuf, France) using the following conditions: initial denaturation cycle for 15 min at 95°C, 15 cycles touchdown PCR [30 s at 95 °C; 60 s at 74 °C, with 0.8 °C drop in temperature each next cycle; 70 s at 72 °C]; 15 cycles long range PCR [30 s at 95 °C; 60 s at 62 °C; 70 s at 72 °C with 5 s increase in time each next cycle]; with a final 10 min at 72 °C. PCR fragments purification, capillary electrophoresis and analysis of electrophoretic profiles were performed as previously described [23].

VNTR	Repeat unit size (in bp)	Multiplex PCR	MVLVA-16_{Orsay}	MVLVA-15_{Orsay}	MLVA-9_{London}	MLVA-$9_{Utrecht}$
ms77	39	PCR1	X	X		X
ms127	15	PCR1	X	X		X
ms142	115	PCR2	X	X		X
ms172	54	PCR1	X	X	X	
ms211	101	PCR2	X	X	X	X
ms212	40	PCR1	X	X		
ms213	103	PCR2	X	X	X	X
ms214	115	PCR1	X	X	X	
ms215	129	PCR2	X	X		X
ms216	113	PCR1	X	X		X
ms217	109	PCR1	X	X	X	X
ms222	101	PCR1	X	X	X	
ms223	106	PCR1	X	X		X
ms61	6	PCR1	X		X	
ms207	6	PCR2	X	X	X	
ms209	6	PCR2	X	X	X	

Table 1. Constitution of the different MLVA schemes

VNTR name	Forward primer sequence (5' → 3')	Reverse primer sequence (5' → 3')	Repeat size (bp)	Expected size (bp) (no of repeats) for strain PAO1	Allele size range (bp) (no of repeat units)
ms77	NED-GGAACAGCAGGTGGCAGT	ACTGGACCGGCCTGTTCC	39	1074 (4)	957-1152 (1U-6U)
ms127	6FAM-CCAGATCCAGCTTGGTCG	CAGGAGGATGCGCTGGAC	15	1120 (8)	1105-1142 (7U-9.5U)
ms142	PET-GTGCCTGCGGAGCTGTTG	CGTTGGAGACGGAGGAGG	115	879 (7)	132-1109 (0.5U-9U)
ms172	PET-CTGCTCAACCTGCAGGTG	GTACGTGACCTGACGTTGG	54	1142 (12)	764-1196 (5U-13U)
ms211	6FAM-GACAAGCGCCAGCCGAAC	GCTGGAACTTCGAACAGG	101	671 (5)	368-1075 (2U-9U)
ms212	6FAM-CGTCGCTGCTCTGATCTG	CAACTGCGCTGAAGTACC	40	340 (9)	100-580 (3U-15U)
ms213	NED-GCTCGGCTACTACATCCTC	GCAAGTGTTGGTGGATCAAC	103	671 (5)	208-1083 (0.5U-9U)
ms214	NED-CTACCTGCTGGCGTTCTG	CCTCCATCATCCTCCTACTGG	115	356 (3)	241-701 (2U-6U)
ms215	VIC-CTAAGGGCCACGGACTTC	CCTCGGCAGCAGGACAGC	129	870 (4)	418-1128 (0.5U-6U)
ms216	VIC-CCACGGTCGTCGACATGCG	GCTACCGGCGGGACAAGC	113	1000 (3)	774-1226 (1U-5U)
ms217	NED-TGCCGTTTGCGTTGTAGG	CTCGAAATGTGGGTGAGC	109	321 (2)	212-757 (1U-6U)
ms222	6FAM-CACTGCTGCGGCTTCGCC	GATGGTGGCGTTGGCCTG	101	696 (2)	546-1100 (0.5U-6U)
ms223	PET-TGAGCTGATCGCCTACTGG	TGGCAATATGCCGGTTCG	106	453 (4)	188-877 (1.5U-8U)
ms61	NED-GATCCGGACGGCGACACT	ACGCCTCTCGCCGACCAG	6	173 (12)	125-203 (4U-17U)
ms207	6FAM-GTTGCCGAAACGGGTGAT	CCCGTCTTCGTCCCGTCC	6	279 (7)	255-333 (3U-16U)
ms209	6FAM-GCTGTTCGGCGACAAGGC	GGTAATGGGCGTGGATATCG	6	110 (6)	80-128 (1U-9U)

Table 2. Oligonucleotide primers used and VNTRs analysed in this study

Data analysis. The typing data file was imported into BioNumerics version 6.6 (Applied-Maths, Sint-Martens-Latem, Belgium). Allele designation for the VNTRs described in previous work [8] was used as published. A complete genotype expressed as the number of repeats at each VNTR was constructed in the order ms77, ms127, ms142, ms172, ms211, ms212, ms213, ms214, ms215, ms216, ms217, ms222, ms223, ms61, ms207, ms209. The coding convention used assigns genotype 4-8-7-12-5-9-5-3-4-3-2-2-4-12-7-6 to the PAO1 genomic sequence NC_002516 [24]. The UPGMA (Unweighted Pair Group Method with Arithmetic mean) clustering method was run within BioNumerics using the categorical coefficient (also called Hamming's distance). The minimum spanning tree was produced in BioNumerics and the logarithm scale was applied when drawing branches. An MLVA type (MT) is a genotype produced by MLVA typing. An MLVA clone (MC) is defined by the grouping of MTs that differ at one or two VNTRs.

Calculation of the predicted in vivo mutation rate. The in vivo mutation rate calculation was based on the time between the first sampling of a MLVA genotype and that of its variant. Growth rates of *P. aeruginosa* in CF lungs were estimated by Yang and colleagues using bacterial ribosome contents in cells isolated from fresh sputum samples [25]. The experimentally observed in vivo doubling time was between 100 and 200 min. Thus, considering a mean generation time of 150 min, the sampling time was converted into a number of generations (number of generations = sampling

time (in min) / 150). Finally, the in vivo mutation rate was obtained by dividing the number of observed genetic events and the calculated number of generations needed to introduce this event. As only single locus mutations were observed, the in vivo mutation rate equals 1 / number of generations between a MLVA type and its variant.

RESULTS

Automated multiplex capillary-based MLVA assay development. VNTR ms61 was added to the previously described MLVA-15$_{Orsay}$ scheme to improve the discriminatory power of the assay. With 6bp repeats, ms61, together with ms207 and ms209, belongs to the category of microsatellites and shows more instability and a larger number of alleles than VNTRs with longer repeats. Using a high precision capillary electrophoresis equipment is mandatory for accurate measurement of microsatellites alleles. To be able to multiplex the PCRs in the MLVA-16$_{Orsay}$ assay, new primer pairs were designed that allow concomitant analysis of two markers labeled with the same dye. The new optimized primers and the high stringency PCR protocol yielded clean electrophoretic profiles without artifactual peaks such as stutter, spike or shoulder peaks. The use of reverse tailed primers which promotes the addition of a non-templated "A" to the product presumably help to avoid the so-called +A peak artifact [26]. Figure 1 shows a typical capillary electrophoresis pattern of the two multiplex PCRs (Figure 1a, PCR1; Figure 1b, PCR2). Half-sized alleles were observed for ms77, ms127 and ms213.

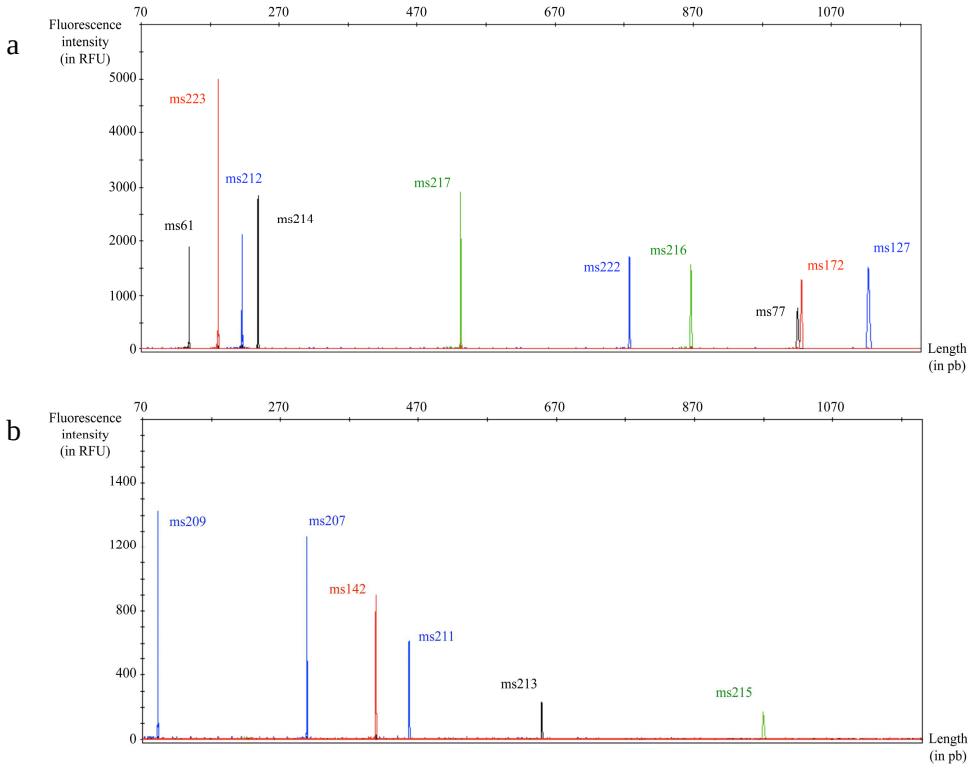


Figure 1. Electrophoregrams showing multiplex PCR amplicons resolved by capillary electrophoresis a. PCR1 ten dye-colored coamplified VNTR loci. b. PCR2 six dye-colored coamplified VNTR loci. Each peak corresponding to a VNTR amplicon is annotated

In order to compare the assays, 98 isolates were typed using the MLVA-15_{Orsay} agarose gel electrophoresis-based method and using the newly developed automatized protocol. This includes 15 previously described isolates and 83 new isolates from patients in the "R. Debré" hospital (Table S1). Very few discrepancies between the assays were observed (Table S2). The agarose-based assay was sometimes less accurate which explains the extra half-sized units observed in some alleles compared to the capillary-based assay. The new set of primers correctly amplified a few previously missing alleles. The genotype of reference strains PA14 and PAO1 were as expected from their genome sequence. The calculated typeability of MLVA-16_{Orsay} was 98%, as missing values occurred for ms142, ms214 and ms223. The absence of amplification could be due to mismatches in the target region of primers or to complete or partial deletion of the locus, but this was not confirmed. Therefore absent data were not considered as a character in the clustering analysis. The genotype number for MLVA-15_{Orsay} and MLVA-16_{Orsay} were respectively 20 and 23 showing that ms61 contribute to 13% of the observed diversity. In a collection of 11 isolates from a patient in the "A. Trousseau" hospital, previously studied by MLVA-15_{Orsay}, one variant at ms61 could be observed (Table S2). MLVA-16_{Orsay} includes all the VNTRs analyzed in MLVA-15_{Orsay}, MLVA-9_{London} and MLVA-$9_{Utrecht}$. The MLVA-9_{London} scheme yielded 22 MLVA genotypes due to the ms61 discriminatory power. The MLVA-$9_{Utrecht}$ method, which excludes the microsatellites ms61, ms207 and ms209, distinguished 19 MLVA genotypes. These schemes allow the distinction of the same MLVA clones and MLVA types (as shown on the dendrogram in Figure S1 for MLVA-16_{Orsay}), but the inferred phylogenic trees differ when using the different protocols as revealed from analysis of congruence (Figure 2).

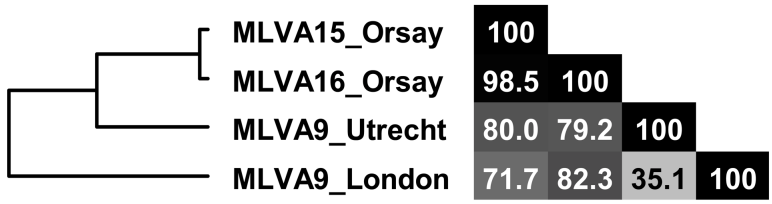


Figure 2. Congruence of MLVA schemes (MLVA-$16_{Orsay,}$ MLVA-$15_{Orsay,}$ MLVA-9_{London} and MLVA-$9_{Utrecht}$)

Longitudinal survey of *P. aeruginosa* colonization of eight CF patients. The 83 isolates from eight CF patients followed during eight years at the "R. Debré" hospital were resolved into 17 MLVA genotypes (MTs) among which 8 were clustered in 4 MCs, as shown on the minimum spanning tree in Figure 3. We performed the comparison of the identified MLVA types with a database that includes nearly 1400 isolates [16,8] and found that none of the clones could be associated to well known lineages. However MT3, MT11 and MT16 isolates were sampled in other French CF centers (Merens et al. manuscript in preparation). MT11 is an endemic CF clone found in Paris, Caen, Lille and Nantes CF centers. The other MTs were not distributed so widely: MT3 and MT16 were present in Paris and Montpellier, MC1 was described in Paris and Besançon. MT13 was exclusively recovered from Parisian CF patients followed in three different centers.

Among the eight CF patients, four were colonized by two or more strains (Figure 3 and Figure S1): RD01 (MT16, followed by MT11), RD02 (MT12 followed by MC2), RD05 (MT3, MT8 and MT17) and RD06 (MT13 followed by MC1). Patient RD01 initially acquired MT16 in 1999, and MT11 only seven years later. Patient RD02 acquired MT12 in 1999, then MC2 in 2002. MT13 was the first colonizing strain of patient RD06 in 2002 but starting in 2006 MC1 isolates were recovered. MT8 was the first strain to persist in patient RD05 in 1999, although MT17 and MT3 isolates were found once in 2003 and 2007. The remaining patients were colonized by a unique MT (Figure 3), and no common MT was shared by the studied patients.

As previously observed, in all patients at least one strain was repeatedly isolated during the study period (Figure 3 and Figure S1). In CF patients RD02, RD07 and RD08 the colonizing strain evolved during the years as shown by the observation of one-locus variants. MT1 was persistent from 2002 to 2006 in patient RD02 and the ms61 variant MT2 was isolated once in 2007. Similarly in patient RD08 an MT15 strain was found 13 times between 1999 and 2006, and the ms61 variant MT14 was isolated twice in 2006. MT6 and MT7 isolates, differing at ms61 were both identified in 2006. Mucoid and non mucoid isolates respectively of genotype MT4 and MT5, differing at ms142 were simultaneously found in patient RD07 from 1999 to 2003.

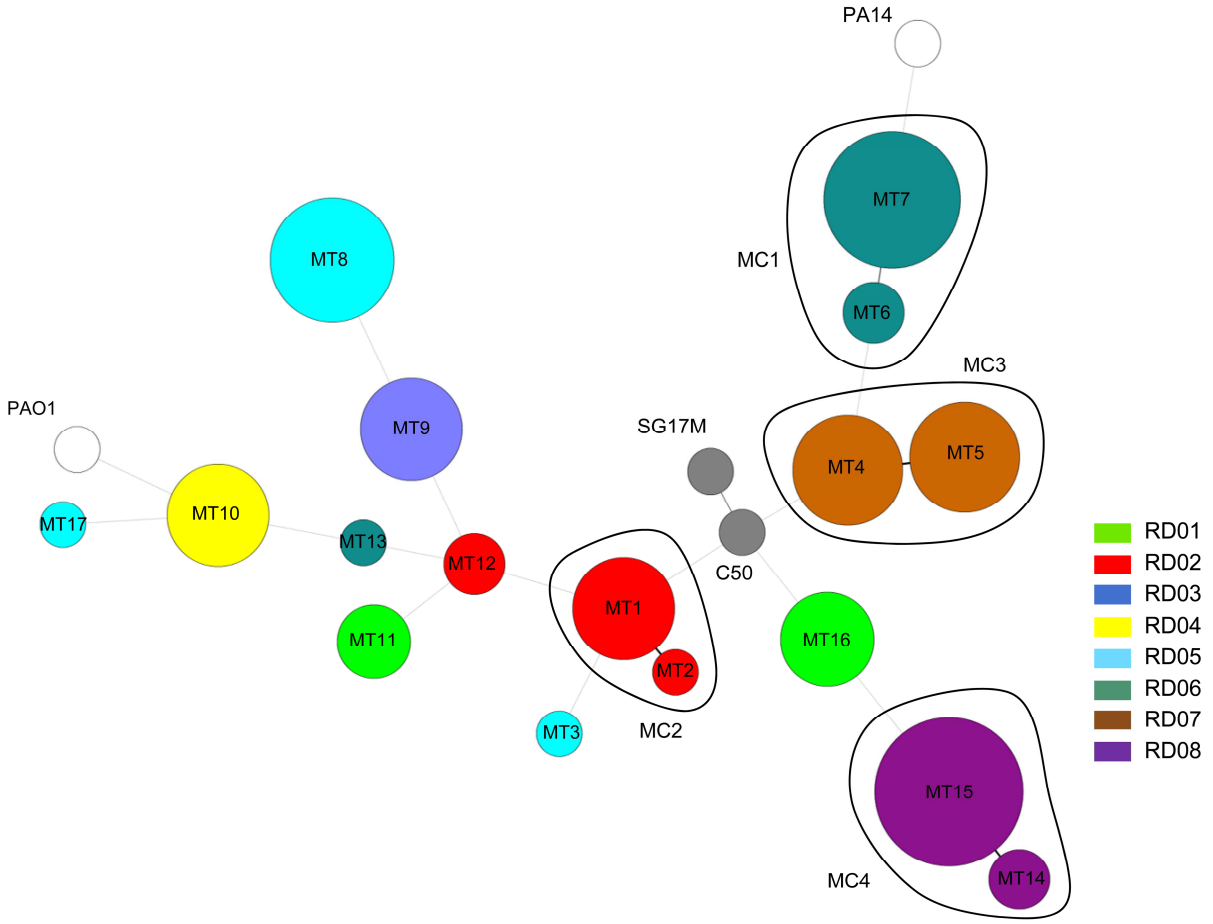
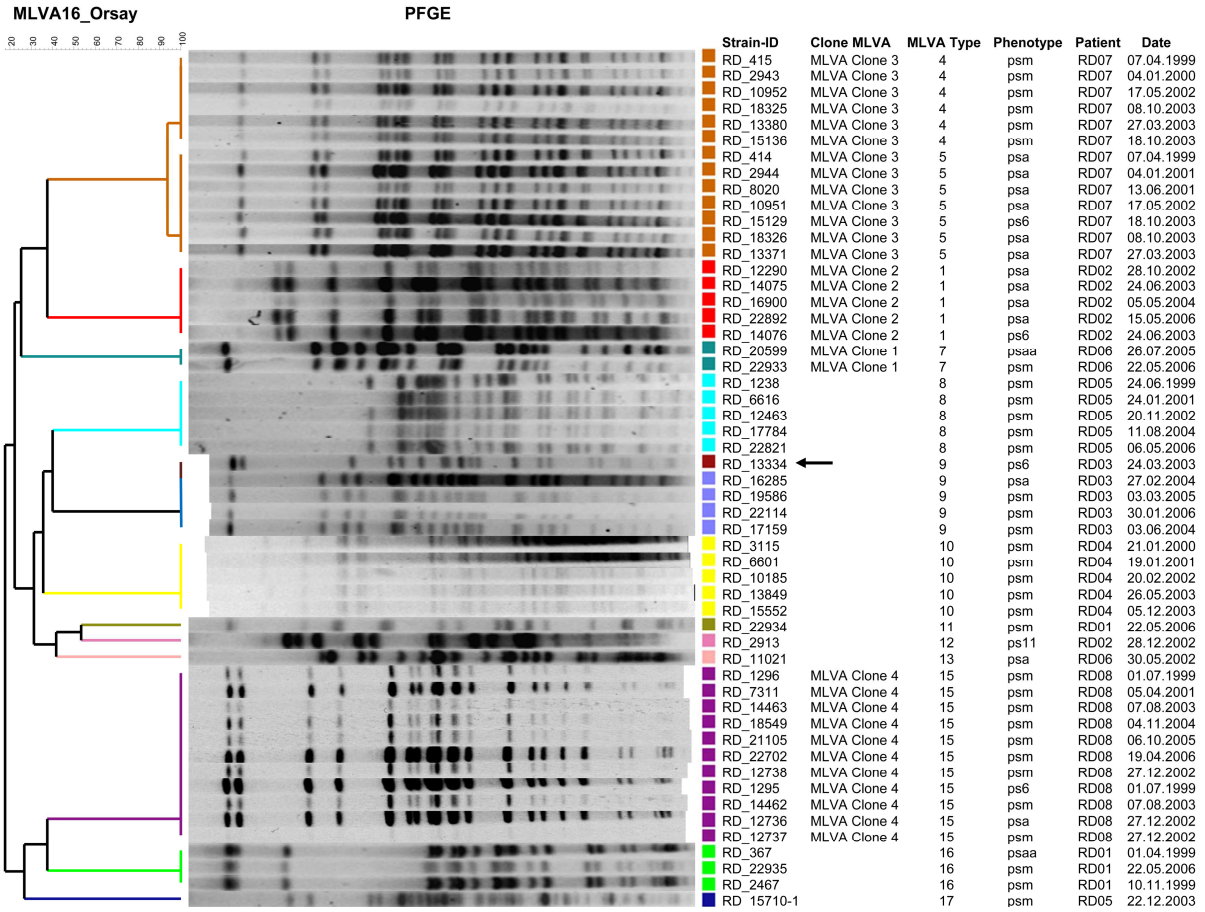


Figure 3. Minimum spanning tree deduced from the clustering analysis of the 83 isolates sampled from eight CF patients for eight years using MLVA-16$_{Orsay}$ (distribution according to patient). Each circle represents an MLVA type (the MT is indicated). The MCs are surrounded.

Comparison of MLVA and PFGE. PFGE is still the reference method for genotyping *P. aeruginosa* in many laboratories and is routinely used in the "Robert Debré" hospital. On Figure 4 a comparison between the two typing methods is shown for 53 isolates for which high quality images were available. Clustering was performed according to MLVA-16$_{Orsay}$ and each PFGE profile is shown on the side for comparison. In this subgroup, MLVA yielded 13 MLVA types where PFGE identified 14 pulsotypes. The two methods were extremely congruent as the MTs corresponded exactly to their associated pulsotypes except for one MT9 isolate, RD_13334 (pointed by an arrow in Figure 4). This isolate was grouped in MT9 together with 4 other isolates from the same patient (RD03) whereas it exhibits a different PFGE pattern. RD_13334 is the first isolate from patient RD03 in 2003. The observation of a different PFGE profile suggests that it experienced some rearrangements during the early period of colonization.

Deep branching dendrogram topologies obtained by MLVA and PFGE were not congruent (data not shown) which is not surprising given that none of these two methods is well-suited for this purpose.



Strain-ID	Clone MLVA	MLVA Type	Phenotype	Patient	Date
RD_415	MLVA Clone 3	4	psm	RD07	07.04.1999
RD_2943	MLVA Clone 3	4	psm	RD07	04.01.2000
RD_10952	MLVA Clone 3	4	psm	RD07	17.05.2002
RD_18325	MLVA Clone 3	4	psm	RD07	08.10.2003
RD_13380	MLVA Clone 3	4	psm	RD07	27.03.2003
RD_15136	MLVA Clone 3	4	psm	RD07	18.10.2003
RD_414	MLVA Clone 3	5	psa	RD07	07.04.1999
RD_2944	MLVA Clone 3	5	psa	RD07	04.01.2001
RD_8020	MLVA Clone 3	5	psa	RD07	13.06.2001
RD_10951	MLVA Clone 3	5	psa	RD07	17.05.2002
RD_15129	MLVA Clone 3	5	ps6	RD07	18.10.2003
RD_18326	MLVA Clone 3	5	psa	RD07	08.10.2003
RD_13371	MLVA Clone 3	5	psa	RD07	27.03.2003
RD_12290	MLVA Clone 2	1	psa	RD02	28.10.2002
RD_14075	MLVA Clone 2	1	psa	RD02	24.06.2003
RD_16900	MLVA Clone 2	1	psa	RD02	05.05.2004
RD_22892	MLVA Clone 2	1	psa	RD02	15.05.2006
RD_14076	MLVA Clone 2	1	ps6	RD02	24.06.2003
RD_20599	MLVA Clone 1	7	psaa	RD06	26.07.2005
RD_22933	MLVA Clone 1	7	psm	RD06	22.05.2006
RD_1238		8	psm	RD05	24.06.1999
RD_6616		8	psm	RD05	24.01.2001
RD_12463		8	psm	RD05	20.11.2002
RD_17784		8	psm	RD05	11.08.2004
RD_22821		8	psm	RD05	06.05.2006
RD_13334		9	ps6	RD03	24.03.2003
RD_16285		9	psa	RD03	27.02.2004
RD_19586		9	psm	RD03	03.03.2005
RD_22114		9	psm	RD03	30.01.2006
RD_17159		9	psm	RD03	03.06.2004
RD_3115		10	psm	RD04	21.01.2000
RD_6601		10	psm	RD04	19.01.2001
RD_10185		10	psm	RD04	20.02.2002
RD_13849		10	psm	RD04	26.05.2003
RD_15552		10	psm	RD04	05.12.2003
RD_22934		11	psm	RD01	22.05.2006
RD_2913		12	ps11	RD02	28.12.2002
RD_11021		13	psa	RD06	30.05.2002
RD_1296	MLVA Clone 4	15	psm	RD08	01.07.1999
RD_7311	MLVA Clone 4	15	psm	RD08	05.04.2001
RD_14463	MLVA Clone 4	15	psm	RD08	07.08.2003
RD_18549	MLVA Clone 4	15	psm	RD08	04.11.2004
RD_21105	MLVA Clone 4	15	psm	RD08	06.10.2005
RD_22702	MLVA Clone 4	15	psm	RD08	19.04.2006
RD_12738	MLVA Clone 4	15	psm	RD08	27.12.2002
RD_1295	MLVA Clone 4	15	ps6	RD08	01.07.1999
RD_14462	MLVA Clone 4	15	psm	RD08	07.08.2003
RD_12736	MLVA Clone 4	15	psa	RD08	27.12.2002
RD_12737	MLVA Clone 4	15	psm	RD08	27.12.2002
RD_367		16	psaa	RD01	01.04.1999
RD_22935		16	psm	RD01	22.05.2006
RD_2467		16	psm	RD01	10.11.1999
RD_15710-1		17	psm	RD05	22.12.2003

Figure 4. Dendrogram deduced from the clustering analysis of 53 isolates using MLVA-16$_{Orsay}$ and comparison with PFGE profiles. Branches were colored according to the pulsotypes defined by PFGE typing.

***In vivo* mutation rate.** Genetic mutations involving ms61 and ms142 were observed leading to the appearance of variants. In the case of ms142 the two variants were contemporary, but for ms61, the less stable VNTR, a time interval could be observed allowing the rough evaluation of a mutation rate. MT2, MT6, and MT14 are most likely ms61-variants of MT1, MT7, and MT15 with interval sampling times of 343, 1555 and 2604 days, respectively. Considering a mean generation time of 150 min, the calculated bacterial generation for the interval sampling times mentioned above are 3293, 14928 and 24999, respectively. The mutation rate based on these numbers would be approximately 3.04×10^{-4}, 6.70×10^{-5} and 4.00×10^{-5} ms61-mutation per generation.

DISCUSSION

Automated capillary-based MLVA system, a powerful genotyping tool. The proposed automated multiplex capillary-based MLVA scheme allows the coamplification of 16 markers in two PCRs multiplex and the automatic MLVA code assignment for the typed isolate. Rare discrepancies were observed between the agarose-based MLVA-15_{Orsay} and MLVA-16_{Orsay} which may be due to the use of new primers pairs and to the electrophoresis device. They have little effect on the typing analysis showing that the two methods can be used to adapt to local settings. The addition of ms61, a very polymorphic microsatellite included in MLVA-9_{London}, to the MLVA-15_{Orsay} scheme increases the discriminatory power of the assay as previously suggested by the studies of Turton and coworkers [17]. Ms61 is present in an ORF predicted to code for a member of the translocation protein in type III secretion family [27]. If it is under selective pressure it might not be suitable for phylogenetic studies but rather to compare potentially epidemiologically related isolates. In the present work which investigates the evolution of strains over time in a limited number of patients, MLVA-16_{Orsay} does not provide an important increase of discrimination over the other schemes. However the possibility to type a large number of VNTRs with different characteristics increases the chance to subtype some clones. An interesting example is provided by the mucoid and non-mucoid isolates of patients RD07 which differ at ms142, whereas ms61 remains unchanged over a period of 5 years.

Using the commercially available typing kit (ceeramTools®, Ceeram, La Chapelle Sur Erdre, France) almost one hundred isolates could be genotyped in less than four days, starting from bacterial colonies and using a single four-capillaries equipment. MLVA is increasingly used for *P. aeruginosa* typing showing that this method is suitable for investigating large and diverse populations of *P. aeruginosa*. Considering its performance and convenience criteria, and the ongoing emergence of draft whole genome sequencing, MLVA could become the first line assay for large-scale *P. aeruginosa* typing and identification of most interesting strains. The described technical improvements are a major breakthrough towards a routine and easy use of MLVA for *P. aeruginosa* genotyping. Resulting data can be queried against existing internet MLVA databases such as the ones maintained at http://mlva.u-psud.fr.

CF patients harbor host-specific genotypes. We investigated the dynamic of *P. aeruginosa* infection in eight CF patients over an eight-year study period. In these patients, host-specific genotypes were observed implying that no patient-to-patient transmission occurred in the "Robert Debré" hospital and that the primary acquisition of *P. aeruginosa* did not arise from a common source. Hygiene guidelines were strictly followed in the hospital, limiting the spread of organisms via handling processes. Hospital environment or hot water systems did not constitute a reservoir for *P. aeruginosa* during the study period.

The persistence of one MLVA type in each patient was observed throughout the study as previously described [11,8], even if few temporal changes in genotype patterns occurred, largely due to ms61 variability. The colonizing genotype appeared very adapted to the CF lung niche resulting in a chronic establishment although multiple extra-contaminations occasionally occurred from different sources. The temporal hierarchical acquisition of the different genotypes was easily observed by focusing on sampling date. In patient RD01 in which MT11 and MT16 persisted together, a different genetic background did not provide any competitive fitness and the two genotypes coexisted suggesting that they might even cooperate. In patient RD02, genotype MT12 was sampled twice in 1999 and 2002 and could be serotyped (the primary-colonization strain), whereas later on it was replaced by MC2 isolates probably showing a specific genetic background for CF lung colonization. It may also be the case for patient RD06 in which the MT13 isolate was sampled once in 2002 and later replaced by MC1 isolates.

The existence of widespread clonal complexes such as clone C, Les or PA14 shown to be highly successful colonizers of CF patients' airways is established. These clusters were not found in the present study but only a small part of the patients followed in the "R. Debré" CRCM were included. In any case one must be careful when attempting to relate an MLVA genotype to an existing clone, particularly when epidemiological observations are missing. MLST might be a necessary complementary method for this purpose and particularly for long term epidemiological studies.

MLVA typing/PFGE. Several schemes for *P. aeruginosa* typing were proposed during the last few years but PFGE remains the reference molecular typing techniques for *P. aeruginosa* in many laboratories as it guarantees a high level of discrimination. In the present work we performed the comparison of the MLVA-16_{Orsay} and PFGE on 53 selected isolates. We highlighted that the two methods were very congruent. MLVA-16_{Orsay} was able to separate PFGE types mostly because of the polymorphism added by VNTR ms61 as previously observed [17].

VNTR as evolutionary markers? VNTR stability has been largely tested in vitro by successive serial passage and the mutation rate was estimated by parallel serial passages experiments (PSPE) [28]. In this study, we analyzed VNTR stability in vivo by studying MLVA pattern changes on different samples from the same patient. We observed that the same MLVA type could be recovered over an eight years period suggesting that VNTR markers are very stable. VNTR ms61 was the most unstable marker. We roughly estimated its in vivo mutation rate between 3.04 x 10-4 and 4.00 x 10-5

mutation per generation. This mutation rate is close to those estimated in vitro for some *Y. pestis* [28], *E. coli* [29] and *B. pseudomallei* [30] microsatellites and is well suited for performing short term epidemiological investigations to track the source and the spread of the pathogen during outbreaks or to assess cross-infections frequencies in a CF center. Low level of congruence between MLVA-16_{Orsay}, MLVA-15_{Orsay}, MLVA-9_{London} and MLVA-$9_{Utrecht}$ indicates that the MLVA method which is affected by homoplasia is not suitable to infer clustering or evolutionary relationships, at least with the currently available data sets and considering the population structure of this species.

Acknowledgements

DS, FLH and BL are employees of Ceeram and hold stocks. This study was performed with the support of the association Vaincre La Mucoviscidose (Grant N° RC0630). The development of tools for the surveillance of bacterial pathogens is supported by the French Direction Générale de l'Armement.

References

1. Saiman L, Siegel J (2004) Infection control in cystic fibrosis. Clin Microbiol Rev 17 (1):57-71

2. Lipuma JJ (2010) The changing microbial epidemiology in cystic fibrosis. Clin Microbiol Rev 23 (2):299-323. doi:23/2/299 [pii] 10.1128/CMR.00068-09

3. Drenkard E, Ausubel FM (2002) Pseudomonas biofilm formation and antibiotic resistance are linked to phenotypic variation. Nature 416 (6882):740-743. doi:10.1038/416740a416740a [pii]

4. Li Z, Kosorok MR, Farrell PM, Laxova A, West SE, Green CG, Collins J, Rock MJ, Splaingard ML (2005) Longitudinal development of mucoid *Pseudomonas aeruginosa* infection and lung disease progression in children with cystic fibrosis. JAMA 293 (5):581-588. doi:293/5/581 [pii] 10.1001/jama.293.5.581

5. Denamur E, Picard B, Decoux G, Denis JB, Elion J (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 (3):275-280

6. Picard B, Denamur E, Barakat A, Elion J, Goullet P (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 (5):313-322

7. Romling U, Fiedler B, Bosshammer J, Grothues D, Greipel J, von der Hardt H, Tummler B (1994) Epidemiology of chronic *Pseudomonas aeruginosa* infections in cystic fibrosis. J Infect Dis 170 (6):1616-1621

8. Vu-Thien H, Corbineau G, Hormigos K, Fauroux B, Corvol H, Clement A, Vergnaud G, Pourcel C (2007) Multiple-locus variable-number tandem-repeat analysis for longitudinal survey of sources of *Pseudomonas aeruginosa* infection in cystic fibrosis patients. J Clin Microbiol 45 (10):3175-3183. doi:JCM.00702-07 [pii] 10.1128/JCM.00702-07

9. Wiehlmann L, Wagner G, Cramer N, Siebert B, Gudowius P, Morales G, Kohler T, van Delden C, Weinel C, Slickers P, Tummler B (2007) Population structure of *Pseudomonas aeruginosa*. Proc Natl Acad Sci U S A 104 (19):8101-8106. doi:0609213104 [pii] 10.1073/pnas.0609213104

10. Speert DP, Campbell ME (1987) Hospital epidemiology of *Pseudomonas aeruginosa* from patients with cystic fibrosis. J Hosp Infect 9 (1):11-21

11. Mahenthiralingam E, Campbell ME, Foster J, Lam JS, Speert DP (1996) Random amplified polymorphic DNA typing of *Pseudomonas aeruginosa* isolates recovered from patients with cystic fibrosis. J Clin Microbiol 34 (5):1129-1135

12. Johnson JK, Arduino SM, Stine OC, Johnson JA, Harris AD (2007) Multilocus sequence typing compared to pulsed-field gel electrophoresis for molecular typing of *Pseudomonas aeruginosa*. J Clin Microbiol 45 (11):3707-3712. doi:JCM.00560-07 [pii] 10.1128/JCM.00560-07

13. Curran B, Jonas D, Grundmann H, Pitt T, Dowson CG (2004) Development of a multilocus sequence typing scheme for the opportunistic pathogen *Pseudomonas aeruginosa*. J Clin Microbiol 42 (12):5644-5649. doi:42/12/5644 [pii] 10.1128/JCM.42.12.5644-5649.2004

14. Garcia-Castillo M, Del Campo R, Morosini MI, Riera E, Cabot G, Willems R, van Mansfeld R, Oliver A, Canton R (2011) Wide dispersion of ST175 clone despite high genetic diversity of carbapenem-nonsusceptible *Pseudomonas aeruginosa* clinical strains in 16 Spanish hospitals. J Clin Microbiol 49 (8):2905-2910. doi:JCM.00753-11 [pii] 10.1128/JCM.00753-11

15. Vergnaud G, Pourcel C (2009) Multiple locus variable number of tandem repeats analysis. Methods Mol Biol 551:141-158

16. Onteniente L, Brisse S, Tassios PT, Vergnaud G (2003) Evaluation of the polymorphisms associated with tandem repeats for *Pseudomonas aeruginosa* strain typing. J Clin Microbiol 41 (11):4991-4997

17. Turton JF, Turton SE, Yearwood L, Yarde S, Kaufmann ME, Pitt TL (2010) Evaluation of a nine-locus variable-number tandem-repeat scheme for typing of *Pseudomonas aeruginosa*. Clin Microbiol Infect 16 (8):1111-1116. doi:CLM3049 [pii] 10.1111/j.1469-0691.2009.03049.x

18. Van der Bij AK, Van Mansfeld R, Peirano G, Goessens WH, Severin JA, Pitout JD, Willems R, Van Westreenen M (2011) First outbreak of VIM-2 metallo-beta-lactamase-producing *Pseudomonas aeruginosa* in The Netherlands: microbiology, epidemiology and clinical outcomes. Int J Antimicrob Agents 37 (6):513-518. doi:S0924-8579(11)00108-7 [pii] 10.1016/j.ijantimicag.2011.02.010

19. van Mansfeld R, Jongerden I, Bootsma M, Buiting A, Bonten M, Willems R (2010) The population genetics of *Pseudomonas aeruginosa* isolates from different patient populations exhibits high-level host specificity. PLoS One 5 (10):e13482. doi:10.1371/journal.pone.0013482

20. Bingen E, Bonacorsi S, Rohrlich P, Duval M, Lhopital S, Brahimi N, Vilmer E, Goering RV (1996) Molecular epidemiology provides evidence of genotypic heterogeneity of multidrug-resistant *Pseudomonas aeruginosa* serotype O:12 outbreak isolates from a pediatric hospital. J Clin Microbiol 34 (12):3226-3229

21. Vu-Thien H, Moissenet D, Valcin M, Dulot C, Tournier G, Garbarg-Chenon A (1996) Molecular epidemiology of *Burkholderia cepacia*, *Stenotrophomonas maltophilia*, and *Alcaligenes xylosoxidans* in a cystic fibrosis center. Eur J Clin Microbiol Infect Dis 15 (11):876-879

22. Grundmann H, Schneider C, Hartung D, Daschner FD, Pitt TL (1995) Discriminatory power of three DNA-based typing techniques for *Pseudomonas aeruginosa*. J Clin Microbiol 33 (3):528-534

23. Sobral D, Le Cann P, Gerard A, Jarraud S, Lebeau B, Loisy-Hamon F, Vergnaud G, Pourcel C (2011) High-throughput typing method to identify a non-outbreak-involved *Legionella pneumophila* strain colonizing the entire water supply system in the town of Rennes, France. Appl Environ Microbiol 77 (19):6899-6907. doi:AEM.05556-11 [pii] 10.1128/AEM.05556-11

24. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FS, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E, Westbrock-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim R, Smith K, Spencer D, Wong GK, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock RE, Lory S, Olson MV (2000) Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. Nature 406 (6799):959-964. doi:10.1038/35023079

25. Yang L, Haagensen JA, Jelsbak L, Johansen HK, Sternberg C, Hoiby N, Molin S (2008) In situ growth rates and biofilm development of *Pseudomonas aeruginosa* populations in chronic lung infections. J Bacteriol 190 (8):2767-2776. doi:JB.01581-07 [pii] 10.1128/JB.01581-07

26. Ballard L, Adams P, Bao Y, Bartley D, Bintzler D, Kasch L, Petukhova L, Rosato C (2002) Strategies for genotyping: Effectiveness of tailing primers to increase accuracy in short tandem repeat determinations. J Biomol Tech 13 (1):20-29

27. Waters RC, O'Toole PW, Ryan KA (2007) The FliK protein and flagellar hook-length control. Protein Sci 16 (5):769-780. doi:16/5/769 [pii] 10.1110/ps.072785407

28. Vogler AJ, Keys CE, Allender C, Bailey I, Girard J, Pearson T, Smith KL, Wagner DM, Keim P (2007) Mutations, mutation rates, and evolution at the hypervariable VNTR loci of *Yersinia pestis*. Mutat Res 616 (1-2):145-158. doi:S0027-5107(06)00321-6 [pii] 10.1016/j.mrfmmm.2006.11.007

29. Noller AC, McEllistrem MC, Shutt KA, Harrison LH (2006) Locus-specific mutational events in a multilocus variable-number tandem repeat analysis of *Escherichia coli* O157:H7. J Clin Microbiol 44 (2):374-377. doi:44/2/374 [pii] 10.1128/JCM.44.2.374-377.2006

30. Price EP, Hornstra HM, Limmathurotsakul D, Max TL, Sarovich DS, Vogler AJ, Dale JL, Ginther JL, Leadem B, Colman RE, Foster JT, Tuanyok A, Wagner DM, Peacock SJ, Pearson T, Keim P (2010) Within-host evolution of *Burkholderia pseudomallei* in four cases of acute melioidosis. PLoS Pathog 6 (1):e1000725. doi:10.1371/journal.ppat.1000725

2.3. *S. aureus* et source de contamination alimentaire

a. Résumé

S. aureus est une bactérie commensale retrouvée dans diverses niches écologiques comme la peau ou les muqueuses des mammifères à sang chaud. Chez l'homme, *S. aureus* se loge principalement au niveau des fosses nasales ; près de 25 % de la population comprendrait des porteurs sains. Cependant *S. aureus* est une bactérie opportuniste pouvant devenir pathogène. La virulence de cette bactérie couplée à sa multirésistance aux antibiotiques assure à *S. aureus* le deuxième rang des pathogènes nosocomiaux avec plus de 30 % des infections nosocomiales recensées. *S. aureus* est une bactérie ubiquitaire qui persiste et se développe dans certains produits alimentaires. Lorsque qu'un aliment souillé est ingéré, le pouvoir pathogène de *S. aureus* s'exprime par la production d'entérotoxines alors responsables d'intoxinations alimentaires caractérisées par des troubles digestifs graves. *S. aureus* représente donc un enjeu majeur pour de nombreux secteurs : clinique, vétérinaire et agro-alimentaire. Les professionnels des secteurs concernés doivent mettre en place les mesures adéquates afin d'identifier les sources, les modes et les vecteurs de contamination. Pour répondre à cette problématique un protocole de génotypage automatisé par MLVA a été développé.

Le protocole MLVA-16_{Orsay}, comprenant l'analyse de 16 marqueurs amplifiés en deux réactions de PCR multiplexe, a été validé sur une collection de référence de 89 souches cliniques. Les différents critères définis par l'ESGEM sont satisfaits et la méthode s'avère plus discriminante que la PFGE, la méthode standard. De plus, le MLVA-16_{Orsay} est congruent avec le MLST suggérant la pertinence du MLVA comme outil de structuration des populations de *S. aureus*. Les complexes clonaux, ou CCs, de la collection de référence sont mis en évidence et le calcul de l'indice de déséquilibre de liaison sur cet échantillon confirme la clonalité de la population.

La méthodologie a été appliquée dans le cadre d'une étude rétrospective évaluant la diversité de la population de *S. aureus* dans le monde animal. Cette étude démontre la versatilité du protocole, adapté aussi bien aux souches humaines qu'animales. Cette étude montre un panorama de la diversité et des particularités de la population de *S. aureus* chez les animaux. Plusieurs CCs sont quasi-exclusivement issus du réservoir animal et certains d'entre eux ont une étroite spécificité d'hôtes (CC9 pour le cochon, CC133 pour les petits ruminants). D'autres CCs ont un spectre d'hôtes élargi illustrant la facette zooanthroponotique du pathogène : adaptation récente du CC5 à la volaille ou circulation favorisée des CC8 ou CC45 par la proximité entre les humains et les animaux domestiques. Ces derniers CCs exhibent une dichotomie très marquée : ils sont majoritaires chez l'Homme (entre 20 et 60 % des isolats cliniques) et quasi-absents chez les animaux de ferme (moins de 2% des isolats récoltés). La connaissance de ce fond génétique a permis de mettre en exergue que les souches impliquées dans des intoxications alimentaires appartenaient à des CCs humains suggérant une origine de contamination exogène, et a fortiori humaine, lors du traitement du produit transformé.

b. ***Article 3 (publié par PlosOne)***

High throughput Multiple Locus Variable Number of Tandem Repeat analysis (MLVA) of *Staphylococcus aureus* from human, animal and food sources.

Daniel Sobral[1,2,3], Stefan Schwarz[4], Dominique Bergonier[5,6], Anne Brisabois[7], Andrea T. Feßler[4], Florence B. Gilbert[8], Kristina Kadlec[4], Benoit Lebeau[3], Fabienne Loisy-Hamon[3], Michaël Treilles[9], Christine Pourcel[1,2], Gilles Vergnaud*[1,2,10]

Univ Paris-Sud, Institut de Génétique et Microbiologie, UMR 8621, Orsay, France[1]

CNRS, Orsay, France[2]

Centre Européen d'Expertise et de Recherche sur les Agents Microbiens (CEERAM), La Chapelle sur Erdre, France[3]

Institute of Farm Animal Genetics, Friedrich-Loeffler-Institut (FLI), Neustadt-Mariensee, Germany[4]

INRA, UMR1225, IHAP, Toulouse, France[5]

Université de Toulouse, INP, ENVT, UMR1225, IHAP, Toulouse, France[6]

ANSES, European Union Community Reference Laboratory for Coagulase Positive Staphylococci, Maisons-Alfort, France[7]

INRA, UR1282 Infectiologie Animale et Santé Publique (IASP), Nouzilly, France[8]

Laboratoire départemental d'analyses de la Manche, Saint-Lô, France[9]

DGA/MRIS- Mission pour la Recherche et l'Innovation Scientifique, Bagneux, France[10]

ABSTRACT

Staphylococcus aureus is a major human pathogen, a relevant pathogen in veterinary medicine, and a major cause of food poisoning. Epidemiological investigation tools are needed to establish surveillance of *S. aureus* strains in humans, animals and food. In this study, we investigated 145 *S. aureus* isolates recovered from various animal species, disease conditions, food products and food poisoning events. Multiple Locus Variable Number of Tandem Repeat (VNTR) analysis (MLVA), known to be highly efficient for the genotyping of human *S. aureus* isolates, was used and shown to be equally well suited for the typing of animal *S. aureus* isolates. MLVA was improved by using sixteen VNTR loci amplified in two multiplex PCRs and analyzed by capillary electrophoresis ensuring a high throughput and high discriminatory power. The isolates were assigned to twelve known clonal complexes (CCs) and a few singletons. Half of the test collection belonged to four CCs (CC9, CC97, CC133, CC398) previously described as mostly associated with animals. The remaining eight CCs (CC1, CC5, CC8, CC15, CC25, CC30, CC45, CC51), representing 46% of the animal isolates, are common in humans. Interestingly, isolates responsible for food poisoning show a CC

distribution signature typical of human isolates and strikingly different from animal isolates, suggesting a predominantly human origin.

INTRODUCTION

Staphylococcus aureus is a common commensal and frequent colonizer of humans and many animal species including companion animals as well as food-producing animals. In humans, the epithelium of the anterior nares is the primary ecological niche. *S. aureus* is also a major pathogen involved in a wide variety of diseases such as purulent skin and subcutaneous infections, pneumonia, endocarditis, abscesses and bacteremia. Moreover, *S. aureus* is an emerging issue in veterinary medicine and a cause of food poisoning by its ability to produce heat-stable enterotoxins [1].

The transfer of *S. aureus* isolates between humans and animals, especially in the case of livestock-associated MRSA ST398, has recently gained particular attention [2]. However, relatively little is known about the more global diversity of *S. aureus* isolates of animal origin [3,4,5,6,7,8,9,10,11,12,13,14,15,16,17]. This limits our ability to identify for example the origin of strains responsible for food poisoning. In order to implement control measures targeted at reservoirs and transmission routes, it is necessary to further improve current knowledge about animal-associated *S. aureus*.

Essentially three techniques are currently used for the large-scale analysis of the diversity of *S. aureus* isolates, namely multi locus sequence typing (MLST), spa typing, and multiple locus variable number of tandem repeats (VNTR) analysis (MLVA). In addition, pulsed field gel electrophoresis (PFGE) is still widely used and considered the gold-standard for typing *S. aureus* isolates. It has a high discriminatory power and it can be used for many bacterial pathogens. It is however not appropriate for routine interlaboratory comparisons [18]. MLST studies allowed the description of major clonal complexes (CC) underlying the *S. aureus* population structure [19,20]. MLST suffers from its relatively high costs and has a moderate discriminatory power. The spa typing is a widely used method in which variations in a highly variable tandem repeat are characterized by sequencing. The Ridom Spaserver http://spaserver.ridom.de allows the designation of spa types [21,22]. The spa typing is a very powerful tool, and is currently the most commonly used first line assay. However it may fail to identify new lineages due to inherent homoplasia and variable evolutionary rate of spa alleles and clustering based on spa data is complex. MLVA was developed more recently. Homoplasia at individual VNTR loci and potentially low variability of specific alleles are compensated at least partly by the use of multiple loci. An assay comprising as little as 8 VNTR loci (called MLVA-$8_{\text{Bilthoven}}$ in the present report) was highly congruent with MLST and able to assign a new isolate to the correct CC for much lower costs [23]. The 8 loci were amplified in two multiplex PCRs and analyzed by capillary electrophoresis. A MLVA assay with 14 loci (MLVA-14_{Orsay}) providing higher discriminatory power was used in a survey of 309 isolates including clinical MRSA isolates, nasal carriage isolates and representatives of the main CCs present in humans [24]. Both schemes can be adapted to low

resolution DNA sizing equipment (such as agarose gels) as well as to higher throughput systems (such as capillary electrophoresis-based devices). MLVA data can be accessed via internet (a list of such databases is maintained on http://minisatellites.u-psud.fr). These databases can be queried even if a subset of loci is used although the discriminatory power and typing assignment precision might then be decreased.

In the present study, we have used MLVA as a first line assay, complemented when necessary by spa typing and MLST data. We have selected 16 loci for the MLVA assay, subsequently called MLVA-16Orsay, which essentially merges MLVA-8Bilthoven and MLVA-14Orsay and we have automated this assay. The products of two multiplex PCR amplifications were resolved by capillary electrophoresis, and the alleles from each of the 16 targeted loci were automatically identified. This expanded MLVA assay was used for the typing of 251 *S. aureus* isolates: the present retrospective investigation included 106 previously typed human clinical isolates, 98 isolates collected from various animal sources and mostly associated with a variety of diseases in these animals, 34 isolates recovered from food products, and 13 enterotoxigenic *S. aureus* from cases of food poisoning.

MATERIALS AND METHODS

Strains. Two hundred and fifty-one isolates were included in the study. One hundred and six are human clinical isolates: ninety isolates from the HARMONY project reference collection kindly provided by Alex van Belkum were used to perform the development and initial validation of the automated MLVA protocol [25]; sixteen isolates were selected among two previously described collections to represent the diversity of clinical *S. aureus* strains from humans [24,26]. Ninety-eight isolates were previously collected from different disease conditions in farm and domestic animals [27,28,29]. Thirty-four isolates were recovered from food [30]. Thirteen isolates were associated with food poisoning (Table 1 and Table S1).

Sample origin	Animal[a]	Food[a]	Food poisoning[a,b]	Total[a]
Companion animals	17 (1)	NA[c]	NA[c]	17 (1)
Poultry	30 (0)	33 (33)	1(0)	64 (33)
Small ruminants	11 (0)	1(0)	7(0)	19 (0)
Swine	32 (5)	0	1(0)	33 (5)
Rodent	2 (0)	0	0	2 (0)
Horse	5 (0)	0	0	5 (0)
Cattle	1(0)	0	3 (0)	4 (0)
Total	98 (5)	34 (33)	12 (0)	144 (39)

Table 1. Overview of the origins of the animal and food isolates
[a] the number of MRSA isolates is indicated between ()
[b] the origin of one of the 13 food poisoning associated isolates is unknown
[c] NA : not applicable

DNA extraction. Strains were cultured overnight at 37°C in Brain Heart Infusion (BHI) or Luria Bertani broth. Genomic DNA was extracted using phenol-chloroform extraction or the DNeasy tissue kit (Qiagen, Courtaboeuf, France) after treatment with lysostaphin (100 mg/l) (Ambi products LLC, USA). Nucleic acid quality and concentration were estimated with an ND-1000 spectrophotometer (NanoDrop, Labtech, Palaiseau, France). Samples diluted in water at 5 ng/µl were used as DNA template for PCR amplification.

Selection of VNTRs and MLVA typing. Twelve loci previously investigated by Pourcel *et al.* and all eight loci used by Schouls *et al.* [23,24] were merged in a single assay comprising 16 loci. The 16 VNTR loci were amplified in two multiplex PCRs using the ceeramTools® Staphylococcus typing kit (Ceeram, La Chapelle sur Erdre, France). PCR reaction 1 amplifies the ten VNTR loci Sa0122, Sa0311, Sa0387, Sa0550, Sa0684, Sa0964, Sa1097, Sa1194, Sa1729, Sa1866. PCR reaction 2 amplifies the six VNTR loci Sa0266, Sa0704, Sa1132, Sa1291, Sa2039, Sa2511 (Tables 2 and 3). VNTRs Sa0122 and Sa0266 are located in the genes *spa* and *coa* respectively. VNTRs Sa0311, Sa1729 and Sa1866 are members of the family of intergenic repeated elements called “STAR” for *S. aureus* repeats [31]. Briefly, the kit includes two primer mixes, one for each multiplex reaction. Forward primers were fluorescently labeled at the 5’ end, reverse primers were synthesized unlabeled and tailed (Applied Biosystems, Courtaboeuf, France) as previously described [32]. Both multiplex PCRs were performed in a final volume of 15 µl using the Qiagen multiplex PCR kit (Qiagen, Courtaboeuf, France). The reactions contained 2 µl template DNA (5 ng/µl), 7.5 µl of 2X multiplex PCR mastermix and 5.5 µl of primer mix. The PCR reactions were run on a Veriti® thermal cycler (Applied Biosystems, Courtaboeuf, France) using the following conditions: initial denaturation cycle for 15 min at 95 °C, 15 cycles touchdown PCR (30 s at 95 °C; 60 s at 75 °C, with 1.0 °C drop in temperature each next cycle, 70 s at 72 °C), 15 cycles long range PCR (30 s at 95 °C, 60 s at 60 °C, 70 s at 72 °C with 5 s increase in time each next cycle), with a final 10 min step at 72 °C. PCR fragments were purified using Qiagen DyeEx plates (Qiagen, Courtaboeuf, France). For each multiplex reaction, 2 µl of purified PCR product were combined with 7.75 µl HiDi formamide and 0.25 µl GS1200LIZ (Applied Biosystems, Courtaboeuf, France). Samples were loaded onto the ABI3130 capillary sequencer using a 50 cm capillary filled with performance-optimized polymer 7 (Applied Biosystems, Courtaboeuf, France) at 60 °C for 6200 s with a running voltage of 12 kV, and an injection time of 10 s at an injection voltage of 1.6 kV.

VNTR[a]	Unit size	Alias[b]	MLVA-16$_{Orsay}$	MVLVA-14$_{Orsay}$[g]	MLVA-10$_{Orsay}$[h]	MLVA-8$_{Bilthoven}$[i]
Sa0122	24	Spa, SIRU21[c], VNTR24_01[d]	X	X	X	X
Sa0266	81	Coa, VNTR81_01[d]	X	X	X	X
Sa0311	55		X	X	X	
Sa0387	55	SIRU1[c]	X			
Sa0550	21	VNTR21_01[d]	X			X
Sa0684	61	VNTR61_02[d]	X			X
Sa0704	67	VNTR67_01[d]	X	X	X	X
Sa0906	56			X		
Sa0964	43	SAV0920-0921[e]	X			
Sa1097	9	Sspa[f], VNTR09_01[d]	X			X
Sa1132	63	VNTR63_01[d]	X	X	X	X
Sa1194	67		X	X	X	
Sa1213	56			X		
Sa1291	64	SIRU13[c]	X	X	X	
Sa1425	58			X		
Sa1729	56		X	X	X	
Sa1756	131	SIRU15[c]		X		
Sa1866	129		X	X	X	
Sa2039	56		X	X	X	
Sa2511	61	VNTR61_01[d]	X			X

Table 2. Constitution of the different MLVA schemes.
a the VNTR locus name reflects genome localisation (in kb) in strain Mu50 refseq NC_002758
b alternative names given in the literature
c from [52]
d from [23]
e from [53]
f from [54]
g MLVA14$_{Orsay}$ corresponds to the 14 panel 1 + panel 2 loci in [24]
h MLVA10$_{Orsay}$ corresponds to the 10 panel 1 loci in [24]
i MLVA8$_{Bilthoven}$ corresponds to the 8 loci used by [23].

VNTR name	PCR[a]	Forward primer sequence (5' → 3')	Reverse primer sequence (5' → 3')	Mu50 size[b]	Allele size range[c]	HGDI* [95% confidence
Sa0122	1	NED-CAGCAGTAGTGCCGTTTG	GCACCAAAAGAGGAAGAC	331 (10)	115-475 (1-16)	0.80 [0.759,0.847]
Sa0266	2	PET-TTGGATATGAAGCGAGACCA	CTTCCGATTGTTCGATGCTT	630 (6)	387-954 (3-10)	0.67 [0.632,0.705]
Sa0311	1	VIC-GTATCAACAAGTGATAGCATCA	AAATGATATTTTCGCAAAATTTATT	316 (3)	206-426 (1-5)	0.64 [0.585,0.691]
Sa0387	1	6FAM-CAAAGTAATAGGCACTACAA	CATTCCAAACATACCATCAC	255 (2)	179-420 (0.5-5)	0.73 [0.685,0.767]
Sa0550	1	PET-GTGACAGATGTAAGACTTAGA	AACTTGATCGACACCAGAGC	847 (5)	784-847,1183-1246 (2-5,21-24)	0.37 [0.257,0.490]
Sa0684	1	6FAM-AGGTATTGGAAGTGAAACAGC	CAACAAGTTGTTCAGCCTGC	1000 (2)	939-1183 (1-5)	0.52 [0.420,0.616]
Sa0704	2	VIC-AAGAGTGTGTAGGGAATGGC	CGATCGCACGATATGGTGG	307 (4)	173-709 (2-10)	0.76 [0.710,0.801]
Sa0964	1	PET-ATCCCAGATTATCCAATACAA	CCAACCTGTTAATCCGATGT	597 (6)	382-769 (1-10)	0.61 [0.539,0.676]
Sa1097	1	PET-GAATTATTGTTATCGCCATTGTC	GCAACTTCTTAAAACAAAATATTG	196 (15)	124-241, 286 (7-20,25)	0.65 [0.558,0.751]
Sa1132	2	6FAM-CTAGTTCAAGCTAGATCAGG	TGGGAGGAATTAATCATGTC	884 (7)	569-1262 (2-13)	0.67 [0.611,0.720]
Sa1194	1	NED-CTGTGTCGGTAGGTTACATT	GGTGCTAAAGTCGATGTAAT	874 (7)	539-1008 (2-9)	0.59 [0.506,0.673]
Sa1291	2	NED-GTCAAGACACAGATATTGCT	GTGTTGCTCTTGAATCATC	870 (4)	678-1254 (1-10)	0.55 [0.442,0.657]
Sa1729	1	6FAM-GTCTCGAATCACTTAACAACG	GACCATGCACTACGTGTTAC	797 (5)	573-853 (1-6)	0.59 [0.522,0.669]
Sa1866	1	VIC-GCTTTACGTGTAATAACACC	GCTTGTTGGTTCAGCTTTAGG	933 (3)	522-1092 (0.5-4)	0.49 [0.400,0.587]
Sa2039	2	6FAM-TATTTCGTTCTACCCCAACT	CATAAATTCAATGTCCTAGGC	275 (3)	163-443 (1-6)	0.65 [0.565,0.726]
Sa2511	2	NED-GGCAAAATGCACATGAAACACT	AAGTCAAGAATATTTAAAATCAATT	370 (3)	248-675 (1-8)	0.82 [0.781,0.860]

Table 3. Oligonucleotide primers used and VNTRs analysed in this study.
(*HGDI calculated on the 90 isolates from the HARMONY collection)
a Multiplex PCR reaction
b Expected amplicon size for strain Mu50 RefSeq NC_002758 (by convention, corresponding number of repeated units for strain Mu50)
c Observed allele size range: amplicons length (number of repeated units)

MLVA data analysis. Each VNTR locus was identified according to color and automatically assigned a size by the GeneMapper software (Applied Biosystems, Courtaboeuf, France). This size was then converted into an allele designation associated with a quality index. Rare intermediate-sized alleles were reported as half-size (.5, Figure 1 and Figure S1) as previously described [24]. New alleles of unexpected size were sequenced. The typing data file was imported into BioNumerics version 6.6 (Applied-Maths, Sint-Martens-Latem, Belgium). Allele designations and allele calling conventions for the VNTRs described in previous work were used as published [24]. The MLVA code is provided in the order corresponding to the genome position in reference strain Mu50 (refseq accession number NC_002758): Sa0122, Sa0266, Sa0311, Sa0387, Sa0550, Sa0684, Sa0704, Sa0964, Sa1097, Sa1132, Sa1194, Sa1291, Sa1729, Sa1866, Sa2039, Sa2511 [33]. Following these conventions, the genotype of the reference strain Mu50 deduced from its genomic sequence is 10-6-3-2-5-2-4-6-15-7-7-4-5-3-3-3 (Table 3).

The diversity (D) index and confidence intervals (CI) were calculated as previously described [32]. The UPGMA (unweighted pair group method with arithmetic mean) clustering method was run using the categorical distance coefficient. A cut-off value of 45% similarity was applied to define clusters [24].

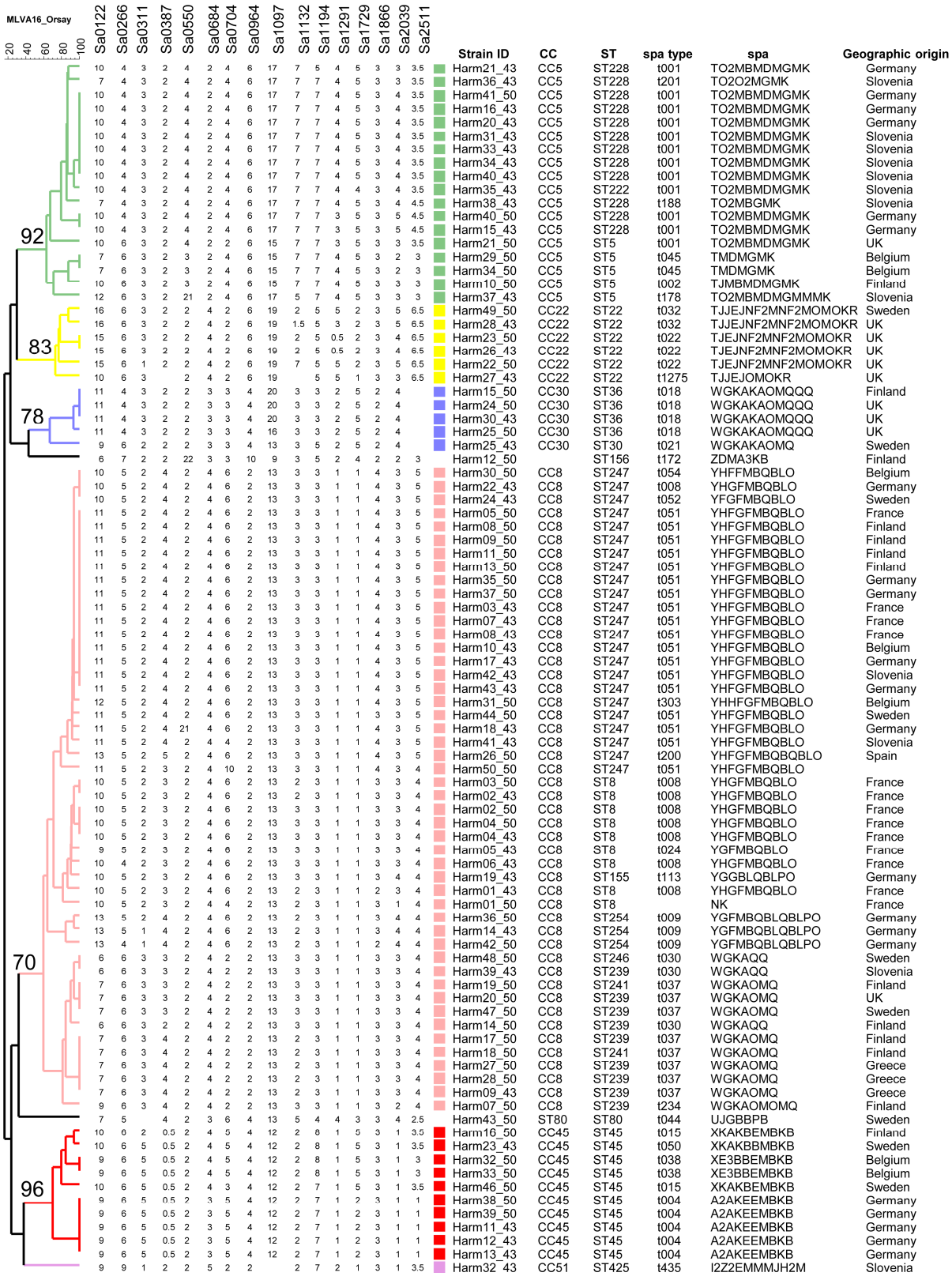
Analysis of linkage disequilibrium, bootstrapping and congruence between different methods. Linkage disequilibrium was measured by using LIAN Ver. 3.5 software accessed at http://guanine.evolbio.mpg.de/ [34]. The Monte-Carlo simulation was run with 100000 iterations. Bootstrap analyses were run with 500 simulations.

Comparison of animal, human and food poisoning isolates. Fischer's exact test was applied to compare the proportions of human-related CCs in three different populations: isolates from animals, isolates from humans and isolates involved in food poisoning. Data from the literature [3,4,5,6,7,8,9,17] and from the present study were analysed to build the population of animal-associated isolates. We combined our results with those described by Wattinger et al. including 20 isolates from food poisoning events, to form the population of 33 isolates involved in food poisoning [35].

DNA sequencing. PCR amplicons were purified using the QIAquick PCR purification (Qiagen, Courtaboeuf, France) and sequenced (Cogenics, UK or Eurofins MWG Operon, Ebersberg, Germany). Sequence data were managed with BioNumerics. The primers and conditions used for the MLST or spa tandem repeat amplification were as described by Enright et al. and Harmsen et al. respectively [21,22]. Alleles and sequence types (STs) were identified using the MLST database (http://www.mlst.net). The spa repeat nomenclature was that of Shopsin et al., and spa types were retrieved from the Ridom SpaServer http://spaserver.ridom.de [36].

RESULTS

Automated multiplex capillary-based MLVA assay development. MLVA-16_{Orsay} integrates the two most recently published MLVA assays, each associated with large databases accessible via internet [23,24] (Table 2). The resulting data can be compared to both data sets. Figure 2 shows a typical capillary electrophoresis pattern of the two multiplex PCRs (Figure 2a PCR1, 10 loci; Figure 2b, PCR2, 6 loci). It preserves the convenient 2-multiplex PCR assay developed by Schouls et al. while including 8 additional VNTR loci [23].



Strain ID	CC	ST	spa type	spa	Geographic origin
Harm21_43	CC5	ST228	t001	TO2MBMDMGMK	Germany
Harm36_43	CC5	ST228	t201	TO2O2MGMK	Slovenia
Harm41_50	CC5	ST228	t001	TO2MBMDMGMK	Germany
Harm16_43	CC5	ST228	t001	TO2MBMDMGMK	Germany
Harm20_43	CC5	ST228	t001	TO2MBMDMGMK	Germany
Harm31_43	CC5	ST228	t001	TO2MBMDMGMK	Slovenia
Harm33_43	CC5	ST228	t001	TO2MBMDMGMK	Slovenia
Harm34_43	CC5	ST228	t001	TO2MBMDMGMK	Slovenia
Harm40_43	CC5	ST228	t001	TO2MBMDMGMK	Slovenia
Harm35_43	CC5	ST222	t001	TO2MBMDMGMK	Slovenia
Harm38_43	CC5	ST228	t188	TO2MBGMK	Slovenia
Harm40_50	CC5	ST228	t001	TO2MBMDMGMK	Germany
Harm15_43	CC5	ST228	t001	TO2MBMDMGMK	Germany
Harm21_50	CC5	ST5	t001	TO2MBMDMGMK	UK
Harm29_50	CC5	ST5	t045	TMDMGMK	Belgium
Harm34_50	CC5	ST5	t045	TMDMGMK	Belgium
Harm10_50	CC5	ST5	t002	TJMBMDMGMK	Finland
Harm37_43	CC5	ST5	t178	TO2MBMDMGMMMK	Slovenia
Harm49_50	CC22	ST22	t032	TJJEJNF2MNF2MOMOKR	Sweden
Harm28_43	CC22	ST22	t032	TJJEJNF2MNF2MOMOKR	UK
Harm23_50	CC22	ST22	t022	TJEJNF2MNF2MOMOKR	UK
Harm26_43	CC22	ST22	t022	TJEJNF2MNF2MOMOKR	UK
Harm22_50	CC22	ST22	t022	TJEJNF2MNF2MOMOKR	UK
Harm27_43	CC22	ST22	t1275	TJJEJOMOKR	UK
Harm15_50	CC30	ST36	t018	WGKAKAOMQQQ	Finland
Harm24_50	CC30	ST36	t018	WGKAKAOMQQQ	UK
Harm30_43	CC30	ST36	t018	WGKAKAOMQQQ	UK
Harm25_50	CC30	ST36	t018	WGKAKAOMQQQ	UK
Harm25_43	CC30	ST30	t021	WGKAKAOMQ	Sweden
Harm12_50		ST156	t172	ZDMA3KB	Finland
Harm30_50	CC8	ST247	t054	YHFFMBQBLO	Belgium
Harm22_43	CC8	ST247	t008	YHGFMBQBLO	Germany
Harm24_43	CC8	ST247	t052	YFGFMBQBLO	Sweden
Harm05_50	CC8	ST247	t051	YHFGFMBQBLO	France
Harm08_50	CC8	ST247	t051	YHFGFMBQBLO	Finland
Harm09_50	CC8	ST247	t051	YHFGFMBQBLO	Finland
Harm11_50	CC8	ST247	t051	YHFGFMBQBLO	Finland
Harm13_50	CC8	ST247	t051	YHFGFMBQBLO	Finland
Harm35_50	CC8	ST247	t051	YHFGFMBQBLO	Germany
Harm37_50	CC8	ST247	t051	YHFGFMBQBLO	Germany
Harm03_43	CC8	ST247	t051	YHFGFMBQBLO	France
Harm07_43	CC8	ST247	t051	YHFGFMBQBLO	France
Harm08_43	CC8	ST247	t051	YHFGFMBQBLO	France
Harm10_43	CC8	ST247	t051	YHFGFMBQBLO	Belgium
Harm17_43	CC8	ST247	t051	YHFGFMBQBLO	Germany
Harm42_43	CC8	ST247	t051	YHFGFMBQBLO	Slovenia
Harm43_43	CC8	ST247	t051	YHFGFMBQBLO	Germany
Harm31_50	CC8	ST247	t303	YHHFGFMBQBLO	Belgium
Harm44_50	CC8	ST247	t051	YHFGFMBQBLO	Sweden
Harm18_43	CC8	ST247	t051	YHFGFMBQBLO	Germany
Harm41_43	CC8	ST247	t051	YHFGFMBQBLO	Slovenia
Harm26_50	CC8	ST247	t200	YHFGFMBQBQBLO	Spain
Harm50_50	CC8	ST247	t051	YHFGFMBQBLO	
Harm03_50	CC8	ST8	t008	YHGFMBQBLO	France
Harm02_43	CC8	ST8	t008	YHGFMBQBLO	France
Harm02_50	CC8	ST8	t008	YHGFMBQBLO	France
Harm04_50	CC8	ST8	t008	YHGFMBQBLO	France
Harm04_43	CC8	ST8	t008	YHGFMBQBLO	France
Harm05_43	CC8	ST8	t024	YGFMBQBLO	France
Harm06_43	CC8	ST8	t008	YHGFMBQBLO	France
Harm19_43	CC8	ST155	t113	YGGBLQBLPO	Germany
Harm01_43	CC8	ST8	t008	YHGFMBQBLO	France
Harm01_50	CC8	ST8		NK	France
Harm36_50	CC8	ST254	t009	YGFMBQBLQBLPO	Germany
Harm14_43	CC8	ST254	t009	YGFMBQBLQBLPO	Germany
Harm42_50	CC8	ST254	t009	YGFMBQBLQBLPO	Germany
Harm48_50	CC8	ST246	t030	WGKAQQ	Sweden
Harm39_43	CC8	ST239	t030	WGKAQQ	Slovenia
Harm19_50	CC8	ST241	t037	WGKAOMQ	Finland
Harm20_50	CC8	ST239	t037	WGKAOMQ	UK
Harm47_50	CC8	ST239	t037	WGKAOMQ	Sweden
Harm14_50	CC8	ST239	t030	WGKAQQ	Finland
Harm17_50	CC8	ST239	t037	WGKAOMQ	Finland
Harm18_50	CC8	ST241	t037	WGKAOMQ	Finland
Harm27_50	CC8	ST239	t037	WGKAOMQ	Greece
Harm28_50	CC8	ST239	t037	WGKAOMQ	Greece
Harm09_43	CC8	ST239	t037	WGKAOMQ	Greece
Harm07_50	CC8	ST239	t234	WGKAOMOMQ	Finland
Harm43_50	ST80	ST80	t044	UJGBBPB	Sweden
Harm16_50	CC45	ST45	t015	XKAKBEMBKB	Finland
Harm23_43	CC45	ST45	t050	XKAKBBMBKB	Sweden
Harm32_50	CC45	ST45	t038	XE3BBEMBKB	Belgium
Harm33_50	CC45	ST45	t038	XE3BBEMBKB	Belgium
Harm46_50	CC45	ST45	t015	XKAKBEMBKB	Sweden
Harm38_50	CC45	ST45	t004	A2AKEEMBKB	Germany
Harm39_50	CC45	ST45	t004	A2AKEEMBKB	Germany
Harm11_43	CC45	ST45	t004	A2AKEEMBKB	Germany
Harm12_43	CC45	ST45	t004	A2AKEEMBKB	Germany
Harm13_43	CC45	ST45	t004	A2AKEEMBKB	Germany
Harm32_43	CC51	ST425	t435	I2Z2EMMMJH2M	Slovenia

Figure 1. Dendrogram of the HARMONY collection using MLVA-16$_{Orsay}$. Color coding is according to MLST clonal complex assignment whereas clustering is done according to the displayed MLVA data. Strain Id, clonal complex, sequence type, spa type, spa code and geographic origin are indicated. MLVA cluster bootstrap values are shown for the main clusters.

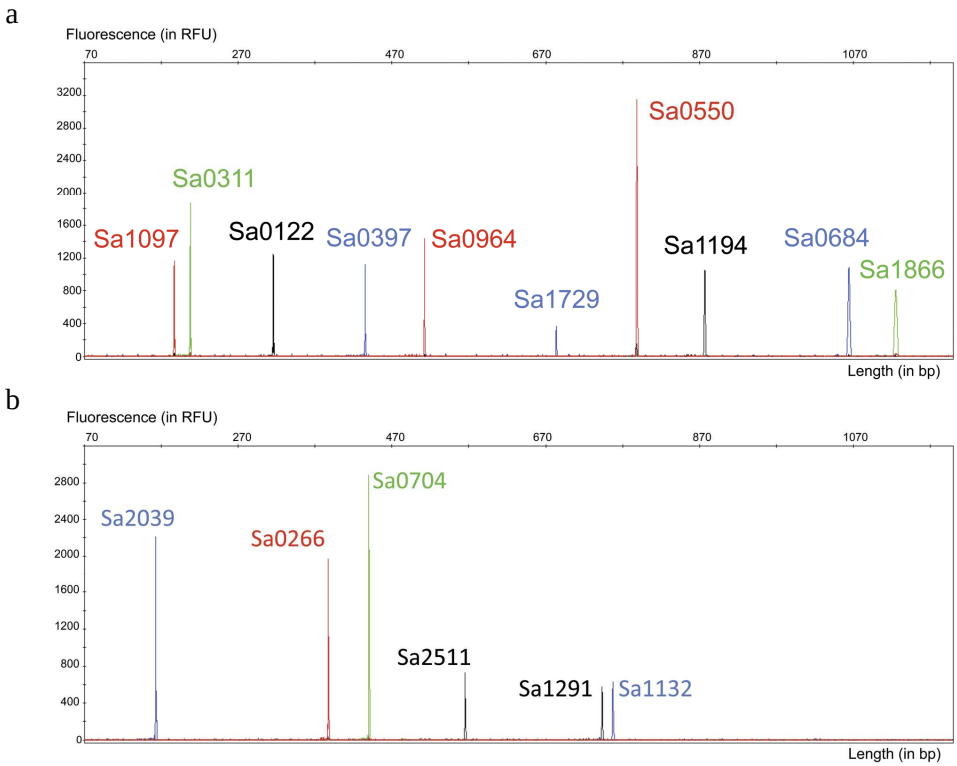


Figure 2. Electrophoregrams showing multiplex PCR amplicons resolved by capillary electrophoresis a. PCR1 ten dye-colored coamplified VNTR loci. b. PCR2 six dye-colored coamplified VNTR loci.

Efficiency of the MLVA-16$_{Orsay}$ protocol. The MLVA-16$_{Orsay}$ scheme was first tested on 90 well-described isolates of the HARMONY reference collection (Figure 1). These isolates represent epidemic or major nosocomial MRSA clones from the mid-1980s to 1998 and were previously investigated by MLVA-14$_{Orsay}$ [24]. Nine missing values were observed in the present study among the expected 1440 values. Sa2511 was not amplified in five closely related isolates. Thus, MLVA-16$_{Orsay}$ provided high typeability (T value = 99.4%). The discriminatory index D was 0.9625, compared to 0.9531 for subset MLVA-8$_{Bilthoven}$. Forty-nine types were identified when using the 16 loci, compared to 41 with the 8 loci assay.

Congruence between MLVA-16$_{Orsay}$ and MLST. Direct comparison between MLST and MLVA clustering based on the 90 isolates of the HARMONY collection showed a strong correlation between these two genotyping methods (Figure 1 and Figure 3), in agreement with previous reports [23,24]. For instance, the congruence between MLST and MLVA-16$_{Orsay}$ was 79.3% and MLVA correctly assigned isolates to their MLST defined CC. CC nodes were supported by high bootstrap values demonstrating the reliability of MLVA-16Orsay clustering for *S. aureus* population investigation (Figure 1). MLVA-16$_{Orsay}$ was much more discriminatory than MLST (MLST distinguishes 17 STs compared to the 49 MLVA-16Orsay types and the discriminatory index D for MLST is 0.8856). The standardized index of linkage association for MLST was 0.349. In comparison the standardized index of linkage association for MLVA-16$_{Orsay}$ was 0.242. The detected linkage disequilibrium was highly significant in both cases (P <10-5). The lower linkage detected by MLVA was previously observed in *Legionella pneumophila* and might be a consequence of homoplasy at VNTR loci [37].

Figure 3. Congruence of MLVA and MLST. Congruence of MLVA schemes (MLVA-16_{Orsay}, MLVA-10_{Orsay}, and MLVA-$8_{Bilthoven}$) and MLST using a Pearson correlation coefficient.

Diversity among *S. aureus* isolates collected from animals. The MLVA-16_{Orsay} assay was used to type 98 animal isolates. A full data set was obtained with one exception: VNTR Sa1291 could not be amplified in isolate sa263. The 98 isolates were resolved into 59 MLVA-16_{Orsay} types (MTs) distributed in 12 clusters and 4 singletons as shown on the minimum spanning tree of Figure 4. MLVA-$8_{Bilthoven}$ resolves 47 MTs (D = 0.9571). In terms of diversity, human-associated (106 isolates) and animal-associated (98 isolates) isolates were similar, with MLVA-16_{Orsay} D values of 0.9727 and 0.9788, respectively.

CC distribution of the animal-associated isolates deduced from previously published or de novo MLST data. The tentative identification of the clusters in the animal and food-product isolates was done by comparison with previously published data [23,24,26]. The spa typing was used to confirm some of the assignments. The largest clusters were CC398 and CC5, which comprised 27% (26 isolates) and 17% (17 isolates) of the typed isolates, respectively (Figure S1). Eleven MLVA-16Orsay genotypes were observed in CC398. Three MLVA-16_{Orsay} clusters comprising 12, 9 and 2 isolates could be assigned to CC9, CC133 and CC97 respectively (Figure S1). Figure 4 shows the distribution of animal strains among the different CCs. The 32 porcine isolates clustered into 5 CCs, CC398 (12 isolates, 38%), CC9 (12 isolates, 38%), CC30 (5 isolates, 15%), CC1 (2 isolates, 6%) and CC97 (1 isolate, 3%). All porcine MRSA isolates belonged to CC398. The 30 isolates from poultry were exclusively MSSA and clustered mainly in 2 CCs: CC398 (13 isolates, 43%) and CC5 (15 isolates, 50%). The 17 isolates from dogs and cats were all MSSA, except for one MRSA CC398 isolate. They were distributed into 4 CCs, CC45 (6 isolates, 35.3%), CC8 (4 isolates, 23.5%), CC15 (3 isolates, 17.6%), CC5 (2 isolates) and a singleton (sa263). The 11 isolates from small ruminants comprised nine CC133 isolates (these MSSA isolates were obtained from sheep of the same German flock suggesting the presence of an epidemic strain) and two non-grouped isolates from mastitis. The five isolates from horses, all MSSA, belonged to three clusters, CC1 (two isolates), CC30 (two isolates) and CC25 (one isolate). Notably, both CC1 isolates shared the same MLVA-16_{Orsay} profile although they were collected in two different countries and 16 years apart. Similarly, the two rodent isolates belonged to the same rare lineage (CC51) although they were collected in two different places and from different disease conditions.

CC distribution among the food-associated isolates of the test collection. Two groups of isolates were recovered from food: 13 isolates originated from cases of food-poisoning and 34 MRSA isolates originated from food not related to poisoning events (food isolates are exclusively MRSA because of the screening procedure [30]). Seven of the 13 food-poisoning associated isolates were collected from dairy products, five were recovered from meat products, and the precise food origin of one isolate was unknown (Figure 4, Table 1 and Table S1). Altogether, 6 isolates belonged to CC8, and single isolates to CC1, CC5, CC25, CC45, CC97, two isolates were not assigned to CCs. The second group was almost exclusively composed of CC398 isolates from poultry meat or poultry meat products (29 of 34 isolates). The remaining five isolates belonged to CC5, CC9 or CC133. Figure S2 shows a minimum spanning tree of all 55 CC398 isolates from animal or food investigated. Figure S3 shows a minimum spanning tree of all isolates based upon MLVA-16_{Orsay} data. Animal and food samples are colored (green for MSSA samples, red for MRSA samples, poultry samples are cross-hatched).

Meta-analysis of human-related *S. aureus* CCs prevalence in food products involved in staphylococcal intoxications. We searched the literature for reports investigating *S. aureus* in animals. We identified 836 farm animal isolates for which a CC assignment is known [3,4,5,6,7,8,9,17]. Ten (1.2 %) and four (0.5%) belonged to CC8 and CC45, respectively. In contrast CC8 and CC45 represent 10-40 % and 10-20% of human isolates respectively [24,26,38,39,40]. Thirty percent (10/33) and twenty one percent (7/33) of the thirty-three isolates involved in food poisoning investigated in this study or by Wattinger *et al.* [35], belonged to CC8 and CC45 respectively. The difference between the proportion of CC8/CC45 in the animal-associated isolates population and the food poisoning isolates is highly significant according to Fisher's exact test ($P<0.05$). In sharp contrast, the proportion of CC8/CC45 among food poisoning isolates is highly similar to the proportion of CC8/CC45 isolates among human isolates. This observation strongly suggested that the isolates associated with food poisoning were mainly of human origin.

High discriminatory power of MLVA-16Orsay in CC398. MLVA-16_{Orsay} distinguished 19 genotypes with a diversity index of 0.8728 (Figure S2 part A). In comparison, 9 genotypes are resolved when using the MLVA-$8_{Bilthoven}$ subset, with a much lower diversity index of 0.6728 (Figure S2 part B). This difference is largely due to locus Sa1291 (Figure S1) which is not otherwise an exceptionally variable locus (Table 3).

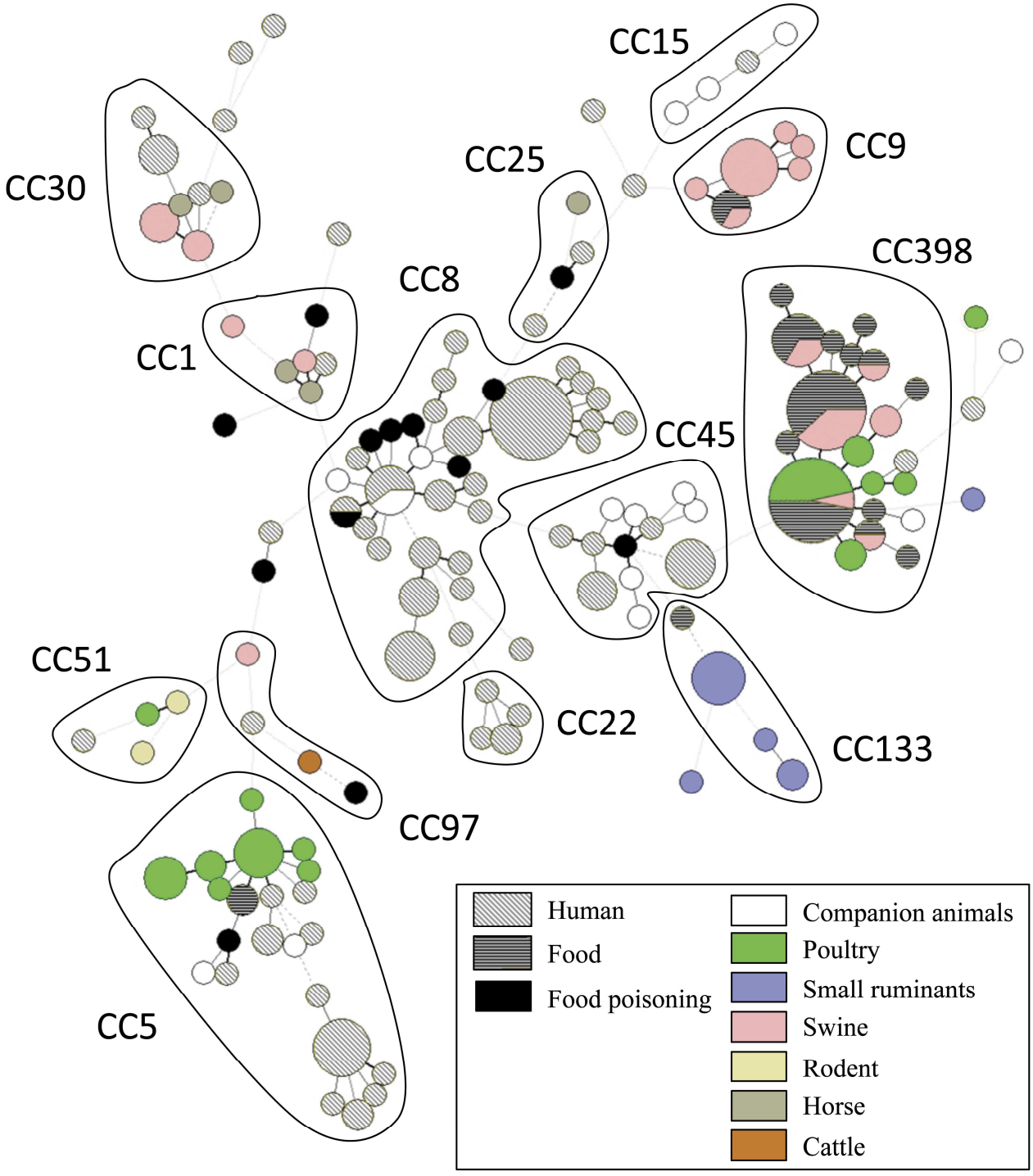


Figure 4. Minimum spanning tree of the 251 *S. aureus* isolates using MLVA-16$_{Orsay}$. Minimum spanning tree of the 251 *S. aureus* isolates (106 human-associated isolates, 98 animal-associated isolates and 47 isolates from food products among which 13 were related with food-poisoning) using MLVA-16$_{Orsay}$. Each circle represents a MLVA genotype. The size of each circle indicates the number of isolates within this MLVA genotype. The different clusters are annotated. The host origin is indicated with a specific color. Human and food isolates are highlighted with two different hatch patterns.

DISCUSSION

The trace-back analysis of food-borne infections requires the availability of appropriate genotyping tools, *i.e.* highly cost-efficient, and fast methods for high-throughput analysis, backed up by relevant and easily accessible typing databases. The present investigation further illustrates the relevance of MLVA for large scale population investigations of *S. aureus*, a major pathogen and source of food intoxications.

Automated capillary-based 16 loci MLVA assay. In the commercially available typing kit (ceeramTools®, Ceeram, La Chapelle sur Erdre, France), 16 loci are amplified in two multiplex PCRs. These 16 loci were chosen in order to provide data directly comparable with previously developed databases based upon the typing of 2 subsets of these loci [23,24]. The assay is able to correctly assign *S. aureus* isolates to defined MLST clonal complexes and further differentiate within these CCs. Homoplasy associated with VNTR loci is presumably efficiently compensated by the analysis of multiple loci. MLVA is equally well adapted for studying *S. aureus* epidemiology regardless of the sample origin (animal or human). The four steps of the MLVA procedure (DNA extraction, amplification of the 16 VNTRs in 2 multiplex PCRs, fragment analysis by capillary electrophoresis, MLVA code assignment) were standardized to be usable and understandable by non-expert users. The resulting data can be queried against freely accessible internet MLVA databases such as http://mlva.u-psud.fr.

MLVA-based diversity analysis of animal-associated *S. aureus*. Ninety-six percent of the isolates were clustered within twelve known CCs. Based on MLVA results, the studied animal-associated isolates were globally as diverse as human-associated ones suggesting that, as a rule, the occurrence of *S. aureus* among animals is not a recent event. Approximately half of the animal isolates belonged to eight well-known human CCs in agreement with previous studies showing the wide host range of some CCs [3,4,5]. However, recent host adaptation, sporadic contamination or the presence of widespread lineages must also be taken into account as possible explanations. Evidence for recent host adaptation is provided by the well-described CC5 isolates in poultry [41]. In the present study, companion animal-associated isolates are almost exclusively found in three main clinical-associated CCs (CC8, CC15 and CC45). This illustrates the spread of *S. aureus* isolates from humans to animals [42].

Animal-adapted clonal complexes. In contrast, the other half of the animal isolates is assigned to CCs not found or uncommon in humans (CC9, CC97, CC133 and CC398) [16]. This observation confirms previous studies suggesting the existence of animal-specific lineages [4,5,6,10,13,14,43]. Recently, Guinane *et al.* have provided evidence that CC133 which is a frequent colonizer of small ruminants, evolved as a host switch from human to ruminant followed by adaptive genome diversification [43]. MSSA-CC9 isolates have been reported in pig farmers and also from infections of swine in France and Germany showing that CC9 is able to change hosts and colonize humans [44,45]. Two MRSA-CC9 from chicken meat were identified [30]. This lineage is very

uncommon and recently emerged from its porcine reservoir in Asia [46,47,48,49]. Clinical human cases due to MRSA-CC9 appeared in the same period [48].

The emerging CC398. In this study, CC398 was found equally in isolates collected from pigs or poultry. Half of the poultry-associated isolates belonged to CC398 whereas CC5, rather than CC398, was previously shown to be the dominant lineage in poultry. This is, to our knowledge, the second report of MSSA CC398 isolates from poultry. This observation might suggest that the lineage was already present among poultry as MSSA and has subsequently evolved as MRSA by independent acquisition of different SCCmec elements. MSSA CC398 could have disseminated from pigs to other food-producing animals, perhaps via farm workers, and the SCCmec cassette could have been acquired in other hosts. Alternatively strains may spontaneously excise part or all of the SCCmec and thus reverse to MSSA [50,51]. Within CC398, various closely related spa types have been described (i.e. t011, t034, t108, t539 and t1793). Schouls and colleagues investigated 216 isolates belonging to CC398, among which 100 were pig-related [23]. They observed little diversity within this complex using MLVA-$8_{Bilthoven}$ (D = 0.721). A similar results was obtained here with MLVA-$8_{Bilthoven}$ (9 genotypes in 55 isolates, D = 0.6728). In contrast, the 16-loci MLVA assay discriminates 19 genotypes (D = 0.8728), suggesting that it might be of high interest to further differentiate CC398 isolates. Due to the higher multiplexing achieved in the MLVA-16_{Orsay} assay as described here compared to MLVA-$8_{Bilthoven}$, the typing of 16 loci instead of 8 does not significantly increase the cost and workload.

MLVA typing as a microbial source tracking tool. The present investigation suggested that *S. aureus* isolates involved in food poisoning are mainly strains found in humans rather than in animals. Among the 33 isolates sampled from food products involved in food poisoning and investigated in this study and by Wattinger *et al.* [35], 10 and 7 belonged to the predominantly human clusters CC8 and CC45, respectively. Given the high human host specificity of CC8 and CC45, this finding provided evidence for the role of humans as a major source of contamination. In sharp contrast, food isolates not associated with food poisoning were almost exclusively assigned to animal-specific clones (CC398, CC9 and CC133) and no CC8 or CC45 isolates were found. In the present study, the CC97 isolate 363F is the only isolate implicated in a food intoxication event with a most likely animal origin. Although these observations need to be substantiated by the analysis of a larger test population, they point towards the role of humans in the contamination of food with enterotoxigenic *S. aureus*.

In conclusion, we merged in this work previous MLVA schemes in a rapid and efficient automated multiplex capillary-based MLVA assay for the high-resolution genotyping of *S. aureus* isolates. The numeric MLVA code is produced automatically and a quality score can be defined facilitating the development of quality controlled databases. The described MLVA-16_{Orsay} assay ensures the same clustering as MLST, assigning similarly *S. aureus* isolates to MLST defined clonal complexes. *S. aureus* MLVA typing data from the 16 loci or any convenient subset can be queried via

http://mlva.u-psud.fr. MLVA is well-suited and compatible for genotyping of animal-associated *S. aureus* as well as human isolates. The present molecular typing analysis provided further insights into the diversity of animal-associated *S. aureus*. We highlighted that the animal-associated population is very diverse suggesting that animal colonization by *S. aureus* is globally ancient. In the present study, the *S. aureus* isolates from animals were divided into human-related and animal-specific CCs. Some CCs are able to switch between a great variety of hosts (i.e. CC5, CC45), whereas others seem to be strongly specific to particular human or animal hosts (i.e. CC9, CC133, CC8). The presented data indicates that *S. aureus* isolates from cases of food poisoning were most likely of human origin.

Acknowledgements

The contributions of SS, ATF and KK are financially supported by the German Federal Ministry of Education and Research (BMBF) through the German Aerospace Center (DLR), grant number 01KI1014D (MedVet-Staph). This study also benefited from the support of the association Vaincre La Mucoviscidose (Grant N° RC0630). The development of tools for the surveillance of bacterial pathogens is supported by the French Direction Générale de l'Armement. We thank Roswitha Becker and Vivian Hensel for excellent technical assistance. We thank Marie-Laure De Buyser from ANSES for her help at the onset of this project.

Conflict of interest

D.S., F.L.-H., and B.L. are employees of Ceeram and hold stocks. Patent licensing arrangements exist with D.S., F.L.-H., B.L., and C.P.

References

1. Le Loir Y, Baron F, Gautier M (2003) *Staphylococcus aureus* and food poisoning. Genet Mol Res 2: 63-76.

2. Smith TC, Pearson N (2011) The emergence of *Staphylococcus aureus* ST398. Vector Borne Zoonotic Dis 11: 327-339.

3. Hasman H, Moodley A, Guardabassi L, Stegger M, Skov RL, et al. (2010) Spa type distribution in *Staphylococcus aureus* originating from pigs, cattle and poultry. Vet Microbiol 141: 326-331.

4. Smyth DS, Feil EJ, Meaney WJ, Hartigan PJ, Tollersrud T, et al. (2009) Molecular genetic typing reveals further insights into the diversity of animal-associated *Staphylococcus aureus*. J Med Microbiol 58: 1343-1353.

5. Sung JM, Lloyd DH, Lindsay JA (2008) *Staphylococcus aureus* host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. Microbiology 154: 1949-1959.

6. Rabello RF, Moreira BM, Lopes RM, Teixeira LM, Riley LW, et al. (2007) Multilocus sequence typing of *Staphylococcus aureus* isolates recovered from cows with mastitis in Brazilian dairy herds. J Med Microbiol 56: 1505-1511.

7. Aires-de-Sousa M, Parente CE, Vieira-da-Motta O, Bonna IC, Silva DA, et al. (2007) Characterization of *Staphylococcus aureus* isolates from buffalo, bovine, ovine, and caprine milk samples collected in Rio de Janeiro State, Brazil. Appl Environ Microbiol 73: 3845-3849.

8. Battisti A, Franco A, Merialdi G, Hasman H, Iurescia M, et al. (2010) Heterogeneity among methicillin-resistant *Staphylococcus aureus* from Italian pig finishing holdings. Vet Microbiol 142: 361-366.

9. Jørgensen HJ, Mork T, Caugant DA, Kearns A, Rorvik LM (2005) Genetic variation among *Staphylococcus aureus* strains from Norwegian bulk milk. Appl Environ Microbiol 71: 8352-8361.

10. Concepcion Porrero M, Hasman H, Vela AI, Fernandez-Garayzabal JF, Dominguez L, et al. (2011) Clonal diversity of *Staphylococcus aureus* originating from the small ruminants goats and sheep. Vet Microbiol.

11. van den Berg S, van Wamel WJ, Snijders SV, Ouwerling B, de Vogel CP, et al. (2011) Rhesus macaques (Macaca mulatta) are natural hosts of specific *Staphylococcus aureus* lineages. PLoS One 6: e26170.

12. Sieber S, Gerber V, Jandova V, Rossano A, Evison JM, et al. (2011) Evolution of multidrug-resistant *Staphylococcus aureus* infections in horses and colonized personnel in an equine clinic between 2005 and 2010. Microb Drug Resist 17: 471-478.

13. de Almeida LM, de Almeida MZ, de Mendonca CL, Mamizuka EM (2011) Novel sequence types (STs) of *Staphylococcus aureus* isolates causing clinical and subclinical mastitis in flocks of sheep in the northeast of Brazil. J Dairy Res 78: 373-378.

14. Sakwinska O, Giddey M, Moreillon M, Morisset D, Waldvogel A, et al. (2011) *Staphylococcus aureus* host range and human-bovine host shift. Appl Environ Microbiol 77: 5908-5915.

15. Lin Y, Barker E, Kislow J, Kaldhone P, Stemper ME, et al. (2011) Evidence of multiple virulence subtypes in nosocomial and community-associated MRSA genotypes in companion animals from the upper midwestern and northeastern United States. Clin Med Res 9: 7-16.

16. Monecke S, Coombs G, Shore AC, Coleman DC, Akpaka P, et al. (2011) A field guide to pandemic, epidemic and sporadic clones of methicillin-resistant *Staphylococcus aureus*. PLoS One 6: e17936.

17. Hata E, Katsuda K, Kobayashi H, Uchida I, Tanaka K, et al. (2010) Genetic variation among *Staphylococcus aureus* strains from bovine milk and their relevance to methicillin-resistant isolates from humans. J Clin Microbiol 48: 2130-2139.

18. te Witt R, van Belkum A, MacKay WG, Wallace PS, van Leeuwen WB (2010) External quality assessment of the molecular diagnostics and genotyping of meticillin-resistant *Staphylococcus aureus*. Eur J Clin Microbiol Infect Dis 29: 295-300.

19. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG (2004) eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol 186: 1518-1530.

20. Robinson DA, Enright MC (2004) Multilocus sequence typing and the evolution of methicillin-resistant *Staphylococcus aureus*. Clin Microbiol Infect 10: 92-97.
21. Enright MC, Day NP, Davies CE, Peacock SJ, Spratt BG (2000) Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. J Clin Microbiol 38: 1008-1015.
22. Harmsen D, Claus H, Witte W, Rothganger J, Turnwald D, et al. (2003) Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for spa repeat determination and database management. J Clin Microbiol 41: 5442-5448.
23. Schouls LM, Spalburg EC, van Luit M, Huijsdens XW, Pluister GN, et al. (2009) Multiple-locus variable number tandem repeat analysis of *Staphylococcus aureus*: comparison with pulsed-field gel electrophoresis and spa-typing. PLoS One 4: e5082.
24. Pourcel C, Hormigos K, Onteniente L, Sakwinska O, Deurenberg RH, et al. (2009) Improved multiple-locus variable-number tandem-repeat assay for *Staphylococcus aureus* genotyping, providing a highly informative technique together with strong phylogenetic value. J Clin Microbiol 47: 3121-3128.
25. Cookson BD, Robinson DA, Monk AB, Murchan S, Deplano A, et al. (2007) Evaluation of molecular typing methods in characterizing a European collection of epidemic methicillin-resistant *Staphylococcus aureus* strains: the HARMONY collection. J Clin Microbiol 45: 1830-1837.
26. Vu-Thien H, Hormigos K, Corbineau G, Fauroux B, Corvol H, et al. (2010) Longitudinal survey of *Staphylococcus aureus* in cystic fibrosis patients using a multiple-locus variable-number of tandem-repeats analysis method. BMC Microbiol 10: 24.
27. Schwarz S, Kadlec K, Strommenger B (2008) Methicillin-resistant *Staphylococcus aureus* and Staphylococcus pseudintermedius detected in the BfT-GermVet monitoring programme 2004-2006 in Germany. J Antimicrob Chemother 61: 282-285.
28. Kadlec K, Ehricht R, Monecke S, Steinacker U, Kaspar H, et al. (2009) Diversity of antimicrobial resistance pheno- and genotypes of methicillin-resistant *Staphylococcus aureus* ST398 from diseased swine. J Antimicrob Chemother 64: 1156-1164.
29. Strommenger B, Kehrenberg C, Kettlitz C, Cuny C, Verspohl J, et al. (2006) Molecular characterization of methicillin-resistant *Staphylococcus aureus* strains from pet animals and their relationship to human isolates. J Antimicrob Chemother 57: 461-465.
30. Fessler AT, Kadlec K, Hassel M, Hauschild T, Eidam C, et al. (2011) Characterization of methicillin-resistant *Staphylococcus aureus* isolates from food and food products of poultry origin in Germany. Appl Environ Microbiol 77: 7151-7157.
31. Cramton SE, Schnell NF, Gotz F, Bruckner R (2000) Identification of a new repetitive element in *Staphylococcus aureus*. Infect Immun 68: 2344-2348.

32. Sobral D, Le Cann P, Gerard A, Jarraud S, Lebeau B, et al. (2011) High-throughput typing method to identify a non-outbreak-involved *Legionella pneumophila* strain colonizing the entire water supply system in the town of Rennes, France. Appl Environ Microbiol 77: 6899-6907.
33. Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, et al. (2001) Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. Lancet 357: 1225-1240.
34. Haubold B, Hudson RR (2000) LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage Analysis. Bioinformatics 16: 847-848.
35. Wattinger L, Stephan R, Layer F, Johler S (2011) Comparison of *Staphylococcus aureus* isolates associated with food intoxication with isolates from human nasal carriers and human infections. Eur J Clin Microbiol Infect Dis.
36. Shopsin B, Gomez M, Montgomery SO, Smith DH, Waddington M, et al. (1999) Evaluation of protein A gene polymorphic region DNA sequencing for typing of *Staphylococcus aureus* strains. J Clin Microbiol 37: 3556-3563.
37. Visca P, D'Arezzo S, Ramisse F, Gelfand Y, Benson G, et al. (2011) Investigation of the population structure of Legionella pneumophila by analysis of tandem repeat copy number and internal sequence variation. Microbiology 157: 2582-2594.
38. Rijnders MI, Deurenberg RH, Boumans ML, Hoogkamp-Korstanje JA, Beisser PS, et al. (2009) Population structure of *Staphylococcus aureus* strains isolated from intensive care unit patients in the netherlands over an 11-year period (1996 to 2006). J Clin Microbiol 47: 4090-4095.
39. Argudin MA, Mendoza MC, Vazquez F, Guerra B, Rodicio MR (2011) Molecular typing of *Staphylococcus aureus* bloodstream isolates from geriatric patients attending a long-term care Spanish hospital. J Med Microbiol 60: 172-179.
40. Vainio A, Koskela S, Virolainen A, Vuopio J, Salmenlinna S (2011) Adapting spa typing for national laboratory-based surveillance of methicillin-resistant *Staphylococcus aureus*. Eur J Clin Microbiol Infect Dis 30: 789-797.
41. Lowder BV, Guinane CM, Ben Zakour NL, Weinert LA, Conway-Morris A, et al. (2009) Recent human-to-poultry host jump, adaptation, and pandemic spread of *Staphylococcus aureus*. Proc Natl Acad Sci U S A 106: 19545-19550.
42. Rich M (2005) Staphylococci in animals: prevalence, identification and antimicrobial susceptibility, with an emphasis on methicillin-resistant *Staphylococcus aureus*. Br J Biomed Sci 62: 98-105.
43. Guinane CM, Ben Zakour NL, Tormo-Mas MA, Weinert LA, Lowder BV, et al. (2010) Evolutionary genomics of *Staphylococcus aureus* reveals insights into the origin and molecular basis of ruminant host adaptation. Genome Biol Evol 2: 454-466.
44. Armand-Lefevre L, Ruimy R, Andremont A (2005) Clonal comparison of *Staphylococcus aureus* isolates from healthy pig farmers, human controls, and pigs. Emerg Infect Dis 11: 711-714.

45. Kehrenberg C, Cuny C, Strommenger B, Schwarz S, Witte W (2009) Methicillin-resistant and -susceptible *Staphylococcus aureus* strains of clonal lineages ST398 and ST9 from swine carry the multidrug resistance gene cfr. Antimicrob Agents Chemother 53: 779-781.

46. Guardabassi L, O'Donoghue M, Moodley A, Ho J, Boost M (2009) Novel lineage of methicillin-resistant *Staphylococcus aureus*, Hong Kong. Emerg Infect Dis 15: 1998-2000.

47. Wagenaar JA, Yue H, Pritchard J, Broekhuizen-Stins M, Huijsdens X, et al. (2009) Unexpected sequence types in livestock associated methicillin-resistant *Staphylococcus aureus* (MRSA): MRSA ST9 and a single locus variant of ST9 in pig farming in China. Vet Microbiol 139: 405-409.

48. Liu Y, Wang H, Du N, Shen E, Chen H, et al. (2009) Molecular evidence for spread of two major methicillin-resistant *Staphylococcus aureus* clones with a unique geographic distribution in Chinese hospitals. Antimicrob Agents Chemother 53: 512-518.

49. Neela V, Mohd Zafrul A, Mariana NS, van Belkum A, Liew YK, et al. (2009) Prevalence of ST9 methicillin-resistant *Staphylococcus aureus* among pigs and pig handlers in Malaysia. J Clin Microbiol 47: 4138-4140.

50. Chlebowicz MA, Nganou K, Kozytska S, Arends JP, Engelmann S, et al. (2010) Recombination between ccrC genes in a type V (5C2&5) staphylococcal cassette chromosome mec (SCCmec) of *Staphylococcus aureus* ST398 leads to conversion from methicillin resistance to methicillin susceptibility in vivo. Antimicrob Agents Chemother 54: 783-791.

51. Boundy S, Zhao Q, Fairbanks C, Folgosa L, Climo M, et al. (2011) Spontaneous SCCmec excision in *Staphylococcus aureus* nasal carriers. J Clin Microbiol.

52. Hardy KJ, Oppenheim BA, Gossain S, Gao F, Hawkey PM (2006) Use of variations in staphylococcal interspersed repeat units for molecular typing of methicillin-resistant *Staphylococcus aureus* strains. J Clin Microbiol 44: 271-273.

53. Gilbert FB, Fromageau A, Gelineau L, Poutrel B (2006) Differentiation of bovine *Staphylococcus aureus* isolates by use of polymorphic tandem repeat typing. Vet Microbiol 117: 297-303.

54. Sabat A, Krzyszton-Russjan J, Strzalka W, Filipek R, Kosowska K, et al. (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: 1801-1804.

2.4. *S. aureus* et adaptation à un écosystème

a. Résumé

S. aureus est un pathogène majeur de nombreux mammifères, notamment des animaux d'élevage, bovins, ovins et caprins. Une relation entre l'hôte et l'expression clinique de la contamination a été mise en évidence ; ainsi il a été observé que *S. aureus* était responsable de pneumonie, infection musculaire, ostéomyélite et septicémie chez les volailles, abcès sous-cutanés, mastite et pododermatite chez le lapin, dermatite et cellulite chez le cheval et septicémie chez les porcs. Cependant, la principale manifestation clinique reste l'infection intramammaire, ou mammite, chez les bovins et les petits ruminants. En France, *S. aureus* est l'une des principales causes de ce type d'infection avec près de 30% des mammites diagnostiquées. Cette pathologie, dont l'enjeu est majeur en élevage laitier, touche près d'un quart des ruminants allaitants et se manifeste par une réduction de la quantité et de la qualité du lait et dans les cas les plus graves par une nécrose de la glande mammaire. Quelles sont les particularités des populations bactériennes déclenchant cette pathologie ?

Dans ce manuscrit, nous décrivons l'analyse génotypique, réalisée par le MLVA-16_{Orsay} automatisé, de 152 isolats issus de mammites de ruminants d'élevage. Nous mettons en évidence que près de 20% des isolats appartiennent à des CCs caractérisés par un tropisme exacerbé pour les glandes mammaires des ruminants. Ces organes représentent donc un écosystème particulier colonisé par une population de *S. aureus* spécifique et notamment par certains clones dont l'émergence semble récente. En effet, et au biais d'échantillonnage près, ces CCs ont une diversité réduite et un nombre d'unités répétées moindre. Même si le modèle évolutif des VNTRs reste à définir, certains auteurs, considèrent que la perte de motifs répétés est un marqueur d'évolution. Cette hypothèse a été validée statistiquement sur l'espèce *M. tuberculosis*. La prudence est cependant de mise lorsque des tentatives phylogénétiques sont réalisées à partir de données de VNTRs.

b. ***Article 4 (en préparation)***

Emergence and evolution of methicillin-susceptible and methicillin-resistant *Staphylococcus aureus* lineages from cases of bovine, ovine and caprine mastitis

Daniel Sobral[a,b,c], Stefan Schwarz[d], Dominique Bergonier[e,f], Florence B. Gilbert[g], Andrea T. Feßler[d], Michaël Treilles[h], Christine Pourcel[a,b], Gilles Vergnaud[a,b,i,*]

[a] Univ Paris-Sud, Institut de Génétique et Microbiologie, UMR 8621, Orsay, France

[b] CNRS, Orsay, France

[c] Ceeram (Centre Européen d'Expertise et de Recherche sur les Agents Microbiens), La Chapelle sur Erdre, France

[d] Institute of Farm Animal Genetics, Friedrich-Loeffler-Institut (FLI), Neustadt-Mariensee, Germany

[e] INRA, UMR1225, IHAP, Toulouse, France

[f] Université de Toulouse, INP, ENVT, UMR1225, IHAP, Toulouse, France

[g] INRA, UR1282, Infectiologie Animale et Santé Publique, Nouzilly, France

[h] Laboratoire départemental d'analyses de la Manche, Route de Bayeux, Saint-Lô, France

[i] DGA/MRIS- Mission pour la Recherche et l'Innovation Scientifique, Bagneux, France

ABSTRACT

Staphylococcus aureus is one of the main etiological agents of mastitis in ruminants. In the present retrospective study, we determined the genetic diversity of individual *S. aureus* isolates from cases of clinical and subclinical mastitis in dairy cattle (n = 118, of which 16 were methicillin-resistant), sheep (n = 18) and goats (n = 16). Using a previously described automated multiple loci variable number of tandem repeats (VNTR) assay (MLVA), the 152 isolates were subdivided into 115 genotypes. This corresponds to a discriminatory index (*D)* value of 0.9936. Comparison with published MLVA data of isolates from diverse human and animal origins revealed the low prevalence (8.5%) of human-related genotypes among the present collection. Eighteen percent of the population belonged to clusters predominantly recovered from mammary gland tissue. MLVA was able to uncover highly specialized lineages some of which displayed a relatively low level of diversity suggesting that they may be of recent emergence.

These findings provide valuable arguments to suggest that specific *S. aureus* lineages have emerged as pathogens particularly adapted to ruminant mammary glands.

INTRODUCTION

Although several bacterial pathogens can cause mastitis, *Staphylococcus aureus* is one of the most important etiologic agents of this disease in cows [1], and the most important in goats and sheep [2,3,4]. Moreover, *S. aureus* infections are chronic, localized deeply in the mammary parenchyma and the bacteria are frequently in an intracellular position. As a consequence the infection may be extremely difficult to cure. Methicillin resistant *S. aureus* (MRSA) isolates are rarely involved in bovine mastitis [5,6]. Most mastitis-associated MRSA were reported to be human contaminants or divergent mecA isolates [6,7,8,9].

During the past decade, the epidemiology of *S. aureus* mastitis in dairy cattle has been studied using various molecular typing methods. Techniques that rely on comparison of electrophoretic patterns, such as PFGE [4,10,11], RAPD analysis [12], ribotyping [11,13] and MLEE [14] proved to be highly discriminatory but interlaboratory comparisons are difficult. Sequence-based typing systems such as MLST or spa typing overcome these problems by producing sharable and storable results [6,15,16,17]. However MLST has a low discriminatory power for a relatively high cost, and the clustering of spa typing results is complex so that these two methods are often combined. These last few years, MLVA schemes taking advantage of tandem repeat polymorphisms were developed for *S. aureus* subtyping [18,19,20] and represent a promising alternative. Recently, we developed an automated capillary-based MLVA assay for *S. aureus* genotyping using 16 loci and demonstrated its suitability for animal-associated *S. aureus* isolates genotyping (Sobral et al., PlosOne, 2012 in press). We previously demonstrated that MLVA, in addition to the much higher discriminatory power, was able to correctly predict clonal complex (CC) assignment and consequently benefit from the strong phylogenetic content provided by MLST analysis [19].

In recent years several MLST-based studies investigating mastitis [4,17,21] have shown the existence of significant host-specificity of *S. aureus* infection and the existence of CCs associated with mammary gland infections. However, in most studies, only MRSA strains were analyzed which did not provide a complete view of strains infecting the mammary gland.

The aim of the present study was to further investigate the diversity among mastitis-associated MRSA and MSSA from cows, goats and sheep. The additional information provided by MLVA and the analysis of VNTR patterns allowed to better identify the evolution and emergence of clones.

MATERIALS AND METHODS

Bacterial strains and DNA preparation. The 152 isolates investigated in this study were obtained from cases of bovine (n=118), ovine (n=18) and caprine (n=16) clinical or subclinical mastitis (Table S1). Among the bovine isolates, 48 were collected all over Germany between 2006 and 2009, 19 in Southern Brazil in 1992 and 1993, and 51 in Western France in 2008 and 2009. Sixteen among the German isolates were previously described as MRSA ST398 [5]. The 19 brazilian isolates

were described by Lange and colleagues [11]. The ovine and caprine isolates were collected in France between 1978 and 2010 (Table S1)

DNA extraction. Strains were cultured overnight at 37°C in Luria Bertani broth. Genomic DNA was extracted by phenol-chloroform extraction or by using the DNeasy tissue kit (Qiagen, Courtaboeuf, France) with lysostaphin (100 mg/L) (Ambi products LLC, USA). Nucleic acid quality and concentration were analyzed using an ND-1000 spectrophotometer (NanoDrop; Labtech; Palaiseau, France). Diluted samples of 5 ng/µl in water were used as DNA template for PCR amplification.

Data production and analysis. The 16 VNTR loci included in MLVA-16Orsay were amplified in two multiplex PCRs using the ceeramTools® Staphylococcus typing kit (Ceeram, La Chapelle sur Erdre, France) as previously described (Sobral et al., PlosOne, 2012 in press). The typing data file was imported into BioNumerics version 6.6 (Applied-Maths, Sint-Martens-Latem, Belgium).
The UPGMA (unweighted pair group method with arithmetic mean) clustering method was run using the categorical coefficient. A cut-off value of 45% similarity was applied to define clusters. Bootstrap analyses were run using BioNumerics 6.6 with 500 simulations.
The primers and condition used for the spa tandem repeat amplification were as described by Harmsen and colleagues [22]. The amplicons were purified using the QIAquick PCR purification (Qiagen, Courtaboeuf, France) and sequenced (Eurofins MWG Operon). The repeat nomenclature was that of Shopsin and colleagues [16], and spa types were retrieved from the Ridom SpaServer http://spaserver.ridom.de.
The primers and conditions used for PCR were as described at http://saureus.mlst.net/ [15]. PCR products were purified using a QIAquick PCR purification kit (Qiagen, Courtaboeuf, France), as recommended by the manufacturer, and sequenced (Beckman-Coulter Genomics). Alleles and sequence types (STs) were identified using the MLST database (http://www.mlst.net).

RESULTS

Diversity of mastitis-associated *S. aureus* isolates. Using MLVA-16_{Orsay} the 152 ruminant isolates from cases of clinical and subclinical mastitis were resolved into 115 MLVA genotypes with an overall diversity index of 0.9936. The 118 bovine, 18 ovine and 16 caprine isolates belonged respectively to 90, 14 and 15 MLVA genotypes. One hundred and thirty four isolates fell into ten main clusters comprising more than 3 isolates each (Figure 1). The clusters were assigned to known MLST defined clonal complexes (CCs) by comparison with previous MLVA data, *spa* typing and MLST analysis of selected strains (Figure 1 and Figure S1). CC97 and CC133 represented the major CCs as they accounted for 22% (34 isolates) and 21% (32 isolates) of the studied collection, respectively. CC1, CC9, CC20, CC130, CC151, CC398 and CC479 together represent another 45% of the isolates. The 16 ST398 isolates are all MRSA collected in 2009 from different locations in Germany and are distributed into 8 MLVA genotypes and 3 *spa* types.

CC8 and CC30, frequently associated with human *S. aureus* infections (Figure 2), were represented by two isolates each, and CC5, CC7, CC22 and CC25 were represented by single isolates (Figure S1). The four remaining singletons were *spa* typed (Figure S1).

Figure 1 shows the distribution of isolates according to host. The two closely related MLVA clusters associated to CC130 comprised exclusively small ruminant-associated isolates. One cluster included sheep-associated isolates and the other contained goat-associated isolates collected from six different herds in 2004. CC1, CC9, CC20, CC97, CC151, CC398 and CC479 included only bovine isolates. CC133 was the only cluster showing complete host diversity; 13 bovine isolates originating from different collection sites (11% of the bovine isolates), 11 sheep isolates (61% of the ovine isolates) and 8 goat isolates (50% of the caprine isolates) belonged to this CC. The remaining small ruminant-associated isolates were singletons or unclustered.

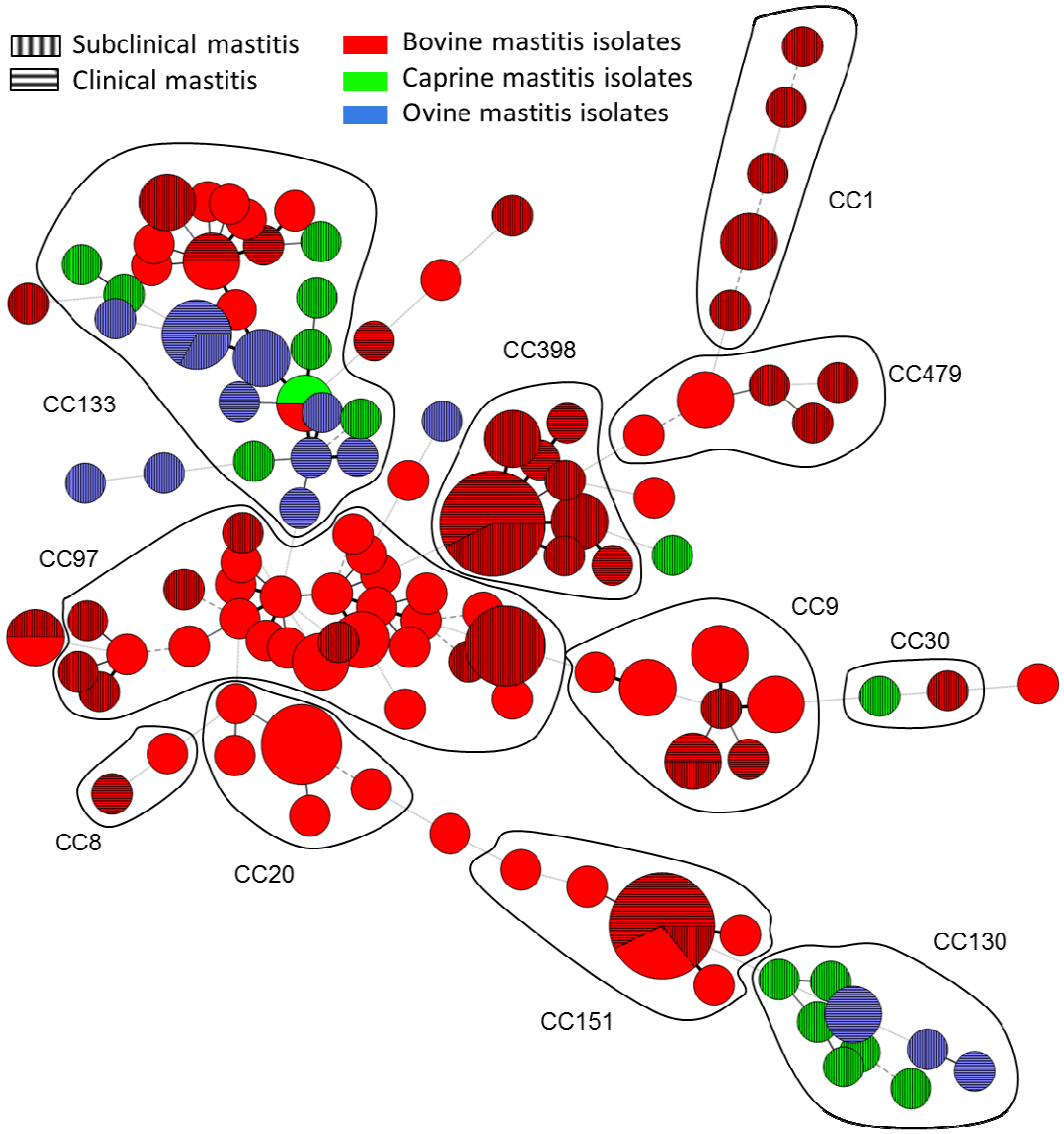


Figure 1. Minimum spanning tree of the 152 *S. aureus* isolates using MLVA-16$_{Orsay}$. Each circle represents a MLVA genotype. The different MLVA clusters are colored according to host. Major CCs are indicated.

In Figure 2, the population structure of *S. aureus* isolates from mastitis was compared to a more global population of *S. aureus* isolates for which MLVA data was available. Some clusters such as CC8, C22 and CC30 are remarkably highly human specific. In contrast, CC133 is exclusively represented by animal isolates, and CC130 by small ruminants mastitis isolates.

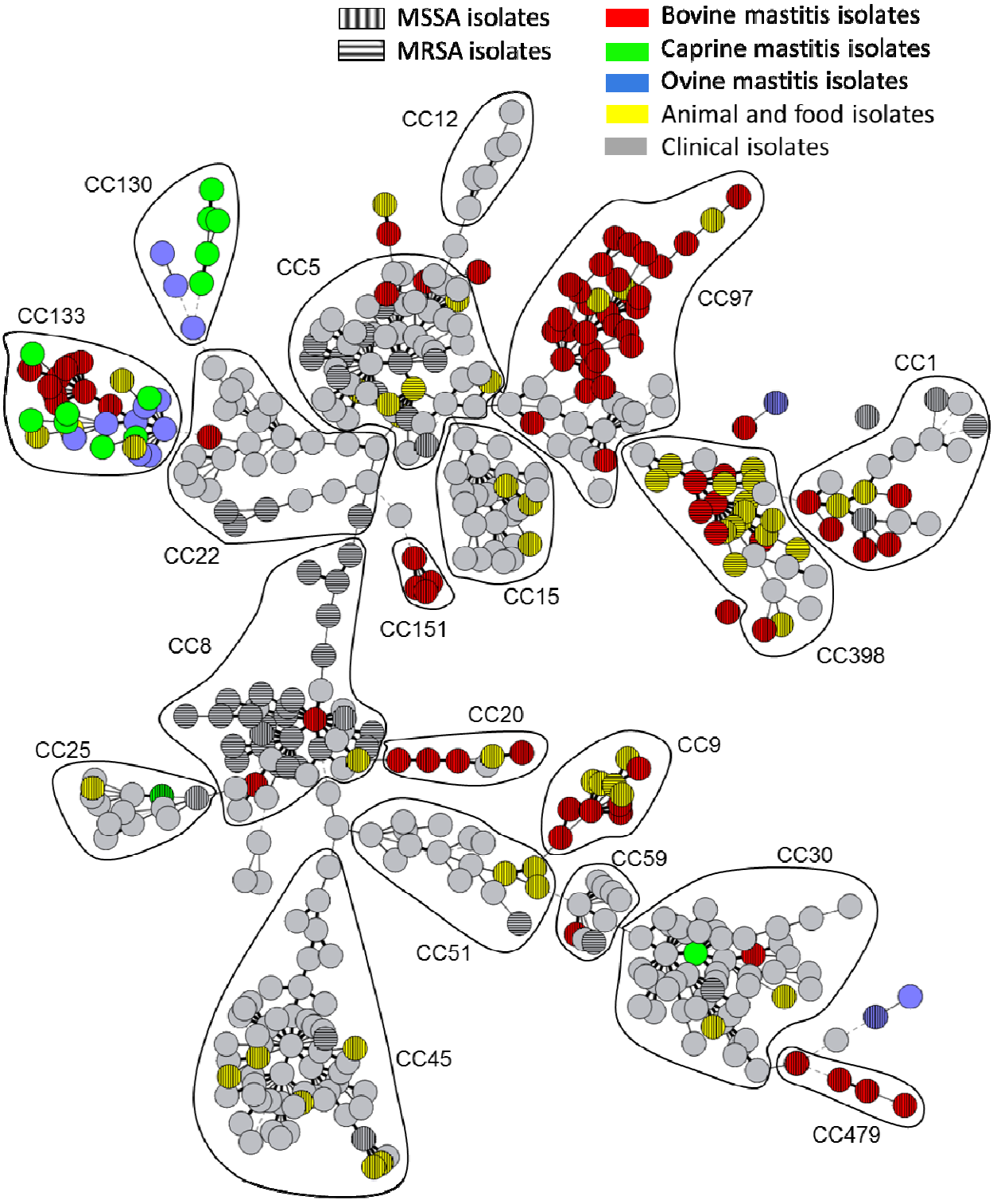


Figure 2. Mastitis isolates superimposed on a background of previously published data from human (grey) and animal (yellow) isolates.

VNTR allele distribution analysis. In order to assess the genetic specificity, based on VNTR pattern analysis, of each lineage (human-predominant [CC1, CC5, CC7, CC8, CC22, CC25, CC30, CC45 and CC51], animal-predominant [CC9, CC20, CC97, CC133 and CC398] and mammary gland-predominant [CC130, CC151 and CC479] [23,24,25,26], we calculated the allelic diversity of each population within each host or tissue related family. Allelic diversity per locus was not significantly different between human-related and animal-specific clusters (Table S2). The observed allelic diversity was considerably reduced in the population of the mammary gland-predominant clusters (Table S2). We also measured the mean number of repetitions per locus and per family. Similarly, no significant difference was observed between the human-predominant clusters and the animal-predominant clusters (Figure S3) in terms of mean number of repetitions per locus. Remarkable differences were noticed when focusing on mammary gland-adapted clusters. This population has a significantly smaller number of repeat units for two VNTRs (Sa0122 and Sa0387) and this observation was most pronounced when considering CC151 (Sa0122, Sa0387, Sa0684, Sa0964, Sa1097, Sa1729, Sa1866 and Sa2511 have smaller number of repeat units in CC151) (Figure S2).

DISCUSSION

The mammary gland, a specialized tissue for infections. In the present collection of isolates, only 13 (8.5%) human-related isolates from 6 different lineages (6 CC1, 2 CC8, 2 CC30, 1 CC5, 1 CC22 and 1 CC25) were observed, thus confirming that the ruminant mammary gland is a very specific niche. Notably, the majority of isolates belong to two major clonal complexes (CC97 and CC133) which represent almost half of the studied collection. MLVA-16Orsay is a very efficient subtytping tool, as 29 and 26 genotypes were observed for a total of 34 CC97 isolates and 31 CC133 isolates respectively. CC133 is commonly sampled from milk produced by small ruminants [23,24] and by cows (15/33 bovine isolates in CC133) [27,28] suffering from mastitis. CC97 is a widespread bovine lineage largely responsible for bovine mastitis cases in Chile, Brazil, Japan and the United States [17,25], and recently also isolated from human [29] and porcine hosts [30]. Our observations corroborate previous findings by Guinane and colleagues who further demonstrated that these lineages shared the pathogenicity island SaPlov2 conferring the ability to coagulate plasma from ruminants in contrast with the tested human or poultry strains [31]. It suggested that these specific *S. aureus* strains were strongly adapted to their host environment.

We identified three additional clusters putatively showing strong preference and adapted to the mammary gland tissue: CC151 (11 isolates), CC130 (10 isolates) and CC479 (6 isolates) which represent 18 % of the present collection. These lineages have almost never been isolated from human beings and, moreover, were exclusively sampled from intramammary infections [28,31,32,33]. Interestingly, whole genome analysis demonstrated that CC130 and CC151 are closely related [31]. This phylogenetic relationship is also detected with MLVA typing. Several studies revealed that only a

few specialized clones are responsible for most of mastitis cases and that these clones have a broad geographical distribution [4,12,14]. In the present study, the CC151-MLVA genotype 2-3-1-2-2-1-9-7-1-0.5-5-2-3-0.5-3-1 was shown to be shared by isolates sampled in Germany and France. Devriese and colleagues had observed specific phenotypic characteristics associated with animal *S. aureus* strains and subsequently emphasized the existence of ecovars adapted to particular host species [34]. Several genetic studies have identified the existence of *S. aureus* genotypes that are associated with cows, sheep, and goats but rarely isolated from humans, suggesting that they are specialized for ruminant hosts [12,14,24,28]. Conversely, a number of CCs show limited host specificity. van Leeuwen and colleagues hypothesized that the specific traits found to be common in bovine strains were related to a tissue rather than to host specificity [35].

Representatives from some animal-predominant clusters have spread recently with apparently neither strong host nor geographical barriers. In this study, we reported 36 isolates belonging to these well-known CCs (16 isolates from CC398, 11 isolates from CC9 and 9 isolates from CC20). ST398 and ST9 are believed to have emerged from their porcine reservoirs and subsequently spread to other hosts. CC20 isolates are frequently described for human carriage [36,37] and infection [38] but are also often sampled from cow mastitic milk [26,39,40].

VNTRs as genetic markers to infer lineage emergence. It has been proposed that a low level of VNTR genetic diversity inside a lineage may reflect recent emergence, and that there exists a tendency toward shortening of tandem repeats array during evolution [41]. Interestingly, we noticed that the mammary gland-predominant lineages (CC130, C151 and CC479) exhibit a smaller repeat unit number per locus and a lower diversity index as compared to other lineages. These observations are most striking in the case of CC151. Modifications in industrial livestock husbandry of dairy ruminants could have led to modified access of microbial flora to mammary glands leading to emergence of some clones such as CC151. Analysis of RF122 strain (ST151) genome sequence provided evidence that this mammary-gland-adapted strain had recently diversified from an ancestor with a supposed human origin through acquisition of mobile genetic elements and gene decay [42,43]. We did not reveal notable differences in VNTR diversity between human and animal lineages arguing for simultaneous colonization of humans and animals. However, because tandem repeat mutation rates have been shown to vary within different lineages (for instance in M. tuberculosis [44], whole genome sequence analysis of well-chosen strains will be necessary to correlate tandem repeats diversity and more neutral genome diversity and evolution in *S. aureus*.

The improvement of our knowledge of mastitis-associated *S. aureus* epidemiology in dairy herds will contribute to the identification of sources and dynamics of spread and transmission of the bacteria. This in turn might help formulate strategies and implement control measures targeted at reservoirs and transmission routes.

Acknowledgements

We thank Fabienne Loisy-Hamon and Benoit Lebeau from Ceeram for their support to this project. This includes the cost of the MLVA typing of all *S. aureus* isolates. The contributions of SS, ATF and KK are financially supported by the German Federal Ministry of Education and Research (BMBF) through the German Aerospace Center (DLR), grant number 01KI1014D (MedVet-Staph). This study also benefited from the support of the association Vaincre La Mucoviscidose (Grant N° RC0630). The development of tools for the surveillance of bacterial pathogens is supported by the French Direction Générale de l'Armement. We thank Roswitha Becker and Vivian Hensel for excellent technical assistance.

Conflict of interest

D.S. is employee of Ceeram and holds stocks. Patent licensing arrangements involving D.S., and C.P. exist with Ceeram.

References

1. Wilson DJ, Gonzalez RN, Das HH (1997) Bovine mastitis pathogens in New York and Pennsylvania: prevalence and effects on somatic cell count and milk production. J Dairy Sci 80: 2592-2598.

2. Bergonier D, de Cremoux R, Rupp R, Lagriffoul G, Berthelot X (2003) Mastitis of dairy small ruminants. Vet Res 34: 689-716.

3. de Almeida LM, de Almeida MZ, de Mendonca CL, Mamizuka EM (2011) Novel sequence types (STs) of *Staphylococcus aureus* isolates causing clinical and subclinical mastitis in flocks of sheep in the northeast of Brazil. J Dairy Res 78: 373-378.

4. Mørk T, Tollersrud T, Kvitle B, Jørgensen HJ, Waage S (2005) Comparison of *Staphylococcus aureus* genotypes recovered from cases of bovine, ovine, and caprine mastitis. J Clin Microbiol 43: 3979-3984.

5. Feßler A, Scott C, Kadlec K, Ehricht R, Monecke S, et al. (2010) Characterization of methicillin-resistant *Staphylococcus aureus* ST398 from cases of bovine mastitis. J Antimicrob Chemother 65: 619-625.

6. Holmes MA, Zadoks RN (2011) Methicillin resistant *S. aureus* in human and bovine mastitis. J Mammary Gland Biol Neoplasia 16: 373-382.

7. Juhasz-Kaszanyitzky E, Janosi S, Somogyi P, Dan A, van der Graaf-van Bloois L, et al. (2007) MRSA transmission between cows and humans. Emerg Infect Dis 13: 630-632.

8. Monecke S, Kuhnert P, Hotzel H, Slickers P, Ehricht R (2007) Microarray based study on virulence-associated genes and resistance determinants of *Staphylococcus aureus* isolates from cattle. Vet Microbiol 125: 128-140.

9. Turkyilmaz S, Tekbiyik S, Oryasin E, Bozdogan B (2010) Molecular epidemiology and antimicrobial resistance mechanisms of methicillin-resistant *Staphylococcus aureus* isolated from bovine milk. Zoonoses Public Health 57: 197-203.

10. Annemuller C, Lammler C, Zschock M (1999) Genotyping of *Staphylococcus aureus* isolated from bovine mastitis. Vet Microbiol 69: 217-224.

11. Lange C, Cardoso M, Senczek D, Schwarz S (1999) Molecular subtyping of *Staphylococcus aureus* isolates from cases of bovine mastitis in Brazil. Vet Microbiol 67: 127-141.

12. Fitzgerald JR, Meaney WJ, Hartigan PJ, Smyth CJ, Kapur V (1997) Fine-structure molecular epidemiological analysis of *Staphylococcus aureus* recovered from cows. Epidemiol Infect 119: 261-269.

13. Larsen HD, Sloth KH, Elsberg C, Enevoldsen C, Pedersen LH, et al. (2000) The dynamics of *Staphylococcus aureus* intramammary infection in nine Danish dairy herds. Vet Microbiol 71: 89-101.

14. Kapur V, Sischo WM, Greer RS, Whittam TS, Musser JM (1995) Molecular population genetic analysis of *Staphylococcus aureus* recovered from cows. J Clin Microbiol 33: 376-380.

15. Enright MC, Day NP, Davies CE, Peacock SJ, Spratt BG (2000) Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. J Clin Microbiol 38: 1008-1015.

16. Shopsin B, Gomez M, Montgomery SO, Smith DH, Waddington M, et al. (1999) Evaluation of protein A gene polymorphic region DNA sequencing for typing of *Staphylococcus aureus* strains. J Clin Microbiol 37: 3556-3563.

17. Smith EM, Green LE, Medley GF, Bird HE, Fox LK, et al. (2005) Multilocus sequence typing of intercontinental bovine *Staphylococcus aureus* isolates. J Clin Microbiol 43: 4737-4743.

18. Hardy KJ, Oppenheim BA, Gossain S, Gao F, Hawkey PM (2006) Use of variations in staphylococcal interspersed repeat units for molecular typing of methicillin-resistant *Staphylococcus aureus* strains. J Clin Microbiol 44: 271-273.

19. Pourcel C, Hormigos K, Onteniente L, Sakwinska O, Deurenberg RH, et al. (2009) Improved multiple-locus variable-number tandem-repeat assay for *Staphylococcus aureus* genotyping, providing a highly informative technique together with strong phylogenetic value. J Clin Microbiol 47: 3121-3128.

20. Schouls LM, Spalburg EC, van Luit M, Huijsdens XW, Pluister GN, et al. (2009) Multiple-locus variable number tandem repeat analysis of *Staphylococcus aureus*: comparison with pulsed-field gel electrophoresis and spa-typing. PLoS One 4: e5082.

21. Kwon NH, Park KT, Moon JS, Jung WK, Kim SH, et al. (2005) Staphylococcal cassette chromosome mec (SCCmec) characterization and molecular analysis for methicillin-resistant *Staphylococcus aureus* and novel SCCmec subtype IVg isolated from bovine milk in Korea. J Antimicrob Chemother 56: 624-632.

22. Harmsen D, Claus H, Witte W, Rothganger J, Turnwald D, et al. (2003) Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for spa repeat determination and database management. J Clin Microbiol 41: 5442-5448.

23. Aires-de-Sousa M, Parente CE, Vieira-da-Motta O, Bonna IC, Silva DA, et al. (2007) Characterization of *Staphylococcus aureus* isolates from buffalo, bovine, ovine, and caprine milk samples collected in Rio de Janeiro State, Brazil. Appl Environ Microbiol 73: 3845-3849.

24. Jørgensen HJ, Mørk T, Caugant DA, Kearns A, Rorvik LM (2005) Genetic variation among *Staphylococcus aureus* strains from Norwegian bulk milk. Appl Environ Microbiol 71: 8352-8361.

25. Rabello RF, Moreira BM, Lopes RM, Teixeira LM, Riley LW, et al. (2007) Multilocus sequence typing of *Staphylococcus aureus* isolates recovered from cows with mastitis in Brazilian dairy herds. J Med Microbiol 56: 1505-1511.

26. Sakwinska O, Giddey M, Moreillon M, Morisset D, Waldvogel A, et al. (2011) *Staphylococcus aureus* host range and human-bovine host shift. Appl Environ Microbiol 77: 5908-5915.

27. Ben Zakour NL, Sturdevant DE, Even S, Guinane CM, Barbey C, et al. (2008) Genome-wide analysis of ruminant *Staphylococcus aureus* reveals diversification of the core genome. J Bacteriol 190: 6302-6317.

28. Smyth DS, Feil EJ, Meaney WJ, Hartigan PJ, Tollersrud T, et al. (2009) Molecular genetic typing reveals further insights into the diversity of animal-associated *Staphylococcus aureus*. J Med Microbiol 58: 1343-1353.

29. Lozano C, Gomez-Sanz E, Benito D, Aspiroz C, Zarazaga M, et al. (2011) *Staphylococcus aureus* nasal carriage, virulence traits, antibiotic resistance mechanisms, and genetic lineages in healthy humans in Spain, with detection of CC398 and CC97 strains. Int J Med Microbiol 301: 500-505.

30. Battisti A, Franco A, Merialdi G, Hasman H, Iurescia M, et al. (2010) Heterogeneity among methicillin-resistant *Staphylococcus aureus* from Italian pig finishing holdings. Vet Microbiol 142: 361-366.

31. Guinane CM, Ben Zakour NL, Tormo-Mas MA, Weinert LA, Lowder BV, et al. (2010) Evolutionary genomics of *Staphylococcus aureus* reveals insights into the origin and molecular basis of ruminant host adaptation. Genome Biol Evol 2: 454-466.

32. Ikawaty R, Willems RJ, Box AT, Verhoef J, Fluit AC (2008) Novel multiple-locus variable-number tandem-repeat analysis method for rapid molecular typing of human *Staphylococcus aureus*. J Clin Microbiol 46: 3147-3151.

33. Sung JM, Lloyd DH, Lindsay JA (2008) *Staphylococcus aureus* host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. Microbiology 154: 1949-1959.

34. Devriese LA (1984) A simplified system for biotyping *Staphylococcus aureus* strains isolated from animal species. J Appl Bacteriol 56: 215-220.

35. van Leeuwen WB, Melles DC, Alaidan A, Al-Ahdal M, Boelens HA, et al. (2005) Host- and tissue-specific pathogenic traits of *Staphylococcus aureus*. J Bacteriol 187: 4584-4591.

36. Ko KS, Lee JY, Baek JY, Peck KR, Rhee JY, et al. (2008) Characterization of *Staphylococcus aureus* nasal carriage from children attending an outpatient clinic in Seoul, Korea. Microb Drug Resist 14: 37-44.

37. Ruimy R, Armand-Lefevre L, Barbier F, Ruppe E, Cocojaru R, et al. (2009) Comparisons between geographically diverse samples of carried *Staphylococcus aureus*. J Bacteriol 191: 5577-5583.

38. Grundmann H, Aanensen DM, van den Wijngaard CC, Spratt BG, Harmsen D, et al. (2010) Geographic distribution of *Staphylococcus aureus* causing invasive infections in Europe: a molecular-epidemiological analysis. PLoS Med 7: e1000215.

39. Bystron J, Podkowik M, Korzekwa K, Lis E, Molenda J, et al. (2010) Characterization of borderline oxacillin-resistant *Staphylococcus aureus* isolated from food of animal origin. J Food Prot 73: 1325-1327.

40. Hata E, Katsuda K, Kobayashi H, Uchida I, Tanaka K, et al. (2010) Genetic variation among *Staphylococcus aureus* strains from bovine milk and their relevance to methicillin-resistant isolates from humans. J Clin Microbiol 48: 2130-2139.

41. Arnold C, Thorne N, Underwood A, Baster K, Gharbia S (2006) Evolution of short sequence repeats in Mycobacterium tuberculosis. FEMS Microbiol Lett 256: 340-346.

42. Herron LL, Chakravarty R, Dwan C, Fitzgerald JR, Musser JM, et al. (2002) Genome sequence survey identifies unique sequences and key virulence genes with unusual rates of amino Acid substitution in bovine *Staphylococcus aureus*. Infect Immun 70: 3978-3981.

43. Herron-Olson L, Fitzgerald JR, Musser JM, Kapur V (2007) Molecular correlates of host specialization in *Staphylococcus aureus*. PLoS One 2: e1120.

44. Comas I, Homolka S, Niemann S, Gagneux S (2009) Genotyping of genetically monomorphic bacteria: DNA sequencing in Mycobacterium tuberculosis highlights the limitations of current methodologies. PLoS One 4: e7815.

3. Conclusion

Les trois premiers articles décrivent le développement et la validation de procédés automatisés de génotypage par MLVA. La comparaison avec les autres méthodes de référence, la PFGE et le MLST, atteste de la puissance du MLVA. Les efforts de standardisation et d'automatisation entrepris permettent de répondre aux nouvelles exigences des épidémiologistes. Le typage entre dans l'ère du haut-débit avec les nouveaux produits développés au cours de cette thèse. Ces articles détaillent l'analyse de petites collections de souches mais ce qui était de l'ordre de l'inenvisageable devient alors désormais possible : génotyper restrospectivement des souchiers entiers de plusieurs milliers de souches est ainsi concevable dans un délai raisonnable. Enfin, ces études explorent la diversité des populations bactériennes et modélisent leur structuration. Il apparait que, quelle que soit la structure de l'espèce, l'adaptation à un nouvel environnement (hôte animal, niche écologique, organe) est l'occasion d'expansions clonales. Ce point sera discuté plus tard. Enfin, le quatrième article s'attache davantage à apprécier les VNTRs en tant qu'outils phylogénétiques. L'homoplasie de certains marqueurs est avérée. L'émergence d'un complexe clonal est proposée sur la base des analyses de données de VNTRs.

Discussion et Perspectives

1. Vers le génotypage de routine

1.1. Automatisation et standardisation des protocoles MLVA

Avec les protocoles développés au cours de mes travaux de thèse, la technologie du MLVA a passé un cap pour trois espèces bactériennes. Désormais, des kits de génotypage sont disponibles et permettent de réaliser des analyses génotypiques de grande ampleur. L'utilisateur n'a plus à jongler avec plusieurs dizaines d'amorces, à multiplier les réactions PCR et couler des dizaines de gels. Le kit est un outil clé en main : le typage d'une centaine de souches, de l'extrait d'ADN au code MLVA, nécessite moins d'une heure de temps de manipulation et pas plus de 48 heures de temps machine selon le type de séquenceur utilisé (quelques heures sur une machine à 96 capillaires). La manipulation du kit en laboratoire ne nécessite aucune expertise particulière, seule l'analyse des données et l'obtention des codes numériques requièrent une certaine expertise. Cependant, un guide de l'utilisateur accompagne le kit et a été rédigé de telle sorte que l'utilisateur puisse obtenir, de manière automatique, les codes numériques en suivant des étapes détaillées pas à pas. Un guide de dépannage sous forme d'arbres décisionnels facilite l'interprétation des données manquantes ou artéfactuelles et guide l'utilisateur vers l'action adéquate à entreprendre.

Un effort de standardisation conséquent a été réalisé. Les VNTRs ont été nommés selon la nomenclature définie dans l'introduction ; cette appellation rassemble toutes les informations nécessaires pour l'identification et la caractérisation du marqueur utilisé. Puis, le codage des allèles est défini selon les recommandations avancées par G. Vergnaud et C. Pourcel [134]. Pour chaque protocole développé, le panel d'allèles a été construit à partir d'une collection représentative de la diversité de l'espèce d'intérêt et enrichi avec les nouveaux allèles identifiés lors des analyses décrites dans les articles. L'ensemble des allèles répertoriés dans la littérature et les différentes bases de données ont été pris en compte pour l'élaboration des panels de multiplexage, autorisant, de surcroît, une marge de manœuvre d'au moins trois allèles, encore non décrits. Afin de contourner l'effet éventuel de modifications de type indel au niveau des zones flanquantes, chaque allèle est associé à un intervalle de confiance empiriquement fixé à 10% de la taille du motif répété. Enfin, le logiciel d'assignation des allèles attribue un indice de confiance à chaque allèle et un score global de qualité est attribué au code numérique final. Cette démarche contribue à la standardisation de la méthodologie afin que l'utilisateur puisse apprécier la validité du typage.

L'automatisation et la standardisation du MLVA sont les prérequis pour son utilisation en routine. Dans le cadre d'études rétrospectives, le typage de souchiers archivant plusieurs milliers d'isolats est dorénavant possible. La connaissance et le suivi de la diversité des populations bactériennes dans plusieurs niches (différentes zones géographiques, réservoir, et/ou organes) et sur plusieurs années sont de véritables indicateurs d'évolution et de puissants outils de prédiction épidémiologiques. La distribution et la prévalence de certains complexes clonaux sont évaluées et

orientent l'épidémiologiste dans son diagnostic. L'industrialisation du MLVA et la mise à disposition de kits ouvrent la voie vers le typage haut-débit mais sont également le moyen de réduire les coûts de manière drastique. L'objectif de cette thèse n'est pas de justifier ni de définir un coût exact des kits de typage ceeramTools, cependant la diminution d'un facteur 10 de celui-ci par rapport au MLST est notable et correspond au prix actuellement proposé par la société Ceeram. De nouvelles optimisations et la production de kits à grande échelle permettront, sans aucun doute et dans les prochains mois, de réduire ces coûts.

Cette avancée technologique n'est valable qu'en raison de l'informativité des VNTRs et de la puissance du MLVA comme outil épidémiologique. Cet intérêt avait préalablement été démontré dans différentes études [64,76,78,140] et nos travaux le confirment.

1.2. Le MLVA, un outil épidémiologique puissant

Pour les trois espèces étudiées, le panel de marqueurs choisi confère une discrimination au moins aussi élevée que celle apportée par la méthode de référence, la PFGE. La stabilité des loci a été évaluée *in vivo* pour deux protocoles : un même génotype de *P. aeruginosa* est persistant dans les bronches d'un patient atteint de mucoviscidose pendant près de huit ans, et, le même génotype de *L. pneumophila* colonise les réseaux d'eau de la ville de Rennes pendant neuf ans. La discrimination élevée ajoutée à la stabilité des VNTRs expliquent la concordance épidémiologique observée par MLVA. Même si l'instabilité des séquences répétées en tandem et leur rôle dans la physiologie bactérienne sont méconnus, il s'avère qu'expérimentalement les VNTRs sélectionnés sont d'excellents marqueurs épidémiologiques : ils sont polymorphes mais ne varient pas trop vite. La révélation de leur polymorphisme par simple PCR contribue à la reproductibilité et à la robustesse du MLVA.

Quelques éléments de discussion plus fondamentaux sur la structuration des populations bactériennes seront abordés dans les parties suivantes, cependant il est temps de noter d'ores et déjà la pertinence du MLVA comme outil de structuration. En effet, le MLVA permet de classer les isolats et de les assigner à des complexes clonaux bien définis. De manière remarquable, ces groupes sont identifiés par des profils caractéristiques qualifiés de signatures. Ainsi, quelques VNTRs permettent un branchement profond des principaux complexes clonaux dans le cas d'espèces à structure clonale, et les caractérisent de manière très spécifique. Chez *S. aureus*, le VNTR Sa1097 permet presque à lui seul d'éclater une population en ses différents complexes clonaux : l'allèle 12U est propre au CC45, l'allèle 13U est quasi-exclusivement identifié chez le CC8 et l'allèle 15U est spécifique du CC398. De même la spécificité et l'exclusivité de certains allèles ont été mises en avant chez *M. tuberculosis* [151,152]. Cependant, l'utilisation d'un seul marqueur pour prédire l'assignation à un clone n'est pas envisageable en raison de l'homoplasie propre aux répétitions en tandem. La définition des différentes lignées d'une espèce peut être réalisée avec robustesse par la combinaison de plusieurs marqueurs d'ancrage rassemblés en un panel minimal. Par exemple, la combinaison de dix VNTRs parmi les

seize utilisés chez *S. aureus* suffit pour démontrer la diversité clonale d'une population. Cette faculté de sous-classification fait du MLVA un puissant outil d'agrégation ou *clustering* c'est-à-dire de constitution de groupes homogènes au sein d'une population.

2. MLVA et agrégation

2.1. L'agrégation

Les études réalisées au cours de cette thèse soulignent l'importance de la maîtrise d'outils d'agrégation en épidémiologie. Dans chaque étude, un des objectifs importants a été de regrouper les génotypes et d'identifier des sous-populations : c'est l'agrégation. Contrairement à la phylogénie, l'agrégation ne vise pas à hiérarchiser une population ou à établir une relation de parenté entre ses membres, mais tente d'organiser ou partitionner cette population en sous-ensembles homogènes. Cette classification en grappes relève de la méthodologie phénétique dont le postulat de base stipule que le degré de ressemblance est corrélé au degré de parenté. Contrairement à l'école cladistique partisane de la hiérarchisation des populations selon l'identification de caractères et le regroupement sur la base d'un état de caractère dérivé, la phénétique suppose de quantifier la ressemblance entre les êtres vivants à classer. L'agrégation par MLVA découle directement de ce procédé et permet la constitution de grappes selon la similitude des codes MLVA *i.e.* le nombre de VNTRs communs. L'agrégation n'est par définition possible que lorsque des groupes d'isolats apparentés existent et par extension lorsque la population est clonale ou mixte. La structure d'une population peut être déterminée par le calcul de l'indice de déséquilibre de liaison des marqueurs disponibles au sein de la population étudiée. Cet indice a été calculé à partir de données MLVA sur quelques bactéries pathogènes dont la structure clonale avait été précédemment démontrée par MLEE ou MLST : *M. tuberculosis* [153], *X. citri* [154], *B. hyodysenteriae* [155], *F. noatunensis* [156] et *V. parahaemolyticus* [157]. Les VNTRs analysés sont dans ces cas en déséquilibre de liaison confirmant ainsi la clonalité des espèces testées. Cependant, compte tenu de l'hétérogénéité des marqueurs utilisés et des populations étudiées, il apparaît périlleux de comparer les indices obtenus. Néanmoins, excepté l'étude concernant *B. hyodysenteriae* (I^S_A = 0.038), l'ordre de grandeur de cet indice est remarquablement conservé (de 0.1345 à 0.59) alors qu'il varie davantage à partir de données MLEE ou MLST [153]. Les VNTRs s'avèrent des marqueurs structurants et permettent l'assignation des différents génotypes à des complexes clonaux connus.

2.2. Méthodes d'agrégation

a. Méthodes de distances : UPGMA et Neighbor-Joining

Le résultat du MLVA est un code numérique comportant autant de nombres que de marqueurs étudiés. Dans une première approche, chaque VNTR est pondéré également et tous les allèles d'un même marqueur sont équidistants les uns des autres. L'utilisation de méthodes permettant la construction d'arbres à partir de ressemblances globales est donc préférée. Le calcul des distances entre les différents génotypes MLVA s'effectue donc par la comparaison et l'évaluation des similitudes et différences observées pour chaque caractère *i.e.* le locus VNTR. La distance la plus communément utilisée est la distance de Hamming, ou catégorique. La similarité S entre deux génotypes est égale au nombre de VNTRs portant le même allèle (M) divisé par le nombre total de marqueurs étudiés (L). Par extension, la distance observée D entre ces deux génotypes comparés est D = 1 – S avec S = M/L. Après calcul de la matrice de distance entre les différents génotypes, les programmes de construction d'arbre procèdent par regroupements successifs depuis la paire des séquences les plus proches aux plus éloignées. Le phénogramme produit est un arbre dans lequel la longueur des branches est corrélée à la distance génétique et représente donc le degré de parenté entre les taxons étudiés. Les extrémités des branches définissent des Unités Taxonomiques Opérationnelles (UTO), groupes d'individus indifférenciés avec les caractères considérés. Les principales méthodes d'agrégation fondées sur des données de distance sont l'UPGMA (*Unweighted Pair Group Method using Averages*) et le *Neighbor Joining*. La méthode agglomérative UPGMA impose que les distances soient ultra-paramétriques et présuppose donc que l'évolution suive le postulat de l'horloge moléculaire [158]. La méthode additive *Neighbor Joining* définit des longueurs des branches telles que les distances déduites de l'arbre soient les plus proches de distances mesurées entre les génotypes [159]. Le *Neighbor Joining* à la différence de la méthode UPGMA autorise un taux de mutation différent sur les branches. Selon la méthode utilisée, l'arbre tracé à partir d'un même jeu de données peut avoir des topologies très différentes. Ainsi, des méthodes statistiques existent pour vérifier la robustesse des arbres. Une approche très répandue pour mesurer la solidité d'une phylogénie consiste à mesurer celle de chacun de ses nœuds par la technique dite de bootstrap [160]. Ce test permet de ré-échantillonner les caractères initiaux, selon un tirage au sort avec remise, introduit dans les programmes de calcul en construisant autant d'arbres que le nombre de ré-échantillonnages demandés. Un arbre consensus est alors calculé indiquant sur chaque nœud de branches le pourcentage d'arbre pour lequel ce même nœud est retrouvé.

b. Arbre de recouvrement minimal et principe de parcimonie

Concept introduit par Edwards et Cavalli-Sforza en 1963, l'application du principe de parcimonie à la phylogénie est à l'origine de l'utilisation du *Minimum Spanning Tree* (MST) ou arbre

de recouvrement minimal. Ainsi, selon Edwards et Cavalli-Sforza, « *l'arbre évolutif à préférer est celui qui invoque la quantité minimum d'évolution* » *i.e.* celui qui implique le moins d'évènements évolutifs. En pratique, l'arbre de recouvrement est celui dont la somme des longueurs des branches est minimale. Ces arbres réticulés définissent des génotypes qualifiés de fondateurs à partir desquels des variants auraient divergé. De plus, cette méthode de *clustering* est beaucoup plus robuste et stable que celles fondées sur l'UPGMA ou le *Neighbor Joining*. Selon la taille de la population, certains génotypes vont être assignés à un agrégat plutôt qu'à un autre. L'arbre de recouvrement minimal n'est pas influencé par la taille de la population analysée ou par l'ajout incrémentiel de nouveaux génotypes : le principe de parcimonie sous-jacent permet la conservation des points de branchements des différents clusters. Les arbres de recouvrement minimal sont graphiquement très informatifs ; il est possible de prédire le fondateur à partir duquel un complexe clonal aurait divergé.

2.3. Identification des complexes clonaux

Les méthodes de distance sont excellentes pour effectuer des analyses d'agrégation ; les génotypes similaires sont regroupés et peuvent former des complexes clonaux pertinents. Comment ceux-ci sont-ils identifiés ? Dans mon étude sur la diversité de l'espèce *S. aureus* chez les animaux, la majorité des profils MLVA étaient nouveaux. J'ai donc dû génotyper quelques membres de ces groupes avec d'autres méthodes comme le typage *spa* ou le MLST pour assigner certains isolats à un complexe clonal. De prime abord, le MLVA seul ne peut être utilisé pour définir les agrégats au sein d'une population. La concordance avec une autre technique telle que le MLST, technique de référence pour la définition de complexes clonaux, est essentielle pour la mise au point et la calibration initiale du MLVA (sélection des marqueurs et de la méthode d'agrégation).

Les complexes clonaux sont définis par l'établissement d'un seuil de similarité ou *cut-off* permettant de rassembler au sein d'un même groupe homogène des génotypes similaires. Ces groupes sont alors qualifiés de clonaux car les allèles des marqueurs étudiés ne sont pas distribués de manière aléatoire au sein de la population et définissent une entité homogène. Le MLST regroupe en complexes clonaux les génotypes qui présentent cinq allèles en commun sur les sept gènes analysés. Selon l'espèce étudiée et le nombre de VNTRs utilisés, le seuil de similarité permettant de définir les complexes clonaux par MLVA peut varier de 45 à 60%. Cette variabilité s'explique par l'hétérogénéité de la vitesse de mutation des répétitions en tandem et par la diversité génétique de l'espèce considérée. La pertinence du MLVA sera démontrée si les agrégats observés sont semblables à ceux identifiés par MLST. Malgré une nécessaire mise au point des paramètres d'agrégation pour chaque nouveau protocole développé, le MLVA s'avère un excellent outil d'identification de complexes clonaux, par exemple grâce aux VNTRs signatures qui se sont maintenus dans les différentes lignées au cours de l'évolution. La compréhension des mécanismes d'évolution des VNTRs pourrait permettre de trouver une règle qui définisse le seuil de similarité à utiliser en fonction

des marqueurs et de leur nombre. Je m'intéresserai donc dans la partie suivante aux différents aspects concernant l'évolution des VNTRs.

3. Evolution des VNTRs et intérêt en épidémiologie

3.1. Problématique

Les méthodes de distance et de parcimonie permettent d'intégrer des modèles de changements évolutifs des marqueurs utilisés. La connaissance de la dynamique d'évolution des VNTRs pourrait ainsi être un atout puissant pour affiner les relations de parentés entre isolats et mieux définir les complexes clonaux. Aujourd'hui, ces données sont traitées comme une chaîne alpha-numérique, les caractères homologues *i.e.* les allèles d'un même VNTR ont le même poids et sont équidistants. En somme, la distance de Hamming donne le même poids à la perte d'un motif répété ou à l'acquisition de neuf. Les mécanismes et probabilités sous-jacentes sont, sans aucun doute, bien différents selon l'évènement mutationnel. L'étude de l'évolution des VNTRs apparaît donc comme un enjeu majeur en épidémiologie d'intervention : le nombre d'évènements de mutation ne sera plus le seul à pouvoir différencier les génotypes, la probabilité de ces évènements interviendra également.

Dans l'étude de suivi des populations de *L. pneumophila* dans les réseaux d'eau rennais, j'ai analysé les profils MLVA des différents variants des fondateurs identifiés par MST et j'ai observé que, parmi les dix variants, neuf résultent d'une mutation sur un seul marqueur dont sept caractérisés par l'acquisition ou la perte d'un seul motif répété soit 78% de mutation simple pas. Cette observation *a posteriori* suggère que les évènements de mutation eux-mêmes sont principalement de ce type. Traduit-elle une réalité biologique ? Si oui a-t-elle été testée expérimentalement ? Afin de valider cette observation, j'ai initié une étude PSPE afin d'évaluer l'instabilité des VNTRs chez *S. aureus*. Cette espèce a été choisie en raison de sa facilité de culture. Sur près de 100.000 générations obtenues, je n'ai observé qu'un seul évènement de mutation : une délétion d'un seul motif répété pour le Sa0122. Cette observation corrobore *a priori* le modèle prédit mais ne peut en aucun cas le valider. Nous verrons que plusieurs modèles ont été proposés et comment ils sont utilisés en épidémiologie. Ces modèles n'ont pas été considérés dans le cadre de mes travaux de thèse mais constituent une perspective majeure de l'utilisation des données de MLVA.

3.2. Modèle évolutif

a. Polarisation

L'évolution des VNTRs est-elle polarisée ? Les variations du nombre de copies répétées des loci VNTRs résultent de mutations de type indel (insertions/délétions) traduisant l'acquisition ou la perte d'une ou plusieurs unités. L'évaluation *in vivo* des taux de mutation de mini et microsatellites

suggère que les acquisitions et délétions surviennent à fréquence égale [104,161]. L'évolution des VNTRs n'est donc pas polarisée. Si je considère uniquement les mutations simple pas, l'apolarité de l'évolution des VNTRs est confirmée chez *L. pneumophila* par le suivi des VACC1, VACC2 et VACC6 (trois acquisitions et quatre délétions). La prédiction du sens d'évolution d'une répétition en tandem n'est pas envisageable *a priori*, seule l'identification du brin mésapparié permet d'évaluer le décalage produit *a posteriori* : le mésappariement du brin néo-synthétisé conduirait à l'addition d'un motif alors que la formation d'une boucle sur le brin matrice engendrerait la perte de motifs [162]. Un autre modèle d'évolution directionnelle suggère que la position du site de décrochage de la polymérase détermine l'acquisition ou la perte de motifs : le décrochage en position 3' de la structure répétée induit un mésappariement du brin néo-synthétisé entraînant une délétion, et inversement un saut de la polymérase en position 5' de la séquence satellite promouvrait les phénomènes d'expansions [163]. En termes mécaniques, les évènements d'insertion et de délétion apparaissent donc de manière équiprobable.

b. Proposition de modèles mathématiques

i) Modèle probabiliste par GSMM

Les mutations génétiques sont des évènements rares, stochastiques et suivent de fait une distribution de Poisson, ou loi des petites probabilités. La loi de Poisson donne la fréquence d'apparition au hasard d'une mutation, *i.e.* la probabilité de distribution d'évènements aléatoires, dans une période donnée. L'occurrence d'évènements d'indels pour les loci VNTRs suit une loi de Poisson. Soit λ la fréquence d'apparition d'un évènement de mutation dans un laps de temps Δ, alors la probabilité qu'il existe exactement *k* évènements est pour tout entier naturel k :

$p(k) = e^{-\lambda}\frac{\lambda^k}{k!}$ avec $\lambda = \Delta \cdot r$ (r étant le taux de mutation pour un intervalle de temps)

Le taux de mutation par génération étant connu par les études PSPE, la distribution des évènements d'indels pour les loci VNTR est donc statistiquement mesurable [139]. L'évaluation expérimentale des taux de mutation couplée à l'approche probabiliste contribue à renforcer les phylogénies inférées à partir de données VNTR et à augmenter leur robustesse en terme d'évolution. Cependant, ce modèle ne se fonde que sur une hypothèse simpliste : le SSMM (*Single Stepwise Mutation Model*) qui considère l'évolution des VNTRs par simple pas *i.e.* acquisition ou perte d'un seul motif. De nombreux travaux rapportent que les indels d'une seule unité sont les plus fréquentes aux loci de type microsatellite mais ne sont pas exclusives [164]. En effet, lors du suivi de populations de *P. aeruginosa* dans les poumons de patients atteints de mucoviscidose, j'ai constaté que le microsatellite ms61 avait muté quatre fois. Sur ces évènements de mutation, deux résultent d'une indel de quatre motifs répétés suggérant l'implication de mécanismes de recombinaison dans l'évolution de ce microsatellite. Cette proposition a été validée expérimentalement : le microsatellite $(GAATTC)_n$ de *E.*

coli est instable lorsque le gène *recA* est délété, ce microsatellite est donc soumis aux systèmes de recombinaisons recA-dépendant [165].

Vogler et collaborateurs ont proposé un modèle dynamique d'évolutions des VNTRs selon un GSMM (*General Stepwise Mutation Model*) qui considère les évolutions multiples [104]. L'étude *in vitro* de l'instabilité des VNTRs montre que, approximativement, 80% des évènements de mutation sont des indels simples, les évènements d'insertions et de délétions étant équiprobables. Les 20% restants sont des évènements multiples pouvant être aussi bien des insertions que des délétions. Même si les 12 VNTRs utilisés pour le génotypage de *L. pneumophila* ne sont pas tous de type microsatellite, l'analyse des variants corrèle les précédentes observations : 78% d'indels simple, 22 d'évènements multiples (une délétion de trois unités et une de quatre). Ces évènements multiples suivent une loi de distribution géométrique, les indels de deux unités étant plus fréquentes que ceux de trois, etc. Le modèle proposé par Vogler et collaborateurs s'apparente à celui admis pour l'évaluation de la dynamique des VNTRs chez l'Homme [166]. Soit n le pas de mutation, entier naturel supérieur ou égal à 2, et P la probabilité de mutation par une simple indel, alors la probabilité pour qu'un VNTR mute d'un pas n est égale à : $P(X = n) = (1-P)^{n-1}$. La variable aléatoire X, pas de mutation, suit une distribution géométrique (g_j) de variance σ^2_g. Ce modèle théorique a été vérifié et validé expérimentalement pour plusieurs espèces bactériennes telles que *E. coli*, *Y. pestis* ou *B. pseudomallei* [41,104,109,161]. Ce modèle est en adéquation avec le mode évolutif par SMM ou bégaiement de la polymérase. La distribution géométrique proposée est donc principalement adaptée à la dynamique des microsatellites, et, par extension, aux petits minisatellites faiblement répétés. Paradoxalement, lors de mes différentes études, les microsatellites ne suivaient pas ce modèle (Lpms39 de *L. pneumophila* et ms61 de *P. aeruginosa* évoluent quasi-exclusivement par indels multiples) alors que les minisatellites de petite taille le corroborent (mutations par simple pas de Lpms35 et Lpms13 de *L. pneumophila* et ms127 de *P. aeruginosa*). Les mutations au niveau d'un grand minisatellite n'ont été observées que deux fois (Lpms31 de *L. pneumophila* et ms142 de *P. aeruginosa*) et impliquent un évènement multiple (deux délétions de trois unités). Le modèle GSMM est donc à nuancer, de nouvelles analyses *in vitro* et *in vivo* doivent le confirmer. La diversité des structures répétées en tandem, notamment des microsatellites, pourrait expliquer la difficulté d'extrapoler un seul et même modèle évolutif pour tous ces loci.

ii) Modèle de l'allèle infini

Le modèle de l'allèle infini décrit chaque allèle différent comme issu d'un seul évènement d'acquisition ou de perte d'une ou plusieurs répétitions. Ce modèle de l'allèle infini, dont découle la théorie neutraliste de l'Evolution proposée par M. Kimura, stipule que chaque allèle, produit par mutation, est nouveau dans la population étudiée. Ces mutations sont sélectivement neutres, apparaissent de manière stochastique et sont éliminées par dérive génétique. Ainsi, un équilibre entre

taux de mutation et vitesse de dérive génétique permet de balancer la perte d'allèles par dérive génétique avec le gain d'allèles par mutation selon la formule $n = 4N_e u + 1$ avec n le nombre d'allèles stabilisés dans la population, N_e la taille de la population et u le taux de mutation. Les allèles sont donc vraisemblablement fixés lors de l'émergence de l'espèce, au moment où l'espèce compte peu d'individus, comme lors du processus de spéciation ou après un phénomène de goulot d'étranglement. Ce modèle est montré comme inapproprié pour l'étude de la dynamique des VNTRs notamment en raison de l'homoplasie de ces loci [167,168]. De plus, ce modèle a été appliqué pour les MIRU-VNTRs de *M. tuberculosis* et estime un taux de mutation en désaccord avec celui calculé expérimentalement [169], lui-même concordant avec le modèle GSMM [170,171]. J'ai évalué le taux de mutation de ms61 de *P. aeruginosa* compris entre 3.10^{-4} et 4.10^{-5} évènements par génération. L'observation d'un évènement de délétion lors de l'étude PSPE menée sur plus de 28000 générations de *S. aureus* me permet d'estimer le taux de mutation de Sa0122 à $3{,}6.10^{-5}$ évènements par génération. Ces taux sont équivalents à ceux mesurés pour d'autres loci chez d'autres espèces bactériennes [104,108,109,161] et à ceux identifiés par [169] lors de l'application du modèle de l'allèle infini. Ce dernier modèle serait-il donc plus adapté ? Reyes et Tanaka ont calculé le taux de mutation sur des loci VNTRs hypervariables exclus des protocoles classiques de génotypage de *M. tuberculosis*, le choix des marqueurs est donc biaisé. De plus, les auteurs justifient la véracité de leur taux de mutation en relevant la similarité avec ceux identifiés chez d'autres espèces (et donc avec ceux que j'ai identifiés). Cependant, la comparaison avec d'autres espèces et d'autres marqueurs est hasardeuse : les VNTRs étudiés entre *M. tuberculosis*, *E. coli*, *Y. pestis*, *P. aeruginosa* ou *S. aureus* sont différents structurellement. De plus, ces espèces ont des histoires évolutives distinctes et n'évoluent pas à la même vitesse. Tenter de rapprocher les taux de mutation de VNTRs hypervariables de *M. tuberculosis* à ceux de VNTRs utilisés en épidémiologie pour d'autres espèces n'est donc pas justifié. La modélisation de la dynamique des VNTRs selon l'allèle infinie reste marginale

3.3. Utilisation du modèle évolutif à des fins épidémiologiques

La détermination d'un modèle biologique d'évolution des répétitions en tandem marquera un tournant dans l'utilisation des données de typage par MLVA. L'épidémiologie de *Y. pestis* est un modèle unique : les seules tentatives de conciliation entre évolution des VNTRs et épidémiologie ont été invoquées pour ce pathogène [109,139,172,173]. Le modèle SSMM suivant une loi de Poisson a été appliqué à l'étude d'un cas épidémique de peste dans une population de chiens de prairies [139]. Plus récemment, en suivant le modèle GSMM, l'origine de la contraction de peste par un couple d'américains a été identifiée dans leur zone d'habitation (Figure 7). L'application du modèle GSMM aux données MLVA traitées par des méthodes de distance permet de prédire les topologies les plus probables en termes d'évolution [109]. L'arbre inféré suggère que la puce identifiée dans le jardin est plus susceptible d'avoir été le vecteur de la maladie [172]. Ainsi, l'utilisation de ce type de modèle

peut conduire à d'apparentes incongruités : un génotype identifié lors d'une épidémie de peste dans une population de chiens de prairie a été confronté à une base de données composée de 1565 profils MLVA et apparaît plus proche génétiquement d'une souche différant à sept loci qu'une autre différant à six [109].

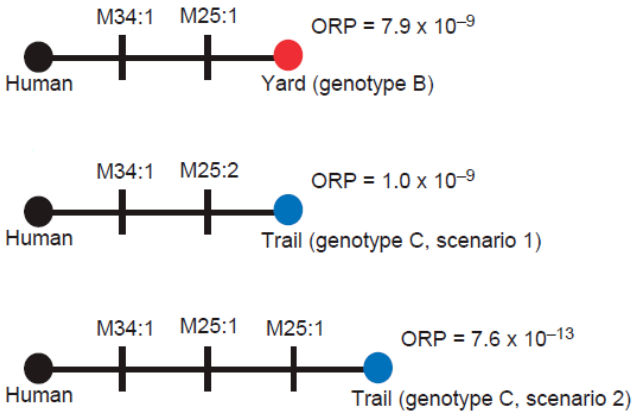


Figure 7. Application du modèle GSSM à l'épidémiologie (extrait de Colman et al., 2009)

Cinq isolats de *Y. pestis* ont été analysés pour l'enquête : un isolat humain de génotype A, trois isolats trouvés dans le jardin de la maison des patients de génotype B présentant deux polymorphismes de simple-répétition et un isolat récolté sur un sentier proche de la zone d'habitation des malades de génotype C présentant également deux polymorphismes, un de simple et un de double répétition. Alors que l'UPGMA avec une distance de Hamming aurait considéré les génotypes B et C équidistants du génotype A, le modèle GSMM les distingue en considérant que la mutation de deux motifs est plus rare que celle de un motif. L'hypothèse évolutive de passage d'un génotype à un autre la plus probable est évaluée par l'établissement d'un rapport de probabilité ou *odd ratio*.

Ces travaux démontrent que l'intégration d'un modèle évolutif est importante pour conduire des investigations épidémiologiques. Ce type de méthodologie constitue une avancée majeure dans l'utilisation des données de MLVA mais reste pour l'instant cantonnée à l'analyse des VNTRs hypervariables de type microsatellite utiles en investigation épidémique. En effet, l'utilisation de tels modèles en macro-épidémiologie n'est pas envisageable compte tenu de l'homoplasie inhérente aux loci VNTR.

4. L'homoplasie et les VNTRs

4.1. Démonstration des phénomènes homoplasiques

Pendant mes travaux de thèse, j'ai pu évaluer l'effet homoplasique auquel était soumise l'évolution des VNTRs en utilisant trois outils : a) l'incongruence, b) l'indice de déséquilibre de liaison et c) le séquençage d'allèles.

a. L'incongruence

La congruence entre la topologie d'arbres obtenus à partir de combinaisons différentes de VNTRs permet rapidement de mettre à jour des phénomènes homoplasiques. Si l'évolution des VNTRs suivait celle de l'espèce considérée, alors, quelle que soit la combinaison des marqueurs utilisés, les arbres inférés devraient être parfaitement congruents. Or, pour *S. aureus*, la congruence entre le MLVA-16_{Orsay} et le MLVA-$8_{Bilthoven}$ ne dépasse pas les 90% bien que les huit VNTRs marqueurs du MLVA-$8_{Bilthoven}$ soient compris dans le MLVA-16_{Orsay}. Lorsque les panels comprennent des marqueurs différents, cette incongruence est plus frappante : pour *P. aeruginosa*, les MLVA-$9_{Utrecht}$ et MLVA-9_{London} ne sont congruents qu'à seulement 35%.

b. Indice de déséquilibre de liaison

Les VNTRs ont été démontrés comme constituant d'excellents outils de structuration des populations. L'analyse du déséquilibre de liaison entre ces marqueurs a confirmé la clonalité de plusieurs espèces bactériennes. Au cours de mes travaux, j'ai également calculé cet indice au sein d'une population très diverse de *S. aureus* avec les données MLVA et MLST. Les deux indices sont significativement non nuls et prouvent la clonalité de l'espèce mais le déséquilibre est moins prononcé pour le MLVA (I^S_A = 0.1003 pour le MLVA et I^S_A = 0.1915 pour le MLST). Bien qu'atténuée, cette observation suggère l'homoplasie des VNTRs. En mesurant l'impact de chaque VNTR sur cet indice, j'ai pu mettre en évidence les marqueurs qui ne contribuaient pas au déséquilibre de liaison. En considérant le postulat de la clonalité de *S. aureus*, il est possible de proposer que les loci contribuant négativement à l'indice de déséquilibre de liaison soient sujets à homoplasie. En retirant ces huit marqueurs du protocole initial, la nouvelle combinaison s'est avérée beaucoup plus congruente avec le MLST (89.3% de congruence contre 79.3 avec le MLVA-16_{Orsay}) et davantage en déséquilibre de liaison (I^S_A = 0.1345). Cette méthodologie constitue un premier crible pour identifier les VNTRs les plus soumis aux phénomènes homoplasiques. Elle ouvre ainsi la voie au développement d'un procédé permettant de choisir, au sein d'un ensemble de loci, les VNTRs les plus pertinents pour inférer des structures de population et identifier des complexes clonaux.

c. Séquençage d'allèles

Le séquençage des différents allèles d'un VNTR peut être utilisé pour détecter les phénomènes homoplasiques. Cette méthode n'est cependant possible que pour les VNTRs qui contiennent une hétérogénéité interne significative. Le séquençage des allèles montrant un nombre de répétions identiques indique si ces allèles ont une histoire évolutive commune ou si ils sont le résultant d'un évènement de convergence. Au cours de mes travaux, j'ai analysé plusieurs données issues du séquençage du gène *spa* (VNTR Sa0122) de *S. aureus*. Le résultat du séquençage est standardisé et

associé à une nomenclature : chaque motif répété est indiqué par une lettre selon sa séquence et la combinaison des différentes lettres caractérise le locus. Le séquençage de ce marqueur montre que l'évolution par mutation du nombre de répétitions est fortement homoplasique. Ainsi par exemple, les isolats sa573 et sa639 présentent des profils MLVA correspondant respectivement aux complexes clonaux CC45 et CC5 alors que leur locus Sa0122 comporte le même allèle avec 10 unités répétées. Cependant, l'analyse des séquences des allèles montre qu'ils ne sont pas identiques : l'isolat sa573 montre un spa type XKAKBEMBKB alors que l'isolat sa639 présente un allèle très différent TJMBMDMGMK. Alors que l'analyse du polymorphisme du nombre de répétition peut être brouillée par l'homoplasie, l'étude de la séquence allélique révèle des informations pertinentes. Le spa type est congruent avec le MLST et permet le plus souvent l'assignation des différents complexes clonaux. De plus, l'analyse de diversité intra-allélique, mesurée par la comparaison par alignement des différents motifs, pourrait permettre de donner quelques clés de compréhension de la dynamique des VNTRs [44].

4.2. Causes de l'homoplasie

Le taux de mutation élevé, l'important polymorphisme et l'apolarité des mécanismes mutationnels de certains VNTRs sont autant de caractéristiques qui rendent ces marqueurs sujets aux effets d'homoplasie [151,174,175,176]. Les VNTRs seraient soumis à des pressions de sélection ; leur évolution ne serait donc pas neutre expliquant alors en partie l'homoplasie observée. Par exemple, chez *L. pneumophila*, au sein des ORFs annotés, 26 minisatellites ont été identifiés ; leurs motifs répétés étaient exclusivement des multiples de trois suggérant ainsi la forte pression de sélection assurant le maintien du cadre de lecture [115]. Parmi les loci utilisés dans le protocole MLVA développé, onze sont intragéniques, dont trois étudiés par [115] et qui pourraient être impliqués dans des phénomènes de variation de phase ou de variation antigénique. A l'inverse, parmi les 16 VNTRs du MLVA-16_{Orsay} pour le typage de *S. aureus*, seuls quatre sont situés dans une phase ouverte de lecture. Nous avons initié une étude préliminaire afin d'évaluer le rôle des VNTRs intergéniques dans la transcription. La diversité de *S. aureus* est grande, chaque complexe clonal présente un métabolisme ou une physiologie propre. Il ne semblait donc pas envisageable de comparer les transcriptomes de deux isolats différents génétiquement. L'enjeu était donc d'identifier des isolats appartenant au même complexe clonal et montrant un polymorphisme de répétitions élevé pour un VNTR. Le comportement de six VNTRs a donc été mesuré. Le niveau de transcription des gènes amont et/ou aval a été évalué par PCR en temps réel et standardisé par rapport à plusieurs gènes de référence. Certains VNTRs (Sa0397, Sa0906, Sa1291) paraissent être impliqués dans la transcription, d'autres non (Sa0964, Sa1729, Sa2039). Quelques résultats de ces expériences préliminaires sont illustrés sur la Figure 8. Ces données ne donnent qu'un aperçu du rôle putatif des VNTRs, des expériences complémentaires doivent être réalisées pour confirmer ces observations.

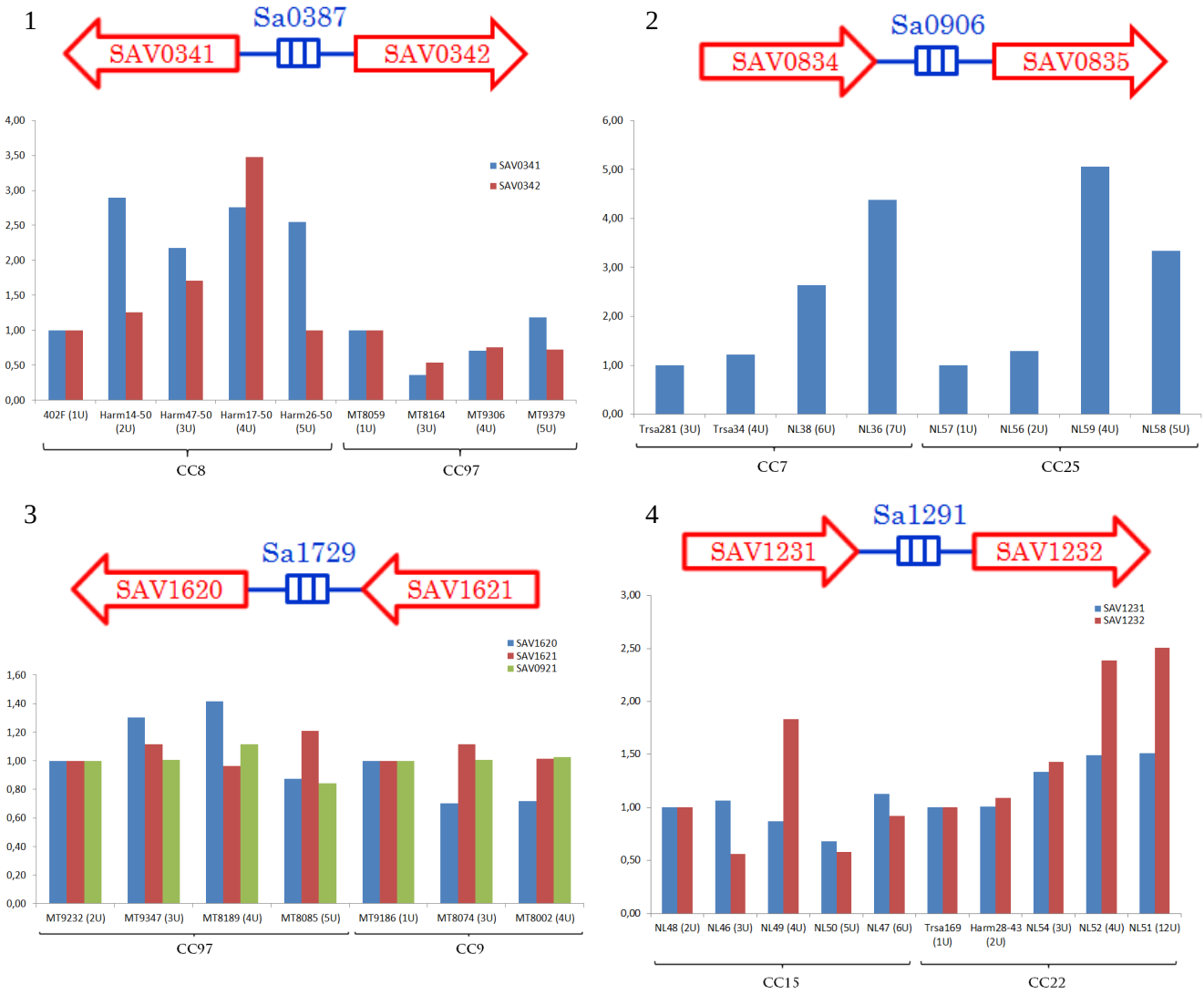


Figure 8. Rôle des VNTRs dans la transcription chez *S. aureus*.

Niveau d'expression par isolat normalisé par rapport à l'isolat comportant l'allèle le plus petit par complexe clonal. Le nombre d'unités répétées du VNTR d'intérêt pour chaque isolat est indiqué entre parenthèses.

1) Impact du polymorphisme de Sa0387 sur la transcription des gènes SAV0341 (codant pour une protéine hypothétique) et SAV0342 (codant pour une protéine ribosomale). Le locus semble jouer un rôle dans la transcription au sein du CC8. SAV0341 est davantage exprimé lorsque Sa0387 comporte plusieurs unités répétées. Le polymorphisme de Sa0387 est fortement corrélé avec le niveau de transcription de SAV0342 ; l'allèle 4U formerait une structure favorable à la transcription de SAV0342.

2) Impact du polymorphisme de Sa0906 sur la transcription du gène SAV0835 (codant pour une protéine hypothétique). Pour les deux CCs testés, le nombre de répétitions de Sa0906 influence la transcription.

3) Impact du polymorphisme de Sa1729 sur la transcription des gènes SAV1620 (codant pour une protéine hypothétique), SAV1621 (codant pour une thiouridylase t-RNA spécifique) et SAV0921 (codant pour une protéine hypothétique). La transcription du gène SAV0921 est utilisée comme témoin. Le polymorphisme du locus Sa1729 n'intervient pas sur le niveau de régulation de la transcription du gène situé en aval.

4) Impact du polymorphisme de Sa1291 sur la transcription des gènes SAV1231 (codant pour une 3-oxoacyl réductase) et SAV1232 (codant pour une protéine de type *acyl carrier*). Le polymorphisme de Sa1291 influe sur la transcription du gène en aval. Cependant, cette régulation est différentielle selon le CC testé : dans le CC15, l'allèle 4U contribue à activer la régulation alors que les autres la défavorisent ; pour le CC22, plus le locus comporte un grand nombre d'unités répétées, plus la transcription est forte.

4.3. Atténuation de l'homoplasie

Empiriquement, les protocoles MLVA sont congruents avec les méthodes de type MLST et permettent parfaitement d'assigner les différents complexes clonaux. L'agrégation ne semble donc pas être affectée par l'effet homoplasique de certains marqueurs. Cet effet d'atténuation peut être expliqué par trois raisons : i) l'augmentation du nombre de marqueurs utilisés permet d'atténuer le fort effet homoplasique induit par certains [177], ii) un ou plusieurs marqueurs caractérisent de manière robuste certains sous-lignages et permettent de les ancrer [152] et iii) la présence de VNTRs hautement polymorphes peu enclins, en termes de probabilité, à converger [176,177]. Cependant, ces suggestions ne sont que des hypothèses et certaines ne sont pas vérifiées dans tous les cas. Par exemple, j'ai montré qu'en enlevant les marqueurs identifiés comme homoplasiques par la méthode de déséquilibre de liaison, le protocole est globalement moins sujet au phénomène de convergence réfutant ainsi l'hypothèse i). L'effet structurant de certains marqueurs aux profils signatures peut être masqué par la convergence d'autres marqueurs ; l'hypothèse ii) n'est vraie que lorsque ces VNTRs sont majoritaires. Enfin, l'étude de l'impact des différents VNTRs dans le déséquilibre de la population de *S. aureus* a révélé une forte corrélation positive entre le polymorphisme et la contribution au déséquilibre. Les marqueurs polymorphes seraient donc moins enclins à la convergence confirmant ainsi l'hypothèse iii). Cependant, le marqueur Sa0122 connu pour son évolution convergente est très polymorphe et n'avait pas été retenu parmi les VNTRs dits homoplasiques.

5. Expansions clonales : causes et conséquences

La structure des espèces bactériennes ciblées par mes travaux de thèse est largement étudiée et de nombreux travaux sont disponibles dans la littérature. Nous n'ambitionnons pas de révolutionner les modèles établis ; cependant un point commun réunit l'ensemble des résultats : quelle que soit l'espèce, de panmictique à clonale, il s'avère que l'adaptation à un nouvel environnement (hôte animal, niche écologique, organe) est l'occasion d'expansions clonales.

5.1. Complexes clonaux et adaptation

Dans les différentes études que j'ai menées, des complexes clonaux ont été identifiés quelle que soit l'espèce analysée. Dans quelques cas, ces agrégats sont directement liés aux propriétés spécifiques d'une niche écologique. Chez *S. aureus*, l'adaptation est hôte-spécifique : certains complexes comme les CC133 et CC9 sont particulièrement adaptés aux hôtes animaux, respectivement les petits ruminants et les cochons, d'autres comme les CC8 ou CC45 sont spécifiques de l'hôte humain. Chez *L. pneumophila*, la diversité limitée retrouvée pourrait être la démonstration de l'adaptation de quelques complexes à l'environnement hydrique artificiel. De surcroît, quelques

complexes clonaux comme le VACC8 exhibent une incroyable adaptation aux réseaux d'eau. Le VACC8 colonise de manière prédominante les réseaux d'eau rennais suggérant que ce complexe serait particulièrement adapté à cet écosystème. Chez *P. aeruginosa*, la recombinaison empêche la structuration clonale. Cependant, lorsqu'un génotype s'avère particulièrement adapté à une niche écologique, celui-ci fonde un complexe qui évolue par expansion clonale. Plusieurs clones largement distribués présentent une identité génétique forte et sont retrouvés exclusivement dans les bronches de patients atteints de mucoviscidose.

Cette colonisation sélective de quelques complexes suggère l'existence d'écotypes chez les bactéries. Ces écotypes sont des groupes génomiquement homogènes composés d'individus adaptés à une niche écologique spécifique. Certains auteurs proposent d'ailleurs de faire évoluer le concept d'espèce bactérienne vers celui d'écotype. Comment émergent les écotypes et comment est assurée leur cohésion génomique ?

5.2. L'écotype bactérien

« Everything is everywhere, but, the environment selects », cet aphorisme plus connu sous le nom de l'hypothèse Baas-Becking traduit l'importance des facteurs écologiques dans l'évolution bactérienne [178]. Ernst Mayr prône une conception de l'espèce qui milite pour l'association du concept d'espèce écologique à la définition génomique de l'espèce. L'espèce écologique correspond à un groupe d'individus, appartenant à la même espèce biologique, ayant acquis une innovation écologique favorable pour permettre son adaptation à une nouvelle niche écologique c'est-à-dire un ensemble de paramètres physico-chimiques et biologiques caractérisant un milieu [179]. Le processus de spéciation est fortement orienté, chez les bactéries, par l'acquisition de nouvelles innovations écologiques adaptatives, par mutation ponctuelle mais surtout par transfert horizontal, augmentant le pouvoir adaptatif bactérien. Le développement de ces nouvelles fonctionnalités métaboliques permet l'adaptation de l'espèce bactérienne à son nouvel habitat, l'espèce devient alors un écotype [3]. La notion d'écotypes proposée par Cohan reprend le concept d'espèces écologiques et le développe [180]. Dans cette conception, la spéciation bactérienne ne résulte pas d'échanges génétiques mais est étroitement liée à l'apparition de phénomènes écologiques particuliers.

L'émergence de nouveaux écotypes implique vraisemblablement des évènements de sélection périodique consécutive à l'adaptation d'un mutant à une nouvelle niche écologique. Cette sélection périodique correspond à un phénomène de balayage sélectif ou *selective sweep*, c'est-à-dire une réduction locale du polymorphisme moléculaire suite à la fixation d'une mutation favorable. Cette mutation diffusera ensuite au sein de la population par le phénomène dit d'effet fondateur. Une mutation génétique peut, par ailleurs, favoriser l'aptitude de l'isolat mutant au sein de sa niche. Dans ce contexte, ce mutant supplantera alors les autres souches de ce même écotype, mais sans compétition avec un autre écotype. Ce mécanisme de sélection périodique purge régulièrement la diversité de

l'écotype (lors de la survenue d'une mutation avantageuse) et conduit au maintien d'une homogénéité au sein d'un écotype. La diversité au sein d'un écotype n'est donc qu'éphémère, elle est régulièrement et drastiquement diminuée par sélection périodique qui favorise un mutant adapté. La sélection périodique a ainsi pour conséquences de limiter la diversité intraécotype et d'augmenter la divergence entre écotypes qui forment avec le temps de réelles entités [3]. La recombinaison homologue entre écotypes peut ralentir l'élimination de la diversité génétique et le transfert latéral de gènes peut faciliter l'émergence continue de nouveaux écotypes avec une concomitante extinction d'autres écotypes concurrents [181]. L'évolution de la diversité des écotypes est modélisée sur la Figure 9. Les balayages sélectifs successifs au sein de chaque écotype augmentent les distances phylogénétiques qui les relient. A terme, les différents écotypes sont suffisamment divergents pour qu'il s'établisse entre eux une barrière génétique donnant naissance à une nouvelle espèce. Ainsi, ce concept d'écotype permet à la fois d'appréhender les causes de la spéciation (les conditions écologiques), les forces de cohésions garantes de l'intégrité de l'espèce (la sélection périodique) et les conséquences de la spéciation bactérienne (l'isolement sexuel).

Les complexes clonaux établis par MLST et MLVA correspondent, de manière remarquable, à des groupes écologiques distincts [3,181,182,183]. Pour *S. aureus* et *N. meningitidis*, ces complexes clonaux ont été apparentés à des écotypes [3,181,183]. Dans le cas de *N. meningitidis* par exemple, les complexes sont autrement appelés *genoclouds* [182]. L'effet de goulot d'étranglement géographique durant une expansion épidémique est à l'origine de la succession et de l'émergence continue de ces *genoclouds*. La diversité interne de ces agrégats n'est que transitoire et résulte de purges génétiques associées aux différentes vagues d'expansion pandémique. La migration et l'installation dans de nouvelles zones géographiques sont un prérequis pour la survie de cet organisme. L'acquisition d'une immunité par les hôtes d'une niche écologique donnée fait perdre toutes les particularités adaptatives du *genocloud* en place, qui sera remplacé par un nouveau *genocloud* adapté au nouveau contexte créant ainsi un nouveau goulot d'étranglement. Il apparaît donc évident que les *genoclouds* sont chacun liés à un écotype.

La limitation de recombinaison entre écotypes et les réguliers évènements de sélection périodique expliquent la corrélation des écotypes et des complexes clonaux et la diversité limitée de ces derniers. Cependant, les observations réalisées au cours de mes travaux suggèrent que les écotypes sont exclusifs dans leur niche respective. Un environnement est constitué d'une multitude de composantes physico-chimiques et biologiques dont les diverses combinaisons peuvent, en théorie, permettre la caractérisation de plusieurs niches écologiques. Les environnements hôte animal, réseaux d'eau rennais et bronches de patients mucoviscidosiques sont donc, *a priori*, ouverts à la colonisation de plusieurs écotypes. Cependant, dans les trois cas, la colonisation est exclusive suggérant que des évènements de compétition important inhibent l'installation d'autres écotypes. La disponibilité minimale de ressources communes aux différents écotypes peut expliquer cet effet compétiteur.

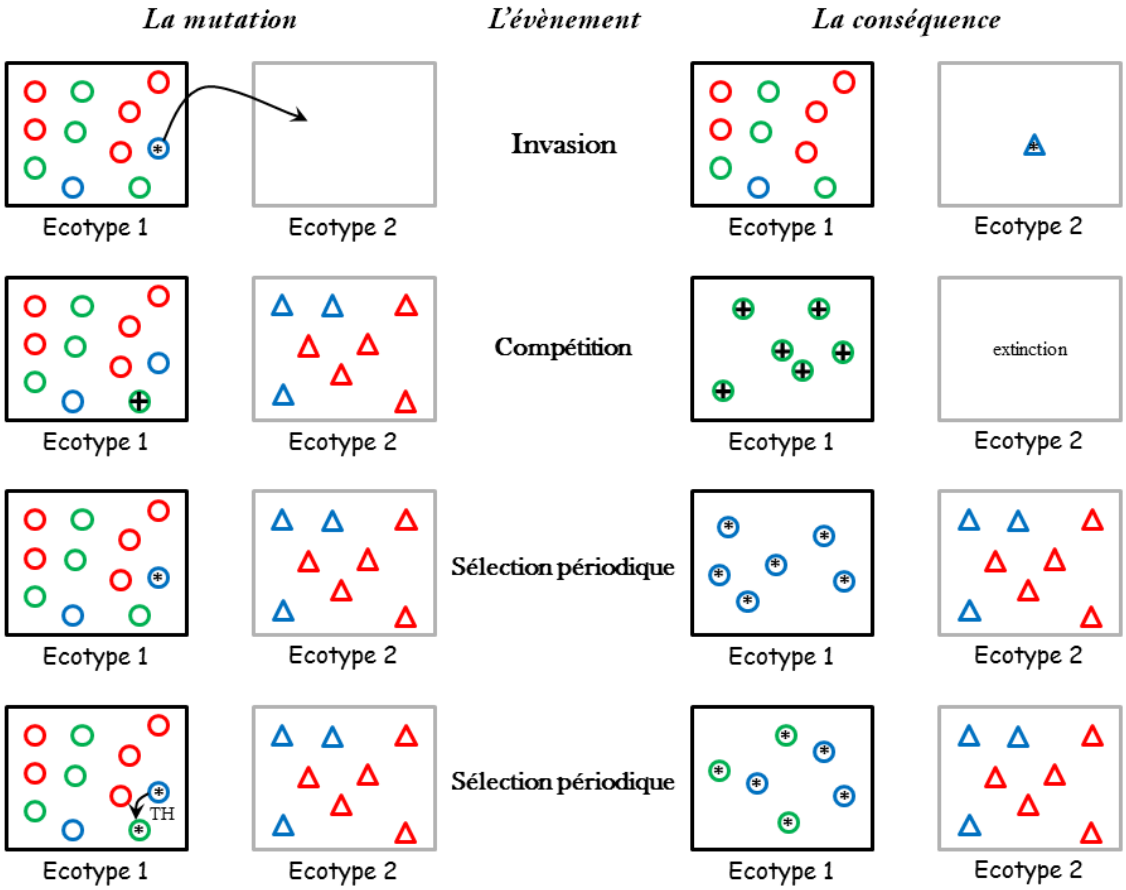


Figure 9. Diversité de l'écotype (adapté de [181])

Les ronds et les triangles symbolisent deux écotypes distincts dont la diversité est représentée par l'occurrence des couleurs.
* mutation adaptative
+ mutation adaptative exceptionnelle, le mutant est compétiteur et adapté à d'autres niches
La dynamique des écotypes est régulée par trois mécanismes principaux enclenchés par l'apparition d'une mutation adaptative. Le mutant peut être amené à pouvoir utiliser de nouvelles ressources et à envahir une nouvelle niche écologique créant alors un nouvel écotype : c'est l'invasion. L'apparition d'une mutation peut favoriser de manière exceptionnelle un génotype qui pourra alors supplanter les autres membres de son écotype et causer l'extinction d'un autre écotype si ce mutant a la capacité d'exploiter les mêmes ressources : c'est la compétition. Un évènement de purge de la diversité favorise le mutant adapté mais n'a pas d'effets sur les autres écotypes : c'est la sélection périodique. Lorsque la mutation adaptative est conférée par transfert horizontal (TH), la sélection périodique a un impact moindre sur la réduction génétique.

6. Conclusion

L'avancée technologique du MLVA permise par mes travaux de thèse a convaincu quelques laboratoires de référence majeurs dans le domaine du génotypage. L'utilisation standardisée de kits de génotypage rend désormais possible la caractérisation de manière systématique et en temps réel des isolats récoltés. Avec ce nouveau procédé, l'étude du polymorphisme des VNTRs s'impose comme la méthode de génotypage la plus adaptée. Le MLVA est à la fois un outil performant pour l'investigation épidémiologique et une méthode efficace d'analyse des structures de population. La généralisation des protocoles automatisés et l'enrichissement des bases de données garantiront la promotion du MLVA et sa plus large utilisation dans les laboratoires de microbiologie. Cependant, les progrès technologiques sont perpétuels et les méthodes qualifiées d'innovantes s'avèrent désuètes peu de temps après. Au moment de l'écriture de ce manuscrit, le MLVA sort à peine de la phase de recherche et commence à être appliqué lors d'études de terrain. Il aura donc fallu près de quinze ans pour que les VNTRs s'imposent comme des marqueurs moléculaires de premier choix. Ces loci sont aujourd'hui les plus commodes et les plus pertinents pour étudier la diversité bactérienne. Cependant, ces dernières années ont été marquées par un accès facilité aux nouvelles techniques de séquençage. L'essor de la bioinformatique a ouvert de nouvelles perspectives et l'analyse d'immenses fichiers de données devient concevable. A terme, le séquençage des génomes complets sera la panacée pour le typage, l'épidémiologie et la phylogénie des bactéries. Les NGS remplaceront-ils les méthodes classiques de typage et notamment le MLVA ? Vraisemblablement oui. Néanmoins, le coût d'un séquençage complet reste encore prohibitif pour une analyse haut-débit et des cribles pertinents permettant de filtrer la masse des données obtenues doivent être conçus. Le MLVA sous sa forme automatisée et standardisée a donc quelques années devant lui. Une connaissance plus approfondie de la biologie des VNTRs et de leurs mécanismes d'évolution contribueront à sélectionner de manière pertinente les loci d'intérêt et ainsi à augmenter la robustesse du MLVA

Bibliographie

1. Mayr E (1942) Systematics and the origin of species from the viewpoint of a zoologist. New York,: Columbia University Press. xiv, 334 p. incl. illus. (incl. maps) tables, diagrs. p.
2. Sonea S, Paniset M (1976) [Towards a new bacteriology]. Rev Can Biol 35: 103-167.
3. Cohan FM (2002) Sexual isolation and speciation in bacteria. Genetica 116: 359-370.
4. Sneath PH (1957) The application of computers to taxonomy. J Gen Microbiol 17: 201-226.
5. Parte A (2012) Bergey's manual of systematic bacteriology. New York: Springer.
6. Stackebrandt E, Frederiksen W, Garrity GM, Grimont PA, Kampfer P, et al. (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Microbiol 52: 1043-1047.
7. Gould SJ (1996) Full house : the spread of excellence from Plato to Darwin. New York: Harmony Books. 244 p. p.
8. Prescott LM, Harley JP, Klein DA (2002) Microbiology. Boston: McGraw-Hill.
9. Monod J (1970) Le hasard et la nécessité; essai sur la philosophie naturelle de la biologie moderne. Paris,: Éditions du Seuil. 197 p. p.
10. Desenclos JC, Vaillant V, Delarocque Astagneau E, Campese C, Che D, et al. (2007) [Principles of an outbreak investigation in public health practice]. Med Mal Infect 37: 77-94.
11. Desenclos JC, Bouvet P, Benz-Lemoine E, Grimont F, Desqueyroux H, et al. (1996) Large outbreak of *Salmonella enterica* serotype paratyphi B infection caused by a goats' milk cheese, France, 1993: a case finding and epidemiological study. BMJ 312: 91-94.
12. de Valk H, Vaillant V, Jacquet C, Rocourt J, Le Querrec F, et al. (2001) Two consecutive nationwide outbreaks of Listeriosis in France, October 1999-February 2000. Am J Epidemiol 154: 944-950.
13. Cheng AC, Jacups SP, Gal D, Mayo M, Currie BJ (2006) Extreme weather events and environmental contamination are associated with case-clusters of melioidosis in the Northern Territory of Australia. Int J Epidemiol 35: 323-329.
14. Al Dahouk S, Hagen RM, Nockler K, Tomaso H, Wittig M, et al. (2005) Failure of a short-term antibiotic therapy for human brucellosis using ciprofloxacin. A study on in vitro susceptibility of Brucella strains. Chemotherapy 51: 352-356.
15. Mermel LA, McCormick RD, Springman SR, Maki DG (1991) The pathogenesis and epidemiology of catheter-related infection with pulmonary artery Swan-Ganz catheters: a prospective study utilizing molecular subtyping. Am J Med 91: 197S-205S.
16. Struelens MJ, Denis O, Rodriguez-Villalobos H (2004) Microbiology of nosocomial infections: progress and challenges. Microbes Infect 6: 1043-1048.
17. Peterson LR, Noskin GA (2001) New technology for detecting multidrug-resistant pathogens in the clinical microbiology laboratory. Emerg Infect Dis 7: 306-311.
18. Scheinbach S, Hong SI (1988) Detection of resident populations of Salmonella and *Escherichia coli* in food ingredients by plasmid analysis. Food Microbiology 5: 235-241.
19. Dodd CE, Chaffey BJ, Waites WM (1988) Plasmid profiles as indicators of the source of contamination of Staphylococcus aureus endemic within poultry processing plants. Appl Environ Microbiol 54: 1541-1549.
20. Hofstra H, van der Vossen JM, van der Plas J (1994) Microbes in food processing technology. FEMS Microbiol Rev 15: 175-183.
21. Dodd CER (1994) The Application of Molecular Typing Techniques to Haccp. Trends in Food Science & Technology 5: 160-164.
22. Bellamy RJ, Freedman AR (2001) Bioterrorism. QJM 94: 227-234.
23. Wheelis M, Rózsa L, Dando M (2006) Deadly cultures : biological weapons since 1945. Cambridge, Mass.: Harvard University Press. xi, 479 p. p.
24. Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, et al. (2000) Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. J Bacteriol 182: 2928-2936.
25. Hoffmaster AR, Fitzgerald CC, Ribot E, Mayer LW, Popovic T (2002) Molecular subtyping of *Bacillus anthracis* and the 2001 bioterrorism-associated anthrax outbreak, United States. Emerg Infect Dis 8: 1111-1116.

26. Harrison TG, Doshi N, Fry NK, Joseph CA (2007) Comparison of clinical and environmental isolates of *Legionella pneumophila* obtained in the UK over 19 years. Clin Microbiol Infect 13: 78-85.
27. Ginevra C, Forey F, Campese C, Reyrolle M, Che D, et al. (2008) Lorraine strain of *Legionella pneumophila* serogroup 1, France. Emerg Infect Dis 14: 673-675.
28. Daurel C, Prunier AL, Chau F, Garry L, Leclercq R, et al. (2007) Role of hypermutability on bacterial fitness and emergence of resistance in experimental osteomyelitis due to Staphylococcus aureus. FEMS Immunol Med Microbiol 51: 344-349.
29. Jaureguy F, Landraud L, Passet V, Diancourt L, Frapy E, et al. (2008) Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. BMC Genomics 9: 560.
30. Spratt BG (2004) Exploring the concept of clonality in bacteria. Methods Mol Biol 266: 323-352.
31. Smith JM, Smith NH, O'Rourke M, Spratt BG (1993) How clonal are bacteria? Proc Natl Acad Sci U S A 90: 4384-4388.
32. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567-1587.
33. Feil EJ, Enright MC (2004) Analyses of clonality and the evolution of bacterial pathogens. Curr Opin Microbiol 7: 308-313.
34. Smith GR (1991) Conjugational recombination in E. coli: myths and mechanisms. Cell 64: 19-27.
35. Smith NH, Beltran P, Selander RK (1990) Recombination of Salmonella phase 1 flagellin genes generates new serovars. J Bacteriol 172: 2209-2216.
36. Moxon ER, Rainey PB, Nowak MA, Lenski RE (1994) Adaptive evolution of highly mutable loci in pathogenic bacteria. Curr Biol 4: 24-33.
37. Go MF, Kapur V, Graham DY, Musser JM (1996) Population genetic analysis of *Helicobacter pylori* by multilocus enzyme electrophoresis: extensive allelic diversity and recombinational population structure. J Bacteriol 178: 3934-3938.
38. Salaun L, Audibert C, Le Lay G, Burucoa C, Fauchere JL, et al. (1998) Panmictic structure of *Helicobacter pylori* demonstrated by the comparative study of six genetic markers. FEMS Microbiol Lett 161: 231-239.
39. Falush D, Kraft C, Taylor NS, Correa P, Fox JG, et al. (2001) Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. Proc Natl Acad Sci U S A 98: 15056-15061.
40. Feil EJ (2004) Small change: keeping pace with microevolution. Nat Rev Microbiol 2: 483-495.
41. Feil EJ, Spratt BG (2001) Recombination and the population structures of bacterial pathogens. Annu Rev Microbiol 55: 561-590.
42. Cohan FM, Koeppel A, Krizanc D (2006) Sequence-based discovery of ecological diversity within Legionella. Legionella: State of the Art 30 Years after Its Recognition: 367-376.
43. Edwards MT, Fry NK, Harrison TG (2008) Clonal population structure of *Legionella pneumophila* inferred from allelic profiling. Microbiology 154: 852-864.
44. Visca P, D'Arezzo S, Ramisse F, Gelfand Y, Benson G, et al. (2011) Investigation of the population structure of *Legionella pneumophila* by analysis of tandem repeat copy number and internal sequence variation. Microbiology 157: 2582-2594.
45. Achtman M (2008) Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu Rev Microbiol 62: 53-70.
46. Gutierrez MC, Brisse S, Brosch R, Fabre M, Omais B, et al. (2005) Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. PLoS Pathog 1: e5.
47. Coombs MC (1976) The Selfish Gene (Book Review). Library Journal 101: 2500.
48. Craig NL (2002) Mobile DNA II. Washington, D.C.: ASM Press. xviii, 1204 p., 1232 p. of plates p.
49. Merlin C, Toussaint A (1999) Transposition bacterial elements. M S-Medecine Sciences 15: Ar1-Ar13.
50. Neidhardt FC, Curtiss R (1996) *Escherichia coli* and Salmonella : cellular and molecular biology. Washington, D.C.: ASM Press.
51. Ochman H, Jones IB (2000) Evolutionary dynamics of full genome content in *Escherichia coli*. EMBO J 19: 6637-6643.

52. Correia A, Pisabarro A, Castro JM, Martin JF (1996) Cloning and characterization of an IS-like element present in the genome of *Brevibacterium lactofermentum* ATCC 13869. Gene 170: 91-94.
53. Buisine N, Tang CM, Chalmers R (2002) Transposon-like Correia elements: structure, distribution and genetic exchange between pathogenic Neisseria sp. FEBS Lett 522: 52-58.
54. Oggioni MR, Claverys JP (1999) Repeated extragenic sequences in prokaryotic genomes: a proposal for the origin and dynamics of the RUP element in *Streptococcus pneumoniae*. Microbiology 145 (Pt 10): 2647-2653.
55. Higgins CF, Ames GF, Barnes WM, Clement JM, Hofnung M (1982) A novel intercistronic regulatory element of prokaryotic operons. Nature 298: 760-762.
56. Bachellier S, Saurin W, Perrin D, Hofnung M, Gilson E (1994) Structural and functional diversity among bacterial interspersed mosaic elements (BIMEs). Mol Microbiol 12: 61-70.
57. Espeli O, Moulin L, Boccard F (2001) Transcription attenuation associated with bacterial repetitive extragenic BIME elements. J Mol Biol 314: 375-386.
58. Khemici V, Carpousis AJ (2004) The RNA degradosome and poly(A) polymerase of *Escherichia coli* are required in vivo for the degradation of small mRNA decay intermediates containing REP-stabilizers. Mol Microbiol 51: 777-790.
59. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315: 1709-1712.
60. Denoeud F, Vergnaud G (2004) Identification of polymorphic tandem repeats by direct comparison of genome sequence from different bacterial strains: a web-based resource. BMC Bioinformatics 5: 4.
61. Wyman AR, White R (1980) A highly polymorphic locus in human DNA. Proc Natl Acad Sci U S A 77: 6754-6758.
62. Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable 'minisatellite' regions in human DNA. Nature 314: 67-73.
63. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27: 573-580.
64. Le Fleche P, Hauck Y, Onteniente L, Prieur A, Denoeud F, et al. (2001) A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*. BMC Microbiol 1: 2.
65. Skolnick MH, White R (1982) Strategies for detecting and characterizing restriction fragment length polymorphisms (RFLP's). Cytogenet Cell Genet 32: 58-67.
66. Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, et al. (1995) Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J Clin Microbiol 33: 2233-2239.
67. Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res 18: 7213-7218.
68. Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18: 6531-6535.
69. Lin JJ, Kuo J, Ma J (1996) A PCR-based DNA fingerprinting technique: AFLP for molecular typing of bacteria. Nucleic Acids Res 24: 3649-3650.
70. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, et al. (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23: 4407-4414.
71. Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17: 6463-6471.
72. Lygo JE, Johnson PE, Holdaway DJ, Woodroffe S, Whitaker JP, et al. (1994) The validation of short tandem repeat (STR) loci for use in forensic casework. Int J Legal Med 107: 77-89.
73. van Belkum A, Melchers WJ, Ijsseldijk C, Nohlmans L, Verbrugh H, et al. (1997) Outbreak of amoxicillin-resistant *Haemophilus influenzae* type b: variable number of tandem repeats as novel molecular markers. J Clin Microbiol 35: 1517-1520.
74. Lindstedt BA, Vardund T, Kapperud G (2004) Multiple-Locus Variable-Number Tandem-Repeats Analysis of *Escherichia coli* O157 using PCR multiplexing and multi-colored capillary electrophoresis. J Microbiol Methods 58: 213-222.

75. Ramisse V, Houssu P, Hernandez E, Denoeud F, Hilaire V, et al. (2004) Variable number of tandem repeats in *Salmonella enterica* subsp. enterica for typing purposes. J Clin Microbiol 42: 5722-5730.
76. Koeck JL, Njanpop-Lafourcade BM, Cade S, Varon E, Sangare L, et al. (2005) Evaluation and selection of tandem repeat loci for *Streptococcus pneumoniae* MLVA strain typing. BMC Microbiol 5: 66.
77. Pourcel C, Minandri F, Hauck Y, D'Arezzo S, Imperi F, et al. (2011) Identification of variable-number tandem-repeat (VNTR) sequences in *Acinetobacter baumannii* and interlaboratory validation of an optimized multiple-locus VNTR analysis typing scheme. J Clin Microbiol 49: 539-548.
78. Pourcel C, Visca P, Afshar B, D'Arezzo S, Vergnaud G, et al. (2007) Identification of variable-number tandem-repeat (VNTR) sequences in *Legionella pneumophila* and development of an optimized multiple-locus VNTR analysis typing scheme. J Clin Microbiol 45: 1190-1199.
79. Wang YW, Watanabe H, Phung DC, Tung SK, Lee YS, et al. (2009) Multilocus variable-number tandem repeat analysis for molecular typing and phylogenetic analysis of *Shigella flexneri*. BMC Microbiol 9: 278.
80. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, 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: 3140-3145.
81. Gaia V, Fry NK, Afshar B, Luck PC, Meugnier H, et al. (2005) Consensus sequence-based scheme for epidemiological typing of clinical and environmental isolates of *Legionella pneumophila*. J Clin Microbiol 43: 2047-2052.
82. Gaia V, Fry NK, Harrison TG, Peduzzi R (2003) Sequence-based typing of *Legionella pneumophila* serogroup 1 offers the potential for true portability in legionellosis outbreak investigation. J Clin Microbiol 41: 2932-2939.
83. Bielaszewska M, Mellmann A, Zhang W, Kock R, Fruth A, et al. (2011) Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. Lancet Infect Dis 11: 671-676.
84. Maslow JN, Mulligan ME, Arbeit RD (1993) Molecular epidemiology: application of contemporary techniques to the typing of microorganisms. Clin Infect Dis 17: 153-162; quiz 163-154.
85. Struelens MJ, De Gheldre Y, Deplano A (1998) Comparative and library epidemiological typing systems: outbreak investigations versus surveillance systems. Infect Control Hosp Epidemiol 19: 565-569.
86. Tibayrenc M (1998) Beyond strain typing and molecular epidemiology: integrated genetic epidemiology of infectious diseases. Parasitol Today 14: 323-329.
87. Keim P, Van Ert MN, Pearson T, Vogler AJ, Huynh LY, et al. (2004) Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales. Infect Genet Evol 4: 205-213.
88. van Belkum A (2007) Tracing isolates of bacterial species by multilocus variable number of tandem repeat analysis (MLVA). FEMS Immunol Med Microbiol 49: 22-27.
89. Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4: 203-221.
90. Schlotterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. Nucleic Acids Res 20: 211-215.
91. van Belkum A, Scherer S, van Alphen L, Verbrugh H (1998) Short-sequence DNA repeats in prokaryotic genomes. Microbiol Mol Biol Rev 62: 275-293.
92. Saveson CJ, Lovett ST (1999) Tandem repeat recombination induced by replication fork defects in *Escherichia coli* requires a novel factor, RadC. Genetics 152: 5-13.
93. Richard GF, Paques F (2000) Mini- and microsatellite expansions: the recombination connection. EMBO Rep 1: 122-126.
94. Shen P, Huang HV (1986) Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. Genetics 112: 441-457.
95. Watt VM, Ingles CJ, Urdea MS, Rutter WJ (1985) Homology requirements for recombination in *Escherichia coli*. Proc Natl Acad Sci U S A 82: 4768-4772.

96. Bi X, Liu LF (1994) recA-independent and recA-dependent intramolecular plasmid recombination. Differential homology requirement and distance effect. J Mol Biol 235: 414-423.
97. Lin Y, Hubert L, Jr., Wilson JH (2009) Transcription destabilizes triplet repeats. Mol Carcinog 48: 350-361.
98. Wierdl M, Dominska M, Petes TD (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics 146: 769-779.
99. Kunkel TA (1992) DNA replication fidelity. J Biol Chem 267: 18251-18254.
100. Radman M, Wagner R (1986) Mismatch repair in *Escherichia coli*. Annu Rev Genet 20: 523-538.
101. Levy DD, Cebula TA (2001) Fidelity of replication of repetitive DNA in mutS and repair proficient *Escherichia coli*. Mutat Res 474: 1-14.
102. Sia EA, Kokoska RJ, Dominska M, Greenwell P, Petes TD (1997) Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. Mol Cell Biol 17: 2851-2858.
103. Debrauwere H, Buard J, Tessier J, Aubert D, Vergnaud G, et al. (1999) Meiotic instability of human minisatellite CEB1 in yeast requires DNA double-strand breaks. Nat Genet 23: 367-371.
104. Vogler AJ, Keys C, Nemoto Y, Colman RE, Jay Z, et al. (2006) Effect of repeat copy number on variable-number tandem repeat mutations in *Escherichia coli* O157:H7. J Bacteriol 188: 4253-4263.
105. Truglio JJ, Croteau DL, Van Houten B, Kisker C (2006) Prokaryotic nucleotide excision repair: the UvrABC system. Chem Rev 106: 233-252.
106. Parniewski P, Bacolla A, Jaworski A, Wells RD (1999) Nucleotide excision repair affects the stability of long transcribed (CTG*CAG) tracts in an orientation-dependent manner in *Escherichia coli*. Nucleic Acids Res 27: 616-623.
107. Feschenko VV, Rajman LA, Lovett ST (2003) Stabilization of perfect and imperfect tandem repeats by single-strand DNA exonucleases. Proc Natl Acad Sci U S A 100: 1134-1139.
108. Noller AC, McEllistrem MC, Shutt KA, Harrison LH (2006) Locus-specific mutational events in a multilocus variable-number tandem repeat analysis of *Escherichia coli* O157:H7. J Clin Microbiol 44: 374-377.
109. Vogler AJ, Keys CE, Allender C, Bailey I, Girard J, et al. (2007) Mutations, mutation rates, and evolution at the hypervariable VNTR loci of *Yersinia pestis*. Mutat Res 616: 145-158.
110. Kroutil LC, Register K, Bebenek K, Kunkel TA (1996) Exonucleolytic proofreading during replication of repetitive DNA. Biochemistry 35: 1046-1053.
111. Tran HT, Keen JD, Kricker M, Resnick MA, Gordenin DA (1997) Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. Mol Cell Biol 17: 2859-2865.
112. Morel P, Reverdy C, Michel B, Ehrlich SD, Cassuto E (1998) The role of SOS and flap processing in microsatellite instability in *Escherichia coli*. Proc Natl Acad Sci U S A 95: 10003-10008.
113. Petes TD, Greenwell PW, Dominska M (1997) Stabilization of microsatellite sequences by variant repeats in the yeast *Saccharomyces cerevisiae*. Genetics 146: 491-498.
114. Cooley MB, Carychao D, Nguyen K, Whitehand L, Mandrell R (2010) Effects of environmental stress on stability of tandem repeats in *Escherichia coli* O157:H7. Appl Environ Microbiol 76: 3398-3400.
115. Coil DA, Vandersmissen L, Ginevra C, Jarraud S, Lammertyn E, et al. (2008) Intragenic tandem repeat variation between *Legionella pneumophila* strains. BMC Microbiol 8: 218.
116. Gauthier MJ, Labedan B, Breittmayer VA (1992) Influence of DNA supercoiling on the loss of culturability of *Escherichia coli* cells incubated in seawater. Mol Ecol 1: 183-190.
117. Majchrzak M, Bowater RP, Staczek P, Parniewski P (2006) SOS repair and DNA supercoiling influence the genetic stability of DNA triplet repeats in *Escherichia coli*. J Mol Biol 364: 612-624.
118. Rocha EP, Matic I, Taddei F (2002) Over-representation of repeats in stress response genes: a strategy to increase versatility under stressful conditions? Nucleic Acids Res 30: 1886-1894.
119. van der Woude MW, Baumler AJ (2004) Phase and antigenic variation in bacteria. Clin Microbiol Rev 17: 581-611, table of contents.

120. De Bolle X, Bayliss CD, Field D, van de Ven T, Saunders NJ, et al. (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: 211-222.
121. Srikhanta YN, Dowideit SJ, Edwards JL, Falsetta ML, Wu HJ, et al. (2009) Phasevarions mediate random switching of gene expression in pathogenic Neisseria. PLoS Pathog 5: e1000400.
122. Hood DW, Deadman ME, Jennings MP, Bisercic M, Fleischmann RD, et al. (1996) DNA repeats identify novel virulence genes in *Haemophilus influenzae*. Proc Natl Acad Sci U S A 93: 11121-11125.
123. Saunders NJ, Peden JF, Hood DW, Moxon ER (1998) Simple sequence repeats in the *Helicobacter pylori* genome. Mol Microbiol 27: 1091-1098.
124. Gravekamp C, Rosner B, Madoff LC (1998) Deletion of repeats in the alpha C protein enhances the pathogenicity of group B streptococci in immune mice. Infect Immun 66: 4347-4354.
125. Madoff LC, Michel JL, Gong EW, Kling DE, Kasper DL (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: 4131-4136.
126. Citti C, Kim MF, Wise KS (1997) Elongated versions of Vlp surface lipoproteins protect *Mycoplasma hyorhinis* escape variants from growth-inhibiting host antibodies. Infect Immun 65: 1773-1785.
127. Ritz D, Lim J, Reynolds CM, Poole LB, Beckwith J (2001) Conversion of a peroxiredoxin into a disulfide reductase by a triplet repeat expansion. Science 294: 158-160.
128. van Ham SM, van Alphen L, Mooi FR, van Putten JP (1993) Phase variation of *H. influenzae* fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. Cell 73: 1187-1196.
129. Martin P, van de Ven T, Mouchel N, Jeffries AC, Hood DW, et al. (2003) Experimentally revised repertoire of putative contingency loci in *Neisseria meningitidis* strain MC58: evidence for a novel mechanism of phase variation. Mol Microbiol 50: 245-257.
130. Akhtar P, Singh S, Bifani P, Kaur S, Srivastava BS, et al. (2009) Variable-number tandem repeat 3690 polymorphism in Indian clinical isolates of *Mycobacterium tuberculosis* and its influence on transcription. J Med Microbiol 58: 798-805.
131. Zheng H, Lu L, Wang B, Pu S, Zhang X, et al. (2008) Genetic basis of virulence attenuation revealed by comparative genomic analysis of *Mycobacterium tuberculosis* strain H37Ra versus H37Rv. PLoS One 3: e2375.
132. Lindstedt BA (2005) Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. Electrophoresis 26: 2567-2582.
133. Lankester ER (1870) II.—On the use of the term homology in modern zoology, and the distinction between homogenetic and homoplastic agreements. Journal of Natural History Series 4 6: 34-43.
134. Vergnaud G, Pourcel C (2009) Multiple locus variable number of tandem repeats analysis. Methods Mol Biol 551: 141-158.
135. Lindstedt BA, Heir E, Gjernes E, 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: 1469-1479.
136. Boxrud D, Pederson-Gulrud K, Wotton J, Medus C, Lyszkowicz E, et al. (2007) Comparison of multiple-locus variable-number tandem repeat analysis, pulsed-field gel electrophoresis, and phage typing for subtype analysis of *Salmonella enterica* serotype Enteritidis. J Clin Microbiol 45: 536-543.
137. Malorny B, Junker E, Helmuth R (2008) Multi-locus variable-number tandem repeat analysis for outbreak studies of *Salmonella enterica* serotype Enteritidis. BMC Microbiol 8: 84.
138. Van Ert MN, Easterday WR, Huynh LY, Okinaka RT, Hugh-Jones ME, et al. (2007) Global genetic population structure of *Bacillus anthracis*. PLoS One 2: e461.
139. Girard JM, Wagner DM, Vogler AJ, Keys C, Allender CJ, et al. (2004) Differential plague-transmission dynamics determine *Yersinia pestis* population genetic structure on local, regional, and global scales. Proc Natl Acad Sci U S A 101: 8408-8413.
140. Le Fleche P, Jacques I, Grayon M, Al Dahouk S, Bouchon P, et al. (2006) Evaluation and selection of tandem repeat loci for a Brucella MLVA typing assay. BMC Microbiol 6: 9.

141. Jiao WW, Mokrousov I, Sun GZ, Guo YJ, Vyazovaya A, et al. (2008) Evaluation of new variable-number tandem-repeat systems for typing *Mycobacterium tuberculosis* with Beijing genotype isolates from Beijing, China. J Clin Microbiol 46: 1045-1049.
142. Grissa I, Bouchon P, Pourcel C, Vergnaud G (2008) On-line resources for bacterial micro-evolution studies using MLVA or CRISPR typing. Biochimie 90: 660-668.
143. Top J, Schouls LM, Bonten MJ, Willems RJ (2004) Multiple-locus variable-number tandem repeat analysis, a novel typing scheme to study the genetic relatedness and epidemiology of *Enterococcus faecium* isolates. J Clin Microbiol 42: 4503-4511.
144. Allix-Beguec C, Harmsen D, Weniger T, Supply P, Niemann S (2008) Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis* complex isolates. J Clin Microbiol 46: 2692-2699.
145. Keys C, Kemper S, Keim P (2005) Highly diverse variable number tandem repeat loci in the *E. coli* O157:H7 and O55:H7 genomes for high-resolution molecular typing. J Appl Microbiol 98: 928-940.
146. Lindstedt BA, Tham W, Danielsson-Tham ML, Vardund T, Helmersson S, et al. (2008) Multiple-locus variable-number tandem-repeats analysis of *Listeria monocytogenes* using multicolour capillary electrophoresis and comparison with pulsed-field gel electrophoresis typing. J Microbiol Methods 72: 141-148.
147. Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, et al. (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.
148. Ciammaruconi A, Grassi S, De Santis R, Faggioni G, Pittiglio V, et al. (2008) Fieldable genotyping of *Bacillus anthracis* and *Yersinia pestis* based on 25-loci Multi Locus VNTR Analysis. BMC Microbiol 8: 21.
149. Schouls LM, Spalburg EC, van Luit M, Huijsdens XW, Pluister GN, et al. (2009) Multiple-locus variable number tandem repeat analysis of *Staphylococcus aureus*: comparison with pulsed-field gel electrophoresis and spa-typing. PLoS One 4: e5082.
150. Sperry KE, Kathariou S, Edwards JS, Wolf LA (2008) Multiple-locus variable-number tandem-repeat analysis as a tool for subtyping *Listeria monocytogenes* strains. J Clin Microbiol 46: 1435-1450.
151. Comas I, Homolka S, Niemann S, Gagneux S (2009) Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. PLoS One 4: e7815.
152. Wada T, Iwamoto T (2009) Allelic diversity of variable number of tandem repeats provides phylogenetic clues regarding the *Mycobacterium tuberculosis* Beijing family. Infect Genet Evol 9: 921-926.
153. Supply P, Warren RM, Banuls AL, Lesjean S, Van Der Spuy GD, et al. (2003) Linkage disequilibrium between minisatellite loci supports clonal evolution of *Mycobacterium tuberculosis* in a high tuberculosis incidence area. Mol Microbiol 47: 529-538.
154. Ngoc LB, Verniere C, Vital K, Guerin F, Gagnevin L, et al. (2009) Development of 14 minisatellite markers for the citrus canker bacterium, *Xanthomonas citri* pv. citri. Mol Ecol Resour 9: 125-127.
155. Hidalgo A, Carvajal A, La T, Naharro G, Rubio P, et al. (2010) Multiple-locus variable-number tandem-repeat analysis of the swine dysentery pathogen, *Brachyspira hyodysenteriae*. J Clin Microbiol 48: 2859-2865.
156. Brevik OJ, Ottem KF, Nylund A (2011) Multiple-locus, variable number of tandem repeat analysis (MLVA) of the fish-pathogen *Francisella noatunensis*. BMC Vet Res 7: 5.
157. Ansede-Bermejo J, Gavilan RG, Trinanes J, Espejo RT, Martinez-Urtaza J (2010) Origins and colonization history of pandemic *Vibrio parahaemolyticus* in South America. Mol Ecol 19: 3924-3937.
158. Sokal RR, Michener CD, Kansas Uo (1958) A statistical method for evaluating systematic relationships: University of Kansas.
159. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406-425.

160. Felsenstein J (1985) Confidence-Limits on Phylogenies - an Approach Using the Bootstrap. Evolution 39: 783-791.
161. Price EP, Hornstra HM, Limmathurotsakul D, Max TL, Sarovich DS, et al. (2010) Within-host evolution of *Burkholderia pseudomallei* in four cases of acute melioidosis. PLoS Pathog 6: e1000725.
162. Bichara M, Wagner J, Lambert IB (2006) Mechanisms of tandem repeat instability in bacteria. Mutat Res 598: 144-163.
163. Morag AS, Saveson CJ, Lovett ST (1999) Expansion of DNA repeats in *Escherichia coli*: effects of recombination and replication functions. J Mol Biol 289: 21-27.
164. Henderson ST, Petes TD (1992) Instability of simple sequence DNA in Saccharomyces cerevisiae. Mol Cell Biol 12: 2749-2757.
165. Pollard LM, Chutake YK, Rindler PM, Bidichandani SI (2007) Deficiency of RecA-dependent RecFOR and RecBCD pathways causes increased instability of the (GAA*TTC)n sequence when GAA is the lagging strand template. Nucleic Acids Res 35: 6884-6894.
166. Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, et al. (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci U S A 91: 3166-3170.
167. Harding RM, Boyce AJ, Clegg JB (1992) The evolution of tandemly repetitive DNA: recombination rules. Genetics 132: 847-859.
168. Shriver MD, Jin L, Chakraborty R, Boerwinkle E (1993) VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. Genetics 134: 983-993.
169. Reyes JF, Tanaka MM (2010) Mutation rates of spoligotypes and variable numbers of tandem repeat loci in *Mycobacterium tuberculosis*. Infect Genet Evol 10: 1046-1051.
170. Grant A, Arnold C, Thorne N, Gharbia S, Underwood A (2008) Mathematical modelling of *Mycobacterium tuberculosis* VNTR loci estimates a very slow mutation rate for the repeats. J Mol Evol 66: 565-574.
171. Wirth T, Hildebrand F, Allix-Beguec C, Wolbeling F, Kubica T, et al. (2008) Origin, spread and demography of the *Mycobacterium tuberculosis* complex. PLoS Pathog 4: e1000160.
172. Colman RE, Vogler AJ, Lowell JL, Gage KL, Morway C, et al. (2009) Fine-scale identification of the most likely source of a human plague infection. Emerg Infect Dis 15: 1623-1625.
173. Lowell JL, Wagner DM, Atshabar B, Antolin MF, Vogler AJ, et al. (2005) Identifying sources of human exposure to plague. J Clin Microbiol 43: 650-656.
174. Brisse S, Pannier C, Angoulvant A, de Meeus T, Diancourt L, et al. (2009) Uneven distribution of mating types among genotypes of *Candida glabrata* isolates from clinical samples. Eukaryot Cell 8: 287-295.
175. Dettman JR, Taylor JW (2004) Mutation and evolution of microsatellite loci in Neurospora. Genetics 168: 1231-1248.
176. Yokoyama E, Hachisu Y, Hashimoto R, Kishida K (2010) Concordance of variable-number tandem repeat (VNTR) and large sequence polymorphism (LSP) analyses of *Mycobacterium tuberculosis* strains. Infect Genet Evol 10: 913-918.
177. Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Mol Ecol 11: 1591-1604.
178. de Wit R, Bouvier T (2006) 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? Environmental Microbiology 8: 755-758.
179. Valen LV (1976) Ecological Species, Multispecies, and Oaks. Taxon 25: 233.
180. Cohan FM, Perry EB (2007) A systematics for discovering the fundamental units of bacterial diversity. Curr Biol 17: R373-386.
181. Nurminsky D (2005) Selective sweep. Georgetown, Tex. New York, N.Y.: Landes Bioscience/Eurekah.com ; Kluwer Academic/Plenum Publishers. 121 p. p.
182. Achtman M (2004) Population structure of pathogenic bacteria revisited. Int J Med Microbiol 294: 67-73.
183. Cohan FM (2001) Bacterial species and speciation. Syst Biol 50: 513-524.

Annexes

Figures et tableaux supplémentaires de l'article 1

Table S1. Characteristics of *L. pneumophila* isolates used in this study.

Source	Sampling date	Equipment	Facility	SBT	ST
Environment	31_03_2000	Hot water supplies	Swimming-pool A		
Environment	26_02_2001	Hot water supplies	Students house A		
Environment	26_02_2001	Hot water supplies	Students house A		
Environment	26_02_2001	Hot water supplies	Students house A		
Environment	26_02_2001	Hot water supplies	Students house A		
Environment	26_03_2001	Hot water supplies	Students house A		
Environment	26_03_2001	Hot water supplies	Students house A		
Environment	07_06_2001	Hot water supplies	Swimming-pool A		
Environment	07_06_2001	Hot water supplies	Swimming-pool B		
Environment	07_06_2001	Hot water supplies	Swimming-pool C		
Environment	11_06_2001	Hot water supplies	Swimming-pool D		
Environment	21_06_2001	Hot water supplies	Swimming-pool E		
Environment	05_07_2001	Hot water supplies	Swimming-pool A		
Environment	19_07_2001	Hot water supplies	Swimming-pool A		
Environment	25_07_2001	Hot water supplies	Swimming-pool A		
Environment	06_08_2001	Hot water supplies	Swimming-pool A		
Environment	13_08_2001	Hot water supplies	Swimming-pool A		
Environment	13_08_2001	Hot water supplies	Swimming-pool A		
Environment	13_08_2001	Hot water supplies	Swimming-pool A		
Environment	13_08_2001	Hot water supplies	Swimming-pool A		
Environment	06_09_2001	Hot water supplies	Health facility A		
Environment	06_09_2001	Hot water supplies	Health facility A		
Environment	06_09_2001	Hot water supplies	Health facility A		
Environment	10_09_2001	Cooling towers	Factory A		
Environment	10_09_2001	Cooling towers	Factory A		
Environment	10_09_2001	Cooling towers	Factory A		
Environment	10_09_2001	Cooling towers	Factory A		
Environment	24_09_2001	Hot water supplies	Mall B		
Environment	24_09_2001	Hot water supplies	Mall B		
Environment	26_09_2001	Cooling towers	NA		
Environment	03_10_2001	Cooling towers	Factory A		
Environment	08_10_2001	Cooling towers	Factory A		
Environment	03_10_2001	Hot water supplies	Geriatric center A	1,4,3,1,1,1,1	1
Environment	03_10_2001	Hot water supplies	Geriatric center A		
Environment	08_10_2001	Cooling towers	Factory B		
Environment	08_10_2001	Hot water supplies	Health facility B		
Environment	08_10_2001	Hot water supplies	Health facility C		
Environment	03_10_2001	Hot water supplies	Geriatric center A		
Environment	18_10_2001	Hot water supplies	Hotel A		
Environment	18_10_2001	Hot water supplies	Hotel B		
Environment	23_10_2001	Hot water supplies	Private A		
Environment	23_10_2001	Hot water supplies	Health facility D		
Environment	23_10_2001	Hot water supplies	Health facility E		
Environment	23_10_2001	Hot water supplies	Private B		

Environment	25_10_2001	Hot water supplies	Health facility F		
Environment	25_10_2001	Hot water supplies	Health facility G		
Environment	30_10_2001	Hot water supplies	Health facility H		
Environment	14_11_2001	Hot water supplies	Private C		
Environment	20_12_2001	Hot water supplies	Swimming-pool A		
Environment	10_01_2002	Hot water supplies	Maintenance facility A	3,6,1,6,14,11,9	114
Environment	10_01_2002	Hot water supplies	Maintenance facility A		
Environment	14_01_2002	Hot water supplies	Hotel C		
Environment	22_01_2002	Cooling towers	Factory B		
Environment	15_01_2002	Hot water supplies	Hotel D		
Environment	13_02_2002	Hot water supplies	Nursing home A		
Environment	13_02_2002	Hot water supplies	Nursing home A		
Environment	13_02_2002	Hot water supplies	Nursing home A		
Environment	13_02_2002	Hot water supplies	Nursing home A		
Environment	19_02_2002	Hot water supplies	Gymnasium C		
Environment	19_02_2002	Hot water supplies	Private D		
Environment	28_05_2002	Cooling towers	Factory B		
Environment	02_06_2002	Hot water supplies	Gymnasium D		
Environment	02_06_2002	Hot water supplies	Gymnasium D		
Environment	11_06_2002	Hot water supplies	Swimming-pool E		
Environment	06_01_2003	Hot water supplies	Students house B		
Environment	06_01_2003	Hot water supplies	Students house A		
Environment	05_05_2003	Cooling towers	Factory B		
Environment	14_05_2003	Hot water supplies	School A		
Environment	14_05_2003	Hot water supplies	School A		
Environment	14_05_2003	Hot water supplies	Gymnasium E		
Environment	15_05_2003	Hot water supplies	Gymnasium F		
Environment	15_05_2003	Hot water supplies	School A		
Environment	15_05_2003	Hot water supplies	Gymnasium F		
Environment	15_09_2003	Hot water supplies	Gymnasium G		
Environment	15_09_2003	Hot water supplies	Gymnasium G		
Environment	15_09_2003	Hot water supplies	Gymnasium G		
Environment	17_09_2003	Hot water supplies	Swimming-pool F	7,6,17,3,13,11,11	59
Environment	17_09_2003	Hot water supplies	Swimming-pool F		
Environment	17_09_2003	Hot water supplies	Gymnasium H		
Environment	17_09_2003	Hot water supplies	Gymnasium H		
Environment	17_09_2003	Hot water supplies	Gymnasium E		
Environment	17_09_2003	Hot water supplies	Gymnasium I		
Environment	17_09_2003	Hot water supplies	Gymnasium J		
Environment	17_09_2003	Hot water supplies	Gymnasium K		
Environment	22_09_2003	Hot water supplies	Students house A		
Environment	28_10_2003	Hot water supplies	Swimming-pool G		
Environment	14_11_2003	Hot water supplies	Swimming-pool G		
Environment	15_12_2003	Hot water supplies	Swimming-pool G		
Environment	15_12_2003	Hot water supplies	Swimming-pool G		
Environment	15_12_2003	Hot water supplies	Swimming-pool G		
Environment	12_01_2004	Hot water supplies	Students house C		
Environment	12_01_2004	Hot water supplies	Students house C		
Environment	12_01_2004	Hot water supplies	Students house C		
Environment	12_01_2004	Hot water supplies	Students house C		
Environment	12_01_2004	Hot water supplies	Students house C		
Environment	12_01_2004	Hot water supplies	Students house C		
Environment	12_01_2004	Hot water supplies	Students house C		
Environment	26_01_2004	Hot water supplies	Swimming-pool B		

Environment	26_01_2004	Hot water supplies	Gymnasium I
Environment	26_01_2004	Hot water supplies	Gymnasium I
Environment	26_01_2004	Hot water supplies	Gymnasium I
Environment	26_01_2004	Hot water supplies	Gymnasium I
Environment	26_01_2004	Hot water supplies	Gymnasium E
Environment	26_01_2004	Hot water supplies	Gymnasium F
Environment	26_01_2004	Hot water supplies	Gymnasium F
Environment	26_01_2004	Hot water supplies	Gymnasium F
Environment	26_01_2004	Hot water supplies	Gymnasium F
Environment	28_01_2004	Hot water supplies	Gymnasium G
Environment	28_01_2004	Hot water supplies	Gymnasium G
Environment	28_01_2004	Hot water supplies	Gymnasium G
Environment	28_01_2004	Hot water supplies	Gymnasium G
Environment	28_01_2004	Hot water supplies	Gymnasium G
Environment	28_01_2004	Hot water supplies	Gymnasium G
Environment	28_01_2004	Hot water supplies	Gymnasium G
Environment	28_01_2004	Hot water supplies	Gymnasium L
Environment	28_01_2004	Hot water supplies	Gymnasium M
Environment	28_01_2004	Hot water supplies	Gymnasium N
Environment	28_01_2004	Hot water supplies	Gymnasium O
Environment	28_01_2004	Hot water supplies	Gymnasium O
Environment	29_01_2004	Hot water supplies	Gymnasium A
Environment	29_01_2004	Hot water supplies	Gymnasium A
Environment	29_01_2004	Hot water supplies	Gymnasium P
Environment	29_01_2004	Hot water supplies	Gymnasium B
Environment	29_01_2004	Hot water supplies	Gymnasium B
Environment	18_02_2004	Hot water supplies	Military facility A
Environment	26_02_2004	Hot water supplies	Private E
Environment	23_03_2004	Hot water supplies	Association A
Environment	23_03_2004	Hot water supplies	Association B
Environment	23_03_2004	Hot water supplies	Association B
Environment	23_03_2004	Hot water supplies	Association B
Environment	23_03_2004	Hot water supplies	Association B
Environment	28_06_2004	Hot water supplies	Swimming-pool A
Environment	28_06_2004	Hot water supplies	Gymnasium L
Environment	28_06_2004	Hot water supplies	Gymnasium L
Environment	28_06_2004	Hot water supplies	Gymnasium M
Environment	28_06_2004	Hot water supplies	Swimming-pool C
Environment	12_07_2004	Hot water supplies	Swimming-pool F
Environment	05_08_2004	Hot water supplies	Gymnasium L
Environment	05_08_2004	Hot water supplies	Gymnasium L
Environment	05_08_2004	Hot water supplies	Gymnasium M
Environment	05_08_2004	Hot water supplies	Swimming-pool A
Environment	05_08_2004	Hot water supplies	Swimming-pool A
Environment	27_09_2004	Hot water supplies	Gymnasium F
Environment	27_09_2004	Hot water supplies	Gymnasium F
Environment	27_09_2004	Hot water supplies	Gymnasium F
Environment	27_09_2004	Hot water supplies	Swimming-pool A
Environment	27_09_2004	Hot water supplies	Swimming-pool C
Environment	27_09_2004	Hot water supplies	Swimming-pool C
Environment	27_09_2004	Hot water supplies	Gymnasium I
Environment	27_09_2004	Hot water supplies	Gymnasium I
Environment	27_09_2004	Hot water supplies	Gymnasium I
Environment	27_09_2004	Hot water supplies	Gymnasium I

Environment	27_09_2004	Hot water supplies	Gymnasium N		
Environment	27_09_2004	Hot water supplies	Gymnasium L		
Environment	20_10_2004	Hot water supplies	Gymnasium Q		
Environment	20_10_2004	Hot water supplies	Gymnasium Q		
Environment	20_10_2004	Hot water supplies	Gymnasium Q		
Environment	20_10_2004	Hot water supplies	Gymnasium Q		
Environment	21_10_2004	Hot water supplies	Gymnasium F		
Environment	21_10_2004	Hot water supplies	Gymnasium R		
Environment	21_10_2004	Hot water supplies	Gymnasium R		
Environment	14_10_2004	Hot water supplies	Laboratory A		
Environment	28_10_2004	Hot water supplies	Swimming-pool A		
Environment	07_12_2004	Hot water supplies	Swimming-pool H		
Environment	02_02_2005	Hot water supplies	Swimming-pool B		
Environment	02_02_2005	Hot water supplies	Swimming-pool B		
Environment	02_02_2005	Hot water supplies	Swimming-pool B		
Environment	02_02_2005	Hot water supplies	Gymnasium S		
Environment	02_02_2005	Hot water supplies	Gymnasium S		
Environment	02_02_2005	Hot water supplies	Gymnasium S		
Environment	02_02_2005	Hot water supplies	Gymnasium A		
Environment	02_02_2005	Hot water supplies	Gymnasium A		
Environment	07_03_2005	Hot water supplies	Military facility A		
Environment	14_03_2005	Hot water supplies	Gymnasium F		
Environment	14_03_2005	Hot water supplies	Gymnasium F		
Environment	16_03_2005	Hot water supplies	Gymnasium T		
Environment	16_03_2005	Hot water supplies	Gymnasium T		
Environment	16_03_2005	Hot water supplies	Gymnasium U		
Environment	03_10_2005	Hot water supplies	Nursing home B	5,1,22,30,6,10,6	74
Environment	03_10_2005	Hot water supplies	Nursing home B		
Environment	28_02_2006	Hot water supplies	Laboratory A		
Environment	28_02_2006	Hot water supplies	Laboratory A		
Environment	23_02_2006	Hot water supplies	Laboratory A		
Environment	03_03_2006	Hot water supplies	Laboratory A		
Environment	03_04_2006	Hot water supplies	Laboratory A		
Environment	17_07_2006	Hot water supplies	INSEE A		
Environment	24_07_2006	Cooling towers	Hospital A		
Environment	24_07_2006	Cooling towers	Hospital A		
Environment	16_08_2006	Cooling towers	Hospital A		
Environment	16_08_2006	Cooling towers	Hospital A		
Environment	30_08_2006	Cooling towers	Hospital A		
Environment	30_08_2006	Cooling towers	Hospital A	1,4,3,1,1,1,1	1
Environment	13_09_2006	Cooling towers	Hospital A		
Environment	03_01_2007	Cooling towers	Hospital A		
Environment	03_01_2007	Cooling towers	Hospital A		
Environment	14_01_2008	Cooling towers	Hospital A		
Environment	16_04_2008	Cooling towers	Hospital A		
Environment	08_10_2008	Cooling towers	Hospital A	1,4,3,1,1,1,1	1
Environment	08_10_2008	Cooling towers	Hospital A		
Environment	17_10_2008	Hot water supplies	Students house A		
Environment	17_10_2008	Hot water supplies	Students house A		
Environment	20_10_2008	Hot water supplies	Military facility B		
Environment	29_10_2008	Hot water supplies	Military facility C		
Environment	29_10_2008	Hot water supplies	Military facility D		
Environment	29_10_2008	Hot water supplies	Military facility E		
Environment	13_11_2008	Hot water supplies	Military facility F		

Environment	19_01_2009	Hot water supplies	Swimming-pool I		
Environment	25_02_2009	Hot water supplies	Swimming-pool I		
Environment	26_02_2009	Hot water supplies	Swimming-pool I		
Environment	02_03_2009	Hot water supplies	Swimming-pool J		
Environment	25_03_2009	Hot water supplies	Swimming-pool J		
Environment	30_03_2009	Hot water supplies	Swimming-pool K		
Environment		Cooling towers	Mall A		
Environment		Cooling towers	Mall A		
Environment		Cooling towers	Mall A		
Environment		Cooling towers	Mall A		
Environment		Cooling towers	Mall A	3,4,1,28,14,11,11	439
Environment		Cooling towers	Night club A		
Environment		Cooling towers	Night club A		
Environment		Cooling towers	Bank A		
Clinical	31_08_2000	Clinical case		3,4,1,28,14,11,1	626
Clinical	15_11_2000	Clinical case		3,4,1,28,14,11,11	439
Clinical	08_09_2000	Clinical case		3,4,1,1,14,11,30	628
Clinical	31_10_2000	Clinical case		3,4,1,1,14,11,30	628
Clinical	03_01_2006	Clinical case		3,4,1,28,14,11,11	439
Environment	22_12_2005	Cooling towers	Bank A		
Environment	16_06_2009	Cooling towers	Hospital A		
Environment	16_06_2009	Cooling towers	Hospital A		

Table S2. Repeatability test on EUL146 using MLVA-12$_{Orsay}$.

EUL146	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	Standard deviation 95%
Lpms01	456,37	456,44	456,37	457,01	456,75	457,13	457,32	457,03	{{455,76;457,85}
Lpms03	837,05	836,86	837,05	837,1	837,3	837,03	837,2	837,22	{836,72;837,49}
Lpms13	762,95	763,06	763,21	763,08	763,34	763,03	763,27	763,3	{762,75;763,56}
Lpms19	122,71	122,74	122,87	122,83	124,09	123,03	122,7	124,05	{121,47;124,78}
Lpms31	695,14	695,07	695,65	695,36	696,28	696,34	696,43	695,09	{694,01;697,33}
Lpms33	688,42	688,29	688,5	688,44	688,78	688,55	688,6	688,62	{688,11;688,94}
Lpms34	383,86	384,16	384,28	384,19	384,39	384,55	384,57	383,84	{383,45;385,01}
Lpms35	451,26	451,37	451,22	451,62	451,9	451,8	451,21	451,3	{450,69;452,23}
Lpms38	260,47	260,47	260,64	260,58	260,79	260,8	260,46	260,52	{260,2;260,98}
Lpms39	128,74	128,86	128,47	128,6	128,64	128,68	128,5	128,52	{128,26;129}
Lpms40	202,41	202,39	202,5	202,42	202,56	202,61	202,63	202,41	{202,22;202,76}
Lpms44	157,76	157,7	157,78	157,87	157,9	157,62	157,58	157,93	{157,41;158,13}

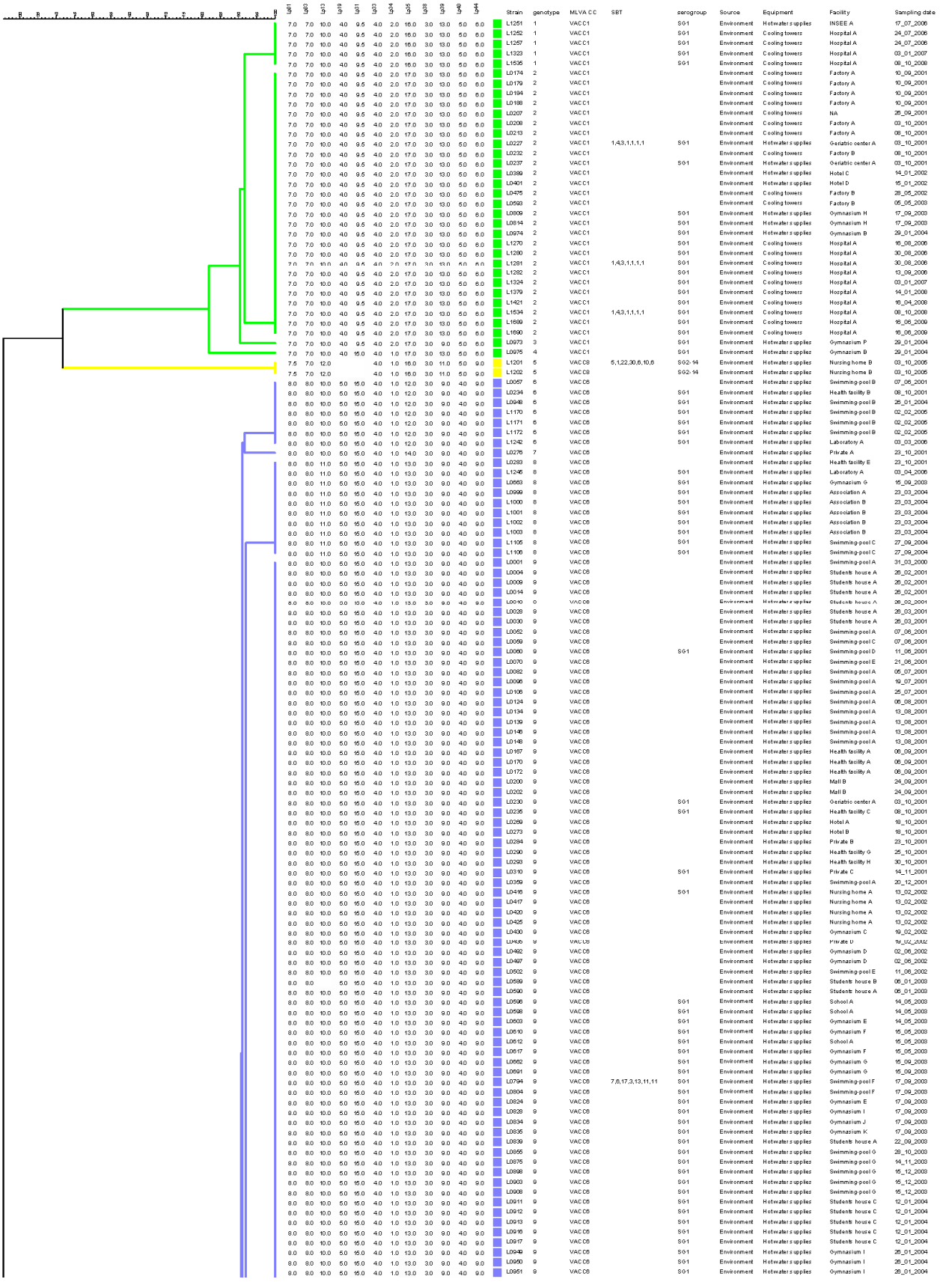
Table S3. Repeatability test on EUL160 using MLVA-12$_{Orsay}$.

EUL160	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	Standard deviation 95%
Lpms01	410,1	410,39	410,18	410,15	409,88	410,28	410,25	409,98	{409,69;410,61}
Lpms03	836,72	836,58	836,69	836,64	836,6	836,7	836,8	836,9	{836,41;837}
Lpms13	620,17	620,76	620,42	620,29	620,53	620,47	620,27	620,56	{619,9;620,96}
Lpms19	122,42	122,25	122,44	122,35	122,39	122,34	122,37	122,49	{122,18;122,58}
Lpms31	871,63	871,64	871,64	871,39	871,61	871,69	871,49	871,51	{871,29;871,86}
Lpms33	437,83	437,78	437,81	437,89	437,68	437,71	437,79	437,69	{437,57;437,98}
Lpms34	504,47	504,76	504,3	504,32	504,7	504,4	504,39	504,72	{503,98;505,04}
Lpms35	556,71	557,01	556,81	556,86	556,83	556,81	556,89	556,93	{556,6;557,11}
Lpms38	316,08	315,96	315,98	315,98	316,02	315,98	316	316,08	{315,88;316,14}
Lpms39	111,23	111,08	111,02	111,13	111,23	111,07	111,27	111,29	{110,88;111,45}
Lpms40	196,22	196,23	196,18	196,15	196,1	196,28	196,16	196,18	{196,03;196,34}
Lpms44	176,01	176,04	175,65	175,72	175,73	175,89	175,92	175,89	{175,46;176,25}

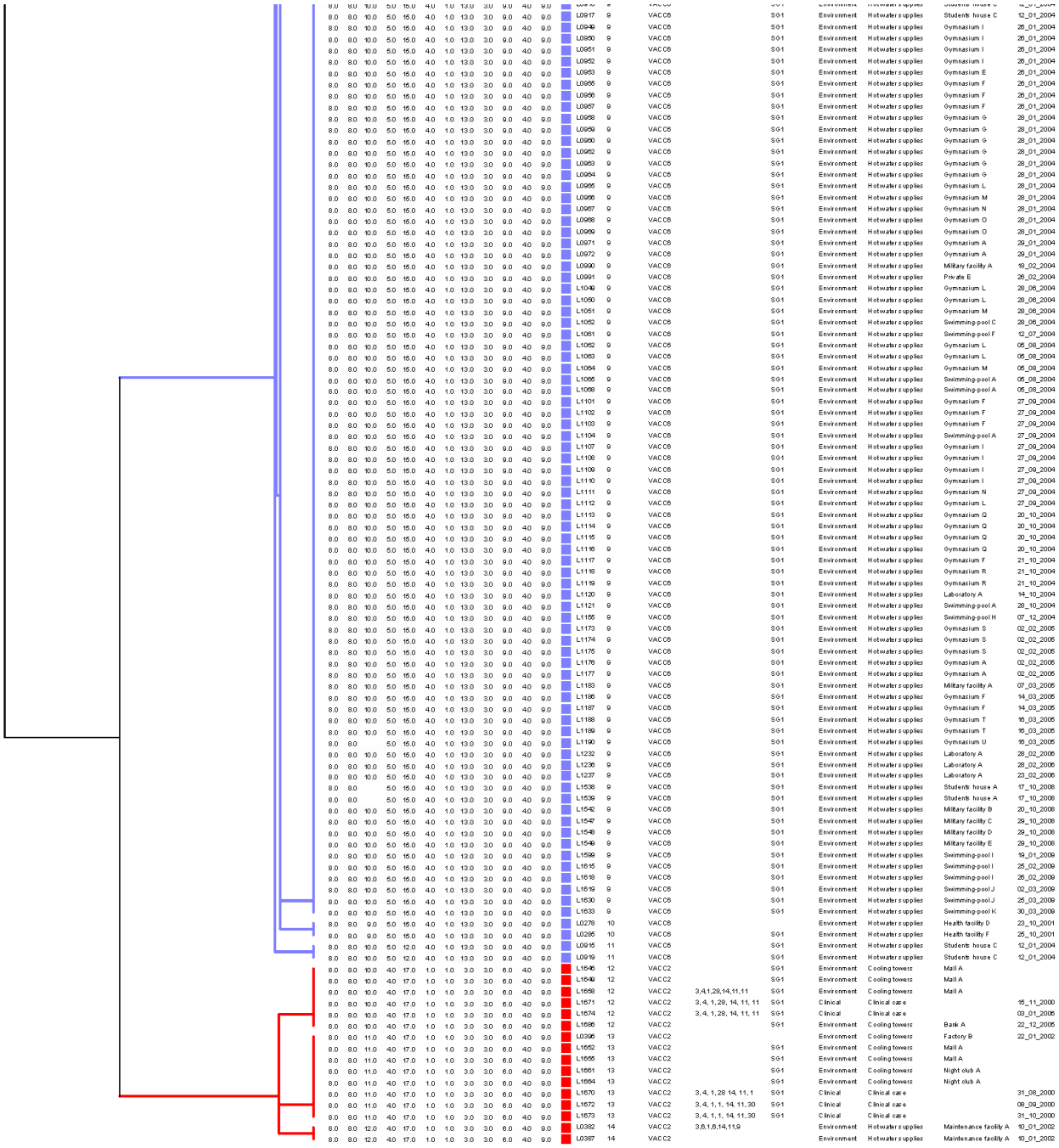
Table S4. Repeatability test on NCTC 11192 using MLVA-12$_{Orsay}$.

NCTC 11192	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	Standard deviation 95%
Lpms01	503,09	503,34	503,6	503,23	503,01	503,39	503,4	503,53	{502,75;503,89}
Lpms03	931,65	931,99	931,34	931,74	931,55	931,89	931,34	931,64	{930,99;932,3}
Lpms13	785,67	785,38	785,57	785,74	785,47	785,48	785,57	785,78	{785,19;785,97}
Lpms19	123,82	123,92	123,74	123,76	123,89	123,92	123,7	123,79	{123,58;124,05}
Lpms31	1023,43	1023,83	1023,68	1023,55	1023,43	1023,88	1023,78	1023,65	{1023,17;1024,14}
Lpms33	313,65	313,76	313,65	313,71	313,66	313,76	313,55	313,73	{313,49;313,88}
Lpms34	258,19	258,2	258,36	258,24	258,18	258,29	258,38	258,14	{258;258,49}
Lpms35	203,37	203,54	203,28	203,35	203,39	203,54	203,29	203,45	{203,12;203,68}
Lpms38	260,47	260,53	260,53	260,72	260,49	260,54	260,59	260,79	{260,26;260,9}
Lpms39	80,53	80,54	80,54	80,46	80,53	80,59	80,59	80,36	{80,31;80,73}
Lpms40	196,25	196,03	196,25	196,29	196,35	196,13	196,27	196,22	{195,94;196,5}
Lpms44	175,86	175,89	175,82	175,8	175,7	175,65	175,99	175,92	{175,51;176,14}

Fig. S1. Dendrogram deduced from the clustering analysis of the 222 isolates from the Rennes' study using MLVA-12$_{Orsay}$. The colours used to show MLVA clusters (VACC) are the same as in Fig. 2.



Strain	genotype	MLVA CC	SBT	serogroup	Source	Equipment	Facility	Sampling date
L1251	1	VACC1		SG1	Environment	Hotwater supplies	INSEE A	17_07_2006
L1252	1	VACC1		SG1	Environment	Cooling towers	Hospital A	24_07_2006
L1257	1	VACC1		SG1	Environment	Cooling towers	Hospital A	24_07_2006
L1323	1	VACC1		SG1	Environment	Cooling towers	Hospital A	03_01_2007
L1535	1	VACC1		SG1	Environment	Cooling towers	Hospital A	08_10_2008
L0174	2	VACC1			Environment	Cooling towers	Factory A	10_09_2001
L0179	2	VACC1			Environment	Cooling towers	Factory A	10_09_2001
L0184	2	VACC1			Environment	Cooling towers	Factory A	10_09_2001
L0188	2	VACC1			Environment	Cooling towers	Factory A	10_09_2001
L0207	2	VACC1			Environment	Cooling towers	NA	25_09_2001
L0208	2	VACC1			Environment	Cooling towers	Factory A	03_10_2001
L0213	2	VACC1			Environment	Cooling towers	Factory A	08_10_2001
L0227	2	VACC1	1,43,1,1,1,1	SG1	Environment	Hotwater supplies	Geriatric center A	03_10_2001
L0232	2	VACC1			Environment	Cooling towers	Factory B	08_10_2001
L0237	2	VACC1		SG1	Environment	Hotwater supplies	Geriatric center A	03_10_2001
L0389	2	VACC1			Environment	Hotwater supplies	Hotel C	14_01_2002
L0401	2	VACC1			Environment	Hotwater supplies	Hotel D	15_01_2002
L0475	2	VACC1			Environment	Cooling towers	Factory B	28_05_2002
L0593	2	VACC1			Environment	Cooling towers	Factory B	05_05_2003
L0809	2	VACC1		SG1	Environment	Hotwater supplies	Gymnasium H	17_09_2003
L0814	2	VACC1		SG1	Environment	Hotwater supplies	Gymnasium H	17_09_2003
L0974	2	VACC1		SG1	Environment	Hotwater supplies	Gymnasium B	29_01_2004
L1270	2	VACC1		SG1	Environment	Cooling towers	Hospital A	16_08_2006
L1280	2	VACC1		SG1	Environment	Cooling towers	Hospital A	30_08_2006
L1281	2	VACC1	1,43,1,1,1,1	SG1	Environment	Cooling towers	Hospital A	30_08_2006
L1282	2	VACC1		SG1	Environment	Cooling towers	Hospital A	13_09_2006
L1324	2	VACC1		SG1	Environment	Cooling towers	Hospital A	03_01_2007
L1379	2	VACC1		SG1	Environment	Cooling towers	Hospital A	14_01_2008
L1421	2	VACC1		SG1	Environment	Cooling towers	Hospital A	16_04_2008
L1534	2	VACC1	1,43,1,1,1,1	SG1	Environment	Cooling towers	Hospital A	08_10_2008
L1689	2	VACC1		SG1	Environment	Cooling towers	Hospital A	16_06_2009
L1690	2	VACC1		SG1	Environment	Cooling towers	Hospital A	16_06_2009
L0973	3	VACC1		SG1	Environment	Hotwater supplies	Gymnasium P	29_01_2004
L0975	4	VACC1		SG1	Environment	Hotwater supplies	Gymnasium B	29_01_2004
L1201	5	VACC8	5,1,22,30,6,10,8	SG2-14	Environment	Hotwater supplies	Nursing home B	03_10_2005
L1202	5	VACC8		SG2-14	Environment	Hotwater supplies	Nursing home B	03_10_2005
L0057	6	VACC6			Environment	Hotwater supplies	Swimming-pool B	07_06_2001
L0234	6	VACC6		SG1	Environment	Hotwater supplies	Health facility B	08_10_2001
L0948	6	VACC6		SG1	Environment	Hotwater supplies	Swimming-pool B	26_01_2004
L1170	6	VACC6		SG1	Environment	Hotwater supplies	Swimming-pool B	02_02_2005
L1171	6	VACC6		SG1	Environment	Hotwater supplies	Swimming-pool B	02_02_2005
L1172	6	VACC6		SG1	Environment	Hotwater supplies	Swimming-pool B	02_02_2005
L1242	6	VACC6		SG1	Environment	Hotwater supplies	Laboratory A	03_03_2006
L0276	7	VACC6			Environment	Hotwater supplies	Private A	23_10_2001
L0283	8	VACC6			Environment	Hotwater supplies	Health facility E	23_10_2001
L1246	8	VACC6		SG1	Environment	Hotwater supplies	Laboratory A	03_04_2006
L0663	8	VACC6		SG1	Environment	Hotwater supplies	Gymnasium G	15_09_2003
L0999	8	VACC6		SG1	Environment	Hotwater supplies	Association A	23_03_2004
L1000	8	VACC6		SG1	Environment	Hotwater supplies	Association B	23_03_2004
L1001	8	VACC6		SG1	Environment	Hotwater supplies	Association B	23_03_2004
L1002	8	VACC6		SG1	Environment	Hotwater supplies	Association B	23_03_2004
L1003	8	VACC6		SG1	Environment	Hotwater supplies	Association B	23_03_2004
L1105	8	VACC6		SG1	Environment	Hotwater supplies	Swimming-pool C	27_09_2004
L1106	8	VACC6		SG1	Environment	Hotwater supplies	Swimming-pool C	27_09_2004
L0001	9	VACC6			Environment	Hotwater supplies	Swimming-pool A	31_03_2000
L0004	9	VACC6			Environment	Hotwater supplies	Students house A	26_02_2001
L0009	9	VACC6			Environment	Hotwater supplies	Students house A	26_02_2001
L0014	9	VACC6			Environment	Hotwater supplies	Students house A	26_02_2001
L0010	9	VACC6			Environment	Hotwater supplies	Students house A	26_02_2001
L0028	9	VACC6			Environment	Hotwater supplies	Students house A	26_03_2001
L0030	9	VACC6			Environment	Hotwater supplies	Students house A	26_03_2001
L0052	9	VACC6			Environment	Hotwater supplies	Swimming-pool A	07_06_2001
L0059	9	VACC6			Environment	Hotwater supplies	Swimming-pool C	07_06_2001
L0060	9	VACC6		SG1	Environment	Hotwater supplies	Swimming-pool D	11_06_2001
L0070	9	VACC6			Environment	Hotwater supplies	Swimming-pool E	21_06_2001
L0082	9	VACC6			Environment	Hotwater supplies	Swimming-pool A	05_07_2001
L0096	9	VACC6			Environment	Hotwater supplies	Swimming-pool A	19_07_2001
L0106	9	VACC6			Environment	Hotwater supplies	Swimming-pool A	25_07_2001
L0124	9	VACC6			Environment	Hotwater supplies	Swimming-pool A	06_08_2001
L0134	9	VACC6			Environment	Hotwater supplies	Swimming-pool A	13_08_2001
L0139	9	VACC6			Environment	Hotwater supplies	Swimming-pool A	13_08_2001
L0145	9	VACC6			Environment	Hotwater supplies	Swimming-pool A	13_08_2001
L0148	9	VACC6			Environment	Hotwater supplies	Swimming-pool A	13_08_2001
L0167	9	VACC6			Environment	Hotwater supplies	Health facility A	06_09_2001
L0170	9	VACC6			Environment	Hotwater supplies	Health facility A	06_09_2001
L0172	9	VACC6			Environment	Hotwater supplies	Health facility A	06_09_2001
L0200	9	VACC6			Environment	Hotwater supplies	Mall B	24_09_2001
L0202	9	VACC6			Environment	Hotwater supplies	Mall B	24_09_2001
L0230	9	VACC6		SG1	Environment	Hotwater supplies	Geriatric center A	03_10_2001
L0235	9	VACC6		SG1	Environment	Hotwater supplies	Health facility C	08_10_2001
L0269	9	VACC6			Environment	Hotwater supplies	Hotel A	18_10_2001
L0273	9	VACC6			Environment	Hotwater supplies	Hotel B	18_10_2001
L0284	9	VACC6			Environment	Hotwater supplies	Private B	23_10_2001
L0290	9	VACC6			Environment	Hotwater supplies	Health facility G	25_10_2001
L0293	9	VACC6			Environment	Hotwater supplies	Health facility H	30_10_2001
L0310	9	VACC6		SG1	Environment	Hotwater supplies	Private C	14_11_2001
L0359	9	VACC6			Environment	Hotwater supplies	Swimming-pool A	20_12_2001
L0416	9	VACC6		SG1	Environment	Hotwater supplies	Nursing home A	13_02_2002
L0417	9	VACC6			Environment	Hotwater supplies	Nursing home A	13_02_2002
L0420	9	VACC6			Environment	Hotwater supplies	Nursing home A	13_02_2002
L0425	9	VACC6			Environment	Hotwater supplies	Nursing home A	13_02_2002
L0430	9	VACC6			Environment	Hotwater supplies	Gymnasium C	19_02_2002
L0435	9	VACC6			Environment	Hotwater supplies	Private D	19_02_2002
L0482	9	VACC6			Environment	Hotwater supplies	Gymnasium D	02_06_2002
L0497	9	VACC6			Environment	Hotwater supplies	Gymnasium D	02_06_2002
L0502	9	VACC6			Environment	Hotwater supplies	Swimming-pool E	11_06_2002
L0589	9	VACC6			Environment	Hotwater supplies	Students house B	06_01_2003
L0590	9	VACC6			Environment	Hotwater supplies	Students house A	06_01_2003
L0596	9	VACC6		SG1	Environment	Hotwater supplies	School A	14_05_2003
L0598	9	VACC6		SG1	Environment	Hotwater supplies	School A	14_05_2003
L0603	9	VACC6		SG1	Environment	Hotwater supplies	Gymnasium E	14_05_2003
L0610	9	VACC6		SG1	Environment	Hotwater supplies	Gymnasium F	15_05_2003
L0612	9	VACC6		SG1	Environment	Hotwater supplies	School A	15_05_2003
L0617	9	VACC6		SG1	Environment	Hotwater supplies	Gymnasium F	15_05_2003
L0662	9	VACC6		SG1	Environment	Hotwater supplies	Gymnasium G	15_09_2003
L0691	9	VACC6		SG1	Environment	Hotwater supplies	Gymnasium G	15_09_2003
L0794	9	VACC6	7,8,17,3,13,11,11	SG1	Environment	Hotwater supplies	Swimming-pool F	17_09_2003
L0804	9	VACC6		SG1	Environment	Hotwater supplies	Swimming-pool F	17_09_2003
L0824	9	VACC6		SG1	Environment	Hotwater supplies	Gymnasium E	17_09_2003
L0828	9	VACC6		SG1	Environment	Hotwater supplies	Gymnasium I	17_09_2003
L0834	9	VACC6		SG1	Environment	Hotwater supplies	Gymnasium J	17_09_2003
L0835	9	VACC6		SG1	Environment	Hotwater supplies	Gymnasium K	17_09_2003
L0839	9	VACC6		SG1	Environment	Hotwater supplies	Students house A	22_09_2003
L0855	9	VACC6		SG1	Environment	Hotwater supplies	Swimming-pool G	29_10_2003
L0875	9	VACC6		SG1	Environment	Hotwater supplies	Swimming-pool G	14_11_2003
L0898	9	VACC6		SG1	Environment	Hotwater supplies	Swimming-pool G	15_12_2003
L0903	9	VACC6		SG1	Environment	Hotwater supplies	Swimming-pool G	15_12_2003
L0908	9	VACC6		SG1	Environment	Hotwater supplies	Swimming-pool G	15_12_2003
L0911	9	VACC6		SG1	Environment	Hotwater supplies	Students house C	12_01_2004
L0912	9	VACC6		SG1	Environment	Hotwater supplies	Students house C	12_01_2004
L0913	9	VACC6		SG1	Environment	Hotwater supplies	Students house C	12_01_2004
L0916	9	VACC6		SG1	Environment	Hotwater supplies	Students house C	12_01_2004
L0917	9	VACC6		SG1	Environment	Hotwater supplies	Students house C	12_01_2004
L0949	9	VACC6		SG1	Environment	Hotwater supplies	Gymnasium I	26_01_2004
L0950	9	VACC6		SG1	Environment	Hotwater supplies	Gymnasium I	26_01_2004
L0951	9	VACC6		SG1	Environment	Hotwater supplies	Gymnasium I	26_01_2004



Figures et tableaux supplémentaires de l'article 2

Table S1. Isolates used in this study.

Strain-ID	Patient-ID	Sampling date	PFGE type	Phenotype
RD_367	RD01	01/04/99	1-1	psaa[1]
RD_22935	RD01	22/05/06	1-5	psm[2]
RD_2467	RD01	10/11/99	1-2	psm
RD_23118	RD01	16/06/06		psa[3]
RD_23119	RD01	16/06/06		psm
RD_24912	RD01	06/03/07		psm
RD_24913	RD01	06/03/07		psm
RD_22934	RD01	22/05/06	1-4	psm
RD_23119	RD01	16/06/06		psm
RD_1857	RD02	09/09/03	2-1	ps5[4]
RD_12290	RD02	28/10/02	2-3	psa
RD_14075	RD02	24/06/03	2-5	psa
RD_16900	RD02	05/05/04	2-7	psa
RD_23261	RD02	30/06/06		psaa
RD_24799	RD02	05/02/07		psa
RD_22892	RD02	15/05/06	2-10	psa
RD_14076	RD02	24/06/03	2-6	ps6[5]
RD_22892b	RD02	15/05/06		psa
RD_12290	RD02	28/10/02		psa
RD_16900	RD02	05/05/04		psa
RD_2913	RD02	28/12/02	2-2	ps11[6]
RD_13334	RD03	24/03/03	3-6	ps6
RD_16285	RD03	27/02/04	3-7	psa
RD_19586	RD03	03/03/05	3-10	psm
RD_22114	RD03	30/01/06	3-13	psm
RD_17159	RD03	03/06/04	3-8	psm
RD_23901	RD03	25/09/06		psm
RD_1603	RD04	13/08/99		psna
RD_3115	RD04	21/01/00	4-2	psm
RD_6601	RD04	19/01/01	4-4	psm
RD_10185	RD04	20/02/02	4-7	psm
RD_13849	RD04	26/03/05	4-8	psm
RD_15552	RD04	05/12/03	4-9	psm
RD_1238	RD05	24/06/99	5-1	psm
RD_6616	RD05	24/01/01	5-3	psm
RD_12463	RD05	20/11/02	5-4	psm
RD_17784	RD05	11/08/04	5-6	psm
RD_21788	RD05	20/12/05		psm

RD_23766	RD05	07/09/06		psm
RD_15710-1	RD05	22/12/03	5-5	psm
RD_15710-2	RD05	22/12/03		psm
RD_22821	RD05	06/05/06	5-7	psm
RD_23192	RD05			psa
RD_24779	RD05	01/02/07		psm
RD_24780	RD05	01/02/07		psa
RD_11021	RD06	30/05/02	1-9	psa
RD_20599	RD06	26/07/05	1-10	psaa
RD_24302	RD06	22/11/06		psa
RD_24301	RD06	22/12/06		psa
RD_22933	RD06	22/05/06	1-13	psm
RD_23269	RD06	04/07/06		psaa
RD_23270	RD06	04/07/06		psaa
RD_23271	RD06	04/07/06		psaa
RD_23291	RD06	04/07/06		psm
RD_24086	RD06	18/10/06		psa
RD_24724	RD06	20/12/06		psm
RD_24725	RD06	23/01/07		psm
RD_24980	RD06	21/03/07		psm
RD_24981	RD06	21/03/07		psa
RD_414	RD07	07/04/99	6-2	psa
RD_415	RD07	07/04/99	6-3	psa
RD_2943	RD07	04/01/00	6-4	psm
RD_2944	RD07	04/01/00	6-5	psa
RD_8020	RD07	13/06/01	6-6	psa
RD_8021	RD07	13/06/01		psm
RD_10951	RD07	17/05/02	6-7	psa
RD_10952	RD07	17/05/02	6-8	psm
RD_15129	RD07	18/10/03	6-11	ps6
RD_18325	RD07	08/10/03	6-13	psm
RD_18326	RD07	08/10/03	6-14	psa
RD_13371	RD07	27/03/03	6-9	psa
RD_13380	RD07	27/03/03	6-10	psm
RD_15136	RD07	18/10/03	6-12	psm
RD_1296	RD08	01/07/99	7-2	psm
RD_7311	RD08	05/04/01	7-3	psm
RD_14463	RD08	07/08/03	7-8	psm
RD_18549	RD08	04/11/04	7-9	psm
RD_21105	RD08	06/10/05	7-10	psm
RD_23604	RD08	16/08/06		psa
RD_23605	RD08	16/08/06		psm
RD_22702	RD08	19/04/06	7-11	psm
RD_24179	RD08	02/11/06		psa

RD_24800	RD08	08/02/07		psa
RD_12738	RD08	27/12/02	7-6	psm
RD_1295	RD08	01/07/99	7-1	ps6
RD_14462	RD08	07/08/03	7-7	psm
RD_12736	RD08	27/12/02	7-4	psa
RD_12737	RD08	27/12/02	7-5	psm
TR_S0702039	CFU24	25/06/07		psa
TR_S0801675	CFU24	09/06/08		psa
TR_S0502351	CFU24	05/10/05		psa
TR_S0502141	CFU24	14/09/05		ps3[7]
TR_S0502280	CFU24	28/09/05		ps6
TR_S0600710	CFU24	15/03/06		psa
TR_S0601347	CFU24	23/05/06		psa
TR_S0602589	CFU24	03/10/06		psa
TR_S0602733	CFU24	05/10/06		psa
TR_S0603715	CFU24	27/12/06		ps1[8]
TR_S0700480	CFU24	06/02/07		psa
PA14	CFU24			
PAO1	CFU24			
SW_C50	CFU24			
SW_SG17M	CFU24			

[1] psaa: autoagglutinable isolate
[2] psm: mucoid isolate
[3] psa: nonagglutinable isolate
[4] ps5: serotype 5 isolate
[5] ps6: serotype 6 isolate
[6] ps11: serotype 11 isolate
[7] ps3: serotype 3 isolate
[8] ps1: serotype 1 isolate

Table S2. Genotyping data * §

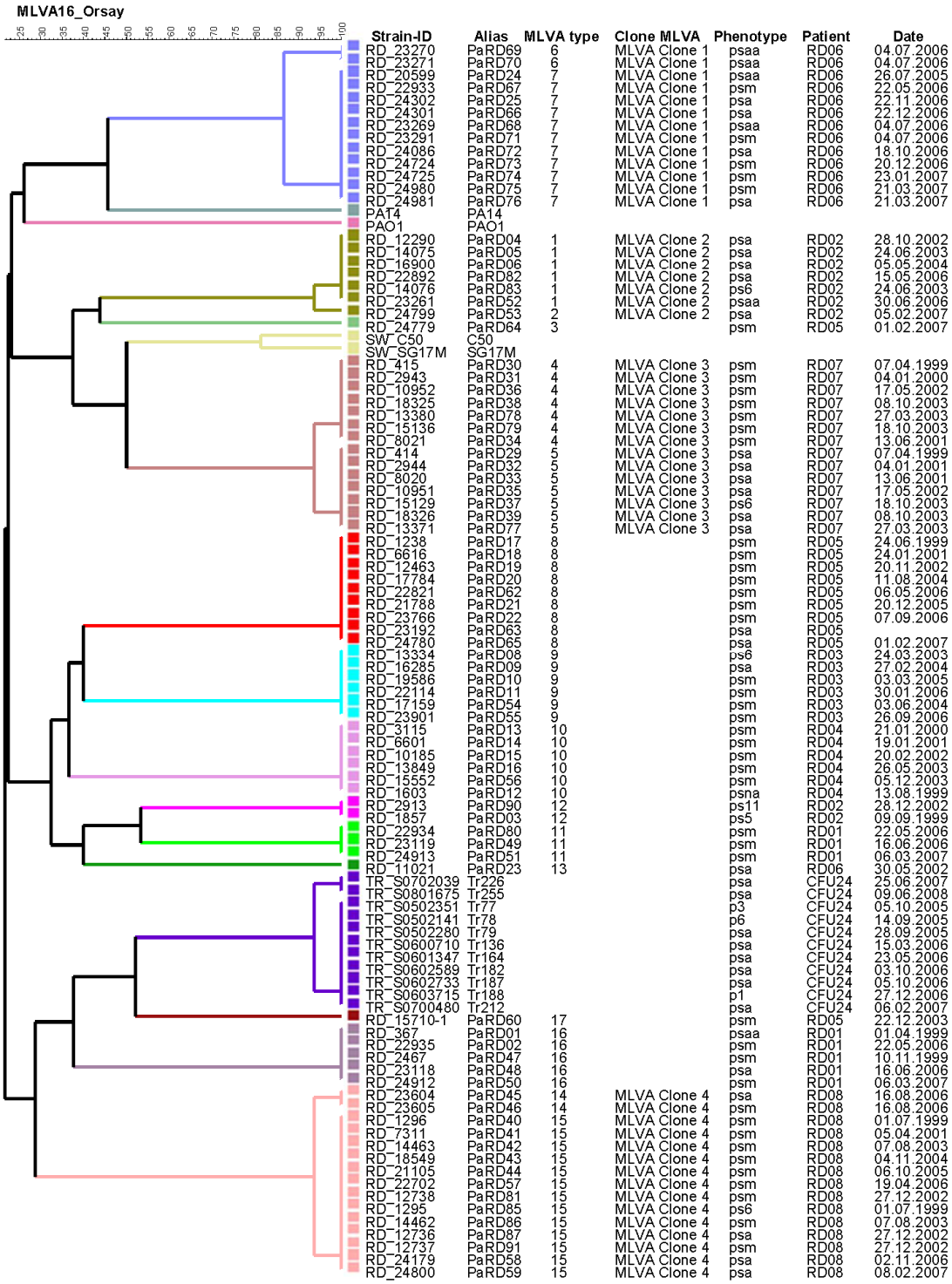
Strain-ID	ms77	ms127	ms142	ms172	ms211	ms212	ms213	ms214	ms215	ms216	ms217	ms222	ms223	ms61	ms207	ms209
RD_23270	2.5/3	7	4	12	2	8	3	5	2	2	2	2	0/2	13	5	4
RD_23271	2.5/3	7	4	12	2	8	3	5	2	2	2	2	0/2	13	5	4
RD_20599	2.5/3	7.5/8	4	12	2	8	3	5	2	2	2	2	0/2	12	5	4
RD_22933	2.5/3	7.5/8	4	12	2	8	3	5	2	2	2	2	0/2	12	5	4
RD_24302	2.5/3	7.5/8	4	12	2	8	3	5	2	2	2	2	0/2	12	5	4
RD_24301	2.5/3	7.5/8	4	12	2	8	3	5	2	2	2	2	0/2	12	5	4
RD_23269	2.5/3	7.5/8	4	12	2	8	3	5	2	2	2	2	0/2	12	5	4
RD_23291	2.5/3	7.5/8	4	12	2	8	3	5	2	2	2	2	0/2	12	5	4
RD_24086	2.5/3	7.5/8	4	12	2	8	3	5	2	2	2	2	0/2	12	5	4
RD_24724	2.5/3	7.5/8	4	12	2	8	3	5	2	2	2	2	0/2	12	5	4
RD_24725	2.5/3	7.5/8	4	12	2	8	3	5	2	2	2	2	0/2	12	5	4
RD_24980	2.5/3	7.5/8	4	12	2	8	3	5	2	2	2	2	0/2	12	5	4
RD_24981	2.5/3	7.5/8	4	12	2	8	3	5	2	2	2	2	0/2	12	5	4/3
PA14	2	9	1	12	2	4	1	5	2	1	5	2	4	12	5	6
PAO1	4	8	7	12	5	9	5	3	4	3	2	2	4	12	7	6
RD_12290	2.5/3	8	3	10	3	6	5	2	5	2	4	3	2	11	12	3
RD_14075	2.5/3	8	3	10	3	6	5	2	5	2	4	3	2	11	12	3
RD_16900	2.5/3	8	3	10	3	6	5	2	5	2	4	3	2	11	12	3
RD_22892	2.5/3	8	3	10	3	6	5	2	5	2	4	3	2	11	12	3
RD_14076	2.5/3	8	3	10	3	6	5	2	5	2	4	3	2	11	12	3
RD_23261	2.5/3	8	3	10	3	6	5	2	5	2	4	3	2	11	12	3
RD_24799	2.5/3	8	3	10	3	6	5	2	5	2	4	3	2	7	12	3
RD_24779	2.5/3	8	3	8.5	8	6	4.5	2	1	2	1	1	3	17	7/4	3
SW_C50	2.5/3	9	4	11	6	6	2	2	5	2	1	3	3	13	8	3
SW_SG17M	2.5/3	9	4	11	6	6	2	2	5	2	1	3	3	11	9	4
RD_415	2.5/3	9	4	10	5	9	2	2	2	2	1	2	2	13	8	3
RD_2943	2.5/3	9	4	10	5	9	2	2	2	2	1	2	2	13	8	3
RD_10952	2.5/3	9	4	10	5	9	2	2	2	2	1	2	2	13	8	3
RD_18325	2.5/3	9	4	10	5	9	2	2	2	2	1	2	2	13	8	3
RD_13380	2.5/3	9	4	10	5	9	2	2	2	2	1	2	2	13	8	3
RD_15136	2.5/3	9	4	10	5	9	2	2	2	2	1	2	2	13	8	3
RD_8021	2.5/3	9	4	10	5	9	2	2	2	2	1	2	2	13	8	3
RD_414	2.5/3	9	7	10	5	9	2	2	2	2	1	2	2	13	8	3
RD_2944	2.5/3	9	7	10	5	9	2	2	2	2	1	2	2	13	8	3
RD_8020	2.5/3	9	7	10	5	9	2	2	2	2	1	2	2	13	8	3
RD_10951	2.5/3	9	7	10	5	9	2	2	2	2	1	2	2	13	8	3
RD_15129	2.5/3	9	7	10	5	9	2	2	2	2	1	2	2	13	8	3
RD_18326	2.5/3	9	7	10	5	9	2	2	2	2	1	2	2	13	8	3
RD_13371	2.5/3	9	7	10	5	9	2	2	2	2	1	2	2	13	8	3
RD_1238	2.5/3	8	1	11	6	9	9/0	0/0	3	1	3	3	3	12	7	3
RD_6616	2.5/3	8	1	11	6	9	9/0	0/0	3	1	3	3	3	12	7	3
RD_12463	2.5/3	8	1	11	6	9	9/0	0/0	3	1	3	3	3	12	7	3
RD_17784	2.5/3	8	1	11	6	9	9/0	0/0	3	1	3	3	3	12	7	3
RD_22821	2.5/3	8	1	11	6	9	9/0	0/0	3	1	3	3	3	12	7	3

RD_21788	2.5/3	8	1	11	6	9	9/0	0/0	3	1	3	3	3	12	7	3
RD_23766	2.5/3	8	1	11	6	9	9/0	0/0	3	1	3	3	3	12	7	3
RD_23192	2.5/3	8	1	11	6	9	9/0	0/0	3	1	3	3	3	12	7	3
RD_24780	2.5/3	8	1	11	6	9	9/0	0/0	3	1	3	3	3	12	7	3/2
RD_13334	2.5/3	8	4	11.5/11	6	9	7	5	1	2	3	3	2	14	4	4
RD_16285	2.5/3	8	4	11.5/11	6	9	7	5/0	1	2	3	3	2	14	4	4
RD_19586	2.5/3	8	4	11.5/11	6	9	7	5	1	2	3	3	2	14	4	4
RD_22114	2.5/3	8	4	11.5/11	6	9	7	5/0	1	2	3	3	2	14	4	4
RD_17159	2.5/3	8	4	11.5/11	6	9	7	5	1	2	3	3	2	14	4	4
RD_23901	2.5/3	8	4	11.5/11	6	9	7	5	1	2	3	3	2	14	4	4
RD_3115	3	8	1	12	3.5	9	5	5	2	1	3	1	2	9	7	7
RD_6601	3	8	1	12	3.5	9	5	5	2	1	3	1	2	9	7	7
RD_10185	3	8	1	12	3.5	9	5	5	2	1	3	1	2	9	7	7
RD_13849	3	8	1	12	3.5	9	5	5	2	1	3	1	2	9	7	7
RD_15552	3	8	1	12	3.5	9	5	5	2	1	3	1	2	9	7	7
RD_1603	3	8	1	12	3.5	9	5	5	2	1	3	1	2	9	7	7
RD_2913	2.5/3	8	4	11	3	5	4	4	4	1	4	3	2	8	4	4
RD_1857	2.5/3	8	4	11	3	5	4	4	4	1	4	3	2	8	4	4
RD_22934	2.5/3	8	0/1	11	2	6	4	4	1	1	2	3	2	12	6	5
RD_23119	2.5/3	8	0/1	11	2	6	4	4	1	1	2	3	2	12	6	5
RD_24913	2.5/3	8	0/1	11	2	6	4	4	1	1	2	3	2	12	6	5
RD_11021	2.5/3	8	1	12	3	5	5	4	6	1	2	1	4	17	4	4
TR_S0702039	2.5/3	9	1	11	2	9	5	6	6	2	2	4	3	14	7	6
TR_S0801675	2.5/3	9	1	11	2	9	5	6	6	2	2	4	3	14	7	6
TR_S0502351	2.5/3	9	1	11	2	9	5	6	6	2	2	4	3	13	7	6
TR_S0502141	2.5/3	9	1	11	2	9	5	6	6	2	2	4	3	13	7	6
TR_S0502280	2.5/3	9	1	11	2	9	5	6	6	2	2	4	3	13	7	6
TR_S0600710	2.5/3	9	1	11	2	9	5	6	6	2	2	4	3	13	7	6
TR_S0601347	2.5/3	9	1	11	2	9	5	6	6	2/0	2	4	3	13	7	6
TR_S0602589	2.5/3	9	1	11	2	9	5	6	6	2	2	4	3	13	7	6
TR_S0602733	2.5/3	9	1	11	2	9	5	6	6	2	2	4	3	13	7	6
TR_S0603715	2.5/3	9	1	11	2	9	5	6	6	2	2	4	3	13	7	6
TR_S0700480	2.5/3	9	1	11	2	9/0	5	6	6/0	2	2	4/0	3	13	7	6
RD_15710-1	2	9	1	11	4	9	5	0/0	4	2	4	1	3	13	9	7
RD_367	2.5/3	9	5	11	2	11	4	5	5	2	4	3	4	14	8	6
RD_22935	2.5/3	9	5	11	2	11	4	5	5	2	4	3	4	14	8	6
RD_2467	2.5/3	9	5	11	2	11	4	5	5	2	4	3	4	14	8	6
RD_23118	2.5/3	9	5	11	2	11	4	5	5	2	4	3	4	14	8	6
RD_24912	2.5/3	9	5	11	2	11	4	5	5	2	4	3	4	14	8	6/7
RD_23604	1.5/2	8	6	9	2	11	7.5/0	5	6	2	4	1	3	16	8	2
RD_23605	1.5/2	8	6	9	2	11	7.5/0	5	6	2	4	1	3	16	8	2
RD_1296	1.5/2	8	6	9	2	11	7.5/0	5	6	2	4	1	3	20	8	2
RD_7311	1.5/2	8	6	9	2	11	7.5/0	5	6	2	4	1	3	20	8	2
RD_14463	1.5/2	8	6	9	2	11	7.5/0	5	6	2	4	1	3	20	8	2
RD_18549	1.5/2	8	6	9	2	11	7.5/0	5	6	2	4	1	3	20	8	2
RD_21105	1.5/2	8	6	9	2	11	7.5/0	5	6	2	4	1	3	20	8	2
RD_22702	1.5/2	8	6	9	2	11	7.5/0	5	6	2	4	1	3	20	8	2
RD_12738	1.5/2	8	6	9	2	11	7.5/0	5	6	2	4	1	3	20	8	2

RD_1295	1.5/2	8	6	9	2	11	7.5/0	5	6	2	4	1	3	20	8	2
RD_14462	1.5/2	8	6	9	2	11	7.5/0	5	6	2	4	1	3	20	8	2
RD_12736	1.5/2	8	6	9	2	11	7.5/0	5	6	2	4	1	3	20	8	2
RD_12737	1.5/2	8	6	9	2	11	7.5/0	5	6	2	4	1	3	20	8	2
RD_24179	1.5/2	8	6	9	2	11	7.5/0	5	6	2	4	1	3	20	8	2
RD_24800	1.5/2	8	6	9	2	11	7.5/0	5	6	2	4	1	3	20	8	2

* When different data were observed for the two assays, the capillary-based MLVA-16_{Orsay} data is shown first followed by / agarose-based MLVA-15_{Orsay}.
§ A 0 allele indicates that no amplicon was observed

Figure S1. Dendrogram deduced from the clustering analysis of 98 isolates using MLVA-16$_{Orsay}$. Branches were colored according to the MLVA clone (MC) or the MLVA type (MT) defined by MLVA-16$_{Orsay}$ typing.



Strain-ID	Alias	MLVA type	Clone MLVA	Phenotype	Patient	Date
RD_23270	PaRD69	6	MLVA Clone 1	psaa	RD06	04.07.2006
RD_23271	PaRD70	6	MLVA Clone 1	psaa	RD06	04.07.2006
RD_20599	PaRD24	7	MLVA Clone 1	psaa	RD06	26.07.2005
RD_22933	PaRD67	7	MLVA Clone 1	psm	RD06	22.05.2006
RD_24302	PaRD25	7	MLVA Clone 1	psa	RD06	22.11.2006
RD_24301	PaRD66	7	MLVA Clone 1	psa	RD06	22.12.2006
RD_23269	PaRD68	7	MLVA Clone 1	psaa	RD06	04.07.2006
RD_23291	PaRD71	7	MLVA Clone 1	psm	RD06	04.07.2006
RD_24086	PaRD72	7	MLVA Clone 1	psa	RD06	18.10.2006
RD_24724	PaRD73	7	MLVA Clone 1	psm	RD06	20.12.2006
RD_24725	PaRD74	7	MLVA Clone 1	psm	RD06	23.01.2007
RD_24980	PaRD75	7	MLVA Clone 1	psm	RD06	21.03.2007
RD_24981	PaRD76	7	MLVA Clone 1	psa	RD06	21.03.2007
PAT4	PA14					
PAO1	PAO1					
RD_12290	PaRD04	1	MLVA Clone 2	psa	RD02	28.10.2002
RD_14075	PaRD05	1	MLVA Clone 2	psa	RD02	24.06.2003
RD_16900	PaRD06	1	MLVA Clone 2	psa	RD02	05.05.2004
RD_22892	PaRD82	1	MLVA Clone 2	psa	RD02	15.05.2006
RD_14076	PaRD83	1	MLVA Clone 2	ps6	RD02	24.06.2003
RD_23261	PaRD52	1	MLVA Clone 2	psaa	RD02	30.06.2006
RD_24799	PaRD53	2	MLVA Clone 2	psa	RD02	05.02.2007
RD_24779	PaRD64	3		psm	RD05	01.02.2007
SW_C50	C50					
SW_SG17M	SG17M					
RD_415	PaRD30	4	MLVA Clone 3	psm	RD07	07.04.1999
RD_2943	PaRD31	4	MLVA Clone 3	psm	RD07	04.01.2000
RD_10952	PaRD36	4	MLVA Clone 3	psm	RD07	17.05.2002
RD_18325	PaRD38	4	MLVA Clone 3	psm	RD07	08.10.2003
RD_13380	PaRD78	4	MLVA Clone 3	psm	RD07	27.03.2003
RD_15136	PaRD79	4	MLVA Clone 3	psm	RD07	18.10.2003
RD_8021	PaRD34	4	MLVA Clone 3	psm	RD07	13.06.2001
RD_414	PaRD29	5	MLVA Clone 3	psa	RD07	07.04.1999
RD_2944	PaRD32	5	MLVA Clone 3	psa	RD07	04.01.2001
RD_8020	PaRD33	5	MLVA Clone 3	psa	RD07	13.06.2001
RD_10951	PaRD35	5	MLVA Clone 3	psa	RD07	17.05.2002
RD_15129	PaRD37	5	MLVA Clone 3	ps6	RD07	18.10.2003
RD_18326	PaRD39	5	MLVA Clone 3	psa	RD07	08.10.2003
RD_13371	PaRD77	5	MLVA Clone 3	psa	RD07	27.03.2003
RD_1238	PaRD17	8		psm	RD05	24.06.1999
RD_6616	PaRD18	8		psm	RD05	24.01.2001
RD_12463	PaRD19	8		psm	RD05	20.11.2002
RD_17784	PaRD20	8		psm	RD05	11.08.2004
RD_22821	PaRD62	8		psm	RD05	06.05.2006
RD_21788	PaRD21	8		psm	RD05	20.12.2005
RD_23766	PaRD22	8		psm	RD05	07.09.2006
RD_23192	PaRD63	8		psa	RD05	
RD_24780	PaRD65	8		psa	RD05	01.02.2007
RD_13334	PaRD08	9		ps6	RD03	24.03.2003
RD_16285	PaRD09	9		psa	RD03	27.02.2004
RD_19586	PaRD10	9		psm	RD03	03.03.2005
RD_22114	PaRD11	9		psm	RD03	30.01.2006
RD_17159	PaRD54	9		psm	RD03	03.06.2004
RD_23901	PaRD55	9		psm	RD03	26.09.2006
RD_3115	PaRD13	10		psm	RD04	21.01.2000
RD_6601	PaRD14	10		psm	RD04	19.01.2001
RD_10185	PaRD15	10		psm	RD04	20.02.2002
RD_13849	PaRD16	10		psm	RD04	26.05.2003
RD_15552	PaRD56	10		psm	RD04	05.12.2003
RD_1603	PaRD12	10		psna	RD04	13.08.1999
RD_2913	PaRD90	12		ps11	RD02	28.12.2002
RD_1857	PaRD03	12		ps5	RD02	09.09.1999
RD_22934	PaRD80	11		psm	RD01	22.05.2006
RD_23119	PaRD49	11		psm	RD01	16.06.2006
RD_24913	PaRD51	11		psm	RD01	06.03.2007
RD_11021	PaRD23	13		psa	RD06	30.05.2002
TR_S0702039	Tr226			psa	CFU24	25.06.2007
TR_S0801675	Tr255			psa	CFU24	09.06.2008
TR_S0502351	Tr77			p3	CFU24	05.10.2005
TR_S0502141	Tr78			p6	CFU24	14.09.2005
TR_S0502280	Tr79			psa	CFU24	28.09.2005
TR_S0600710	Tr136			psa	CFU24	15.03.2006
TR_S0601347	Tr164			psa	CFU24	23.05.2006
TR_S0602589	Tr182			psa	CFU24	03.10.2006
TR_S0602733	Tr187			psa	CFU24	05.10.2006
TR_S0603715	Tr188			p1	CFU24	27.12.2006
TR_S0700480	Tr212			psa	CFU24	06.02.2007
RD_15710-1	PaRD60	17		psm	RD05	22.12.2003
RD_367	PaRD01	16		psaa	RD01	01.04.1999
RD_22935	PaRD02	16		psm	RD01	22.05.2006
RD_2467	PaRD47	16		psm	RD01	10.11.1999
RD_23118	PaRD48	16		psa	RD01	16.06.2006
RD_24912	PaRD50	16		psm	RD01	06.03.2007
RD_23604	PaRD45	14	MLVA Clone 4	psa	RD08	16.08.2006
RD_23605	PaRD46	14	MLVA Clone 4	psm	RD08	16.08.2006
RD_1296	PaRD40	15	MLVA Clone 4	psm	RD08	01.07.1999
RD_7311	PaRD41	15	MLVA Clone 4	psm	RD08	05.04.2001
RD_14463	PaRD42	15	MLVA Clone 4	psm	RD08	07.08.2003
RD_18549	PaRD43	15	MLVA Clone 4	psm	RD08	04.11.2004
RD_21105	PaRD44	15	MLVA Clone 4	psm	RD08	06.10.2005
RD_22702	PaRD57	15	MLVA Clone 4	psm	RD08	19.04.2006
RD_12738	PaRD81	15	MLVA Clone 4	psm	RD08	27.12.2002
RD_1295	PaRD85	15	MLVA Clone 4	ps6	RD08	01.07.1999
RD_14462	PaRD86	15	MLVA Clone 4	psm	RD08	07.08.2003
RD_12736	PaRD87	15	MLVA Clone 4	psa	RD08	27.12.2002
RD_12737	PaRD91	15	MLVA Clone 4	psm	RD08	27.12.2002
RD_24179	PaRD58	15	MLVA Clone 4	psa	RD08	02.11.2006
RD_24800	PaRD59	15	MLVA Clone 4	psa	RD08	08.02.2007

Figures et tableaux supplémentaires de l'article 3

Table S1. Isolates used in this study.

Strain ID	Host	Physiological origin	Year	*spa* type	Oxa	Origin
7	Swine	Skin infection	2004	t1419	MSSA	Drewitz, Germany
19	Swine	Skin infection	2004	t011	MSSA	Kissing, Germany
495	Swine	Skin infection	2004	t337	MSSA	Germany
582	Swine	Skin infection	2004	t1430	MSSA	Mettingen, Germany
809	Swine	Skin infection	2004	t2112	MSSA	Harriehausen, Germany
825	Swine	Skin infection	2004	t034	MRSA	Hodenhagen, Germany
963	Swine	Skin infection	2004	t034	MSSA	Obermützkow, Germany
1921	Swine	Skin infection	2004	t011	MRSA	Nuthe-Urstromtal, Germany
119	Swine	Skin infection	2004	t337	MSSA	Aldersbach, Germany
1022	Swine	Skin infection	2004	t337	MSSA	Solingen, Germany
FAL 226	Swine	Skin infection	2005	t337	MSSA	Heek, Germany
2567	Swine	Skin infection	2005	t964	MSSA	Oschersleben, Germany
2791	Swine	Skin infection	2005	t034	MSSA	Nemsdorf-Göhrendorf, Germany
124	Swine	Genital tract infection	2004	t1939	MSSA	Mönchershofe, Germany
296	Swine	Genital tract infection	2004	t011	MRSA	Kraft, Germany
324	Swine	Genital tract infection	2004	t011	MSSA	Schönermark, Germany
528	Swine	Genital tract infection	2004	t318	MSSA	Kalefeld, Germany
1061	Swine	Genital tract infection	2004	t034	MSSA	Schnatzling, Germany
1213	Swine	MMA-syndrome	2004	t318	MSSA	Bergen, Germany
1231	Swine	MMA-syndrome	2004	t127	MSSA	Armsen, Gemany
1251	Swine	MMA-syndrome	2004	t021	MSSA	Westen, Germany
1295	Swine	Urinary tract infection	2004	t011	MSSA	Sandbeiendorf, Germany
2187	Swine	MMA-syndrome	2005	t011	MRSA	Serbohlsdorf, Germany
2296	Swine	Genital tract infection	2005	t011	MRSA	Kraft, Germany
2594	Swine	MMA-syndrome	2005	t337	MSSA	Scheeßel, Germany
2171	Swine	Genital tract infection	2005	t011	MSSA	Ertingen, Germany
2533	Swine	Genital tract infection	2005	t318	MSSA	Baienfurt, Germany
2920	Swine	MMA-syndrome	2005	t899	MSSA	Bad Fallingbostel, Germany
2926	Swine	MMA-syndrome	2006	t337	MSSA	Reeßum, Germany
2962	Swine	MMA-syndrome	2006	t337	MSSA	Oppershausen, Germany
3036	Swine	MMA-syndrome	2006	t021	MSSA	Oppershausen, Germany
MT9186	Swine	Cervical swab	2009		MSSA	Manche, France
285	Chicken	Septicaemia	2004	t002	MSSA	Rietberg, Germany
287	Chicken	Septicaemia	2004	t002	MSSA	Delbrück, Germany
298	Chicken	Septicaemia	2004	t002	MSSA	Rietberg, Germany
508	Chicken	Septicaemia	2004	t002	MSSA	Mennewitz, Germany
604	Turkey	Septicaemia	2004	t034	MSSA	Steinhagen, Germany
606	Chicken	Septicaemia	2004	t9844	MSSA	Rheda-Wiedenbrück, Germany

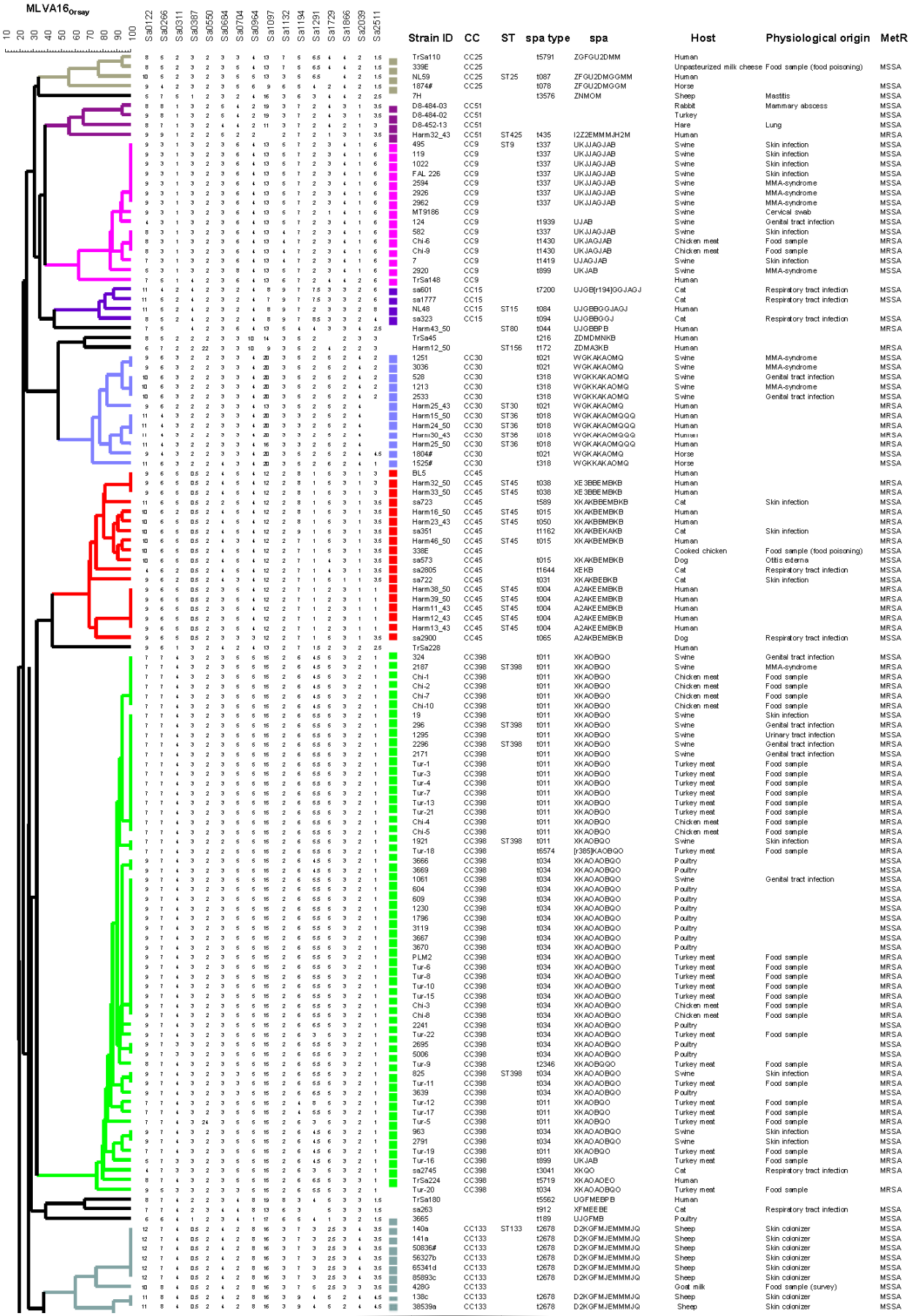
609	Turkey	Septicaemia	2004	t034	MSSA	Stadtlohn, Germany
923	Chicken	Septicaemia	2004	t9844	MSSA	Sprachbrücken, Germany
1230	Turkey	Septicaemia	2004	t034	MSSA	Ebersdorf, Germany
1465	Chicken	Septicaemia	2004	t9844	MSSA	Rüssel, Germany
1796	Turkey	Septicaemia	2004	t034	MSSA	Lorup, Germany
2020	Chicken	Septicaemia	2004	t002	MSSA	Fladder, Germany
2240	Turkey	Septicaemia	2004	t002	MSSA	Dötlingen, Germany
2241	Turkey	Septicaemia	2004	t034	MSSA	Wardenburg, Germany
2303	Chicken	Septicaemia	2004	t002	MSSA	Rietberg, Germany
2695	Turkey	Septicaemia	2004	t034	MSSA	Wadersloh, Germany
3118	Chicken	Septicaemia	2004	t2308	MSSA	Simmerhausen, Germany
3119	Turkey	Septicaemia	2004	t034	MSSA	Rheda-Wiedenbrück, Germany
3122	Chicken	Septicaemia	2004	t002	MSSA	Nieheim-Oeynhausen, Germany
3639	Turkey	Septicaemia	2004	t034	MSSA	Recke, Germany
3664	Chicken	Septicaemia	2004	t214	MSSA	Ostbevern, Germany
3665	Turkey	Septicaemia	2004	t189	MSSA	Reken Großreken, Germany
3666	Turkey	Septicaemia	2004	t034	MSSA	Steinhagen, Germany
3667	Turkey	Septicaemia	2004	t034	MSSA	Rheda-Wiedenbrück, Germany
3669	Turkey	Septicaemia	2004	t034	MSSA	Steinhagen, Germany
3670	Turkey	Septicaemia	2004	t034	MSSA	Rheda-Wiedenbrück, Germany
5006	Turkey	Septicaemia	2004	t034	MSSA	Verl, Germany
5008	Chicken	Septicaemia	2004	t2308	MSSA	Barnstedt, Germany
D8-522-13	Poultry	Unknown	1972		MSSA	Unknown
D8-484-02	Turkey	Unknown	1989		MSSA	Unknown
138c	Sheep	Skin colonizer	2010	t2678	MSSA	Neustadt, Germany
140a	Sheep	Skin colonizer	2010	t2678	MSSA	Neustadt, Germany
141a	Sheep	Skin colonizer	2010	t2678	MSSA	Neustadt, Germany
38539a	Sheep	Skin colonizer	2010	t2678	MSSA	Neustadt, Germany
50836#	Sheep	Skin colonizer	2010	t2678	MSSA	Neustadt, Germany
56327b	Sheep	Skin colonizer	2010	t2678	MSSA	Neustadt, Germany
56417b	Sheep	Skin colonizer	2010	t2678	MSSA	Neustadt, Germany
65341d	Sheep	Skin colonizer	2010	t2678	MSSA	Neustadt, Germany
85893c	Sheep	Skin colonizer	2010	t2678	MSSA	Neustadt, Germany
9A	Sheep	Mastitis	1995		MSSA	Pyrénées-Atlantiques, France
7H	Sheep	Mastitis	1996		MSSA	Pyrénées-Atlantiques, France
sa263	Cat	Respiratory tract infection	2004	t912	MSSA	Salzgitter, Germany
sa323	Cat	Respiratory tract infection	2004	t094	MSSA	Borkwalde, Germany
sa325	Cat	Skin infection	2004	t008	MSSA	Niedergörsdorf, Germany
sa351	Cat	Skin infection	2004	t1162	MSSA	Berlin, Germany
sa485	Cat	Otitis externa	2004	t008	MSSA	München, Germany
sa601	Cat	Respiratory tract infection	2004	t7200	MSSA	Berlin, Germany
sa639	Cat	Respiratory tract infection	2004	t002	MSSA	Berlin, Germany
sa722	Cat	Skin infection	2004	t031	MSSA	Loonig, Germany
sa723	Cat	Skin infection	2004	t589	MSSA	Hamburg, Germany
sa1777	Cat	Respiratory tract infection	2004		MSSA	Hamburg, Germany
sa2745	Cat	Respiratory tract infection	2005	t3041	MRSA	Berlin, Germany

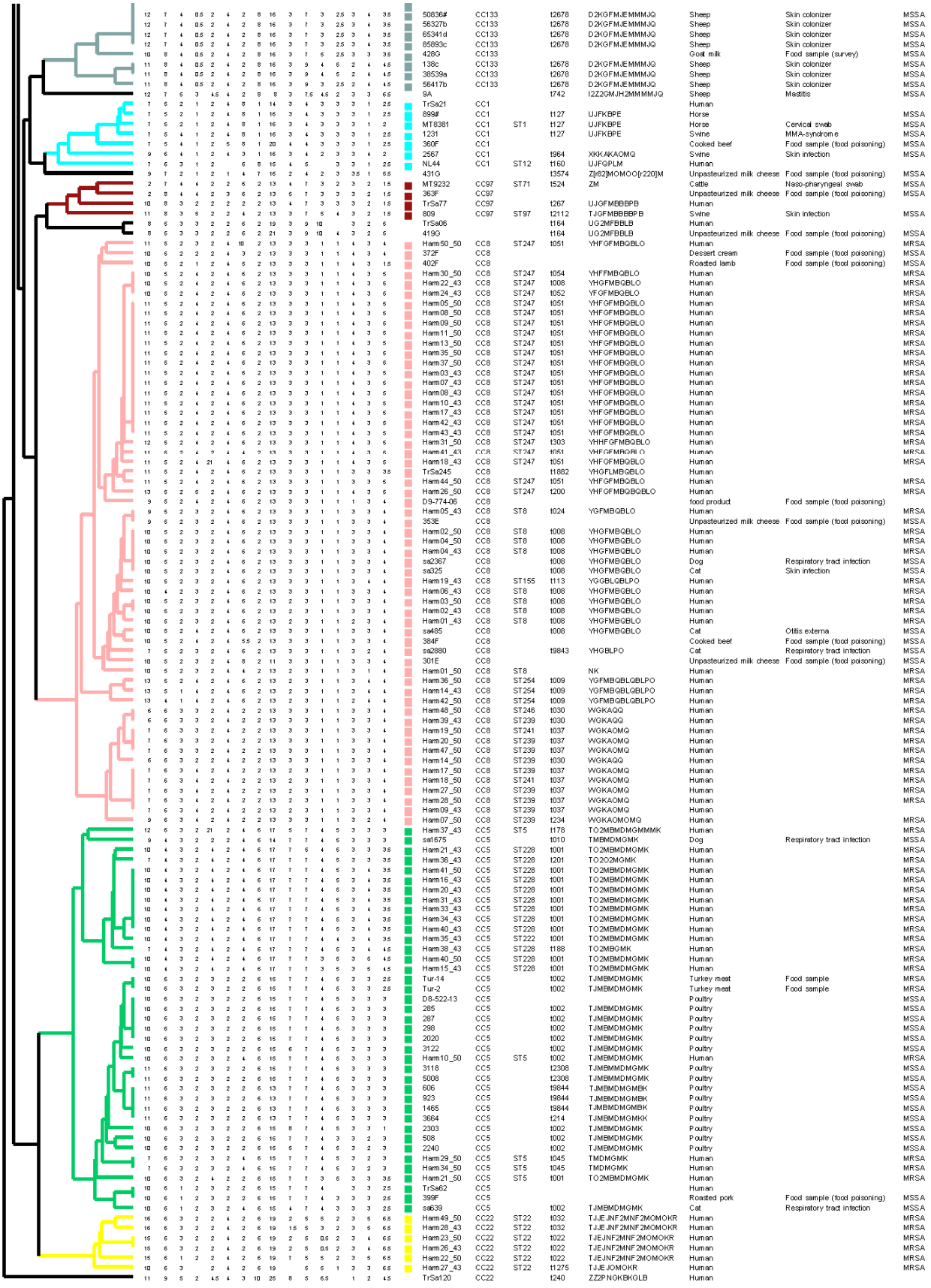
sa2805	Cat	Respiratory tract infection	2005	t1644	MSSA	Leitzkau, Germany
sa2880	Cat	Respiratory tract infection	2008	t9843	MSSA	Geesthacht, Germany
899#	Horse	Unknown	1992	t127	MSSA	Germany
1525#	Horse	Unknown	1992	t318	MSSA	Germany
1804#	Horse	Unknown	1992	t021	MSSA	Germany
1874#	Horse	Unknown	1992	t078	MSSA	Germany
MT8381	Horse	Cervical swab	2008		MSSA	Manche, France
sa573	Dog	Otitis externa	2004	t015	MSSA	34329 Nieste, Germany
sa1675	Dog	Respiratory tract infection	2004	t010	MSSA	22850 Norderstedt, Germany
sa2367	Dog	Respiratory tract infection	2005	t008	MSSA	22523 Hamburg, Germany
sa2900	Dog	Respiratory tract infection	2005	t065	MSSA	22523 Hamburg, Germany
MT9232	Cattle	Naso-pharyngeal swab	2009		MSSA	Manche, France
D8-452-13	Hare	Lung	1988		MSSA	Puy de Dôme, France
D8-484-03	Rabbit	Mammary abscess	1989		MSSA	Vendée, France
PLM2	Turkey meat	Food sample	2009	t034	MRSA	Rhineland-Palatinate, Germany
Tur-1	Turkey meat	Food sample	2009	t002	MRSA	Rhineland-Palatinate, Germany
Tur-2	Turkey meat	Food sample	2009	t011	MRSA	Rhineland-Palatinate, Germany
Tur-3	Turkey meat	Food sample	2009	t011	MRSA	Rhineland-Palatinate, Germany
Tur-4	Turkey meat	Food sample	2009	t011	MRSA	Rhineland-Palatinate, Germany
Tur-5	Turkey meat	Food sample	2009	t011	MRSA	Rhineland-Palatinate, Germany
Tur-6	Turkey meat	Food sample	2009	t034	MRSA	Rhineland-Palatinate, Germany
Tur-7	Turkey meat	Food sample	2009	t011	MRSA	Rhineland-Palatinate, Germany
Tur-8	Turkey meat	Food sample	2009	t034	MRSA	Rhineland-Palatinate, Germany
Tur-9	Turkey meat	Food sample	2009	t2346	MRSA	Rhineland-Palatinate, Germany
Tur-10	Turkey meat	Food sample	2009	t034	MRSA	Rhineland-Palatinate, Germany
Tur-11	Turkey meat	Food sample	2009	t034	MRSA	Rhineland-Palatinate, Germany
Tur-12	Turkey meat	Food sample	2009	t011	MRSA	Rhineland-Palatinate, Germany
Tur-13	Turkey meat	Food sample	2009	t011	MRSA	Rhineland-Palatinate, Germany
Tur-14	Turkey meat	Food sample	2009	t002	MRSA	Rhineland-Palatinate, Germany
Tur-15	Turkey meat	Food sample	2009	t034	MRSA	Rhineland-Palatinate, Germany
Tur-16	Turkey meat	Food sample	2009	t899	MRSA	Rhineland-Palatinate, Germany
Tur-17	Turkey meat	Food sample	2009	t011	MRSA	Rhineland-Palatinate, Germany
Tur-18	Turkey meat	Food sample	2009	t6574	MRSA	Rhineland-Palatinate, Germany
Tur-19	Turkey meat	Food sample	2009	t011	MRSA	Rhineland-Palatinate, Germany
Tur-20	Turkey meat	Food sample	2009	t034	MRSA	Rhineland-Palatinate, Germany
Tur-21	Turkey meat	Food sample	2009	t011	MRSA	Rhineland-Palatinate, Germany
Tur-22	Turkey meat	Food sample	2009	t034	MRSA	Rhineland-Palatinate, Germany
Chi-1	Chicken meat	Food sample	2009	t011	MRSA	Rhineland-Palatinate, Germany
Chi-2	Chicken meat	Food sample	2009	t011	MRSA	Rhineland-Palatinate, Germany
Chi-3	Chicken meat	Food sample	2009	t034	MRSA	Rhineland-Palatinate, Germany
Chi-4	Chicken meat	Food sample	2009	t011	MRSA	Rhineland-Palatinate, Germany
Chi-5	Chicken meat	Food sample	2009	t011	MRSA	Rhineland-Palatinate, Germany
Chi-6	Chicken meat	Food sample	2009	t1430	MRSA	Rhineland-Palatinate, Germany
Chi-7	Chicken meat	Food sample	2009	t011	MRSA	Rhineland-Palatinate, Germany
Chi-8	Chicken meat	Food sample	2009	t034	MRSA	Rhineland-Palatinate, Germany
Chi-9	Chicken meat	Food sample	2009	t1430	MRSA	Rhineland-Palatinate, Germany

Chi-10	Chicken meat	Food sample	2009	t011	MRSA	Rhineland-Palatinate, Germany
428G	Goat milk	Food sample	2002		MSSA	Saône et Loire, France
338E	Cooked chicken	Food sample (food poisoning)	1989		MSSA	Paris, France
419G	Sheep UMC*	Food sample (food poisoning)	2001		MSSA	Puy de Dôme, France
301E	Sheep UMC	Food sample (food poisoning)	1997		MSSA	Aveyron, France
353E	Sheep UMC	Food sample (food poisoning)	1981		MSSA	Pyrénées-Atlantiques, France
339E	Sheep UMC	Food sample (food poisoning)	1986		MSSA	Landes, France
363F	Sheep UMC	Food sample (food poisoning)	1998		MSSA	Cantal, France
431G	Sheep UMC	Food sample (food poisoning)	2002		MSSA	Pyrénées-Atlantiques, France
372F	Dessert cream	Food sample (food poisoning)	2001		MSSA	Val de Marne, France
360F	Cooked beef	Food sample (food poisoning)	1983		MSSA	Oise, France
384F	Cooked beef	Food sample (food poisoning)	1983		MSSA	Doubs, France
399F	Roasted pork	Food sample (food poisoning)	2001		MSSA	Ille et Vilaine, France
402F	Roasted lamb	Food sample (food poisoning)	2001		MSSA	Gironde, France
D9-774-06		Food sample (food poisoning)	1992		ND	USA

*UMC: Unpasteurized milk cheese

Figure S1. Dendrogram deduced from the clustering of the 251 *S. aureus* animal-associated isolates and human strains from the HARMONY collection using MLVA-16$_{Orsay}$.





50836#	CC133		t2678	D2KGFMJEMMMJQ	Sheep	Skin colonizer	MSSA
56327b	CC133		t2678	D2KGFMJEMMMJQ	Sheep	Skin colonizer	MSSA
65341d	CC133		t2678	D2KGFMJEMMMJQ	Sheep	Skin colonizer	MSSA
85893c	CC133		t2678	D2KGFMJEMMMJQ	Sheep	Skin colonizer	MSSA
428G	CC133				Goat milk	Food sample (survey)	MSSA
138c	CC133		t2678	D2KGFMJEMMMJQ	Sheep	Skin colonizer	MSSA
38639a	CC133		t2678	D2KGFMJEMMMJQ	Sheep	Skin colonizer	MSSA
56417b	CC133		t2678	D2KGFMJEMMMJQ	Sheep	Skin colonizer	MSSA
9A			t742	I2Z2GMJH2MMMMJQ	Sheep	Mastitis	MSSA
TrSa21	CC1				Human		
899#	CC1		t127	UJFKBPE	Horse		MSSA
MT8381	CC1	ST1	t127	UJFKBPE	Horse	Cervical swab	MSSA
1231	CC1		t127	UJFKBPE	Swine	MMA-syndrome	MSSA
360F	CC1				Cooked beef	Food sample (food poisoning)	MSSA
2567	CC1		t964	XKKAKAOMQ	Swine	Skin infection	MSSA
NL44	CC1	ST12	t160	UJFQPLM	Human		
431G			t3574	Z[r82]MOMOO[r220]M	Unpasteurized milk cheese	Food sample (food poisoning)	MSSA
MT9232	CC97	ST71	t524	ZM	Cattle	Naso-pharyngeal swab	MSSA
363F	CC97				Unpasteurized milk cheese	Food sample (food poisoning)	MSSA
TrSa77	CC97		t267	UJGFMBBBPB	Human		
809	CC97	ST97	t2112	TJGFMBBBPB	Swine	Skin infection	MSSA
TrSa06			t164	UG2MFBBLB	Human		
419G			t164	UG2MFBBLB	Unpasteurized milk cheese	Food sample (food poisoning)	MSSA
Ham50_50	CC8	ST247	t051	YHFGFMBQBLO	Human		MRSA
372F	CC8				Dessert cream	Food sample (food poisoning)	MSSA
402F	CC8				Roasted lamb	Food sample (food poisoning)	MSSA
Ham30_50	CC8	ST247	t054	YHFFMBQBLO	Human		MRSA
Ham22_43	CC8	ST247	t008	YHGFMBQBLO	Human		MRSA
Ham24_43	CC8	ST247	t052	YFGFMBQBLO	Human		MRSA
Ham05_50	CC8	ST247	t051	YHFGFMBQBLO	Human		MRSA
Ham08_50	CC8	ST247	t051	YHFGFMBQBLO	Human		MRSA
Ham09_50	CC8	ST247	t051	YHFGFMBQBLO	Human		MRSA
Ham11_50	CC8	ST247	t051	YHFGFMBQBLO	Human		MRSA
Ham13_50	CC8	ST247	t051	YHFGFMBQBLO	Human		MRSA
Ham35_50	CC8	ST247	t051	YHFGFMBQBLO	Human		MRSA
Ham37_50	CC8	ST247	t051	YHFGFMBQBLO	Human		MRSA
Ham03_43	CC8	ST247	t051	YHFGFMBQBLO	Human		MRSA
Ham07_43	CC8	ST247	t051	YHFGFMBQBLO	Human		MRSA
Ham08_43	CC8	ST247	t051	YHFGFMBQBLO	Human		MRSA
Ham10_43	CC8	ST247	t051	YHFGFMBQBLO	Human		MRSA
Ham17_43	CC8	ST247	t051	YHFGFMBQBLO	Human		MRSA
Ham42_43	CC8	ST247	t051	YHFGFMBQBLO	Human		MRSA
Ham43_43	CC8	ST247	t051	YHFGFMBQBLO	Human		MRSA
Ham31_50	CC8	ST247	t303	YHHFGFMBQBLO	Human		MRSA
Ham41_43	CC8	ST247	t061	YHFGFMRQRLO	Human		MRSA
Ham18_43	CC8	ST247	t051	YHFGFMBQBLO	Human		MRSA
TrSa245	CC8		t1882	YHGFLMBQBLO	Human		
Ham44_50	CC8	ST247	t051	YHFGFMBQBLO	Human		MRSA
Ham26_50	CC8	ST247	t200	YHFGFMBQBQBLO	Human		MRSA
D9-774-06	CC8				food product	Food sample (food poisoning)	
Ham05_43	CC8	ST8	t024	YGFMBQBLO	Human		MRSA
353E	CC8				Unpasteurized milk cheese	Food sample (food poisoning)	MSSA
Ham02_50	CC8	ST8	t008	YHGFMBQBLO	Human		MRSA
Ham04_50	CC8	ST8	t008	YHGFMBQBLO	Human		MRSA
Ham04_43	CC8	ST8	t008	YHGFMBQBLO	Human		MRSA
sa2367	CC8		t008	YHGFMBQBLO	Dog	Respiratory tract infection	MSSA
sa325	CC8		t008	YHGFMBQBLO	Cat	Skin infection	MSSA
Ham19_43	CC8	ST155	t113	YGGBLQBLPO	Human		MRSA
Ham06_43	CC8	ST8	t008	YHGFMBQBLO	Human		MRSA
Ham03_50	CC8	ST8	t008	YHGFMBQBLO	Human		MRSA
Ham02_43	CC8	ST8	t008	YHGFMBQBLO	Human		MRSA
Ham01_43	CC8	ST8	t008	YHGFMBQBLO	Human		MRSA
sa485	CC8		t008	YHGFMBQBLO	Cat	Otitis externa	MSSA
384F	CC8				Cooked beef	Food sample (food poisoning)	MSSA
sa2880	CC8		t9843	YHGBLPO	Cat	Respiratory tract infection	MSSA
301E	CC8				Unpasteurized milk cheese	Food sample (food poisoning)	MSSA
Ham01_50	CC8	ST8		NK	Human		MRSA
Ham36_50	CC8	ST254	t009	YGFMBQBLQBLPO	Human		MRSA
Ham14_43	CC8	ST254	t009	YGFMBQBLQBLPO	Human		MRSA
Ham42_50	CC8	ST254	t009	YGFMBQBLQBLPO	Human		MRSA
Ham48_50	CC8	ST246	t030	WGKAQG	Human		MRSA
Ham39_43	CC8	ST239	t030	WGKAQG	Human		MRSA
Ham19_50	CC8	ST241	t037	WGKAOMQ	Human		MRSA
Ham20_50	CC8	ST239	t037	WGKAOMQ	Human		MRSA
Ham47_50	CC8	ST239	t037	WGKAOMQ	Human		MRSA
Ham14_50	CC8	ST239	t030	WGKAQG	Human		MRSA
Ham17_50	CC8	ST239	t037	WGKAOMQ	Human		MRSA
Ham18_50	CC8	ST241	t037	WGKAOMQ	Human		MRSA
Ham27_50	CC8	ST239	t037	WGKAOMQ	Human		MRSA
Ham28_50	CC8	ST239	t037	WGKAOMQ	Human		MRSA
Ham09_43	CC8	ST239	t037	WGKAOMQ	Human		MRSA
Ham07_50	CC8	ST239	t234	WGKAOMOMQ	Human		MRSA
Ham37_43	CC5	ST5	t178	TO2MBMDMGMMMK	Human		MRSA
sa1675	CC5		t010	TMBMDMGMK	Dog	Respiratory tract infection	MRSA
Ham21_43	CC5	ST228	t001	TO2MBMDMGMK	Human		MRSA
Ham36_43	CC5	ST228	t201	TO2O2MGMK	Human		MRSA
Ham41_50	CC5	ST228	t001	TO2MBMDMGMK	Human		MRSA
Ham16_43	CC5	ST228	t001	TO2MBMDMGMK	Human		MRSA
Ham20_43	CC5	ST228	t001	TO2MBMDMGMK	Human		MRSA
Ham31_43	CC5	ST228	t001	TO2MBMDMGMK	Human		MRSA
Ham33_43	CC5	ST228	t001	TO2MBMDMGMK	Human		MRSA
Ham34_43	CC5	ST228	t001	TO2MBMDMGMK	Human		MRSA
Ham40_43	CC5	ST228	t001	TO2MBMDMGMK	Human		MRSA
Ham35_43	CC5	ST222	t001	TO2MBMDMGMK	Human		MRSA
Ham38_43	CC5	ST228	t188	TO2MBGMK	Human		MRSA
Ham40_50	CC5	ST228	t001	TO2MBMDMGMK	Human		MRSA
Ham15_43	CC5	ST228	t001	TO2MBMDMGMK	Human		MRSA
Tur-14	CC5		t002	TJMBMDMGMK	Turkey meat	Food sample	MRSA
Tur-2	CC5		t002	TJMBMDMGMK	Turkey meat	Food sample	MRSA
D8-522-13	CC5				Poultry		MSSA
285	CC5		t002	TJMBMDMGMK	Poultry		MSSA
287	CC5		t002	TJMBMDMGMK	Poultry		MSSA
298	CC5		t002	TJMBMDMGMK	Poultry		MSSA
2020	CC5		t002	TJMBMDMGMK	Poultry		MSSA
3122	CC5		t002	TJMBMDMGMK	Poultry		MSSA
Ham10_50	CC5	ST5	t002	TJMBMDMGMK	Human		MRSA
3118	CC5		t2308	TJMBMMDMGMK	Poultry		MSSA
5008	CC5		t2308	TJMBMMDMGMK	Poultry		MSSA
606	CC5		t9844	TJMBMDMGMBK	Poultry		MSSA
923	CC5		t9844	TJMBMDMGMBK	Poultry		MSSA
1465	CC5		t9844	TJMBMDMGMBK	Poultry		MSSA
3664	CC5		t214	TJMBMDMGMKK	Poultry		MSSA
2303	CC5		t002	TJMBMDMGMK	Poultry		MSSA
508	CC5		t002	TJMBMDMGMK	Poultry		MSSA
2240	CC5		t002	TJMBMDMGMK	Poultry		MSSA
Ham29_50	CC5	ST5	t045	TMDMGMK	Human		MRSA
Ham34_50	CC5	ST5	t045	TMDMGMK	Human		MRSA
Ham21_50	CC5	ST5	t001	TO2MBMDMGMK	Human		MRSA
TrSa62	CC5				Human		
399F	CC5				Roasted pork	Food sample (food poisoning)	MSSA
sa639	CC5		t002	TJMBMDMGMK	Cat	Respiratory tract infection	MSSA
Ham49_50	CC22	ST22	t032	TJJEJNF2MNF2MOMOKR	Human		MRSA
Ham28_43	CC22	ST22	t032	TJJEJNF2MNF2MOMOKR	Human		MRSA
Ham23_50	CC22	ST22	t022	TJEJNF2MNF2MOMOKR	Human		MRSA
Ham26_43	CC22	ST22	t022	TJEJNF2MNF2MOMOKR	Human		MRSA
Ham22_50	CC22	ST22	t022	TJEJNF2MNF2MOMOKR	Human		MRSA
Ham27_43	CC22	ST22	t1275	TJJEJOMOKR	Human		MRSA
TrSa120	CC22		t240	ZZ2PNGKBKGLB	Human		

Figure S2. Minimum spanning tree showing the relative discriminatory power of MLVA-$8_{Bilthoven}$ and MLVA-16_{Orsay} for typing CC398.

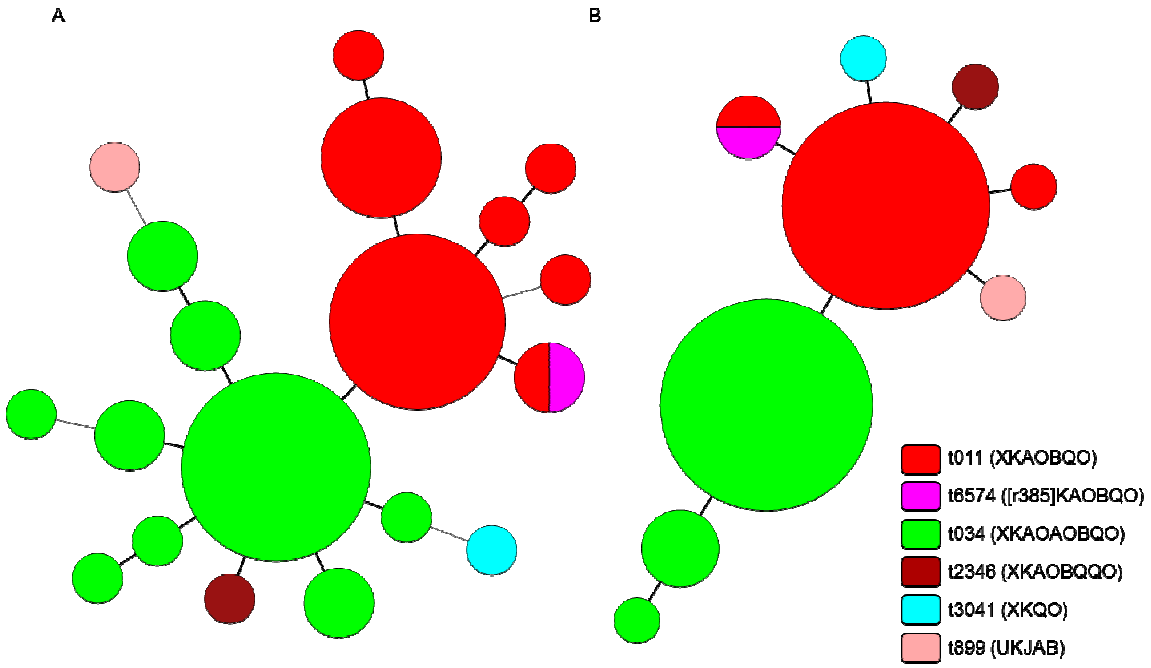
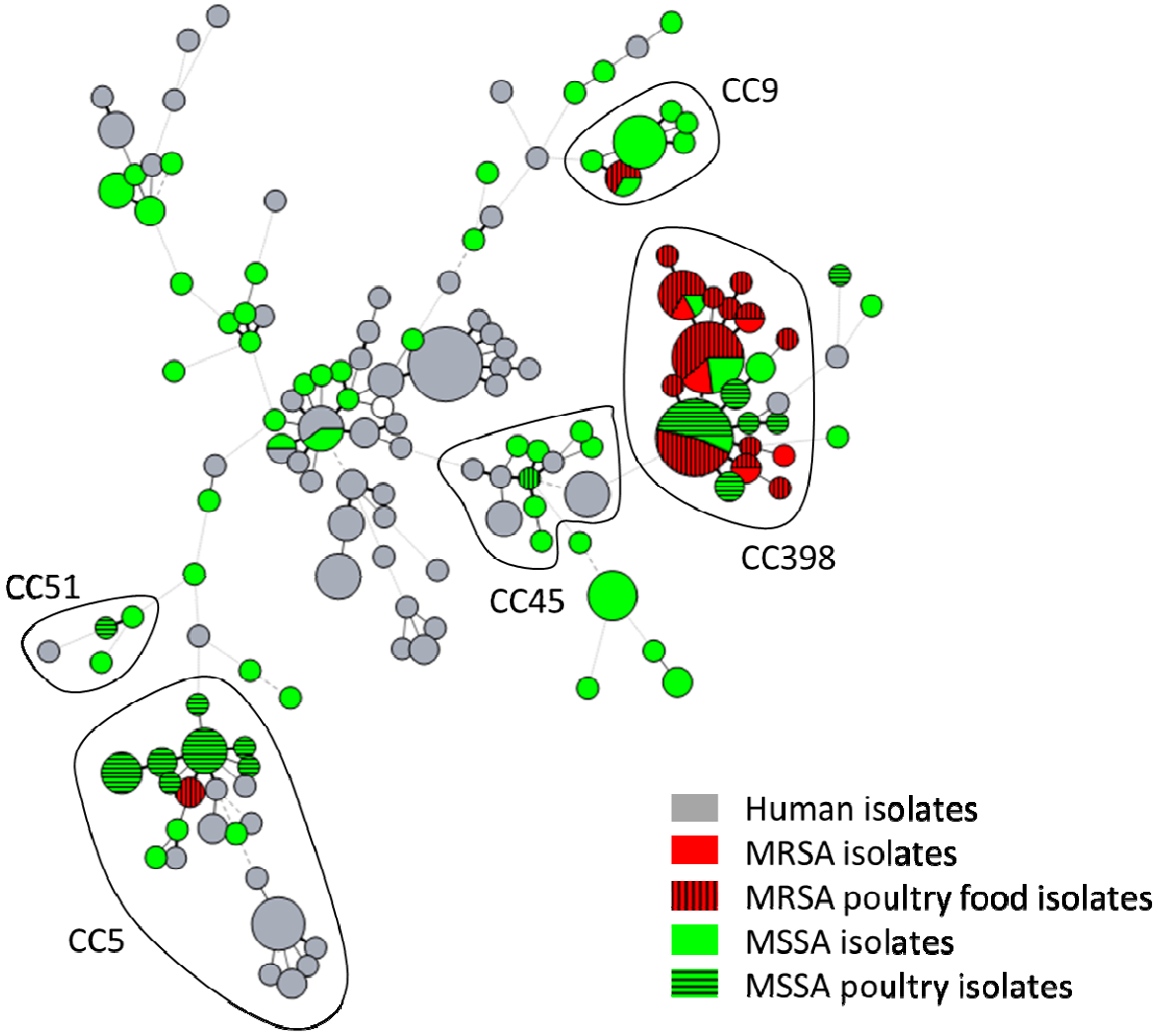


Figure S3. Minimum spanning tree for the poultry isolates.
The minimum spanning tree is identical to the one shown in Figure 4, with the following color code. All human isolates are grayed, MSSA animal and food isolates are shown in green, MRSA isolates in red. Poultry isolates collecting from living animals are cross-hatched with horizontal lines. Poultry isolates from food products are cross-hatched with vertical lines. The MSSA CC45 poultry isolate was collected from a cooked chicken involved in a food poisoning event.



Figures et tableaux supplémentaires de l'article 4

Table S1. Isolates used in this study.

Strain ID	Sampling date	Geographic origin	CC	MLVA type	Host	Physiological origin	MetR	Collaboration
17B	2003	Aveyron, France	CC130	1	Sheep	Subclinical mastitis		INRA-ENVT Toulouse, D. BERGONIER
79F	1998	Côtes d'Armor, France	CC130	2	Sheep	Clinical mastitis		INRA-ENVT Toulouse, D. BERGONIER
74P	2010	Aveyron, France	CC130	3	Sheep	Clinical mastitis		INRA-ENVT Toulouse, D. BERGONIER
27B	1997	Aveyron, France	CC130	3	Sheep	Clinical mastitis		INRA-ENVT Toulouse, D. BERGONIER
084G	2004	Banon, Alpes de Haute Provence, France	CC130	4	Goat	Subclinical mastitis		INRA Nouzilly, F. GILBERT
Touffue D	2004	Banon, Alpes de Haute Provence, France	CC130	5	Goat	Subclinical mastitis		INRA Nouzilly, F. GILBERT
3D	2005	Banon, Alpes de Haute Provence, France	CC130	6	Goat	Subclinical mastitis		INRA Nouzilly, F. GILBERT
6D	2004	Banon, Alpes de Haute Provence, France	CC130	6	Goat	Subclinical mastitis		INRA Nouzilly, F. GILBERT
20D	2005	Banon, Alpes de Haute Provence, France	CC130	7	Goat	Subclinical mastitis		INRA Nouzilly, F. GILBERT
Timili D	2004	Banon, Alpes de Haute Provence, France	CC130	8	Goat	Subclinical mastitis		INRA Nouzilly, F. GILBERT
MT8773	2008	Montmartin-en-Graignes, Manche, France	CC151	9	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT9353	2009	Remilly-sur-Lozon, Manche, France	CC151	9	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
91166	2006-2009	54608 Mützenich, Germany	CC151	9	Cattle	Clinical mastitis	MSSA	FLI, S. SCHWARZ
91594	2006-2009	35104 Lichtenfels, Germany	CC151	9	Cattle	Clinical mastitis	MSSA	FLI, S. SCHWARZ
91597	2006-2009	34513 Waldeck, Germany	CC151	9	Cattle	Clinical mastitis	MSSA	FLI, S. SCHWARZ
91844	2006-2009	99819 Marksuhl, Germany	CC151	9	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
93861	2006-2009	34359 Reinhardshagen, Germany	CC151	9	Cattle	Clinical mastitis	MSSA	FLI, S. SCHWARZ
MT8259	2008	Saint-Pierre-Langers, Manche, France	CC151	10	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
93225	2006-2009	95473 Haag, Germany	CC151	11	Cattle	mastitis	MSSA	FLI, S. SCHWARZ
90104	2006-2009	31535 Neustadt, Germany	CC151	12	Cattle	mastitis	MSSA	FLI, S. SCHWARZ
92917	2006-2009	27607 Langen, Germany	CC151	13	Cattle	mastitis	MSSA	FLI, S. SCHWARZ
91051	2006-2009	38871 Langen, Germany	CC479	14	Cattle	mastitis	MSSA	FLI, S. SCHWARZ
91866	2006-2009	07751 Sulza, Germany	CC479	14	Cattle	mastitis	MSSA	FLI, S. SCHWARZ
91920	2006-2009	68642 Bürstadt, Germany	CC479	15	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
90139	2006-2009	63628 Bad Soden-Salmünster, Germany	CC479	16	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
90102	2006-2009	31535 Neustadt, Germany	CC479	17	Cattle	mastitis	MSSA	FLI, S. SCHWARZ
92966	2006-2009	07937 Silberfeld, Germany	CC479	18	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
BZ8	1992-1993	Rio Grande do Sul, Brazil	CC1	19	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
BZ12	1992-1993	Rio Grande do Sul, Brazil	CC1	19	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
BZ6	1992-1993	Rio Grande do Sul, Brazil	CC1	20	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
BZ10	1992-1993	Rio Grande do Sul, Brazil	CC1	21	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
BZ3	1992-1993	Rio Grande do Sul, Brazil	CC1	22	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
BZ9	1992-1993	Rio Grande do Sul, Brazil	CC1	23	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
15D	2004	Banon, Alpes de Haute Provence, France	CC30	24	Goat	Subclinical mastitis		INRA Nouzilly, F. GILBERT
BZ19	1992-1993	Rio Grande do Sul, Brazil	CC30	25	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
MT9350	2009	Remilly-sur-Lozon, Manche, France		26	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
BZ23	1992-1993	Rio Grande do Sul, Brazil		27	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
9A	1995	Pyrénées-Atlantiques, France		28	Sheep	Subclinical mastitis	MSSA	INRA-ENVT Toulouse, D. BERGONIER
58H	2000	Aveyron, France		29	Sheep	Subclinical mastitis		INRA-ENVT Toulouse, D. BERGONIER
BZ4	1992-1993	Rio Grande do Sul, Brazil	CC133	30	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ

BZ15	1992-1993	Rio Grande do Sul, Brazil	CC133	30	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
MT8168	2008	Acqueville, Manche, France	CC133	31	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
91793	2006-2009	31275 Lehrte, Germany	CC133	32	Cattle	mastitis	MSSA	FLI, S. SCHWARZ
93822	2006-2009	64832 Babenhausen, Germany	CC133	32	Cattle	Clinical mastitis	MSSA	FLI, S. SCHWARZ
MT8251	2008	Saint-Romphaire, Manche, France	CC133	33	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
91592	2006-2009	35625 Hüttenberg, Germany	CC133	34	Cattle	Clinical mastitis	MSSA	FLI, S. SCHWARZ
90088	2006-2009	36466 Wiesenthal, Germany	CC133	35	Cattle	mastitis	MSSA	FLI, S. SCHWARZ
90105	2006-2009	46348 Raesfeld, Germany	CC133	36	Cattle	mastitis	MSSA	FLI, S. SCHWARZ
MT8141	2008	Lessay, Manche, France	CC133	37	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8064	2008	Auvers, Manche, France	CC133	38	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8604	2008	Camprond, Manche, France	CC133	39	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
34L	2003	Aveyron, France	CC133	40	Sheep	Subclinical mastitis		INRA-ENVT Toulouse, D. BERGONIER
63P	2010	Aveyron, France	CC133	41	Sheep	Clinical mastitis		INRA-ENVT Toulouse, D. BERGONIER
52B	1997	Aveyron, France	CC133	42	Sheep	Clinical mastitis		INRA-ENVT Toulouse, D. BERGONIER
91537	2006-2009	82547 Eurasburg, Germany	CC133	43	Cattle	mastitis	MSSA	FLI, S. SCHWARZ
232D	1983	Vienne, France	CC133	43	Goat	Mastitis	MSSA	ANSES, M.-L. DE BUYSER
2G	2005	Banon, Alpes de Haute Provence, France	CC133	44	Goat	Subclinical mastitis		INRA Nouzilly, F. GILBERT
56C	1998	Pyrénées-Atlantiques, France	CC133	45	Sheep	Subclinical mastitis		INRA-ENVT Toulouse, D. BERGONIER
29A	2003	Pyrénées-Atlantiques, France	CC133	45	Sheep	Clinical mastitis		INRA-ENVT Toulouse, D. BERGONIER
34A	1997	Pyrénées-Atlantiques, France	CC133	45	Sheep	Clinical mastitis		INRA-ENVT Toulouse, D. BERGONIER
61B	1998	Pyrénées-Atlantiques, France	CC133	46	Sheep	Subclinical mastitis		INRA-ENVT Toulouse, D. BERGONIER
74D	1996	Pyrénées-Atlantiques, France	CC133	46	Sheep	Subclinical mastitis		INRA-ENVT Toulouse, D. BERGONIER
47B	2003	Aveyron, France	CC133	47	Sheep	Clinical mastitis		INRA-ENVT Toulouse, D. BERGONIER
3187D	2004	Banon, Alpes de Haute Provence, France	CC133	48	Goat	Subclinical mastitis		INRA Nouzilly, F. GILBERT
D4-113-17	1978	Charente-Maritime, France	CC133	49	Goat	Subclinical mastitis		INRA Nouzilly, F. GILBERT
D8-660-22	1993	Deux-Sèvres, France	CC133	50	Goat	Subclinical mastitis		INRA Nouzilly, F. GILBERT
D9-786-04	1995	Cher, France	CC133	51	Goat	Subclinical mastitis		INRA Nouzilly, F. GILBERT
2A	1996	Pyrénées-Atlantiques, France	CC133	52	Sheep	Clinical mastitis		INRA-ENVT Toulouse, D. BERGONIER
34G	2005	Sainte-Maure, Indre et Loire, France	CC133	53	Goat	Subclinical mastitis		INRA Nouzilly, F. GILBERT
358G	2005	Sainte-Maure, Indre et Loire, France	CC133	54	Goat	Subclinical mastitis		INRA Nouzilly, F. GILBERT
1F	2001	Aveyron, France	CC133	55	Sheep	Subclinical mastitis		INRA-ENVT Toulouse, D. BERGONIER
MT8477	2008	Courcy, Manche, France	CC5	56	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
90181	2006-2009	37276 Meinhard, Germany	CC22	57	Cattle	Clinical mastitis	MSSA	FLI, S. SCHWARZ
BZ5	1992-1993	Rio Grande do Sul, Brazil		58	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
MT8066	2008	Saint-Laurent-de-Terregatte, Manche, France	CC20	59	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8099	2008	Saint-Pair-sur-Mer, Manche, France	CC20	59	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8314	2008	Sartilly, Manche, France	CC20	59	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT9030	2009	Coudeville-sur-Mer, Manche, France	CC20	59	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8434	2008	Carquebut, Manche, France	CC20	60	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8348	2008	Quetteville, Calvados, France	CC20	61	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8582	2008	Fougerolles-du-Plessis, Mayenne, France	CC20	62	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8333	2008	Saint-Cyr-du-Bailleul, Manche, France	CC20	63	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
92554	2006-2009	87459 Pfronten, Germany	CC7	64	Cattle	mastitis	MSSA	FLI, S. SCHWARZ
90178	2006-2009	35274 Kirchhain, Germany	CC8	65	Cattle	Clinical mastitis	MSSA	FLI, S. SCHWARZ
91794	2006-2009	31515 Wunstorf, Germany	CC8	66	Cattle	mastitis	MSSA	FLI, S. SCHWARZ
MT9379	2009	Saint-Brice-de-Landelles, Manche, France	CC97	67	Cattle	Mastitis	MSSA	LDA50, M. TREILLES

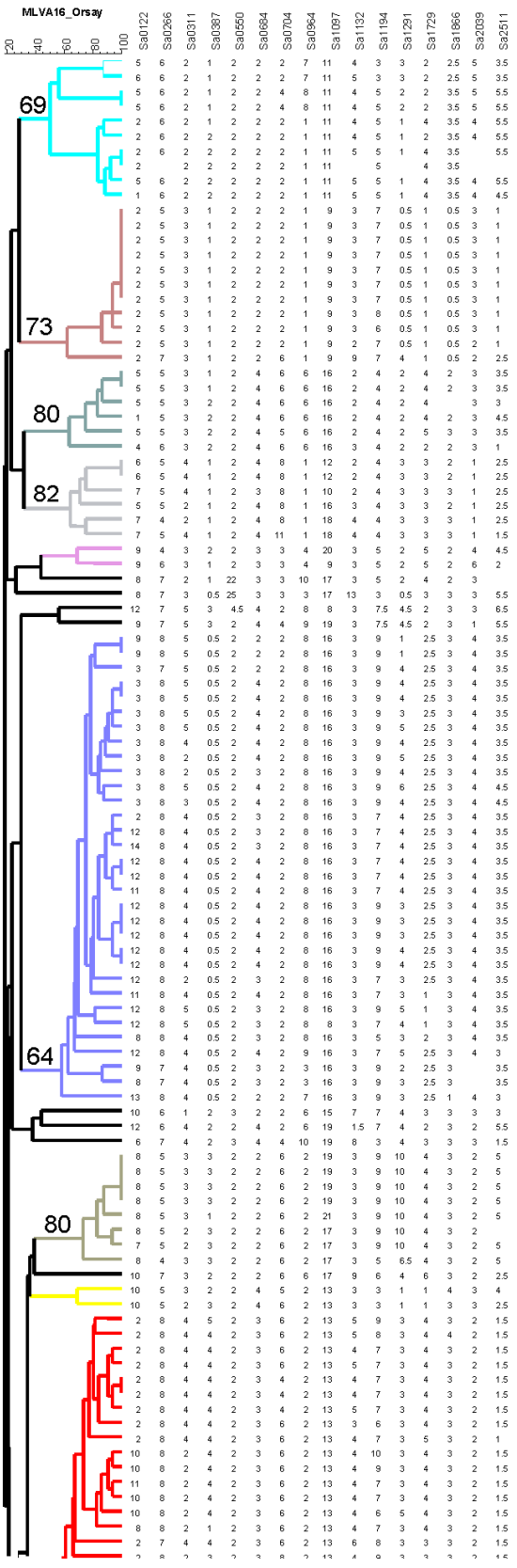
MT9385	2009	Feugères, Manche, France	CC97	68	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8189	2008	Muneville-sur-Mer, Manche, France	CC97	69	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT9342	2009	Rocheville, Manche, France	CC97	70	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8187	2008	Muneville-sur-Mer, Manche, France	CC97	71	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8763	2008	Parigny, Manche, France	CC97	71	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8130	2008	Pont-Hébert, Manche, France	CC97	72	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT9094	2009	Saint-Martin-le-Bouillant, Manche, France	CC97	73	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8085	2008	Hauteville-la-Guichard, Manche, France	CC97	74	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8671	2008	Saint-Jean-des-Champs, Manche, France	CC97	75	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT9306	2009	Le Teilleul, Manche, France	CC97	76	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8134	2008	Cambernon, Manche, France	CC97	77	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8775	2008	Saint-Lô, Manche, France	CC97	78	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8741	2008	Saint-Georges-de-Bohon, Manche, France	CC97	79	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8059	2008	Saint-Sauveur-la-Pommeraye, Manche, France	CC97	80	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT9347	2009	Bacilly, Manche, France	CC97	81	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
90745	2006-2009	36115 Ehrenberg, Germany	CC97	82	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
MT8395	2008	Le Mesnil-Villeman, Manche, France	CC97	83	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
90756	2006-2009	34508 Willingen, Germany	CC97	84	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
MT8269	2008	Baudreville, Manche, France	CC97	85	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8765	2008	Saint-Jean-des-Champs, Manche, France	CC97	86	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
BZ11	1992-1993	Rio Grande do Sul, Brazil	CC97	87	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
92275	2006-2009	07381 Bodelwitz, Germany	CC97	88	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
BZ13	1992-1993	Rio Grande do Sul, Brazil	CC97	89	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
MT8360	2008	Montgothier, Manche, France	CC97	90	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
BZ1	1992-1993	Rio Grande do Sul, Brazil	CC97	91	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
BZ16	1992-1993	Rio Grande do Sul, Brazil	CC97	91	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
BZ17	1992-1993	Rio Grande do Sul, Brazil	CC97	91	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
BZ18	1992-1993	Rio Grande do Sul, Brazil	CC97	91	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
MT8164	2008	Drubec, Calvados, France	CC97	92	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8343	2008	Drubec, Calvados, France	CC97	92	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
91591	2006-2009	36157 Ebersberg, Germany	CC97	93	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
MT9279	2009	Subligny, Manche, France	CC97	94	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
BZ21	1992-1993	Rio Grande do Sul, Brazil	CC97	95	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
MT9315	2009	Saint-Michel-de-Montjoie, Manche, France		96	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
BZ7	1992-1993	Rio Grande do Sul, Brazil		96	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
MT9359	2009	Reffuveille, Manche, France		97	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
7H	1996	Pyrénées-Atlantiques, France		98	Sheep	Subclinical mastitis	MSSA	INRA-ENVT Toulouse, D. BERGONIER
MT8065	2008	Saint-Laurent-de-Terregatte, Manche, France	CC9	99	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8074	2008	Saint-Laurent-de-Terregatte, Manche, France	CC9	99	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
92344	2006-2009	65589 Hademar, Germany	CC9	100	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
MT8041	2008	Chèvreville, Manche, France	CC9	101	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8042	2008	Chèvreville, Manche, France	CC9	101	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
90758	2006-2009	36110 Schlitz, Germany	CC9	102	Cattle	Subclinical mastitis	MSSA	FLI, S. SCHWARZ
91586	2006-2009	36145 Hofbieber, Germany	CC9	102	Cattle	Clinical mastitis	MSSA	FLI, S. SCHWARZ
90145	2006-2009	37296 Ringgau, Germany	CC9	103	Cattle	Clinical mastitis	MSSA	FLI, S. SCHWARZ
MT8002	2008	Saint-Jean-de-Daye, Manche, France	CC9	104	Cattle	Mastitis	MSSA	LDA50, M. TREILLES

MT8682	2008	Saint-Côme-du-Mont, Manche, France	CC9	104	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
MT8101	2008	Saint-Côme-du-Mont, Manche, France	CC9	105	Cattle	Mastitis	MSSA	LDA50, M. TREILLES
M51	2009	74589 Satteldorf, Germany	CC398	106	Cattle	Subclinical mastitis	MRSA	FLI, S. SCHWARZ
M53	2009	Baden-Württemberg, Germany	CC398	106	Cattle	Subclinical mastitis	MRSA	FLI, S. SCHWARZ
M1	2009	32257 Bünde, Germany	CC398	107	Cattle	Clinical mastitis	MRSA	FLI, S. SCHWARZ
M3	2009	89296 Osterberg, Germany	CC398	107	Cattle	Clinical mastitis	MRSA	FLI, S. SCHWARZ
M5	2009	27446 Farven, Germany	CC398	107	Cattle	Clinical mastitis	MRSA	FLI, S. SCHWARZ
M12	2009	49328 Melle, Germany	CC398	107	Cattle	Clinical mastitis	MRSA	FLI, S. SCHWARZ
M40	2009	Baden-Württemberg, Germany	CC398	107	Cattle	Subclinical mastitis	MRSA	FLI, S. SCHWARZ
M50	2009	Baden-Württemberg, Germany	CC398	107	Cattle	Subclinical mastitis	MRSA	FLI, S. SCHWARZ
M60	2009	Baden-Württemberg, Germany	CC398	107	Cattle	Subclinical mastitis	MRSA	FLI, S. SCHWARZ
M7	2009	Bayern, Germany	CC398	108	Cattle	Subclinical mastitis	MRSA	FLI, S. SCHWARZ
M44	2009	Baden-Württemberg, Germany	CC398	108	Cattle	Subclinical mastitis	MRSA	FLI, S. SCHWARZ
M58	2009	Baden-Württemberg, Germany	CC398	109	Cattle	Subclinical mastitis	MRSA	FLI, S. SCHWARZ
M2	2009	32469 Petershagen, Germany	CC398	110	Cattle	Clinical mastitis	MRSA	FLI, S. SCHWARZ
M11	2009	26655 Westerstede, Germany	CC398	111	Cattle	Clinical mastitis	MRSA	FLI, S. SCHWARZ
M6	2009	26901 Lorup, Germany	CC398	112	Cattle	Clinical mastitis	MRSA	FLI, S. SCHWARZ
M9	2009	Bayern, Germany	CC398	113	Cattle	Subclinical mastitis	MRSA	FLI, S. SCHWARZ
32D	2005	Banon, Alpes de Haute Provence, France	CC25	114	Goat	Subclinical mastitis	MSSA	INRA Nouzilly, F. GILBERT
MT9057	2009	Herqueville, Manche, France		115	Cattle	Mastitis	MSSA	LDA50, M. TREILLES

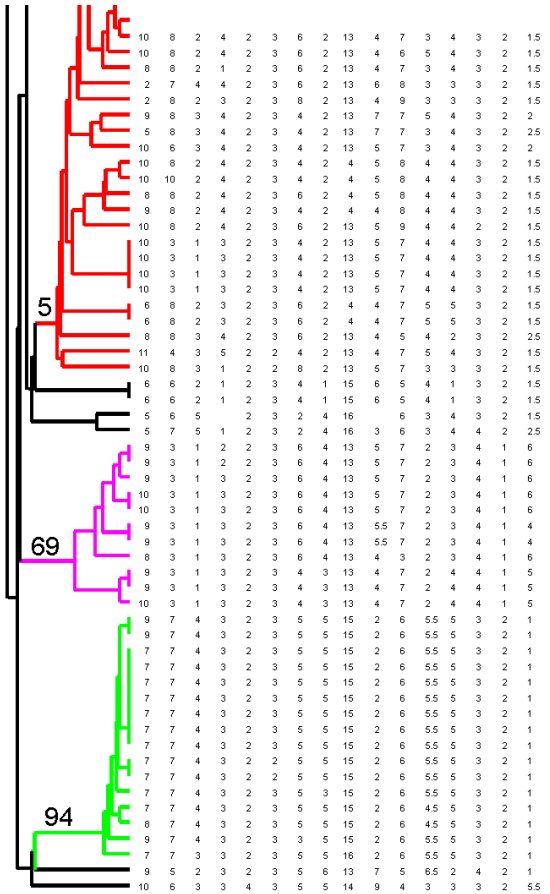
Table S2. Allelic richness and diversity per locus and for each family (human-related, animal-specific, mammary gland-adapted).

VNTR	Human-predominant (Allelic richness - Allelic diversity [standard deviation of 5%])	Animal-predominant (Allelic richness - Allelic diversity [standard deviation of 5%])	Mammary gland-predominant (Allelic richness - Allelic diversity [standard deviation of 5%])
Sa0122	12 - 0.7833 [0.7446,0.8219]	13 - 0.8474 [0.8226,0.8722]	5 - 0.6182 [0.4758,0.7607]
Sa0311	5 - 0.6683 [0.6272,0.7093]	5 - 0.6346 [0.5711,0.6981]	2 - 0.4843 [0.3880,0.5807]
Sa0387	7 - 0.7482 [0.7115,0.7849]	6 - 0.5829 [0.5211,0.6446]	2 - 0.4615 [0.3406,0.5825]
Sa0550	4 - 0.3981 [0.3127,0.4836]	2 - 0.0106 [0.0000,0.0315]	1 - 0.0000 [0.0000,0.0000]
Sa0684	4 - 0.6125 [0.5601,0.6648]	3 - 0.4388 [0.3626,0.5150]	2 - 0.3590 [0.1812,0.5368]
Sa0964	6 - 0.7033 [0.6737,0.7329]	7 - 0.7393 [0.7096,0.7689]	4 - 0.5641 [0.3873,0.7409]
Sa1097	15 - 0.7823 [0.7308,0.8338]	8 - 0.7130 [0.6829,0.7430]	3 - 0.6724 [0.6173,0.7274]
Sa1194	7 - 0.6965 [0.6575,0.7354]	8 - 0.6967 [0.6659,0.7274]	6 - 0.7721 [0.7001,0.8440]
Sa1729	7 - 0.7121 [0.6725,0.7517]	8 - 0.7490 [0.7157,0.7824]	4 - 0.6980 [0.6283,0.7677]
Sa1866	4 - 0.4908 [0.4141,0.5675]	4 - 0.2906 [0.2132,0.3680]	6 - 0.7436 [0.6439,0.8433]
Sa0266	6 - 0.6475 [0.6130,0.6820]	7 - 0.6805 [0.6433,0.7178]	4 - 0.5726 [0.4606,0.6847]
Sa0704	9 - 0.8115 [0.7878,0.8351]	6 - 0.7378 [0.7111,0.7646]	4 - 0.5185 [0.3347,0.7024]
Sa1132	10 - 0.7040 [0.6612,0.7467]	7 - 0.7296 [0.6986,0.7605]	6 - 0.7835 [0.7020,0.8649]
Sa1291	9 - 0.6957 [0.6427,0.7487]	11 - 0.8314 [0.8079,0.8549]	6 - 0.7607 [0.6759,0.8455]
Sa2039	6 - 0.6880 [0.6370,0.7390]	5 - 0.5326 [0.4660,0.5993]	5 - 0.6610 [0.4947,0.8272]
Sa2511	14 - 0.8760 [0.8555,0.8966]	11 - 0.7719 [0.7338,0.8099]	7 - 0.7692 [0.6651,0.8733]
MLVA-16_{Orsay}	105 - 0.9863 [0.9794,0.9931]	106 - 0.9755 [0.9651,0.9859]	18 - 0.9316 [0.8595,1.0000]

Figure S1. Dendrogram deduced from the clustering of the 152 *S. aureus* mastitis-associated isolates using MLVA-16$_{Orsay}$. The color code reflects MLVA clusters when using the 45% cutoff.

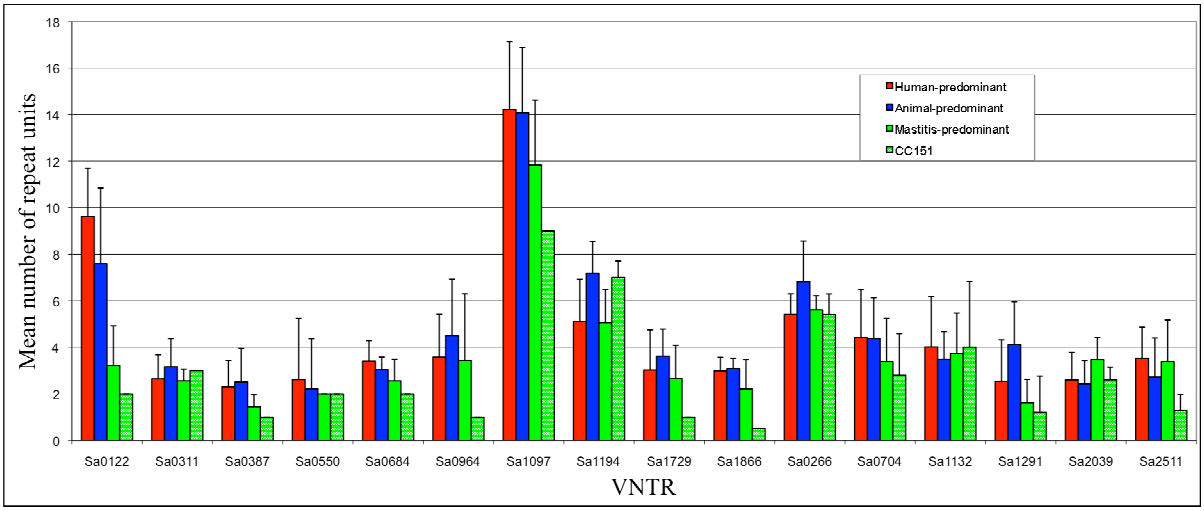


Strain ID	CC	ST	spa	spa type	MT	MetR	Host	Geographic origin	Sampling date
17B	CC130		Z[r82]MOM	t1773	1		Sheep	Aveyron, France	2003
79F	CC130		Z[r82]MMOM		2		Sheep	Côtes d'Armor, France	1998
74P	CC130	ST2011	Z[r82]MOM	t1773	3		Sheep	Aveyron, France	2010
27B	CC130				3		Sheep	Aveyron, France	1997
084G	CC130				4		Goat	Banon, Alpes de Haute Provence, France	2004
Touffue D	CC130	ST130	ZM	t524	5		Goat	Banon, Alpes de Haute Provence, France	2004
3D	CC130				6		Goat	Banon, Alpes de Haute Provence, France	2005
6D	CC130				6		Goat	Banon, Alpes de Haute Provence, France	2004
20D	CC130		Z[r82]MOM	t1773	7		Goat	Banon, Alpes de Haute Provence, France	2005
Timili D	CC130		Z	t528	8		Goat	Banon, Alpes de Haute Provence, France	2004
MT8773	CC151	ST504	ZB	t529	9	MSSA	Cattle	Montmartin-en-Graignes, Manche, France	2008
MT9353	CC151				9	MSSA	Cattle	Remilly-sur-Lozon, Manche, France	2009
91166	CC151				9	MSSA	Cattle	54608 Mützenich, Germany	2006-2009
91594	CC151				9	MSSA	Cattle	35104 Lichtenfels, Germany	2006-2009
91597	CC151				9	MSSA	Cattle	34513 Waldeck, Germany	2006-2009
91844	CC151				9	MSSA	Cattle	99819 Marksuhl, Germany	2006-2009
93861	CC151				9	MSSA	Cattle	34359 Reinhardshagen, Germany	2006-2009
MT8259	CC151				10	MSSA	Cattle	Saint-Pierre-Langers, Manche, France	2008
93225	CC151				11	MSSA	Cattle	95473 Haag, Germany	2006-2009
90104	CC151				12	MSSA	Cattle	31535 Neustadt, Germany	2006-2009
92917	CC151		ZB	t529	13	MSSA	Cattle	27607 Langen, Germany	2006-2009
91051	CC479	ST479	ZDMNQ	t2873	14	MSSA	Cattle	38871 Langen, Germany	2006-2009
91866	CC479				14	MSSA	Cattle	07751 Sulza, Germany	2006-2009
91920	CC479				15	MSSA	Cattle	68642 Bürstadt, Germany	2006-2009
90139	CC479		Z	t528	16	MSSA	Cattle	63628 Bad Soden-Salmünster, Germany	2006-2009
90102	CC479		ZDMNQ	t2873	17	MSSA	Cattle	31535 Neustadt, Germany	2006-2009
92966	CC479		ZDMM	t3104	18	MSSA	Cattle	07937 Silberteld, Germany	2006-2009
BZ8	CC1		UJFKPE	t922	19	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
BZ12	CC1				19	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
BZ6	CC1				20	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
BZ10	CC1		UKBPE	t114	21	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
BZ3	CC1				22	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
BZ9	CC1				23	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
15D	CC30				24		Goat	Banon, Alpes de Haute Provence, France	2004
BZ19	CC30				25	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
MT9350			ZDMDMNKB	t216	26	MSSA	Cattle	Remilly-sur-Lozon, Manche, France	2009
BZ23			ZDMJQDMO	t518	27	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
9A			I2Z2GMJH2MMMMJQt742		28	MSSA	Sheep	Pyrénées-Atlantiques, France	1995
58H			I2Z2GJH2MMMM	t3584	29		Sheep	Aveyron, France	2000
BZ4	CC133				30	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
BZ15	CC133				30	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
MT8168	CC133				31	MSSA	Cattle	Acqueville, Manche, France	2008
91793	CC133				32	MSSA	Cattle	31275 Lehrte, Germany	2006-2009
93822	CC133				32	MSSA	Cattle	64832 Babenhausen, Germany	2006-2009
MT8251	CC133				33	MSSA	Cattle	Saint-Romphaire, Manche, France	2008
91592	CC133				34	MSSA	Cattle	35625 Hüttenberg, Germany	2006-2009
90088	CC133				35	MSSA	Cattle	36466 Wiesenthal, Germany	2006-2009
90105	CC133				36	MSSA	Cattle	46348 Raesfeld, Germany	2006-2009
MT8141	CC133		D2JG	t1403	37	MSSA	Cattle	Lessay, Manche, France	2008
MT8064	CC133				38	MSSA	Cattle	Auvers, Manche, France	2008
MT8604	CC133				39	MSSA	Cattle	Camprond, Manche, France	2008
34L	CC133				40		Sheep	Aveyron, France	2003
63P	CC133				41		Sheep	Aveyron, France	2010
52B	CC133		D2KGFMJEMMMMMJQt4560		42		Sheep	Aveyron, France	1997
91537	CC133				43	MSSA	Cattle	82547 Eurasburg, Germany	2006-2009
232D	CC133				43	MSSA	Goat	Vienne, France	1983
2G	CC133				44		Goat	Banon, Alpes de Haute Provence, France	2005
56C	CC133				45		Sheep	Pyrénées-Atlantiques, France	1998
29A	CC133				45		Sheep	Pyrénées-Atlantiques, France	2003
34A	CC133				45		Sheep	Pyrénées-Atlantiques, France	1997
61B	CC133				46		Sheep	Pyrénées-Atlantiques, France	1998
74D	CC133				46		Sheep	Pyrénées-Atlantiques, France	1996
47B	CC133				47		Sheep	Aveyron, France	2003
3187D	CC133				48		Goat	Banon, Alpes de Haute Provence, France	2004
D4-113-17	CC133				49		Goat	Charente-Maritime, France	1978
D8-660-22	CC133				50		Goat	Deux-Sèvres, France	1993
D9-786-04	CC133				51		Goat	Cher, France	1995
2A	CC133				52		Sheep	Pyrénées-Atlantiques, France	1996
34G	CC133				53		Goat	Sainte-Maure, Indre et Loire, France	2005
358G	CC133				54		Goat	Sainte-Maure, Indre et Loire, France	2005
1F	CC133		D2KGFMJEMMMMJQt544		55		Sheep	Aveyron, France	2001
MT8477	CC5				56	MSSA	Cattle	Courcy, Manche, France	2008
90181	CC22		TJEJNCMOMOKR	t005	57	MSSA	Cattle	37276 Meinhard, Germany	2006-2009
BZ5			UEBBPB	t2815	58	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
MT8066	CC20	ST389	UQ2MFBBLB	t164	59	MSSA	Cattle	Laurent-de-Terregatte, Manche, France	2008
MT8099	CC20				59	MSSA	Cattle	Saint-Pair-sur-Mer, Manche, France	2008
MT8314	CC20				59	MSSA	Cattle	Sartilly, Manche, France	2008
MT9030	CC20				59	MSSA	Cattle	Coudeville-sur-Mer, Manche, France	2009
MT8434	CC20				60	MSSA	Cattle	Carquebut, Manche, France	2008
MT8348	CC20				61	MSSA	Cattle	Guetteville, Calvados, France	2008
MT8582	CC20		UQ2MFBLB	t731	62	MSSA	Cattle	Fougerolles-du-Plessis, Mayenne, France	2008
MT8333	CC20				63	MSSA	Cattle	Saint-Cyr-du-Bailleul, Manche, France	2008
92554	CC7		UJFMBGJAGJ	t091	64	MSSA	Cattle	87459 Pfronten, Germany	2006-2009
90178	CC8				65	MSSA	Cattle	35274 Kirchhain, Germany	2006-2009
91794	CC8				66	MSSA	Cattle	31515 Wunstorf, Germany	2006-2009
MT9379	CC97				67	MSSA	Cattle	Saint-Brice-de-Landelles, Manche, France	2009
MT9385	CC97				68	MSSA	Cattle	Feugères, Manche, France	2009
MT8189	CC97				69	MSSA	Cattle	Muneville-sur-Mer, Manche, France	2008
MT9342	CC97				70	MSSA	Cattle	Rocheville, Manche, France	2009
MT8187	CC97				71	MSSA	Cattle	Muneville-sur-Mer, Manche, France	2008
MT8763	CC97				71	MSSA	Cattle	Parigny, Manche, France	2008
MT8130	CC97				72	MSSA	Cattle	Pont-Hébert, Manche, France	2008
MT9094	CC97				73	MSSA	Cattle	Saint-Martin-le-Bouillant, Manche, France	2009
MT8085	CC97				74	MSSA	Cattle	Hauteville-la-Guichard, Manche, France	2008
MT8671	CC97		UJGFMBBBPB	t267	75	MSSA	Cattle	Saint-Jean-des-Champs, Manche, France	2008
MT9306	CC97				76	MSSA	Cattle	Le Teilleul, Manche, France	2009
MT8134	CC97				77	MSSA	Cattle	Cambernon, Manche, France	2008
MT8775	CC97				78	MSSA	Cattle	Saint-Lô, Manche, France	2008
MT8741	CC97				79	MSSA	Cattle	Saint-Georges-de-Bohon, Manche, France	2008
MT8059	CC97				80	MSSA	Cattle	Saint-Sauveur-la-Pommeraye, Manche, F.	2008
MT9347	CC97				81	MSSA	Cattle	Bacilly, Manche, France	2009
90745	CC97				82	MSSA	Cattle	36115 Ehrenberg, Germany	2006-2009



10	8	2	4	2	3	6	2	13	4	7	3	4	3	2	1.5	MT8775	CC97				78	MSSA	Cattle	Saint-Lô, Manche, France	2008
10	8	2	4	2	3	6	2	13	4	6	5	4	3	2	1.5	MT8741	CC97				79	MSSA	Cattle	Saint-Georges-de-Bohon, Manche, France	2008
8	8	2	1	2	3	6	2	13	4	7	3	4	3	2	1.5	MT8059	CC97				80	MSSA	Cattle	Sauveur-la-Pommeraye, Manche, France	2008
2	7	4	4	2	3	6	2	13	6	8	3	3	3	2	1.5	MT9347	CC97				81	MSSA	Cattle	Bacilly, Manche, France	2009
2	8	2	3	2	3	8	2	13	4	9	3	3	3	2	1.5	90745	CC97				82	MSSA	Cattle	36115 Ehrenberg, Germany	2006-2009
9	8	3	4	2	3	4	2	13	7	7	5	4	3	2	2	MT8395	CC97				83	MSSA	Cattle	Le Mesnil-Villeman, Manche, France	2008
5	8	3	4	2	3	4	2	13	7	7	3	4	3	2	2.5	90756	CC97		UKBPB	t2844	84	MSSA	Cattle	34508 Willingen, Germany	2006-2009
10	6	3	4	2	3	4	2	13	5	7	3	4	3	2	2	MT8269	CC97		UJGFMB	t189	85	MSSA	Cattle	Baudreville, Manche, France	2008
10	8	2	4	2	3	4	2	4	5	8	4	4	3	2	1.5	MT8765	CC97				86	MSSA	Cattle	Saint-Jean-des-Champs, Manche, France	2008
10	10	2	4	2	3	4	2	4	5	8	4	4	3	2	1.5	BZ11	CC97				87	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
8	8	2	4	2	3	6	2	4	5	8	4	4	3	2	1.5	92275	CC97				88	MSSA	Cattle	07381 Bodelwitz, Germany	2006-2009
9	8	2	4	2	3	4	2	4	4	8	4	4	3	2	1.5	BZ13	CC97		UJGFMBBPB	t359	89	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
10	8	2	4	2	3	6	2	13	5	9	4	4	2	2	1.5	MT8360	CC97				90	MSSA	Cattle	Montgothier, Manche, France	2008
10	3	1	3	2	3	4	2	13	5	7	4	4	3	2	1.5	BZ1	CC97				91	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
10	3	1	3	2	3	4	2	13	5	7	4	4	3	2	1.5	BZ16	CC97				91	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
10	3	1	3	2	3	4	2	13	5	7	4	4	3	2	1.5	BZ17	CC97				91	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
10	3	1	3	2	3	4	2	13	5	7	4	4	3	2	1.5	BZ18	CC97		UJGFMBBBPB	t267	91	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
6	8	2	3	2	3	6	2	4	4	7	5	5	3	2	1.5	MT8164	CC97		UKBBPB	t1201	92	MSSA	Cattle	Drubec, Calvados, France	2008
6	8	2	3	2	3	6	2	4	4	7	5	5	3	2	1.5	MT8343	CC97				92	MSSA	Cattle	Drubec, Calvados, France	2008
8	8	3	4	2	3	6	2	13	4	5	4	2	3	2	2.5	91591	CC97		UJGFMBPB	t224	93	MSSA	Cattle	36157 Ebersberg, Germany	2006-2009
11	4	3	5	2	2	4	2	13	4	7	5	4	3	2	1.5	MT9279	CC97		TJGFMBBBBPB	t2112	94	MSSA	Cattle	Subligny, Manche, France	2009
10	8	3	1	2	2	8	2	13	5	7	3	3	3	2	1.5	BZ21	CC97		UJGFMBBBPB	t267	95	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
6	6	2	1	2	3	4	1	15	6	5	4	1	3	2	1.5	MT9315			UJGFMB	t189	96	MSSA	Cattle	Saint-Michel-de-Montjoie, Manche, France	2009
6	6	2	1	2	3	4	1	15	6	5	4	1	3	2	1.5	BZ7					96	MSSA	Cattle	Rio Grande do Sul, Brazil	1992-1993
5	6	5		2	3	2	4	16		6	3	4	3	2	1.5	MT9359					97	MSSA	Cattle	Reffuveille, Manche, France	2009
5	7	5	1	2	3	2	4	16	3	6	3	4	4	2	2.5	7H			ZNMOM	t3576	98	MSSA	Sheep	Pyrénées-Atlantiques, France	1996
9	3	1	2	2	3	6	4	13	5	7	2	3	4	1	6	MT8065	CC9				99	MSSA	Cattle	Laurent-de-Terregatte, Manche, France	2008
9	3	1	2	2	3	6	4	13	5	7	2	3	4	1	6	MT8074	CC9				99	MSSA	Cattle	Laurent-de-Terregatte, Manche, France	2008
9	3	1	3	2	3	6	4	13	5	7	2	3	4	1	6	92344	CC9				100	MSSA	Cattle	65589 Hademar, Germany	2006-2009
10	3	1	3	2	3	6	4	13	5	7	2	3	4	1	6	MT8041	CC9		UKJJJAGJAB	t3131	101	MSSA	Cattle	Chèvreville, Manche, France	2008
10	3	1	3	2	3	6	4	13	5	7	2	3	4	1	6	MT8042	CC9				101	MSSA	Cattle	Chèvreville, Manche, France	2008
9	3	1	3	2	3	6	4	13	5.5	7	2	3	4	1	4	90758	CC9				102	MSSA	Cattle	36110 Schlitz, Germany	2006-2009
9	3	1	3	2	3	6	4	13	5.5	7	2	3	4	1	4	91586	CC9				102	MSSA	Cattle	36145 Hofbieber, Germany	2006-2009
8	3	1	3	2	3	6	4	13	4	3	2	3	4	1	6	90145	CC9				103	MSSA	Cattle	37296 Ringgau, Germany	2006-2009
9	3	1	3	2	3	4	3	13	4	7	2	4	4	1	5	MT8002	CC9		UKJJAGJAB	t337	104	MSSA	Cattle	Saint-Jean-de-Daye, Manche, France	2008
9	3	1	3	2	3	4	3	13	4	7	2	4	4	1	5	MT8682	CC9				104	MSSA	Cattle	Saint-Côme-du-Mont, Manche, France	2008
10	3	1	3	2	3	4	3	13	4	7	2	4	4	1	5	MT8101	CC9				105	MSSA	Cattle	Saint-Côme-du-Mont, Manche, France	2008
9	7	4	3	2	3	5	5	15	2	6	5.5	5	3	2	1	M51	CC398	ST398	XKAOAOBQO	t034	106	MRSA	Cattle	74589 Satteldorf, Germany	2009
9	7	4	3	2	3	5	5	15	2	6	5.5	5	3	2	1	M53	CC398	ST398	XKAOAOBQO	t034	106	MRSA	Cattle	Baden-Württemberg, Germany	2009
7	7	4	3	2	3	5	5	15	2	6	5.5	5	3	2	1	M1	CC398	ST398	XKAOBQO	t011	107	MRSA	Cattle	32257 Bünde, Germany	2009
7	7	4	3	2	3	5	5	15	2	6	5.5	5	3	2	1	M3	CC398	ST398	XKAOBQO	t011	107	MRSA	Cattle	89296 Osterberg, Germany	2009
7	7	4	3	2	3	5	5	15	2	6	5.5	5	3	2	1	M5	CC398	ST398	XKAOBQO	t011	107	MRSA	Cattle	27446 Farven, Germany	2009
7	7	4	3	2	3	5	5	15	2	6	5.5	5	3	2	1	M12	CC398	ST398	XKAOBQO	t011	107	MRSA	Cattle	49328 Melle, Germany	2009
7	7	4	3	2	3	5	5	15	2	6	5.5	5	3	2	1	M40	CC398	ST398	XKAOBQO	t011	107	MRSA	Cattle	Baden-Württemberg, Germany	2009
7	7	4	3	2	3	5	5	15	2	6	5.5	5	3	2	1	M50	CC398	ST398	XKAOBQO	t011	107	MRSA	Cattle	Baden-Württemberg, Germany	2009
7	7	4	3	2	3	5	5	15	2	6	5.5	5	3	2	1	M60	CC398	ST398	XKAOBQO	t011	107	MRSA	Cattle	Baden-Württemberg, Germany	2009
7	7	4	3	2	2	5	5	15	2	6	5.5	5	3	2	1	M7	CC398	ST398	XKAOBQO	t011	108	MRSA	Cattle	Bayern, Germany	2009
7	7	4	3	2	2	5	5	15	2	6	5.5	5	3	2	1	M44	CC398	ST398	XKAOBQO	t011	108	MRSA	Cattle	Baden-Württemberg, Germany	2009
7	7	4	3	2	3	5	3	15	2	6	5.5	5	3	2	1	M58	CC398	ST398	XKAOBQO	t011	109	MRSA	Cattle	Baden-Württemberg, Germany	2009
7	7	4	3	2	3	5	5	15	2	6	4.5	5	3	2	1	M2	CC398	ST398	XKAOBQO	t011	110	MRSA	Cattle	32469 Petershagen, Germany	2009
8	7	4	3	2	3	5	5	15	2	6	4.5	5	3	2	1	M11	CC398	ST398	XGKAOBQO	t2576	111	MRSA	Cattle	26655 Westerstede, Germany	2009
9	7	4	3	2	3	3	5	15	2	6	5.5	5	3	2	1	M6	CC398	ST398	XKAOAOBQO	t034	112	MRSA	Cattle	26901 Lorup, Germany	2009
7	7	3	3	2	3	5	5	16	2	6	5.5	5	3	2	1	M9	CC398	ST398	XKAOBQO	t011	113	MRSA	Cattle	Bayern, Germany	2009
9	5	2	3	2	3	5	6	13	7	5	6.5	2	4	2	1	32D	CC25				114	MSSA	Goat	Banon, Alpes de Haute Provence, France	2005
10	6	3	3	4	3	5	5	14	9	4	3	3	3	2	5.5	MT9057			WGKKAKA[r348]M		115	MSSA	Cattle	Herqueville, Manche, France	2009

Figure S2. Mean number of repeat units per locus and for each family (human-related, animal-specific, mammary gland-adapted).



Résumé

Les espèces bactériennes exhibent plusieurs états de structure de populations pouvant varier de clonale à panmictique selon l'importance des transferts horizontaux et la nature de leur écosystème. Dans mon travail de thèse, je me suis intéressé à trois espèces bactériennes, *Staphylococcus aureus*, *Legionella pneumophila* et *Pseudomonas aeruginosa* qui reflètent trois situations différentes. Afin de pouvoir décrire de façon rapide de grandes collections de souches, j'ai utilisé comme marqueurs de diversité le polymorphisme de séquences répétées en tandem appelées VNTRs, pour *Variable Number Tandem Repeat*. La méthode MLVA, ou *Multiple Loci VNTR Analysis*, est une méthode de typage moléculaire qui s'appuie sur l'étude concomitante du polymorphisme de plusieurs loci VNTRs. Dans un premier temps, j'ai conçu des protocoles de typage automatisés pour les trois espèces considérées, puis j'ai appliqué ces outils pour traiter de questions d'épidémiologie.

S. aureus, espèce à structure clonale, est un pathogène majeur responsable notamment de toxi-infections alimentaires collectives (TIAC). Les travaux réalisés ont permis de démontrer la spécificité d'hôte de certains complexes clonaux et l'origine humaine des cas de TIAC. *L. pneumophila* est un pathogène de l'environnement dont la structure de population est atypique : présumée panmictique dans la nature, la bactérie semble connaître une évolution clonale lorsque son écosystème est restreint, dans un milieu anthropique par exemple. L'étude épidémiologique menée sur la population de *L. pneumophila* dans la ville de Rennes a mis en évidence la présence d'un écotype, non impliqué dans les cas cliniques épidémiques, particulièrement adapté aux réseaux d'eau. *P. aeruginosa*, modèle de bactérie panmictique, colonise les bronches de patients atteints de mucoviscidose. Le suivi longitudinal de patients indique que les souches installées sont persistantes et quasi-exclusives de la niche qu'elles occupent.

L'exploration de cette diversité du monde bactérien est un préalable à l'investigation épidémiologique des maladies infectieuses. Avec un même outil moléculaire de première intention, cette thèse retrace l'épidémiologie et la structure de trois espèces bactériennes très différentes. L'adaptation à un nouvel environnement (hôte animal, niche écologique, organe) est l'occasion d'expansions clonales.

Abstract

Bacterial species exhibit diversity in their population structure varying from clonal to panmictic according to the abundance of horizontal transfer and the nature of their ecosystem. During my PhD, I focused on three bacterial species, *Staphylococcus aureus*, *Pseudomonas aeruginosa* and *Legionella pneumophila*, which reflect three different situations. To perform the characterisation of large strain collections, I studied the polymorphism of molecular markers called VNTRs for Variable Number Tandem Repeat. MLVA (Multiple Loci VNTR Analysis) is a PCR based typing method that relies on the concomitant analysis of several VNTRs loci. Initially, I designed automated typing protocols for the three species, then I applied these tools to address issues of epidemiology.

S. aureus, a clonal species, is a major cause of food poisoning. The present work confirmed the existence of host-specific clonal complexes and demonstrated the predominantly human origin of foodborne disease cases. *L. pneumophila* is an environmental pathogen whose population structure is atypical: it is presumed panmictic in the environment but the bacterium expands clonally when the ecosystem is restricted, in an anthropogenic habitat for instance. A long-term epidemiological monitoring of *L. pneumophila* populations in the city of Rennes highlighted the presence of an ecotype, not involved in epidemic cases, particularly adapted to hot water supply systems. *P. aeruginosa*, a well-described panmictic bacterium, colonizes CF patients' airways. The longitudinal monitoring of patients provided evidence that the settled strains were persistent and exhibited strong exclusivity for the occupied niche.

Exploring the bacterial world diversity is a prerequisite for epidemiological investigation of infectious diseases. Using a first-line molecular tool, these works trace the epidemiology and the population structure of three bacterial species. The adaptation to a new environment (animal host, ecological niche, organ) generally results in clonal expansions.