

UNIVERSITE DE LA REUNION

Ecole Doctorale des Sciences, Technologies et Santé E.D. n°542
UMR Peuplements Végétaux et Bio-agresseurs en Milieu Tropical, CIRAD
Laboratoire de la Santé des Végétaux, ANSES

THESE

Présentée par

Damien Richard

Pour obtenir le diplôme de Doctorat en Sciences

Spécialité : génomique, épidémiologie moléculaire, phytopathologie, bactériologie

**Microévolution et adaptation à une pression de sélection anthropique chez
Xanthomonas citri pv. *citri*, une bactérie pathogène des agrumes : dynamique
du compartiment plasmidique**

Soutenue le 05.03.2019 devant le jury composé de

Marie Agnès Jacques	Directrice de recherche, INRA	Rapporteur
Francois Balloux	Professeur, University college, Londres	Rapporteur
Nicolas Radomski	Chercheur, ANSES	Examineur
Nathalie Becker	Maître de conférences HDR, MNHN	Examinatrice
Stéphane Poussier	Professeur, Université de La Réunion	Président
Olivier Pruvost	Chercheur HDR, CIRAD	Directeur de Thèse
Pierre Lefeuvre	Chercheur, CIRAD	Invité

Remerciements

En premier lieu, j'adresse tous mes remerciements à Olivier Pruvost (CIRAD) mon directeur de thèse et à Aude Chabirand, co-encadrante (ANSES). Merci à tous les deux de m'avoir fait confiance.

Olivier, je n'oublierai pas tout le support que tu m'as apporté, et la quantité astronomique de temps que tu as passé à me conseiller et à relire ma prose. Je pense sincèrement qu'il y a un monde en « être doctorant » et « être doctorant encadré par Olivier ».

Pierre Lefeuvre, merci de t'être tant impliqué dans cette thèse. J'ai beaucoup apprécié nos travaux en binôme et ton extrême réactivité lors de ceux-ci ! Merci à Adrien Rieux pour ses nombreux conseils. Merci à vous deux d'avoir supporté mes apparitions continues dans votre bureau.

J'exprime également ma gratitude à Benoît Facon, Virginie Ravigné et Isabelle Robène pour leur appui scientifique.

J'adresse tous mes remerciements à Bruno Hostachy pour son appui et à Gilles Cellier pour sa camaraderie et ses nombreux conseils.

Merci à Stéphanie Javegny et Claudine Boyer qui ont su me guider durant les manipulations de laboratoire et qui réalisent au Cirad 3P un travail titanesque et de grande qualité.

Je remercie vivement Nils Tribot, Aurélie Montbel-Fontaine et Ines Daras qui se sont impliqués dans ce projet de recherche à travers des stages de quelques mois.

Un grand merci également aux membres des différents comités de thèse avec qui j'ai pu discuter librement : Benoit Barres, Benoit Doublet et Philippe Roumagnac.

Merci également aux évaluateurs pour leur relecture: Francois Balloux et Marie Agnès Jacques que je ne connais pas encore à l'écriture de ce manuscrit ; mais aussi la sympathique et souriante Nathalie Becker, et Nicolas Radomski, fervent promoteur de la bioinformatique que j'ai rencontré à Paris pendant un cycle de conférences de l'Anses.

Un petit mot pour Jacques David et Grégoire Marandel, qui j'imagine seront contents de savoir que leurs conseils et avis ont été déterminants dans mon choix de me lancer dans l'aventure de la thèse.

Pour finir, toutes mes pensées pour Philippe Prior.

Liste des sigles

CC : Composant Connecté

CuR : Résistant au cuivre

CuS : Sensible au cuivre

EPS : ExoPolySaccharide

HGT : Transfert horizontal de gènes (*Horizontal Gene Transfer*)

ICE : Élément conjugatif intégré (*Integrated Conjugative Element*)

IS : Séquence insertionnelle (*Insertion Sequence*)

LPS : LipoPolySaccharide

MGE : Élément génétique mobile (*Mobile Genetic Element*)

MLVA : Analyse de répétitions en tandem multi-site (*MultiLocus Variable number of tandem repeat Analysis*)

MOB : Système de MOBilité d'un plasmide

PAMP : Motif moléculaire associé à un pathogène (*Pathogen-Associated Molecular Pattern*)

ROS : Espèces réactives oxygénées (*Reactive Oxygen Species*)

T2SS : Système de sécrétion de type 2 (*Type 2 Secretion System*)

T3SS : Système de sécrétion de type 3 (*Type 3 Secretion System*)

T4CP : Système de couplage de type 4 (*Type 4 Coupling Protein*)

T4SS : Système de sécrétion de type 4 (*Type 4 Secretion System*)

TALE : Effecteur activateur de transcription (*Transcription Activator-Like Effector*)

Xcc : *Xanthomonas citri* pv. *citri*

Sommaire

Contexte général.....	1
Revue bibliographique	3
I L'adaptation	3
I.1 L'adaptation en génétique.....	3
I.2 Les forces évolutives.....	4
I.3 Le cas des bactéries dites « monomorphes »	7
II Le transfert horizontal de gènes	8
II.1 Généralités	8
II.2 Distribution des gènes chez les individus	9
II.3 Facteurs contraignant les transferts horizontaux de gènes.....	12
II.4 Approches pour l'inférence de transferts de gènes	13
II.5 Réseau de similarité de gènes	14
II.6 Une diversité d'éléments génétiques mobiles	21
II.7 Le plasmide, véhicule pour les transferts horizontaux de gènes.....	27
III Substances antimicrobiennes en agriculture	40
III.1 Généralités.....	40
III.2 Les métaux lourds	41
III.3 Le cuivre.....	44
IV Une bactérie phytopathogène comme modèle d'étude : la bactérie monomorphe <i>Xanthomonas citri</i> pv. <i>citri</i>	55
IV.1 La plante-hôte : les agrumes.....	55
IV.2 Le chancre asiatique des agrumes : généralités.....	56
IV.3 Pouvoir pathogène.....	56
IV.4 Structure génétique.....	61

IV.5	Epidémiologie	63
IV.6	Méthodes de lutte	64
	Objectifs généraux.....	69
	Chapitre I Déterminisme génétique de la résistance au cuivre chez Xcc.....	71
I	Emergence de souches de Xcc résistantes au cuivre sur deux territoires ultramarins Européens	72
II	Séquençage de souches de Xanthomonadaceae résistantes au cuivre.....	73
III	Caractérisation des supports génomiques de la résistance au cuivre	74
IV	Phénotypes associés aux plasmides de CuR chez Xcc.....	75
V	Conclusions du chapitre	76
	Note 1 First report of <i>Xanthomonas citri</i> pv. <i>citri</i> pathotype A causing Asiatic citrus canker in Martinique, France	77
	Note 2 First report of copper-resistant <i>Xanthomonas citri</i> pv. <i>citri</i> pathotype A causing Asiatic citrus canker in Réunion, France.....	79
	Annnonce de génome 1 Complete genome sequences from a long read technology of six copper-resistant <i>Xanthomonas citri</i> pv. <i>citri</i> strains causing Asiatic citrus canker.....	81
	Annnonce de génome 2 Complete genome sequences from a long read technology of six copper-resistant <i>Xanthomonas</i> strains causing bacterial spot of solaneous plants: <i>X. gardneri</i> , <i>X. euvesicatoria</i> and <i>X. vesicatoria</i>	83
	Annnonce de génome 3 Complete genome sequence of a copper-resistant bacterium from the citrus phyllosphere, <i>Stenotrophomonas</i> sp. strain LM091, obtained using long-read technology	85
	Publication 1 Adaptation of genetically monomorphic bacteria: evolution of copper resistance through multiple horizontal gene transfers of complex and versatile mobile genetic elements	87
	Chapitre II La génomique pour caractériser l’histoire évolutive de Xcc dans l’Océan Indien	107
	Publication 2 Population-genomic insights into the evolution of the crop pathogenic bacterium <i>Xanthomonas citri</i> pv. <i>citri</i>	109

Chapitre III Analyse des transferts horizontaux chez les Xanthomonadaceae	137
Publication 3 The structure of bacterial communities as seen through the sharing of Xanthomonadaceae plasmid genes.....	139
Discussion générale.....	165
I La lutte contre les phyto bactérioses résistantes.....	165
II Le caractère monomorphe : un taux de mutation faible?	167
III Les plasmides, véhicules pour l'échange de gènes	169
IV Deux types de mobilité : focus sur les transposons Tn3 chez <i>Xanthomonas</i>	171
V Le rôle de réservoir de gènes des bactéries commensales.....	173
Littérature citée	177
Annexes.....	197

Contexte général

Les maladies des plantes occasionnent chaque année des pertes économiques de l'ordre de 40% de la production en Afrique et en Asie, et de 20% pour les pays développés (Somerville & Briscoe 2001). Dans les cas les plus extrêmes, ces maladies peuvent s'étendre à de larges portions d'un continent et mettre en péril la sécurité alimentaire, particulièrement lorsqu'elles touchent une des cultures piliers de l'alimentation comme le blé, le maïs, la pomme de terre ou le riz (Strange & Scott 2005). Les organismes phytopathogènes endémiques d'une région y infectent une ou plusieurs cultures de manière récurrente. Au contraire, les maladies dites émergentes sont caractérisées par leur apparition ou l'augmentation marquée de leur prévalence, dont les causes peuvent être variables et parfois multiples. Ces causes peuvent être l'introduction de leur agent causal sur un nouveau territoire, leur expansion suite à des changements environnementaux tels que ceux causés par le réchauffement climatique, le contournement d'une résistance variétale ou intrinsèque de leur plante hôte, un changement d'hôte ou encore le développement d'une résistance à un ou plusieurs pesticides (Anderson *et al.* 2004). L'agriculture intensive, et notamment la monoculture (McDonald & Stukenbrock 2016), ainsi que l'explosion des échanges de marchandises agricoles et alimentaires et des flux humains mondiaux (Hulme 2009) ont largement facilité l'émergence de bio-agresseurs des cultures et leur expansion à de grandes échelles. Parmi les bioagresseurs émergents des plantes cultivées, les virus représentent le groupe le plus fréquent, suivi des champignons et des bactéries phytopathogènes qui représentent environ un sixième des cas d'émergence. Les cas d'émergence bactérienne sont le plus souvent liés à l'établissement de populations pathogènes préadaptées à leur(s) hôte(s) dans une nouvelle région. Cependant, ces émergences résultent parfois de mécanismes adaptatifs bactériens, qui peuvent être responsables du contournement de résistances végétales ou du développement de résistance à un pesticide (Anderson *et al.* 2004).

Pour lutter contre les bactéries pathogènes des plantes, les recherches publique et privée travaillent sur plusieurs thématiques : développement de pratiques agronomiques adaptées, prophylaxie, amélioration et sélection des plantes, élaboration de pesticides naturels et de synthèse ou encore développement de mesures de biocontrôle. Le contrôle des bactéries phytopathogènes repose souvent sur des stratégies dites intégrées combinant plusieurs pressions indépendantes ayant chacune un effet réducteur sur les effectifs de pathogènes. En pratique, la lutte chimique basée

majoritairement sur l'utilisation de métaux lourds ou parfois d'antibiotiques reste une composante majeure de gestion des maladies bactériennes des plantes. La toxicité de ces molécules repose sur la perturbation de différents mécanismes cellulaires : inhibition de la réplication ou de la transcription, inhibition de la synthèse protéique, inhibition de la synthèse de la paroi bactérienne, dégradation membranaire ou encore inhibition de la biosynthèse de métabolites essentiels (Lemire *et al.* 2013). Une pression de sélection récurrente peut conduire à l'apparition de génotypes résistants au sein d'une population pathogène. Un caractère adapté peut être causé par des mécanismes comme les mutations ponctuelles ou l'acquisition d'un ou plusieurs gènes de résistance. L'histoire de l'agriculture comporte de nombreuses illustrations de l'émergence de souches résistantes à des pesticides. La résistance à la streptomycine chez les agents de la gale bactérienne des solanées et le feu bactérien des pomacées restent des exemples emblématiques de l'émergence de résistances chez des bactéries phytopathogènes (Jones 1982). De la même manière, des cas de résistance au cuivre ont été décrits depuis plusieurs décennies chez plusieurs genres de bactéries phytopathogènes (Cooksey 1990). Plus récemment, des souches résistantes au cuivre de *Xanthomonas citri* pv. *citri* (bactérie pathogène des agrumes) ont été détectées dans deux territoires ultramarins Français. L'apparition de cette résistance de manière concomitante dans différentes régions du monde ainsi que l'hypothèse du support plasmidique de ce phénotype suggèrent une étiologie complexe. Cette dernière constitue l'objet d'étude de ce travail de thèse.

Revue bibliographique

Revue bibliographique

I L'adaptation

I.1 L'adaptation en génétique

L'adaptation d'une population d'organismes vivants a été définie comme une réponse à l'environnement (Fisher 1930). Ce processus améliore la performance du phénotype pour un environnement donné. En biologie évolutive, thématique dans laquelle s'inscrit cette thèse, des définitions plus précises et récentes sont proposées: "Lorsque l'on considère l'échelle populationnelle et les relations interspécifiques, l'adaptation comme processus est un succès reproductif différentiel entre variants obtenu par sélection naturelle. L'adaptation comme résultat est un état dérivé de caractère conférant un avantage testable dans un milieu donné" (Lecointre 2012); "L'adaptation peut-être définie de manière complète comme un caractère nouveau apparu chez un organisme et maintenu par la sélection naturelle" (Heams *et al.* 2009).

L'adaptation, ainsi définie comme héritable, relève principalement de la génétique. Les facteurs épigénétiques peuvent également être pris en considération mais ne sont pas développés dans cette thèse. La génétique a pour objet d'étude le génome, qui code l'information nécessaire au développement des êtres vivants. Des modifications au sein du génome peuvent donc faire varier le phénotype d'un individu. La transmission complète ou partielle du génome d'un individu à un autre permet la diffusion de ces nouveaux phénotypes dans la population. L'adaptation (utilisée ici et dans l'ensemble du manuscrit au sens génétique du terme) repose donc sur l'apparition de variations au sein du génome, créant différents génotypes. L'exemple emblématique d'adaptation génétique est celui de la phalène du bouleau, un papillon de nuit dont la coloration des ailes s'est assombrie avec l'augmentation de la pollution industrielle du XIX^{ème} siècle. La coloration sombre, dont le déterminisme génétique repose sur la présence d'un allèle dominant, permet en effet un meilleur camouflage du papillon sur un support sali par des dépôts de poussières industrielles (Kettlewell 1958). L'adaptation par le biais de la variation du génome a dans ce cas été rapide (en quelques années) et localisée (autour de certaines grandes villes industrielles de Grande-Bretagne). Dans le cas des bactéries, leurs temps de génération sont courts, et leurs génomes présentent généralement des taux de mutation élevés. Ainsi, l'adaptation à une nouvelle

contrainte environnementale peut opérer en quelques jours voire quelques heures (Baym *et al.* 2016).

I.2 Les forces évolutives

Au début du XX^{ème} siècle, l'existence de mécanismes à l'origine de la diversité génétique entre individu est connue (Muller 1927), et des théories expliquant l'évolution, comme le Darwinisme, ont été formulées. La combinaison des deux, nécessaire pour appréhender la génétique évolutive, sera faite au milieu du siècle, entre autres par J. Huxley, qui publie en 1942 le livre *Evolution : the modern synthesis* (Huxley 1942). Le néodarwinisme est né : la diversité héritable est créée au sein des populations, et la fréquence des différents variants est modifiée par la sélection naturelle. Cinq forces évolutives formalisent aujourd'hui les mécanismes de création et de variation en fréquence des allèles au sein d'une population (Figure 1).

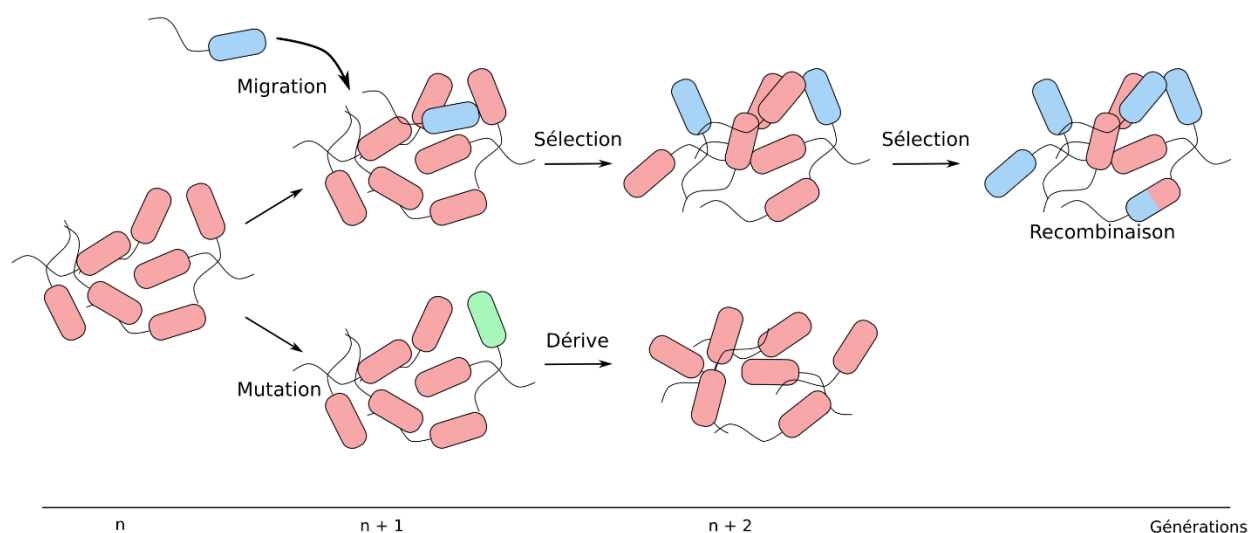


Figure 1 Représentation de l'action des forces évolutives sur une population bactérienne des générations n à $n+2$. La couleur de la bactérie représente un génotype. Les génotypes recombinants sont représentés par des bactéries de plusieurs couleurs.

I.2.a *La mutation*

Le terme mutation est ici utilisé au sens large, et comprend donc la substitution d'un nucléotide par un autre ainsi que les insertions/délétions de nucléotides. Les génomes bactériens ont généralement un taux de substitution de l'ordre de 10^{-7} substitution par site et par an (Duchene *et al.* 2016) tandis qu'il est de l'ordre de 10^{-9} chez l'humain (Scally & Durbin 2012) et jusqu'à 10^{-3} pour certains virus à ARN (Jenkins *et al.* 2002). On considère que l'évolution par mutation est un processus graduel,

qui modifie le phénotype au fil des générations. Cependant, dans certains cas, une seule ou quelques mutations peuvent être suffisantes pour, par exemple, dans le cas d'interactions hôte-pathogène, conférer un phénotype résistant à l'hôte ou au contraire la virulence au parasite (Dimijian 1999; Ebert 1998; Hurtle *et al.* 2003).

I.2.b Sélection

La sélection naturelle est le processus selon lequel les allèles avantageux pour la survie et la reproduction d'une population dans un environnement donné augmentent en fréquence tandis que ceux qui sont désavantageux voient au contraire leur fréquence diminuer. En outre, la force de la sélection appliquée sur un allèle ou une combinaison d'allèles dépend de la variation de valeur sélective¹ entre l'individu porteur et la moyenne de cette valeur sélective pour la population. L'action de cette force évolutive permet le déplacement du phénotype d'une population en direction d'un nouvel optimum, engendrant des niveaux de différenciation qui relèvent alors de l'adaptation, comme dans le cas de la phalène du bouleau présenté plus haut (voir I.1.).

I.2.c Migration

En génétique, la migration correspond à un déplacement de matériel génétique d'une population à une autre. Ce flux de gènes peut se présenter sous la forme d'un mouvement de gamètes (*e.g.* du pollen transporté par le vent), d'individus (*e.g.* une graine transportée par un oiseau) ou encore de fragments d'ADN (cas particulier traité en II). Chez les bactéries, la présence de migrations récurrentes entre deux populations a tendance à homogénéiser leurs fréquences alléliques jusqu'à ce que les deux populations deviennent indiscernables et n'en forment plus qu'une. La migration d'un nombre restreint d'individus vers une nouvelle niche écologique crée une population nouvelle qui ne présente qu'une fraction de la variabilité génétique de la population fondatrice.

I.2.d Dérive génétique

Lors de la colonisation d'une nouvelle niche par un faible nombre d'individus, l'effet de fondation, c'est à dire la dérive génétique associée à la migration, sera très fort. Cet effet de fondation peut amener la fixation d'allèles rares dans la nouvelle population, et ainsi générer une population

¹ La valeur sélective illustre de manière quantitative le succès reproducteur d'un individu. En génétique des populations, cette mesure correspond au nombre de descendants viables et fertiles que produit en moyenne chaque individu de ce génotype.

divergente. Cette force évolutive représente la variation de la fréquence des allèles dans une population suite à des phénomènes aboutissant à l'échantillonnage aléatoire d'individus, comme par exemple dans le cas d'un goulet d'étranglement. La spéciation par la seule action de la dérive génétique est difficile à démontrer dans la nature, mais son action, couplée à celle de la sélection naturelle, a un impact considérable sur l'évolution des espèces. En effet, la sélection n'opère que si plusieurs génotypes ont des valeurs sélectives différentes. Le caractère aléatoire de la dérive génétique permet en revanche l'exploration de la diversité de génotypes adaptés ayant des valeurs sélectives équivalentes. Le passage de manière transitoire par l'un d'eux peut permettre d'accéder par sélection à un génotype de valeur sélective plus forte. La perte aléatoire de certains allèles suite à des événements de dérive est d'autant plus rapide que la taille de la population est faible.

I.2.e Recombinaison et régime de reproduction

Chez les eucaryotes, la recombinaison homologue¹ lors de la méiose crée de nouvelles combinaisons alléliques. Si deux individus parents non clonaux ne sont pas homozygotes, leur descendance ne sera pas non plus homozygote. Elle présentera de plus une association d'allèles différente de celle du génome de ses parents. La fréquence d'apparition de nouvelles combinaisons dépend du régime de reproduction, qui influence la dynamique évolutive d'une population en déterminant comment les allèles s'échangent dans la population. Si les allèles peuvent s'échanger librement entre individus on parle de régime panmictique et s'ils ne s'échangent pas du tout on parle de régime clonal. Le régime de reproduction d'une population correspond à sa capacité à effectuer de la recombinaison au sens large (homologue et hétérologue²). Seul, il n'a un effet que sur l'association des allèles mais pas sur la fréquence de ceux-ci dans la population. Par contre, lorsqu'un gène est soumis à une sélection positive, la taille du segment d'ADN sur laquelle se porte la sélection varie en fonction du régime de reproduction. En effet, elle se limite au gène (et aux régions avoisinantes en déséquilibre de liaison) dans le cas d'un régime de reproduction panmictique, alors qu'elle concerne l'intégralité du génome comprenant le gène adaptatif dans le cas d'un régime clonal. La stricte clonalité apparaît ici être un frein à l'adaptation.

¹ En génétique, la recombinaison correspond à la cassure et au réassemblage de deux brins d'ADN. Lorsque la recombinaison est qualifiée d'homologue, les séquences de nucléotides impliquées sont (homologues) ou similaires (homéologues).

² La recombinaison hétérologue, par opposition à la recombinaison homologue, implique des séquences de nucléotides différentes. Généralement, ce processus correspond à l'intégration d'un fragment d'ADN dans un autre.

Les bactéries se reproduisent par mitose¹, donc de façon clonale, sans échange de matériel génétique entre individus durant la phase de reproduction. La réalité est cependant plus complexe, et l'idée selon laquelle les bactéries ne sont pas strictement clonales a émergé dans les années 1940 (Tatum & Lederberg 1947). En effet, les régimes de reproduction des populations bactériennes s'étendent selon un continuum allant de panmictique à clonal et varient selon les groupes d'individus étudiés (Maynard Smith *et al.* 1993). En réalité, il est peu pertinent de parler de régime de reproduction dans le cas des populations bactériennes, puisque la recombinaison à l'origine des nouvelles combinaisons alléliques a lieu en dehors de la phase de reproduction (Istock *et al.* 1992; Spratt *et al.* 2001). Chez les bactéries, la recombinaison non-méiotique peut être homologue ou non-homologue. Plusieurs voies de recombinaison homologue existent et sont souvent également impliquées dans la réparation de l'ADN (Rocha *et al.* 2005). Associées à des régions nucléotidiques répétées, ces processus peuvent provoquer des réarrangements chromosomiques et ainsi générer de la diversité génotypique. Les événements de recombinaison inter-individus sont appelés transferts horizontaux de gènes et reposent sur plusieurs mécanismes détaillés en II.6. Certaines espèces comme *Helicobacter pylori* présentent une structure de population compatible avec un assortiment des allèles au hasard et une recombinaison libre (panmixie) alors que d'autres, comme *Mycobacterium tuberculosis* présentent de très forts déséquilibres de liaison² entre loci, suggérant une population clonale (Sheppard *et al.* 2018). Des régimes de reproduction transitoires peuvent également apparaître comme dans le cas où un seul génotype d'une bactérie panmictique devient épidémique (clonalité épidémique). L'épidémie possède alors une base génétique très étroite en lien avec le faible nombre d'individus fondateurs. L'étude génétique de ces individus épidémiques montre une structure clonale (Smith *et al.* 2000).

I.3 Le cas des bactéries dites « monomorphes »

Les études de diversité génétique bactérienne ont montré que certaines lignées présentaient cependant un nombre de polymorphismes très faible, voire nul. Les bactéries dont la structure génétique ne peut être résolue par l'analyse des séquences nucléotidiques de quelques gènes ont alors été qualifiées de monomorphes (Achtman 2008). Les bactéries monomorphes comprennent

¹ La mitose correspond à la division d'une cellule mère en deux cellules filles. Ces dernières sont génétiquement strictement identiques. Ce type de reproduction est non sexué.

² Le déséquilibre de liaison indique une fréquence des allèles de deux loci différente de la fréquence théoriquement obtenue avec une association aléatoire de ces allèles. Ce déséquilibre reflète une association préférentielle entre deux allèles, et reflète donc l'absence de recombinaison entre ces deux allèles, ou leur forte contre-sélection.

un grand nombre de lignées épidémiques et pathogènes de l'humain, comme *Mycobacterium tuberculosis*, *Yersinia pestis* ou *Escherichia coli* O157:H7 (Achtman 2008)), ou phytopathogènes comme *Pseudomonas syringae* pv. *tomato* (Cai *et al.* 2011) ou *Xanthomonas citri* pv. *citri* (Leduc *et al.* 2015). L'avènement des technologies de séquençage à haut débit et le grand nombre de génomes bactériens complets actuellement disponibles ont cependant mis en évidence un continuum de niveau de diversité de séquence chez les bactéries plutôt qu'une séparation franche entre les bactéries monomorphes et les autres (Achtman 2008). Malgré un faible niveau de diversité de séquence, plusieurs lignées monomorphes ont montré de grandes capacités adaptatives parfois causées par des mutations ponctuelles mais provenant souvent d'évènements de transfert horizontal de gènes (Behlau *et al.* 2013; Galimand *et al.* 1997; Nonaka *et al.* 2012).

II Le transfert horizontal de gènes

II.1 Généralités

De nombreux organismes procaryotes possèdent la faculté d'échanger du matériel génétique de manière horizontale (c'est-à-dire, autrement que par descendance), ce qui correspond à de la recombinaison inter-individus (voir I.2.e). Des portions entières de génome peuvent être échangées par des bactéries d'espèces, voire de familles différentes. Des exemples d'échanges entre règnes du vivant ont même été mis en évidence (Jaramillo *et al.* 2015; Keeling & Palmer 2008). Ce phénomène est appelé « transfert horizontal de gènes » (HGT, pour *Horizontal Gene Transfer*), et a des implications considérables sur la structure génétique des populations bactériennes, leur dynamique évolutive ainsi que leur capacité d'adaptation (Fraser *et al.* 2009; Gogarten *et al.* 2002; Gogarten & Townsend 2005; Koonin *et al.* 2001).

L'échange, entre organismes, d'éléments génétiques codant pour des traits adaptatifs peut permettre une adaptation rapide, dans les cas où les gènes transmis sont adaptés à la niche écologique de l'organisme receveur. La majorité des transferts horizontaux bactériens impliquent non pas un gène mais un ou plusieurs clusters de gènes codant souvent pour des voies métaboliques complètes. Dans cette situation, les transferts permettent alors non pas une adaptation pas-à-pas, mais un véritable « bon en avant » adaptatif (du terme anglais « *quantum leaps* ») (De la Cruz & Davies 2000; Paauw *et al.* 2010).

Les gènes acquis horizontalement peuvent représenter une part non-négligeable d'un génome bactérien (Lapierre & Gogarten 2009). L'étude de Lawrence and Ochman (1997) a d'abord permis de montrer que le génome d'*Escherichia coli* comportait environ 600kb¹ d'ADN acquis horizontalement à un taux estimé alors de 31kb par million d'année. Depuis, de nombreux travaux ont permis d'étendre ce constat (McDaniel *et al.* 2010; Nakamura *et al.* 2004; Ochman *et al.* 2000; Welch *et al.* 2002). Kloesges *et al.* (2011) ont aussi montré que, chez les bactéries, près de 75 % des familles de gènes ont été échangées horizontalement dans un passé récent. D'autres estimations montrent, suivant la méthode utilisée ou le jeu de données, des proportions de gènes transférés variant entre 0 et 17 % pour les bactéries (Garcia-Vallvé *et al.* 2000; Lawrence & Ochman 1997; Ochman *et al.* 2000). Quel que soit le taux exact de transfert, il apparaît maintenant très clairement que la composante horizontale de l'évolution des génomes est essentielle pour décrire l'évolution d'un grand nombre de groupes bactériens (Halary *et al.* 2010; Popa & Dagan 2011; Spratt *et al.* 2001) (Figure 2).

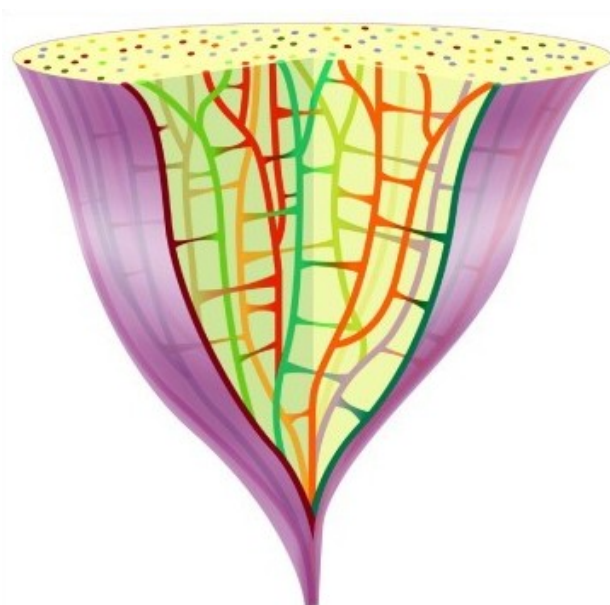


Figure 2 Représentation schématique de l'arbre phylogénétique des génomes microbiens incluant les événements de transfert horizontal. Les différentes couleurs représentent les différentes lignées bactériennes, et les connexions horizontales entre lignées représentent les événements de transfert horizontal de matériel génétique. Figure tirée de Popa & Dagan (2011).

II.2 Distribution des gènes chez les individus

¹ Le kb, de l'anglais *kilo-base pair*, mesure la taille d'une molécule d'ADN et correspond à 1000 nucléotides

L'existence de transferts horizontaux de gènes implique qu'une souche bactérienne peut acquérir un gène ou plusieurs gènes nouveaux, conduisant alors à un contenu en gènes variable entre souches parfois phylogénétiquement proches. Ainsi, certains gènes sont encodés dans tous les génomes des bactéries d'un niveau taxonomique donné (et forment le génome de base) alors que d'autres ne sont encodés que chez une fraction de ceux-ci (formant le génome dit accessoire). L'ensemble de ces gènes constitue le pan-génome de l'espèce. Le qualificatif d'« accessoire » est trompeur puisque ces gènes sont souvent essentiels à l'adaptation de bactéries à diverses conditions écologiques ou à la compétition inter-espèces (Wiedenbeck & Cohan 2011). Chez les espèces à pan-génome dit « ouvert », une part importante des gènes seront non conservés, à l'inverse des espèces à pan-génome dit « fermé » (Figure 3). Le séquençage de trois souches de *Escherichia coli* a par exemple mis en évidence que seuls 39% du total des gènes sont encodés chez chaque souche (Welch *et al.* 2002), alors que huit génomes de *Bacillus anthracis* séquençés (génétiquement distincts mais appartenant possiblement à un même clade (Tettelin *et al.* 2005)) ont révélé un contenu en gènes identique (Medini *et al.* 2005). Le cas d'un pan-génome strictement fermé est cependant rarement observé, et est peut-être conditionné par un échantillonnage trop restreint.

De manière plus générale, lors du séquençage d'un grand nombre de génomes de souches bactériennes de la même espèce, l'analyse de la fréquence des gènes dans la population montre que ceux-ci sont généralement présents à forte ou faible fréquence mais rarement à des fréquences intermédiaires, formant ainsi une distribution en forme de « U » communément appelée « *U-shape* » en anglais (Haegeman & Weitz 2012; Lobkovsky *et al.* 2013). Si les mécanismes sous-jacents à la formation de cette distribution, et en particulier le rôle de la sélection, restent encore discutés. Il est toutefois considéré que les gènes présents à faible fréquence ont été acquis horizontalement (Bolotin & Hershberg 2015). Le nombre de gènes rares, vraisemblablement acquis horizontalement, est très élevé au sein de très nombreux groupes bactériens (Gogarten & Townsend 2005). Ces gènes seraient particulièrement associés à des fonctions adaptatives et permettraient aux populations bactériennes de répondre rapidement aux changements environnementaux (Fraser *et al.* 2009).

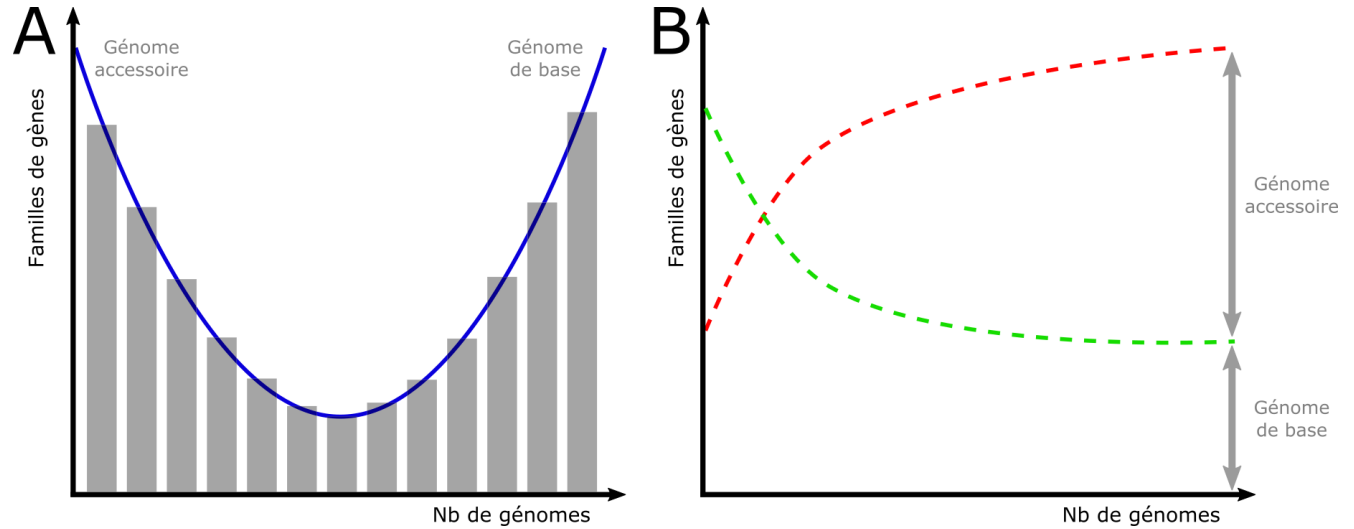


Figure 3 (A) Distribution du nombre de familles de gènes trouvées en fonction du nombre d'individus pris en compte. Cette distribution en forme de « U » (pointillés verts), communément appelée « *U-shape* » en anglais, est caractéristique de la présence d'une grande diversité de familles de gènes dans un nombre limité d'individus (les gènes dits accessoires) ainsi que dans un grand nombre d'individus (le génome de base), tandis qu'un moindre nombre de familles de gènes est présent dans une proportion d'individus moyenne. Cette distribution a été empiriquement observée dans de nombreux cas (Collins & Higgs 2012; Haegeman & Weitz 2012; Koonin 2015; Lobkovsky *et al.* 2013). **(B)** Evolution du nombre de familles de gènes retrouvées dans tous les génomes (en vert, le génome de base) et du nombre de familles de gènes totales (en rouge, le génome accessoire) en fonction du nombre de souches séquencées.

II.3 Facteurs contraignant les transferts horizontaux de gènes

Si l'importance des échanges horizontaux dans l'évolution microbienne n'est plus à démontrer, il apparaît aussi clairement que les flux de gènes ne se font pas librement entre tous les individus. Alors que des facteurs tels que le %GC¹ des génomes ou les taux d'usage des codons semblent en partie structurer les échanges de gènes entre génomes bactériens, un des grands débats sur les barrières aux transferts horizontaux repose sur l'importance relative de la parenté entre individus (proximité phylogénétique ou taxonomique) et de leur écologie respective. Ainsi de nombreuses études ont, ces dernières années, tour à tour tenté de démontrer que l'un ou l'autre façon de manière prioritaire la structure des échanges horizontaux de gènes entre bactéries (McInerney 2013; Polz *et al.* 2013; Skippington & Ragan 2012). Il apparaît toutefois difficile de conclure sur l'importance de chacune de ces barrières (taxonomique ou écologique) et de pouvoir étendre ces résultats à l'ensemble du microbiome (Popa & Dagan 2011) (Figure 4).

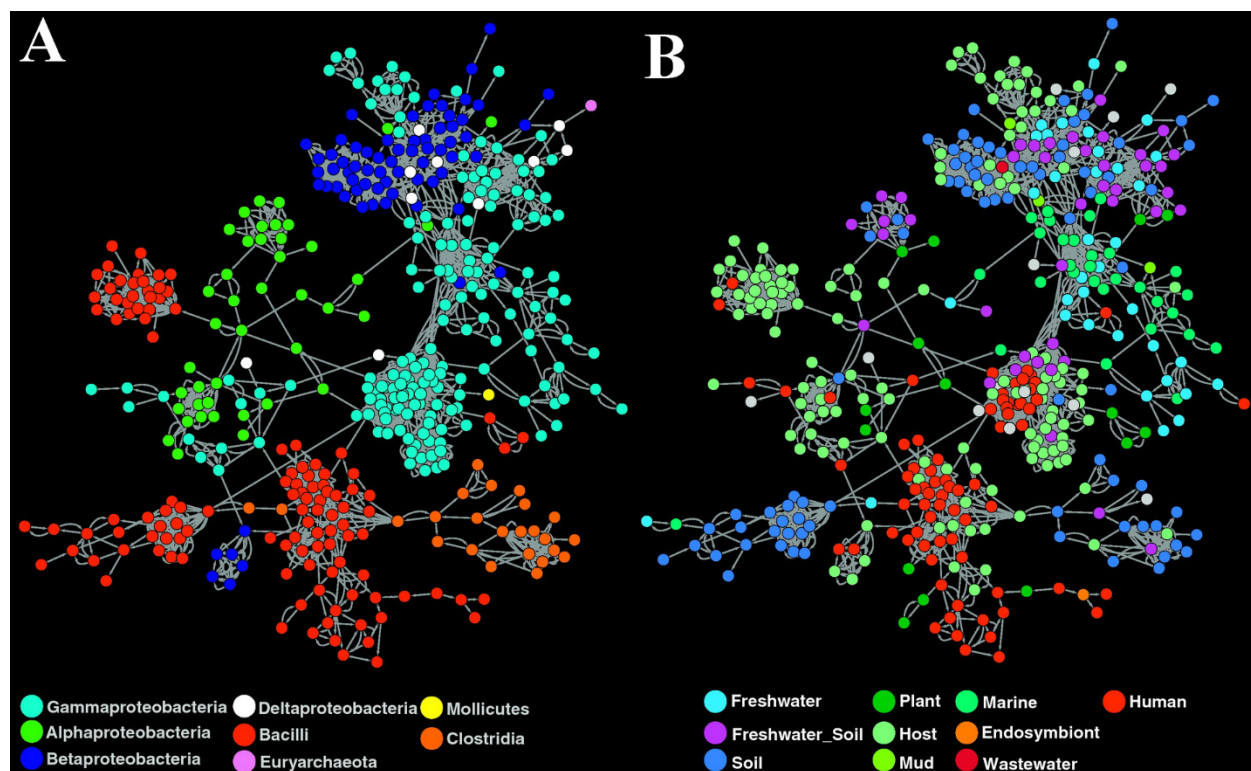


Figure 4 Représentation d'un réseau dirigé matérialisant les transferts horizontaux de gènes (liens gris) entre bactéries (cercles colorés). Alors qu'en (A), la couleur des nœuds correspond à la classification taxonomique

¹ Le % GC d'un génome ou d'une séquence en indique le pourcentage de guanine et de cytosine.

détaillée dans la légende, elle représente en (B) la niche écologique de l'organisme. Tiré de Popa and Dagan (2011)

La construction et l'analyse de réseaux d'échanges de gènes ont également permis d'identifier le type de vecteur d'ADN mobilisé pour les transferts horizontaux au sein de différents groupes bactériens. Les bactéries du genre *Streptococcus* échangent par exemple préférentiellement de l'ADN par le biais des bactériophages alors que celles du genre *Xanthomonas* le font majoritairement par des plasmides (Halary *et al.* 2010). De manière plus générale, les transferts horizontaux chez les protéobactéries (dont *Xanthomonas* fait partie) semblent majoritairement s'effectuer par conjugaison entre taxons différents partageant une même niche écologique (Kloesges *et al.* 2011).

En outre, les plasmides connectent un grand nombre de genres bactériens et semblent donc jouer un rôle prépondérant dans les échanges horizontaux entre bactéries (Halary *et al.* 2010). L'histoire évolutive des bactéries apparaît donc intimement liée aux transferts de plasmides. La reconstruction et l'analyse de ces transferts pourraient ainsi permettre de mieux comprendre l'évolution des communautés bactériennes (Corel *et al.* 2016).

II.4 Approches pour l'inférence de transferts de gènes

L'histoire évolutive des bactéries peut être appréhendée par la comparaison des séquences génétiques conservées entre plusieurs individus. Pour cela, la création d'un arbre phylogénétique où les nœuds externes représentent des souches contemporaines et les nœuds internes leurs ancêtres, permet à elle seule de comprendre une grande variété de phénomènes évolutifs. Cependant, cette vision de l'évolution bactérienne, pourtant très utilisée, reste restrictive, puisqu'elle ne permet que l'étude des séquences conservées entre les souches considérées et se limite à la composante verticale de leurs relations de parenté. À l'heure des technologies de séquençage à haut débit, l'étude de l'histoire évolutive des bactéries à l'échelle génomique devient possible, et peut maintenant inclure l'étude du signal apporté par les événements de gain et de perte de gènes (Dagan 2011).

En biologie, les réseaux sont souvent utilisés pour représenter des interactions complexes entre différentes entités, qui peuvent être des protéines et leurs substrats, les organismes d'un réseau trophique, les voies de signalisation ou de régulation cellulaire ou encore des réseaux neuronaux.

Les réseaux ont la particularité de ne pas limiter le nombre d'interactions entre nœuds et constituent donc un très bon outil pour appréhender les événements d'évolution réticulée tels que les hybridations, la recombinaison, la fusion de génomes ou le transfert horizontal de gènes (Huson & Scornavacca 2011).

Plusieurs types d'approches ont été utilisés pour mettre en évidence la composante horizontale de l'évolution bactérienne. Les réseaux *split* et *reticulate* forment deux grandes familles de représentations qui permettent d'inclure une composante horizontale à un arbre phylogénétique (Huson & Bryant 2006). Ces deux méthodes se basent par exemple sur la combinaison de multiples phylogénies de gènes (mais peuvent aussi être inférées à partir de matrice de distances ou d'autres données, décrites par Huson and Scornavacca (2011)), qui peuvent ne pas être concordantes. Les événements d'évolution réticulée y sont représentés par des liens reliant les branches de la phylogénie. La reconstruction d'états ancestraux de présence/absence de gènes le long d'une phylogénie permet également d'inférer les HGT et leur direction. D'autres approches ne s'appuient en revanche pas du tout sur les concepts de phylogénie et peuvent même ne représenter que des événements de transferts horizontaux. Ainsi, un intérêt croissant pour les réseaux de similarité de gènes est constaté ces dernières années, notamment du fait de la simplicité de leur mise en œuvre (Corel *et al.* 2016).

II.5 Réseau de similarité de gènes

La méthode des réseaux de similarité de gènes consiste généralement à représenter chaque individu par un nœud et à établir une connexion entre deux individus lorsque la comparaison de leur contenu en gène montre une similarité entre séquences correspondant aux critères d'identité et de longueur, ou d'*e-value*¹ de BLAST (comme dans Bolduc *et al.* (2017)), préalablement fixés (Figure 5). Si les liens ont un poids, en plus d'indiquer une interaction entre nœuds, ils indiqueront aussi leur importance quantitative. Enfin, les liens peuvent être dirigés, indiquant alors la direction du transfert (*e.g.* transfert de gènes d'une souche donneuse à une souche receveuse).

¹ L'*e-value* associée à un résultat d'homologie lors d'une comparaison Blast entre une séquence et une base de données de séquences est un paramètre qui décrit le nombre de résultats de recherche (hits) de ce niveau d'homologie attendus par chance lors de la recherche de cette séquence dans une base de donnée de même taille. Une *e-value* proche de zéro signifie donc que l'on ne s'attend pas à observer une telle homologie par chance.

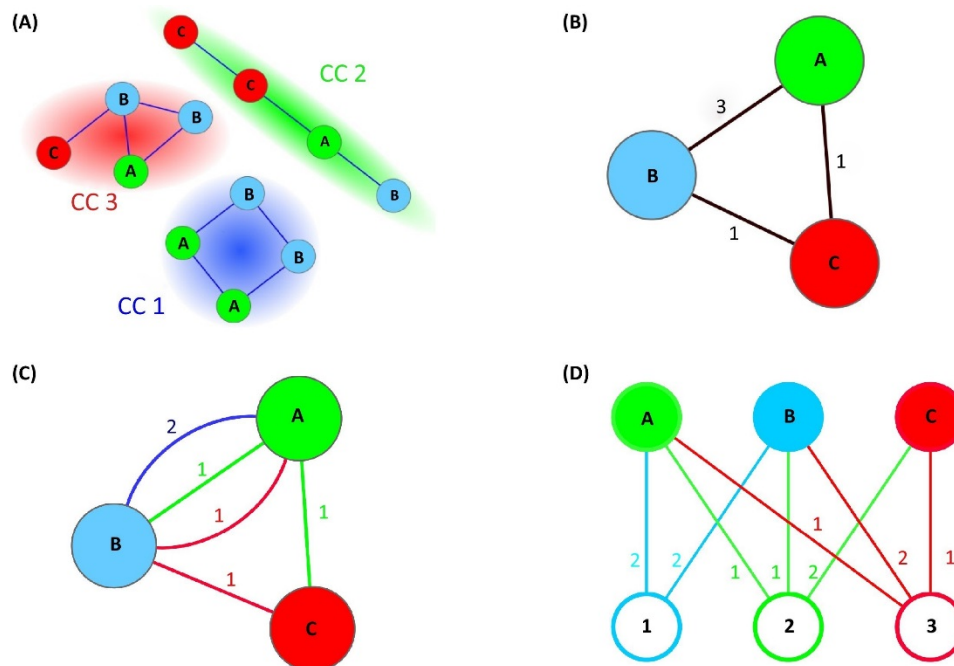


Figure 5 Différentes représentations d'un même réseau d'échanges de gènes entre plusieurs individus, notés A, B et C. (A) représentation d'un réseau de similarité de gènes : chaque nœud représente un gène, la couleur et le nom de chaque nœud indique l'individu duquel il provient. Deux nœuds sont connectés par un lien si les deux gènes possèdent une identité supérieure aux critères définis (pourcentage d'identité et longueur). Ce réseau comporte trois composants connectés (CC) correspondant à des familles de gènes distinctes. (B) Les nœuds représentent cette fois les individus et le poids des liens est fonction du nombre de gènes partagés. Les liens ne détaillent en revanche pas quelles familles de gènes sont communes entre deux génomes contrairement à la représentation en (C), où chaque famille de gènes est représentée par une couleur de lien différente. (D) Dans cette dernière représentation, deux type de nœuds existent, avec les individus (en haut) et les familles de gènes en bas. Les liens ne peuvent ici connecter que des nœuds de classes différentes. Figure tirée de Corel *et al.* (2016).

Le seuil de similarité choisi pour la construction des liens entre individus n'est pas anodin puisqu'il influe directement sur l'âge des événements de transferts identifiables. En effet, dans le cas d'un réseau construit selon l'identité entre séquences et considérant l'accumulation temporelle de mutations dans les séquences nucléotidiques, le seuil d'identité utilisé conditionnera l'ancienneté des transferts mis en évidence (i.e. une identité haute correspondant à des transferts récents) (Halary *et al.* 2010). Ainsi, les réseaux de similarité de gènes ont permis à plusieurs reprises d'appréhender l'histoire évolutive d'organismes, et ce à des échelles spatio-temporelles très variables. Une analyse de réseau supporte par exemple la théorie de la naissance des eucaryotes par des phénomènes endosymbiotiques, alors que l'ancienneté et la multiplicité de ces événements ne permettent pas de créer un seul arbre phylogénétique décrivant ce processus (Alvarez-Ponce *et al.* 2013). Les réseaux

de similarité de gènes ont également permis d'identifier des familles de gènes fréquemment échangées et leur évolution après leur transfert (Kloesges *et al.* 2011), les facteurs limitant les transferts horizontaux de gènes (Popa & Dagan 2011), et bien d'autres événements évolutifs (Corel *et al.* 2016).

Deux types d'évènements peuvent être à l'origine d'homologie de séquences entre individus : l'apparentement et les HGT. Pour exclure des liens du réseau l'homologie due à l'apparentement et ne garder que l'homologie due aux transferts horizontaux, plusieurs approches sont couramment utilisées. La conservation des régions génomiques adjacentes à un élément génétique mobile (voir II.6) entre deux individus indique qu'ils l'ont probablement héritée d'un ancêtre commun (Hooper *et al.* 2009). La recherche des meilleurs homologues bidirectionnels (*bidirectional best hit*) est souvent employée pour trouver les gènes orthologues (hérités d'un ancêtre commun) par paire de génomes, mais cette approche introduit des erreurs dans quelques cas particuliers détaillés par Gabaldón and Koonin (2013). Dans un échantillonnage taxonomiquement vaste, les gènes échangés horizontalement sont plus faciles à identifier puisqu'ils possèdent des niveaux d'homologie élevés alors que leurs génomes hôtes sont plus divergents (Fondi & Fani 2010). Cette dernière stratégie, basée sur l'hypothèse d'horloge moléculaire, a été utilisée par Smillie *et al.* (2011) qui considèrent que les éléments génétiques très conservés (identité nucléotidique > 99%) entre individus montrant une séquence d'ARN ribosomal 16S divergente (identité nucléotidique < 97%) proviennent d'un événement de transfert horizontal. Ces méthodes peuvent être utilisées de façon conjointe ou adaptées en fonction des caractéristiques du jeu de données ainsi que l'ancienneté et la nature des éléments génétiques mobiles ciblés par l'étude. L'objectif est dans tous les cas l'identification de toutes les paires de génomes qui ont échangé un élément génétique de façon horizontale.

Les premières représentations des transferts horizontaux entre plusieurs génomes bactériens datent de 2005, et analysent l'ensemble des génomes bactériens (moins de 200) dont la séquence nucléotidique était disponible à l'époque (Beiko *et al.* 2005; Kunin *et al.* 2005). Depuis, l'augmentation exponentielle du nombre de génomes séquencés rend leur étude exhaustive techniquement difficile à réaliser. En conséquence, les réseaux d'échange de gènes sont généralement utilisés pour répondre à des questions de recherche précises qui ne requièrent qu'une fraction des génomes disponibles. Les approches basées sur les réseaux d'échanges de gènes

s'appliquent donc généralement à des jeux de données limités par la taxonomie (McGowan & Fitzpatrick 2017), la niche écologique (Smillie *et al.* 2011), la nature du véhicule génétique (Hooper *et al.* 2009; Tamminen *et al.* 2012), la fonction des gènes (Fondi & Fani 2010) ou le nombre d'individus par niveau taxonomique (Dagan *et al.* 2008). Cette dernière stratégie, qui peut dans certains cas consister à étudier un individu par genre au sein d'une famille, présente le double avantage de limiter l'éventuelle surreprésentation de certains taxons très étudiés, mais limite également la prise en compte d'homologies transmises par descendance plutôt que par transfert horizontal.

Alors que la simple observation d'un réseau peut permettre une interprétation dans le cas de l'étude d'un petit nombre d'individus, cela devient plus problématique lorsque le jeu de données est plus important et complexe. Il est alors essentiel d'analyser la structure du réseau et des patrons de connexions entre individus à l'aide de descripteurs statistiques. Ceux-ci permettent, entre autres, de définir la topologie du réseau, de définir des sous-groupes plus connectés entre eux qu'ils ne le sont avec le reste du réseau (appelés modules ou communautés) ou encore de décrire de manière précise les caractéristiques de connexion de chacun des nœuds ou groupes de nœuds.

Il existe plusieurs topologies de réseaux, qui sont caractéristiques de leur fonctionnement. Dans un réseau aléatoire (*random* ou encore de Erdős–Rényi), tous les nœuds ont approximativement le même nombre de connections (Barabasi & Oltvai 2004), ce qui n'est a priori jamais le cas des réseaux biologiques (Barabási & Albert 1999). En effet, les réseaux biologiques (tels que ceux des voies de biosynthèse ou des interactions entre protéines) possèdent généralement un faible nombre de nœuds fortement connectés et un grand nombre de nœuds faiblement connectés : c'est la topologie dite *scale-free*. Toutefois, alors qu'un grand nombre d'études suggéraient que cette topologie était presque universelle au vivant, quel que soit le modèle ou l'échelle, des études récentes remettent en question la véracité de ce constat. Parfois, les nœuds qui présentent un grand nombre de connections sont également très connectés entre eux et s'organisent en clusters, créant ainsi un réseau avec une forte modularité : c'est la topologie hiérarchique. Les réseaux qui représentent un métabolisme cellulaire sont souvent organisés selon cette topologie (Ravasz *et al.* 2002). Enfin, que le réseau soit hiérarchique ou *scale-free*, il peut également être de type *small-world*. Dans un réseau *small-world*, deux nœuds pris au hasard ne sont séparés que par un petit nombre de nœuds. Cette caractéristique dépend de la présence de *hubs* (organisés en modules ou

non) très connectés et qui raccourcissent donc le chemin entre n'importe quelle paire de nœuds. Des grandes villes comme Paris ou New-York jouent par exemple ce rôle de *hub* dans un réseau représentant le transport aérien de passagers, dans lequel deux villes sont rarement séparées par plus de quelques aéroports.

Ensuite, dans un réseau, des groupes de nœuds (appelés communautés) sont souvent plus connectés entre eux qu'ils ne le sont avec les autres. Leur nombre et leur composition constituent des paramètres topologiques importants du réseau et participent à l'interprétation des données biologiques étudiées. La délimitation de ces sous-ensembles de nœuds aide également à la compréhension du fonctionnement global du réseau (Girvan & Newman 2002). La méthode de Louvain permet de détecter cette organisation en communautés. Elle est basée sur la notion de modularité du réseau, un indicateur mesurant à quel point un groupe de nœuds est plus connecté qu'avec une distribution aléatoire des liens entre eux. Cette méthode repose sur la répétition itérative de deux étapes : la première crée des groupes de nœuds en maximisant leur modularité, et la deuxième considère un réseau dont les nœuds sont les groupes précédemment créés. L'algorithme répète les deux étapes jusqu'à ce que l'ajout de groupes n'augmente plus la modularité du réseau (Blondel *et al.* 2008).

Dans certains réseaux, des groupes de nœuds sont encore plus fortement indépendants en n'étant aucunement reliés au reste du réseau. Ces groupes sont appelés composants connectés (CC) et définissent des groupes qui ne sont aucunement liés (Figure 6).

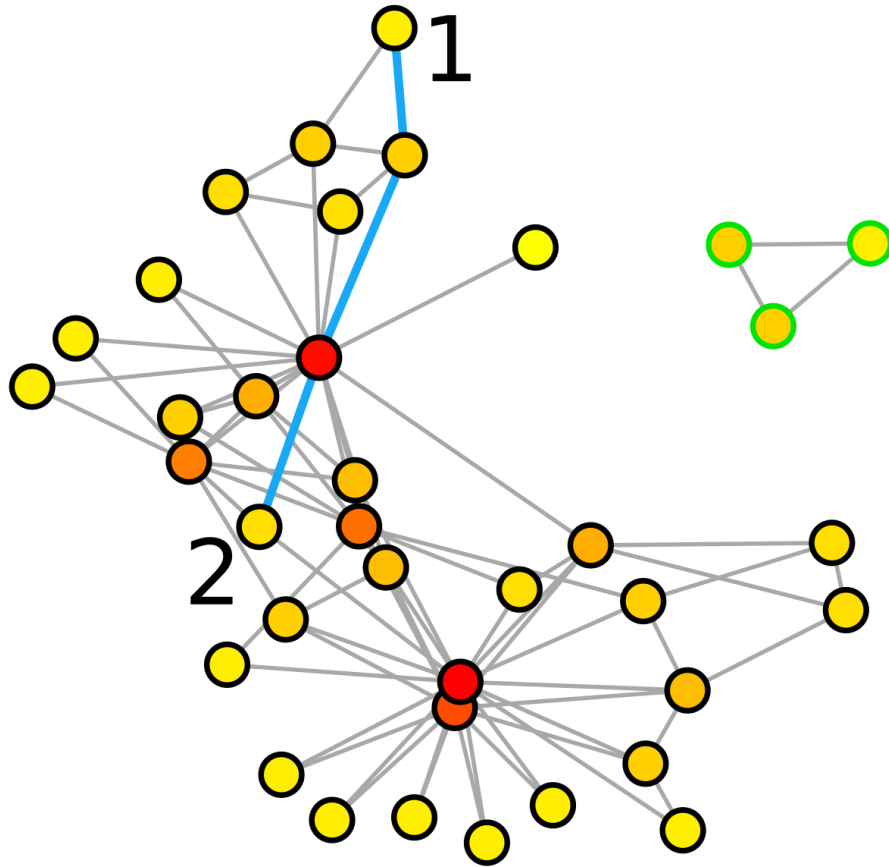


Figure 6 Représentation d'un réseau de similarité de gènes simulé comprenant 37 individus. Ce réseau comprend deux composants connectés : celui dont les nœuds sont cerclés de noirs et celui dont les nœuds sont cerclés de verts. La couleur de chaque nœud est fonction de son degré, c'est-à-dire le nombre de liens qu'il possède : de jaune pour un faible degré à rouge pour un fort degré. Les liens représentés en traits épais et bleus représentent le plus court chemin (diamètre) entre les nœuds « 1 » et « 2 ».

Enfin, l'analyse de la structure d'un réseau ou d'un sous-réseau peut se faire à l'aide d'indicateurs spécifiques liés à l'importance de chacun des nœuds dans le réseau (Valente *et al.* 2008). Dans le domaine des sciences sociales, un grand nombre d'indicateurs visant à caractériser les réseaux ont été développés au cours du temps (deux travaux majeurs étant Freeman (1977) et Newman (2005)). Ces indicateurs permettent de comparer de manière quantitative des nœuds appartenant à différentes catégories biologiques. Dans le cas d'un réseau d'échanges de gènes où les nœuds représentent des individus et les liens le partage de familles de gènes, la distribution des valeurs des indicateurs peut être mise en relation avec la classification taxonomique des génomes ou la fonction des gènes partagés.

Encart 1 Les indicateurs locaux

Degré simple d'un nœud : nombre de connections qu'il a avec d'autres nœuds. Des variations de cette définition existent, puisque le degré d'un nœud peut également être normalisé en prenant en compte la globalité du réseau, soit en fonction de la valeur maximale possible soit en fonction de la valeur maximale atteinte au sein du réseau étudié. L'étude de la distribution des degrés est informative puisqu'elle permet d'apprécier si la structure du réseau dépend de tous les nœuds de façon identique ou si un ou plusieurs nœuds ont une implication plus importante dans sa structuration (*i.e.* ont un degré plus élevé).

Proximité (*closeness*) : éloignement moyen du nœud considéré avec tous les autres nœuds. Plus un nœud est central plus sa valeur de proximité sera faible.

Intermédierité (*betweenness*) : proportion des plus courts diamètres entre toutes les paires de nœuds du réseau qui passent par le nœud considéré. Plus un nœud possède une haute valeur d'intermédierité, plus il est central.

Centralité de Bonacich : permet de prendre en compte, dans la valeur de centralité d'un nœud, les valeurs de centralités des nœuds auquel il est connecté. C'est sur ce principe que fonctionne le système page-rank, utilisé par les moteurs de recherche sur internet.

Les indicateurs les plus utilisés en analyse de réseau peuvent être classés en deux groupes : les indicateurs locaux et les indicateurs globaux (revus par (Ducruet 2010b) et (Ducruet 2010a), respectivement). Les indicateurs locaux (Encart 1), aussi appelés centralités, ne caractérisent qu'un nœud mais leur application à tous les nœuds d'un réseau permet de calculer des indicateurs globaux (Encart 2).

Encart 2 Les indicateurs globaux

Ordre : nombre de nœuds

Taille : nombre de liens

Densité du réseau : ratio entre le nombre de liens possibles et le nombre de liens existants.

Longueur caractéristique du chemin : moyenne des plus courts diamètres entre chaque paire de nœuds. Cette mesure permet de quantifier à quel point le réseau est de type *small-world*.

Diamètre : correspond au plus long des plus courts chemins, donne une indication sur la compacité du réseau.

II.6 Une diversité d'éléments génétiques mobiles

L'existence de bactéries présentant un pan-génome ouvert suggère qu'elles ont intégré de l'ADN étranger à leur propre matériel génétique. L'acquisition de gènes peut se faire par l'intermédiaire de plusieurs processus (Popa & Dagan 2011) (Figure 7). La pénétration d'ADN libre extracellulaire à l'intérieur d'une cellule bactérienne et son intégration au génome est appelée transformation. La transduction intervient par l'intermédiaire de bactériophages (qui sont des virus bactériens) ayant intégré dans leur génome des fragments de génome bactérien. Ces phages injectent alors parfois ces ADN aux nouvelles bactéries cibles en même temps que leur propre ADN. La conjugaison, quant à elle, correspond au passage direct d'un plasmide ou d'un autre élément génétique depuis une cellule donneuse à une cellule receveuse, généralement par le biais d'un pilus. D'autres mécanismes moins bien caractérisés existent : le relargage dans l'environnement de structures génétiques semblables à des phages appelées agents de transfert de séquences nucléotidiques (Popa & Dagan 2011), ou encore le transfert de protéines et d'ADN par le biais de structures intercellulaires telles que les nanotubes (Dubey & Ben-Yehuda 2011).

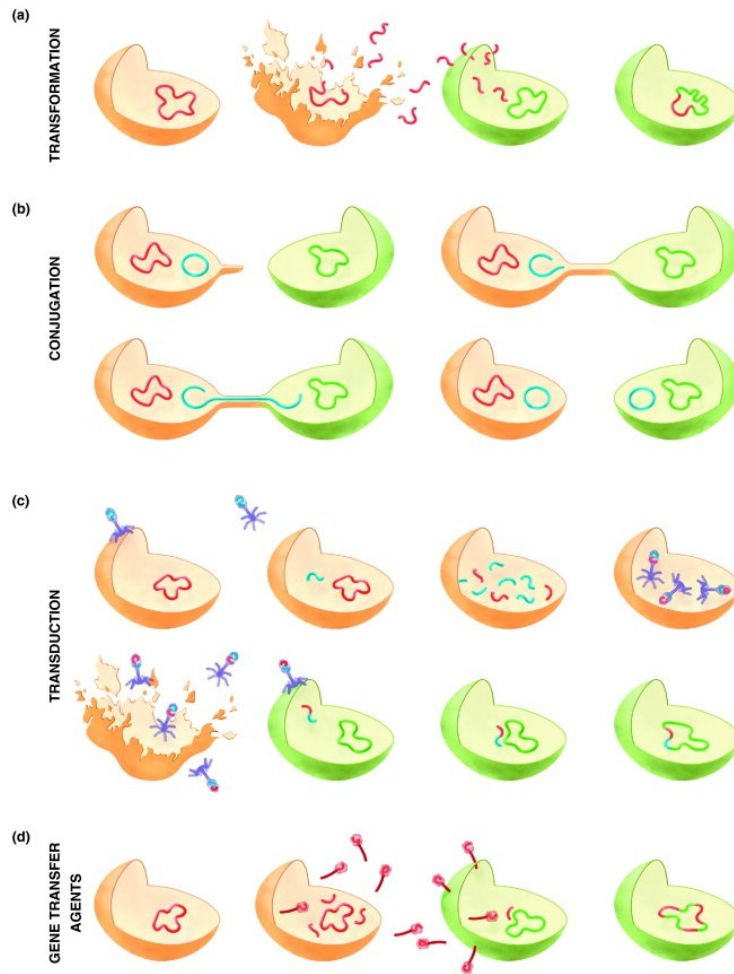


Figure 7 Schématisation des différents mécanismes pouvant aboutir à l'intégration d'ADN étranger au sein d'un génome bactérien. (a) transformation. L'entrée d'ADN libre dans le cytoplasme est possible lorsqu'une cellule bactérienne est dans un état de compétence¹. L'incorporation de la séquence d'ADN au sein du génome de la bactérie receveuse se fait par recombinaison. (b) conjugaison. Les plasmides sont des molécules d'ADN circulaires et extra-chromosomiques qui peuvent souvent être échangées par un mécanisme impliquant la réplication du plasmide et son transfert d'une cellule à une autre. (c) transduction. Les phages sont des virus qui infectent les bactéries. En s'incorporant au génome de la bactérie hôte, le génome du phage capture parfois des fragments du génome bactérien, qui peuvent être transmis à un hôte infecté par la suite. (d) les agents de transferts de gènes. Ces structures génétiques sont semblables aux phages et véhiculent de l'ADN d'un hôte à un autre mais n'incorporent pas leur matériel génétique dans le génome bactérien. Les agents de transfert de gènes ne semblent pas avoir de caractère pathogène. Figure tirée de Popa and Dagan (2011).

Les éléments génétiques mobiles (MGE pour *Mobile Genetic Element*) encodent des protéines qui assurent leur mobilité. Ils peuvent présenter une mobilité intracellulaire, c'est-à-dire d'un locus à

¹ La compétence d'une bactérie indique sa capacité à prélever une molécule d'ADN libre dans son environnement et à l'intégrer dans son propre génome par le phénomène de transformation.

un autre dans une même cellule, ou une mobilité intercellulaire et ainsi être impliqués dans les mécanismes de transfert horizontal présentés précédemment (Figure 8).

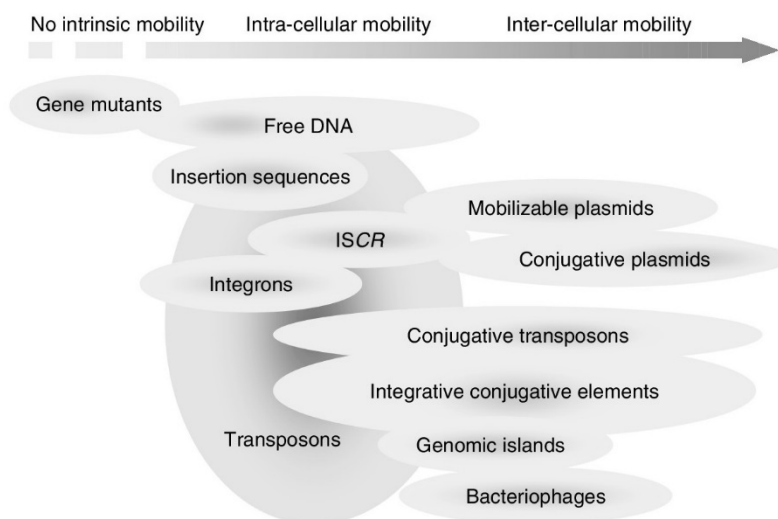


Figure 8 Représentation schématique du type de mobilité des éléments génétiques impliqués dans le transfert horizontal. Les liaisons physique ou fonctionnelle sont indiquées par la superposition des éléments. Certains éléments possèdent une mobilité intra et inter-cellulaire. Ainsi, leurs largeurs sur la figure sont des approximations de leur implication dans les deux types de mobilité. ISCR, de l'anglais *insertion sequences with common regions*: séquences insérées à régions communes. Figure tirée de Boerlin and Reid-Smith (2008)

II.6.a Les éléments à mobilité intracellulaire

Les séquences insertionnelles (IS), sont des MGE de très petite taille généralement composés uniquement du(es) gène(s) nécessaire(s) à leur transposition. Elles peuvent être présentes en quantité importante dans les génomes bactériens. Leur fréquence de transposition est variable : certaines IS nécessitent un évènement rare (le changement de cadre de lecture par le ribosome) pour la traduction de leur transposase, alors que d'autres semblent pouvoir transposer librement (Siguier *et al.* 2015). Le taux de transposition des IS dans une population bactérienne semble dépendre du stade évolutif de cette population. En effet, lorsqu'une bactérie s'est récemment spécialisée à un hôte, le taux de transposition des IS est particulièrement important (Siguier *et al.* 2014). Les sites d'insertion des IS peuvent être très spécifiques, seulement limités par des contraintes locales comme le taux de GC de la région nucléotidique réceptrice, ou préférentiellement situés dans les régions promotrices de gènes. Ainsi, les IS peuvent influencer le phénotype de la cellule en inactivant un ou plusieurs gènes ou en modifiant une région promotrice

(Casacuberta & González 2013). La présence d'IS dans un génome bactérien favorise également la recombinaison homologue qui peut donner lieu à des réarrangements chromosomiques, des délétions ou le passage d'ADN d'un compartiment génomique à un autre (Siguier *et al.* 2015). L'importance des IS comme acteurs de l'évolution a été démontrée de nombreuses fois et motive leur caractérisation approfondie. C'est pourquoi une classification en familles leur a été attribuée et que leurs mécanismes de transposition ont été étudiés. Malgré la diversité d'éléments de type IS, ces mécanismes sont étonnamment peu divers, et parfois très semblables à ceux retrouvés pour certains transposons. La distinction entre IS et transposon est historiquement basée sur la présence ou l'absence de gènes accessoires (Siguier *et al.* 2009). Cependant, il a été montré qu'une séquence d'ADN flanquée de deux IS peut subir une translocation d'un locus à un autre, à la manière des transposons. Certains transposons comme Tn5 ou Tn10 sont flanqués de deux copies (partielles ou complètes) de la même IS : au moins l'un d'eux comporte une transposase et le motif nucléotidique qu'elle reconnaît (répétition inversée : IR ou répétition directe : DR) tandis que l'autre peut ne comporter que l'IR ou la DR (Bennett 2004). Le transposon ainsi formé est appelé transposon composite. Les auteurs du domaine sont unanimes : la plasticité de ces deux familles d'éléments génétiques rend leur classification difficile. Les transposons qui n'appartiennent pas à la famille des transposons composites sont appelés transposons complexes, une famille très diverse d'éléments génétiques qui peuvent comporter une grande quantité de gènes accessoires (non-impliqués dans le processus de transposition). Les transposons Tn3, la multitude de Tn3-like ou encore les Tn21 en font partie et sont très étudiés car ils encodent souvent des gènes de résistance à différents antibiotiques (Bennett 2004). Le mode de transposition des transposons de la famille des Tn3 est particulier puisqu'il comprend une étape d'intégration complète du réplicon donneur au réplicon receveur (Nicolas *et al.* 2015). A cette étape dite de cointégration, deux copies du transposon sont présentes dans la molécule circulaire unique formée par la fusion du réplicon donneur et du réplicon receveur. L'étape finale de résolution sépare les deux réplicons à un locus de résolution bien précis. Le fonctionnement parfois imparfait de ce processus de transposition complexe est responsable de la formation de réplicons plasmidiques chimères (Gochez *et al.* 2018) mais peut aussi permettre la capture de gènes supplémentaires. Par exemple, l'existence de loci de résolution alternatifs peut causer une mauvaise réattribution de l'ADN propre à chaque réplicon lors de la résolution (Yano *et al.* 2013). La présence de motifs nucléotidiques qui ressemblent aux

IR (répétitions inversées¹) peut également redéfinir les bornes du transposon, incluant alors les gènes situés à une de leur extrémité (Naas *et al.* 2008). La recombinaison homologue est également souvent impliquée dans l'acquisition de nouveaux gènes par les Tn3. Ce processus est possible grâce à la présence de régions communes entre éléments génétiques, comme les sites de résolution ou les IR (Grindley 2002; Nicolas *et al.* 2015; Partridge 2011). Les IS, très répandus dans certains génomes bactériens, fournissent également une profusion de sites de recombinaison potentiels.

Des gènes accessoires sont également souvent associés à des loci où la recombinaison est très fréquente, constituant une entité génétique appelée intégron. Ces plateformes d'intégration de gènes sont composées d'une recombinase, appelée intégrase, d'un site de recombinaison et d'un promoteur qui pourra agir sur l'expression des gènes intégrés à la plateforme. Les gènes intégrés se présentent sous la forme de « cassettes », chacune composée d'un gène et d'un site de recombinaison qui sera reconnu par l'intégrase (Gillings 2014; Siguier *et al.* 2015). Elles présentent de façon transitoire une forme circulaire lors de leur copie d'un locus à un autre.

Les éléments à mobilité intracellulaire ne doivent généralement leur mobilité qu'à un petit nombre de gènes (*e.g.* transposase, recombinase ou encore résolvase) qui permettent leur déplacement au sein des compartiments génomiques d'une cellule par « copier-coller » ou « couper-coller ». Cependant, leur transmission intercellulaire doit être assistée par un élément génétique à mobilité intercellulaire.

II.6.b Les éléments à mobilité intercellulaire

Les éléments à mobilité intercellulaire comprennent les bactériophages, les agents de transfert de gènes et différents éléments conjugatifs. L'infection d'une bactérie par un phage peut conduire à sa lyse ou, le plus souvent, à un état de lysogénie. Ce type d'infection, non létal pour la bactérie, peut parfois conduire à des modifications du phage qui le rendent bénéfique pour la bactérie. Des séquences phagiques peuvent en effet être détournées en molécules bactériocines. Le phage peut également jouer le rôle de transporteur intercellulaire d'ADN (*i.e.* transduction). Par exemple, il peut être mobilisé par une région du génome bactérien impliquée dans la pathogénicité pour l'encapsidation et l'export extracellulaire de cette dernière (Canchaya *et al.* 2003). De manière plus

¹ Les IR (*Inverted Repeats*), répétitions inversées, sont des motifs nucléotidiques répétés de part et d'autres des transposons et la plupart du temps nécessaires au caractère mobile de ces derniers

générale, le transfert horizontal de gènes bactériens par l'intermédiaire de phages est lié à l'incorporation accidentelle d'ADN bactérien dans le génome du phage et à la transmission de celui-ci à une autre cellule-hôte.

Les agents de transfert de gènes sont des phages dégénérés, qui ont perdu leur capacité à transférer une copie fonctionnelle de leur génome à un nouvel hôte. La particule produite ressemble à celle d'un phage, mais l'ADN qu'elle encapsule est majoritairement originaire de la bactérie hôte. En outre, la particule n'est libérée dans le milieu extra-cellulaire que lors de la mort de la cellule-hôte (Lang *et al.* 2012).

Les éléments conjugatifs doivent généralement leur capacité de transfert d'une cellule à une autre à des pili ou à des canaux de transfert (Burrus *et al.* 2002). Ces éléments sont très divers et ont une grande variété de noms : les plasmides conjugatifs ou mobilisables, les éléments conjugatifs intégrés (ICE), ainsi que les transposons conjugatifs ou mobilisables. Les termes d'îlots génomiques, d'îlots de pathogénicité ou de résistance sont également rencontrés dans la littérature mais ne reposent aucunement sur la conservation d'une structure génomique ou leur apparemment.

Burrus *et al.* (2002) proposent d'appeler ICE tous les éléments génétiques capables d'excision par recombinaison site-spécifique, de transfert par conjugaison et d'intégration dans un génome naïf. Cette définition inclut donc les transposons conjugatifs et mobilisables, qui possèdent la faculté supplémentaire de se déplacer par transposition au sein d'un génome. Les ICE sont intégrés au génome de leur hôte et sont donc répliqués en même-temps que ce dernier. Les ICE transmissibles de façon autonome par conjugaison encodent la machinerie cellulaire nécessaire à leur excision du génome hôte, leur réplication, leur transfert par conjugaison et leur insertion dans un réplicon¹ d'une cellule receveuse. Un ICE est flanqué de deux motifs nucléotidiques particuliers. La recombinaison entre ces deux loci produit une forme circulaire de l'ICE où le site *attP* est créé au point de fermeture, tandis qu'un site *attB* est généré sur le réplicon duquel s'est excisé l'ICE. La réplication par cercle roulant (*rolling circle*) produit un brin linéaire complémentaire dont le transfert par conjugaison s'effectue à la manière des plasmides (voir II.7.c). Dans la cellule donneuse, la recombinaison entre *attB* et *attP* réintègre l'ICE à son locus d'origine. Dans la cellule

¹ Un réplicon est une molécule d'ADN ou d'ARN, ou une région d'ADN ou d'ARN qui peut se multiplier par réplication à partir d'une seule origine de réplication.

receveuse, l'ICE est circularisé et un deuxième brin est synthétisé pour former une molécule circulaire double brin. L'ICE est intégré au réplicon receveur au niveau d'un site *attB* (très conservé et présent dans un grand nombre de bactéries (McGrath & Pembroke 2004)), ou parfois à l'extrémité 3' d'un locus ARNt¹, selon le type et la spécificité de l'intégrase encodée par l'ICE (Wozniak & Waldor 2010). La plupart du temps, l'appareil de conjugaison des ICE est réprimé, mais des facteurs environnementaux ou intracellulaires peuvent l'activer et ainsi permettre leur mouvement intercellulaire (Burrus 2017). La grande majorité des ICE encodent, en plus des gènes de leur appareil conjugatif, des gènes accessoires qui sont souvent associés à l'adaptation et améliorent la survie de l'hôte dans certains environnements. Parmi les fonctions encodées par les gènes accessoires des ICE, on retrouve la résistance aux antibiotiques, aux métaux lourds, aux phages, des facteurs de virulence, d'utilisation de sources de carbone alternatives ou encore de fixation de l'azote atmosphérique (Burrus 2017).

Les plasmides sont des éléments génétiques très divers aussi bien en termes de taille, de mobilité, de phénotype ou de persistance dans les génomes. Ils sont à eux seuls l'objet de la section suivante II.7.

II.7 Le plasmide, véhicule pour les transferts horizontaux de gènes

II.7.a Généralités

Les plasmides sont des molécules d'ADN extra-chromosomiques. Ils sont généralement circulaires, de taille inférieure au chromosome, double brin et peuvent coder pour des protéines « accessoires », c'est-à-dire que ces protéines ne sont pas indispensables à la vie de la bactérie en conditions normales, mais peuvent procurer à la bactérie un avantage sélectif dans des environnements particuliers (Wiedenbeck & Cohan 2011). Certains taxons possèdent des facteurs de virulence indispensables encodés sur des plasmides, comme *Yersinia pseudotuberculosis* (Galyov *et al.* 1993), *Agrobacterium tumefaciens* (Kim *et al.* 2006) ou encore *Xanthomonas citri* pv. *citri* (Da Silva *et al.* 2002).

¹ Les ARNt, ou acides ribonucléiques de transfert, sont les intermédiaires entre les ARN messagers (qui sont des copies de portions d'ADN codant pour un ou plusieurs gènes) et les ribosomes auxquels ils apportent les acides aminés qui vont constituer les protéines.

Les plasmides sont des réplicons qui possèdent leur propre appareil de réplication. Burrus *et al.* (2002) ne considèrent d'ailleurs comme étant de véritables plasmides que les éléments génétiques dont la transmission verticale dépend de la réplication d'une molécule d'ADN circulaire. Lorsque l'appareil de réplication est identique entre deux plasmides, ceux-ci ne peuvent être hérités de façon stable dans une lignée : ils sont alors dits « incompatibles » (Couturier *et al.* 1988). Cette particularité a donné lieu à une classification basée sur le type d'appareil de réplication. Les plasmides sont transmis verticalement par descendance (Meacock & Cohen 1980), mais ils peuvent également être transmis horizontalement, partiellement ou totalement, par les quatre processus décrits précédemment (II.6) (Frischer *et al.* 1994). La conjugaison participe cependant beaucoup plus aux transferts plasmidiques que les trois autres processus, puisque bon nombre de plasmides codent pour tout ou partie d'un appareil de conjugaison. Ce dernier est composé d'un système de mobilité (MOB) et d'un système de sécrétion de type IV (T4SS). Les plasmides totalement dépourvus d'appareil de conjugaison (environ la moitié des plasmides connus) sont non-mobilisables, tandis que ceux qui n'encodent que pour le MOB sont mobilisables : ils peuvent être transférés par conjugaison en utilisant le T4SS d'un autre élément génétique de la cellule (Smillie *et al.* 2010) (Figure 9).

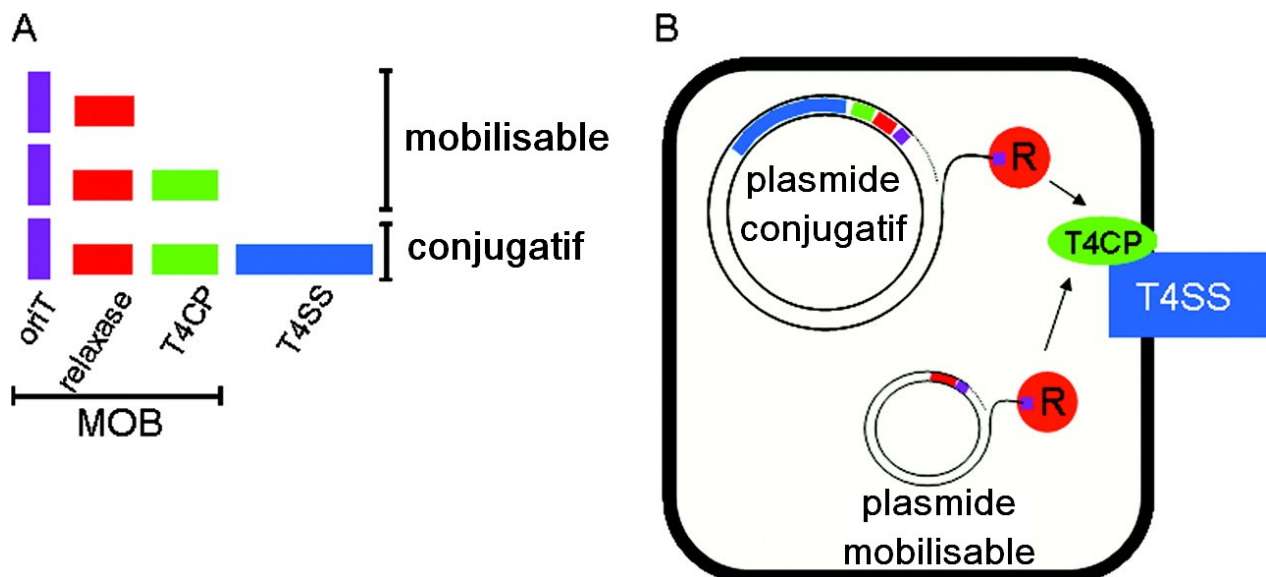


Figure 9 (A) Représentation schématique des principaux constituants génétiques impliqués dans la mobilité plasmidique. Les plasmides possédant un système MOB complet et un système de sécrétion de type IV (T4SS) sont conjugalifs tandis que ceux qui n'encodent pas de T4SS mais tout ou partie d'un système de mobilité (MOB) sont mobilisables. (B) Représentation des principales interactions entre constituants génétiques de l'appareil conjugal durant la conjugaison. Figure adaptée d'après Smillie *et al.* (2010).

II.7.b Groupes d'incompatibilité

La réplication d'un plasmide repose généralement sur trois éléments génétiques : l'origine de réplication *oriV*, un locus où est initiée la réplication ; une protéine impliquée dans l'initiation de la réplication (souvent la protéine Rep) ainsi que des gènes contrôlant la réplication. Il existe trois types de systèmes de réplication chez les plasmides : le type *theta*, le déplacement du brin (*strand displacement*), et le cercle roulant (*rolling circle*) (Del Solar *et al.* 1998). La détermination des groupes d'incompatibilité, basée sur le séquençage du locus *rep* par PCR (Carattoli *et al.* 2005; Couturier *et al.* 1988) ou sur des tests d'insertion dans une gamme de bactéries contenant déjà un plasmide de groupe connu, a été réalisée indépendamment pour plusieurs taxons comme *Enterobacteriaceae*, *Pseudomonas* ou *Staphylococcus* (Shintani *et al.* 2015). Ces auteurs ont entrepris la détermination du groupe d'incompatibilité *in silico* de la totalité des 4 602 plasmides dont les séquences étaient disponibles sur GenBank à l'époque. Moins de la moitié ont pu être classés dans les groupes d'incompatibilité connus (environ 200), certains étant classés dans plusieurs groupes à cause de la redondance des groupes d'incompatibilité entre différentes familles bactériennes ou de la présence de plusieurs réplicons sur certains plasmides (Shintani *et al.* 2015). Pour certains taxons ou pour les plasmides ayant plusieurs systèmes de réplication, cette classification possède donc des lacunes et il est courant de ne pas pouvoir classer un plasmide. Aucun plasmide isolé à partir de bactéries de l'ordre des Xanthomonadales n'a par exemple pu être classé dans les groupes d'incompatibilité connus (Shintani *et al.* 2015).

II.7.c Systèmes de mobilité

La conjugaison plasmidique repose sur deux systèmes. Le système T4SS, ou système de sécrétion de type IV, permet l'élaboration d'un canal de transfert nécessaire au transfert intercellulaire à proprement parler. Le système MOB ou système de mobilité prend en charge les mécanismes cellulaires liés au clivage du plasmide dans la cellule donneuse, au transport d'une forme simple brin vers le pilus du T4SS, et une fois le plasmide transféré, à la reconstitution d'une forme double brin circulaire. Les groupes phylogénétiques de ces deux systèmes ne sont pas corrélés car ils évoluent indépendamment, empêchant leur étude simultanée (Smillie *et al.* 2010). Le MOB et le T4SS pour le plasmide F de l'organisme modèle *E. coli* (présenté en Figure 10) sont les systèmes les mieux caractérisés à l'heure actuelle.

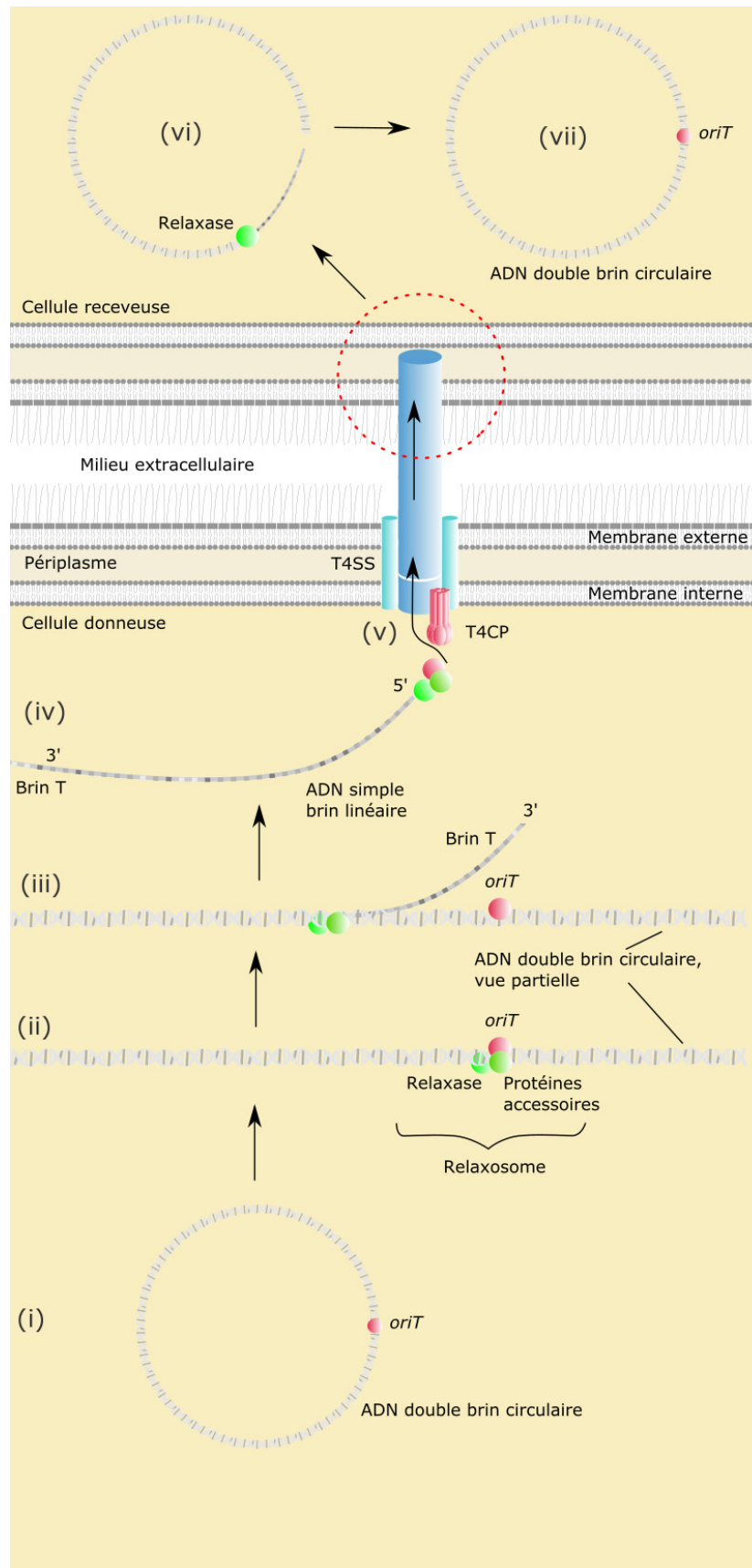


Figure 10 Mécanisme de transfert du plasmide modèle F par conjugaison via un système de sécrétion de type IV (T4SS) entre deux bactéries Gram-. (i) Le plasmide circulaire comprend un locus *oriT*, ainsi qu'une multitude de gènes *tra* codant pour des protéines impliquées dans la conjugaison. (ii) Le relaxosome se fixe à l'origine de transfert (*oriT*) et ouvre un des deux brins (appelé brin T pour brin transféré) du plasmide double brin. (iii) La relaxase (TraI) désenroule ensuite progressivement la double-hélice de l'ADN de l'extrémité 5' à l'extrémité 3' du brin T. Il est supposé que durant ce processus une copie du brin T est synthétisée par l'ADN polymérase III en utilisant comme ADN guide le brin de l'ADN-plasmidique non ouvert (cette synthèse n'a jamais été démontrée (Dostal *et al.* 2011)). (iv) Une fois entièrement synthétisé, le brin T est guidé vers le T4CP (Gomis-Ruth & Coll 2006; Llosa *et al.* 2002). (v) Le T4CP connecte le relaxosome au T4SS qui peut comprendre un pilus, dans lequel circule le complexe ADN-relaxase (tous deux liés de façon covalente) ainsi que des molécules TraI additionnelles jusqu'à la cellule receveuse (Dostal *et al.* 2011). (vi) Dans la cellule receveuse, la relaxase migre de l'extrémité 5' à l'extrémité 3' du brin T et la synthèse du brin complémentaire est initiée, selon un mécanisme mal connu (Parker & Meyer 2005). (vii) La relaxase réassemble ensuite les deux extrémités double brin pour former un plasmide double brin circulaire (Dostal *et al.* 2011). La zone entourée de pointillés rouges matérialise la méconnaissance de la nature du mécanisme permettant le passage du brin T à travers les membranes de la cellule receveuse et la limite de pénétration du pilus conjugal à travers celles-ci. Adapté de Alvarez-Martinez and Christie (2009).

Le système de mobilité (MOB) d'un plasmide est composé d'une origine de transfert *oriT*, d'une relaxase et éventuellement d'un système de couplage de type IV (T4CP). La relaxase, seule ou en association avec d'autres protéines, se lie avec le locus *oriT* et sépare l'ADN double brin en simple brin. Le complexe formé impliquant le plasmide sous forme simple brin et le plus souvent quelques protéines auxiliaires, est appelé relaxosome¹. Il est ensuite connecté au T4SS par l'intermédiaire du système de couplage T4CP en vue d'être transféré (De La Cruz *et al.* 2010). VirD4 joue le rôle de T4CP chez la majorité des bactéries tandis que ce rôle est assuré par TcpA chez certains groupes de Firmicutes (Guglielmini *et al.* 2014). Durant la phase finale de la conjugaison, localisée dans la cellule receveuse, la relaxase reconnaît les extrémités du plasmide linéaire transféré et le re-circularise (De La Cruz *et al.* 2010). Tous les plasmides mobilisables ou conjuguatifs possèdent une relaxase et très rarement plusieurs (Garcillan-Barcia *et al.* 2009), ce qui constitue donc un critère de classification intéressant. Une telle classification a ainsi été formalisée et comprend six groupes MOB, ainsi qu'une variété de sous-groupes (Garcillan-Barcia *et al.* 2011; Garcillan-Barcia *et al.* 2009; Smillie *et al.* 2010).

II.7.d Systèmes de sécrétion de type IV

Le système de sécrétion de type IV (T4SS) est l'ultime système nécessaire à la conjugaison. Cependant, tous les plasmides ne l'encodent pas puisque certains produisent leur propre relaxosome mais utilisent le T4SS porté par un autre plasmide. Ces plasmides sont qualifiés de

¹ Le relaxosome est un complexe nucléo-protéique composé du plasmide simple brin, de la relaxase et de protéines auxiliaires se fixant à l'ADN

mobilisables. Chez les bactéries Gram -, le T4SS est généralement composé de 10 à 20 gènes qui produisent des protéines nécessaires à l'élaboration d'un canal de transfert appelé pilus, qui va rejoindre les deux cellules bactériennes lors de la conjugaison. Le pilus est connecté au relaxosome par la T4CP, qui forme un pore dans la membrane bactérienne (De La Cruz *et al.* 2010). Les T4SS des bactéries Gram + sont moins bien connus, la présence d'un pilus faisant d'ailleurs débat (Goessweiner-Mohr *et al.* 2013; Goessweiner-Mohr *et al.* 2014b). Les différences de membrane plasmique et de paroi entre bactéries Gram + et Gram - expliquent la différence de structure de leur appareil de sécrétion T4SS (Goessweiner-Mohr *et al.* 2014a). Bien que le T4SS soit un constituant fondamental du processus de conjugaison bactérienne, ce système de sécrétion est également utilisé par certaines bactéries pathogènes pour injecter des facteurs de virulence dans les cellules de leurs hôtes (Elisabeth *et al.* 2018).

Le nombre important de gènes composant le T4SS, les fréquents remaniements, pertes/gains de gènes et la recombinaison entre différents systèmes compliquent la mise en place d'une classification des systèmes de sécrétion. De plus, la nomenclature des gènes impliqués n'est pas standardisée : des protéines de même nom ne sont pas forcément de proches homologues et deux protéines homologues n'ont pas forcément le même nom suivant le taxon qui l'encode. Selon Guglielmini *et al.* (2011), tous les T4SS encodent une protéine de la famille VirB4. Cependant, sa détection en dehors des protéobactéries est rendue délicate par le fait que certaines T4CP (*e.g.* VirD4) peuvent posséder une homologie semblable voire plus élevée avec VirB4 qu'avec les gènes du système T4SS (Smillie *et al.* 2010). Huit classes de T4SS correspondant à huit clades monophylétiques des protéines de la famille VirB4 ont toutefois été établies. Quatre classes sont presque exclusivement retrouvées chez les protéobactéries (appelés MPFF, MPFI, MPFG, MPFT dans Smillie *et al.* (2010)), leur contenu en gènes et leur synténie¹ étant relativement conservés au sein des classes. Les T4SS retrouvés en dehors des protéobactéries sont également répartis en quatre classes : MPFFATA, MPFFA, MPFB, MPFC (Guglielmini *et al.* 2013).

Si la conservation de certains éléments des systèmes de conjugaison plasmidique (*i.e.* relaxase et T4SS) permet dans une certaine mesure de classer et d'étudier les plasmides, il a toutefois été démontré que les plasmides d'un groupe T4SS ne présentent pas toujours le même type de relaxase

¹ La synténie correspond à l'ordre des gènes sur un brin d'ADN.

(Figure 11). Ajoutant à cela que tous les allèles de chaque système ne sont pas caractérisés, il est donc courant de ne pas pouvoir assigner un plasmide à un groupe connu.

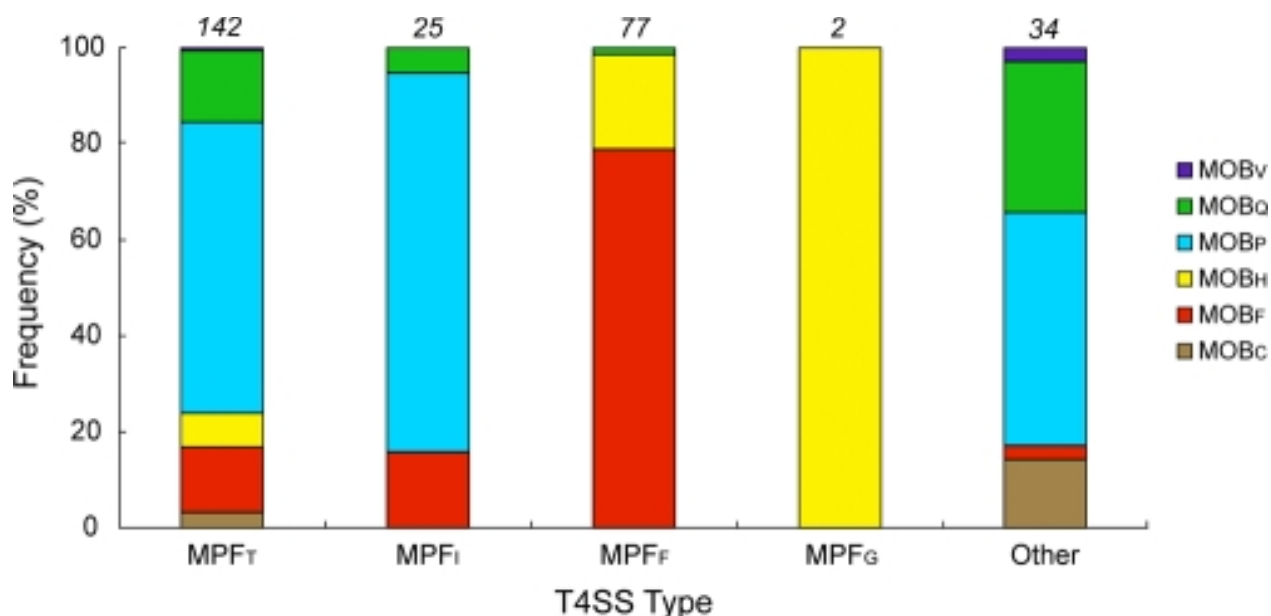


Figure 11 Distribution des familles de relaxases (classées en groupes MOB) en fonction du type de système de sécrétion de type IV (T4SS) encodé sur 280 plasmides. Tiré de Smillie *et al.* (2010).

II.7.e Un deuxième niveau de mobilité pour certains gènes plasmidiques

Les éléments à mobilité intercellulaire comportent souvent des MGE à mobilité intracellulaire (Liebert *et al.* 1999). En associant les échanges entre réplicons et les échanges entre cellules, l'imbrication de ces deux types d'éléments permet théoriquement l'échange entre cellules bactériennes de n'importe quelle séquence d'ADN.

Ne possédant pas de système de mobilité intercellulaire, les intégrons sont des éléments mobiles ne pouvant être transférés horizontalement seuls. Ils sont cependant fréquemment rencontrés sur des éléments à mobilité intercellulaires tels que les ICE ou les plasmides et sont alors dénommés intégrons mobiles. Souvent observés en association avec des systèmes de toxine-antitoxine (voir II.7.g), ils favorisent le maintien de leur réplicon-hôte ainsi que leur propre maintien dans la lignée bactérienne (Cambray *et al.* 2010). Les intégrons comportent une recombinase et peuvent intégrer des cassettes de gènes (voir II.6.a). Ces éléments présentent une structuration génétique plus corrélée à l'environnement de leurs hôtes qu'à la taxonomie de ces derniers (Gillings 2014). Cette structure suggère l'importance capitale des transferts horizontaux dans leur histoire évolutive.

Un intégron encodé sur un transposon Tn402 et portant des cassettes de gènes de résistance à l'amoxicilline, la streptomycine, les sulfonamides, la tétracycline et les ions mercure a été observé chez plusieurs genres de bactéries nosocomiales (Gillings 2014; Schluter *et al.* 2003). Ce type de transposon est très souvent rencontré sur des plasmides. Le succès évolutif de cette combinaison d'éléments génétiques mobiles a été renforcé par l'utilisation successive de plusieurs molécules antimicrobiennes, dont les gènes de résistance correspondants ont été progressivement insérés dans l'intégron sous forme de cassettes (Cambray *et al.* 2010). Une proportion importante des bactéries nosocomiales Gram- présentent aujourd'hui un intégron du même type (40% en moyenne et jusqu'à 76% pour *E. coli*) (Martinez-Freijo *et al.* 1998; van Essen-Zandbergen *et al.* 2007).

Par ailleurs, certains transposons transposent plus fréquemment sur un plasmide que sur un chromosome, pourtant de taille plus importante (Bennett 2004). La raison de ce paradoxe reste inconnue. L'association transposon - plasmide apparaît comme une voie importante pour la dissémination des gènes de résistance aux antibiotiques (Bennett 2004; Davies 1997; Liebert *et al.* 1999) voire de toxines (Okinaka *et al.* 1999).

II.7.f Les plasmides, véhicules pour les gènes adaptatifs

L'implication des plasmides dans l'adaptation bactérienne dépasse largement le cadre de la diffusion des gènes de résistance : les gènes accessoires portés par des plasmides peuvent coder pour une grande diversité de fonctions. Les plus courantes sont les résistances aux toxiques, mais d'autres voies métaboliques sont transmises par le biais des plasmides. C'est le cas du catabolisme, c'est-à-dire la dégradation d'un substrat pour produire de l'énergie; de la production de toxines; de fonctions génomiques modifiées, telle que l'augmentation du taux de mutation; ou encore de l'interaction avec des cellules eucaryotes, avec par exemple l'apparition de virulence ou l'induction de la production de nodules (détails et références en Tableau 1). Le phénotype apporté par la seule présence d'un plasmide peut être si différent du phénotype sans ce dernier qu'il conduit parfois l'hôte à changer drastiquement de niche écologique, comme dans le cas de l'acquisition de pathogénicité. En ce sens, les plasmides peuvent être considérés comme des déterminants majeurs de la spéciation (De la Cruz & Davies 2000).

De façon générale, les fonctions des gènes accessoires des plasmides augmentent la valeur sélective de la cellule hôte dans un environnement donné, contribuant à leur maintien au sein des populations.

Tableau 1 Exemples de phénotypes adaptatifs encodés sur des plasmides, classés en quatre groupes de fonctions. Les cas associés à des bactéries phytopathogènes sont soulignés, et ceux associés au genre *Xanthomonas* sont en gras. Les références suivies de † sont citées par Thomas (2003).

Résistance à des conditions environnementales défavorables <u>Cuivre</u> (Richard <i>et al.</i> 2017a) <u>Streptomycine</u> (Hyun <i>et al.</i> 2012; McManus <i>et al.</i> 2002) Bactériophage (Klaenhammer & Sanozky 1985) UV (Lodwick <i>et al.</i> 1990; Sundin <i>et al.</i> 2000) Températures supraoptimales (Bojer <i>et al.</i> 2010) Système de restriction-modification (Chopin <i>et al.</i> 1984)	Métabolisme Catabolisme (Chatterjee <i>et al.</i> 1981) Synthèse de polyketide (Mochizuki <i>et al.</i> 2003) Synthèse de caroténoïdes (Mochizuki <i>et al.</i> 2003) <u>Prototrophie en thiamine</u> (McGhee & Jones 2000)
Interactions avec des cellules eucaryotes <u>Effecteurs de type III</u> (Barash & Manulis-Sasson 2009; Brunings & Gabriel 2003) <u>Sécrétion de phytohormones</u> (Stes <i>et al.</i> 2011) <u>Induction de galles</u> (Platt <i>et al.</i> 2014) <u>Synthèse d'enzymes pectinolytiques</u> (Gonzalez <i>et al.</i> 1997) <u>Synthèse de coronatine</u> (Arnaud & Hwang 2015)	Production de composés toxiques <u>Antibiotique</u> (Kirby & Hopwood 1977) † (Kim <i>et al.</i> 2006) Bactériocine (Van Belkum <i>et al.</i> 1992) † Delta-endotoxine (González <i>et al.</i> 1982) † Enterotoxine (Scotland <i>et al.</i> 1983) † Neurotoxine (Zhou <i>et al.</i> 1995) †

II.7.g *Fréquence des plasmides au sein des populations*

Si l'avantage sélectif conféré par un plasmide de résistance à un toxique est évident sous pression de sélection (*i.e.* en présence du toxique), il apparaît aussi que ce même plasmide représentera un coût en l'absence de pression. Dans ce cas, le plasmide diminuerait peu à peu en fréquence au sein de la population (Harrison & Brockhurst 2012), à la fois par la sélection des clones non-porteurs et par la ségrégation aléatoire des plasmides qui, en fonction du nombre de copies du plasmide dans la cellule, conduit à la formation de cellules-filles non-porteuses (Nordstrom & Gerdes 2003).

Le coût de valeur sélective associé au squelette plasmidique est dépendant de la taille du plasmide et de son nombre de copies cellulaires, même si le plasmide ne porte, en plus de son squelette, que de l'ADN non codant (Glick 1995). La réplication de l'ADN plasmidique et les processus cellulaires impliqués dans sa réparation suffisent donc à imposer un coût mesurable. À cela s'ajoute le coût, plus important, représenté par la production des protéines plasmidiques accessoires (Rozkov *et al.* 2004) ou des protéines liées à la conjugaison. La formation de pili conjugatifs peut en outre amener une perte supplémentaire en valeur sélective, puisqu'un grand nombre de phages ciblent cette structure (Jalasvuori *et al.* 2011). *In vitro* et sans pression de sélection, le coût de valeur sélective associé à deux plasmides de multirésistance ciblés est de 6% et de 21% chez *E. coli* (Dahlberg & Chao 2003). Pourtant, les mêmes auteurs n'ont observé aucune lignée bactérienne sans plasmide après plus de 1000 générations.

On imagine donc qu'en cas de présence de la pression de sélection sur le long-terme, l'intégration chromosomique des gènes apportant un réel gain de valeur sélective devrait se substituer au plasmide entier (Harrison & Brockhurst 2012). Les gènes adaptatifs plasmidiques sont souvent encodés sur des éléments génétiques à mobilité intracellulaire (transposons, intégrons) (Bennett 2004; Harrison & Brockhurst 2012; Petrovski & Stanisich 2011). Ainsi, des transferts de gènes adaptatifs depuis un plasmide vers un chromosome (ainsi que la réciproque) sont possibles par l'intermédiaire d'éléments à mobilité intracellulaire (voir II.7.e) et ont déjà été observés (Bennett & Richmond 1976; Dobrindt *et al.* 2004; Ravatn *et al.* 1998). La bactérie présentant ce type d'intégration s'affranchirait en effet des coûts métaboliques liés au maintien du plasmide et à la production des protéines de son squelette.

Le fait de posséder des traits adaptatifs encodés sur des plasmides a nécessairement des avantages par rapport à des systèmes équivalents encodés sur un chromosome. Ces avantages peuvent se situer au niveau de la rapidité de l'adaptation que permet l'acquisition d'un plasmide, la mobilité de ces éléments génétiques ou encore la répartition du coût de valeur sélective au sein d'une population. En effet, pour une pression de sélection discontinue dans le temps, si seule une fraction de la population porte un plasmide, et donc subit le coût de valeur sélective associé durant les périodes sans pression de sélection, cette portion de la population peut jouer le rôle de réservoir de traits adaptatifs pour le reste lorsque la pression de sélection réapparaît (Heuer *et al.* 2008). Dans un système comprenant plusieurs pressions de sélection discontinues dans le temps et l'espace, les réponses adaptatives correspondantes peuvent être présentes dans une population bactérienne sans que tous les individus qui la composent n'aient à en supporter les coûts en valeur sélective (Eberhard 1990; Heuer *et al.* 2008; Slater *et al.* 2008).

En outre, les plasmides possèdent souvent des systèmes génétiques permettant d'assurer leur maintien au sein des populations. Trois mécanismes complémentaires et souvent présents de façon concomitante sur un même plasmide permettent ce maintien.

Le premier de ces systèmes permet d'éviter une ségrégation aléatoire des plasmides durant la division de leur cellule-hôte. Ces systèmes de partition agissent durant la réplication du plasmide, en distribuant spatialement les copies au sein de la cellule avant la division (Nordstrom & Gerdes 2003).

Deuxièmement, la réplication de plusieurs plasmides identiques au sein d'une cellule donne parfois naissance à des dimères ou des multimères plasmidiques. Or ce phénomène diminue le nombre de plasmides dans la cellule et augmente ainsi la probabilité d'apparition d'une cellule-fille non porteuse du plasmide, particulièrement dans le cas de plasmides à faible nombre de copies. Avant la division de la cellule, un système de recombinaison site-spécifique permet alors de séparer les éventuels multimères en monomères (Sengupta & Austin 2011), ce qui favorise encore une fois le maintien du plasmide dans les cellules-filles.

Troisièmement, une grande diversité de systèmes dits de « toxine-antitoxine » sont encodés par des plasmides et limitent également leur perte. Généralement, ils impliquent la production intracellulaire d'une toxine relativement stable dans le temps, et d'une antitoxine qui est

naturellement dégradée de façon plus rapide. Lors de la division cellulaire d'une bactérie, les cellules-filles n'ayant pas reçu le plasmide seront, une fois l'anti-toxine dégradée, soumise aux effets de la toxine présente dans le cytoplasme hérité de la cellule-mère (Van Melderer & Saavedra De Bast 2009). Ces systèmes égoïstes assurent le transfert vertical des parasites intracellulaires que sont les plasmides. L'effet « addictif » procuré par le système toxine-antitoxine est parfois contrecarré par l'intégration au chromosome du système toxine-antitoxine (Van Melderer & Saavedra De Bast 2009).

Plusieurs auteurs rapportent la conservation de plasmides non porteurs de systèmes d'addiction (toxine-antitoxine) et parfois non mobiles dans certaines lignées génétiques même en l'absence de pression de sélection maintenant le plasmide (Modi & Adams 1991; Modi *et al.* 1991). Cette persistance va souvent de pair avec une diminution du coût de valeur sélective imposée par le plasmide sur son hôte. Cette amélioration est attribuée à un phénomène de co-évolution (Harrison & Brockhurst 2012). La co-évolution est la conséquence de mutations compensatoires (incluant mutations ponctuelles, délétions et toute autre modification de l'ADN), impliquées dans l'adaptation du plasmide à son hôte et de l'hôte à son plasmide. Au cours des générations, l'association hôte/plasmide devient ainsi plus performante. Le coût de valeur sélective lié à la présence du plasmide peut alors chuter dramatiquement, jusqu'à la limite de détection au laboratoire, expliquant le fait que certaines études mentionnent l'absence de coût pour certains plasmides évolués (Dahlberg & Chao 2003). La coévolution peut avoir un niveau de spécificité variable, selon que les mutations compensatoires de l'une des deux parties n'améliorent la valeur sélective de l'ensemble qu'avec l'entité avec laquelle il a co-évolué ou avec une plus grande diversité d'éléments (Harrison & Brockhurst 2012). Les modifications génétiques associées aux mutations compensatoires comprennent l'absence ou la réduction du taux de conjugaison, la perte de gènes ou la diminution de leur expression, ou encore la diminution du nombre de copies du plasmide par cellule (Harrison & Brockhurst 2012).

La conservation d'un plasmide au sein d'une population bactérienne apparaît donc comme un équilibre entre des phénomènes stochastiques, l'effet de systèmes intrinsèques au plasmide promouvant sa conservation et l'impact de la pression de sélection sur le coût de valeur sélective du plasmide.

III Substances antimicrobiennes en agriculture

III.1 Généralités

Les plantes cultivées sont la cible de nombreux microorganismes : principalement mycètes, bactéries et virus. Les techniques mises en œuvre pour lutter contre ces agents pathogènes sont très variées en fonction du type d'organisme, de la nature de son éventuel vecteur et d'une multitude d'autres paramètres. Toutefois, dans la majorité des cas, la lutte comprend une composante chimique. Plus précisément, les antibactériens utilisés en agriculture comprennent des métaux lourds et des antibiotiques.

Si historiquement, les antibiotiques sont des substances produites par des microorganismes et qui peuvent inhiber ou tuer d'autres microorganismes (Madigan *et al.* 1997), ce terme inclut aujourd'hui également des substances produites chimiquement. L'utilisation de ces molécules en agriculture s'est développée de façon importante à partir des années 1950 pour traiter, entre autres, le feu bactérien des pommiers et des poiriers (McManus *et al.* 2002). La faible phytotoxicité des antibiotiques explique alors l'engouement du monde agricole pour ces substances dans le cadre de la mise en place des programmes de protection phytosanitaire. Des traitements utilisant comme molécule active l'oxytetracycline, la streptomycine, la gentamycine ou l'acide oxolinique ont depuis vu le jour (Mellon *et al.* 2001). Cependant, le nombre de molécules antibiotiques utilisées en agriculture est limité par rapport aux domaines de la santé humaine ou animale. Cette faible prévalence des traitements antibiotiques en agriculture a plusieurs origines. Tout d'abord, le prix de ces molécules est souvent prohibitif, n'autorisant leur utilisation que pour des cultures à très forte valeur ajoutée. Ensuite, en cas de maladie endémique, des méthodes alternatives de lutte sont plutôt recherchées, afin de limiter les coûts pour les agriculteurs, leur dépendance vis-à-vis des produits antibiotiques et afin d'éviter l'acquisition de résistances par le pathogène ciblé. Enfin, la meilleure connaissance des mécanismes de résistance des bactéries pathogènes et la découverte d'homologies entre les déterminants génétiques de la résistance chez des bactéries associées à l'humain, au sol, aux plantes et aux animaux a soulevé des hypothèses alarmantes quant à la capacité de dispersion des résistances (McManus *et al.* 2002). Motivés par la crise de l'antibiorésistance en médecine humaine, un grand nombre de nations ont choisi de préserver l'activité antibiotique des molécules connues pour un usage en santé humaine en mettant en place des restrictions (voire interdictions totales) concernant les molécules autorisées ainsi que sur le

nombre de traitements annuels en agriculture et en élevage (Casewell *et al.* 2003; Lipsitch *et al.* 2002; Stockwell & Duffy 2012; Witte 1998).

III.2 Les métaux lourds

III.2.a *Utilisation en agriculture*

Des sels de métaux lourds peuvent constituer le principe actif d'antibactériens. En agriculture, les métaux lourds les plus utilisés sont le cuivre, le fer, le mercure, l'arsenic et l'antimoine, l'usage de ces trois derniers étant en forte décroissance (Hobman & Crossman 2015). Par exemple, les produits organomercuriques, longtemps utilisés pour traiter les semences de céréales, sont peu à peu abandonnés à cause de leur haute toxicité (Leroux & Gardan 2002). Les sels de cuivre sont quant à eux abondamment utilisés contre une large gamme de bactéries phytopathogènes affectant le système aérien des plantes (Leroux & Gardan 2002). Les composés cupriques, sous forme de bouillie bordelaise, d'hydroxyde de cuivre, d'oxyde cuivreux ou d'oxychlorure de cuivre, sont de loin les plus utilisés pour traiter les graines ou les plantes contre des champignons, comme le mildiou, ou pour lutter contre les phyto bactérioses, et ce depuis les années 1920 (Leroux & Gardan 2002; Smith 1985). L'addition d'un dithiocarbamate, comme le mancozèbe, ou de chlorure de fer au produit phytosanitaire permet une amélioration relative de l'efficacité du cuivre, notamment vis-à-vis de souches bactériennes résistantes (Canteros *et al.* 2017; Lee *et al.* 1993; Scheck & Pscheidt 1998).

III.2.b *Toxicité des métaux lourds*

Bien que beaucoup de métaux lourds soient des molécules nécessaires aux microorganismes car incorporés à des enzymes ou des cofacteurs, plusieurs mécanismes les rendent toxiques à plus forte concentration (Sharma & Dietz 2009). Cinq principaux mécanismes de toxicité sont définis dans la revue de Lemire *et al.* (2013).

Tout d'abord, les métaux lourds affecteraient les membranes des bactéries en modifiant la sélectivité de leur imperméabilité. La chaîne respiratoire peut par exemple être bloquée, faute de pouvoir générer un potentiel transmembranaire par l'expulsion de Na⁺ (Bragg & Rainnie 1974). Certains métaux lourds pourraient également endommager les chaînes lipidiques constitutives des

membranes bactériennes (Hong *et al.* 2012). La toxicité membranaire des métaux lourds, difficile à dissocier des effets d'autres dommages intracellulaires, est encore peu caractérisée.

Les métaux lourds interfèrent également avec l'absorption de certains nutriments, par des mécanismes qui ne sont pas entièrement compris (Pereira *et al.* 2008). L'effet toxique comprend vraisemblablement la dérégulation de l'expression des gènes gouvernant l'absorption ainsi que l'inhibition compétitive qu'exerce le métal lourd avec le nutriment pour l'entrée dans la cellule.

Dans le cas particulier du fer, la réaction de Fenton (*i.e.* une réaction d'oxydo-réduction) provoque une dégradation de l'ADN (Keyer & Imlay 1996). Cette génotoxicité n'a étonnamment pas été observée avec d'autres métaux lourds, même s'ils sont aussi impliqués dans la réaction de Fenton.

La majorité des métaux lourds cause également des dommages oxydatifs aux cellules. Les métaux dits « redox-inactifs », comme le cobalt ou le nickel, se fixent à certains antioxydants et les inactivent, augmentant ainsi la concentration intracellulaire en espèces réactives oxygénées (ROS). Les métaux « redox-actifs » catalysent quant à eux de manière directe des réactions d'oxydo-réduction et donc la formation de ROS. La toxicité des ROS, tel que le peroxyde d'hydrogène, n'est pas entièrement comprise mais elle inclurait la peroxydation des lipides, l'oxydation des protéines et la dégradation des acides nucléiques (Dupont *et al.* 2011).

Enfin, les métaux lourds peuvent bloquer des voies métaboliques de manière directe, comme lorsqu'ils remplacent un métal structural ou catalytique au sein d'une enzyme, ou de manière indirecte par la production de ROS qui peuvent oxyder et inactiver certains sites d'enzymes (Anjem & Imlay 2012). La substitution d'un atome de Zinc par un atome de Nickel au sein d'une enzyme produite par *E. coli* provoque par exemple un blocage de la glycolyse (Macomber *et al.* 2011).

III.2.c Les lignes de défense des bactéries contre les métaux lourds

Une variété de mécanismes protège les bactéries des molécules toxiques. Zhou *et al.* (2015) proposent une séparation de ces mécanismes en trois groupes, qu'ils appellent « lignes de défense ».

La première correspond à l'organisation des bactéries en biofilms, dans lesquels les cellules adhèrent les unes aux autres et sont protégées dans une matrice d'exopolysaccharides (EPS) qu'elles sécrètent.

Cette organisation en matrice protège les bactéries des substances antimicrobiennes par minimisation du contact physique avec la substance (Zhou *et al.* 2015). À l'interface entre le milieu extérieur et le biofilm, certaines molécules toxiques ne peuvent pas y pénétrer du fait de leur taille ou d'interactions électrostatiques ou hydrophobiques, peuvent être adsorbées au biofilm, voire être dégradées, alors que d'autres toxiques arrivent à se diffuser dans la matrice d'EPS (Zhou *et al.* 2015).

L'environnement que constitue le biofilm présente une variabilité spatiale. En effet, les couches les plus profondes de la structure présentent une concentration en oxygène et nutriments plus faible. Les bactéries des différentes couches du biofilm présentent donc des niveaux d'activité métabolique variables, maximisant la probabilité qu'une fraction d'entre elles se trouve dans un état de moindre sensibilité aux molécules antibiotiques (Costerton *et al.* 1999).

Au sein du biofilm, certaines cellules bactériennes présentent un état cellulaire dit de persistance, à la manière des spores. Un traitement antibiotique prolongé peut tuer une grande partie des bactéries d'un biofilm, mais les cellules persistantes y survivent parfois et peuvent ainsi reconstituer la population bactérienne lorsque les conditions sont redevenues propices (Qu *et al.* 2010).

Lorsqu'elles sont soumises à des conditions environnementales contraignantes, les bactéries activent des mécanismes de réponse au stress. Ces derniers sont très variés, et comprennent la variation du niveau d'expression de certains gènes ou le ralentissement du métabolisme (Poole (2012). Les mécanismes de réponse au stress ne sont pas propres aux biofilms, mais cette organisation présente l'avantage de retarder la pénétration du toxique, laissant aux bactéries plus de temps pour mettre en place les mécanismes de réponse au stress.

La seconde ligne de défense comprend la paroi cellulaire, la membrane cellulaire ainsi que les pompes d'efflux qui y sont insérées. Une plus forte imperméabilité, un changement dans la conformation de ces barrières, l'action de pompes qui évacuent le toxique vers le milieu extérieur ou la séquestration extracellulaire ou périplasmique peuvent limiter la concentration intracellulaire en antimicrobiens.

La troisième ligne de défense entre en jeu lorsque le toxique est présent dans le milieu intracellulaire. Pour échapper à l'action du métal, la bactérie peut modifier les voies métaboliques

ciblées par celui-ci, réparer les molécules endommagées, synthétiser des protéines de protection, ou encore dégrader, précipiter ou séquestrer les molécules antimicrobiennes (Lemire *et al.* 2013). Ces fonctions peuvent être natives, apparaître par mutation du matériel génétique existant ou être acquises par transfert horizontal de gènes (Zhou *et al.* 2015).

De manière générale, les mécanismes bactériens de résistances aux métaux lourds sont les mêmes que les mécanismes de résistances aux antibiotiques. Certains systèmes génétiques encodant des pompes d'efflux confèrent en outre un phénotype résistant pour les deux types d'antimicrobiens (Staehlin *et al.* 2016). Une forte corrélation entre la présence de métaux lourds et la persistance de déterminants génétiques de résistance à des antibiotiques a même été mise en évidence (Ji *et al.* 2012; Poole 2017; Seiler & Berendonk 2012; Wales & Davies 2015). Une hypothèse expliquant cette corrélation est l'utilisation concomitante d'antibiotiques et de métaux lourds pour lutter contre certaines bactérioses en agriculture ou dans les promoteurs de croissance utilisés en élevage. Si les molécules antibiotiques se dégradent rapidement, la persistance des métaux lourds maintient une pression de sélection sur le ou les supports génétiques des différentes résistances (Ji *et al.* 2012; Ogiyama *et al.* 2005). La durée de persistance dans les sols agricoles du cuivre, principal métal lourd utilisé en agriculture, n'est pas connue. Cependant, une seule application de cuivre sur un sol modifie ses propriétés biologiques et physico-chimiques pour au moins 5 ans (Wakelin *et al.* 2010). Cette rémanence pose de nombreux problèmes environnementaux et socio-économiques (de Boer *et al.* 2012; Thrupp 1991; Van-Zwieten *et al.* 2004; Wightwick *et al.* 2010).

III.3 Le cuivre

Dans des conditions physiologiques, le cuivre existe dans deux états d'oxydation : le Cu^+ (CuI) et le Cu^{2+} (CuII) (Stevenson *et al.* 2013). Le cuivre est un métal dit « redox-actif » qui fait partie des métaux lourds les plus toxiques pour les bactéries. En milieu aérobie, il catalyse de façon directe la formation des ROS par les réactions de Fenton et Haber-Weiss (Nies 1999). Chez *E. coli*, le cuivre, contrairement au fer, n'induit pas de dégâts à l'ADN et la production de ROS est très majoritairement détectée dans le périplasme (Macomber *et al.* 2007). En contact avec une solution cuprique, la concentration en cuivre intracellulaire de cellules d'*E. coli* augmente, et la mortalité progresse à mesure que les membranes cellulaires se dégradent (Santo *et al.* 2011). L'action oxydative du cuivre se fait également de manière indirecte par fixation à certains antioxydants cellulaires, et par une cascade de réactions chimiques débutant par l'oxydation de composés

comportant du fer et générant ultimement des ROS par réaction de Fenton et Haber-Weiss (Dupont *et al.* 2011).

Les propriétés chimiques du cuivre lui confèrent la capacité de former des complexes extrêmement stables avec un certain nombre de ligands. Le cuivre peut en effet se substituer à d'autres métaux lourds indispensables en tant que cofacteurs à de nombreuses métalloprotéines, dont la fonction sera altérée (Stevenson *et al.* 2013). La forte capacité du cuivre à se lier à des enzymes et à former des complexes stables avec elles est vue comme un élément essentiel de sa toxicité. Par exemple chez *E. coli*, le cuivre est capable de se substituer au fer dans une classe de protéines appelées protéines fer-soufre, ce qui empêche la synthèse de certains acides aminés et inactive les enzymes de type déshydratases (Macomber & Imlay 2009). D'autres mécanismes sont probablement à l'origine de la toxicité du cuivre, puisque la supplémentation des acides aminés manquants à *E. coli* en milieu cuprique ne suffit pas à sa survie (Macomber & Imlay 2009).

III.3.a Homéostasie du cuivre

Bien que toxique à forte concentration, le cuivre est nécessaire à la croissance bactérienne. Il sert en effet de cofacteur pour des enzymes importantes pour le métabolisme bactérien telle que la cytochrome c oxydase, une enzyme terminant la chaîne respiratoire, ou la superoxyde dismutase, composante essentielle du mécanisme d'élimination des radicaux libres (Stevenson *et al.* 2013). Le contrôle de la localisation et de la concentration du cuivre intracellulaire est alors indispensable. Pour prévenir l'implication du cuivre dans des réactions chimiques néfastes pour la cellule, des enzymes appelées métallochaperones le séquestrent et le transportent jusqu'aux enzymes qui en ont besoin (Stevenson *et al.* 2013). La régulation fine de la teneur en cuivre intracellulaire est quant à elle possible grâce à des systèmes de pompe régulant l'homéostasie. Les gènes codant pour des systèmes de résistance et ceux codant pour des systèmes d'homéostasie sont très similaires. De plus, certains systèmes de résistance et d'homéostasie peuvent être des paralogues : ils descendent d'un même gène ancestral dont les fonctions ont été modifiées au cours du temps (Staehlin *et al.* 2016). Ces deux constats créent une certaine confusion et il n'est pas toujours évident d'attribuer un gène à l'un ou l'autre de ces deux mécanismes.

Les systèmes d'homéostasie les mieux caractérisés sont ceux présent chez l'organisme modèle *E. coli*, qui comporte trois systèmes chromosomiques impliqués dans le contrôle de la concentration intracellulaire en cuivre (Figure 12). Chez *E. coli*, l'oxydase périplasmique CueO oxyde le CuI en CuII, moins toxique, en utilisant l'oxygène comme donneur d'électron, réaction qui n'est donc possible qu'en milieu aérobie (Grass & Rensing 2001; Singh *et al.* 2004). La régulation de la transcription de *cueO* est effectuée par *cueR* (Outten *et al.* 2000). Un deuxième système comprend CopA qui est une pompe de transport de cations de la famille des transporteurs de métaux mous qui pompe le CuI du cytoplasme au périplasma (Rensing *et al.* 2000; Rensing & Grass 2003). Tout comme *cueO*, *copA* est sous le contrôle du régulateur de transcription CueR et dépend de la présence de CuI. Le CuI périplasmique est ensuite transporté en dehors de la cellule par le système Cus. Un opéron de quatre gènes *cusCFBA* encode pour des protéines requises pour l'efflux extracellulaire de cuivre tandis que les deux gènes de l'opéron *cusRS* encodent pour l'histidine kinase CusS, et le régulateur CusR (Franke *et al.* 2003). Ces deux dernières protéines forment un système de senseur/régulateur qui active les gènes *cusCFBA*, dont la transcription est dépendante de la concentration en cuivre (Franke *et al.* 2003). Trois protéines Cus forment un canal à travers le périplasma : CusA est une pompe de la membrane interne, CusB est une protéine périplasmique et CusC est une protéine de la membrane externe (Hernandez- Montes *et al.* 2012) (Figure 12). CusF transporte les ions CuI du périplasma jusqu'au canal de transfert.

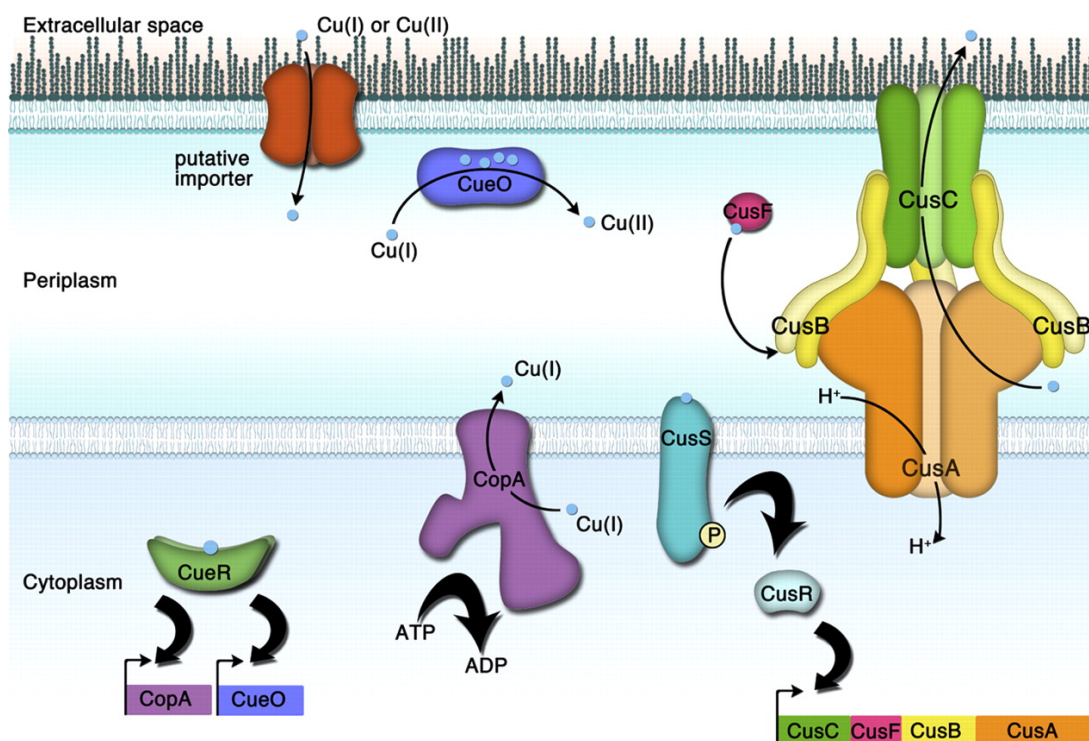


Figure 12 Mécanismes de l'homéostasie du cuivre chez *E. coli*. Le mode de pénétration du cuivre dans la cellule n'est pas connu. Le CuI du cytoplasme est transporté dans le périplasma par CopA, actif en aérobiose. Dans le périplasma, le CuI est oxydé par CueO en CuII, moins toxique. L'association CopA/CueO permet donc de stocker le cuivre dans le périplasma sous sa forme la moins toxique pour la cellule. CueR active la transcription de CopA et de CueO en présence de cuivre. En anaérobiose (et en aérobiose lorsque la concentration intracellulaire en cuivre est forte (Outten *et al.* 2001)), le détecteur de CuI périplasmique CusS ajoute un groupement phosphate à CusR qui active la transcription de *cusCFBA*. Ce dernier excrète le CuI périplasmique vers le milieu extérieur. Figure tirée de Kim *et al.* (2011).

Le système *cue* est activé à des concentrations en cuivre bien en deçà de celles nécessaires à l'expression de *cus*. Par ailleurs, les niveaux d'expression de chaque système après induction cuprique sont fortement différents (12x pour *cue* et 800x pour *cus*) (Outten *et al.* 2001). L'hypothèse selon laquelle *cue* est le système de régulation primaire et que *cus* n'est activé que si la concentration en cuivre intracellulaire est trop élevée pour que *cue* le gère efficacement a donc été émise. Il a été estimé chez *E. coli* que le système *cue* gère en conditions aérobies des concentrations en cuivre intracellulaire allant jusqu'à environ 50µM en stockant le CuII dans le périplasma. Lorsque les concentrations en cuivre intracellulaire sont supérieures, ou en conditions anaérobies, le système *cus* devient l'acteur majeur en permettant l'efflux extracellulaire du cuivre (Outten *et al.* 2001).

L'attribution d'une fonction d'homéostasie ou de résistance aux différents homologues du système *cus* n'est pas aisée. En effet, le support génomique de *cus* est variable, le phénotype de la souche

porteuse peut être résistant (CuR) ou sensible au cuivre (CuS), et il est parfois associé à d'autres systèmes. Il est en effet présent sur le chromosome de souches de *E. coli* qui présentent un phénotype sensible au cuivre (la souche K12, nommée ED8739 dans Rouch *et al.* (1985) et Tetaz and Luke (1983)). Le système *cus* a également été retrouvé sur un plasmide qui confère un phénotype résistant à une concentration en cuivre trois fois supérieure (Rouch *et al.* 1985), mais qui encode également le système Pco (voir III.3.b) (Staehlin *et al.* 2016) dont l'implication dans la résistance au cuivre est avérée (Tetaz & Luke 1983). Il est donc difficile d'attribuer l'augmentation de la résistance au seul système Pco ou à son association avec la version plasmidique du système Cus, qui peut avoir changé de fonction (paralogie). La variation de phénotype entre souches *cus*⁺ et souches *cus*⁻ obtenues par mutagenèse dirigée ou entre souches *cus*⁻ et souches *cus*⁺ obtenues par complémentation pourrait permettre de conclure sur le rôle de du système Cus. La conservation de la séquence d'acides aminés de CusCFBA chromosomiques (*E. coli*, accession GenBank U00096) et de leur équivalent plasmidique (*E. coli* étudié par Staehlin *et al.* (2016), accession GenBank KC146966) est de 72%, 47%, 67% et 87% sur >95% de leur longueur, respectivement, ce qui autorise la suspicion d'une modification de fonction (Figure 13).

Chez *Xanthomonas*, les gènes chromosomiques *cohL*, *cohA* et *cohB* régulent l'homéostasie du cuivre (Behlau *et al.* 2011). À cause de l'homologie entre ces gènes et ceux d'un cluster de gènes plasmidiques conférant un phénotype résistant au cuivre (voir III.3.b et Figure 13), ces gènes ont été appelés *copLAB* par plusieurs auteurs de façon erronée, décrivant les souches encodant ce système chromosomique comme résistantes au cuivre (Hsiao *et al.* 2011; Palmieri *et al.* 2010; Teixeira *et al.* 2008). Behlau *et al.* (2011) ont donc renommé les gènes chromosomiques, portés aussi bien par les souches sensibles que par les souches résistantes, en *cohLAB*. CohA est une oxydase dont l'induction de l'expression par la présence de cuivre est dépendante de la présence de CohL (Hsiao *et al.* 2011). CohA est présente dans le cytoplasme tandis que CohB est présente dans la membrane cytoplasmique (Teixeira *et al.* 2008).

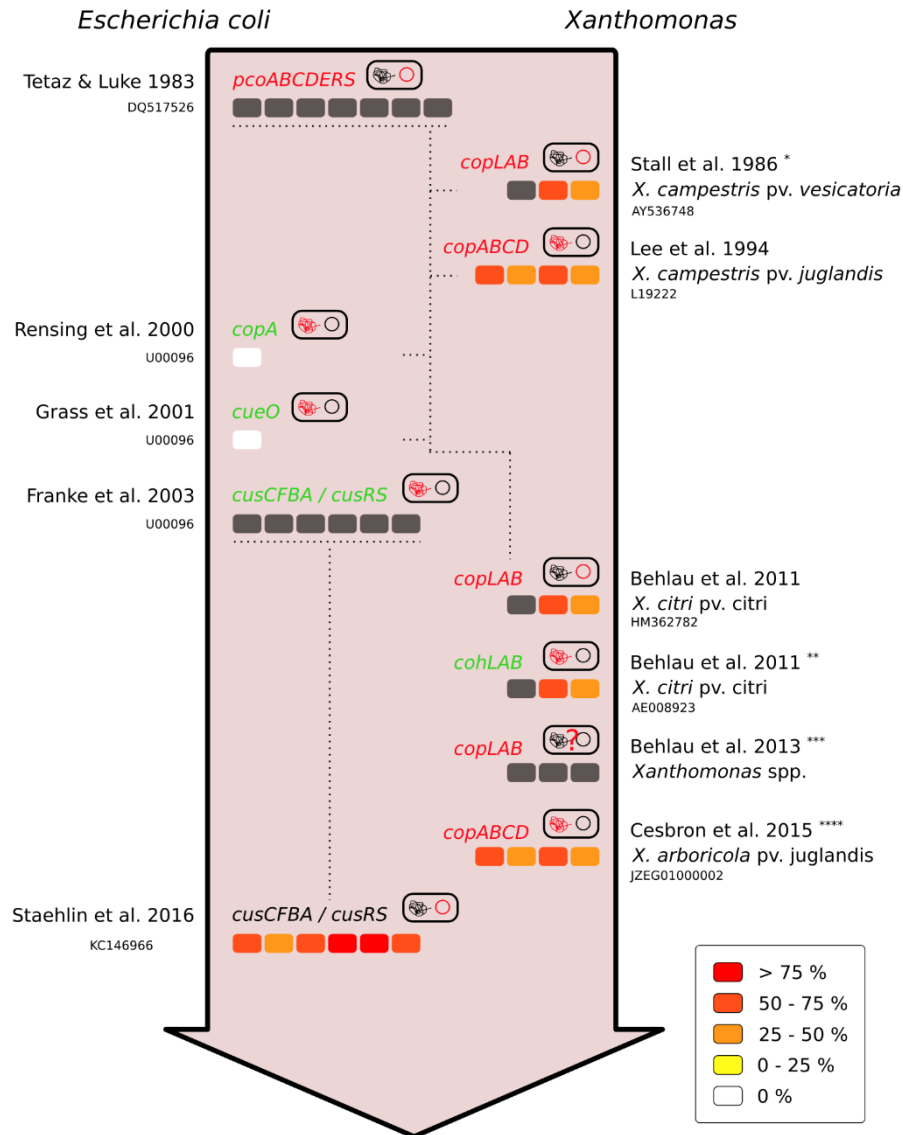


Figure 13 Frise chronologique des premières descriptions des systèmes d'homéostasie (police verte) et de résistance (police rouge, noire si non défini) au cuivre chez *Escherichia coli* (à gauche) et chez *Xanthomonas* spp. (à droite). L'identité entre la séquence d'acides aminés encodée par chaque gène et leur homologue du même nom chez *pcoABCDERS* est représentée par un code couleur (voir légende), sauf *copL* et *cohL* (pas d'homologue) et les gènes des deux systèmes *cus* qui ont été comparés entre eux. * Les auteurs n'ont pas nommé le système génétique responsable du phénotype résistant au cuivre, cependant nous l'avons déterminé par Blastn comme étant *copLAB* (données non présentées). ** *cohLAB* a été décrit antérieurement aux travaux de Behlau *et al.* (2011) mais cette étude est la première dans laquelle ils sont attribués à un mécanisme d'homéostasie et non de résistance, et dans laquelle les gènes sont mentionnés sous ce nom. *** Behlau *et al.* (2013) ont mis en évidence la présence de *copLAB* dans de multiples espèces de *Xanthomonas* : *X. citri* pv. *citri*, *X. euvesicatoria* pv. *citrumelo*, *X. vesicatoria*, *X. euvesicatoria*, *X. perforans*, *X. gardneri*, *X. arboricola* pv. *juglandis*. **** Chez *X. arboricola* pv. *juglandis* CFBP7179, le système *copABCD* est encodé sur un ICE.

III.3.b La résistance au cuivre

Chez l'organisme modèle *E. coli*

De façon additionnelle aux systèmes d'homéostasie du cuivre, certaines souches de *E. coli* peuvent encoder un système de résistance au cuivre Pco, généralement présent sur un plasmide et permettant à la bactérie de survivre à des concentrations en cuivre supérieures à celles gérées par les systèmes d'homéostasie (Figure 14). Ce système comporte les gènes *pcoABCDERS* (Tetaz & Luke 1983). Selon Zimmermann et al. (2012), PcoE serait une protéine chélatant préférentiellement le CuI dans le périplasme et qui constituerait une première ligne de défense précédant temporellement l'expression de *pcoABCD*. Elle aurait également un rôle de transporteur de CuI vers PcoA (Rensing & Grass 2003). PcoA est une oxydase multi-cuivre (*multicopper oxidase*) qui oxyde le CuI en CuII, moins toxique. PcoB n'est pas bien caractérisée mais serait une protéine de la membrane externe interagissant avec PcoA et PcoC. On ne sait pas à ce jour si PcoB permet d'exporter le cuivre de façon extracellulaire. PcoC est un transporteur de cuivre périplasmique possédant un site fixant le CuI et un site fixant le CuII. PcoD quant à elle, est une protéine membranaire. Ces deux dernières protéines sont nécessaires à l'expression d'un phénotype de résistance complète (Hernandez- Montes *et al.* 2012; Hobman & Crossman 2015). Ce cluster de gènes est régulé par PcoRS mais également par leurs paralogues chromosomiques CusRS (Brown *et al.* 1995; Munson *et al.* 2000). Chez *E. coli*, la résistance au cuivre est donc un système complexe impliquant des gènes plasmidiques qui confèrent une résistance accrue au cuivre, couplés à des gènes chromosomiques impliqués dans l'homéostasie du cuivre et qui servent de régulateurs (Cooksey 1993).

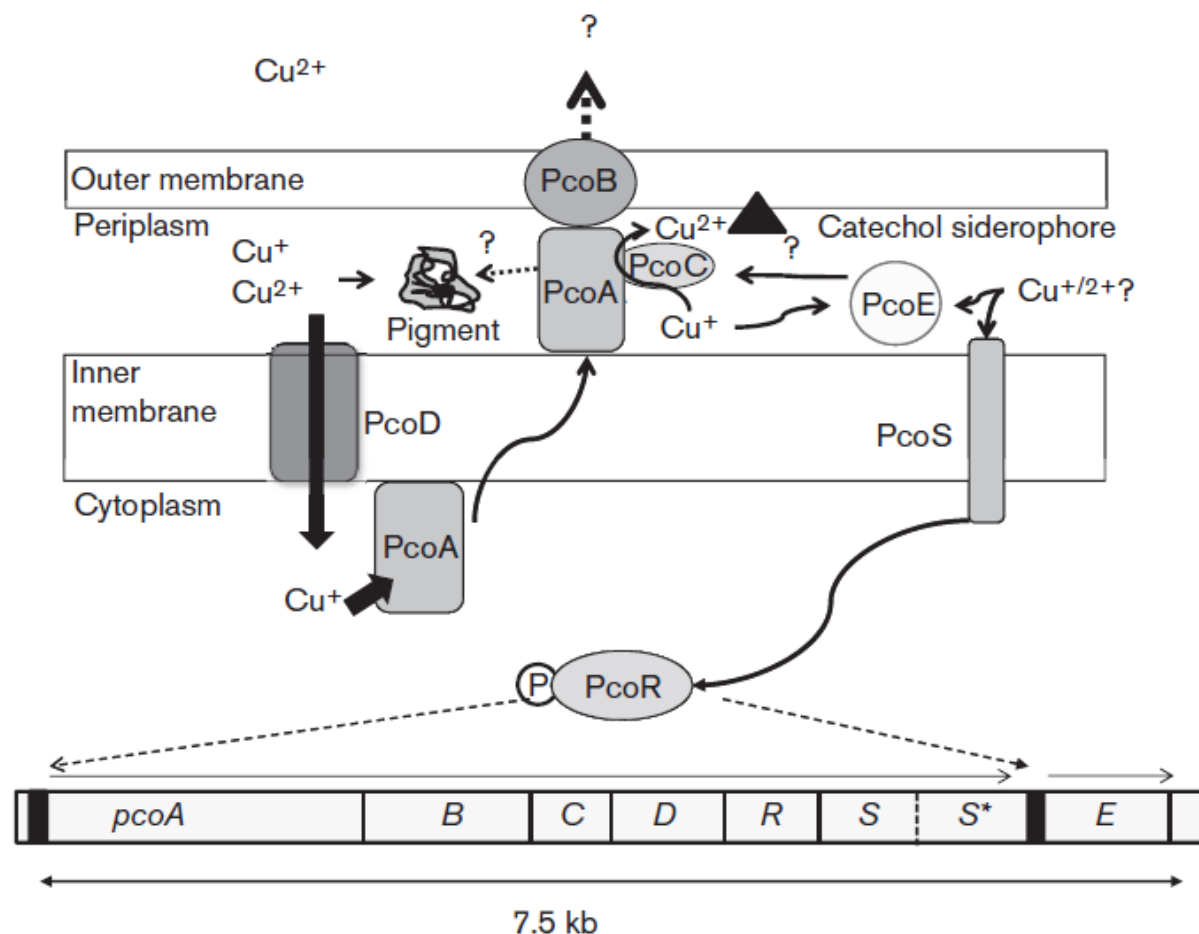


Figure 14 Fonctionnement du mécanisme de résistance au cuivre Pco porté par le plasmide pRJ1004 chez *E. coli*. Le mode de pénétration du cuivre dans la cellule n'est pas connu. Le système de détecteur/régulateur PcoRS active la transcription de pcoABCDRE en présence de cuivre. La localisation et la fonction des autres gènes est détaillée dans le texte. Des composés sécrétés par la bactérie, comme des pigments ou des catechol siderophores¹, et ayant une très forte affinité avec le cuivre sont potentiellement impliqués dans la résistance au cuivre conférée par pRJ1004, mais leurs mécanismes d'actions restent incompris. Figure tirée de Hobman and Crossman (2015).

Les mécanismes de résistance au cuivre ont d'abord été étudiés chez *E. coli* mais les phénomènes de résistance à ce métal ont été observés chez de multiples familles de bactéries et notamment des bactéries phytopathogènes. Le premier cas de suspicion de résistance chez ce type de bactéries semble être lié à l'observation de l'inefficacité des traitements cupriques contre la gale bactérienne dans les années 1960 aux États-Unis (Stall *et al.* 1986).

¹ Les catechol siderophores sont des composés ayant une très forte affinité avec le fer. Le complexe fer-siderophore peut être transporté à travers les membranes cellulaires.

Chez les bactéries phytopathogènes

Des cas de bactéries phytopathogènes résistantes au cuivre¹ ont été décrits chez *Pseudomonas* et *Xanthomonas*² dans les années 1980, mais aussi plus récemment chez *Agrobacterium tumefaciens*³ (Cooksey 1990; Nawapan *et al.* 2009). Chez quelques autres bactéries phytopathogènes comme *Erwinia amylovora* ou *Burkholderia gladioli*, des souches résistantes ont été décrites sans que les supports génétiques de celles-ci ne puissent être déterminés (Al-Daoude *et al.* 2009; Keith *et al.* 2005).

La nomenclature des protéines Cop impliquées dans la résistance au cuivre chez les *Xanthomonas* est hasardeuse. Deux systèmes distincts comportent des gènes de même nom mais non identiques : CopLAB comprend les gènes *copLABMGCDF* et CopABCD comprend les gènes *copABCD* (Figure 15).

Le système CopABCD

Le système de résistance au cuivre CopABCD a tout d'abord été décrit chez *Pseudomonas syringae* (Mellano & Cooksey 1988), et est homologue à PcoABCDE. *Coli* (74% d'identité en acide-aminés sur 99% de la protéine, 60% sur 80%, 59% sur 99% et 39% sur 98% pour PcoA/CopA, PcoB/CopB, PcoC/CopC, PcoD/CopD, respectivement). CopABCD a ensuite été identifié chez *Xanthomonas arboricola* pv. *juglandis* (Lee *et al.* 1994). Ce système comprend deux protéines périplasmiques CopA (une oxydase multi-cuivre (*multicopper oxidase*)) et CopC (fixant le cuivre), ainsi que deux protéines membranaires CopB (possiblement une protéine d'export du cuivre) et CopD (Hobman & Crossman 2015). Insérés seuls dans une souche naïve, CopA et CopB ne lui confèrent qu'un phénotype partiellement résistant tandis que CopCD ne modifient pas à eux seuls le phénotype. CopABCD conjointement permettent d'obtenir un niveau de résistance comparable à celui de la souche épidémique (Lee *et al.* 1994).

¹ Dans ce manuscrit, nous ne considérons comme résistantes au cuivre que les souches bactériennes possédant un système génétique différent des systèmes d'homéostasie connus et montrant une nette élévation de leur concentration minimale inhibitrice par rapport à des souches de référence.

² *Pseudomonas* et *Xanthomonas* appartiennent à la classe des Gammaprotéobactéries. Le genre *Xanthomonas* comprend plus de 200 espèces pathogènes de plantes cultivées telles que le riz, les agrumes ou les tomates (Hayward 1993) Les bactéries du genre *Pseudomonas* sont ubiquitaires. Majoritairement commensales, certaines souches de *Pseudomonas* soient phytopathogènes.

³ *Agrobacterium tumefaciens* appartient à la classe des Alphaprotéobactéries

Le système CopLAB

Un plasmide portant des déterminants génétiques de la résistance au cuivre différents de *copABCD* a été décrit chez une souche de *Xanthomonas* responsable de la gale bactérienne des solanacées (Potnis *et al.* 2015; Stall *et al.* 1986), puis chez différentes espèces de *Xanthomonas* et de *Pseudomonas* pathogènes ou commensales (Behlau *et al.* 2013; Bender & Cooksey 1986). Chez *Xanthomonas*, les systèmes impliqués dans la résistance au cuivre et leur support génétique sont variables (Basim *et al.* 2005; Bender *et al.* 1990; Cooksey *et al.* 1990; Richard *et al.* 2017c; Stall *et al.* 1986). Toutefois, le système *copLAB* comprenant de 6 à 8 gènes et encodé sur un plasmide constitue le cas le plus courant chez ce genre (incluant *X. arboricola*, *X. campestris*, *X. citri*, *X. euvesicatoria*, *X. gardneri* et *X. vesicatoria*) (Behlau *et al.* 2017; Behlau *et al.* 2013). Le système comprend l'oxydase multi-cuivre CopA, une protéine fixant le cuivre et CopB, un régulateur CopL, une oxydase impliquée dans le transport d'électrons CopM, une protéine d'export CopG, et deux transporteurs transmembranaires CopC et CopD (Behlau *et al.* 2011 ; Voloudakis *et al.* 2005). *copA*, *copB*, *copC* et *copD* du cluster de gènes *copLABMGCDF* encodent pour des protéines homologues à PcoABCD_{*E. coli*} (64%, 45%, 34% et 34% d'identité en acide-aminés, pour PcoA/CopA, PcoB/CopB, PcoC/CopC, PcoD/CopD, respectivement). Chez *Xanthomonas*, bien que les gènes *copABCD* possèdent une certaine homologie avec *pcoABCD* de *E. coli* et *Pseudomonas syringae*, le régulateur *copL* est original et semble remplacer *pcoRS* (Voloudakis *et al.* 2005). Sur la base de travaux de mutagenèse, il a été montré que les trois protéines CopLAB sont indispensables au phénotype résistant au cuivre, et que CopM augmente le seuil de tolérance (Behlau *et al.* 2011). L'insertion du système CopLAB complet dans une souche de *X. perforans* sensible ne confère cependant pas un niveau de résistance au cuivre aussi élevé que celui des souches possédant un plasmide de résistance au cuivre dans son intégralité (Behlau *et al.* 2011), suggérant la présence d'un autre système de résistance. Le système Cus, non transféré dans *X. perforans* lors de l'étude mentionnée précédemment, est un bon candidat pour ce rôle, puisqu'il est présent sur le plasmide conférant la résistance au cuivre chez *Xanthomonas citri* pv. *citri* A44 (nommée LM180 dans Richard *et al.* (2017c)).

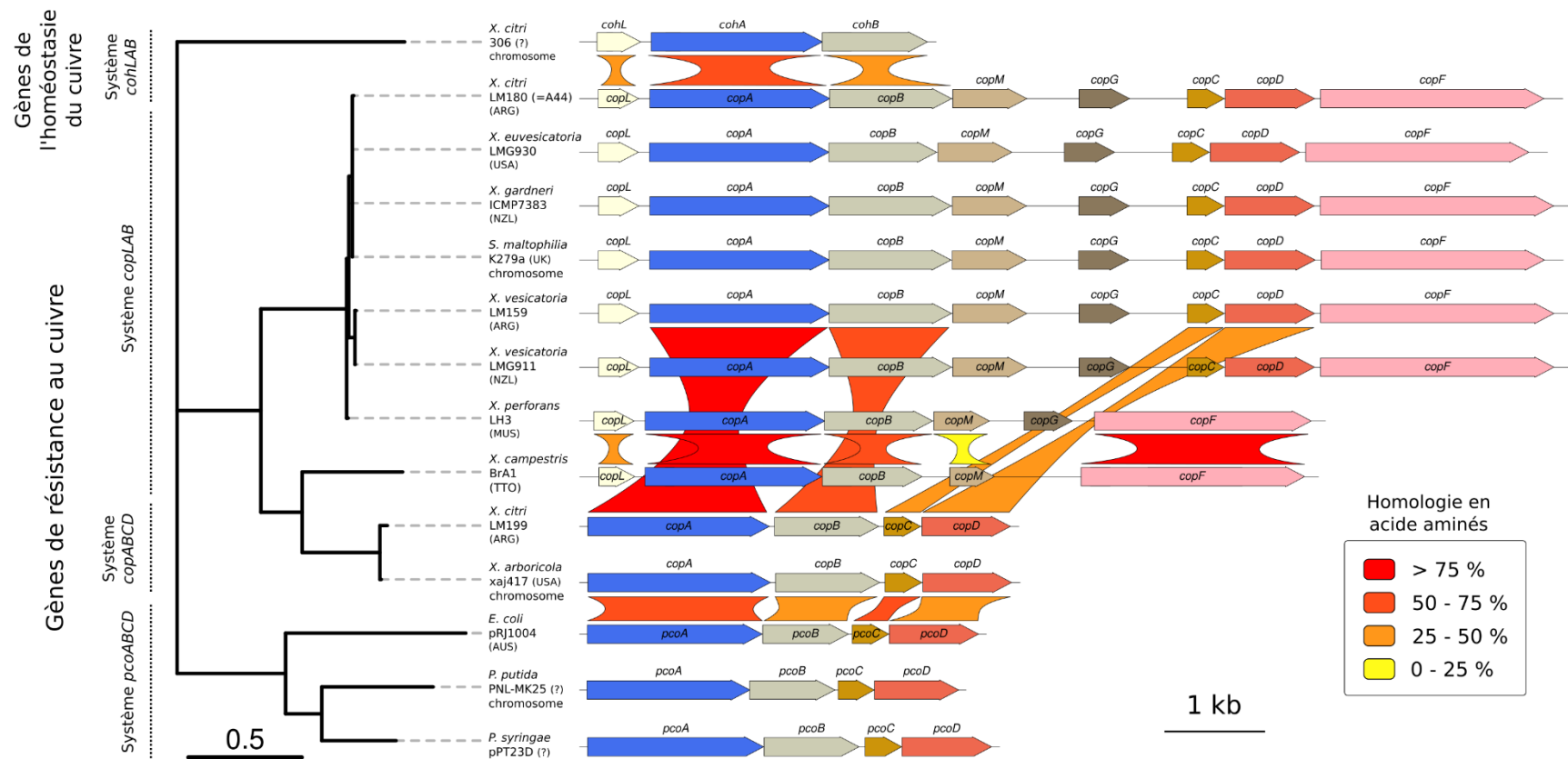


Figure 15 Représentation graphique de l'organisation des gènes impliqués dans l'homéostasie et la résistance au cuivre chez *Xanthomonas* et comparaison de ceux-ci avec les systèmes connus chez d'autres organismes. Sauf mention contraire, les systèmes sont plasmidiques. L'alignement utilisé pour calculer l'arbre de distances provient de la concaténation des alignements des homologues de *copA* et *copB*, seuls gènes présents dans tous les systèmes. L'identité entre les séquences d'acides aminés encodées par les gènes de plusieurs systèmes est représentée par un dégradé de blanc à rouge détaillé dans la légende. La matrice présentant toutes les valeurs d'identité est présente en Annexe I.

IV Une bactérie phytopathogène comme modèle d'étude : la bactérie monomorphe *Xanthomonas citri* pv. *citri*

IV.1 La plante-hôte : les agrumes

Les plantes connues sous le nom « d'agrumes » comprennent trois genres principaux de la famille des Rutaceae : *Citrus*, *Poncirus* et *Fortunella*. Originaire du Sud-Est asiatique, le genre *Citrus* regroupe la majorité des espèces commerciales d'agrumes, *Poncirus* ne comprenant qu'une seule espèce parfois utilisée comme porte-greffe et *Fortunella* regroupant des espèces à petits fruits appelés kumquats (Wu *et al.* 2018). Les agrumes commerciaux sont des hybrides primaires, secondaires, voire plus complexes principalement issus de trois espèces : les cédratiers (*C. medica*), les mandariniers communs (*C. reticulata*) et les pamplemoussiers (*C. maxima*) (Barrett & Rhodes 1976). Cultivés et consommés en Chine depuis au moins quatre millénaires (Gmitter & Hu 1990), les variétés d'agrumes connaissent une véritable expansion mondiale au XV^{ème} siècle, où ils sont introduits sur les continents africains et américains ainsi que dans le bassin méditerranéen, où leur culture à grande échelle se développera à partir du XIX^{ème} siècle (Ramón-Laca 2003). L'agrumiculture est aujourd'hui devenue la première production fruitière mondiale, avec plus de 140 pays producteurs totalisant 146 millions de tonnes annuelles en 2016 (données FaoStat).

De nombreux organismes appartenant à tous les domaines du vivant sont des bioagresseurs des agrumes. Il peut s'agir d'insectes ravageurs, par exemple la mineuse des agrumes *Phyllocnistis citrella* (Heppner 1993), ou vecteurs de maladies, tels que le puceron brun des agrumes *Toxoptera citricida* (Tsai 2008), principal vecteur du virus de la tristeza (CTV) (Bar-Joseph & Lee 1989), ou encore le psylle asiatique *Diaphorina citri* et le psylle africain (*Trioza erytreae*) qui transmettent la maladie du greening ou HLB (Gottwald 2010). Plusieurs espèces de champignons telles que *Alternaria* spp. (Timmer *et al.* 2003), *Guignardia citricarpa* (Yonow *et al.* 2013), *Pseudocercospora angolensis* (Yesuf 2013) et d'oomycètes comme *Phytophthora* spp. (Graham & Timmer 1992) infectent également les agrumes. Les pathogènes des agrumes comprennent aussi plusieurs bactéries : le phytoplasme responsable de la maladie des balais de sorcières (Bové 1986), la bactérie *Candidatus Liberibacter* responsable du HLB (Gottwald 2010), *Xanthomonas citri* pv. *citri* et pv. *aurantifolii* respectivement responsables du chancre Asiatique et du chancre Sud-Américain (Graham *et al.* 2004) ou encore *Xylella fastidiosa* responsable de la chlorose variéguee (Chang *et al.* 1993). Enfin plusieurs virus, dont le plus

important est le CTV, sont également la cause de maladies chez ce groupe végétal (Moreno *et al.* 2008).

IV.2 Le chancre asiatique des agrumes : généralités

Au cours du XX^{ème} siècle, l'uniformisation de l'agriculture, regroupant les aspects variétaux, culturels, temporels et géographiques, et l'intensification des échanges au niveau mondial ont facilité l'émergence et la dispersion géographique des agents pathogènes. La spécialisation d'hôte chez ces derniers a également été favorisée (Stukenbrock & McDonald 2008).

Le chancre Asiatique des agrumes fait partie des agents pathogènes ayant émergé pendant cette phase d'uniformisation. La bactérie qui cause cette maladie est *Xanthomonas citri* pv. *citri* (Xcc), de la classe des Gammaprotéobactéries. La zone d'origine du pathosystème (hôte et bactérie) est l'Asie (Pruvost *et al.* 2014; Wu *et al.* 2018), et des cas d'expansion géographique de la bactérie hors de son aire d'origine ont été mentionnés à partir du début du 20^e siècle.

Le chancre Asiatique des agrumes est une maladie d'importance majeure présente dans la plupart des régions agrumicoles du monde (Graham *et al.* 2004). Les conséquences néfastes de cette maladie sur l'agrumiculture sont multiples. Sur les espèces et cultivars sensibles, elle occasionne une baisse du rendement et de la qualité des fruits ainsi qu'une défoliation partielle des arbres. La maladie a également un fort impact indirect car elle représente un frein à l'exportation des fruits, conséquence du statut d'agent pathogène de quarantaine de Xcc. Les derniers grands bassins de production indemnes de maladie sont le bassin Méditerranéen, l'Afrique australe, le Mexique et la Californie.

IV.3 Pouvoir pathogène

IV.3.a *Symptômes et infection*

La bactérie pénètre initialement une plante saine par les stomates en présence d'eau ou par des blessures. Les symptômes sont observés sur feuilles, fruits et rameaux, y compris les épines (Figure 16). Les jeunes organes encore en croissance sont les plus prônes à l'entrée de la bactérie par les stomates, rendant les plantes greffées sur des porte-greffes à forte vigueur végétative particulièrement sensibles (Pruvost *et al.* 2015). La pénétration stomatale d'une seule cellule bactérienne semble suffisante pour déclencher une infection (Gottwald & Graham 1992). Les blessures causées par les pratiques agricoles, le vent ou les insectes comme la mineuse asiatique des agrumes permettent cependant l'infection de tous les organes

indépendamment de leur stade phénologique. Aux premiers stades d'infection, la maladie se présente, sur la face inférieure des feuilles, sous la forme de petites taches huileuses de un à quelques millimètres.

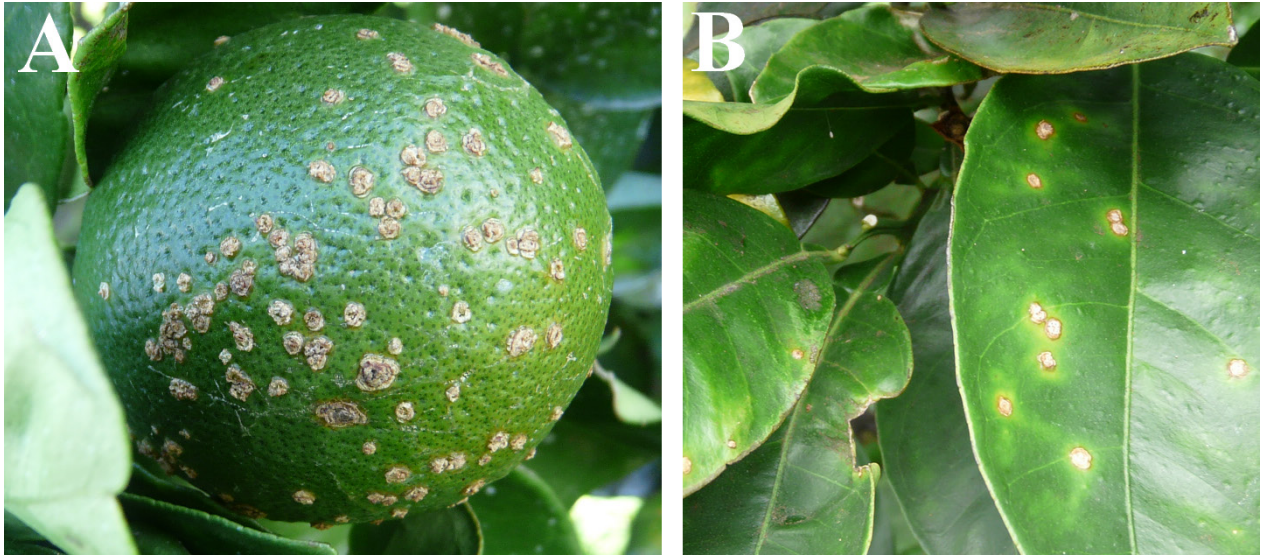


Figure 16 Symptômes causés par le chancre citrique sur (A) fruit et (B) feuille de *Citrus* sp. Photographie Damien Richard.

Sous l'effet de l'hyperplasie et de l'hypertrophie des cellules de l'hôte (Figure 17), les symptômes circulaires prennent ensuite du volume pour former des pustules, qui peuvent être entourées d'un halo chlorotique. Les lésions prennent ensuite une couleur brun clair à foncé avec un aspect liégeux et rêche au toucher (lésions chancreuses). Les lésions les plus âgées prennent un aspect cratériforme. La chute des feuilles et des fruits peut se produire sur des arbres très touchés, mais la maladie ne provoque pas la mort de l'arbre. L'hypertrophie et

l'hyperplasie entraînent des ruptures d'épiderme qui facilitent l'exsudation rapide et importante des populations bactériennes (Pruvost *et al.* 2002).

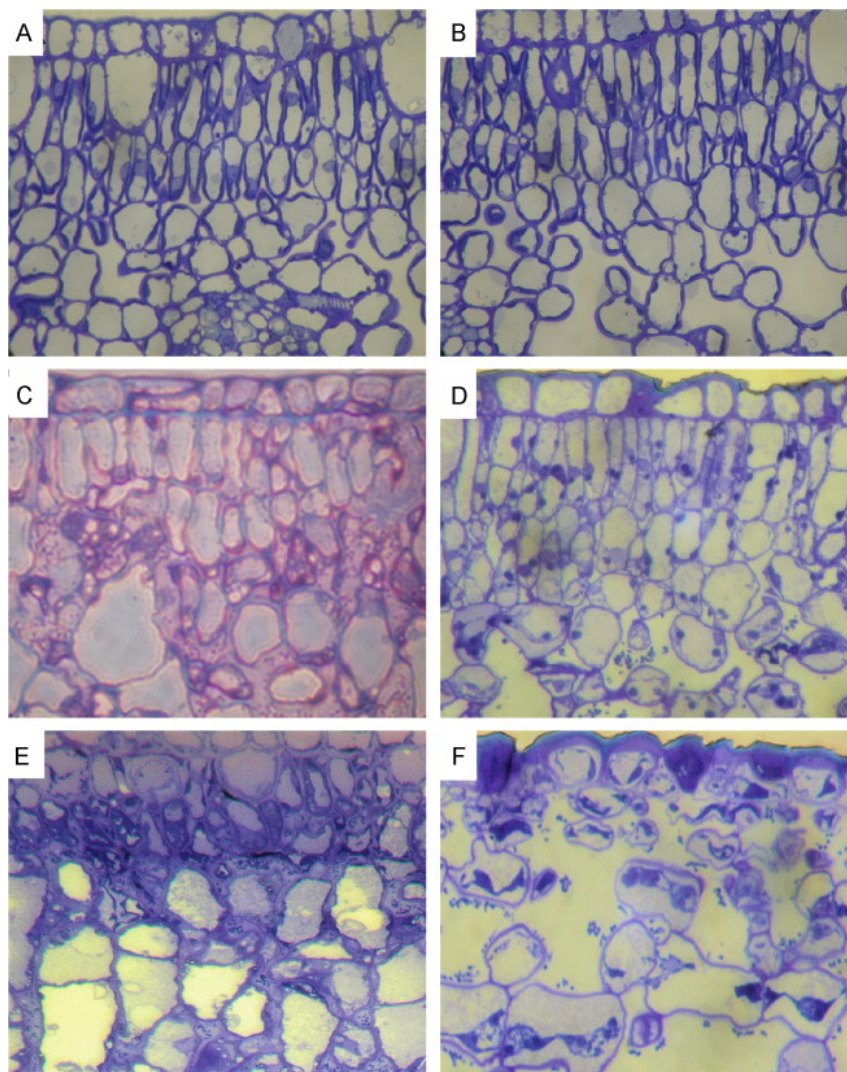


Figure 17 Coupes microscopiques de feuilles saines et de feuilles inoculées par *Xanthomonas citri* pv. *citri* pathotype A. Coupes de feuilles de pomelo cv. Duncan (*C. paradisi*) (A,C,E) et de kumquat (*Fortunella margarita*) cv. Nagami (B,D,F) trois jours (C,D) et sept jours (E,F) après inoculation d'une solution de Xcc A à 5×10^5 cfu/ml. L'hypertrophie, qui correspond à l'augmentation du volume des cellules, est observable en C,D,E et F, tandis que l'hyperplasie, qui correspond à l'augmentation du nombre de cellules, apparaît peu sur cette illustration. Figure tirée de Khalaf *et al.* (2007).

IV.3.b Bases génétiques du pouvoir pathogène

Entre le relargage d'inoculum sur un hôte sain et la formation de symptômes visibles, le processus d'infection de Xcc mobilise une multitude de facteurs de virulence.

Déposée à la surface d'une feuille, une cellule de Xcc doit tout d'abord s'y fixer puis se déplacer jusqu'à une ouverture de la plante. À ce stade précoce de l'infection, deux mécanismes sont

particulièrement importants pour la survie et la virulence de Xcc : l'adhésion et l'attraction chimiotactique, tous deux régulés par le système chimiosensitif (Mhedbi-Hajri *et al.* 2011). L'adhésion de la bactérie à la surface de la feuille est la toute première étape du phénomène d'agrégation qui conduit ultimement à la formation de biofilm. L'étape d'adhésion repose sur l'excrétion de protéines à la surface de la cellule bactérienne. L'attraction chimiotactique repose quant à elle sur la détection de signaux environnementaux (*e.g.* concentration en toxiques, en oxygène ou en nutriments) par des récepteurs spécifiques. Ces derniers peuvent être de type MCP (*Methyl-accepting Chemotaxis Protein*), STCRS (*Sensors of Two-component Regulatory System*) ou encore TBDT (*TonB-Dependent Transporter*) (Mhedbi-Hajri *et al.* 2011). Le signal est ensuite transmis au moteur du flagelle, qui va permettre le déplacement de la cellule vers une zone avec une concentration supérieure en molécules bénéfiques ou inférieure en molécules néfastes à la bactérie. Les espèces et pathovars de *Xanthomonas* présentent globalement des répertoires uniques de gènes impliqués dans les étapes précoces de l'interaction plante-bactérie (Mhedbi-Hajri *et al.* 2011).

Les exopolysaccharides (EPS) excrétés par Xcc forment une matrice protectrice autour des bactéries, structurant le biofilm (Rigano *et al.* 2007). Les biofilms sont connus pour fournir une protection aux bactéries contre diverses agressions chimiques ou physiques telles que les stress environnementaux, les mécanismes de défense de l'hôte ou encore les antimicrobiens (Rigano *et al.* 2007). Ainsi, des mutants ne produisant pas d'EPS présentent une sensibilité accrue à divers stress incluant les stress oxydatifs (Li & Wang 2012). Le rôle des EPS dans la rétention de l'eau autour des cellules bactériennes a été démontré chez *Xanthomonas campestris* (Denny 1999). La matrice intercellulaire formée par les EPS favorise le déplacement du phytopathogène à l'intérieur des tissus de son hôte. Même si toutes ces propriétés n'ont pas été démontrées expérimentalement chez Xcc, les différentes propriétés des EPS les rendent importants pour sa pathogénicité. Les mutants qui en sont dépourvus possèdent une agressivité réduite (Yan & Wang 2012).

Les lipopolysaccharides (LPS) sont des constituants majeurs de la membrane externe des bactéries, et représentent une autre classe de molécules impliquées dans le pouvoir pathogène (Büttner & Bonas 2010; Pruvost *et al.* 2015; Ryan *et al.* 2011). En effet, des mutants de Xcc ne produisant pas de LPS ne forment pas de biofilms, présentent une réduction de la motilité et de la résistance aux stress environnementaux ainsi qu'une agressivité réduite (Li & Wang 2011; Petrocelli *et al.* 2012). Les LPS sont détectés par le système immunitaire inné de la plante qui active la réponse immunitaire basale (aussi appelée PTI, pour *PAMP Triggered Immunity*).

Cette réponse comprend des modifications cellulaires visant à stopper la prolifération de l'agent pathogène : la production de composés antibactériens, le renforcement de la paroi cellulaire, l'inhibition de l'entrée de glucose dans la cellule ou encore des cascades de voies métaboliques qui peuvent aboutir à la formation de peroxyde d'hydrogène à l'intérieur de la cellule végétale (Ma & Berkowitz 2007).

Une fois à l'intérieur des tissus végétaux, plusieurs mécanismes de sécrétion sont importants pour le pouvoir pathogène de Xcc. Le T2SS (Système de Sécrétion de Type II) exporte en effet des enzymes de dégradation telles que des cellulases, des protéases ou encore des lipases qui dégradent la paroi cellulaire végétale (Büttner & Bonas 2010). En même temps, la bactérie va tenter d'inhiber la réponse immunitaire basale et de manipuler la cellule-hôte à son profit. Si cette interaction se produit, l'hôte est sensible à Xcc, et leur relation peut être qualifiée de « hôte », ou encore de « compatible ». Ce dialogue entre la bactérie et la plante est notamment possible grâce au système de sécrétion de type III (T3SS), à travers duquel Xcc injecte des effecteurs à la cellule végétale (Ryan *et al.* 2011). Certains de ces effecteurs sont reconnus par la plante, ce qui entraîne l'activation de mécanismes de défense (aussi appelés *effector-triggered immunity* ou ETI) (Jones & Dangl 2006a). Les effecteurs TALE (pour *Transcription Activator-Like Effector*) sont des protéines capables de reconnaître des motifs nucléotidiques de régions promotrices et qui jouent le rôle de régulateur de transcription eucaryote, avec une spécificité variable. Chez Xcc, un de ces effecteurs (PthA4 ou ses variants) est indispensable à l'induction des symptômes du chancre Asiatique des agrumes. Sa seule présence est suffisante pour induire l'hyperplasie et l'hypertrophie cellulaire chez son hôte (Duan *et al.* 1999). Chez Xcc souche 306, *pthA4* et ses trois variants (*pthA1*, *pthA2* et *pthA3*) sont encodés sur deux plasmides, le premier étant indispensable au pouvoir pathogène de Xcc (Brunings & Gabriel 2003). PthA4 modifie la transcription de plusieurs gènes chez les agrumes, causant les symptômes caractéristiques du chancre asiatique. *csLOB1*, dont la transcription est modifiée par PthA4, est chez les agrumes un régulateur des gènes impliqués dans l'expansion cellulaire. *csSWEET1* est quant à lui un transporteur extracellulaire de sucre (glucose et sucrose), dont la surexpression fournit des nutriments aux cellules bactériennes (Hu *et al.* 2016b). Une comparaison de profils transcriptomiques entre la souche Xcc 306 et son équivalent dépourvu du gène *pthA4* (mutant de délétion) a été réalisée. Les résultats ont permis d'attribuer à *PthA4* la diminution de la transcription de gènes impliqués dans plusieurs fonctions comme la défense contre les agents pathogènes, et l'augmentation de celle de gènes impliqués dans diverses

fonctions comme la dégradation des parois cellulaires, la multiplication cellulaire ou la synthèse de protéines (Hu *et al.* 2016b).

IV.3.c Gamme d'hôtes

La classification infraspécifique associée aux bactéries phytopathogènes s'organise souvent en pathovars eux-mêmes parfois subdivisés en pathotypes. Le pathovar regroupe des souches ayant des caractéristiques communes de pouvoir pathogène (symptômes, gamme d'hôtes, niche colonisée). La gamme d'hôtes de bactéries d'un même pathovar ne s'étend généralement pas au-delà d'une famille botanique. Par exemple, *Xanthomonas citri* pv. *mangiferaeindicae* est pathogène de quelques plantes de la famille des Anacardiaceae alors que *Xanthomonas citri* pv. *citri* est pathogène des Rutaceae. Au sein d'un pathovar infectant une famille botanique, des groupes de souches peuvent présenter une gamme d'hôte variable et induire différents types de réponse immunitaire. Cette variabilité intra-pathovar définit des pathotypes. Xcc est subdivisé en 3 pathotypes : A, A* et A^w. Le pathotype A a une gamme d'hôtes assez large qui comprend tous les agrumes commerciaux du genre *Citrus*, ainsi que quelques espèces de genres proches. Les souches du pathotype A* sont pathogènes du limettier mexicain (*C. aurantifolia* – hôte d'isolement de loin le plus fréquent) mais également de quelques espèces taxonomiquement proches comme le porte-greffe *C. macrophylla*, le limettier de Tahiti (*C. latifolia*) ou le limettier doux (*C. limettioides*). Le pathotype A^w est également virulent sur limettier Mexicain et *C. macrophylla* mais a la particularité d'induire une réaction hypersensible sur des espèces non-hôtes (Brunings & Gabriel 2003; Rybak *et al.* 2009).

IV.4 Structure génétique

La structure génétique de Xcc a été caractérisée de manière globale par AFLP (pour *Amplified Fragment Length Polymorphism*) (Bui Thi Ngoc *et al.* 2009) et par un schéma de marqueurs minisatellites (MLVA31, pour *MultiLocus Variable number of tandem repeat Analysis*). Ces analyses ont permis de regrouper la diversité mondiale en 4 lignées génétiques (Pruvost *et al.* 2014). De façon intéressante, la structuration génétique de Xcc correspond à la classification par la gamme d'hôtes : les pathotypes A*, A^w et A correspondent aux lignées 4, 3, et au regroupement des lignées 1 et 2, respectivement (Figure 18).

Un autre schéma MLVA, basé sur l'analyse de 14 loci microsatellites, permet d'apprécier les relations génétiques entre souches à une échelle évolutive beaucoup plus petite. Ces marqueurs rendent en effet possible l'analyse de la structure spatiale et des paramètres de dispersion de

Xcc A. Ils ont par exemple été utilisés pour mettre en évidence une lignée épidémique au Vietnam, où la maladie est endémique (Verniere *et al.* 2014).

Enfin, une analyse récente sur la base de données de génomes complets a confirmé la structure génétique du pathovar préalablement établie par géotypage partiel. Elle a décrit un ensemble de processus et de caractéristiques des pathotypes de Xcc, incluant des mutations ponctuelles, des îlots génomiques et/ou de recombinaison ou encore des événements de pseudogénisation¹ (Gordon *et al.* 2015). Les auteurs n'ont observé qu'un faible nombre de gènes impliqués dans des événements évolutifs pathotype-spécifiques. Ces derniers sont majoritairement présents à des loci possédant des traces de recombinaison et/ou de transfert horizontal de gènes. Ils ont également identifié des SNP non-synonymes et pathotype-spécifiques dans des gènes connus pour être impliqués dans le pouvoir pathogène.

Une expansion géographique majeure de Xcc en dehors de sa zone de répartition originelle asiatique a lieu depuis le début du 20^{ème} siècle. Le pathotype A, et plus spécifiquement la lignée génétique 1, est majoritairement responsable de cette expansion (Pruvost *et al.* 2014). Cette lignée s'est disséminée dans l'Océan Indien, en Afrique de l'Ouest, en Amérique du Sud et du Nord, et en Océanie. Les émergences de la dernière décennie ont, quant à elles, principalement concerné le continent africain (Balestra *et al.* 2008; Derso *et al.* 2009; Leduc *et al.* 2015; Leduc *et al.* 2011). En contraste avec la situation prédominante au 20^{ème} siècle, ces émergences ont concerné plusieurs lignées génétiques avec parfois des populations fondatrices probablement originaires de plusieurs pays distincts (Leduc *et al.* 2015). Cette structure illustre l'augmentation du risque de contamination de territoires indemnes de maladie probablement liée à l'augmentation des échanges internationaux.

¹ Un pseudogène est un gène devenu non fonctionnel à cause de mutations génétiques. Il ne conduit donc pas à l'expression d'une protéine fonctionnelle.

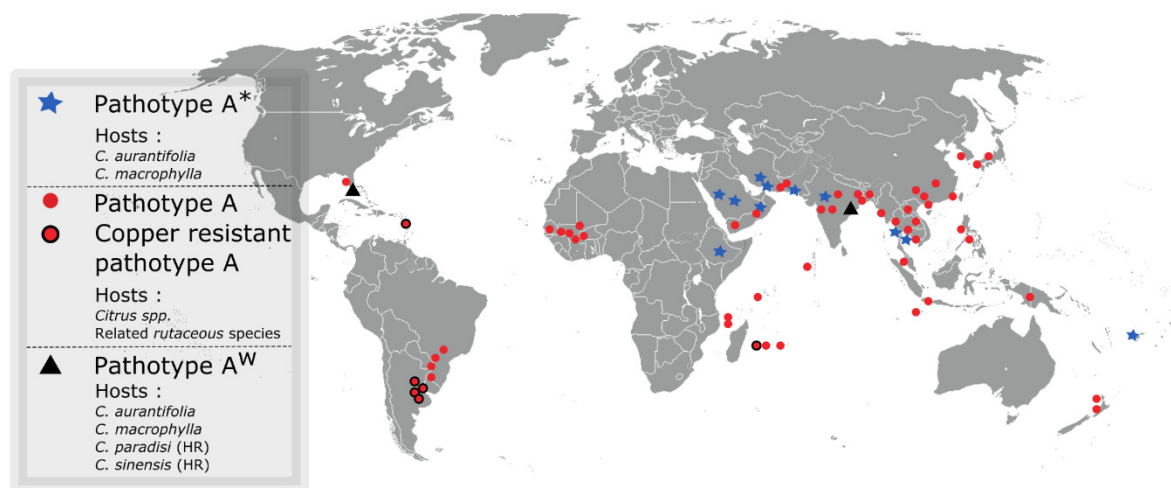


Figure 18 Association entre la gamme d'hôtes, les pathotypes, le phénotype vis-à-vis du cuivre et la distribution géographique de Xcc.

IV.5 Epidémiologie

IV.5.a *Survie de la bactérie*

Xcc survit principalement dans les symptômes chancreux qu'il provoque sur agrumes (Graham *et al.* 2004). Après l'infection initiale, le nombre de bactéries augmente rapidement dans les symptômes. Environ huit jours après l'infection, le nombre de bactéries atteint son maximum, de l'ordre de 10^7 bactéries par symptôme. L'épiderme de la plante rompt, ce qui facilite la libération des bactéries à la surface de la plante (Brunings & Gabriel 2003). En présence d'eau libre, ces dernières peuvent alors être disséminées et induire des infections secondaires (Pruvost *et al.* 2002). Les densités de populations de Xcc dans les symptômes diminuent au cours du temps. Cette diminution dépend des conditions environnementales locales. Dans les régions où il existe une saison hivernale marquée, une baisse drastique des populations allant jusqu'à 4 unités logarithmiques a été observée (Koizumi & Kochinotsu 1977; Stall *et al.* 1980). Dans un environnement tropical, la baisse d'effectif enregistrée dans des symptômes âgés de 18 mois était au maximum d'1,5 log (Pruvost *et al.* 2002). La survie de Xcc à l'intérieur de symptômes présents sur des rameaux pendant plusieurs années a été observée (Graham *et al.* 2004).

Toutefois, Xcc a une aptitude limitée à la survie asymptomatique sur ou en dehors de ses hôtes (Graham *et al.* 2004). Sa capacité de survie asymptomatique est grandement conditionnée par la formation d'un biofilm à la surface des organes végétaux qui protège la bactérie des stress environnementaux tels que la dessiccation, les rayonnements ultraviolets ou l'action de molécules antibactériennes (Cubero *et al.* 2011; Sena-Vélez *et al.* 2015). La survie de Xcc a été très largement abordée par dénombrement des populations cultivables. Des travaux ont suggéré

l'existence d'un état viable non cultivable¹ lorsque Xcc est placé en conditions de stress nutritif et en présence de cuivre. Cet état serait réversible (Golmohammadi *et al.* 2013). L'importance biologique des populations viables non cultivables reste méconnue.

IV.5.b *Dissémination naturelle et humaine de la bactérie*

La dissémination de Xcc sur de courtes distances, au sein des rameaux d'un même arbre, ou entre plants de pépinière, généralement peu espacés, se produit par éclaboussures d'eau contaminée. À l'échelle d'un verger, la dispersion de la bactérie est possible lors d'épisodes associant du vent à une vitesse supérieure à 8 m.s⁻¹ et de la pluie (Serizawa & Inoue 1975). Au sein d'une exploitation agricole, la dissémination peut également avoir une origine humaine lors des opérations d'entretien effectuées avec des outils et équipements agricoles non désinfectés. De plus, la dissémination sur des distances bien plus importantes, de l'ordre du kilomètre à la dizaine de kilomètres, a été décrite après des événements cycloniques (Irey *et al.* 2006). Cependant, l'introduction de l'inoculum primaire dans une zone géographique précédemment saine est essentiellement causée par l'action humaine (Graham *et al.* 2004; Pruvost *et al.* 2015). Les variétés sélectionnées d'agrumes sont généralement multipliées par greffage. Aussi, le commerce mondial ou le transport illégal de variétés sous forme de greffons a constitué la cause majeure de l'expansion géographique de la maladie sur de très grandes distances, bien que plusieurs pays aient déjà adopté une législation sur le transport des agrumes. La contamination des pépinières, qui fournissent localement des plants aux agriculteurs, est une des causes majeures de la dissémination locale de la maladie. Cela a été le cas aussi bien dans l'aire d'origine de la bactérie (Verniere *et al.* 2014) qu'en zone d'émergence (Leduc *et al.* 2015). Le génotypage avec des marqueurs microsatellites s'est révélé à plusieurs reprises efficace pour déterminer la source d'inoculum primaire à l'origine de l'émergence de Xcc dans de nouveaux territoires (Leduc *et al.* 2015; Verniere *et al.* 2014).

IV.6 Méthodes de lutte

IV.6.a *Mesures préventives*

Dans l'Union Européenne, l'introduction et la dissémination de *Xanthomonas citri* pv. *citri* sont interdites (Annexe II-A Chapitre I de la Directive 2000/29 CE). L'introduction et la circulation

¹ En bactériologie, l'état viable non cultivable désigne un état de très faible activité métabolique. Le faible nombre de divisions cellulaires ne permet donc pas de former une colonie visible sur un milieu de culture. Les cellules bactériennes à l'état viable non cultivable sont en revanche bien vivantes et peuvent, si l'environnement devient favorable, devenir cultivables.

des plants d'agrumes venant de pays tiers sont également interdites. Des exigences particulières sont imposées concernant les importations de fruits d'agrumes provenant de pays tiers : provenance de zones obligatoirement exemptes de maladie, contrôle visuel de l'absence de symptômes ou système de certification équivalent (Annexe IV-A Chapitre I de la Directive 2000/29 CE). Dans les DOM Français, des exigences particulières sont également requises pour l'introduction d'agrumes. L'arrêté du 31 juillet 2000 établissant la liste des organismes nuisibles aux végétaux soumis à des mesures de lutte obligatoire, prévoit, pour les Antilles et la Guyane, l'inscription du chancre citrique en annexe A (qui stipule la liste des organismes pour lesquels la lutte est obligatoire, de façon permanente, sur tout le territoire). Pour la Réunion, l'arrêté préfectoral 2011-1479 du 30 septembre 2011 interdit l'introduction et la dissémination de *Xanthomonas citri* pv. *citri*, classé comme organisme nuisible présent sur le territoire de la Réunion et dont la dissémination doit être limitée. Ces mesures préventives visent à empêcher l'entrée du pathogène sur un territoire, et ne sont efficaces qu'en association avec des organismes de contrôle en mesure d'appliquer des protocoles de détection.

IV.6.b Eradication

Lors de l'émergence de la maladie dans une zone géographique précédemment saine, la stratégie de lutte dépend de l'importance économique des agrumes dans la zone, de la situation économique du pays concerné, et de la précocité de détection du(es) foyer(s) de maladie. Lorsque que le risque associé à l'émergence de Xcc est élevé et que sa détection est précoce, la meilleure option est l'éradication. Cette méthode de lutte consiste à arracher et à brûler les arbres infectés ainsi que des plantes-hôtes présentes dans la zone du foyer d'émergence. Ces opérations sont généralement supervisées par les services de protection des végétaux qui placent également la zone en quarantaine (pas d'échanges de matériel végétal dans et à partir de cette zone et interdiction de replantation). Xcc présente des caractéristiques qui rendent cette méthode de lutte radicale appropriée. En effet, Xcc (i) a une courte survie hors de son hôte, (ii) n'a pas de vecteur non humain efficace, (iii) présente des symptômes facilement identifiables, autorisant un diagnostic précis et rapide, (iv) a une gamme d'hôtes restreinte à une culture pérenne à haute valeur ajoutée, (v) est virulent sur un grand nombre de variétés d'agrumes, rendant les autres mesures de contrôle de la maladie très coûteuses et modérément efficaces. La maladie a ainsi été éradiquée avec succès sur plusieurs territoires : USA au début du XX^{ème} siècle (Schoulties *et al.* 1987), Afrique du Sud (Calavan 1956), Nouvelle Zélande (Dye 1969) et Australie (Gambley *et al.* 2009). Cependant, la réémergence du pathogène dans ces territoires est toujours possible, comme cela a été le cas en Floride où les autorités ont, malgré la mise en

œuvre de moyens considérables pendant plus de 10 ans, échoué à éradiquer la maladie (Gottwald & Irey 2007). Un autre exemple concerne l'Australie, où la maladie a été de nouveau détectée en 2018 dans deux États (non publié au moment de la rédaction de ce manuscrit), après plusieurs réussites d'éradication (Broadbent 1995; Gambley *et al.* 2009). Dans une zone préalablement indemne de Xcc où l'agrumiculture est économiquement importante et qui possède des services de protection des végétaux efficaces, les analyses coût-bénéfice ont montré la pertinence économique de l'option d'éradication (Alam & Rolfe 2006; Jetter *et al.* 2000; Schubert *et al.* 2001). Malgré des efforts considérables, l'échec de l'éradication en Floride mentionné précédemment est à mettre en relation avec l'existence dans cette zone de conditions environnementales très favorables aux épidémies. De plus, le rayon d'éradication autour des foyers détectés a initialement, c'est-à-dire dans les années 1990, été insuffisant au vu des paramètres de dispersion de la bactérie aujourd'hui mieux connus. Enfin, de nombreuses actions en justice des propriétaires d'arbres détruits ont fortement ralenti le processus (Centner & Ferreira 2012; Gottwald *et al.* 2001; Irey *et al.* 2006).

IV.6.c *La lutte intégrée*

Lorsque l'éradication est impossible, comme dans les régions où le pathogène est endémique, des méthodes de lutte intégrée existent et peuvent prévenir ou limiter l'apparition de la maladie (Graham *et al.* 2004). Cette stratégie, la plus commune à l'échelle mondiale, consiste à mettre en place une multitude de mesures visant à limiter l'apparition, le développement et la dissémination d'un agent pathogène. Ces mesures peuvent être de nature très variée mais leur choix est rationnel : il prend en compte aussi bien les facteurs économiques qu'environnementaux ou toxicologiques (Kogan 1998).

Composante chimique de la lutte

L'usage d'antibiotiques ou plus souvent de pesticides à base de cuivre fait partie de la lutte chimique contre ce pathogène. Ces molécules sont utilisées massivement pour protéger les organes pendant leur phase de forte sensibilité : les feuilles durant les poussées végétatives et les fruits durant les 3 mois suivant la nouaison.

En France, les traitements autorisés sur les agrumes pour traiter les bactérioses ne comprennent que deux formulations à base de cuivre : l'Heliocuire (dont la molécule active est l'hydroxyde de cuivre) et le Novicure (dont la molécule active est le sulfate de cuivre tribasique). Tout comme la législation française, la législation européenne n'autorise pas l'utilisation de molécules antibiotiques en agriculture. Les molécules actives autorisées en agriculture et

utilisables pour le contrôle de Xcc ne comprennent ainsi que des formulations à base de cuivre : le sulfate de cuivre tribasique, l'hydroxyde de cuivre, l'oxyde de cuivre, l'oxychloride de cuivre et la bouillie bordelaise (sulfate de cuivre neutralisé par de la chaux éteinte).

Les pesticides de la famille des néonicotinoïdes présentent plusieurs propriétés qui les rendent efficaces. En effet, lorsqu'ils sont appliqués sur le sol, ils induisent une résistance accrue des *Citrus* à Xcc sur des périodes allant jusqu'à 24 semaines (Francis *et al.* 2009; Graham & Myers 2011, 2013). Des insecticides néonicotinoïdes sont aussi régulièrement utilisés sur les agrumes pour lutter contre la mineuse des agrumes (FERENCE *et al.* 2018). Bien qu'elle ne soit pas vectrice de Xcc, cette chenille cause des blessures de l'épiderme végétal en y creusant des galeries, ce qui favorise l'infection bactérienne. L'imidaclopride, qui produit une molécule analogue à l'acide isonicotinique et est un inducteur de l'immunité acquise des plantes, apporte une protection équivalente à un programme de pulvérisation de cuivre ou de streptomycine (Graham & Myers 2013).

L'usage fréquent de certaines molécules a cependant rapidement entraîné l'émergence de souches résistantes. Par exemple, les producteurs en Corée du Sud utilisent la streptomycine pour lutter contre Xcc. En 2012, une étude moléculaire a montré qu'une grande partie des souches de Xcc prélevées dans des vergers traités à la streptomycine possèdent une résistance à cet antibiotique (Hyun *et al.* 2012). Les bases génétiques de cette résistance sont variables puisqu'il peut s'agir d'une mutation ponctuelle sur le gène chromosomique *rpsL* ou de l'intégration d'un plasmide portant le système *strA-strB*. En Argentine, un phénotype de résistance au cuivre a également émergé depuis 1994 (Canteros *et al.* 2017). Les souches résistantes étaient porteuses d'un large plasmide (>200kb) transmissible par conjugaison (Behlau *et al.* 2011 2012). L'utilisation de cuivre additionné de mancozèbe a été efficace en Argentine pour contrôler partiellement des souches résistantes au cuivre, et en traitement préventif pour empêcher l'émergence de la résistance (Canteros *et al.* 2017).

Des traitements alternatifs appelés stimulateurs de défenses naturelles, qui activent le système immunitaire acquis des plantes même en l'absence de pathogènes, ont également montré leur efficacité. Chez les plantes, l'acide salicylique a un rôle de signal local et systémique pour activer la résistance contre les pathogènes biotrophes (Jones & Dangl 2006b). Ainsi, l'acibenzolar-S-méthyl est un homologue fonctionnel de l'acide salicylique et est couramment utilisé comme stimulateur de défenses naturelles (Oostendorp *et al.* 2001). L'utilisation de bactériophages pour contrôler la maladie est également une piste de recherche pour le

développement de nouveaux traitements, mais n'a pas encore montré de réelle efficacité au champ (Balogh *et al.* 2008).

Méthodes prophylactiques

La prophylaxie comprend l'ensemble des mesures de prévention de l'arrivée de la maladie ou de sa propagation. Les mesures prophylactiques qui concernent Xcc sont en lien avec les deux modes de transmissions privilégiés de la maladie qui sont la diffusion humaine de plants contaminés et l'association de pluie et de forts vents (voir IV.5.b). L'adaptation des pratiques culturales, avec la suppression de l'irrigation par aspersion, la mise en place de haies brise vent ou encore la désinfection des outils et équipements agricoles, en font partie. L'utilisation de plants certifiés indemnes de maladie et de plusieurs variétés d'agrumes avec des niveaux de résistance variables est également préconisée. En effet, bien que toutes sensibles, les espèces et variétés commerciales d'agrumes présentent des niveaux de sensibilité variables. Les kumquats (*Fortunella* spp.) font partie des plus résistantes, le citronnier (*C. limon*) et le mandarinier (*C. reticulata*) sont sensibles et modérément sensibles, respectivement alors que le combava (*C. hystrix*) et le pomelo (*C. paradisi*) font partie des plus sensibles (Schubert *et al.* 2001). L'obtention de cultivars résistants par les techniques classiques d'hybridation et de sélection n'est pas aisé à cause de leur temps de génération long et de caractéristiques génomiques et biologiques telles que de la polyploïdie de certaines variétés (Davey *et al.* 2005). Le gène *csLOB1* est chez la plante la cible de *pthA4*, un effecteur TALE important dans le pouvoir pathogène de Xcc (voir IV.3.b). Des plantes transgéniques résistantes grâce à la mutation de *csLOB1* (Jia *et al.* 2016; Jia *et al.* 2017) ou de sa région promotrice (Peng *et al.* 2017), ont été obtenues, mais aucune application agronomique n'existe à ce jour. Les niveaux de résistance de cybrides, dont le noyau cellulaire provient d'une espèce et les organites cytoplasmiques d'une autre, sont actuellement en cours d'évaluation (Omar *et al.* 2017).

Objectifs généraux

Xanthomonas citri pv. *citri* est la bactérie responsable du chancre Asiatique des agrumes qui constitue une forte contrainte pour l'agrumiculture mondiale. L'utilisation de pesticides à base de cuivre représente une composante importante des stratégies intégrées utilisées pour lutter contre cette bactérie. À La Réunion, île du Sud-ouest de l'océan Indien, le chancre Asiatique a été décrit depuis les années 1960 mais a probablement préexisté pendant une période dont l'étendue demeure inconnue. À partir des années 2010, des niveaux de maladie élevés y ont été observés malgré l'application de produits cupriques et l'absence de conditions environnementales favorables au développement d'épidémies. De façon concomitante, un foyer épidémique résistant au cuivre a été détecté en Martinique (archipel des petites Antilles, Caraïbes), alors que cette île, ainsi que tous les territoires de l'arc Antillais, étaient jusqu'alors indemnes de maladie. Par PCR, les gènes de résistance au cuivre *copLAB* ont été détectés à la fois chez les souches réunionnaises et martiniquaises. Ces mêmes gènes avaient été précédemment décrits chez Xcc en Argentine, où la résistance au cuivre a été détectée à partir des années 1990, ainsi que chez d'autres espèces de *Xanthomonas* phytopathogènes. Le fonctionnement de ce cluster de gènes de résistance reste mal connu. Lors de sa description en Argentine, un support plasmidique a été suggéré mais n'a pas été caractérisé. Ces observations suggèrent la présence de liens épidémiologiques entre les souches d'Argentine, de La Réunion et de Martinique.

Afin de comprendre ces liens épidémiologiques, mais aussi pour mieux caractériser le support génétique des gènes de résistance au cuivre, nous avons entrepris une étude de diversité génétique sur une collection de souches de Xcc résistantes au cuivre provenant des trois régions où ce phénotype a été décrit (Chapitre I). À l'aide d'une technologie de séquençage de troisième génération, nous avons obtenu le génome complet de certaines de ces souches ainsi que de celui d'une souche commensale isolée de la phyllosphère d'agrumes, résistante au cuivre et taxonomiquement proche du genre *Xanthomonas*. Le génotypage de ces souches a permis de suggérer l'existence d'une relation épidémiologique entre les foyers de résistance au cuivre des deux territoires ultramarins Français. Au-delà de confirmer et de caractériser le support plasmidique de la résistance chez Xcc, nos analyses ont permis d'identifier l'inclusion des gènes causatifs au sein d'un transposon. La présence de ce transposon sur différents supports plasmidiques chez *Xanthomonas* ainsi que son intégration au chromosome de certaines

bactéries de la phyllosphère révèlent une dynamique évolutive profondément dépendante des transferts horizontaux.

Nous avons alors souhaité quantifier l'importance de ces mécanismes de transferts horizontaux relativement à l'évolution par mutation (Chapitre II). Nous avons caractérisé ces deux formes d'évolution génomique chez *Xcc*, de l'échelle de l'Océan Indien jusqu'à des échelles géographiques très restreintes telles que le verger. Le séquençage complet de 225 souches originaires de l'océan Indien et isolées de 1970 à 2017 a permis d'établir le taux de mutation de *Xcc* et de définir le contenu en gènes de chaque individu. La dynamique de dispersion du plasmide de résistance au cuivre à l'échelle de l'Océan Indien a également pu être appréciée grâce à un échantillonnage comprenant des souches sensibles et des souches résistantes au cuivre. Nous avons ainsi pu décrire de manière précise l'histoire évolutive de *Xcc*, ainsi que la prévalence de chacune des deux formes d'évolution, les SNP d'une part et les variations de contenu en gènes d'autre part, et ce à plusieurs échelles. En relation avec le caractère monomorphe et épidémique de *Xcc* ainsi que la biologie de cette bactérie, ces résultats fournissent un éclairage nouveau sur son évolution génomique. Nous avons montré que la variation de contenu en gènes est une forme d'évolution importante chez *Xcc*. Une meilleure caractérisation des organismes avec lesquels *Xcc*, et plus largement les bactéries phytopathogènes, peuvent échanger du matériel génétique pourrait en conséquence améliorer la compréhension de leur évolution.

Nous avons enfin analysé le partage de gènes plasmidiques encodés dans l'ensemble des génomes connus des bactéries de la famille des Xanthomonadaceae grâce à une approche basée sur la construction d'un réseau d'échanges de gènes. Cette méthode permet d'apprécier la structure d'un groupe d'individus en se basant uniquement sur les gènes plasmidiques qu'ils partagent. L'objectif était d'identifier le ou les facteurs structurant ce partage (Chapitre III). Nos résultats ont mis en évidence l'importance des échanges de gènes impliqués dans la résistance à des antibiotiques et à des métaux lourds et ont montré l'existence d'une barrière taxonomique nette. Les résultats de ces travaux ont souligné une nouvelle fois la nécessité de considérer les communautés de bactéries dans leur ensemble dans le cadre de la gestion des résistances.

Chapitre I Déterminisme génétique de la résistance au cuivre chez Xcc

Chapitre I Déterminisme génétique de la résistance au cuivre chez Xcc

Les espèces phytopathogènes du genre *Xanthomonas* présentent souvent une forte spécialisation d'hôte, liée à l'uniformité génétique et environnementale des systèmes agricoles contemporains (Jacques *et al.* 2016). Les répertoires d'effecteurs de type III ont un rôle central dans le pouvoir pathogène de nombreuses espèces de *Xanthomonas* et évoluent en réponse aux systèmes de défense des plantes. La durabilité de résistances monogéniques présentes chez le riz, la tomate ou le piment vis-à-vis des *Xanthomonas* respectivement responsables de la bactériose vasculaire du riz et de la gale bactérienne des solanées à graines s'est souvent avérée faible, mettant parfois en évidence l'importance de plasmides dans ces contournements (McDonald & Linde 2002). *Xanthomonas citri* pv. *citri* (Xcc) est une bactérie monomorphe qui possède une distribution mondiale et pour laquelle la lutte chimique repose depuis un siècle environ sur l'utilisation de pesticides à base de cuivre. En arboriculture comme en maraîchage, le cuivre est en effet fréquemment utilisé pour lutter contre *Xanthomonas* et constitue même la seule substance homologuée pour lutter contre certaines bactérioses (Giraud M *et al.* 2007).

L'île de la Réunion (Sud-Ouest de l'Océan Indien) comporte quelques 300 hectares dédiés à l'agrumiculture (DAAF 2014). Le chancre citrique y a un caractère endémique puisque Xcc y a été observé régulièrement depuis les années 1960 bien que la maladie ait vraisemblablement longuement précédé sa description (Brun, 1971). Les premières analyses ont été entreprises dans un verger de combava (*Citrus hystrix*) où les niveaux de maladie étaient exceptionnellement élevés au vu des conditions environnementales particulièrement défavorables au développement d'épidémies. En 2014, le système génétique de résistance au cuivre *copLAB* a été détecté par PCR sur des souches réunionnaises collectées à partir de 2010. Ce système avait été décrit comme associé au phénotype résistant lors de la première mise en évidence de résistance au cuivre chez Xcc en Argentine au début des années 1990 (Behlau *et al.* 2011 ; Canteros *et al.* 2008).

Parallèlement à cette détection, un foyer de chancre citrique a été détecté en Martinique (archipel des petites Antilles, Caraïbes) qui était jusqu'alors indemne de la maladie (**Note 1**, (Richard *et al.* 2016a)). Suite à ses constats, des échantillons des trois foyers de résistance au cuivre ont été analysés afin de déterminer s'il existe un lien épidémique entre ces souches.

I Emergence de souches de Xcc résistantes au cuivre sur deux territoires ultramarins Européens

L'analyse par minisatellites et microsatellites de souches issues de La Réunion (n=219, 1978-2017), de Martinique (n=20, 2014) et d'Argentine (n=111, 1994-2015) a révélé que les souches de Xcc résistantes au cuivre provenant des trois régions où la résistance au cuivre a été décrite (Argentine, Martinique et Réunion) appartiennent à la lignée génétique 1 du pathotype A, qui est la lignée prédominante mondialement (Pruvost *et al.* 2014). Ces souches se structurent en deux groupes génétiques fortement différenciés : l'un comprenant les souches d'Argentine et l'autre comprenant les souches de Martinique et de la Réunion. A la Réunion, nous avons observé une très forte proximité génétique entre des souches CuR et des souches CuS d'une part, et entre des souches CuS historiques isolées depuis la fin des années 1970 et des souches contemporaines CuR d'autre part. Aussi, des souches CuR et CuS cohabitent fréquemment au sein d'un même verger. La distribution de la résistance dans les différents bassins de production Réunionnais ainsi que la détection de souches CuR dans la principale pépinière de production de plants d'agrumes de la Réunion illustrent une dissémination humaine de Xcc. La distribution de plants contaminés par les pépinières locales est le principal mode de dissémination de Xcc sur de longues distances (Leduc *et al.* (2015); Verniere *et al.* (2014); **Note 2**, Richard *et al.* (2017d)). L'analyse de la diversité des souches Martiniquaises, toutes résistantes au cuivre, suggère quant à elle l'introduction sur l'île de souches possédant déjà ce phénotype.

Nos résultats, obtenus à l'aide de données microsatellites, soulèvent l'hypothèse (i) de l'acquisition du phénotype CuR par des souches de Xcc Réunionnaises par transfert horizontal plutôt que par établissement de nouvelles souches, et (ii) d'une relation épidémiologique entre les souches de la Réunion et de Martinique.

II Séquençage de souches de *Xanthomonadaceae* résistantes au cuivre

La présence de *copLAB* dans au moins trois systèmes épidémiques contrastés (Martinique, Réunion et Argentine) ainsi que son insertion dans un plasmide conjugatif pour au moins une souche Argentine (Behlau *et al.* 2012) suggèrent le caractère transmissible de ce phénotype. Des travaux antérieurs effectués sur plusieurs espèces de *Xanthomonas* phytopathogènes avaient montré la présence de plusieurs systèmes génétiques conférant un phénotype de résistance au cuivre dans ce genre bactérien. En effet, la présence indépendante de deux systèmes génétiques, *copABCD* et *copLAB* dans un même taxon a été décrite chez *X. arboricola* pv. *juglandis* (responsable de la bactériose du noyer) (Behlau *et al.* 2013; Cesbron *et al.* 2015; Giovanardi *et al.* 2016; Lee *et al.* 1994), et chez une espèce de *Xanthomonas* responsable de la gale bactérienne (nommée selon la nomenclature en vigueur à l'époque de la publication d'origine et non réassignée à un taxon valide aujourd'hui) de la tomate et du piment (Behlau *et al.* 2013; Cooksey *et al.* 1990). Le système *copLAB* a également été identifié dans plusieurs souches d'une espèce d'un genre proche (*Stenotrophomonas maltophilia*) responsables d'infections nosocomiales en milieu hospitalier (Crossman *et al.* 2008). Ces systèmes ont un support génétique variable puisqu'ils peuvent être plasmidiques, chromosomiques ou localisés sur des éléments mobiles comme les ICEs (pour « Integrative and Conjugative Elements ») (Cesbron *et al.* 2015).

A la lumière de cette variabilité, une approche par séquençage de génome complet a paru pertinente pour élucider sans a priori les bases génétiques de la résistance au cuivre chez Xcc. Certaines caractéristiques des technologies de séquençage « *long-read* » les rendent particulièrement adaptées à la reconstruction de plasmides complets. Nous avons ainsi séquencé six souches de Xcc résistantes au cuivre avec la technologie long-reads PacBio (**Annonce de génome 1**, Richard *et al.* (2017b)). Il a semblé pertinent de procéder également au séquençage de représentants des trois espèces de *Xanthomonas* responsables de la gale bactérienne de la tomate et du piment où cette résistance a été initialement décrite et mise en évidence de façon répétée (**Annonce de génome 2**, Richard *et al.* (2017a)), ainsi que d'une souche résistante du genre proche *Stenotrophomonas* isolée de la phyllosphère d'agrumes (**Annonce de génome 3**, Richard *et al.* (2016b)).

III Caractérisation des supports génomiques de la résistance au cuivre

L'obtention de données génomiques de ces 13 souches de Xanthomonadaceae qui présentent un phénotype de résistance au cuivre nous a permis d'aborder des questions de recherche liées à l'adaptation de Xcc (**Publication 1**, Richard *et al.* (2017c)). L'analyse de ces séquences a mis en évidence trois systèmes de gènes possiblement impliqués dans le phénotype résistance au cuivre. Ces derniers sont présents en deux configurations : *copLAB* en association avec *cusABC* et *copABCD* en association avec *cusABC*, *copABCD* étant jusque-là encore inconnu chez Xcc. Les trois systèmes sont toujours encodés au sein d'un transposon proche du type Tn3, lui-même présent soit sur le chromosome chez la souche commensale de *Stenotrophomonas*, soit sur un plasmide (*Xanthomonas*). Le squelette plasmidique dans lequel est intégré Tn3_{pLH201.1} peut également être variable, comme nous l'avons mis en évidence pour une souche de *Xanthomonas euvesicatoria*. Au sein du transposon Tn3_{pLH201.1}, les gènes *cop* sont associés à d'autres gènes à caractère adaptatif. En effet, nous avons mis en évidence plusieurs autres clusters de gènes potentiellement impliqués dans la résistance à des métaux lourds : *Czc* pour Cuivre Zinc Cadmium, *Cus* pour cuivre, *Ars* pour l'arsenic. La séquence codant pour la transposase du transposon Tn3 est très conservée entre les souches mais la structure du transposon est très variable, avec de nombreux réarrangements et insertions-délétions de gènes entre les souches.

Dans le but de mettre en évidence l'éventuelle existence de systèmes non connus et impliqués dans la résistance au cuivre chez Xcc à la Réunion, nous avons phénotypé une large collection de souches positives et négatives en PCR *copLAB* pour déterminer si ces dernières présentent un phénotype sensible ou résistant. Associée à des PCR *copABCD* et aux résultats des PCR *copLAB* déjà disponibles, la réalisation de tests basés directement sur le phénotype n'a pas mis en évidence de système de résistance autre de *copLAB* chez Xcc à la Réunion (détails dans la Note 2).

IV Phénotypes associés aux plasmides de CuR chez Xcc

Nous avons montré que le transposon plasmidique qui encode pour les gènes de résistance au cuivre encode également d'autres clusters de gènes de résistance à des métaux lourds. L'annotation de ces gènes est toutefois parfois incertaine puisque associée à de faibles homologies avec des systèmes connus et validés expérimentalement. Le phénotype de résistance aux différents métaux lourds visés par les différents systèmes de résistance détectés a donc été caractérisé expérimentalement en cultivant toutes les souches sur des milieux enrichis en métal lourd (détails dans l'Annonce de génome 1 et en Annexe II). Si les phénotypes de résistance à l'arsenic semblent corrélés à la présence du système Ars dans les génomes analysés, ce n'est pas le cas pour le système Czc vis-à-vis de la résistance au zinc, cadmium et cobalt.

Nos résultats d'analyse *in silico* ont suggéré une dispersion taxonomique de pLH201.1 globalement limitée au genre *Xanthomonas*. Cependant, le faible nombre de génomes de Xanthomonadaceae n'appartenant pas au genre *Xanthomonas* disponibles sur les bases de données de NCBI a pu induire un biais d'analyse. La confirmation de ces résultats obtenus *in silico* a donc été faite de façon expérimentale, en testant la capacité du plasmide à se transmettre par conjugaison. Nous avons ainsi évalué le taux de conjugaison de pLH201.1 entre Xcc et plusieurs souches présentant un éloignement taxonomique variable avec Xcc LH201 : (i) différentes souches de *Xanthomonas* (comprenant Xcc, d'autres pathovars de l'espèce *X. citri* ainsi que d'autres espèces) et (ii) d'autres genres de Xanthomonadaceae.

Nos résultats ont validé le caractère fonctionnel de l'appareil conjugatif encodé par pLH201.1 ainsi que la transférabilité du phénotype de résistance au cuivre depuis une souche CuR vers une souche CuS. Nous avons également mis en évidence le transfert de pLH201.1 dans différentes espèces de *Xanthomonas*, comprenant des bactéries commensales isolées de la phyllosphère d'agrumes mais aussi des espèces pathogènes provenant de niches écologiques diverses (incluant les agrumes). Une limite taxonomique assez forte au transfert de pLH201.1 par conjugaison semble en outre exister, puisque nous n'avons pas observé de transfert dans des souches non-*Xanthomonas* de la famille des Xanthomonadaceae. De la même manière que lors de l'étude *in silico*, la faible disponibilité en ressources bactériennes sensibles au cuivre dans cette famille a limité le nombre de souches que nous avons pu tester. La croissance fastidieuse de certaines souches ou la difficulté d'obtenir des mutants pour la contre-sélection des souches donneuses ont également représenté un frein important à l'expérimentation.

L'absence de transfert de pLH201.1 dans certaines souches présentant un éloignement taxonomique à Xcc LH201 identique à celui d'autres souches ayant intégré le plasmide (mêmes taxons) est surprenant (Annexe III). Ce résultat contrasté pourrait provenir de l'incompatibilité entre pLH201.1 et des plasmides de ces souches réfractaires à la conjugaison.

V Conclusions du chapitre

Nous avons montré à la fois par analyse génomique, par analyse de réseau d'échanges de gènes, et par approche expérimentale (conjugaison *in vitro*) que des transferts horizontaux de gènes directs ou indirects permettent possiblement de connecter plusieurs niches écologiques. En effet, des plasmides proches de pLH201.1 ou le transposon de résistance au cuivre porté par pLH201.1 ont été retrouvés chez des bactéries environnementales, associées au Solanées ou encore associées aux agrumes. Dans le même réseau d'échanges de gènes, la faible dispersion taxonomique des gènes du plasmide de résistance au cuivre suggère la présence d'une barrière taxonomique au transfert de pLH201.1 par conjugaison. Ces résultats ont été confirmés par nos tests de conjugaison *in vitro*. Nous avons enfin montré que le transfert par conjugaison de pLH201.1 est souche-dépendant au sein d'un même pathovar, et avéré uniquement au sein du genre *Xanthomonas*, qu'il s'agisse de souches pathogènes des agrumes ou d'autres familles de plantes, ou de souches commensales isolées de la phyllosphère d'agrumes. Dans le cas de la résistance au cuivre, l'adaptation de Xcc repose fortement sur un événement de transfert horizontal. Après avoir abordé l'évolution de Xcc à travers l'acquisition d'un phénotype particulier, nous avons souhaité évaluer l'importance de cette forme d'évolution sans a priori, c'est-à-dire en analysant le contenu en gènes de souches épidémiques de Xcc sans restreindre l'étude aux gènes à caractère adaptatif.

First Report of *Xanthomonas citri* pv. *citri* Pathotype A Causing Asiatic Citrus Canker in Martinique, France

September 2016, Volume 100, Number 9

Page 1946

<https://doi.org/10.1094/PDIS-02-16-0170-PDN>

D. Richard, C. Boyer, S. Javegny, K. Boyer, P. Grygiel and O. Pruvost, CIRAD UMR PVBMT, F-97410 Saint-Pierre, La Réunion, France; **A.L. Rioualec, V. Rakotobe and J. Iotti**, DAAF Service de l'Alimentation, F-97264 Fort de France, Martinique; **R. Picard**, FREDON, F-97224 Ducos, Martinique; **C. Vernière**, CIRAD UMR BGPI, F-34398 Montpellier Cedex 5, France; **C. Audusseau, C. François and V. Olivier**, ANSES Laboratoire de la Santé des Végétaux, F-49044 Angers, France; **A. Moreau and A. Chabirand**, ANSES Laboratoire de Santé des Végétaux, F-97410 Saint-Pierre, La Réunion, France.

Asiatic canker, caused by *Xanthomonas citri* pv. *citri*, is a major threat to worldwide citriculture. Three pathotypes differing in host range and hypersensitive reactions towards citrus species have been defined. Whereas pathotypes A^w and A^{*} have a restricted host range, *X. citri* pv. *citri* pathotype A infects a broader range including most commercial citrus species and hybrids and can cause important economic losses in tropical and sub-tropical areas. Pathotype A strains, especially those assigned to lineage 1, were implicated in the major geographical expansion of *X. citri* pv. *citri* during the 20th century from their native area, Asia (Pruvost et al. 2014). *X. citri* pv. *citri* is listed as a quarantine pathogen in the European Union (EU) – Directive 2000/29/EC annex II A1.

Martinique (France) is an outermost region of the EU in the eastern Caribbean Sea. Canker lesions were first observed at Morne Rouge, Martinique in June 2014 on grapefruit (*Citrus paradisi*), mandarin (*C. reticulata*), Tahiti lime (*C. latifolia*), Valencia and Washington Navel oranges (*C. sinensis*). Official diagnostics, including bacterial isolations on YPGA or KC semi-selective medium (Pruvost et al. 2005), PCR-based identification with 4/7 primers (Hartung et al. 1996) and pathogenicity tests, were performed following the EPPO standard PM7/44 (<http://www.eppo.int/>) and identified isolates as *X. citri* pv. *citri*. Three strains isolated in Martinique in 2014 from grapefruit or Tahiti lime were further characterized (LL074-4, LL077-2 and LL079).

Multilocus sequence analysis (MLSA) targeting six housekeeping genes (*atpD*, *dnaK*, *efp*, *gltA*, *gyrB*, and *lepA*) (Almeida et al. 2010; Bui Thi Ngoc et al. 2010) identified Martinique strains as *X. citri* pv. *citri* with 100% sequence identity to the type strain LMG 9322. Using MLVA-31 targeting 31 minisatellites, Martinique strains were assigned to lineage 1

composed of pathotype A strains (Pruvost et al. 2014). All strains were inoculated by a detached leaf assay onto Mexican lime SRA 140 (*C. aurantifolia*), sweet orange Washington Navel SRA 102 and grapefruit Henderson SRA 336 (Bui Thi Ngoc et al. 2010). All inoculated leaves produced typical erumpent, callus-like tissue at wound sites. *Xanthomonas*-like colonies were re-isolated from lesions that had developed. Boiled suspensions were assayed by PCR with 4/7 primers and produced the expected amplicon, fulfilling Koch's postulates. No lesions developed on the negative control consisting of sterile 0.01M tris buffer pH 7.2. This is the first report of *X. citri* pv. *citri* in Martinique and to our knowledge in the Caribbean region. Surprisingly, all strains collected to date in Martinique grew on YPGA supplemented with 300 mg L⁻¹ copper sulfate even when no extensive copper spray programs have been used, suggesting that copper-resistant strains may have been introduced. Disease is contained by tree removal and burning and the situation is presently under apparent control although positive trees were sporadically detected in 2015 in backyards or small orchards at Le Lorrain and Saint Pierre. An extensive surveillance program is currently implemented in Martinique for quarantine pathogens of citrus.

References

- (1) N. F. Almeida et al. Phytopathology 100:208, 2010. (2) L. Bui Thi Ngoc et al. Int. J. Syst. Evol. Microbiol. 60:515, 2010. (3) J. S. Hartung et al. Appl. Environ. Microbiol. 59:1143,1993. (4) O. Pruvost et al. J. Appl. Microbiol. 99:803, 2005. (5) O. Pruvost et al. PLoS ONE 9(6): e98129. 2014.

First Report of Copper-resistant *Xanthomonas citri* pv. *citri* Pathotype A Causing Asiatic Citrus Canker in Réunion, France

March 2017, Volume 101, Number 3

Page 503

<http://dx.doi.org/10.1094/PDIS-09-16-1387-PDN>

D. Richard CIRAD - Anses, Plant Health Laboratory, Tropical Pests and Diseases Unit - Université de la Réunion, UMR PVBMT, F-97410 Saint-Pierre, Reunion Island, France; **N. Tribot, C. Boyer, M. Terville, K. Boyer, S. Javegny, M. Roux-Cuvelier** and **O. Pruvost**, CIRAD UMR PVBMT, F-97410 Saint-Pierre, La Réunion, France; **A. Moreau** and **A. Chabirand**, ANSES Laboratoire de Santé des Végétaux, F-97410 Saint-Pierre, La Réunion, France; **C. Vernière**, CIRAD UMR BGPI, F-34398 Montpellier Cedex 5, France.

Xanthomonas citri pv. *citri* causes Asiatic citrus canker (ACC), which induces erumpent, callus-like lesions on all aerial organs, and consequently defoliation, premature fruit drop, and twig dieback under high inoculum pressure. Present in most tropical and subtropical citrus producing regions, the disease can lead to important decreases in yield and limitations to export markets due to the quarantine status of the pathogen in some countries (Graham et al., 2004). Strains assigned to genetic lineage 1 within pathotype A, which has the widest host range comprising most citrus cultivars and citrus relatives, were shown to be primarily involved in the worldwide expansion of the bacterium over the 20th century (Pruvost et al. 2014).

ACC has been considered endemic in Réunion Island since the 1970s, but was likely present earlier and was not subjected to any eradication attempt. Most citrus growers on the island frequently apply copper-based pesticides as part of integrated pest management programs. A recurring lack of ACC control in several citrus groves was investigated in 2014 using our lab collection of authenticated *X. citri* pv. *citri* strains. 501 strains from active epidemics as well as 123 historical strains isolated from 1978 onward were tested for copper resistance using PCR assays targeting *copL* (Behlau et al., 2013). 115 strains produced the expected amplicon whereas others (including all historical strains) and the negative control (H₂O) did not. Phenotypic tests were carried out on low-complexing mineral salts casitone-yeast extract-glycerol-agar (CYE) plates (Zevenhuizen et al., 1979) flooded with bacterial suspensions ($\approx 5 \times 10^6$ cfu ml⁻¹) in sterile tris buffer (pH 7.2), allowed to dry for 15 min., inoculated with a sterile CuSO₄, 5H₂O solution (10mg ml⁻¹) with an easySpiral[®] plater used to create a gradient of log-increasing copper concentrations towards the plate center and incubated for two days at 28°C. Using this assay, *copL*⁺ strains grew over a radius (measured from the plate periphery

where the concentration of copper are lowest) > 23 mm whereas *copL*⁻ strains growth radius were constantly lower. Copper-resistant strains (CRS) had minimum inhibitory concentrations for copper sulfate ranging from 128 to 256 $\mu\text{l ml}^{-1}$ (Pruvost et al., 1998). A subset of 20 CRS and 40 copper-susceptible strains (CSS) (including historical strains) were all assigned to genetic lineage 1 by minisatellite typing (performed as reported by Pruvost et al. 2014) and split into two multilocus haplotypes shared also with CSS. CRS were widespread in most citrus production areas of the island where they coexisted with CSS. The oldest CRS dated from 2010. Because no extreme hurricane hit the island between 2010 and 2014, the fast spread of CRS over the island was very likely related to their presence in a major citrus nursery, as confirmed by our data.

In Réunion, the European *Conformitas Agraria Communitatis* (CAC) standards (directive 92/34/EEC) certify the disease-free status of citrus nursery plants (assessed by visual inspections for ACC). CAC is presently recommended but not compulsory. This report suggests that a strict implementation of this procedure would have helped limiting the spread of CRS in Réunion.

References:

- Behlau, F.**, et al. 2013. *Phytopathology* 103:409-418.
- Graham, J. H.**, et al., 2004. *Mol. Plant Pathol.* 5:1-15.
- Pruvost, O.**, et al. 1998. *J. Appl. Microbiol.* 84:15-24.
- Pruvost, O.**, et al., 2014. *PLoS ONE* 9: e98129.
- Zevenhuizen, L.P.T.M.**, et al., 1979. *Microb. Ecol.* 5:139-164.



Complete Genome Sequences of Six Copper-Resistant *Xanthomonas citri* pv. *citri* Strains Causing Asiatic Citrus Canker, Obtained Using Long-Read Technology

Damien Richard,^{a,b,c} Claudine Boyer,^a Christian Vernière,^{a,d} Blanca I. Canteros,^e Pierre Lefeuvre,^a Olivier Pruvost^a

Cirad, UMR PVBMT, St.-Pierre, Réunion, France^a; ANSES, Plant Health Laboratory, St.-Pierre, Réunion, France^b; Université de la Réunion, UMR PVBMT, St.-Denis, Réunion, France^c; Cirad, UMR BGPI, Montpellier, France^d; INTA, Estación Experimental Agropecuaria Bella Vista, Buenos Aires, Argentina^e

ABSTRACT The gammaproteobacterium *Xanthomonas citri* pv. *citri* causes Asiatic citrus canker. Pathotype A strains have a broad host range, which includes most commercial citrus species, and they cause important economic losses worldwide. Control often relies on frequent copper sprays. We present here the complete genomes of six *X. citri* pv. *citri* copper-resistant strains.

Asiatic citrus canker, caused by *Xanthomonas citri* pv. *citri*, causes important economic losses in most tropical and subtropical areas. Its development impacts crop yields and also limits exports because of the quarantine status of the pathogen in some countries. The pathogen is currently classified into three pathotypes. Whereas pathotypes A^w and A^{*} have a restricted host range and geographical distribution, pathotype A has a wider host range, which has resulted in major geographical expansion of the pathogen (1). In the early stages of emergence, eradication is feasible through tree removal, burning, and quarantine restrictions, whereas integrated pest management (IPM) is conducted after the disease has spread in an area. IPM typically includes repetitive copper applications. Copper-resistant (Cu^R) strains were first described in 1994 in Argentina (2), and 20 years later in the French islands of Réunion (3) and Martinique (4). To investigate the genetic basis of this resistance, six Cu^R strains were sequenced using long-read PacBio RSII technology with one single-molecule real-time (SMRT) cell per strain. Three strains originated from Réunion (LH201, LH276, and LJ207-7), two from Argentina (LM199 and LM180 = A44), and one from Martinique (LL074-4). Minisatellite typing revealed that all Cu^R strains were assigned to genetic lineage 1 within pathotype A (1).

The *de novo* genome assembly was conducted using the SMRT Analysis HGAP version 2.3 protocol. Circularization of the contigs was attempted using a combination of the Minimus assembler (5) and the SMRT Analysis resequencing version 1 protocol. The protein-coding sequences of the plasmid that confers Cu^R to LH201 were predicted using the MaGe genome annotation platform. In order to improve the assembly of transcription activator-like genes, additional assembly steps were performed in a manner similar to that as previously described (6).

The chromosomes of the six strains were fully reconstructed, with sizes ranging from 5,148,274 to 5,198,476 bp. Depending on the strains, two to four additional contigs were circularized into plasmids, with sizes ranging from 28,949 to 249,697 bp. All six strains hosted a plasmid carrying Cu^R genetic determinants. These plasmids were highly conserved among strains (despite remote geographical origins), apart from LM199.

Received 4 January 2017 Accepted 25 January 2017 Published 23 March 2017

Citation Richard D, Boyer C, Vernière C, Canteros BI, Lefeuvre P, Pruvost O. 2017. Complete genome sequences of six copper-resistant *Xanthomonas citri* pv. *citri* strains causing Asiatic citrus canker, obtained using long-read technology. Genome Announc 5: e00010-17. <https://doi.org/10.1128/genomeA.00010-17>.

Copyright © 2017 Richard et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Olivier Pruvost, olivier.pruvost@cirad.fr.

TABLE 1 Characteristics of six *Xanthomonas citri* pv. citri strains

Strain	Other number(s)	Accession no.	Copper resistance location	Place of isolation	Yr of isolation	Host
LM199	<i>X. citri</i> pv. citri 15-4632 (INTA) ^a	MSQV00000000	Plasmid	Argentina	2015	Orange
LM180	<i>X. citri</i> pv. citri 03-1638 (INTA); A44 ^b	MSQW00000000	Plasmid	Argentina	2003	Grapefruit
LL074-4		CP018847 to CP018849	Plasmid	Martinique	2014	Grapefruit
LJ207-7		CP018850 to CP018853	Plasmid	Réunion	2012	Kaffir lime
LH276		CP018854 to CP018857	Plasmid	Réunion	2010	Kaffir lime
LH201		CP018858 to CP018860	Plasmid	Réunion	2010	Kaffir lime

^aINTA, Instituto Nacional de Tecnología Agropecuaria.

^bAs designated by Behlau et al. (7).

Whereas the other five strains hosted the *copLAB* gene system previously described for *Xanthomonas citri* pv. citri (7), LM199 encoded *copABCD*, which has already been described on the chromosome of *X. arboricola* pv. juglandis (8). Additional putative resistance gene clusters were found on the Cu^R plasmid: *cusAB/smmD*, *czcABCD*, and *arsBHC*. However, LM199 lacked *czcABCD* and *arsBHC*. MICs were evaluated as previously described (9): zinc chloride, 32 mg/L; copper sulfate, 128 to 256 mg/L; sodium arsenite, 8 to 16 mg/L (LM199), 64 mg/L (LH201), 128 mg/L (LL074-4), >128 mg/L (LH276, LM180, and LJ207-7); cadmium sulfate, 6.4 to 12.8 mg/L; and cobalt chloride 16 to 32 mg/L. Because strain LM199 (which did not contain the putative *czcABCD* region) and other sequenced strains (containing *czcABCD*) shared similar MIC values for cadmium sulfate, zinc chloride, and cobalt chloride, the function of this efflux pump remains unclear. These sequences highlight the potential of microbial adaptation through horizontal gene transfer and the mosaic structure of the genetic elements carrying adaptive genes.

Accession number(s). The six closed genome sequences have been deposited in GenBank. The genome accession numbers are summarized in Table 1.

ACKNOWLEDGMENTS

The European Regional Development Fund (ERDF), Région Réunion, the French Agropolis Foundation (Labex Agro-Montpellier, E-SPACE project number 1504-004), Anses, and Cirad provided financial support.

REFERENCES

- Pruvost O, Magne M, Boyer K, Leduc A, Tourterel C, Drevet C, Ravigné V, Gagnevin L, Guérin F, Chiroleu F, Koebernik R, Verdier V, Vernière C. 2014. A MLVA genotyping scheme for global surveillance of the citrus pathogen *Xanthomonas citri* pv. citri suggests a worldwide geographical expansion of a single genetic lineage. *PLoS One* 9:e98129. <https://doi.org/10.1371/journal.pone.0098129>.
- Canteros BI, Rybak M, Gochez A, Velazquez P, Rivadeneira M, Mitidieri M, Garran S, Zequeira L. 2008. Occurrence of copper resistance in *Xanthomonas axonopodis* pv. citri in Argentina. *Phytopathology* 98:S30–S31.
- Richard D, Tribot N, Boyer C, Terville M, Boyer K, Javegny S, Roux-Cuvelier M, Pruvost O, Moreau A, Chabirand A, Vernière C. 2017. First report of copper-resistant *Xanthomonas citri* pv. citri pathotype A causing Asiatic citrus canker in Réunion, France. *Plant Dis* [Epub ahead of print.] <https://doi.org/10.1094/PDIS-09-16-1387-PDN>.
- Richard D, Boyer C, Javegny S, Boyer K, Grygiel P, Pruvost O, Rioualec AL, Rakotobe V, Iotti J, Picard R, Vernière C, Audusseau C, François C, Olivier V, Moreau A, Chabirand A. 2016. First report of *Xanthomonas citri* pv. citri pathotype A causing Asiatic citrus canker in Martinique, France. *Plant Dis* 100:1946. <https://doi.org/10.1094/PDIS-02-16-0170-PDN>.
- Sommer DD, Delcher AL, Salzberg SL, Pop M. 2007. Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics* 8:64. <https://doi.org/10.1186/1471-2105-8-64>.
- Booher NJ, Carpenter SC, Sebra RP, Wang L, Salzberg SL, Leach JE, Bogdanove AJ. 2015. Single molecule real-time sequencing of *Xanthomonas oryzae* genomes reveals a dynamic structure and complex TAL (transcription activator-like) effector gene relationships. *Microb Genom* 1:1–22. <https://doi.org/10.1099/mgen.0.000032>.
- Behlau F, Canteros BI, Minsavage GV, Jones JB, Graham JH. 2011. Molecular characterization of copper resistance genes from *Xanthomonas citri* subsp. citri and *Xanthomonas alfalfae* subsp. citrumelonis. *Appl Environ Microbiol* 77:4089–4096. <https://doi.org/10.1128/AEM.03043-10>.
- Pereira UP, Gouran H, Nascimento R, Adaskaveg JE, Goulart LR, Dandekar AM. 2015. Complete genome sequence of *Xanthomonas arboricola* pv. juglandis 417, a copper-resistant strain isolated from *Juglans regia* L. *Genome Announc* 3(5):e01126-15. <https://doi.org/10.1128/genomeA.01126-15>.
- Pruvost O, Couteau A, Perrier X, Luisetti J. 1998. Phenotypic diversity of *Xanthomonas* sp. *mangiferaeindicae*. *J Appl Microbiol* 84:115–124. <https://doi.org/10.1046/j.1365-2672.1997.00328.x>.



Complete Genome Sequences of Six Copper-Resistant *Xanthomonas* Strains Causing Bacterial Spot of Solaneous Plants, Belonging to *X. gardneri*, *X. euvesicatoria*, and *X. vesicatoria*, Using Long-Read Technology

Damien Richard,^{a,b,c} Claudine Boyer,^a Pierre Lefevre,^a Blanca I. Canteros,^d Shyam Beni-Madhu,^e Perrine Portier,^f Olivier Pruvost^a

CIRAD, UMR PVBMT, St.-Pierre, Réunion, France^a; ANSES, Plant Health Laboratory, St.-Pierre, Réunion, France^b; Université de la Réunion, UMR PVBMT, St.-Denis, Réunion, France^c; INTA, Estación Experimental Agropecuaria Bella Vista, Buenos Aires, Argentina^d; FAREI, Le Réduit, Mauritius^e; INRA, UMR1345 IRHS Institut de Recherche en Horticulture et Semences, Beaucauzé, France^f

ABSTRACT *Xanthomonas vesicatoria*, *Xanthomonas euvesicatoria*, and *Xanthomonas gardneri* cause bacterial spot disease. Copper has been applied since the 1920s as part of integrated management programs. The first copper-resistant strains were reported some decades later. Here, we fully sequenced six *Xanthomonas* strains pathogenic to tomato and/or pepper and having a copper-resistant phenotype.

Bacterial spot disease causes yield loss and impairs fruit quality in most tomato- and pepper-producing areas. The disease is caused by several distinct *Xanthomonas* species: *X. vesicatoria*, *X. euvesicatoria*, *X. gardneri*, and *X. perforans*, with *X. perforans* being recently reclassified as *X. euvesicatoria* (1). These phylogenetically distant species display similar symptoms and host range (2). Tomato (*Lycopersicon esculentum*), sweet pepper (*Capsicum annuum*), and chili pepper (*Capsicum frutescens*) are the main natural host species. The large-scale spread of the pathogens occurs through contaminated seeds, requiring strict disease management strategies to be imposed. A research effort is being made on biocontrol strategies, such as bacteriophage use, because resistant cultivars are scarce and rapidly overcome, and bacterial resistance to antimicrobial compounds is often reported. For example, widely used on pepper since the 1920s in the United States (3), a failure of copper sprays to control bacterial spot disease was first observed in 1968 (4), and the first associated genetic determinant (i.e., a transmissible plasmid) was reported more than 20 years later (2, 5).

In order to obtain a better understanding of copper resistance determinants and their spread among xanthomonads, we sequenced six tomato and pepper pathogen strains displaying a copper-resistant phenotype: *X. euvesicatoria* LMG930 and LH3, *X. vesicatoria* LMG911 and LM159, and *X. gardneri* JS749-3 and ICMP7383.

The long-read PacBio RSII technology was used to fully sequence the six strains, using one single-molecule real-time (SMRT) cell for each strain. Assembly of the raw reads was then performed using a SMRT Analysis HGAP version 2.3 protocol, and circularization of the contigs was done using a combination of the Minimus assembler (6) and the SMRT Analysis resequencing version 1 protocol. Assembly of transcription activator-like genes was improved using a custom version of the method previously described (7).

We obtained six closed chromosomes with sizes ranging from 4,969,893 bp to 5,313,102 bp, along with two to four closed plasmids per strain (31,328 bp to

Received 15 December 2016 **Accepted** 23 December 2016 **Published** 23 February 2017

Citation Richard D, Boyer C, Lefevre P, Canteros BI, Beni-Madhu S, Portier P, Pruvost O. 2017. Complete genome sequences of six copper-resistant *Xanthomonas* strains causing bacterial spot of solaneous plants, belonging to *X. gardneri*, *X. euvesicatoria*, and *X. vesicatoria*, using long-read technology. Genome Announc 5:e01693-16. <https://doi.org/10.1128/genomeA.01693-16>.

Copyright © 2017 Richard et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Olivier Pruvost, olivier.pruvost@cirad.fr.

TABLE 1 Characteristics of the six sequenced *Xanthomonas* strains^a

Strain	Other no.	Species	Copper resistance location	Country of isolation	Yr of isolation	Host	Chromosome size (bp)	G+C content (%)	ANI score (%) ^b	Accession no.
LMG930		<i>X. euvesicatoria</i>	Plasmid	United States	1969	Pepper	5,079,107	64.7	99.98	CP018463 to CP018467
LMG911	CFBP2537	<i>X. vesicatoria</i>	Plasmid	New Zealand	1955	Tomato	5,110,163	64.3	99.95	CP018725 to CP018727
LM159	Bv-5-4a (INTA)	<i>X. vesicatoria</i>	Plasmid	Argentina	1987	Pepper	5,086,726	64.2	99.88	CP018468 to CP018471
LH3	Xv5 (FAREI), CFBP7993	<i>X. euvesicatoria</i> ^c	Plasmid	Mauritius	2010	Tomato	4,969,893	64.9	98.68 (99.98)	CP018472 to CP018476
JS749-3		<i>X. gardneri</i>	Plasmid	Réunion	1997	Tomato	5,158,913	63.7	99.91	CP018728 to CP018730
ICMP7383	CFBP7999	<i>X. gardneri</i>	Plasmid	New Zealand	1980	Tomato	5,313,102	63.5	98.29	CP018731 to CP018734

^aLMG strains are from the Belgian Coordinated Collections of Microorganisms (BCCM/LMG), University of Ghent, Belgium. ICMP, International Collection of Microorganisms from Plants, Landcare Research, Auckland, New Zealand; INTA, Instituto Nacional de Tecnología Agropecuaria; FAREI, Food and Agricultural Research and Extension Institute; CFBP, Collection Française de Bactéries Associées aux Plantes.

^bANI scores were calculated using type strains of each species: LMG27970 for *X. euvesicatoria*; ATCC 35937 for *X. vesicatoria*; and ATCC 19865 for *X. gardneri*. Value in parentheses is for the type strain of *X. perforans* (CFBP7293).

^cSynonym of *X. perforans*.

222,061 bp) and uncircularized contigs. Average nucleotide identity (ANI) scores with type strains were used to confirm species assignment (Table 1). The six strains possessed the previously described copper resistance gene system *copLAB* (8) carried on a conjugative plasmid. Interestingly, the plasmid was very well conserved among five strains (dating from 1955 to 2010), while the other (LMG930) had no similarities apart from the copper resistance gene cluster. On five strains, this plasmid also comprised the *cusAB-smmD*, *czcABCD*, and *arsBHCR* resistance gene cluster. LH3 solely lacked *arsBHCR*. Evaluation of the minimum inhibitory concentrations was achieved on casitone yeast extract glycerol (CYE) medium, as previously reported (9, 10): these were zinc chloride, 16 to 32 mg/liter; copper sulfate, 128 to 256 mg/liter; sodium arsenite, 16 mg/liter (LH3), 128 to 512 mg/liter (other strains); cadmium sulfate, 6.4 to 12.8 mg/liter; and cobalt chloride, 16 to 32 mg/liter.

The very high similarity of the plasmids hosting the copper resistance adaptive trait among four species pathogenic to solaneous plants suggests pervasive events of horizontal gene transfer at the niche level.

Accession number(s). The six closed genome sequences have been deposited at GenBank. The genome accession numbers are shown in Table 1.

ACKNOWLEDGMENTS

Géraldine Taghouti and Marion Le Saux (INRA) provided part of the data.

REFERENCES

- Constantin EC, Cleenwerck I, Maes M, Baeyen S, Van Malderghem C, De Vos P, Cottyn B. 2016. Genetic characterization of strains named as *Xanthomonas axonopodis* pv. *dieffenbachiae* leads to a taxonomic revision of the *X. axonopodis* species complex. *Plant Pathol* 65:792–806. <https://doi.org/10.1111/ppa.12461>.
- Potnis N, Timilsina S, Strayer A, Shantharaj D, Barak JD, Paret ML, Vallad GE, Jones JB. 2015. Bacterial spot of tomato and pepper: diverse *Xanthomonas* species with a wide variety of virulence factors posing a worldwide challenge. *Mol Plant Pathol* 16:907–920. <https://doi.org/10.1111/mpp.12244>.
- Higgins BB. 1922. The bacterial spot of pepper. *Phytopathology* 12: 501–516.
- Marco GM, Stall R. 1983. Control of bacterial spot of pepper initiated by strains of *Xanthomonas campestris* pv. *vesicatoria* that differ in sensitivity to copper. *Plant Dis* 67:779–781. <https://doi.org/10.1094/PD-67-779>.
- Stall RE, Loschke DC, Jones JB. 1986. Linkage of copper resistance and avirulence loci on a self-transmissible plasmid in *Xanthomonas campestris* pv. *vesicatoria*. *Phytopathology* 76:240–243. <https://doi.org/10.1094/Phyto-76-240>.
- Sommer DD, Delcher AL, Salzberg SL, Pop M. 2007. Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics* 8:64. <https://doi.org/10.1186/1471-2105-8-64>.
- Booher NJ, Carpenter SC, Sebra RP, Wang L, Salzberg SL, Leach JE, Bogdanove AJ. 2015. Single molecule real-time sequencing of *Xanthomonas oryzae* genomes reveals a dynamic structure and complex TAL (transcription activator-like) effector gene relationships. *Microb Genom* 1. <https://doi.org/10.1099/mgen.0.000032>.
- Behlau F, Canteros BI, Minsavage GV, Jones JB, Graham JH. 2011. Molecular characterization of copper resistance genes from *Xanthomonas citri* subsp. *citri* and *Xanthomonas alfalfae* subsp. *citrumelonis*. *Appl Environ Microbiol* 77:4089–4096. <https://doi.org/10.1128/AEM.03043-10>.
- Zevenhuizen LPTM, Dolfing J, Eshuis EJ, Scholten-Koerselman IJ. 1979. Inhibitory effects of copper on bacteria related to the free ion concentration. *Microb Ecol* 5:139–146. <https://doi.org/10.1007/BF02010505>.
- Pruvost O, Couteau A, Perrier X, Luisetti J. 1998. Phenotypic diversity of *Xanthomonas* sp. *mangiferaeindicae*. *J Appl Microbiol* 84:115–124. <https://doi.org/10.1046/j.1365-2672.1997.00328.x>.



Complete Genome Sequence of a Copper-Resistant Bacterium from the Citrus Phyllosphere, *Stenotrophomonas* sp. Strain LM091, Obtained Using Long-Read Technology

Damien Richard,^{a,b,c} Claudine Boyer,^a Pierre Lefeuvre,^a Olivier Pruvost^a

Cirad, UMR PVBMT, St.-Pierre, Réunion, France^a; ANSES, Plant Health Laboratory, St.-Pierre, Réunion, France^b; Université de la Réunion, UMR PVBMT, St.-Denis, Réunion, France^c

The *Stenotrophomonas* genus shows great adaptive potential including resistance to multiple antimicrobials, opportunistic pathogenicity, and production of numerous secondary metabolites. Using long-read technology, we report the sequence of a plant-associated *Stenotrophomonas* strain originating from the citrus phyllosphere that displays a copper resistance phenotype.

Received 13 October 2016 Accepted 24 October 2016 Published 15 December 2016

Citation Richard D, Boyer C, Lefeuvre P, Pruvost O. 2016. Complete genome sequence of a copper-resistant bacterium from the citrus phyllosphere, *Stenotrophomonas* sp. strain LM091, obtained using long-read technology. *Genome Announc* 4(6):e01327-16. doi:10.1128/genomeA.01327-16.

Copyright © 2016 Richard et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Olivier Pruvost, olivier.pruvost@cirad.fr.

The genus *Stenotrophomonas* (Gammaproteobacteria) is composed of ubiquitous bacteria mostly found associated with plants and soil (1). Members of the genus *Stenotrophomonas* are known for their wide ecological niche range and their great phenotypic versatility including opportunistic pathogenicity to humans and resistance to many antimicrobials. For *S. maltophilia*, this high adaptability has been associated with a large number of genomic islands differing between strains and conferring them various adaptive traits (1).

Here, we sequenced the *Stenotrophomonas* sp. strain LM091 that was collected in Réunion (South West Indian Ocean) from the citrus phyllosphere. The strain possesses a copper-resistant phenotype genetically related to that reported for different plant pathogens in the *Xanthomonas* genus (2).

Sequencing was operated with PacBio RSII, using one single-molecule real-time (SMRT) cell. Resulting raw reads were assembled *de novo* using SMRT Analysis HGAP v2.3 protocol with default parameters followed by a circularization step based on a combination of the minimus assembler (3) and the SMRT Analysis resequencing v1 protocol. The assembly yielded a single circularized contig of 4,317,450 bp with a G+C content of 66.7%, a mean coverage of 229, and a predicted number of coding sequences (CDS) of 3,800.

The 16S rDNA sequence was 100% identical to that of *S. rhizophila* strain QL-P4 and 99.8% identical to that of *S. rhizophila* type strain DSM14405. As advised in Figueras et al. (4), we calculated the average nucleotide identity (ANI) (program available at <https://github.com/DamienFr/>) scores with all the chromosome sequences of type *Stenotrophomonas* strains available to date. As the ANI scores were always inferior to the 95% species ANI cutoff value (ranging from 80.2 to 87.6%), this strain could not be assigned to any of the valid *Stenotrophomonas* species nor to putative new species recently studied (5). The most closely related species were *S. rhizophila* type strain DSM14405 (87.6%) and *S. maltophilia* type strain MTCC434 (84.3%).

LM091 comprises a 43-kb long chromosome-borne Tn3-like transposon encoding for the *copLAB* gene system. This system, demonstrated to confer a copper-resistant phenotype, was previously identified on plasmids from several *Xanthomonas* (6). In Réunion, this adaptive transposon was also found in the same niche in the citrus pathogen *Xanthomonas citri* pv. *citri*, suggesting putative transfers of genetic elements between these two genera, as previously proposed (7). The nucleotide sequence of the copper-resistance transposon of LM091 is 97.7% identical to that of *Xanthomonas citri* pv. *citri*. The sequence presented here is part of the sequencing effort necessary to further understand the putative implication of strains of *Stenotrophomonas* in the transfer of genes conferring resistance to antimicrobials.

Accession number(s). The genome sequence has been deposited at GenBank under the GenBank accession no. [CP017483](#).

ACKNOWLEDGMENT

We thank J.A. Bellón Clavel for his contribution.

FUNDING INFORMATION

This work, including the efforts of Damien Richard, Claudine Boyer, Pierre Lefeuvre, and Olivier Pruvost, was funded by European Regional Development Fund (ERDF). This work, including the efforts of Damien Richard, Claudine Boyer, Pierre Lefeuvre, and Olivier Pruvost, was funded by Région Réunion. This work, including the efforts of Damien Richard, was funded by Agence Nationale de Sécurité Sanitaire de l'Alimentation, de l'Environnement et du Travail (ANSES). This work, including the efforts of Damien Richard, Claudine Boyer, Pierre Lefeuvre, and Olivier Pruvost, was funded by Agropolis Fondation (1504-004). This work, including the efforts of Damien Richard, Claudine Boyer, Pierre Lefeuvre, and Olivier Pruvost, was funded by Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD).

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Richard et al.

REFERENCES

1. Ryan RP, Monchy S, Cardinale M, Taghavi S, Crossman L, Avison MB, Berg G, van der Lelie D, Dow JM. 2009. The versatility and adaptation of bacteria from the genus *Stenotrophomonas*. *Nat Rev Microbiol* 7:514–525. <http://dx.doi.org/10.1038/nrmicro2163>.
2. Jacques MA, Arlat M, Boulanger A, Boureau T, Carrère S, Cesbron S, Chen NW, Cociancich S, Darrasse A, Denancé N, Fischer-Le Saux M, Gagnevin L, Koebnik R, Lauber E, Noël LD, Pieretti I, Portier P, Pruvost O, Rieux A, Robène I, Royer M, Szurek B, Verdier V, Vernière C. 2016. Using ecology, physiology, and genomics to understand host specificity in *Xanthomonas*. *Annu Rev Phytopathol* 54:163–187. <http://dx.doi.org/10.1146/annurev-phyto-080615-100147>.
3. Sommer DD, Delcher AL, Salzberg SL, Pop M. 2007. Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics* 8:64. <http://dx.doi.org/10.1186/1471-2105-8-64>.
4. Figueras MJ, Beaz-Hidalgo R, Hossain MJ, Liles MR. 2014. Taxonomic affiliation of new genomes should be verified using average nucleotide identity and multilocus phylogenetic analysis. *Genome Announc* 2(6): e00927–14. <http://dx.doi.org/10.1128/genomeA.00927-14>.
5. Patil PP, Midha S, Kumar S, Patil PB. 2016. Genome sequence of type strains of genus *Stenotrophomonas*. *Front Microbiol* 7:309. <http://dx.doi.org/10.3389/fmicb.2016.00309>.
6. Behlau F, Hong JC, Jones JB, Graham JH. 2013. Evidence for acquisition of copper resistance genes from different sources in citrus-associated xanthomonads. *Phytopathology* 103:409–418. <http://dx.doi.org/10.1094/PHTO-06-12-0134-R>.
7. Behlau F, Canteros BI, Jones JB, Graham JH. 2012. Copper resistance genes from different xanthomonads and citrus epiphytic bacteria confer resistance to *Xanthomonas citri* subsp. *citri*. *Eur J Plant Pathol* 133:949–963. <http://dx.doi.org/10.1007/s10658-012-9966-8>.

MICROBIAL LOCAL ADAPTATION

Adaptation of genetically monomorphic bacteria: evolution of copper resistance through multiple horizontal gene transfers of complex and versatile mobile genetic elements

D. RICHARD,*†‡  V. RAVIGNÉ,* A. RIEUX,* B. FACON,§¶ C. BOYER,* K. BOYER,* P. GRYGIEL,* S. JAVEGNY,* M. TERVILLE,* B. I. CANTEROS,** I. ROBÈNE,* C. VERNIÈRE,†† A. CHABIRAND,† O. PRUVOST* and P. LEFEUVRE*

*UMR PVBMT, CIRAD, F-97410 St Pierre, Réunion, France, †Plant Health Laboratory, ANSES, F-97410 St Pierre, Réunion, France, ‡Université de la Réunion, UMR PVBMT, F-97490 St Denis, Réunion, France, §INRA, UMR PVBMT, F-97410 St Pierre, Réunion, France, ¶INRA, UMR CBGP, F-34090 Montpellier, France, **INTA, Estación Experimental Agropecuaria Bella Vista, Bella Vista, Argentina, ††CIRAD, UMR BGPI, F-34398 Montpellier, France

Abstract

Copper-based antimicrobial compounds are widely used to control plant bacterial pathogens. Pathogens have adapted in response to this selective pressure. *Xanthomonas citri* pv. *citri*, a major citrus pathogen causing Asiatic citrus canker, was first reported to carry plasmid-encoded copper resistance in Argentina. This phenotype was conferred by the *copLAB* gene system. The emergence of resistant strains has since been reported in Réunion and Martinique. Using microsatellite-based genotyping and *copLAB* PCR, we demonstrated that the genetic structure of the copper-resistant strains from these three regions was made up of two distant clusters and varied for the detection of *copLAB* amplicons. In order to investigate this pattern more closely, we sequenced six copper-resistant *X. citri* pv. *citri* strains from Argentina, Martinique and Réunion, together with reference copper-resistant *Xanthomonas* and *Stenotrophomonas* strains using long-read sequencing technology. Genes involved in copper resistance were found to be strain dependent with the novel identification in *X. citri* pv. *citri* of *copABCD* and a *cus* heavy metal efflux resistance–nodulation–division system. The genes providing the adaptive trait were part of a mobile genetic element similar to Tn3-like transposons and included in a conjugative plasmid. This indicates the system's great versatility. The mining of all available bacterial genomes suggested that, within the bacterial community, the spread of copper resistance associated with mobile elements and their plasmid environments was primarily restricted to the *Xanthomonadaceae* family.

Keywords: contemporary adaptation, heavy metal resistance, mobile DNA, plasmid, *Stenotrophomonas*, *Xanthomonas*

Received 5 June 2016; revision received 28 November 2016; accepted 8 December 2016

Correspondence: D. Richard and P. Lefeuvre, Pôle de Protection des Plantes, UMR PVBMT, Station Ligne Paradis, 7 chemin de l'Irat, 97410 Saint Pierre, Réunion, France, Fax: +262 262 499 293; E-mails: damien.richard@cirad.fr and pierre.lefeuvre@cirad.fr

Introduction

The dynamics of genome evolution differ considerably across the bacterial diversity spectrum. Some bacteria display a level of polymorphism that makes it possible to distinguish strains from a few housekeeping genes.

On the other hand, genetically monomorphic bacteria appear very similar and require full genome sequencing to uncover some diversity (Achtman 2008). Far from being anecdotal cases, genetically monomorphic bacteria include numerous successful species, including several human pathogens, such as *Mycobacterium tuberculosis*, *Yersinia pestis* or *Escherichia coli* O157:H7 (Achtman 2008), as well as plant pathogens, such as *Pseudomonas syringae* pv. *tomato* (Cai *et al.* 2011) or *Xanthomonas citri* pv. *citri* (Leduc *et al.* 2015).

Since the discovery of genetically monomorphic bacteria, they have been considered to be an evolutionary puzzle (Achtman 2008). Genetic variability drives adaptation, which would suggest that these bacteria have a low adaptive potential. However, many have caused worldwide epidemics, demonstrating an ability to thrive in contrasted environmental conditions. For years, a variety of antibiotics have been used to control genetically monomorphic bacteria that are pathogenic to humans. Yet, the bacteria have regularly succeeded in evolving antibiotic resistance through various molecular mechanisms (Davies & Davies 2010). Genome sequencing projects have led to the discovery of the few polymorphisms responsible for some of the bacterial adaptations. For instance, ciprofloxacin resistance in *Y. pestis* and rifampicin resistance in *M. tuberculosis* are due to the mutation of a single gene on their chromosome (Telenti *et al.* 1993; Hurtle *et al.* 2003). However, in many cases, the genes responsible for adaptation were actually horizontally acquired from the bacterial mobile gene pool (Galimand *et al.* 1997; Nonaka *et al.* 2012; Behlau *et al.* 2013).

Horizontal gene transfer (HGT, Bryant *et al.* (2012)) is the incorporation of novel DNA into the host genome. In many bacterial species, HGT is so frequent that the high number of alien genes compared with the number of core genes of a species has led to the development of the concept of an open pan-genome (Garcia-Vallve *et al.* 2000; Chaudhuri *et al.* 2010). HGT can occur through conjugation (i.e. DNA transfer via plasmids or integrative conjugative elements, ICE), transformation (i.e. the uptake of naked DNA from the environment), transduction (i.e. the incorporation of DNA through a bacteriophage) and various gene transfer agents (e.g. phage-like structures that are produced by a donor cell and released into the environment (Popa & Dagan 2011)). Although we do not yet fully understand the mechanism involved, nanotubes were recently highlighted because they allow DNA and protein transfer between bacterial cells (Dubey & Ben-Yehuda 2011). Using network theory-based analyses of shared gene content among bacterial species, plasmids were shown to be of prime importance in HGT (Halary *et al.* 2010). In genetically monomorphic bacteria, alien genes and, therefore,

traces of HGT were shown to be largely restricted to bacteriophages and plasmids (Achtman 2012).

Plasmids are composed of backbone genes, essential to their evolutionary dynamics (i.e. genes involved in conjugation, replication control and stable inheritance). The presence of assisted or autonomous mobility genetic elements determines the plasmids' ability to be mobilizable or conjugative (Smillie *et al.* 2010). Due to their relative stability in the plasmid genome, backbone genes may show traces of adaptation to bacterial hosts (Norberg *et al.* 2011). Besides backbone genes, plasmids typically carry accessory genes encoding traits that may be adaptive for the bacterial host (Pansegrau *et al.* 1994; Sen *et al.* 2013). Accessory genes are often located on mobile genetic elements themselves, either transposons (Trefault *et al.* 2004; Haines *et al.* 2007) or integrons (Tennstedt *et al.* 2005), defining hot spots of insertions within the plasmid genome. These elements are generally autonomous and can be transferred between genera, families and even kingdoms (Heinemann & Sprague 1989). Plasmid-encoded accessory genes enrich the species' gene pool and provide new adaptive traits in response to environmental modifications, such as the use of a new antibiotic or the colonization of a new ecological niche (e.g. Hobman & Crossman 2015; Ochman *et al.* 2000; Weinthal *et al.* 2007). Therefore, it is now obvious that in genetically monomorphic pathogens as in others, in order to further our understanding of bacterial adaptation and ecology, we need to understand plasmid evolutionary dynamics, that is deciphering the genomic structure of plasmids and the mechanisms involved in plasmid spread among bacteria. While several studies have already documented the history of specific accessory genes (Perry & Wright 2013), very few studies have documented the evolution of the whole plasmid genome for a given ecological function (but see Norberg *et al.* (2011) for a study of the backbone genome of IncP-1 plasmid family).

Among the traits typically coded by plasmid accessory genes, antibiotic and heavy metal resistance has a tremendous impact on human, animal and plant health (Davies & Davies 2010; Seiler & Berendonk 2012). In agriculture worldwide, fungal and bacterial pathogens that cause damage to vegetable and fruit crops have been controlled chemically for over a century, with the intensive use of copper-based pesticides (Smith 1985). This has led to the development of copper resistance (Cu^R) in several plant pathogenic bacterial species, primarily in the *Pseudomonas* and *Xanthomonas* genera, two members of the Gammaproteobacteria class (Cooksey *et al.* 1990). For example, in most citrus-growing areas, Asiatic citrus canker represents a significant risk that is both direct (yield decrease, alteration of fruit quality and partial defoliation) and indirect (restrictions to

fresh fruit export because of the quarantine status of the pathogen) (Graham *et al.* 2004; Gottwald *et al.* 2007). The disease is caused by the genetically monomorphic bacterium *X. citri* pv. *citri*, which is disseminated through contaminated water (splashing and wind-driven rains), tools and plant materials (particularly over long distances). Given its Asiatic origin, *X. citri* pv. *citri* has geographically expanded and become established in many citrus-producing areas, including Florida (USA), South America and, recently, Africa (Pruvost *et al.* 2014; Leduc *et al.* 2015). In regions where *X. citri* pv. *citri* has been established for a long time, control typically involves integrated pest management strategies that partly rely on repetitive applications of copper-based pesticides (Graham *et al.* 2004). *X. citri* pv. *citri* copper resistance first emerged in 1994 in Argentina and it has persisted. There were no other reports of *X. citri* pv. *citri* copper resistance until it was discovered in the French islands of Réunion (a French territory in the South West Indian Ocean) and Martinique (a French territory in the Eastern Caribbean Sea) in 2014, two decades later (Richard *et al.* 2016, 2017). The causative genes of Argentinian *X. citri* pv. *citri* copper resistance, identified as *copLAB*, were shown to be located on a large transmissible plasmid (Behlau *et al.* 2011, 2012), suggesting that HGT has a role. Indeed, exactly the same gene set conferring copper resistance was identified in several other *Xanthomonas* species and pathovars. In addition, the hypothesis of an HGT is endorsed by evidence of the major incongruence between the phylogeny of the concatenated *copLAB* genes of *Xanthomonas* species and the species' phylogeny (Behlau *et al.* 2013).

Although it is tempting to conclude that the mechanistic basis of copper resistance has been elucidated in *X. citri* pv. *citri*, a recent study in a close relative species has revealed that the genetic basis for copper resistance at the infraspecific level can be variable. In *Xanthomonas arboricola* pv. *juglandis*, copper resistance could be conferred by plasmid-borne *copLAB* (Giovanardi *et al.* 2016), as well as by a *copABCD* gene set located on the chromosome (Lee *et al.* 1994), which is sometimes associated with an ICE structure (Cesbron *et al.* 2015). The *copABCD* system has been described in detail on a plasmid from other bacterial species, such as the tomato pathogen *P. syringae* pv. *tomato* (Mellano & Cooksey 1988) or *E. coli* (Brown *et al.* 1995). The diversity of gene sets involved in copper resistance as well as the diversity of the physical supports of these genes raise the question of the nature and origin of copper resistance in *X. citri* pv. *citri* populations.

This study aimed to improve our understanding of the genetic basis for the adaptation to copper in *X. citri* pv. *citri*. To achieve this, we gathered an unprecedented

collection of strains from the three known outbreaks of copper resistance to date: Argentina, Réunion and Martinique. By combining population genetics based on microsatellite and minisatellite markers, comparative genomics and network analysis, we address the following questions: How do copper-resistant strains from the three outbreaks relate to each other? Are the genes responsible for copper resistance similar for the different outbreaks of *X. citri* pv. *citri*? What is the genomic environment of these genes and how does it vary within pathovar *citri* and among xanthomonads? We discuss the evolutionary implications of our findings and propose future lines of research to increase our understanding of how adaptive genes spread in bacterial communities.

Materials and methods

Bacterial strains

The study included a total of 350 *Xanthomonas citri* pv. *citri* Cu^R strains from the three known outbreaks of copper resistance (Argentina, *n* = 111; Réunion, *n* = 219; and Martinique, *n* = 20). Argentinian strains originated from a collection held at INTA and were collected from 1994 (the first outbreak) to 2015. Starting in early 2014, when copper resistance was discovered in Réunion, strains were collected in the field directly from active epidemics. In addition, a large collection of citrus strains from Réunion (held at the PVBMT laboratory and including strains collected since 1978) was tested for copper resistance. This revealed resistance in 219 Cu^R strains from all locations in Réunion, as well as the oldest copper-resistant strains found, dating from 2010. Martinique was a disease-free territory until the 2014 outbreak (Richard *et al.* 2016). Every strain characterized was Cu^R and 20 were collected from the first disease focus in Morne Rouge.

In addition to Cu^R haplotypes (*n* = 115 identified among 219 strains), all known haplotypes corresponding to copper-susceptible (Cu^S) *X. citri* pv. *citri* strains sampled in Réunion (*n* = 524) from all the citrus-producing areas were also included in the study to assess the genetic relatedness between Cu^R and Cu^S strains on the island. This extensive dataset was built from strain collections sampled over two periods, 1978–1997 (historical strains) and 2009–2015 (contemporary epidemics).

Finally, as Cu^R xanthomonads were originally reported in strains causing bacterial spot of tomato and pepper and no genomic resources were available for such strains, reference Cu^R strains from the four species causing this disease were included in this study. They originated from Argentina, Mauritius, New Zealand and the USA (Table 1).

Table 1 Characteristics of copper-resistant bacterial strains used for long-read sequencing

Strain	Other numbers	Accession	Genus	Species	Pathovar	Copper-resistance location	Country	Date	Host
LMG930		CP018463-CP018467	<i>Xanthomonas</i>	<i>euvasicatoria</i>		Plasmid	USA	1969	Pepper
LMG911		CP018725-CP018727	<i>Xanthomonas</i>	<i>vesicatoria</i>		Plasmid	New Zealand	1955	Tomato
LM091		CP017483	<i>Stenotrophomonas</i>	sp.		Chromosome	Réunion	2015	Tangor
LM199	Xcc 15-4632 (INTA)	MSQV000000000	<i>Xanthomonas</i>	<i>citri</i>	<i>citri</i>	Plasmid	Argentina	2015	Orange
LM180	Xcc 03-1638 (INTA); A44*	MSQW000000000	<i>Xanthomonas</i>	<i>citri</i>	<i>citri</i>	Plasmid	Argentina	2003	Grapefruit
LM159	Bv-5-4a (INTA)	CP018468-CP018471	<i>Xanthomonas</i>	<i>vesicatoria</i>		Plasmid	Argentina	1987	Pepper
LL074-4		CP018847-CP018849	<i>Xanthomonas</i>	<i>citri</i>	<i>citri</i>	Plasmid	Martinique	2014	Grapefruit
LJ207-7		CP018850-CP018853	<i>Xanthomonas</i>	<i>citri</i>	<i>citri</i>	Plasmid	Réunion	2012	Kafir lime
LH3		CP018472-CP018476	<i>Xanthomonas</i>	' <i>perforans</i> ' [†]		Plasmid	Maurice	2010	Tomato
LH276		CP018854-CP018857	<i>Xanthomonas</i>	<i>citri</i>	<i>citri</i>	Plasmid	Réunion	2010	Kafir lime
LH201		CP018858-CP018860	<i>Xanthomonas</i>	<i>citri</i>	<i>citri</i>	Plasmid	Réunion	2010	Kafir lime
JS749-3		CP018728-CP018730	<i>Xanthomonas</i>	<i>gardenii</i>		Plasmid	Réunion	1997	Tomato
ICMP7383		CP018731-CP018734	<i>Xanthomonas</i>	<i>gardenii</i>		Plasmid	New Zealand	1980	Tomato

ICMP (International Collection of Microorganisms from Plants, Landcare Research, Auckland, New Zealand), BCCM/LMG (Belgian Coordinated Collections of Microorganisms, University of Ghent, Belgium), INTA (Instituto Nacional de Tecnología Agropecuaria).

*As designated by Behlau *et al.* (2011).

[†]X. '*perforans*' was reclassified as X. *euvasicatoria* (Constantin *et al.* 2016).

Evaluation of copper resistance

To assess copper resistance, we used the growth test on YPGA (yeast extract 7 g/L, casein peptone 7 g/L, glucose 7 g/L, agar 18 g/L, pH 7.2), supplemented with $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ (470 mg/L), and PCR using *copL* primers. We slightly modified the PCR protocol reported previously (Behlau *et al.* 2013) using GoTaq Flexi polymerase (Promega), a lower primer concentration (5 pmol/ μL) and a higher annealing temperature (66 °C). In the case of negative results using *copL* primers, *copA* and *copB* primers were also used, as reported previously (Behlau *et al.* 2013).

Minisatellite and microsatellite genotyping

To determine how the studied strains relate to the known diversity of *X. citri* pv. *citri* worldwide, a subset of available strains from Argentina (47 Cu^{R} and 45 Cu^{S}), Martinique (9 Cu^{R}) and Réunion (64 Cu^{R} and 103 Cu^{S}) were genotyped using 31 minisatellites (MLVA-31) developed for global epidemiology analyses and compared with the known worldwide diversity (<http://www.biopred.net/MLVA/>). Genotyping and discriminant analysis of principal components were performed as reported earlier (Pruvost *et al.* 2014). MLVA fragment sizes were transformed to tandem repeat numbers (rounding to the superior integer). Manhattan distances between strains were then calculated and multidimensional scaling (MDS) plots were computed using the BIOS2MDS R package (Pele *et al.* 2012). Categorical minimum-spanning trees (MST) were built using the algorithm recommended for tandem repeat data combining global optimal eBURST (GOEBURST) and Euclidean distances in PHYLOVIZ v1.0 (Francisco *et al.* 2012).

The relatedness between strains was assessed at smaller spatiotemporal scales (i.e. at the country level) by examining the diversity of all *X. citri* pv. *citri* strains using a set of 14 microsatellites (MLVA-14), which represent the most discriminative *X. citri* pv. *citri* genotyping technique available for infrapathotype typing (Bui Thi Ngoc *et al.* 2009). Population genetics analyses were performed on two different datasets, as explained above. Moreover, allelic richness was computed using a rarefaction method with the HIERFSTAT R package (Goudet 2005). Population pairwise R_{ST} values were computed for estimating genetic differentiation among Cu^{R} populations at the country level. Significance was tested with 1000 permutations using ARLEQUIN 3.5.2.2 (Excoffier *et al.* 2005). The first dataset, which consisted of all Cu^{R} strains from the three resistance outbreaks, allowed us to decipher the genetic relationships between the three outbreaks. The second dataset, consisting of all Cu^{R} and Cu^{S} strains from Réunion (the

only region where we have extensively sampled local populations), enabled us to examine the relationships between the Cu^{R} and Cu^{S} strains that were isolated both before and during the Cu^{R} *X. citri* pv. *citri* epidemics. When comparing Cu^{R} and Cu^{S} strains from Réunion, the optimal number of clusters in the dataset was assessed by computing silhouette scores from multiple K-means runs using the BIOS2MDS R package.

DNA sequencing and assembly

Based on strain diversity analyses and the Cu profiles (i.e. the combination of the growth test and PCR results), six *X. citri* pv. *citri* strains from Argentina, Martinique and Réunion were selected for sequencing (Table 1). In addition, six other reference strains of Cu^{R} xanthomonads causing bacterial spot disease on tomato and/or pepper and a Cu^{R} strain of *Stenotrophomonas* sp., a commensal bacterium collected from the citrus phyllosphere in Réunion (i.e. the same ecological niche as *X. citri* pv. *citri*), were also selected for genome sequencing. In total, 13 genomes were completely sequenced using the long-read PacBio RSII technology, using one SMRT cell for each strain (Table 1). Assembly of the resulting raw reads was made using SMRT ANALYSIS HGAP v. 2.3 protocol with default parameters. An additional step of contig circularization was conducted using a combination of minimus assembler (Sommer *et al.* 2007) and the SMRT ANALYSIS resequencing v. 1 protocol.

Sequence annotations

Sequence data obtained from the 13 strains were screened for known copper-resistance systems using BLASTN and BLASTP algorithms. As for most strains, Cu^{R} genes were found on a 230-kb plasmid with strong homology. We arbitrarily selected strain LH201 as a reference. Hereafter, this plasmid is referred to as pLH201.1. We carried out pLH201.1 CDS prediction and automatic gene annotation using the MAGE genome annotation platform (Vallenet *et al.* 2006) and then manual curation of the plasmid annotations using the extensive set of tools and databases available via the platform.

Specific plasmid features were characterized in more detail. We conducted a search of plasmid toxin-antitoxin systems using the online databases TADB (Shao *et al.* 2011) and RASTA (Sevin & Barloy-Hubler 2007). The conjugative apparatus of pLH201.1 (encoded by 16 *tra* genes dispatched in three clusters) was blasted against amino acid sequences of (i) the plasmid database of NCBI and (ii) the ICEberg database (Wozniak & Waldor 2010). We only retained hits

that displayed homology with at least one gene (on at least 70% of its length) from each of the three gene clusters involved in conjugation. Finally, ISfinder (Siguier *et al.* 2006) was used to detect known insertion sequences.

We used CIRCOS (Krzywinski *et al.* 2009) to plot a graphical representation of the plasmid. We computed GC% and GC skew $[(G+C)/(G-C)]$ using a custom perl script available at <https://github.com/DamienFr/GC-content-in-sliding-window>.

The search for homologous sequences

In our search for sequences that are homologous to pLH201.1 among all the available sequences, we queried the Whole Genome Shotgun (WGS) and Non-Redundant (NR) public nucleotide databases. Results were automatically retrieved using BioPerl modules (SEARCHIO, SeqIO and EUTILITIES).

We first used CYTOSCAPE v. 3.3.0 to build a network of the gene sequences that were homologous with pLH201.1. We retained all the sequences that had a hit with at least one of the pLH201.1 genes (with 95% nucleotide identity over 95% of gene length). All the hits were blasted against each other to obtain the full network structure. The network was then plotted using an edge-weighted spring-embedded layout, which positions the nodes closer if the number of genes they share is higher.

We adopted a clustering strategy based on average nucleotide identity (ANI) to determine the diversity of sequences related to pLH201.1 and eliminate the redundancy of the hit dataset. Nucleotide sequences that shared more than 10 genes with pLH201.1 (with 70% nucleotide identity over 70% of gene length) were clustered using a modified ANI algorithm (available at https://github.com/DamienFr/Clustering_with_ANI) that takes into account the identity of the homologous region in relation to the size of the complete sequence. For each sequence that was representative of a cluster, CDS prediction was performed using PRODIGAL (Hyatt *et al.* 2010). Sequences were then compared using a combination of MAUVE (Darling *et al.* 2004), MAFFT (Katoh & Standley 2013) and the APE (Popescu *et al.* 2012) and GENOPLOT (Guy *et al.* 2010) R packages.

Results

In this study, we combined PCR screening, micro- and minisatellite typing and comparative genomics to place the strains that caused the outbreak into the known diversity and to characterize copper-resistance genetic and genomic support.

Preliminary characterization of Cu^R outbreak strains and placement within the pathovar diversity

Most Cu^R *Xanthomonas citri* pv. *citri* strains (98%) produced the expected amplicon when assayed by PCR with *copL* primers. Notably, seven strains from Argentina grew on YPGA supplemented with CuSO₄ although they were PCR negative when assayed with *copL*, *copA* and *copB* primers.

Strains sampled from the three outbreaks were assessed by MLVA-31 minisatellites and were all assigned to the same genetic cluster (formerly called DAPC1 in Pruvost *et al.* (2014)).

At the spatiotemporal evolutionary scale defined by these markers, strains from the three origins formed a single clonal complex (CC), that is a network of haplotypes composed of single-locus variants. A total of 34 haplotypes were delineated with no clear structure in terms of geographical origin or copper phenotype (Fig. S1, Supporting information; information on haplotypes available at <http://www.biopred.net/MLVA/>). Strains from Martinique were assigned to haplotypes also including strains from Réunion. Some MLVA-31 haplotypes were shared by Cu^R strains from Martinique and Réunion. Most Cu^R strains from Réunion were assigned to two frequent haplotypes (#173 and 175), which also contained Cu^S strains of the same origin. Similarly, haplotype #51, the most frequent haplotype among Argentinian strains, included strains that differed in terms of the copper phenotype and/or *cop* PCR.

Relationships between outbreaks

Using microsatellite data (MLVA-14), we first characterized the genetic diversity of Cu^R *X. citri* pv. *citri* strains originating from the three resistance outbreaks. A total of 83, eight and 115 haplotypes were identified among the 111, 20 and 219 Cu^R strains from Argentina, Martinique and Réunion, respectively. Using a rarefaction method ($n = 20$), strains from Argentina, Martinique and Réunion showed a mean allelic richness of 3.96, 1.71 and 2.76, respectively. Two clusters were identified according to their distribution on the first axis of the MDS plot (68.0% of the total variance). Their identification was endorsed by the silhouette score derived from multiple *K*-means runs (Fig. 1A). Axes 2–5 contributed much less to total variance, ranging from 6.2% to 2.2%. Cluster 1 included all strains from Martinique and Réunion, while cluster 2 only had Argentinian strains. Cu^R strains that were PCR negative for *copLAB* also grouped in cluster 2 without any apparent substructure. Within cluster 1, strains from Martinique and Réunion were moderately but significantly differentiated

COPPER RESISTANCE IN MONOMORPHIC BACTERIA 2137

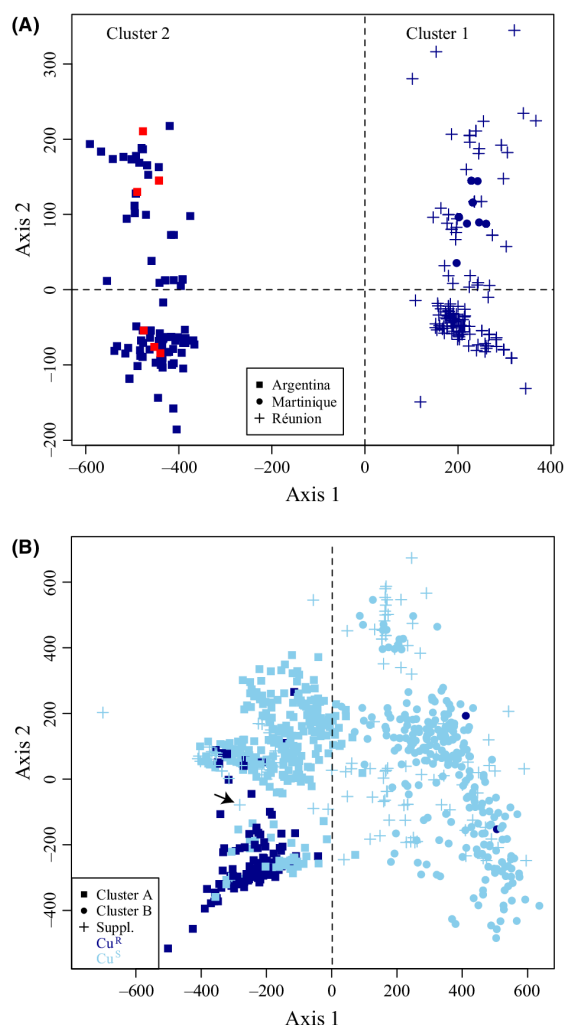


Fig. 1 Genetic diversity derived from microsatellite data of *Xanthomonas citri* pv. *citri* strains differing in geographical origin and their susceptibility to copper. (A) Manhattan distances represented by multidimensional scaling (MDS) among copper-resistant haplotypes from Argentina, Martinique and Réunion. The MDS plot showing axes 1 and 2 represented 68.0% and 6.2% of the total variation, respectively. Blue and red dots indicate strains bearing the copLAB and copABCD system, respectively, as confirmed by PCR. (B) Manhattan distances represented by multidimensional scaling (MDS) among copper-resistant and copper-susceptible haplotypes from Réunion, sampled between 2009 and 2015. Historical strains sampled between 1978 and 1997 were included as supplementary (suppl.) individuals in the MDS analysis. The MDS plot showing axes 1 and 2 represented 27.5% and 14.8% of the total variation, respectively. The black arrow localizes strain D07, which was identified as the historical strain genetically closest to the main clonal complex of epidemic copper-resistant strains.

($R_{ST} = 0.22$; $P < 0.001$). In contrast, strains from Argentina (i.e. cluster 2) were strongly differentiated ($R_{ST} = 0.85$; $P < 0.001$) from strains collected in Martinique and Réunion (i.e. cluster 1).

Relationships between copper-resistant and copper-susceptible strains

Microsatellite data (MLVA-14) were also used to decipher the genetic relatedness of the Cu^R and Cu^S strains from Réunion. Silhouette scores derived from multiple K-means runs suggested a lack of strong structure in the dataset. For $K > 2$, strain assignment probabilities to clusters were often low (i.e. < 0.5). We decided to retain $K = 2$ because minimal strain assignment probabilities to clusters were > 0.8 . All but two Cu^R strains were assigned to cluster A, together with Cu^S strains. In contrast, only two Cu^R strains were assigned to cluster B, including almost exclusively Cu^S strains (Fig. 1B). In a few cases, Cu^R and Cu^S strains shared the same haplotype. Most Cu^R strains were genetically closely related. They were structured as five CCs and 27 singletons, that is haplotypes sharing no single-locus variation with others (Fig. S2, Supporting information). Among these, a major CC (CC1) comprised 64 haplotypes, representing 152 strains (69%). An additional set of 48 strains (37 haplotypes), corresponding to smaller CC and singletons, consisted of double-locus variants of CC1. These strains did not join the main CC, which is probably due to incomplete sampling. Both clusters included historical (1978–1997) and contemporary strains (≥ 2009). The Cu^S D07 strain (sampled from satsuma mandarin in 1989) was the closest historical strain to the main group of Cu^R strains (i.e. a double-locus variant).

Sequencing

After PacBio reads assembly of the 13 fully sequenced strains, we obtained from one to 18 contigs per strain (see Table S1, Supporting information for details). It is important to note that all the chromosome sequences were successfully circularized and 34 of the remaining 70 contigs were circularized into plasmids. No plasmid was detected in the copper-resistant commensal strain of *Stenotrophomonas* sp. isolated from citrus in Réunion and one to four plasmids were detected in xanthomonads, depending on the strain.

Features of plasmids associated with copper resistance in xanthomonads

Cop genes were plasmid-borne for all the *Xanthomonas* strains sequenced. Consistent with previous WGS data

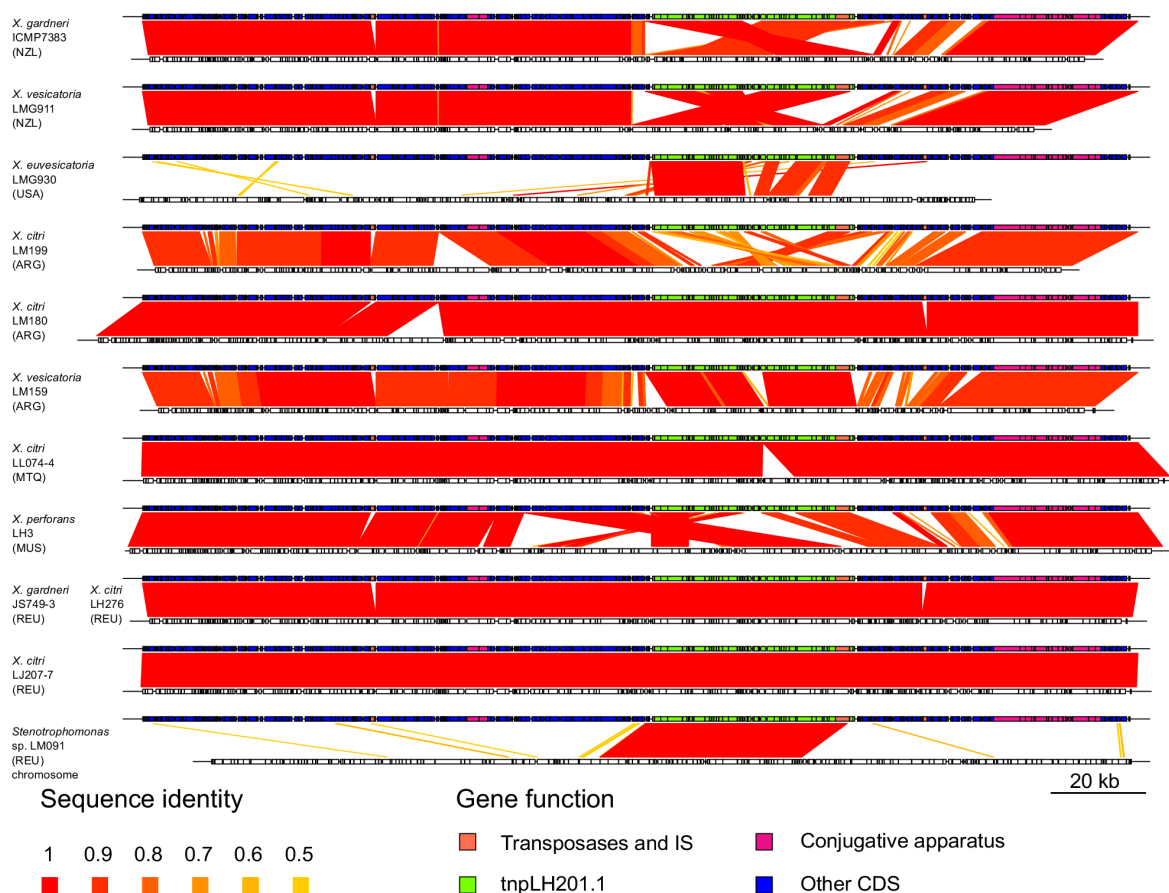


Fig. 2 Alignments between pLH201.1 and sequences carrying pLH201.1 homologues from strains sequenced in this study. All sequences are plasmids, unless specified otherwise. On pLH201.1, blocks represent curated gene prediction (MAGE), whereas on other genomes the blocks represent uncurated Prodigal gene prediction. The pLH201.1 genes are categorized into several groups according to the caption. The comparison zone shows homologous nucleotide sequences. Colour varies according to genetic distance between homologous blocks defined by a MAUVE alignment.

from *Stenotrophomonas maltophilia*, these genes were present on the chromosome of the strain of *Stenotrophomonas* sp. isolated from citrus phyllosphere (Crossman *et al.* 2008; Davenport *et al.* 2014; Pak *et al.* 2015). A MAUVE comparison suggested that, with the exception of *Xanthomonas euvesicatoria* LMG930, all other plasmids bearing *cop* genes were genetically related (Fig. 2). Consequently, the plasmid from the *X. citri* pv. *citri* strain LH201 was arbitrarily selected as our reference. MAGE annotation of this plasmid revealed two tRNA and 258 CDS, 176 of which (68%) encoded proteins of unknown function (Fig. 3). We identified the plasmid replication initiator gene *trfA* near a GC-skew switch. This commonly indicates the origin of replication (*oriV*) (Grigoriev 1998), suggesting that the *oriV* locus is located nearby. We revealed an 87-amino acid-long conserved domain of a putative HigB-like addiction module killer toxin (*e*-value < 10^{-10})

(Schuessler *et al.* 2013). A 222-amino acid-long putative transcriptional regulator of the xenobiotic response element family, which might serve as an antitoxin protein, was found downstream and antisense to the toxin (*e*-value < 10^{-8}).

The pLH201.1 encodes for all the apparatus required for conjugation with 16 Tra proteins, located at two different regions of the plasmid (region 1: 69 675 – 74 152; region 2: 182 783 – 205 787) and organized in at least three operons (Fig. 3). Using the NCBI plasmid and the ICEberg ICE databases, we conducted a search at the amino acid level, keeping only sequences that matched at least one Tra protein from each of the three pLH201.1 conjugative operons.

These 16 pLH201.1 Tra proteins shared best amino acid identity (AAI) (from 20% to 58%) and organization with some IncA/C plasmids and SXT/R391-related

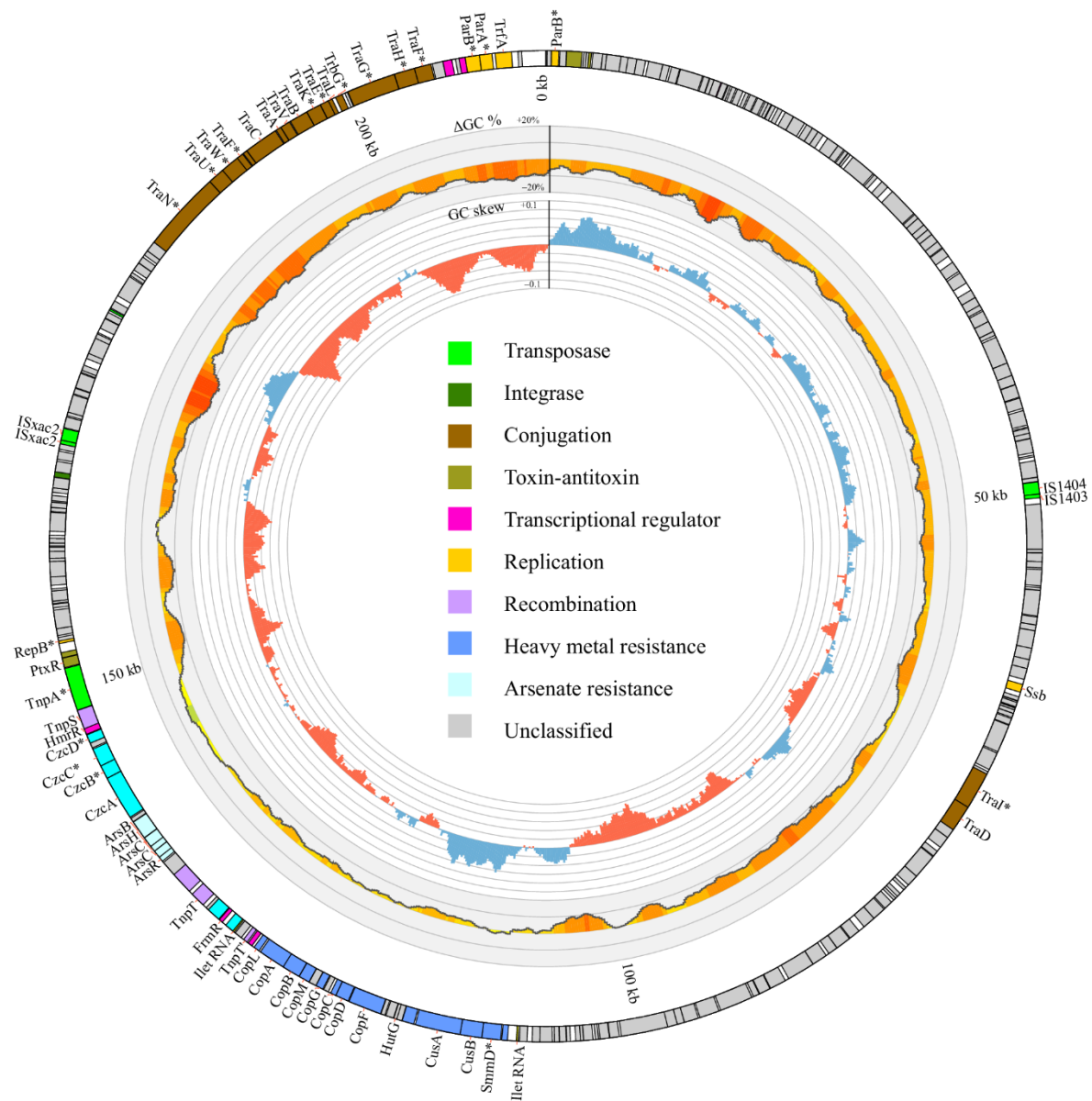


Fig. 3 Circular representation of the plasmid pLH201.1. Circles, from the outer to the inner, represent: (1) predicted genes and their function according to the caption at the centre of the circle (putative genes are followed by *); (2) difference between mean GC content of the chromosome of LH201 and pLH201.1's % of GC content in a 2000-bp sliding window with a 200-bp step; and (3) GC skew using a 4000-bp sliding window with a 200-bp step.

ICE conjugative apparatus (Fricke *et al.* 2009; Carraro *et al.* 2014) (Table S2, Supporting information). On the 52 plasmid hits (120 kb to 582 kb) and 16 ICE hits, not a single Tra homologue displayed a AAI superior to 60%. Twelve matched ICEs belonged to the SXT/R391 family and the remaining four were from the SPI-7 family (one sequence) or unclassified (three sequences). Most of the plasmid hits ($n = 23$) were

annotated as IncA/C plasmids, while the remainder were spread in groups H, P, T, J, F, multireplicon and unknown plasmid incompatibility groups. All the plasmids conferred multidrug resistance and were found in five bacterial families (*Pseudomonadaceae*, *Enterobacteriaceae*, *Aeromonadaceae*, *Burkholderiaceae* and *Vibrionaceae*) with contrasted ecological niches and geographical origins.

As the protein sequences of conjugative relaxases are useful for plasmid classification, we assessed the AAI of the pLH201.1 relaxase (TraI_{pLH201.1}) to that of each of the six MOB groups defined in the literature (Garcillan-Barcia *et al.* 2009). TraI_{pLH201.1} solely matched the MOB_H type (*e*-value < 10⁻⁴) with its typical amino acid signature ((HQ)-x2-PASE-x-HHH-x3-GG-x3-H-x-L and (LV)-x-HD-(AVLI)-GK). MOB_H relaxases are scarce and have only been reported in large plasmids (>60 kb) (Smillie *et al.* 2010) and found in the incompatibility groups IncH, IncJ, IncT, IncP7 and IncA/C (Garcillan-Barcia *et al.* 2009). Of all known MOB_H clades and subclades, the TraI_{pLH201.1} appeared to be most closely related to the MOB_{H121} subclade of MOB_{H12} (Alvarado *et al.* 2012). TraI_{pLH201.1} also displayed a conserved domain of the PFL_4751 family of ICE relaxases (*e*-value = 2.20 e⁻²¹), required for transfer of the SXT and R391 ICE types (Daccord *et al.* 2010). Similarly, the conjugative coupling factor TraD was homologous to the SXT-TraD domain (*e*-value = 0) found in conjugative transposon-like mobile genetic elements (Beaber *et al.* 2002) and various groups of plasmids including IncA/C (Fernandez-Alarcon *et al.* 2011). TraD_{pLH201.1} also shared good AAI with the TrwB coupling factor (*e*-value of 2.35 e⁻¹⁹) from IncW conjugative plasmids (Gomis-Ruth *et al.* 2002).

Globally, the content and organization of pLH201.1 conjugative apparatus clearly shared similarities with that of IncA/C plasmids and SXT/R391 ICE. However, the low AAI levels indicate that it may be a new system that has not yet been described.

Cop genes are part of a Tn3-like transposon

In IncA/C plasmids, genes associated with adaptive traits are often found as part of complex transposons (Harmer & Hall 2014) and display a higher GC content (Zhang *et al.* 2014). Globally, pLH201.1 had a GC content of 59.2%, lower than the 64.8% of the chromosome, but displayed a higher local GC content (63.3%) in the ~108- to 152-kb region. GC-skew profiles presented several variations. These lines of evidence suggest that the plasmid shows a mosaic structure. On pLH201.1, this region (hereafter referred to as TnpLH201.1, located at 108 034–151 931 bp) was surrounded by two inverted repeats of 34 bp. It contained genes that are syntenic and similar to genes from the plasmid-encoded *Xanthomonas* TnXo19, a Tn3-like transposon (Niu *et al.* 2015). This includes a transposon-related cointegrate protein *tnpT*, a cointegrate protein *tnpS*, a transposase *tnpA*, a DNA recombination protein and a DNA helicase, that all display nucleotide identity (NI) between 70% and 90% with their TnXo19 homologues, on 94%, 70%, 26%, 100% and 100% of TnPLH201.1's gene length,

respectively. *tnpT*, *tnpS* and *tnpA* were all shown to be involved in the transposition of some Tn3-family transposons (Tsuda & Iino 1988; Yano *et al.* 2013). Inside TnpLH201.1, we found an additional copy of the 34-bp repeat, which formed two direct copies separated by 43 897 bp at one extremity of TnpLH201.1 and one inverted copy at the other extremity. This pattern, typical of composite transposons, has also been reported on a Tn3-like transposon from *Pseudomonas putida* (Lauf *et al.* 1998).

Gene content of the Tn3-like transposons

The pLH201.1 contained two clusters of heavy metal resistance genes that were included in TnpLH201.1. Apart from LM199, all the strains displayed a TnpLH201.1 homologue with a globally conserved and syntenic gene content, however sometimes showing rearrangements (Figs 2 and S3, Supporting information). The first cluster was delineated by two 34-bp direct repeats and encompassed several genes, including the previously reported *copLABMGCDF* genes involved in copper resistance in Argentinian *X. citri* pv. *citri* populations (Behlau *et al.* 2011) and *cusAB/smmD* present in *S. maltophilia* (Crossman *et al.* 2008). The *copLABMGCDF* region was identical (100%) between the five sequenced *X. citri* pv. *citri* strains that possessed this system and with the 10 328-bp region, which was described in *X. citri* pv. *citri* strain A44 (i.e. LM180 in the present study) and known to be functionally involved in Cu^R (Behlau *et al.* 2011) (Figs S3 and S4, Supporting information). Interestingly, the Cu^R *X. citri* pv. *citri* strain from Argentina LM199 failed to produce PCR amplicons using primer pairs specific to *copL*, *copA* and *copB* of the *copLAB* system. Its genome sequence displayed a plasmid backbone extremely similar to that of pLH201.1 but with a distinct copper transposon region (Fig. 2), hereafter called TnpLM199. Indeed, the annotation of TnpLM199 revealed the presence of the alternate copper-resistance system *copABCD*. TnpLM199 however displayed a transposition apparatus (*tnpA*, *tnpT* and *TnpS*) similar to that of TnpLH201.1 (NI above 90% on more than 80% of each gene's length) and comprised almost identical 34-bp inverted repeats at its extremities. The nucleotide sequences of *copA*, *copB*, *copC* and *copD* from LM199 respectively showed a NI of 97, 98, 98 and 98% with those of the known chromosomal system of *X. arboricola* pv. *juglandis* Xaj417 isolated from walnut (Pereira *et al.* 2015). Copper-resistance gene nomenclature is quite ambiguous: whereas *copA* and *copB* from *copABCD* respectively share 98 and 63% NI (both on 74% of the gene length) with their *copLAB* homologues, CopC_{copABCD} and CopD_{copABCD} amino acid sequences only display low

COPPER RESISTANCE IN MONOMORPHIC BACTERIA 2141

AAI levels with their *copLAB* counterparts (34% on 97% of gene's length) (see Fig. S4, Supporting information). PCR primers targeting the four genes were designed (Table S3, Supporting information). The seven Cu^R *X. citri* pv. *citri* strains from Argentina that were *copLAB* negative by PCR all produced amplicons of the expected size for *copABCD*. The system found in pLM199, as in *X. arboricola* pv. *juglandis*, did not encode for CopRS, unlike *E. coli* and *P. syringae*. However, a transcriptional regulator that belongs to the MerR family was found close to the *copABCD* cluster. This regulator family has been reported to respond to environmental stimuli, including heavy metals. It also controls the expression of *copA* in *E. coli* (Stoyanov *et al.* 2001; Brown *et al.* 2003).

In TnpLH201.1, genes homologous to *cusAB/smmD*, a heavy metal efflux resistance-nodulation-division (HME-RND) (>95% AAI) of *S. maltophilia*, were identified. These sequences corresponded to that of known copper/silver efflux pumps. On all strains from Réunion and Martinique (five *X. citri* pv. *citri*, one *X. gardneri* and one *Stenotrophomonas* sp.) and in one Argentinian strain (LM180), this cluster was surrounded by two almost perfect 907-bp-long direct repeats. We only found one copy on other Cu^R-bearing DNA molecules (plasmids for all strains apart from *Stenotrophomonas* sp. LM091) for all the other strains and none in LH3. Interestingly, on TnpLH201.1, the two copies of 907 bp comprised an Ile tRNA (UAU anticodon), which is absent on the chromosome of *X. citri* pv. *citri* LH201. Codon usage revealed a bias in the use of the ATA codon between chromosomal genes (1.4% of the Ile-encoding codons) and plasmid-borne genes (6.5%) (data not shown).

The second cluster contained genes also involved in heavy metal resistance: *czcA*, *czcB*, *czcC* and *czcD* (which encode a cobalt/zinc/cadmium efflux system), and three genes involved in arsenic resistance, an arsenate reductase (*arsC*), a NADPH-dependent FMN reductase (*arsH*) and an arsenite/antimonite transporter (*arsB*). Again, *czc* and *ars* genes were found to be highly (>95%) and moderately (30–67%) identical, respectively, to those of *S. maltophilia*. The latter genes were located downstream of the *arsR* transcriptional regulator, which is induced by arsenite and antimonite (Wang *et al.* 2004). This cluster also encoded a copper-dependent transcription regulator *hmrR* from the MerR family, which is highly identical (97%) to that of *S. maltophilia*.

The region between the two clusters was surrounded by 66-bp-long direct repeats and encoded a metal chelator protein [also described in TnXo19 from *X. oryzae* pv. *oryzae*, (Niu *et al.* 2015)], a heavy metal efflux protein, a copper-sensing transcriptional repressor (which binds to a gene promoter and to copper with a higher

affinity to copper) and a metal binding exoribonuclease (which might help degrade the transcriptional repressor in the presence of copper).

Slight variations were observed between the TnpLH201.1 homologues. For example, pLL074-4 displayed a 6 945-bp insertion at position 154 316. This insertion contained 40-bp inverted repeats at both ends. We found an identical transposon (100% NI) in a *X. citri* pv. *citri* strain jx-6 plasmid pXAC33, which encoded a transposase, a resolvase and the twitching motility protein PilT. In addition, pLMG930 displayed an inserted gene (99% NI with *S. maltophilia* iron permease), a deletion of the *ars* cluster and several gene rearrangements in the region between the *copLAB* and the *czcABCD* clusters (Fig. S3, Supporting information).

Therefore, TnpLH201.1 appeared to be a hotspot of insertions, deletions and rearrangements, consistent with previous data on *Xanthomonas* Tn3-like transposons (Ferreira *et al.* 2015; Niu *et al.* 2015). Two different Tn3-like transposons encoded for copper resistance. Other genes putatively involved in heavy metal resistance were conserved in the sequenced plasmids, except for a single strain (*X. euvesicatoria* LMG930).

TnpLH201.1 is found in various genomic environments

Homologues of the transposon TnpLH201.1 were found in diverse genomic environments within the other sequenced strains (see Figs 2, S5 and Table S4, Supporting information for values of nucleotide divergence between the blocks of nucleotide identity defined in Fig. 2). First, we found that TnpLH201.1 was integrated in plasmids that were highly homologous and syntenic to the pLH201.1. These conserved plasmids were present in *X. citri* pv. *citri* strains LM180, LH276, LJ207-7 and LL74-4 (from the three regions studied) and from other *Xanthomonas* species pathogenic to solanaceous species: *X. gardneri* JS749-3 (Réunion) and *X. vesicatoria* LM159 (Argentina). Then, we found that highly homologous copies of the transposon were also integrated in rearranged pLH201.1 homologues present in strains of other *Xanthomonas* species pathogenic to solanaceous species: *X. 'perforans'* LH3 (synonym *X. euvesicatoria*; Mauritius), *X. gardneri* ICMP7383 (New Zealand) and *X. vesicatoria* LMG911 (New Zealand) (Table 1). We also observed a highly similar transposon homologue that was integrated in a markedly different plasmid environment (*X. euvesicatoria* LMG930, USA). The conjugative apparatus of pLMG930 displayed homology to that of pBVIE04 from *Burkholderia vietnamensis* G4, an ecologically versatile rice root-associated nitrogen-fixing betaproteobacterium (Chiarini *et al.* 2006). Indeed, genes of pBVIE04 involved in conjugation are located in three

separate genomic regions, each of which are very well conserved in pLMG930, sharing 93%, 90% and 92% NI, on 7 068, 5 037 and 7 792 bp, respectively. Lastly, TnpLH201.1 was integrated in the chromosome of the citrus-associated strain of *Stenotrophomonas* sp. LM091 (Réunion) with a NI of 97.7% on 43 276 bp.

Networks of gene sharing

As the clues indicated a mosaic structure for the Cu^R plasmid, we searched for pLH201.1 homologues in the public NCBI databases NR and WGS. Networks of gene sharing (Fig. 4) revealed that pLH201.1 homologues present in *X. citri* pv. *citri* could only be identified from *X. gardneri* and *X. vesicatoria*, consistent with data produced in this study. In contrast, genes homologous to TnpLH201.1 were detected from 14 species included in five genera (*Xanthomonas*, *Stenotrophomonas*, *Pseudoxanthomonas*, *Pelomonas* and *Pseudomonas*). Globally, Cu^R gene homologues were found further apart in the taxonomy than plasmid backbone homologues.

Thirty percent of the genomes represented on the network only shared one or two genes with pLH201.1, 86% of which (26% of the total) only matched known insertion sequences.

After clustering 180 NCBI sequences sharing more than 10 genes with pLH201.1 (with 70% NI over 70% of gene length), we obtained 62 clusters (Fig. S6, Supporting information). Three patterns of homology emerged. The first pattern (two clusters) comprised homologues to the complete pLH201.1. The second pattern was found with clusters whose backbone region displayed NI with the entire pLH201.1 backbone but differed from TnpLH201.1 (three clusters). Within this group, all clusters displayed a highly similar backbone region and a conserved gene content, which suggests that they are closely related. However, their accessory gene regions were different with the lack of *copLABMGCDF* (*X. 'perforans'*, contig Accession no. JZUY01000051, that can be circularized with a 21-bp perfect-match overlap), or the incorporation of other gene clusters coding traits such as cobalt efflux or ion transport (*X. 'perforans'*, Accession no. JZVH01000033). In addition, this cluster comprised a contig of a *X. euvesicatoria* pv. *allii* strain from Réunion matching the whole backbone region of pLH201.1 (Accession no. JOJQ01000000) that we were unable to circularize. Finally, the third pattern consisted of multiple clusters that only shared NI with TnpLH201.1 or parts of it. However, the distance between them was sufficient to form distinct clusters.

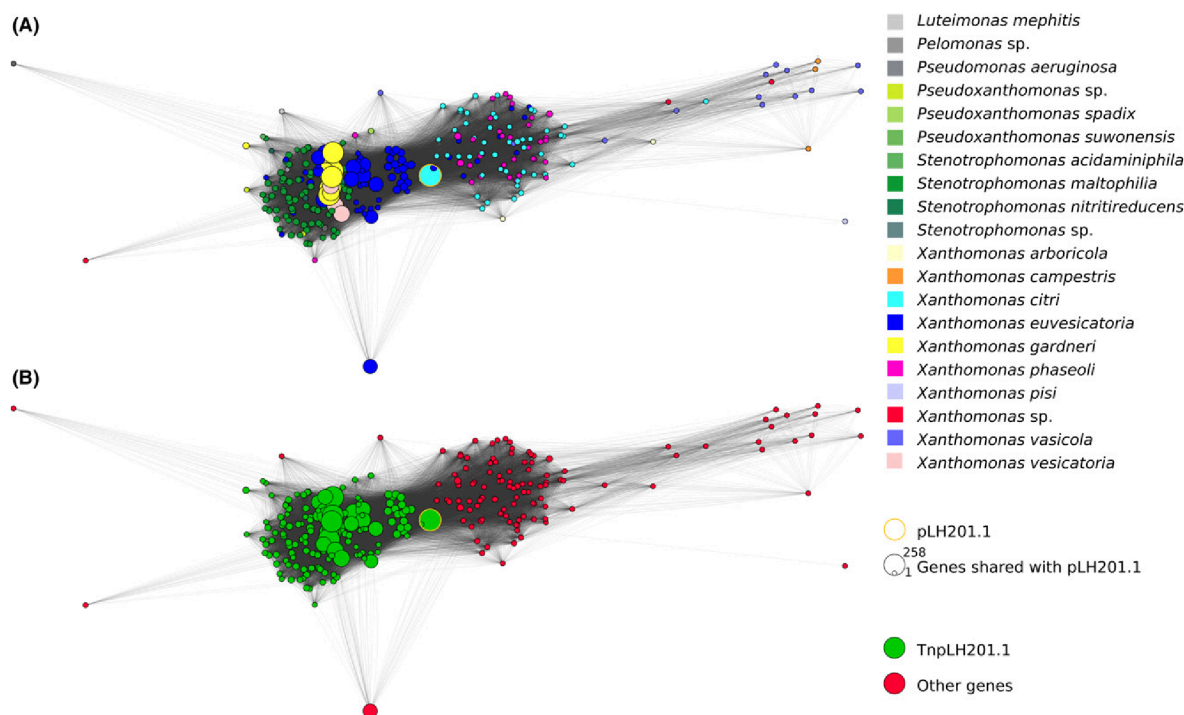


Fig. 4 Network of all NCBI genomes sharing homologous genes with pLH201.1 (nucleotide identity > 95% on >95% of pLH201.1 gene length). Edges appear closer if the number of genes they share is higher and diameter of the nodes is proportional to the number of genes shared with pLH201.1. (A) Nodes are coloured depending on the taxonomy of the organisms; (B) nodes are coloured in green if at least one gene in the sequence is homologous with TnpLH201.1, otherwise they are red.

This confirmed that highly similar TnpLH201.1 homologues insert in diverse genomic environments.

Discussion

In response to the use of copper-based antimicrobial compounds to control plant bacterial pathogens, copper-resistant strains have emerged repeatedly in different parts of the world. Determinants of copper resistance have often been reported to be plasmid-borne, as in the case of *Xanthomonas citri* pv. *citri*, the causal agent of Asiatic citrus canker (Behlau *et al.* 2012). Until now, the understanding of the ecology of these resistance determinants and the evolution of the associated plasmids has been limited by the lack of genomic data. In the present study, we provide a comparative genomic analysis of plasmids associated with Cu^R in several *Xanthomonas* species, including *X. citri* pv. *citri*.

Multiple acquisitions of copper resistance in X. citri pv. citri

An unprecedented collection of *X. citri* pv. *citri* Cu^R strains from all the geographical areas where its emergence has been reported (Argentina, Réunion and Martinique) was genotyped using two complementary sets of markers (minisatellites and microsatellites). Cu^R strains were genetically related, so much so that they were assigned to a single lineage (DAPC1) and formed a single clonal complex based on minisatellite data. This lineage corresponded to the wide host range pathotype A, known to be responsible for the worldwide emergence of Asiatic citrus canker in the 20th century from Asia (i.e. its native origin) (Pruvost *et al.* 2014). Nevertheless, within this clonal complex, microsatellite analysis revealed a substantial differentiation into two genetic clusters likely with no epidemiological link. One cluster included Argentinian strains, while the other encompassed all strains from Réunion and Martinique. Genomic data confirmed the differentiation between French and Argentinian *X. citri* pv. *citri* strains (with intergroup genetic divergence tenfold higher than intragroup one, see Table S4, Supporting information for details). In Réunion, the genetic relatedness between Cu^R contemporary outbreak strains and Cu^S 'historical' strains isolated two to four decades ago suggested that the establishment of Cu^R strains on the island was unlikely to be the result of the recent introduction of genetically distinct strains. Instead, our data support the hypothesis that strains, which were genetically similar to the Cu^S populations characterized in the early years of the disease in Réunion, acquired a pLH201.1-like plasmid from a presently unknown source (i.e. Asiatic canker was reported for the first time in Réunion in

1968 – Brun (1971)). On the contrary, the weak genetic divergence between strains from Réunion and Martinique shows a possible epidemiological link between them (Fig. S5 and Table S4, Supporting information).

Contrasting with their epidemiological structure and genetic divergence, the Argentinian LM180 strain and the French ones displayed extremely similar plasmids (Fig. S5 and Table S4, Supporting information), suggesting independent copper-resistance acquisition by these two groups of strains, and showing the mobility of plasmid-encoded adaptive traits at very large geographical scales.

This mobility and the scenario of independent acquisition was further supported by the fact that in Réunion, a Cu^R *X. gardneri* strain (a pathogen of tomato and pepper) was found to carry a copy of pLH201.1 (average of 0.05 different nucleotides per kb) and a *X. euvesicatoria* pv. *allii* strain contained a contig in its genome (Gagnevin *et al.* 2014), which corresponded to the pLH201.1 backbone. The latter could not be circularized and, therefore, we were unable to confirm that it was a plasmid which lacked TnpLH201.1, despite the fact that the strain was PCR negative for *copLAB* and had a Cu^S phenotype. Finally, in Argentina, some Cu^R strains varying in *copLAB* PCR amplification were not genetically differentiated based on microsatellite data. This suggests the independent acquisition of two distinct copper-resistance systems within Argentinian lineages.

Several putative copper-resistance systems in X. citri pv. citri

For the Argentinian *X. citri* pv. *citri* A44 (LM180), Cu^R genes primarily shown to be experimentally functional comprise a transcriptional regulator (*copL*) and two copper-binding proteins (*copAB*) (Behlau *et al.* 2011). PCR tests provided evidence that most known Cu^R *X. citri* pv. *citri* strains possess this *copLAB* system. Using transposon mutagenesis, Behlau *et al.* (2011) demonstrated that the disruption of the *copLAB* genes was sufficient to lower the copper-resistance level to that of Cu^S strains. However, when inserted in a Cu^S strain of a closely related species, *X. 'perforans'*, the *copLAB* system alone did not confer the level of copper resistance of wild-type Cu^R strains. The authors suggested that this was due to the fact that their recipient strain had a different genetic background (Behlau *et al.* 2011). In 12 of 13 Cu^R sequenced strains, we identified an additional gene cluster that could be involved in copper resistance and may explain the partial phenotype restoration observed by Behlau *et al.* (2011). The HME-RND system, which is also present in *S. maltophilia* (Crossman *et al.* 2008) and forms a channel through the periplasm, is composed of an inner membrane pump (here CusA),

a periplasmic protein (CusB) and an outer membrane protein (SmmD) (Routh *et al.* 2011). The *cusAB/smmD* is not widespread among xanthomonads. It is not present in the draft genome of the strain complemented with *copLAB* by Behlau *et al.* for their functional analysis (data not shown) and, to date, has solely been detected from the draft genome of the Cu^R *X. vesicatoria* ATCC 35937 and mentioned in the NR database in *S. maltophilia* and *S. acidaminiphila*. Here, we added *X. citri* pv. *citri*, *X. gardneri*, *X. euvesicatoria* and *Stenotrophomonas* sp. to the list of species with both the *copLAB* and the *cusAB/smmD* systems.

In contrast to the majority of strains that have been studied, some *X. citri* pv. *citri* strains from Argentina were Cu^R and PCR positive for *copABCD* but not for *copLAB*. One of the strains (LM199 from Argentina) was sequenced and revealed that its plasmid hosts a different Tn3-like transposon (i.e. containing *copABCD*) in a genetically related backbone. Both the *copLAB* and the *copABCD* systems were reported from distinct strains of the walnut pathogen *X. arboricola* pv. *juglandis* on plasmid and chromosome, respectively (Lee *et al.* 1994; Behlau *et al.* 2013). Hence, we were able to establish that at least two distinct *cop* systems were associated with copper resistance in Argentinian *X. citri* pv. *citri*. Why polymorphism exists in copper-resistance systems is intriguing. Currently, we lack the necessary elements to test whether it is adaptive, that is occurs in response to environmental variations in copper concentration or fortuitous and driven by bioavailability.

The importance of HGT for the adaptation of genetically monomorphic bacteria

For all the studied *X. citri* pv. *citri* strains, copper-resistance systems were found on closely related plasmids of approximately 230 kb in size. Extensive annotation of the Réunion *X. citri* pv. *citri* plasmid pLH201.1 revealed that it bears all the genetic elements required for conjugation, confirming *in vitro* tests (data not shown) and previous data on strain A44 (LM180) from Argentina (Behlau *et al.* 2012). The pLH201.1 showed no strong homology to plasmids described previously. However, its relaxase is such that it belongs to the MOB_{H12} plasmid family. The content and organization of its conjugative apparatus clearly have similarities with IncA/C plasmids and SXT/R391 ICE. IncA/C plasmids have a very broad host range and are found in very diverse environments and geographical areas. Recently, an unknown MOB_{H12} plasmid from a marine environment, which also has similarities with IncA/C plasmids and SXT/R391, was reported (Nonaka *et al.* 2012). This suggests that the MOB_{H12} plasmid family could be wider than previously thought.

Plasmids can confer a broad range of adaptive traits to their host, such as antibiotic resistance (Ochman *et al.* 2000), heavy metal resistance (Hobman & Crossman 2015), UV tolerance, hormone production, pathogenicity determinants and toxin production (Vivian *et al.* 2001; Sundin 2007). These adaptations can lead to the colonization of new ecological niches. They may even be responsible for major evolutionary events, such as the emergence of new pathogenic populations. For example, different allelic forms of the pPATH plasmid have transformed strains of the commensal bacterial species *Pantoea agglomerans* into gall-forming pathogens of gypsophila and beet (Weinthal *et al.* 2007). Our results strongly support the acquisition of a new adaptive phenotype through plasmid incorporation by different *X. citri* pv. *citri* populations. However, the mechanisms of xanthomonad adaptation through HGT might not be restricted to plasmid acquisition. Indeed, chromosomally encoded resistance was reported on *Stenotrophomonas* and xanthomonads causing bacterial spot of tomato and pepper or bacterial blight of walnut. In addition, several genomic islands, including genes of plasmid origin, were detected on the chromosome of *X. citri* pv. *citri* (Gordon *et al.* 2015).

Importance of plasmid-borne mobile genomic elements

In the present study, we provide evidence that copper-resistance gene clusters on pLH201.1 were encoded on Tn3-like transposon (referred to here as TnpLH201.1). Tn3-like transposons have been reported for other plasmids in *Xanthomonas* (Ferreira *et al.* 2015; Niu *et al.* 2015), as well as other genera (Lauf *et al.* 1998). IncA/C plasmids often carry a complex transposon-based cluster of resistance genes involved in the spread of multidrug resistance between bacteria (Harmer & Hall 2014).

This feature could mitigate the apparently limited host range of pLH201.1, by providing a second layer of mobility. Indeed, three distinct species (*X. euvesicatoria*, *S. maltophilia* and *Stenotrophomonas* sp.) harbour a transposon almost identical to the one hosting the *copLAB* gene system in *X. citri* pv. *citri* (>99% NI) in a genomic environment that is markedly different from that of pLH201.1 (i.e. a 179-kb plasmid for *X. euvesicatoria* and the chromosome for the two other species). This supports the hypothesis that the transposon is a source of mobility for the Cu^R gene cluster. Moreover, the *copABCD* system found on LM199 has 98% NI with that encoded on the chromosome of *X. arboricola* pv. *juglandis*, while encoded on a pLH201.1-related plasmid (NI of 90% on 85% of pLM199 length, see Fig. 2 and Table S4, Supporting information). The pLM199 comprised all the 16 genes from the pLH201.1 conjugative

gene set. In this regard, the transposon TnpLH201.1 can be considered as an autonomous vehicle. Indeed, it encodes for Cu^R proteins, transcriptional regulators, a transposition apparatus and a single tRNA. The latter, which is required for the transcription of the genes encoded on the transposon, is absent on the chromosome.

To lower the fitness cost of plasmid carriage, chromosomal genome and plasmids co-evolve (Harrison & Brockhurst 2012). This process could limit the spread of entire alien plasmids and, instead, favour the incorporation of the transposon into plasmids that are already present within a restricted taxonomic group and therefore already adapted to their host.

Barriers to HGT and importance of reservoir bacteria

Optima of genome functioning leave strong imprints, such as GC content and codon usage. These differences tend to limit exchange of DNA between distantly related bacteria (Popa & Dagan 2011). Indeed, within networks of shared DNA among bacterial genomes (with 95% NI), *Xanthomonas* species tend to form an isolated cluster. Only plasmids with lower NI within the species connect *Xanthomonas* to some distantly related bacterial genera (Halary *et al.* 2010). Our network approach with pLH201.1 yielded similar results. We identified complete or nearly complete pLH201.1 homologues, as well as genes involved in Cu^R or conjugation, primarily in the Xanthomonadaceae family.

The present study has provided evidence that distantly related *Xanthomonas* species (e.g. the tomato pathogen *X. gardneri* and the citrus pathogen *X. citri* pv. *citri*) carry the same plasmid. *Xanthomonas* is a bacterial genus largely composed of plant pathogenic bacteria with a high degree of host specialization (Leyns *et al.* 1984). As *Xanthomonas* lineages that have a different host range colonize distinct ecological niches, they would not be expected to share DNA directly through conjugation because it requires cell-to-cell contact. However, in some agricultural contexts (intercropping, for example), the physical proximity of plant species contaminated with distinct bacterial pathogens could facilitate contact and HGT. In addition, extreme weather events that have been reported to spread bacterial cells over long distances (Irey *et al.* 2006) may contribute to the mixing of xanthomonad populations. A key factor for gene transfer between populations probably lies in the ability of *Xanthomonas* to survive transiently on plant surfaces, in natural plant openings or even on nonhost plant species (Robinson & Callow 1986). In fact, xanthomonads were reported to form mixed-biofilm structures on plant surfaces (Jacques *et al.* 2005; Cubero *et al.* 2011), which have been recognized as highly

favourable to HGT within the phyllosphere (Van Elsas *et al.* 2003).

Indirect transfer of copper resistance between xanthomonads may occur. Different reservoirs of bacteria resistant to antimicrobials can be found in different environmental compartments that interact and share interfaces (Nesme *et al.* 2014). Indeed, a resistome to environmental or industrial copper does exist and combines different genes associated with copper resistance (He *et al.* 2010). By tracking the dispersal and availability of this type of resistance in the natural environment or agro-ecosystem and linking it to other settings, we should be able to understand and predict how the ecosystem functions (Vieites *et al.* 2009).

Following sporadic reporting over several decades (Vauterin *et al.* 1996), commensal xanthomonads are now being more carefully characterized in terms of taxonomy or taxonomic placement (Triplett *et al.* 2015) or pathogenicity gene repertoires and mobile genetic elements (Cesbron *et al.* 2015). The extent to which these commensal *Xanthomonas* strains or commensal bacteria act as reservoirs or hubs for adaptive genes is still unknown. In the context of increased HGT frequencies between phylogenetically related species, the significance of *Stenotrophomonas* (and other genera in the Xanthomonadaceae family) as a major source of adaptive genes for xanthomonads in agricultural ecosystems has largely been underrated. At least two commensal *Stenotrophomonas* species displayed a highly identical copy of TnpLH201.1. Despite its relative individual insignificance as a pathogen, *S. maltophilia* is of major relevance in terms of plant, animal and human health because it constitutes a gene reservoir that is available for gene transfer within the community. Indeed, the panoply of resistance genes that it harbours could provide a source of antibacterial resistance determinants that are transferable to bacterial pathogens, such as the copper-resistance system presented here or other types of resistance relating to human health reported previously (Crossman *et al.* 2008). Our study highlights the importance of conducting further research on entire microbial communities in order to improve our understanding of the emergence of pathogenic bacteria.

Acknowledgements

We would like to thank B. Barrès, B. Doublet and P. Roumagnac for their helpful contribution to discussions, as well as all those who provided help during field prospection. The European Regional Development Fund (ERDF) and European Agricultural Fund for Rural Development (EAFRD), Conseil Départemental de la Réunion, Région Réunion, État français, the French Agropolis Foundation (Labex Agro – Montpellier,

2146 D. RICHARD ET AL.

E-SPACE project number 1504-004), ANSES and CIRAD provided financial support.

References

- Achtman M (2008) Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annual Review of Microbiology*, **62**, 53–70.
- Achtman M (2012) Insights from genomic comparisons of genetically monomorphic bacterial pathogens. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, **367**, 860–867.
- Alvarado A, Garcillan-Barcia MP, de la Cruz F (2012) A degenerate primer MOB typing (DPMT) method to classify gamma-proteobacterial plasmids in clinical and environmental settings. *PLoS One*, **7**, e40438.
- Beaber JW, Hochhut B, Waldor MK (2002) Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. *Journal of Bacteriology*, **184**, 4259–4269.
- Behlau F, Canteros BI, Minsavage GV, Jones JB, Graham JH (2011) Molecular Characterization of Copper Resistance Genes from *Xanthomonas citri* subsp. *citri* and *Xanthomonas alfalfae* subsp. *citrumelonis*. *Applied and Environmental Microbiology*, **77**, 4089–4096.
- Behlau F, Canteros BI, Jones JB, Graham JH (2012) Copper resistance genes from different xanthomonads and citrus epiphytic bacteria confer resistance to *Xanthomonas citri* subsp. *citri*. *European Journal of Plant Pathology*, **133**, 949–963.
- Behlau F, Hong JC, Jones JB, Graham JH (2013) Evidence for acquisition of copper resistance genes from different sources in citrus-associated xanthomonads. *Phytopathology*, **103**, 409–418.
- Brown NL, Barrett SR, Camakaris J, Lee BT, Rouch DA (1995) Molecular genetics and transport analysis of the copper-resistance determinant (pco) from *Escherichia coli* plasmid pRJ1004. *Molecular Microbiology*, **17**, 1153–1166.
- Brown NL, Stoyanov JV, Kidd SP, Hobman JL (2003) The MerR family of transcriptional regulators. *FEMS Microbiology Reviews*, **27**, 145–163.
- Brun J (1971) Le chancre bactérien des Citrus. *Fruits*, **26**, 533–540.
- Bryant J, Chewapreecha C, Bentley SD (2012) Developing insights into the mechanisms of evolution of bacterial pathogens from whole-genome sequences. *Future Microbiology*, **7**, 1283–1296.
- Bui Thi Ngoc L, Verniere C, Vital K et al. (2009) Development of 14 minisatellite markers for the citrus canker bacterium, *Xanthomonas citri* pv. *citri*. *Molecular Ecology Resources*, **9**, 125–127.
- Cai R, Lewis J, Yan S et al. (2011) The plant pathogen *Pseudomonas syringae* pv. *tomato* is genetically monomorphic and under strong selection to evade tomato immunity. *PLoS Pathogens*, **7**, e1002130.
- Carraro N, Sauve M, Matteau D et al. (2014) Development of pVCR94DeltaX from *Vibrio cholerae*, a prototype for studying multidrug resistant IncA/C conjugative plasmids. *Frontiers in Microbiology*, **5**, 44.
- Cesbron S, Briand M, Essakhi S et al. (2015) Comparative genomics of pathogenic and nonpathogenic strains of *Xanthomonas arboricola* unveil molecular and evolutionary events linked to pathoadaptation. *Frontiers in Plant Science*, **6**, 1126.
- Chaudhuri RR, Sebaihia M, Hobman JL et al. (2010) Complete genome sequence and comparative metabolic profiling of the prototypical Enterococcal *Escherichia coli* strain 042. *PLoS One*, **5**, e8801.
- Chiarini L, Bevivino A, Dalmastrì C, Tabacchioni S, Visca P (2006) *Burkholderia cepacia* complex species: health hazards and biotechnological potential. *Trends in Microbiology*, **14**, 277–286.
- Constantin EC, Cleenwerck I, Maes M et al. (2016) Genetic characterization of strains named as *Xanthomonas axonopodis* pv. *dieffenbachiae* leads to a taxonomic revision of the *X. axonopodis* species complex. *Plant Pathology*, **65**, 792–806.
- Cooksey DA, Azad HR, Cha JS, Lim CK (1990) Copper resistance gene homologs in pathogenic and saprophytic bacterial species from tomato. *Applied and Environment Microbiology*, **56**, 431–435.
- Crossman LC, Gould VC, Dow JM et al. (2008) The complete genome, comparative and functional analysis of *Stenotrophomonas maltophilia* reveals an organism heavily shielded by drug resistance determinants. *Genome Biology*, **9**, R74.
- Cubero J, Gell I, Johnson EG, Redondo A, Graham JH (2011) Unstable green fluorescent protein for study of *Xanthomonas citri* subsp. *citri* survival on citrus. *Plant Pathology*, **60**, 977–985.
- Daccord A, Ceccarelli D, Burrus V (2010) Integrating conjugative elements of the SXT/R391 family trigger the excision and drive the mobilization of a new class of *Vibrio* genomic islands. *Molecular Microbiology*, **78**, 576–588.
- Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, **14**, 1394–1403.
- Davenport KW, Daligault HE, Minogue TD et al. (2014) Complete genome sequence of *Stenotrophomonas maltophilia* type strain 810-2 (ATCC 13637). *Genome Announcements*, **2**, e00974–e00914.
- Davies J, Davies D (2010) Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews*, **74**, 417–433.
- Dubey GP, Ben-Yehuda S (2011) Intercellular nanotubes mediate bacterial communication. *Cell*, **144**, 590–600.
- Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, **1**, 47–50.
- Fernandez-Alarcon C, Singer RS, Johnson TJ (2011) Comparative genomics of multidrug resistance-encoding IncA/C plasmids from commensal and pathogenic *Escherichia coli* from multiple animal sources. *PLoS One*, **6**, e23415.
- Ferreira RM, de Oliveira AC, Moreira LM et al. (2015) A TALE of transposition: Tn3-like transposons play a major role in the spread of pathogenicity determinants of *Xanthomonas citri* and other xanthomonads. *MBio*, **6**, e02505–e02514.
- Francisco AP, Vaz C, Monteiro PT et al. (2012) PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics*, **13**, 87.
- Fricke WF, Welch TJ, McDermott PF et al. (2009) Comparative genomics of the IncA/C multidrug resistance plasmid family. *Journal of Bacteriology*, **191**, 4750–4757.
- Gagnevin L, Bolot S, Gordon JL et al. (2014) Draft genome sequence of *Xanthomonas axonopodis* pv. *allii* strain CFBP 6369. *Genome Announcements*, **2**, e00727–e00714.

COPPER RESISTANCE IN MONOMORPHIC BACTERIA 2147

- Galimand M, Guiyoule A, Gerbaud G *et al.* (1997) Multidrug resistance in *Yersinia pestis* mediated by a transferable plasmid. *New England Journal of Medicine*, **337**, 677–680.
- Garcia-Vallve S, Romeu A, Palau J (2000) Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Research*, **10**, 1719–1725.
- Garcillan-Barcia MP, Francia MV, de la Cruz F (2009) The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiology Reviews*, **33**, 657–687.
- Giovanardi D, Bonneau S, Gironde S *et al.* (2016) Morphological and genotypic features of *Xanthomonas arboricola* pv. *juglandis* populations from walnut groves in Romagna region, Italy. *European Journal of Plant Pathology*, **145**, 1–16.
- Gomis-Ruth FX, Moncalian G, de la Cruz F, Coll M (2002) Conjugative plasmid protein TrwB, an integral membrane type IV secretion system coupling protein. Detailed structural features and mapping of the active site cleft. *Journal of Biological Chemistry*, **277**, 7556–7566.
- Gordon JL, Lefeuve P, Escalon A *et al.* (2015) Comparative genomics of 43 strains of *Xanthomonas citri* pv. *citri* reveals the evolutionary events giving rise to pathotypes with different host ranges. *Bmc Genomics*, **16**, 1098.
- Gottwald TR, Bassanezi RB, Amorim L, Bergamin-Filho A (2007) Spatial pattern analysis of citrus canker-infected plantings in Sao Paulo, Brazil, and augmentation of infection elicited by the Asian leafminer. *Phytopathology*, **97**, 674–683.
- Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**, 184–186.
- Graham JH, Gottwald TR, Cubero J, Achor DS (2004) *Xanthomonas axonopodis* pv. *citri*: factors affecting successful eradication of citrus canker. *Molecular Plant Pathology*, **5**, 1–15.
- Grigoriev A (1998) Analyzing genomes with cumulative skew diagrams. *Nucleic Acids Research*, **26**, 2286–2290.
- Guy L, Kultima JR, Andersson SG (2010) genoPlotR: comparative gene and genome visualization in R. *Bioinformatics*, **26**, 2334–2335.
- Haines AS, Jones K, Batt SM, Kosheleva IA, Thomas CM (2007) Sequence of plasmid pBS228 and reconstruction of the IncCP-1 alpha phylogeny. *Plasmid*, **58**, 76–83.
- Halary S, Leigh JW, Cheaib B, Lopez P, Baptiste E (2010) Network analyses structure genetic diversity in independent genetic worlds. *Proceedings of the National Academy of Sciences, USA*, **107**, 127–132.
- Harmer CJ, Hall RM (2014) pRMH760, a precursor of A/C(2) plasmids carrying blaCMY and blaNDM genes. *Microbial Drug Resistance*, **20**, 416–423.
- Harrison E, Brockhurst MA (2012) Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends in Microbiology*, **20**, 262–267.
- He LY, Zhang YF, Ma HY *et al.* (2010) Characterization of copper-resistant bacteria and assessment of bacterial communities in rhizosphere soils of copper-tolerant plants. *Applied Soil Ecology*, **44**, 49–55.
- Heinemann JA, Sprague GF (1989) Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature*, **340**, 205–209.
- Hobman JL, Crossman LC (2015) Bacterial antimicrobial metal ion resistance. *Journal of Medical Microbiology*, **64**, 471–497.
- Hurtle W, Lindler L, Fan W *et al.* (2003) Detection and identification of ciprofloxacin-resistant *Yersinia pestis* by denaturing high-performance liquid chromatography. *Journal of Clinical Microbiology*, **41**, 3273–3283.
- Hyatt D, Chen GL, Locascio PF *et al.* (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, **11**, 119.
- Irey M, Gottwald TR, Graham JH, Riley TD, Carlton G (2006) Post-hurricane analysis of citrus canker spread and progress towards the development of a predictive model to estimate disease spread due to catastrophic weather events. *Plant Health Progress*, doi:10.1094/PHP-2006-0822-1001-RS.
- Jacques MA, Josi K, Darrasse A, Samson R (2005) *Xanthomonas axonopodis* pv. *phaseoli* var. *fusca* is aggregated in stable biofilm population sizes in the phyllosphere of field-grown beans. *Applied and Environment Microbiology*, **71**, 2008–2015.
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780.
- Krzywinski M, Schein J, Birol I *et al.* (2009) Circos: an information aesthetic for comparative genomics. *Genome Research*, **19**, 1639–1645.
- Lauf U, Muller C, Herrmann H (1998) The transposable elements resident on the plasmids of *Pseudomonas putida* strain H, Tn5501 and Tn5502, are cryptic transposons of the Tn3 family. *Molecular and General Genetics*, **259**, 674–678.
- Leduc A, Traore YN, Boyer K *et al.* (2015) Bridgehead invasion of a monomorphic plant pathogenic bacterium: *Xanthomonas citri* pv. *citri*, an emerging citrus pathogen in Mali and Burkina Faso. *Environmental Microbiology*, **17**, 4429–4442.
- Lee YA, Henderson M, Panopoulos NJ, Schroth MN (1994) Molecular-cloning, chromosomal mapping, and sequence analysis of copper resistance genes from *Xanthomonas campestris* pv. *juglandis* – homology with small blue copper proteins and multicopper oxidase. *Journal of Bacteriology*, **176**, 173–188.
- Leyns F, De Cleene M, Swings JG, De Ley J (1984) The host range of the genus *Xanthomonas*. *Botanical Review*, **50**, 308–356.
- Mellano MA, Cooksey DA (1988) Nucleotide sequence and organization of copper resistance genes from *Pseudomonas syringae* pv. *tomato*. *Journal of Bacteriology*, **170**, 2879–2883.
- Nesme J, Cecillon S, Delmont TO *et al.* (2014) Large-scale metagenomic-based study of antibiotic resistance in the environment. *Current Biology*, **24**, 1096–1100.
- Niu XN, Wei ZQ, Zou HF *et al.* (2015) Complete sequence and detailed analysis of the first indigenous plasmid from *Xanthomonas oryzae* pv. *oryzicola*. *Bmc Microbiology*, **15**, 233.
- Nonaka L, Maruyama F, Miyamoto M *et al.* (2012) Novel conjugative transferable multiple drug resistance plasmid pAQU1 from *Photobacterium damsela* subsp. *damsela* isolated from marine aquaculture environment. *Microbes and Environments*, **27**, 263–272.
- Norberg P, Bergstrom M, Jethava V, Dubhashi D, Hermansson M (2011) The IncP-1 plasmid backbone adapts to different host bacterial species and evolves through homologous recombination. *Nature Communications*, **2**, 268.
- Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature*, **405**, 299–304.

- Pak TR, Altman DR, Attie O *et al.* (2015) Whole-genome sequencing identifies emergence of a quinolone resistance mutation in a case of *Stenotrophomonas maltophilia* bacteremia. *Antimicrobial Agents and Chemotherapy*, **59**, 7117–7120.
- Pansegau W, Lanka E, Barth PT *et al.* (1994) Complete nucleotide sequence of Birmingham IncP-Alpha plasmids – compilation and comparative analysis. *Journal of Molecular Biology*, **239**, 623–663.
- Pele J, Becu JM, Abdi H, Chabbert M (2012) Bios2mds: an R package for comparing orthologous protein families by metric multidimensional scaling. *BMC Bioinformatics*, **13**, 133.
- Pereira UP, Gouran H, Nascimento R *et al.* (2015) Complete genome sequence of *Xanthomonas arboricola* pv. *juglandis* 417, a copper-resistant strain isolated from *Juglans regia* L. *Genome Announc*, **3**, e01126–01115.
- Perry JA, Wright GD (2013) The antibiotic resistance “mobilome”: searching for the link between environment and clinic. *Frontiers in Microbiology*, **4**, 138.
- Popa O, Dagan T (2011) Trends and barriers to lateral gene transfer in prokaryotes. *Current Opinion in Microbiology*, **14**, 615–623.
- Popescu AA, Huber KT, Paradis E (2012) ape 3.0: new tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics*, **28**, 1536–1537.
- Pruvost O, Magne M, Boyer K *et al.* (2014) A MLVA genotyping scheme for global surveillance of the citrus pathogen *Xanthomonas citri* pv. *citri* suggests a worldwide geographical expansion of a single genetic lineage. *PLoS One*, **9**, e98129.
- Richard D, Boyer C, Javegny S *et al.* (2016) First report of *Xanthomonas citri* pv. *citri* pathotype A causing asiatic citrus canker in Martinique, France. *Plant Disease*, **100**, 1946.
- Richard D, Tribot N, Boyer C *et al.* (2017) First report of copper-resistant *Xanthomonas citri* pv. *citri* pathotype A causing Asiatic citrus canker in Réunion, France. *Plant Disease*, doi:10.1094/PDIS-1009-1016-1387-PDN.
- Robinson JN, Callow JA (1986) Multiplication and spread of pathovars of *Xanthomonas campestris* in host and non-host plants. *Plant Pathology*, **35**, 169–177.
- Routh MD, Zalucki Y, Su CC *et al.* (2011) Efflux pumps of the resistance-nodulation-division family: a perspective of their structure, function, and regulation in gram-negative bacteria. *Advances in Enzymology and Related Areas of Molecular Biology*, **77**, 109–146.
- Schuessler DL, Cortes T, Fivian-Hughes AS *et al.* (2013) Induced ectopic expression of HigB toxin in *Mycobacterium tuberculosis* results in growth inhibition, reduced abundance of a subset of mRNAs and cleavage of tmRNA. *Molecular Microbiology*, **90**, 195–207.
- Seiler C, Berendonk TU (2012) Heavy metal driven co-selection of antibiotic resistance in soil and water bodies impacted by agriculture and aquaculture. *Frontiers in Microbiology*, **3**, 399.
- Sen DY, Brown CJ, Top EM, Sullivan J (2013) Inferring the evolutionary history of IncP-1 plasmids despite incongruence among backbone gene trees. *Molecular Biology and Evolution*, **30**, 154–166.
- Sevin EW, Barloy-Hubler F (2007) RASTA-Bacteria: a web-based tool for identifying toxin-antitoxin loci in prokaryotes. *Genome Biology*, **8**, R155.
- Shao Y, Harrison EM, Bi D *et al.* (2011) TADB: a web-based resource for Type 2 toxin-antitoxin loci in bacteria and archaea. *Nucleic Acids Research*, **39**, D606–D611.
- Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M (2006) ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Research*, **34**, D32–D36.
- Smillie C, Garcillan-Barcia MP, Francia MV, Rocha EPC, de la Cruz F (2010) Mobility of plasmids. *Microbiology and Molecular Biology Reviews*, **74**, 434.
- Smith IM (1985) *Fungicides for crop protection: 100 years of progress: proceedings of the Bordeaux Mixture Centenary Meeting: Bordeaux, France, 5th-7th September 1985*, Bordeaux, France.
- Sommer DD, Delcher AL, Salzberg SL, Pop M (2007) Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics*, **8**, 64.
- Stoyanov JV, Hobman JL, Brown NL (2001) CueR (YbbI) of *Escherichia coli* is a MerR family regulator controlling expression of the copper exporter CopA. *Molecular Microbiology*, **39**, 502–511.
- Sundin GW (2007) Genomic insights into the contribution of phytopathogenic bacterial plasmids to the evolutionary history of their hosts. *Annual Review of Phytopathology*, **45**, 129–151.
- Telenti A, Imboden P, Marchesi F *et al.* (1993) Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. *Lancet*, **341**, 647–650.
- Tennstedt T, Szczepanowski R, Krahn I, Puhler A, Schluter A (2005) Sequence of the 68,869 bp IncP-1 alpha plasmid pTB11 from a waste-water treatment plant reveals a highly conserved backbone, a Tn402-like integron and other transposable elements. *Plasmid*, **53**, 218–238.
- Trefault N, De la Iglesia R, Molina AM *et al.* (2004) Genetic organization of the catabolic plasmid pJP4 from *Ralstonia eutropha* JMP134 (pJP4) reveals mechanisms of adaptation to chloroaromatic pollutants and evolution of specialized chloroaromatic degradation pathways. *Environmental Microbiology*, **6**, 655–668.
- Triplett LR, Verdier V, Campillo T *et al.* (2015) Characterization of a novel clade of *Xanthomonas* isolated from rice leaves in Mali and proposal of *Xanthomonas maliensis* sp. nov. *Antonie van Leeuwenhoek*, **107**, 869–881.
- Tsuda M, Iino T (1988) Identification and characterization of Tn4653, a transposon covering the toluene transposon Tn4651 on TOL plasmid pWW0. *Molecular and General Genetics*, **213**, 72–77.
- Vallenet D, Labarre L, Rouy Z *et al.* (2006) MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Research*, **34**, 53–65.
- Van Elsas JD, Turner S, Bailey MJ (2003) Horizontal gene transfer in the phytosphere. *New Phytologist*, **157**, 525–537.
- Vauterin L, Yang P, Alvarez A *et al.* (1996) Identification of non-pathogenic *Xanthomonas* strains associated with plants. *Systematic and Applied Microbiology*, **19**, 96–105.
- Vieites JM, Guazzaroni ME, Belouqui A, Golyshin PN, Ferrer M (2009) Metagenomics approaches in systems microbiology. *FEMS Microbiology Reviews*, **33**, 236–255.
- Vivian A, Murillo J, Jackson RW (2001) The roles of plasmids in phytopathogenic bacteria: mobile arsenals? *Microbiology*, **147**, 763–780.
- Wang G, Kennedy SP, Fasiludeen S, Rensing C, DasSarma S (2004) Arsenic resistance in *Halobacterium* sp. strain NRC-1 examined by using an improved gene knockout system. *Journal of Bacteriology*, **186**, 3187–3194.
- Weinthal DM, Barash I, Panijel M *et al.* (2007) Distribution and replication of the pathogenicity plasmid pPATH in diverse

- populations of the gall-forming bacterium *Pantoea agglomerans*. *Applied and Environment Microbiology*, **73**, 7552–7561.
- Wozniak RA, Waldor MK (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nature Reviews Microbiology*, **8**, 552–563.
- Yano H, Genka H, Ohtsubo Y *et al.* (2013) Cointegrate-resolution of toluene-catabolic transposon Tn4651: determination of crossover site and the segment required for full resolution activity. *Plasmid*, **69**, 24–35.
- Zhang WJ, Xu XR, Schwarz S *et al.* (2014) Characterization of the IncA/C plasmid pSECE2 from *Escherichia coli* of swine origin that harbours the multiresistance gene *cfr*. *Journal of Antimicrobial Chemotherapy*, **69**, 385–389.

Data accessibility

The MLVA-31 and MLVA-14 data generated in this study are available in the *Xanthomonas citri* genotyping database (<http://www.biopred.net/MLVA/>) and at <https://agritrop.cirad.fr/582853/>, respectively. Sequences produced in this study are deposited in the GenBank database (Table 1).

P.L., V.R., C.V. and O.P. conceived and designed the study. B.I.C. provided essential bacterial strains and produced phenotypic data on Argentinian strains. D.R., C.B., K.B., P.G., S.J., M.T., B.I.C., C.V. and O.P. carried out the experiments. O.P., C.V. and V.R. conducted population biology analyses. D.R. and P.L. conducted genomic data analyses. B.F., I.R. and A.R. helped to design and carry out the study. D.R., V.R., A.R., B.F., I.R., A.C., C.V., O.P. and P.L. wrote the manuscript. All the authors have read and approved the final manuscript.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Categorical minimum spanning tree of DAPC1 strains from Argentina, Martinique and Réunion, which differ in their susceptibility to copper (268 strains – 34 haplotypes), based on minisatellite data.

Fig. S2 Categorical minimum spanning tree of copper-resistant *Xanthomonas citri* pv. *citri* strains from Réunion based on microsatellite data.

Fig. S3 Alignments between TnpLH201.1 and sequences carrying TnpLH201.1 homologues from strains sequenced in this study and *S. maltophilia* K279a.

Fig. S4 Distance tree of the copper-resistance region of all the sequenced strains as well as some sequences coding for known copper-resistance systems extracted from public database along with and graphical representations of the copper-resistance genes organisation.

Fig. S5 Heatmap representation of the genetic divergence between six sequenced strains.

Fig. S6 Clusters of sequences with homologues to pLH201.1 genes.

Table S1 Size (in bp) of the contigs obtained for each of the 13 sequenced strains

Table S2 plasmids and ICE having similar conjugative apparatus with pLH201.1.

Table S3 Primers and PCR conditions used for amplification of the copABCD system of *Xanthomonas citri* pv. *citri*.

Table S4 Genetic divergence over the homologous sequence tracks defined in Fig. 2.

Chapitre II La génomique pour caractériser l'histoire évolutive de Xcc dans l'Océan Indien

Chapitre II La génomique pour caractériser l'histoire évolutive de Xcc dans l'Océan Indien

Chez les souches de Xanthomonadaceae étudiées dans le Chapitre I, la relative conservation du transposon portant les gènes de résistance au cuivre, la variabilité de son environnement génomique ainsi que le caractère composite des plasmides qui le portent ont montré la nature variable du contenu en gènes et le rôle important des transferts horizontaux (HGT) dans l'adaptation de Xcc.

Chez les bactéries, la caractérisation des HGT a souvent été liée à l'acquisition d'un phénotype nouveau, comme cela a été le cas dans ce Chapitre I. Les événements de transferts de gènes entre bactéries sont pourtant courants et ne se limitent pas aux cas connus d'intégration d'éléments mobiles conférant un phénotype nouveau : ils font partie intégrante de l'évolution des génomes bactériens. Au regard du nombre d'études basées sur les SNP pour décrire l'évolution bactérienne, la prise en compte du contenu en gènes reste encore rare (Holt *et al.* 2013; Zhou *et al.* 2013; Zhou *et al.* 2014), tout comme les méthodes de séquençage et d'analyse de données de séquençage pour y parvenir. Des premières études sur des modèles phytopathogènes ont été publiées récemment (Chen *et al.* 2018; Jibrin *et al.* 2018; McCann *et al.* 2017). Dans le cas de Xcc, les variations de contenu en gène et leur importance dans l'histoire évolutive sont encore peu étudiées.

Alors que des méthodes de génotypage basées sur des répétitions en tandem comportent des limitations liées à des phénomènes d'homoplasie¹, l'étude de l'intégralité des SNP issus de séquençages complets permet de représenter les relations phylogénétiques entre individus de manière bien plus précise. L'étude concomitante du contenu en gène permet de caractériser l'importance de ce type de variation au regard de la mutation.

C'est ainsi que le Chapitre II est consacré à l'analyse de l'histoire évolutive de Xcc à l'échelle de l'océan Indien, en prenant en compte plusieurs formes d'évolution génomiques. Pour ce faire nous avons procédé au séquençage complet d'une collection de souches historiques et contemporaines de la Réunion, des îles voisines ainsi que du reste du monde (**Publication 2**).

¹ L'homoplasie est l'observation de deux états alléliques identiques qui ne provient pas d'un ancêtre commun. Dans les cas des marqueurs microsatellites, la version allélique correspond à l'ajout ou la perte d'un motif répété. Un locus microsatellite peut comprendre n motifs peut muter à n+1 et ensuite retourner à n au cours de son histoire évolutive, masquant un signal phylogénétique.

Dans la diversité mondiale, les souches de la région Sud-Ouest de l'Océan Indien (SWIO) forment un clade monophylétique. Au sein de ce clade, nous avons montré l'existence d'une accumulation progressive des SNP dans le temps, ce qui nous a permis d'en dater tous les nœuds. L'introduction de la bactérie dans la zone SWIO a pu être datée au début du 19^{ème} siècle. Cette période correspond à celle d'expéditions menées par le botaniste Pierre Poivre pour introduire de nombreuses espèces végétales, dont plusieurs espèces d'arbres fruitiers, dans l'archipel des Mascareignes en provenance de différentes régions d'Asie. Elle est également concomitante de flux migratoires d'un grand nombre de travailleurs asiatiques vers la région. L'analyse du contenu en gènes des souches SWIO a révélé la présence de 699 gènes « accessoires » dont la moitié a un support plasmidique. Le séquençage de souches Réunionnaises phylogénétiquement et géographiquement très proches nous a permis de conclure que des variations de contenu en gène sont observées à des échelles très réduites. En effet, la présence de gènes variables entre souches très apparentées suggère un taux d'apparition élevé. Nous avons également mis en évidence la présence d'une variabilité génétique élevée à l'échelle du verger, aussi bien en termes de SNP que de variations de contenu en gènes. Cette faible structuration spatiale est probablement en lien avec l'existence d'évènements d'introduction multiples de la bactérie dans les vergers et/ou des inocula polyclonaux associés à chaque évènement. Le mosaïsme de plusieurs plasmides souvent impliqués dans le caractère pathogène de Xcc a été observé. Il souligne l'importance des évènements de recombinaison dans l'évolution de ce compartiment génomique. Les évènements de perte ou de gain de gènes chromosomiques dans le chromosome ne comprennent quant à eux le plus souvent qu'un nombre faible de gènes. Lorsque ces variations concernent un nombre plus important de gènes, ceux-ci sont regroupés dans quelques régions génomiques qui présentent parfois des similarités avec des ICE.

Nous supposons que le dynamisme du contenu en gène a un rôle d'autant plus important dans l'évolution qu'il est observé chez une bactérie monomorphe, c'est-à-dire présentant un taux de mutation faible.

Population-genomic insights into the evolution of the crop pathogenic bacterium *Xanthomonas citri* pv. *citri*

D. Richard^{1,2,3,†}, O. Pruvost¹, C. Boyer¹, A. Rieux¹, P. Lefeuvre¹

Affiliations:

¹ Cirad, UMR PVBMT, F-97410 St Pierre, Réunion, France ; ² ANSES, Plant Health Laboratory, F-97410 St Pierre, Réunion, France; ³ Université de la Réunion, UMR PVBMT, F-97490 St Denis, Réunion, France ; [†]Corresponding author: damien.richard@cirad.fr

Address: Pôle de Protection des Plantes, UMR PVBMT, Station Ligne Paradis, 7 chemin de l'Irat, 97410 Saint Pierre, Réunion, France

Abstract

Various biological processes generate genetic variability and allow bacterial evolution. Among them, SNP appear in the genome in a time-dependent manner. Based on the amount of changes separating heterochronous samples, tip-dating analysis can be used to infer the substitution rate of an organism and to date nodes within a phylogeny. Horizontal gene transfer also plays a crucial role in bacterial evolution and speciation as it allows the exchange and spread of adaptive traits between species. Focusing on an epidemic *Xanthomonas citri* pv. *citri* (Xcc) lineage in the South West Indian Ocean, we herein evaluate the relative contributions of two different forms of molecular evolution: SNP and gene content variation. To do so, we sequenced 222 phylogenetically related phytopathogenic Xcc strains originating from one Caribbean and ten SWIO islands and spanning 39 years of evolution. We dated the introduction of Xcc in the SWIO region at -200 years and hypothesized that some historical migration events could be causal. We evidenced high levels of gene content variations even at very small phylogenetic and geographic scales and associated this finding to Xcc's mode of dissemination. Whereas at a global scale SNP evolution appears as predominant, gene content variation appeared as the driver of Xcc diversity at more local scale. In the light of our results, it seems crucial to take into account gene content variation in order to depict more thoroughly the evolutionary history of monomorphic bacteria.

Keywords: accessory genome, mobile DNA, molecular clock, microevolution, *Xanthomonas citri* pv. *citri*

Introduction

Evolution refers to the accumulation of genetic variation over time. This process drives speciation and adaptation. Strains that belong to the so-called genetically monomorphic bacteria display very low levels of genetic polymorphism and often appear genetically uniform when only a few loci are analyzed (*e.g.*, when using MultiLocus Sequence Typing, MLST). Genome-wide analyses are required for reliably understanding their structure (Achtman 2008). Nonetheless, their adaptive potential is highlighted by the success of several human pathogens such as *Salmonella enterica* serovar Typhi, *Mycobacterium tuberculosis* or *Yersinia pestis*, which showed evidences of adaptive evolution caused by different types of genetic polymorphisms (Achtman 2008; Eldholm & Balloux 2016). Indeed, bacterial genomes not only experience mutations (*sensu largo*, including single nucleotide polymorphism but also nucleotide insertion/deletion), but also recombination, which can imply the incorporation of foreign DNA from a different taxon into their genome (Didelot & Maiden 2010).

A large diversity of genetic elements mediate intercellular horizontal gene transfer (HGT), sometimes in association with intracellular recombination. They have been reported by numerous authors as playing a key role in bacterial evolution (Gogarten & Townsend 2005; Polz *et al.* 2013; Wiedenbeck & Cohan 2011). Strains of the same bacterial species harbor an accessory genome content of variable size, only having in common genes that belong to what is called the core-genome. A meta-analysis suggested that strains of a species sometimes differ by up to 35% of their gene content (Konstantinidis *et al.* 2006). Although horizontally acquired DNA can impose a substantial metabolic burden to its bacterial host, it sometimes greatly improves the fitness of a pathogen. This is because it provides the bacterium with new functions such as enhanced in-host multiplication, colonization of a new ecological niche or antimicrobial resistance (McManus *et al.* 2002; Swarup *et al.* 1991; Thynne *et al.* 2015). Indeed, several pathogens have emerged following the acquisition of a single plasmid (Barash & Manulis-Sasson 2009; Kado 2015; Pilla & Tang 2018).

The emergence of epidemic bacterial pathogens is favored by several factors such as the uniformity of their environment and the increased international movement of humans and goods due to globalization (Baker *et al.* 2018; McDonald & Stukenbrock 2016). Agro-ecosystems are human-created uniform environments, where crops often having restricted genetic basis are grown at high densities (McDonald & Stukenbrock 2016). Moreover, seeds, cuttings or eatable goods from these crops are often exchanged on the worldwide market, possibly enhancing

dispersal of associated pathogens. A great number of phytopathogenic bacteria successfully thrived in cultivated environments and now cause severe economic and social issues all around the world (Mansfield *et al.* 2012). Studies aiming at characterizing the genetic basis of adaptation of plant pathogenic bacteria mainly focused on the genetic determinants of host-pathogen interactions (including host specialization) that arise through mutations (Cruz *et al.* 2000), gene loss or inactivation (Kousik & Ritchie 1996), HGT (Chen *et al.* 2018; Llop *et al.* 2012; Valinsky *et al.* 1998) and to a lesser extent the genetic basis of antibiotic resistance (Sundin & Wang 2018). SNP-based phylogenetic studies indeed only offer a partial view of bacterial evolutionary history. Instead, the concomitant study of multiple types of polymorphisms often allows a better understanding of pathogen transmission, geographical expansion and host specificity. The use of such integrative approaches on emerging plant pathogenic bacteria is however in its early steps (Baltrus *et al.* 2017; Goss 2015; Jacques *et al.* 2016; Monteil *et al.* 2016; Vinatzer *et al.* 2014).

Xanthomonas citri pv. *citri* (Xcc) is a pathogenic bacterium causing Asiatic citrus canker (ACC), a major disease that has an Asian origin and emerged over the XXth century in all tropical and subtropical areas of mass citrus production. This bacterium is part of the so-called monomorphic group, and a single genetic lineage of it mostly caused a major worldwide expansion (Pruvost *et al.* 2014). The genetic diversity of Xcc has first been studied using a number of genotyping methods (Graham *et al.* 2004; Pruvost *et al.* 2015) that display multiple limitations (Ranjbar *et al.* 2014). More recently, genome-wide studies evidenced the important role played by HGT events in the evolution of pathogenicity in Xcc and other *Xanthomonas* species (Chen *et al.* 2018; Ferreira *et al.* 2015; Gordon *et al.* 2015). In response to massive copper applications used for ACC control, Xcc showed adaptive potential despite its monomorphic nature, mediated by the integration of a copper-resistance conjugative plasmid (Richard *et al.* 2017b). Apart from early elements, the characterization of the evolutionary history of Xcc is nonetheless still scarce. No information about Xcc SNP substitution rate is available, nor about the relative importance of gene content variations relative to SNP.

The potential of using SNP along with gene content variations to decipher evolutionary history of bacterial pathogens was clearly illustrated in two recent studies. McCann *et al.* (2017) showed the potential of the combined use of these two approaches for deciphering the emergence of the kiwifruit canker bacterium, *Pseudomonas syringae* pv. *actinidiae*. Chen *et al.* (2018) evidenced the importance of gene acquisition events in the pathogenicity and speciation of bean-pathogenic *Xanthomonas* lineages. Focusing on a monophyletic cluster of Xcc strains

causing recurrent ACC outbreaks in the South West Indian Ocean, we herein evaluate the relative contributions of two different forms of molecular evolution in an attempt to have a better understanding of the underlying mechanisms responsible for the emergence, spread and adaptation of a crop pathogenic bacterium at several spatio-temporal scales.

Materials and Methods

Bacterial strains

SWIO strains studied herein were selected as strains representative of the genetic diversity on the basis of a preliminary analysis (data not shown) of microsatellite data (MLVA-14) of (i) 281 strains collected in the Comoros archipelago (Anjouan, Grande Comore, Mayotte and Moheli) from 1993 until 2013, (ii) 1460 strains collected in the Mascarene archipelago (Mauritius, Réunion and Rodrigues) from 1978 until 2016 (Richard *et al.*, 2017; this study), and (iii) 50 strains collected in the Seychelles archipelago (Mahé) from 2003 until 2017. The MLVA-14 typing scheme was previously found to display a discriminatory power high enough to reveal polymorphism at small spatio-temporal scales (Leduc *et al.* 2015; Verniere *et al.* 2014). Genotyping was performed as reported previously (Bui Thi Ngoc *et al.* 2009). Amplicon sizes were transformed to tandem repeat numbers (rounding to the superior integer). Multidimensional scaling plots were built from a matrix of Manhattan distances between strains using the BIOS2MDS R package (Pele *et al.* 2012). PHYLOVIZ v2.0 was used to produce categorical minimum-spanning trees using the algorithm recommended for tandem repeat data combining global optimal eBURST (goeBURST) and Euclidean distances (Nascimento *et al.* 2016). Strain clusters were defined as haplotype networks linking up to quintuple-locus variants.

A total of 225 SWIO strains (Fig. 1) was further selected from all genetic clusters conservatively defined so that each cluster contained haplotype networks linking up to quintuple-locus variants. Of these SWIO strains, 105 were within-grove replicates, collected from 14 Réunion citrus groves, with a mean number of 7.5 strains per site. These 225 strains, along with 60 Xcc pathotype A strains from 16 remote countries in Asia, Oceania, West Africa, North and South America and the Caribbean region and the Xcc pathotype A* strain JK002-10 were further selected for full genome sequencing. Notably, the strain collection included copper-resistant (CuR, n=44) strains from all three territories (Argentina, Reunion, Martinique) where this phenotype was reported (Table S1, Supporting information).

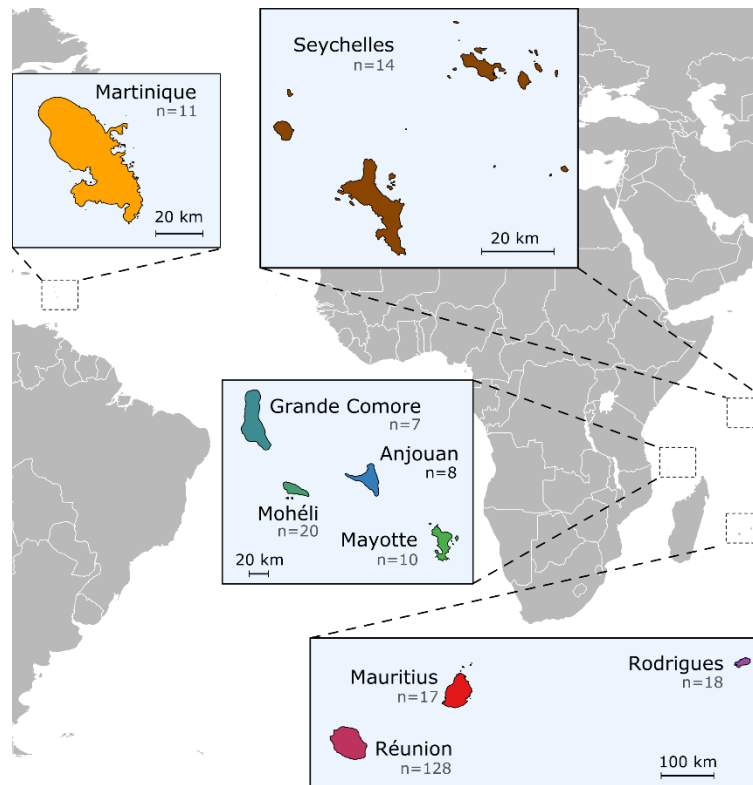


Fig. 1. Location of the South West Indian Ocean islands and Martinique from where the majority of the strains studied in this were sampled.

DNA extraction and sequencing

Of the 285 strains, 20 were previously sequenced as described in the corresponding publication (Gordon *et al.* 2015). DNA of the 265 remaining strains was extracted from colonies grown on YPGA media using the Promega Wizard kit. Illumina sequencing was performed by GATC Biotech on HiSeq 4000 instrument (2x150bp) using the genomic Nextera XT protocol.

Read alignment, SNP detection and filtration

The bioinformatics pipeline used to process the Illumina raw reads is detailed in Fig. S1, Supporting information. Briefly, after a quality control trimming step using Trimmomatic v. 0.36 (Bolger *et al.* 2014), reads were aligned against the chromosome of Xcc strain IAPAR 306 (GenBank accession NC_003919.1) with BWA-MEM v. 0.7.15 (Li 2013). Duplicated reads were removed using PicardTools v. 2.7. Indel realignment and SNP calling were performed using Freebayes 0.9.21-5 (Garrison & Marth 2012). The SNP set (n=20,650) was filtrated with a custom perl script (see Fig. S1, Supporting information for details). In order to exclude regions that undergone homologous recombination from the analysis, a detection of recombinant

regions was performed with ClonalFrameML (Didelot & Wilson 2015) and RDP4 (Martin *et al.* 2015) using default parameters.

Core and accessory genome assignation

In order to estimate the gene content of each strain from the SWIO clade, we used an approach that combined mapping and de novo assembly. Trimmed reads were mapped with BWA-MEM against six high-quality PacBio consensus chromosomes and plasmids from our previous work (Richard *et al.* 2017b). Per strain, unmapped reads were then assembled using Spades v. 3.6.2 (Bankevich *et al.* 2012) using default parameters (see Table S2, Supporting information for assemblies details) and genes were predicted from the resulting contigs using prodigal v. 2.6.3 (Hyatt *et al.* 2010). Contigs corresponding to human or virus DNA and most probably originating from contaminants were identified using Blastn against NCBI databases and discarded. Genes with a coverage inferior to 20 or detected as potential chimeras using the Uchime algorithm of the VSEARCH package were also discarded (Rognes *et al.* 2016). These predicted genes were added to the 29,164 ones from the six pacbio references mentioned above and together represented the total geneset of the SWIO clade (n=35,527 genes). The total geneset was then clustered using MMseqs2 linclust using parameters -c 0.5 and --min-seq-id 0.9 (Steinegger & Söding 2017). A total of 5,046 gene clusters was obtained. A gene with a coverage of 20X over at least 60% of its length (on either the six reference genomes or de novo assemblies) was considered present in a strain. Gene clusters were then attributed to the strains according to their validated gene content. Among these 5,046 gene clusters, it was important to distinguish between genes associated to chromosome and to plasmid. To this end we use a network approach where gene clusters were associated depending on the number of times they were found on the same contig. After a detection of the communities in this network (*i.e.* groups of gene clusters frequently co-occurring on the same contig), gene cluster with unknown location but within communities with good support for either a chromosomal or plasmidic support were given the location of the community. Gene clusters displaying the same pattern of presence/absence among all the strains were merged into groups representing putative single events of gene movement. We then reconstructed the ancestral states of presence/absence of those groups using the Ape R package (Paradis *et al.* 2004). We considered the gain of any given gene group when its presence probability increased of more than 0.5 along a branch and did the same approach for gene loss. For any given clade, we therefore defined (i) the number of gain and loss events, (ii) the number of gained and lost genes, given that a gene can be gained multiple times and (iii) the number of variable genes.

Gene presence/absence detection power estimation

To validate our gene content analysis method, we applied our gene content pipeline to three replicates of two strains (Xcc LH201 and Xcc LE50) both already sequenced using PacBio. Importantly, distinct bacterial culture and DNA extraction were performed for each of the replicates. Six gene clusters encoding for unknown proteins displayed a contrasted presence/absent pattern within the LE50 replicates and none for LH201. These results highlight that whereas our methodology is globally effective to reconstruct gene content, variations between strains may be in part due to the methodology. A focus on events including numerous genes or on events that display phylogenetic signal (*i.e.* when closely related individuals present with the same event) may surmount this limitation.

Phylogenetics and molecular clock analysis

Defining the model of evolution of the polymorphisms studied is an important step of a phylogenetic study. Different sets of sites can display independent models of molecular evolution, and this information is useful for precise Bayesian inference of phylogenies. Using PartitionFinder v.2.1.1 (Lanfear *et al.* 2016), we therefore tested if our set of SNP displayed such heterogeneity by testing different clustering schemes: synonymous vs non-synonymous; codon position 1 vs 2 vs 3 vs non-coding; the grouping of the two previous; singleton vs. non-singleton. None of the tested clustering schemes improved the BIC. According to PartitionFinder results, the model of evolution best fitting our dataset was GTR+G+I.

A maximum likelihood tree was obtained with RAXML (Kozlov *et al.* 2018) using a GTR+G+I model of evolution. The strain Xcc JK002-10 (A* pathotype) was used as an outgroup. The tree was edited using the ape R package (Paradis *et al.* 2004). The strength of temporal signal in our data was investigated by computing the linear regression between sample age and root-to-tip distances at every internal node of the tree using the tipdatingbeast R package (Rieux & Khatchikian 2017). Internal nodes at which both the linear regression was statistically significant and the slope positive were assumed to contain detectable amount of evolutionary change and were considered as thorough evolutionary scales for tip-dating inferences to be performed. At those scales, a date-randomization test (DRT, 20 independent randomization) was performed to confirm the presence of thorough temporal signal (Duchene *et al.* 2015).

Tip-dating inferences were then performed using BEAST v1.8.4 (Drummond & Rambaut 2007) within clades at which sufficient temporal signal was detected. Nucleotide substitution rate was

estimated with a general time-reversible (GTR) substitution model of evolution and rate variation among sites was modeled with a discrete gamma distribution with four rate categories and invariant sites (GTR+G+I). We assumed an uncorrelated log-normal relaxed clock to account for rate variation among lineages. To minimize prior assumptions about demographic history, we first used an extended Bayesian skyline plot (EBSP) approach to integrate data over different coalescent histories. The inspection of the demographic reconstruction revealed that the demography followed an exponential growth, which was thus used as a tree prior. Three independent chains were ran for 100,000,000 steps and sampled every 10,000 steps. The first 1,000 samples (10%) were discarded as burn-in. Convergence to the stationary distribution and sufficient sampling and mixing were checked by inspection of posterior samples using Tracer (effective sample size >200). After combination of the three runs, a maximum clade credibility (MCC) tree was obtained using TreeAnnotator.

Results

Sequencing, SNP calling and de novo assembly

We sequenced 285 strains, of which 225 originated from the Indian Ocean region. Strains were selected in order to be representative of the genetic diversity of the region, based on microsatellite data. After alignment against Xcc IAPAR 306, the 285 sequenced citrus pathogenic Xcc strains displayed a mean coverage of 232 (details in Table S2, Supporting information). The SNP calling yielded 20,650 SNP which were filtrated on quality, coverage and other parameters (see Fig. S1, Supporting information) resulting in 7,005 high confidence SNP. Recombination analysis using ClonalframeML detected a 136kb recombinant region comprising 233 SNP in seven strains originating from Mali, Senegal, Bangladesh and India. This recombinant region was concordant with the RDP4 detection of a 197 kb (315 SNP) recombinant region that entirely overlapped the region detected with ClonalFrameML. These 315 SNP were thus excluded from the global dataset but were kept for the SWIO analysis, as these strains did not present any intra-group recombination event. Within the SWIO subgroup, 3,602 out of the 7,005 SNP were monomorphic. On the global ML phylogeny inferred from the SNP set (Fig. S2, Supporting information), all SWIO strains were included in one monophyletic clade (later on called SWIO clade). Consistent with our previous microsatellite-based hypothesis that copper-resistant Xcc strains emerging in Martinique (French West Indies) may have originated from Réunion (Richard *et al.* 2017b), eleven strains from Martinique all clustered within the SWIO clade together with most copper-resistant strains isolated from Réunion (Fig. 2). The 20 strains from the Northern part of the Indian Ocean (Maldives and Seychelles, later on called NIO) were gathered in four clades that were genetically distant from the SWIO strains and more related to strains from the Indian sub-continent. Consistent with SNP data (Fig. S2, Supporting information), multidimensional scaling plots built from microsatellite data separated strains from NIO and SWIO into distinct genetic structures (Fig. S3A and Table S1, Supporting information).

Strong geographic structure at the island scale

At the island scale, a strong geographical structure was apparent within the SWIO clade with strains originating from the same or closely located islands being grouped in subclades (Fig. 2). Likely relating the homoplastic nature of tandem repeat allelic states, this geographical structure revealed from SNP data was less apparent with the microsatellite data. Moreover, it was

difficult, using microsatellite typing, to assess the genetic relatedness of many singletons (*i.e.*, strains that differed from their closest relative at > 5 microsatellite loci - Fig. S2, Supporting information) to other SWIO strains. This is because microsatellites have a low ability to reliably establish deep genetic relationships among strains. All the 27 strains from the Comoros archipelago were grouped in the SNP clade number 2 (Fig. 2). All but two Mauritius strains formed the well-defined clade number 3 on Fig. 2. One of the two remaining strains was included within the Rodrigues clade (otherwise monophyletic, see clade number 3 of Fig. 2) while the second one was a close relative to the group of Reunion including copper-resistant and -susceptible strains and the last one was a distant outgroup of the whole SWIO clade. In Reunion, for which an extensive strain sampling from 14 groves was performed, strains were genetically diverse and formed several clearly defined phylogenetic clades. Interestingly, the major clade from Reunion, that comprised 51 out of 128 strains from this island, gathered most ($n=39$) of the copper-resistant strains, while the remaining ones ($n=5$), also from Réunion, were distributed into three distinct subclades (clade number 6 on Fig. 2). In Réunion, no spatial structure was apparent and all but one groves hosted strains from two or more of the Réunion subclades (Fig. S2, Supporting information).

Temporal signal and tip dating

Whereas there was no significant temporal signal for the global phylogeny, the regression between the root-to-tip distances and the node ages was significant (p -value 2.3×10^{-4}) at the root of the SWIO clade with a slope of 3.2×10^{-5} and a R^2 of 0.06 (Fig. S4B, Supporting information). The temporal signal at the SWIO clade scale was further confirmed by a date-randomization test for which no overlap between the inferred tree root age 95%HPD of the real and the date-randomized datasets was observed (Fig. S4A, Supporting information).

Table 1. Inferred dates and substitution rates of SWIO root and of five nodes associated with subclades of interest.

Node number	Inferred node date	Date 95% confidence interval	Individuals	Variable genes	Variable SNP	Subst. rate SNP
1	1818	1762-1868	222	699	3403	8.4E-8
2	1906	1882-1927	48	205	812	8.8E-8
3	1931	1914-1947	19	222	363	9.0E-8
4	1949	1934-1963	12	143	134	8.8E-8
5	1959	1946-1971	27	147	288	9.4E-8
6	1965	1956-1973	62	358	318	1.0E-7

A BEAST analysis was thus conducted using a log-normal relaxed clock and an exponential growth tree model. Overall, the ML and the Bayesian trees had a similar topology (Fig. S5, Supporting information). The tree root age was inferred at 200 years (95% HPD: 147-257 years) since the most recent sample dating from 2017 and the mean substitution rate was estimated at 8.4×10^{-8} substitutions per site per year (95% HPD: 6.9×10^{-8} - 1.0×10^{-7}), which scaled to 0.43 substitutions per genome per year (95%HPD: 0.35-0.51). The estimated standard deviation of rate variation among branches was 0.26 (95% HPD 0.18-0.34). Whereas the exclusion of the zero from this interval (*i.e.* no variation between branches) rejects the strict clock, its low value nevertheless indicated a low variation of the substitution rate between branches of the tree. The substitution rates of five well defined SWIO subclades (detailed in Fig. 2 and Table 1) were also inferred in two ways. First, the inference was performed globally with the full dataset. Rates and node ages were then extracted and summarized per clade. Secondly, five distinct inferences were performed with datasets restricted to each clade (Fig. S6, Supporting information). Confidence intervals of all the inferred substitution rates overlapped.

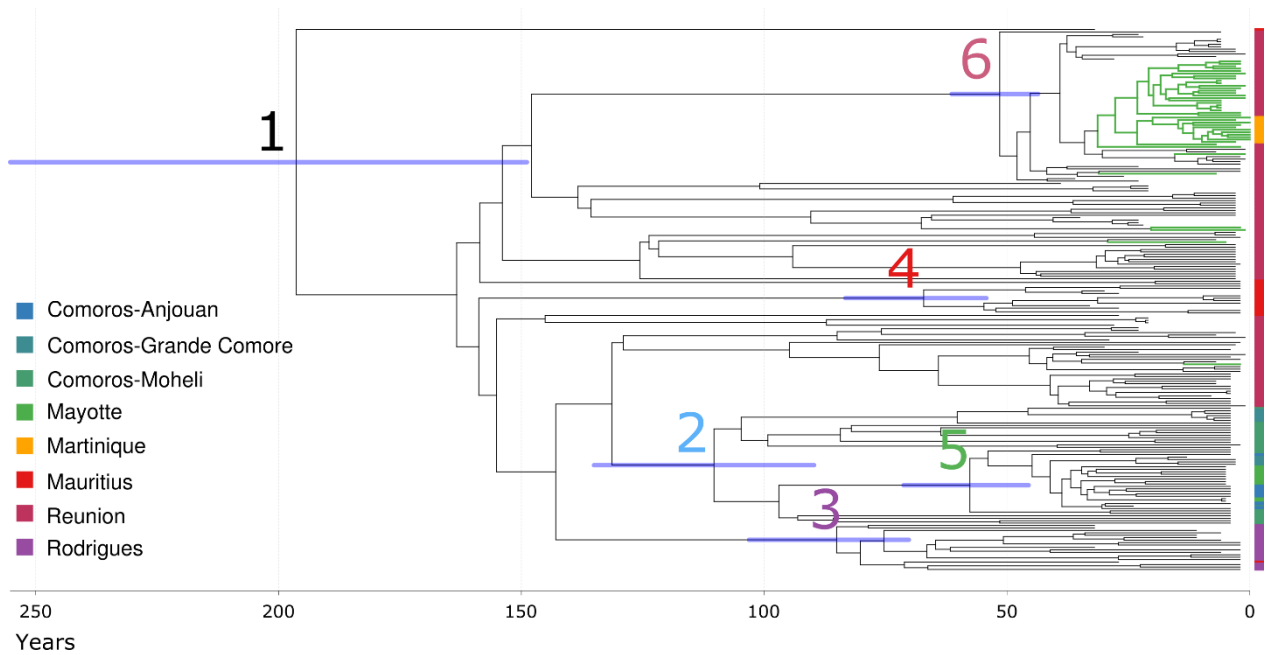


Fig. 2. A dated phylogeny of the SWIO clade of *Xanthomonas citri* pv. *citri*. Node bars represent 95% highest posterior density (HPD) for node ages estimated with tip-calibration. Tips are colored according to sampling location. Strains at the tip of bold green branches are copper resistant.

A variable accessory genome content

The 222 strains originating from the SWIO were *de novo* assembled, which yielded a mean number of 418 contigs comprising a mean of 5,325,952 nucleotides. Among the 5,046 genes detected for all the 222 SWIO strains, 4,347 were detected in all the strains (core-genome), while 699 were absent from at least one strain (accessory genome), none being specifically present or absent in a subclade. Of these, 336 were assigned to the chromosome compartment, 339 to the plasmid compartment and 24 remained unassigned (Fig. 3). Gene clusters displaying the same pattern of presence/absence among all the strains were merged into 180 groups representing putative single events of gene movement. Among these groups, 25 solely comprised one gene of which 14 were found in only one individual. The possible errors associated with our gene detection pipeline are likely part of these 14 groups.

We then reconstructed the presence/absence ancestral states of these 180 groups along the phylogeny and thus the history of gene gain and loss. The SWIO phylogeny comprised 551 gain events and 1,810 loss events (of which 14 (0.6%) corresponded to the possibly erroneous singletons mentioned above). The plasmid and the chromosome genomic compartments displayed different dynamics of gene gain and loss. Plasmid gain and loss events comprised more genes than chromosomal ones. Indeed, plasmid events comprised a mean number of 16

genes, versus two for chromosomal genes. Most lost genes belonged to the chromosomal compartment (2,438 chromosomal genes were lost along the entire phylogeny and 564 of plasmid origin) while most gained genes had a plasmid origin (chromosome : 462; plasmid : 1,483) (Fig. 4 A).

Gain, loss and mosaic structure of known plasmids

Several gene movements comprised parts or all of the genes of known plasmids such as a known copper resistance plasmid and three other plasmids. The nature of our sequencing data prevented us from assembling complete circular plasmids. Using each strain's gene content, we however could assess the presence of the genes of each known plasmid in each strain.

High-quality circular plasmid sequences homologous to pXac64 (GenBank accession CP018856.1, from strain Xcc LH276) and pXac47 (GenBank accession CP018853.1, from strain Xcc LJ207-7) were used as genomic reference for each plasmid's gene content and order.

Most of pXac64 genes were conserved among all strains sampled in the SWIO region. 49 strains from Mauritius, Réunion and Rodrigues however showed a large deletion of 26 genes mostly encoding for plasmid-backbone functions such as conjugation, partitioning or plasmid maintenance (Fig. 3 and Fig. S7, Supporting information). Strains sampled from a same Réunion grove sometimes displayed different pXac64 gene contents.

Most strains from all SWIO islands hosted a pXac47-like plasmid. Genes encoded on pXac47 were mostly absent from 77 strains (Fig. 3 and Fig. S7, Supporting information). Of these, 61 were phylogenetically related and corresponded to all but one individual of the well-defined clade 6, which grouped copper-resistant and -susceptible strains from Réunion. 16 other strains presenting with this deletion were spread in the phylogeny and originated from Réunion, Mauritius, Moheli, Mayotte and Rodrigues. pXac47 complete gene set was found in all islands. Eight groves out of 16 included both plasmid-bearing and plasmid-lacking strains. Loss of this plasmid was inferred to have occurred nine times along the phylogeny.

An original 39.8kb circularized plasmid encoding for 40 genes was found in a single strain from Mauritius. No other strain presented with these genes, which mostly encoded for proteins involved in conjugation, partitioning and plasmid maintenance usually found in *Xylella fastidiosa* or *Xanthomonas oryzae* (GenBank accession CP014330.2 and CP007810.1, respectively).

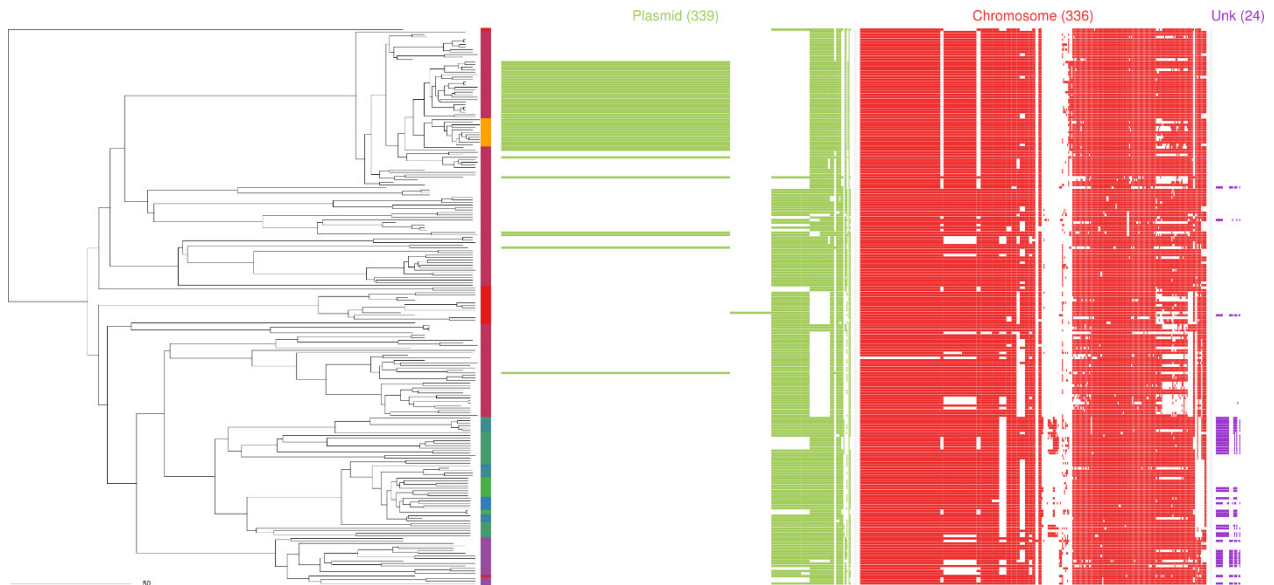


Fig. 3. Bayesian phylogeny of the SWIO clade. The matrix at the right of the tree comprise one column for each one of the 699 genes that varied in presence/absence among the 222 SWIO strains. A colored box is printed in front of a strain if it encodes for this gene family. Gene families are separated according to their genetic compartment of origin: plasmid (green), chromosome (red), undefined (purple).

A highly polymorphic chromosomal region

Using gene positions of a high quality circular chromosome sequence (GenBank accession CP018854.1, from strain Xcc LH276) as reference, we located 206 of the 336 variable SWIO chromosomal gene clusters into four regions (Table 2), none of these corresponded to recombinant loci as detected by our analysis and that of Gordon *et al.* (2015). The remaining 130 genes were individually spread along the chromosome, or had no homolog in the chosen reference.

One of these regions (region 3 in Table 2) appeared particularly prone to gene deletion and showed five main patterns of variation that comprised from two to 78 genes. All five patterns had in common the deletion of five genes located directly downstream to a toxin-antitoxin system and a gene encoding a protein involved in replication (Table 2). Interestingly, this five-gene region was inferred to have been independently lost 19 times in the phylogeny. No less than 13 genes encoding for ICE related proteins and eight genes encoding for proteins involved in conjugation were found downstream of the deletions. These genes were conserved in all but one strain. However no known complete ICE or genes with evident adaptive functions were annotated.

Region	Position on Xcc LH276 chromosome (CP018854)	Individuals lacking region	Genes	Times lost in the phylogeny	Times gained in the phylogeny	Main detected fonctions
1	2224234-2234692	1	7	1	0	Integrase
2	2547120-2550354	1	2	1	0	Chemotaxis
3	2817666-2819774	84	5	19	2	Replication
	2830567-2861258	19	18	13	0	Stress-induced efflux transport
	2862047-2876642	18	14	12	0	--
	2884264-2976635	1	78	1	0	ICE, Conjugation
	3000841-3003260	16	3	1	0	DNA repair
4	4172586-4176990	1	4	1	0	DNA repair
	4183190-4187237	5	5	3	0	--
	4187883-4190106	4	3	3	0	Integrase

Table 2. Characteristics of four chromosomal regions presenting with important gene content variations.

Microevolutionary gene variations

At different scales, the number of gene variation events remained in the range of the number of SNP (Fig. 4A). However, because an event can result in the insertion or deletion of several genes, the number of variable genes of a group of strains can be much more important than the number of variable SNP. We first observed this pattern at small phylogenetic scales (see subclades 4 and 6 Fig. 4B). Levels of SNP (from five to 335) and variable genes (from six to 385) as well as the SNP/variable genes ratio (from 0.2 to 5.9) varied greatly (Fig. 4C) between the groves. At the grove scale, the diversity of plasmid content was grove-dependent. While it was found homogeneous in four groves, strains hosted variable plasmid contents in other ones (*i.e.*, up to four different versions of pXac64-like plasmid within a single grove, variable presence of pXac47 in eight groves, variable presence of pLH201.1 in six groves). The most variable plasmid content was found in grove 2 where six distinct plasmid profiles were detected among ten strains.

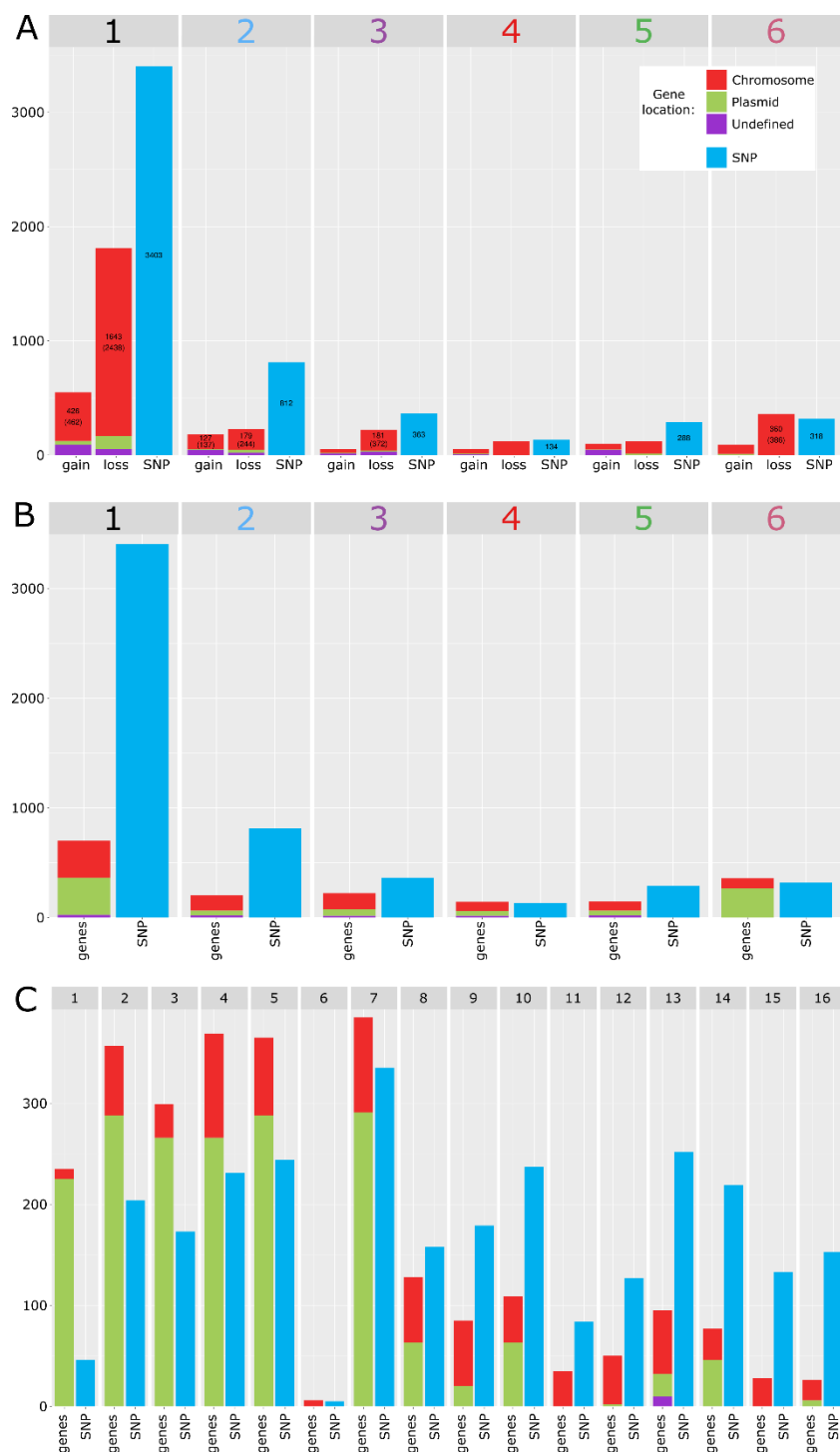


Fig. 4. (A) Barplot representing the number of gene gain events (left bar), gene loss events (center bar) and SNP (right bar) of the SWIO root and of five subclades of interest. Numbers of events are indicated when superior to 125. Numbers in brackets indicates the number of gained and lost genes. The following barplots represent the number of variable genes and the number of variable SNP among the strains of (i) five subclades of interest (B) and (ii) each sample Réunion grove (C). Bars are horizontally split into portions whose colors depend on the genomic compartment of the gene family as stated above.

Discussion

In this study, we considered a population genomic approach to gain insights into the underlying mechanisms associated with the emergence, adaptation and evolution of *Xanthomonas citri* pv. *citri* within the South West Indian Ocean (SWIO) area. To this purpose, we combined the inference of a dated phylogeny with a gene content reconstruction. On top of worldwide diversity and diversity at the country scale, the sampling comprised multiple strains from a more restricted geographic range, allowing the characterization of the evolution of the bacterium from the macro-evolutionary (the SWIO area) to the micro-evolutionary (single citrus groves) scales.

We first built a ML tree using 6,690 SNP excluding recombining regions and showed that Xcc strains from the Comoros and the Mascarene archipelagos (SWIO) formed a well-defined monophyletic cluster distinct from strains originating from all other regions (NIO and rest of the world). Despite SNP and microsatellite data structures being often congruent, a stronger phylogenetic structure was obtained as expected with SNP relative to microsatellite data (Fig. S2). Importantly, we evidenced the presence of temporal signal in the accumulation of SNP at the evolutionary scale of the SWIO clade, which allowed the use of a tip-dating approach to calibrate the phylogenetic tree using strain sampling dates (ranging from 1978 to 2017) (Rieux & Balloux 2016).

Dating the root of the SWIO clade

We hence used the models implemented in BEAST to jointly infer the substitution rate (number of SNP per site per year) and dates associated with the nodes of the tree within the SWIO clade (Rieux & Balloux 2016). The age of the ancestor of all strains can be used as an indicator of Xcc's introduction date in the SWIO. This introduction was estimated to 1818 (95% HPD: 1762-1868). We found no writings dating precisely the first disease emergence in the region, though the earliest report of the disease was in 1917 in Mauritius (Aubert 2014). Xcc and its main host genus, Citrus, originated in Asia (Pruvost *et al.* 2014; Wu *et al.* 2018) and both were most likely spread by human-mediated movements of plants or plant propagative material. The time frame inferred herein yielded two main hypotheses on the origin of the pathogen in the SWIO. On the one hand, the French botanist, Christian missionary and colonial administrator Pierre Poivre (1719-1786) organized introductions from several Asian countries of numerous plant species, including many fruit trees (possibly including Citrus infected by Xcc) in the

Mascarene archipelago (at the time all islands being French possessions) starting mid-eighteenth century (Du Pont de Nemours 1797). On the other hand, after the abolition of slavery in Mauritius and Réunion in 1835 and 1848, respectively, several hundred thousand indentured laborers were brought from several Asian countries (most numerous from India) to these islands to build a workforce involved in agriculture (Campbell 1923; Carter 2002; Govindin 1994). This active flow of goods and humans from the Asiatic continent may have been associated with the introduction of Xcc in the SWIO area. The emergence of phytopathogenic pathogens due to migratory events is in fact a known phenomenon as previously illustrated with the European emergence of *Phytophthora infestans*, the fungus that triggered the 1845 Irish potato famine (Yoshida *et al.* 2013). Rice blast caused by *Magnaporthe oryzae* and wheat yellow rust fungus caused by *Puccinia striiformis* are other examples of migrations driving the emergence of internationally devastating pathogens of major staple crops (McDonald & Stukenbrock 2016). More generally, introduction of pre-adapted pathogens into new territories is the main driver of the emergence of crop diseases, including that caused by bacteria (Anderson *et al.* 2004). We were not able to clearly identify the geographic origin of the strains that first migrated to the SWIO region. Indeed, if the SWIO clade is well-defined and evidence a clear genetic coherence, more genomic resources from Asia are needed for understanding its branching in the worldwide diversity.

The inferred substitution rate for SWIO Xcc strains was of 8.4×10^{-8} substitutions per site year⁻¹ (HPD : 6.9×10^{-8} - 1.0×10^{-7}), consistent with data obtained from various bacterial taxa, which presented with rates between 10^{-8} and 10^{-5} , although being from the lowest part of the known rates (Duchene *et al.* 2016). The rate estimated herein was found close to that calculated using a node dating approach for the monomorphic emerging Psa3 lineage of *Pseudomonas syringae* pv. *actinidiae* responsible for kiwifruit canker disease (McCann *et al.* 2017). The rate estimated for Xcc was in the same range than that of other monomorphic bacterial pathogens sometimes sampled over a similar time frame (Achtman 2012; Duchene *et al.* 2016). Our study represents to our knowledge the first estimation of a plant pathogenic bacterium substitution rate obtained through tip dating. Beside putting its evolutionary history on a timescale, the characterization of the rate of evolution of a pathogen is paramount to apprehend its evolutionary past and predict its propensity to adapt. Altogether, these new information will help understand the past history of Xcc and implement future integrated pest management approaches.

Within SWIO genetic structure

The small length of the branches at the basis of the SWIO clade are indicative of the ancient introduction of a restricted diversity of Xcc in SWIO and no other incoming migration of genetically distant strains since then. Nevertheless, determining if the extant diversity is the outcome of a single or several introductions remains difficult to define. The very fact that all the SWIO strains formed a monophyletic clade is indicative that regardless of the number of introduction events, the imported strains were genetically and probably geographically closely related. Indeed, polyclonal introductions of Xcc strains are likely frequent (Leduc *et al.* 2015). In such scenario, the more the introduced strains were genetically diverse, the more the actual introduction into the region would be more recent than the MRCA of all the SWIO strains. Nevertheless, if the subclade structure of the Réunion strains is due to multiple introduction events, the timing of the phylogeny is globally incompatible with any introduction that led to the establishment of strains after 1973 (the lower limit of the 95%HPD of the subclade 6, Table 1) which could illustrate (i) the efficiency of regulations restricting the importation of citrus plants and plant parts, or (ii) in some instances the low need of citrus budwood importation from remote sources due to the minor economic significance of local citrus industries.

It is also difficult to establish which island was first contaminated. Two points prevented its identification on the basis of the node ages. First, unbalanced geographic sampling increases TMRCA ages of moderately sampled clades. Secondly, the polyphyletic structure of both Réunion and Mauritius strains suggests a polyclonal nature of the primary inoculum and/or the presence of migratory events between the islands.

Inter-island migration events were illustrated by the genetic structure of two groups of strains. First, all strains originating from Rodrigues (a tiny remote Mauritian territory) grouped in a clade that also comprised a Mauritian strain. This mixing could indicate migration between the two islands, that are likely to occur given that Rodrigues and Mauritius are politically and economically linked. The inferred date of the clade root was 1931 (HPD : 1914 – 1947) that corresponds to the first description of the disease in Rodrigues, dated at 1937 (Wiehe, 1941). In Mauritius, a report suggests the presence of the pathogen as early as 1917 (Aubert 2014) whereas the inferred date of the root of the main Mauritian clade was 1949 (HPD : 1934 – 1963). Secondly, we found no clear genetic delineation between the strains from the Comoros archipelago, with Grande Comore strains spread in the clade. The geographical proximity of the islands (<75 km) and the absence of plant movement regulations between the three islands

of the Union of the Comoros may explain this apparent mixing. The clade grouping all the strains of the Comoros archipelago was dated at 1906 (1882-1927), which slightly preceded the first description of the disease in Moheli in 1966 (Brun, 1971).

Long distance dispersal of Xcc is mainly achieved through human-associated dispersal of contaminated citrus budwood or sometimes whole plants (Leduc *et al.* 2015). This dispersal mode is illustrated at the worldwide scale by the positioning of all Martinique strains within a Réunion clade. Martinique, which was a disease-free area until a few years ago (Richard *et al.* 2016), seems to have been contaminated by strains originating from Réunion. These two islands are distant of more than 13,000 km but are both overseas regions of France, which implies exchange of goods and regular human flow.

Microevolution and biology of the pathogen

We observed a polyphyletic structure of the strains originating from most of the 14 Réunion groves, revealing admixture of individuals at this spatial scale (Supplementary Fig. S2). This illustrates the multiplicity of inoculum inputs during the life of citrus groves brought by the plants used for grove establishment, major wind-driven rain episodes (*i.e.*, storms, hurricanes) and/or grove maintenance operations. Strains from one single grove indeed contained up to 300 SNP and varied on the presence of up to 350 gene clusters (Fig. 4C). These results emphasize that the orchards constitute physical contact points between genetically divergent bacterial strains in areas of disease endemicity. This might be crucial for HGT-mediated evolution because it increases the likelihood that a genetic recombination event brings new alleles to a genome. The Indian subcontinent was shown to host most of Xcc's genetic diversity known to date (Pruvost *et al.* 2014). Thorough analyses of HGT at microgeographical scales in Xcc's native area where distinct genetic lineages occur in sympatry are important topics for future research.

HGT causes phylogenetically-related strains to sometimes harbor a different gene content (Fig. 4A) and can provide a larger panel of accessory genes to local genotypes through introduction of foreign genes. Some of these genes may bring adaptive traits to populations bearing them and even intensify their epidemic character. An epidemic copper-resistant clade in Réunion (which comprised strains from eleven of the 14 studied groves) likely consist of local strains that integrated an adaptive plasmid from an unidentified donor. The evolutionary importance of horizontally transferred genes in xanthomonads was suggested at the species (Chen *et al.* 2018) and the pathovar scales (Gordon *et al.* 2015).

The SNP/variable genes ratio (that is 4 at the SWIO root) indeed sometimes considerably favored variable genes at low geographic and phylogenetic scales, being as low as 0.2 in a grove (Fig. 4C) and 1 for to subclades (Fig. 4B). As stated by multiple authors, SNP measured at very small scales reflect different biological processes (*i.e.* close to the non-lethal mutation rate) than those measured over longer terms (substitution rate, reflecting non-costly mutations (Duchene *et al.* 2016; Ho *et al.* 2011)). Our results did not evidence such a time-dependency of the rate of SNP evolution between the SWIO root and the six analyzed subclades (Fig. S6, Supporting information). In the same manner, we postulate that accessory genes that are variable at a local scale represent a fraction of the local gene pool for Xcc adaptation. The study of this dynamic gene pool would benefit from sampling at even higher resolution such as at the scale of the diseased tree, the symptomatic branch or the canker lesion.

The dynamic character of the plasmid compartment

At the SWIO scale, several plasmids presented a mosaic structure, and even varied in their presence amongst related strains. Some Xcc virulence determinants, including the major pathogenicity gene *pthA4* are known to be plasmid-encoded. Rearrangements between resident plasmids are apparently common in Xcc, sometimes even leading to chimeric plasmids grouping genes from all of the three described plasmids bearing TALE genes (Gochez *et al.* 2018). Our sequencing data didn't allow full plasmid reconstruction. However, it gave insights on the extent to which recombination events modify plasmid gene content based on the circular sequences of some plasmid forms previously obtained (Richard *et al.* 2017a). We further confirmed the mosaic structure of pathogenicity-related plasmids in Xcc. The high number of genes encoding for transposases and recombinases in those plasmids may drive their fast evolution (this study and (Ferreira *et al.* 2015)).

The large plasticity of plasmids bearing TALE genes on the present study contrasts with the gene content uniformity of the two other plasmids frequently detected in our dataset (pLH201.1 and pXac47-like). The genetic uniformity of the copper resistance plasmid, which contrasts with its mosaic nature at broader geographic and phylogenetic scale (Richard *et al.* 2017b), may be due to its recent introduction in the region, where the phenotype has been first observed in strains isolated in 2010 (Richard *et al.* 2017c).

References

- Achtman M (2008) Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu. Rev. Microbiol.* **62**, 53-70.
- Achtman M (2012) Insights from genomic comparisons of genetically monomorphic bacterial pathogens. *Philos Trans R Soc Lond B Biol Sci* **367**, 860-867.
- Anderson PK, Cunningham AA, Patel NG, *et al.* (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in ecology & evolution* **19**, 535-544.
- Aubert B (2014) Vergers de la Réunion et de l'Océan Indien. In: *Hommes et fruits en pays du Sud* (ed. CIRAD M, France), pp. 111-165.
- Baker S, Thomson N, Weill F-X, Holt KE (2018) Genomic insights into the emergence and spread of antimicrobial-resistant bacterial pathogens. *Science* **360**, 733-738.
- Baltrus DA, McCann HC, Guttman DS (2017) Evolution, genomics and epidemiology of *Pseudomonas syringae*: challenges in bacterial molecular plant pathology. *Molecular plant pathology* **18**, 152-168.
- Bankevich A, Nurk S, Antipov D, *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology* **19**, 455-477.
- Barash I, Manulis-Sasson S (2009) Recent evolution of bacterial pathogens: the gall-forming *Pantoea agglomerans* case. *Annual Review of Phytopathology* **47**, 133-152.
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120.
- Bui Thi Ngoc L, Verniere C, Vital K, *et al.* (2009) Development of 14 minisatellite markers for the citrus canker bacterium, *Xanthomonas citri* pv. *citri*. *Molecular Ecology Resources* **9**, 125-127.
- Campbell PC (1923) *Chinese coolie emigration to countries within the British empire*.
- Carter MT, K. (2002) *Coolitude: an anthology of the Indian labour diaspora*.
- Chen NW, Serres-Giardi L, Ruh M, *et al.* (2018) Horizontal gene transfer plays a major role in the pathological convergence of *Xanthomonas* lineages on common bean. *Bmc Genomics* **19**, 606.
- Cruz CMV, Bai J, Oña I, *et al.* (2000) Predicting durability of a disease resistance gene based on an assessment of the fitness loss and epidemiological consequences of avirulence gene mutation. *Proceedings of the National Academy of Sciences* **97**, 13500-13505.
- Didelot X, Maiden MCJ (2010) Impact of recombination on bacterial evolution. *Trends in Microbiology* **18**, 315-322.
- Didelot X, Wilson DJ (2015) ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Computational Biology* **11**, e1004041.
- Drummond AJ, Rambaut A (2007) BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 214.
- Du Pont de Nemours PS (1797) *Oeuvres complètes de P. Poivre, intendant des isles de France et de Bourbon, correspondant de l'académie des sciences, etc.*
- Duchene S, Duchene D, Holmes EC, Ho SY (2015) The performance of the date-randomization test in phylogenetic analyses of time-structured virus data. *Molecular Biology and Evolution* **32**, 1895-1906.
- Duchene S, Holt KE, Weill FX, *et al.* (2016) Genome-scale rates of evolutionary change in bacteria. *Microb Genom* **2**, e000094.
- Eldholm V, Balloux F (2016) Antimicrobial resistance in *Mycobacterium tuberculosis*: the odd one out. *Trends in Microbiology* **24**, 637-648.

- Ferreira RM, de Oliveira AC, Moreira LM, *et al.* (2015) A TALE of transposition: Tn3-like transposons play a major role in the spread of pathogenicity determinants of *Xanthomonas citri* and other xanthomonads. *MBio* **6**, e02505-02514.
- Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.
- Gochez AM, Huguet-Tapia JC, Minsavage GV, *et al.* (2018) Pacbio sequencing of copper-tolerant *Xanthomonas citri* reveals presence of a chimeric plasmid structure and provides insights into reassortment and shuffling of transcription activator-like effectors among *X. citri* strains. *Bmc Genomics* **19**, 16.
- Gogarten JP, Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* **3**, 679-687.
- Gordon JL, Lefeuvre P, Escalon A, *et al.* (2015) Comparative genomics of 43 strains of *Xanthomonas citri* pv. *citri* reveals the evolutionary events giving rise to pathotypes with different host ranges. *Bmc Genomics* **16**, 1098.
- Goss EM (2015) Genome-enabled analysis of plant-pathogen migration. *Annual Review of Phytopathology* **53**, 121-135.
- Govindin S-S (1994) *Les engagés indiens: Ile de la Réunion, XIXe siècle*.
- Graham JH, Gottwald TR, Cubero J, Achor DS (2004) *Xanthomonas axonopodis* pv. *citri*: factors affecting successful eradication of citrus canker. *Mol. Plant Pathol.* **5**, 1-15.
- Ho SY, Lanfear R, Bromham L, *et al.* (2011) Time-dependent rates of molecular evolution. *Molecular Ecology* **20**, 3087-3101.
- Hyatt D, Chen GL, Locascio PF, *et al.* (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *Bmc Bioinformatics* **11**, 119.
- Jacques MA, Arlat M, Boulanger A, *et al.* (2016) Using ecology, physiology, and genomics to understand host specificity in *Xanthomonas*. *Annual Review of Phytopathology* **54**, 163-187.
- Kado CI (2015) Historical events that spawned the field of plasmid biology. In: *Plasmids: Biology and Impact in Biotechnology and Discovery*, pp. 3-11. American Society of Microbiology.
- Konstantinidis KT, Ramette A, Tiedje JM (2006) The bacterial species definition in the genomic era. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **361**, 1929-1940.
- Kousik C, Ritchie D (1996) Race shift in *Xanthomonas campestris* pv. *vesicatoria* within a season in field-grown pepper. *Phytopathology (USA)*.
- Kozlov A, Darriba D, Flouri T, Morel B, Stamatakis A (2018) RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. *bioRxiv*, 447110.
- Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B (2016) PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* **34**, 772-773.
- Leduc A, Traore YN, Boyer K, *et al.* (2015) Bridgehead invasion of a monomorphic plant pathogenic bacterium: *Xanthomonas citri* pv. *citri*, an emerging citrus pathogen in Mali and Burkina Faso. *Environmental Microbiology* **17**, 4429-4442.
- Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.
- Llop P, Barbé S, López MM (2012) Functions and origin of plasmids in *Erwinia* species that are pathogenic to or epiphytically associated with pome fruit trees. *Trees (Berlin, Germany : West)* **26**, 31-46.
- Mansfield J, Genin S, Magori S, *et al.* (2012) Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular plant pathology* **13**, 614-629.

- Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: detection and analysis of recombination patterns in virus genomes. *Virus evolution* **1**.
- McCann HC, Li L, Liu Y, *et al.* (2017) Origin and evolution of the kiwifruit canker pandemic. *Genome Biology and Evolution* **9**, 932-944.
- McDonald BA, Stukenbrock EH (2016) Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. *Phil. Trans. R. Soc. B* **371**, 20160026.
- McManus PS, Stockwell VO, Sundin GW, Jones AL (2002) Antibiotic use in plant agriculture. *Annual Review of Phytopathology* **40**, 443-465.
- Monteil CL, Yahara K, Studholme DJ, *et al.* (2016) Population-genomic insights into emergence, crop adaptation and dissemination of *Pseudomonas syringae* pathogens. *Microb Genom* **2**, e000089.
- Nascimento M, Sousa A, Ramirez M, *et al.* (2016) PHYLOViZ 2.0: providing scalable data integration and visualization for multiple phylogenetic inference methods. *Bioinformatics* **33**, 128-129.
- Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289-290.
- Pele J, Becu JM, Abdi H, Chabbert M (2012) Bios2mds: an R package for comparing orthologous protein families by metric multidimensional scaling. *Bmc Bioinformatics* **13**, 133.
- Pilla G, Tang CM (2018) Going around in circles: virulence plasmids in enteric pathogens. *Nature Reviews Microbiology*, 1.
- Polz MF, Alm EJ, Hanage WP (2013) Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet* **29**, 170-175.
- Pruvost O, Magne M, Boyer K, *et al.* (2014) A MLVA genotyping scheme for global surveillance of the citrus pathogen *Xanthomonas citri* pv. *citri* suggests a worldwide geographical expansion of a single genetic lineage. *Plos One* **9**, e98129.
- Pruvost O, Robene I, Escalon A, *et al.* (2015) The dynamic world of the genus *Xanthomonas*. In: *Virulence mechanisms of plant pathogenic bacteria* (ed. Wang Nian JJBSGWFFHSARCDLFLHJH), pp. 381-385. The American Phytopathological Society, Saint Paul, États-Unis.
- Ranjbar R, Karami A, Farshad S, Giammanco G, Mammina C (2014) Typing methods used in the molecular epidemiology of microbial pathogens: a how-to guide. *New Microbiologica* **37**, 1-15.
- Richard D, Boyer C, Javegny S, *et al.* (2016) First report of *Xanthomonas citri* pv. *citri* pathotype A causing asiatic citrus canker in Martinique, France. *Plant Disease* **100**, 1946-1946.
- Richard D, Boyer C, Vernière C, *et al.* (2017a) Complete genome sequences of six copper-resistant *Xanthomonas citri* pv. *citri* strains causing asiatic citrus canker, obtained using long-read technology. *Genome Announcements* **5**.
- Richard D, Ravigne V, Rieux A, *et al.* (2017b) Adaptation of genetically monomorphic bacteria: evolution of copper resistance through multiple horizontal gene transfers of complex and versatile mobile genetic elements. *Mol Ecol* **26**, 2131-2149.
- Richard D, Tribot N, Boyer C, *et al.* (2017c) First report of copper-resistant *Xanthomonas citri* pv. *citri* pathotype A causing asiatic citrus canker in Réunion, France. *Plant Disease*.
- Rieux A, Balloux F (2016) Inferences from tip-calibrated phylogenies: a review and a practical guide. *Molecular Ecology* **25**, 1911-1924.
- Rieux A, Khatchikian CE (2017) TipDatingBeast: An R package to assist the implementation of phylogenetic tip-dating tests using BEAST. *Molecular Ecology Resources* **17**, 608-613.

- Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584.
- Steinegger M, Söding J (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature biotechnology* **35**, 1026.
- Sundin GW, Wang N (2018) Antibiotic resistance in plant-pathogenic bacteria. *Annual Review of Phytopathology*.
- Swarup S, De Feyter R, Brlansky RH, Gabriel DW (1991) A pathogenicity locus from *Xanthomonas citri* enables strains from several pathovars of *X. campestris* to elicit cankerlike lesions on citrus. *Phytopathology* **81**, 802-809.
- Thynne E, McDonald MC, Solomon PS (2015) Phytopathogen emergence in the genomics era. *Trends in Plant Science* **20**, 246-255.
- Valinsky L, Manulis S, Nizan R, Ezra D, Barash I (1998) A pathogenicity gene isolated from the pPATH plasmid of *Erwinia herbicola* pv. *gypsophillae* determines host specificity. *Mol Plant Microbe Interact* **11**, 753-762.
- Verniere C, Ngoc LBT, Jarne P, *et al.* (2014) Highly polymorphic markers reveal the establishment of an invasive lineage of the citrus bacterial pathogen *Xanthomonas citri* pv. *citri* in its area of origin. *Environmental Microbiology* **16**, 2226-2237.
- Vinatzer BA, Monteil CL, Clarke CR (2014) Harnessing population genomics to understand how bacterial pathogens emerge, adapt to crop hosts, and disseminate. *Annual Review of Phytopathology* **52**, 19-43.
- Wiedenbeck J, Cohan FM (2011) Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiology Reviews* **35**, 957-976.
- Wu GA, Terol J, Ibanez V, *et al.* (2018) Genomics of the origin and evolution of *Citrus*. *Nature* **554**, 311.
- Yoshida K, Schuenemann VJ, Cano LM, *et al.* (2013) The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *Elife* **2**, e00731.

Chapitre III Analyse des transferts horizontaux chez les Xanthomonadaceae

Chapitre III Analyse des transferts horizontaux chez les *Xanthomonadaceae*

Les résultats des chapitres précédents ont mis en évidence l'importance des gènes plasmidiques dans l'évolution de plusieurs espèces de *Xanthomonas*. Très étudiés chez les organismes nosocomiaux et souvent responsables de leur résistance aux antibiotiques, les transferts horizontaux de gènes plasmidiques sont moins bien caractérisés chez les bactéries phytopathogènes et les bactéries associées aux plantes en général. Les résultats du Chapitre I ont mis en évidence une limite taxonomique à la diffusion du plasmide de résistance au cuivre. Dans le Chapitre II, l'analyse du contenu en gènes de plus de 200 souches de Xcc a permis d'identifier plusieurs plasmides connus chez *Xanthomonas*. Cette analyse n'a cependant mis en évidence qu'un seul plasmide non classiquement inféodé au genre *Xanthomonas*, et aucun en dehors de la famille. Nous avons également révélé l'importance du mosaïsme plasmidique suggérant l'implication d'éléments génétiques imbriqués.

Au vu de ces résultats, une caractérisation de la dynamique d'échanges des gènes plasmidiques chez les bactéries phytopathogènes de la famille des *Xanthomonadaceae* a été menée par une méthode de réseaux. L'application de ce type d'approche a par ailleurs déjà permis de caractériser plusieurs aspects des dynamiques évolutives bactériennes (Halary *et al.* 2010 ; Kloesges *et al.* 2011; Popa & Dagan 2011; Popa *et al.* 2011).

C'est ainsi que nous avons construit un réseau de similarité de gènes comprenant l'ensemble des gènes plasmidiques connus de la famille des *Xanthomonadaceae* ainsi que leurs homologues disponibles les bases de données NCBI afin (i) d'évaluer l'importance de l'échange de gènes plasmidiques (ii) de définir les facteurs qui structurent ou les limitent ces échanges et (iii) d'en identifier les véhicules. La recherche de gènes homologues à ceux encodés par les plasmides de 160 souches de *Xanthomonadaceae* a ainsi inclus dans notre analyse 481 souches de *Xanthomonas*, soit 61,5% des génomes du genre présents dans les bases de données utilisées. Les caractéristiques du jeu de données ainsi créé suggèrent à elles seules la forte prévalence des gènes plasmidiques au sein de la famille. L'inclusion de la majorité des souches du jeu de données au sein d'un même composant connecté illustre théoriquement les possibilités d'échanges de gènes plasmidiques entre souches de la famille, avec un ou plusieurs intermédiaires.

Au sein du réseau, des sous-structures, appelées communautés, regroupent des souches bactériennes partageant plus de gènes plasmidiques entre elles qu'avec les autres membres du réseau. Nous avons interprété la faible diversité taxonomique de ces communautés ainsi que la faible proportion de souches non-Xanthomonadaceae dans la totalité du réseau comme traduisant une forte limitation taxonomique à l'échange de gènes plasmidiques. Malgré la faible caractérisation de la niche écologique des souches étudiées, nos données suggèrent également un rôle de l'habitat dans la structuration des échanges de gènes plasmidiques.

Le rôle des gènes plasmidiques dans l'adaptation bactérienne est illustré par les fonctions des gènes plasmidiques fréquemment échangés au sein de la famille, qui comprennent (i) la résistance au cuivre et à la streptomycine, tous deux utilisés en agriculture, (ii) des déterminants de l'interaction hôte-pathogène et (iii) des fonctions liées à la mobilité d'éléments génétiques tels que les plasmides ou les transposons. L'importance des communautés bactériennes et des éléments génétiques mobiles est souvent mentionnée lors d'études visant à comprendre les mécanismes de résistance des bactéries. Nos résultats montrent que ces constats, provenant souvent du domaine de la santé humaine, sont également valables pour les bactéries phytopathogènes. Nous ajoutons que l'échange de gènes entre les compartiments plasmidiques et chromosomiques lié à l'association entre éléments génétiques à mobilité inter et intra-cellulaire est probablement une composante importante de l'adaptation bactérienne.

The structure of bacterial communities as seen through the sharing of Xanthomonadaceae plasmid genes

D. Richard^{1,2,3,†}, P. Roumagnac⁴, O. Pruvost¹, P. Lefeuvre^{1,†}

¹ Cirad, UMR PVBMT, F-97410 St Pierre, Réunion, France

² ANSES, Plant Health Laboratory, F-97410 St Pierre, Réunion, France

³ Université de la Réunion, UMR PVBMT, F-97490 St Denis, Réunion, France

⁴ Cirad, UMR BGPI, F-34000 Montpellier, France

[†]Corresponding authors: damien.richard@cirad.fr

Abstract

Whereas horizontal gene transfer has played a cardinal role on the evolution of bacterial lineages over large evolutionary scales, it is also known to provide fast adaptation potential. Plasmids, that frequently bear adaptive genes, provide an efficient vehicle for gene sharing among bacterial populations. Understanding the structure of plasmid gene sharing events is then crucial to apprehend their implication in bacterial evolution. We herein use a gene sharing bipartite network to characterize the sharing of plasmid genes among the Xanthomonadaceae, a family encompassing plant pathogens, opportunistic human pathogens responsible for nosocomial infections and commensal species. Interestingly several antibiotic and metal resistance genes were encoded on the plasmidome of the Xanthomonadaceae and were part of the most shared gene set, strongly supporting the ability of plasmid to function as adaptive elements. Often debated, the extent to which bacterial characteristics limit the sharing of bacterial genes is in our dataset strongly driven by taxonomy. Despite often being mobile, plasmids surprisingly didn't represent a well-supported transfer unit. On the contrary, the versatility of plasmid genes and their organization into sub-elements is depicted. Importantly, marked connections between such sub-elements and chromosomal compartments were detected indicating that gene sharing between plasmid and chromosome is not rare. Globally our study highlights the importance of human-mediated selective pressure over the adaptive gene set that circulate within the Xanthomonadaceae populations.

Keywords: plasmid, horizontal gene transfer, network of gene sharing, *Xanthomonas*, Xanthomonadaceae

Introduction

One of the most striking revelations after two decades of whole bacterial genome sequencing is the great variability in gene content that two closely related bacteria can exhibit (Baltrus *et al.* 2017; Cordero & Polz 2014; Hyun *et al.* 2012; Shapiro *et al.* 2012; Tenaillon *et al.* 2010). Seminal work on genome content evolution soon indicated that these variations can arise at an unexpectedly rapid pace relative to mutations (Ochman *et al.* 2000). Later, more holistic studies extended this discovery to the whole bacterial history and clearly demonstrated that a great part of bacterial genomes has been acquired by horizontal gene transfer (HGT) (Dagan *et al.* 2008; Wiedenbeck & Cohan 2011) to such extent that it was paramount in several speciation events (De la Cruz & Davies 2000; Gogarten & Townsend 2005; Ochman *et al.* 2000). HGT is also an important actor of bacterial evolution over shorter evolutionary timescales, allowing strains to acquire beneficial traits from different lineages in one generation (Thomas & Nielsen 2005). This fast adaptive mechanism allows rapid adaptation to environmental changes, making HGT events often involved in the acquisition of bio-degradation pathways, determinants of resistance or virulence (De la Cruz & Davies 2000).

In bacteria, four mechanisms of horizontal gene transfer have been identified to date: transformation, conjugation, transduction and more recently gene transfer agents (Lang *et al.* 2017; Thomas & Nielsen 2005). While a great body of work has been devoted to understanding the functioning of the genetic machinery involved in each process (Hynes *et al.* 2016; Popa & Dagan 2011; Seitz & Blokesch 2013; Smillie *et al.* 2010; Touchon *et al.* 2017), the factors that delineate the transfer community are still to be fully apprehended. Most notably, whereas there is evidence that characteristics such as taxonomy, genome GC content, codon usage or ecological niche might restrict gene sharing between individuals, the relative importance of each of these barriers remain the subject of an intense debate (McInerney 2013; Polz *et al.* 2013; Popa & Dagan 2011; Popa *et al.* 2011; Skippington & Ragan 2012).

Among the HGT vehicles, plasmids were depicted as cardinal for bacteria (Corel *et al.* 2016; Halary *et al.* 2010; Popa & Dagan 2011). Their intercellular mobility and their capacity to transmit ready-to-use functional pathways is often highlighted as the cause of their success. Some bacterial genera are prone to acquire genes through phage integration (*i.e.*, transduction, such as in *Streptococcus*) or transformation (Dorer *et al.* 2010). In other taxa, including the *Xanthomonas* genus, HGT was demonstrated to mainly occur through plasmid exchange (Halary *et al.* 2010). The *Xanthomonas* genus comprises numerous plant pathogenic species

that collectively cause diseases on a wide range of crops in agro-ecosystems, suggesting an ability to adapt to contrasting environmental conditions (Leyns *et al.* 1984). Species typically unify several pathologically highly specialized lineages (*i.e.*, the so-called pathovars), several of which were reported as genetically monomorphic, with no or very little polymorphism revealed when only a few loci are analyzed (*e.g.*, when using MultiLocus Sequence Typing, MLST) (Garita-Cambronero *et al.* 2016; Nakato *et al.* 2018; Pruvost *et al.* 2014). Plasmids encoding adaptive genes including essential virulence genes (Chen *et al.* 2018; Yang 1994), antibiotics (Hyun *et al.* 2012; Minsavage *et al.* 1990) or metal resistance (Richard *et al.* 2017) are ubiquitous within all members of this genus. Plasmids are likely to have played a key-role in adaptive evolution of xanthomonads. Within the Xanthomonadaceae family (a family belonging to Gammaproteobacteria), several plasmid-borne resistances to antibiotics were reported in *Stenotrophomonas maltophilia*, an opportunistic pathogen associated with nosocomial human infections (Avison *et al.* 2001; Kanamori *et al.* 2015). Similarly, *Xylella fastidiosa* also hosts plasmids, at least one of which contains genes involved in resistance to toxic compounds (Rogers & Stenger 2012).

Using comparative genomics, the analysis of the molecular basis of copper resistance in *Xanthomonas citri* pv. *citri* strains revealed the plasmidic support of a copper resistance gene cluster. The extension of the resistance further than the *X. citri* species revealed that this gene cluster was encoded in several bacteria from the Xanthomonadaceae family, either integrated in a highly homologous or distinct plasmid backbone (*Xanthomonas* spp.) or in the chromosome (*Stenotrophomonas* spp.) (Richard *et al.* 2017). Besides confirming the high prevalence of plasmids within the family, these findings suggest a large modularity of the vehicles. In order to define whether these findings can be extended to the Xanthomonadaceae plasmidome and to properly estimate the barriers mostly influencing the plasmid transfer patterns, we analyzed the sharing of homologous plasmid genes using a gene-sharing network. After having estimated the degree to which plasmids are conserved and shared within the family, we assessed the influence of the ecological niche and the taxonomy over the patterns of gene sharing. We latter intend to estimate the genericity of links between plasmid encoded genes and chromosome.

Materials and methods

Plasmid reference sequences

The complete set of plasmid sequences from the twelve genera of the Xanthomonadaceae family (*sensu* Naushad *et al.* (2015)) were downloaded on November 2017 from the NCBI non-redundant (NR) and Whole Genome Shotgun (WGS) databases (n=441, obtained from 160 distinct individuals (*i.e.* NCBI biosamples)). For suspicion of being chromosomal rather than plasmidic, contigs encoding for genes typically chromosome-encoded (*e.g.* those encoding ribosomal proteins) were discarded, as well as sequences that were not assembled as circular molecules. The remaining 140 individuals comprised 305 contigs (Supplementary Table S1) from which 18,715 genes were predicted using Prodigal v2.6.2 (Hyatt *et al.* 2010) with default parameters. These genes were then clustered into 8,197 groups of 100% identical sequences using the derep_fulllength algorithm available in VSEARCH v1.9.5 (Rognes *et al.* 2016).

Search of homologous sequences in NCBI

Clustered sequences were used as a query for a BLASTn search against NR and WGS databases (as of November 2017) using default parameters except for max_target_seqs that was set to 10,000,000. All NR and WGS contigs matching at least one gene on a minimum of 90% of its length and with a minimum of 90% of identity were downloaded. In order to evaluate the robustness of the gene sharing network to the similarity threshold we used, we constituted subsets of the BLASTn hit table with different levels of stringency for the minimum identity (%id) and the length percentage of the matched query (%len). Six combinations (%id / %len) of these two filters were used: 90/90; 95/90; 95/95; 98/90; 98/98 and 99.8/90. Genes were extracted based on GenBank gene annotation when available or conversely after structural annotation using Prodigal v2.6.2 as described above. Including the query genes, a total of 87,421 genes from 14,576 contigs and 10,112 biosamples were obtained (Supplementary Table S2). Along with the available GenBank functional annotations, all these genes were functionally annotated using a BLASTp search against the Swiss-Prot and TrEMBL databases (O'Donovan *et al.* 2002).

Detection of gene sharing relations

The whole set of gene sequences along with plasmidic query genes were then clustered using the "linclust" algorithm of MMseqs2 (Steinegger & Söding 2017). The coverage (c parameter)

and the minimum identity (min-seq-id parameter) were both set according to values of the corresponding BLASTn filters (%id and %len) previously described. Gene clusters that did not comprise any plasmidic query gene were removed, resulting in a total of 8,197 gene clusters (at >99.8% identity along >90% of query length, the number of gene clusters obtained with other thresholds are detailed in Supplementary Table S2). Genes that were assigned to the same cluster were considered homologous.

Bipartite network of gene sharing

Bipartite networks were then constructed using the R package igraph (Csardi & Nepusz 2006). Two classes of nodes coexist in a bipartite network: individuals and gene clusters. Whereas no direct link exists between two nodes of the same class, individuals sharing a gene cluster are connected to this same gene cluster node and conversely a gene cluster is connected to all the individuals in which it was detected. In order to reduce the redundancy of the dataset due to unbalanced sequencing efforts and to limit the inclusion of vertical inheritance rather than horizontal transfer, individuals from the same species and containing the same gene clusters were merged. This made our network conservative in that it globally represents the minimum number of horizontal transfer that must have occurred to explain the current plasmid distribution. Community detection (*i.e.* groups of nodes more connected between each other than they are to the rest of the network) was performed using the Louvain algorithm (Blondel *et al.* 2008) as implemented in the igraph R package. Importantly, as the Louvain algorithm is not suitable for bipartite network community detection, the network was first projected onto the biosamples (*i.e.* only biosample nodes remained and the weight of an edge between two biosamples corresponded to the number of gene clusters they share).

Chromosomal integration inference

In order to identify putative chromosomal support of genes from the plasmid query set, analysis of the size of the contigs encoding these genes was performed. Because of the relative unfinished nature of the genomic project available on the WGS and NR databases and in order to avoid over-estimation of the contig sizes due to improper scaffolding (*e.g.* chromosomal scaffolds that would contain plasmidic contigs), all contigs were split at every occurrence of long N sequences (indicative of contig junction) before estimating the size of the genomic fragment that displays similarities with plasmid genes. We then searched for instances where more than five genes from a given query plasmid contig were found within a single contig of

more than 1,000 genes. Importantly, a permutation test was implemented to ensure that these genes were clustered (encoded close to each other) within both bearing contigs. This first scan allowed the detection of the largest and most conserved Inter-Genomic Compartment Gene sharing (IGCG) between a chromosome and a plasmid. Every IGCG was then analyzed manually and extended to other individuals displaying similar but weakest signals.

Results & discussion

The Xanthomonadaceae Plasmids database

The query dataset comprised 18,715 genes encoded on 305 plasmidic contigs from 140 Xanthomonadaceae individuals (*i.e.* NCBI biosamples) (Supplementary Table S1). Individuals belonged to 19 Xanthomonadaceae species, mostly *X. citri* (49%), *X. fastidiosa* (15%) and *X. campestris* (9%). The remaining 16 species represent 5% or less of the database each. Plasmid size was very variable, with small (<10kb) to intermediate size plasmids (<100kb) being the most represented (20% and 73% respectively). Nevertheless, 22 plasmids larger than 100kb were obtained and represented 29% of the analyzed gene pool. Whereas 53 individuals hosted a single plasmid, all others had more than one plasmid with a mean of 2.9 plasmids and a maximum of nine plasmids for three strains of *X. campestris*.

On the 18,715 proteins, 41% presented close homologs in the Swiss-Prot and TrEMBL database and were annotated. Most remained uncharacterized. Interestingly, when annotation was possible, enrichments in avirulence (n=399), plasmid persistence (n=1,157), transposition (n=2,077) and conjugation (n=2,533) functions were detected. Conjugative apparatus genes were detected in 59% of the plasmids while a toxin-antitoxin system was present in 50%.

Table 1: Number of biosample per taxonomical groups sharing genes with the Xanthomonadaceae plasmids at the 99.8% identity and 90% length threshold.

Taxonomical level	Taxonomy	Nb of biosample sharing Xanthomonadaceae plasmid genes	Nb of biosample in Nr and WGS databases	% of Nr/WGS biosamples included in the network
Family	Enterobacteriaceae	6036	164587	3,7%
	Vibrionaceae	370	5680	6,5%
	Pseudomonaceae	707	10810	6,5%
	Moraxellaceae	1127	6386	17,6%
	Xanthomonadaceae	621	1506	41,2%
Genus	<i>Salmonella</i>	1444	93616	1,5%
	<i>Burkholderia</i>	93	3611	2,6%
	<i>Escherichia</i>	1793	54468	3,3%
	<i>Aeromonas</i>	20	501	4,0%
	<i>Pseudomonas</i>	708	10867	6,5%
	<i>Vibrio</i>	370	5307	7,0%
	<i>Citrobacter</i>	37	512	7,2%
	<i>Serratia</i>	35	402	8,7%
	<i>Proteus</i>	18	202	8,9%
	<i>Shigella</i>	513	5493	9,3%
	<i>Corynebacterium</i>	109	1120	9,7%
	<i>Acinetobacter</i>	1126	6279	17,9%
	<i>Klebsiella</i>	1748	9685	18,0%
	<i>Stenotrophomonas</i>	92	489	18,8%
	<i>Sphingobium</i>	21	102	20,6%
	<i>Enterobacter</i>	484	1946	24,9%
	<i>Xyllela</i>	46	94	48,9%
	<i>Xanthomonas</i>	481	782	61,5%

These systems were detected on plasmids from intermediate to large size. Our query dataset comprised four antibiotic-resistance (acriflavin, beta-lactam, bleomycin and streptomycin) genes encoded on three plasmids. Genes conferring acriflavin resistance, which encode a multi-drug efflux system that is believed to protect bacteria against a broad range of dyes, detergents, antibiotics and plant phytoalexins, have been previously detected in *X. arboricola* and *X. campestris*, sometimes in association with an integrative conjugative element (ICE) on

which copper resistance genes were identified (Cesbron *et al.* 2015; He *et al.* 2006). The reasons why bleomycin resistance was over-represented are presently unknown and warrant further investigation. In addition, seven metal-resistance (arsenic, cadmium, cobalt, copper, mercury, silver and zinc) genes were identified from 13 plasmids. Although heavy metals other than copper are of little or no use in agriculture (Hobman & Crossman 2015), the over-representation of resistance to these compounds can likely be explained by the frequent co-localization of systems encoding for different metals on a same plasmid (Gullberg *et al.* 2014; Staehlin *et al.* 2016).

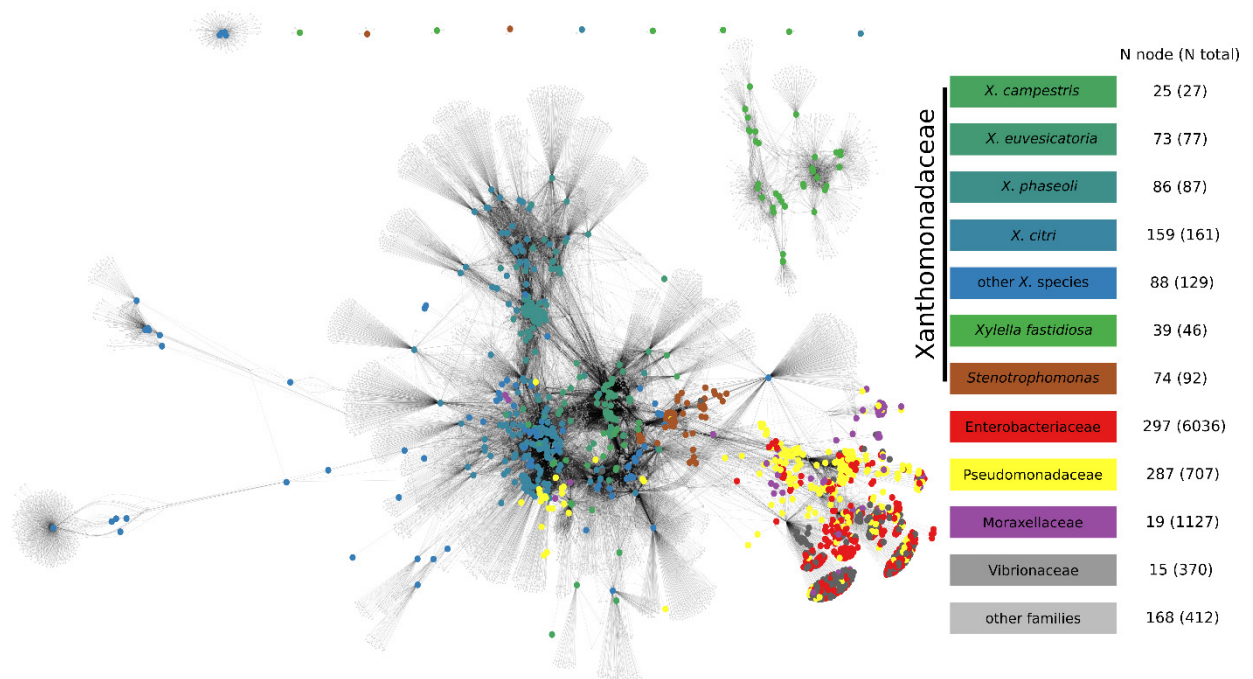


Figure 1: Gene sharing bipartite network of the Xanthomonadaceae plasmidic genes. Large nodes represent bacterial strains whereas smaller sized grey nodes represent gene clusters. Bacterial nodes are colored according to the taxonomy of the strain it represents. An edge between a bacterial node and a cluster gene node indicates that the bacterial genome comprises a gene of this cluster. When multiple individuals of the same bacterial species encode for the same gene families, only one node is represented on the network. For several taxon levels, the number of unique bacterial nodes is given on the right side of the legend while the total number of corresponding bacteria is in brackets.

A total of 86 plasmids from 11 species encoded for genes associated to pathogenicity. Among these, 42 gene clusters corresponded to TALE effector genes that were encoded on 28 plasmids. Collectively, our results suggest an over-representation of genes associated with (i) intrinsic plasmid functions such as partitioning, conjugation, replication and maintenance, or mobile genetic element functions such as transposition, (ii) resistance against anthropic pesticide (copper, streptomycin) used in agriculture, and (iii) host-pathogen interactions.

Plasmid genes largely connect individuals within the Xanthomonadaceae family

In order to explore the robustness of the gene sharing patterns depending on the threshold used to define gene clusters, we used different clustering thresholds, from relaxed to more stringent ones. Importantly, whereas more genes and individuals were obtained for the less stringent parameter combination (Supplementary Table S2), the global pattern of gene sharing remains fully consistent from one threshold to the other. Similar network structures were globally obtained with a main connected component (CC, *i.e.* group of connected nodes) that gathered most of the individuals and only weakly connected non-Xanthomonadaceae species. Depending on the threshold, the number of those non-Xanthomonadaceae species varied markedly. While the proportion of Xanthomonadaceae species within the network varied from 38 to 53%, the numbers of individuals remained stable. It is indeed the non-Xanthomonadaceae prevalence that dramatically decreased with higher stringency.

Depending on the clustering threshold used, the global gene-sharing network comprised nine to 13 distinct CC. Regardless of the clustering threshold, every network comprised a main CC that gathered most of the strains (from 85.9% to 99.7%). Stringent clustering thresholds resulted in the gathering of most of the *Xylella fastidiosa* strains in an independent CC. Strains from this separated CC were connected to the main CC with less stringent clustering threshold, the number of genes involved remaining restricted nonetheless. The rest of the individuals (0.2% representing 17 individuals) were allocated into 11 CCs, two of which only shared four plasmids each and nine of which included only one plasmid, revealing the absence of homologous genes in the genomes hosted in the NR/WGS databases (see the unconnected components on the top of Figure 1). These 17 individuals belonged to *Stenotrophomonas*, *Xylella* and *Xanthomonas* genera and host small plasmids of less than four genes. In this study, we did not attempt to disentangle vertical and horizontal inheritance of genes. However, we only studied genes of plasmid origin and we merged individuals of the same species comprising the same gene clusters. The presence of phylogenetically related strains in different CC illustrates that our methodology didn't capture homology due to vertical inheritance. Most gene sharing events depicted here should therefore correspond to HGT. Globally, our results suggest that the Xanthomonadaceae plasmidome is highly mobile within the family.

Using our most conservative 99.8%id / 90%len threshold (meaning that in order to be included in the network, a gene had to match query genes with 99.8% identity on 90% of their length), 9,359 individuals representing 1,422 distinct species profiles, and coding for 8,197 gene clusters

were included in the network. Interestingly, while no gene was shared by all individuals, 3,944 (48%) genes were unique to an individual. The remaining 4,253 gene clusters were retrieved in average in 12 individuals with a maximum of 404 individuals. Additionally, gene clusters that were shared the most tended to be associated with plasmid stabilization, transposition, mercury resistance, streptomycin resistance, cellular signalization, response to oxidative stresses, replication, avirulence and conjugation. This stringent threshold yielded the biggest number of CC. Indeed, 36 query plasmids from eleven individuals had no connections with any other replicon.

Degree of plasmid conservation within the community

In the global network, most of the gene sharing events involved partial plasmid portions and most often small numbers of genes (Figure 2A), indicative of plasmid mosaicism, a common feature that has been documented in the *Xanthomonas* genus (Gochez *et al.* 2018; Richard *et al.* 2017). Several of the Xanthomonadaceae plasmids in our database were apparent chimeras relative to each other (see the module over the diagonal on Supplementary Figure 1B).

Only a fraction of the individuals shared full or almost full plasmids. Full plasmid sharing communities (clusters of red within Figure 2B) were in fact restricted to small numbers of individuals. In order to visualize plasmid sharing rather than gene sharing, networks representing the sharing of almost full (>90%) plasmids were computed (Figure 3). Importantly, whereas the existence of a link implied the sharing of >90% of the plasmid genes, we had no indication that they were actually carried on a single plasmid or if a similar gene content was encoded on distinct plasmid variants.

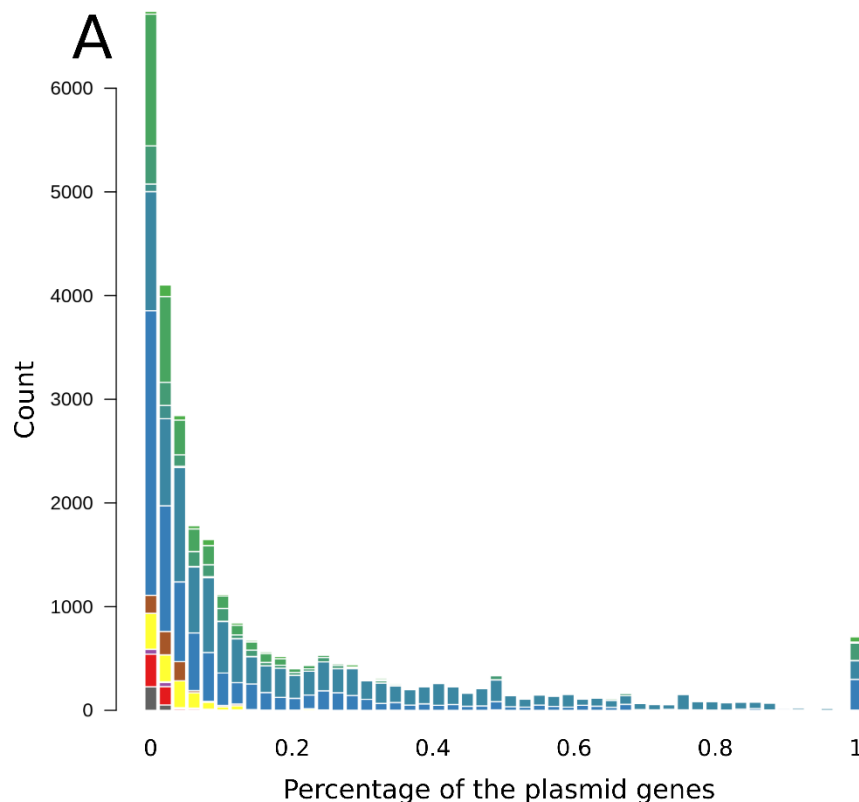
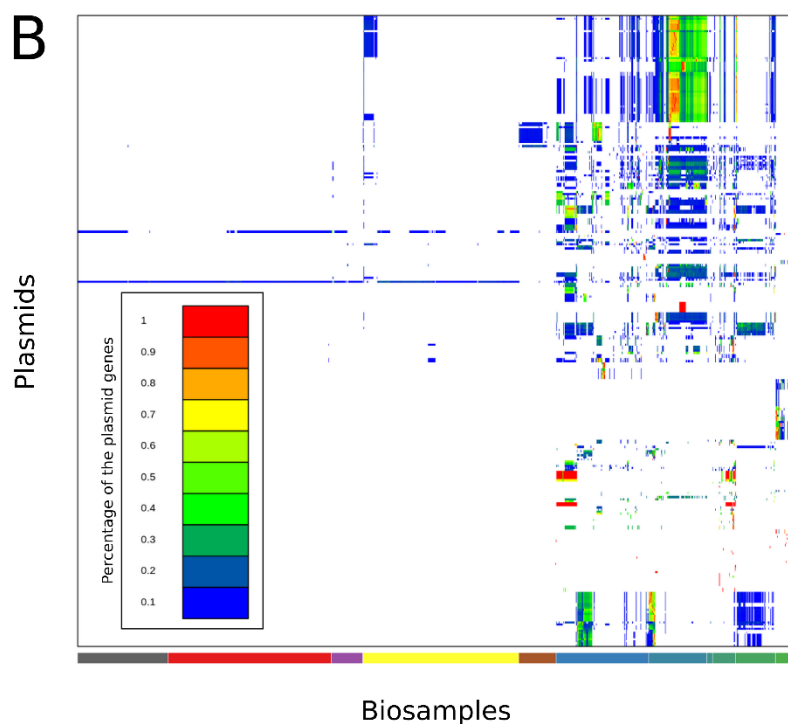


Figure 2: (A) Distribution of the proportion of shared plasmid genes for every pair of connected individuals.

The percentage of the plasmid genes shared is indicated on the x-axis, whereas the count is indicated of the y-axis.



Vertical bars are colored according to the taxonomy of the strain that shares the plasmid genes as depicted in Figure 1.

(B) Matrix representation of the proportion of plasmid genes shared between individuals.

Plasmids are represented in rows and individuals in columns. The proportion of shared plasmid genes is colored according to the scale on the bottom left of the matrix. Bacterial

taxonomy is represented with the colored line at the bottom of the figure according to the color defined in Figure 1.

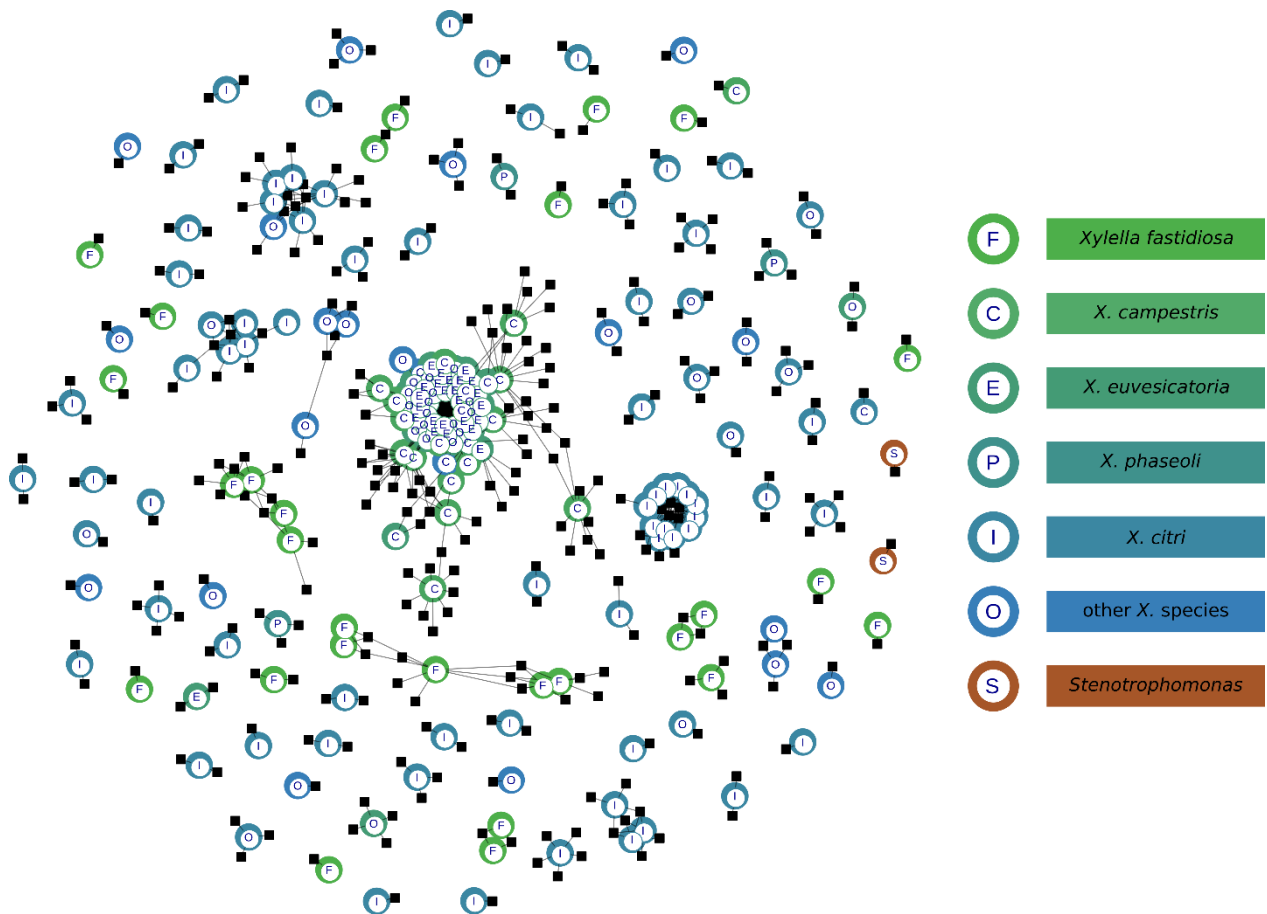


Figure 3: Bipartite network of full plasmid sharing. Plasmid nodes are represented with black squares and bacterial nodes are represented with colored circles. Edges represent events where a bacterial strain encodes at least 90% of the genes of a given plasmid. Colors and letters over bacterial nodes are function of the taxonomy of the bacterial genome, as depicted in the legend on the right of the figure.

Whereas several bacterial families were included in the global gene sharing network, only individuals of the Xanthomonadaceae family remained in the full plasmid sharing networks regardless of the threshold. This network was composed of 97 CC (99.8%id / 90%len threshold), with several small CC and only two groups that contained more than ten individuals. The first group encompassed 63 strains and 68 plasmids belonging to *X. euvesicatoria* and *X. campestris* pv. *campestris* (which infect Solanaceae and Brassicaceae plant families, respectively (Leyns *et al.* 1984)). This grouping is surprising because physical contact between both species is probably rare, therefore requiring an intermediate. Despite the narrow host-range of most phytopathogenic *Xanthomonas*, *X. campestris* pv. *raphani* displays the uncommon feature of being pathogenic to both plant families (Fargier & Manceau 2007; Punina *et al.* 2007), and could have an important role in this plasmid exchange hub. Only two available *X. campestris* pv. *raphani* genomes are available and were described as plasmid-free. More genome data is therefore required to address this hypothesis (Roux *et al.* 2015). The second

group, that comprised twelve strains of *Xanthomonas citri* pv. *citri*, shared pXCAW19 and pXCAW58 plasmids. This suggests plasmid horizontal transfer between *X. citri* pv. *citri* pathotypes, consistent with previous data that evidenced a large role of horizontal transfer, plasmid insertion and recombination in the genomic evolution of *X. citri* pv. *citri* pathotypes (Gordon *et al.* 2015).

Taxonomy shapes plasmid gene transfer

The use of plasmid genes from the Xanthomonadaceae family as queries for network construction implied that it only included bacterial genomes encoding close homologues of these genes. It therefore delineated the extent to which the genes from this family are present within other bacterial taxa. Importantly, the plasmid gene network was strongly restricted by taxonomic barriers whatever the identity threshold level used. Our approach retrieved 621 of the 1506 (41%) Xanthomonadaceae genomes available in the WGS and NR databases from GenBank. When focusing solely on the *Xanthomonas* genus, this percentage reached 84% (see Supplementary Table S2 for details). Conversely, beside the absence of full plasmid sharing events beyond the Xanthomonadaceae family, inspection of the full bipartite network (Figure 1) demonstrated that inter-family gene sharing was rare and restricted to a small number of genes. Gene sharing events between Xanthomonadaceae and non-Xanthomonadaceae individuals comprised a mean number of 2.4 genes (against 19.6 for intra-Xanthomonadaceae events). For families other than Xanthomonadaceae, the highest proportion of individuals included in the network was recorded for the Moraxellaceae, Vibrionaceae and Enterobacteriaceae (all of which are Gammaproteobacteria) for which, respectively 17.6, 6.5 and 3.7% of the individuals present in the databases were recovered (Table 1). It suggests that the plasmid gene pool is mostly restricted to the Xanthomonadaceae but highly shared within this family.

Going one step further, the analysis of the taxonomic features of the communities defined within the network was also indicative of taxonomic gene sharing restrictions. Communities (also frequently described as modules) group nodes (in this case individuals) that share more edges between each other than with the rest of the network. The six communities detected within the main CC of the full bipartite network (Figure 4) were mostly organized by species taxonomy, with every community being composed of one distinct main species (see acronym of the species at the bottom panel for the sections representing more than 5% of the individuals count). For example, community 1, 2 and 3, mostly comprised strains belonging to *X. phaseoli*,

X. euvesicatoria and *X. citri* species, respectively. Moreover, the second biggest CC of the network (community 7) was uniquely composed of 34 *Xylella fastidiosa* strains.

Rather than taxonomy itself, shared intrinsic characteristics of the genomes within taxonomical groups are thought to limit HGT. Indeed, the occurrence of HGT events between two cells is first conditioned by the success of the transfer itself. Several systems can limit this transfer. First, the mating-pair formation apparatus (MPFA), such as the well-known conjugative pilus, generates the junction between one cell to another, allowing plasmid transfer. The specificity of the interaction between the MPFA and the membrane constituents can be high, preventing the interaction between non-compatible cells (Thomas & Nielsen 2005). Secondly, the restriction-modification systems, comprising an endonuclease that cleaves particular DNA motifs and a modification enzyme which methylates these same motifs, restrict the acquisition of foreign DNA. These two systems tend to be shared by closely-related bacteria (Wiedenbeck & Cohan 2011). Similarly, CRISPR systems also protect bacteria against phage and plasmid DNA (Horvath & Barrangou 2010). Moreover, in order to be maintained into a lineage, transferred genes must be regulated and expressed efficiently (Popa *et al.* 2011). The taxonomical limit of HGT therefore probably includes multiple factors such as GC percentage and codon usage, whose adequation between donor and recipient cell favor usability of transmitted genes by the recipient's cellular machinery. These barriers tend to limit gene exchange between bacteria and make such exchanges between extremely divergent organisms rare (Popa *et al.* 2011). As a consequence, conjugation between bacteria of different Gram types is believed not to occur naturally (Sørensen *et al.* 2005).

Ecological niche: a secondary barrier of HGT events?

While the network structure is apparently strongly restricted by taxonomy, a finer analysis of the community group composition also suggested the implication of ecological niche in structuring gene exchange. However, the limited availability of information regarding ecological niches of the strains within the network restricted the possibilities of interpretations. Among Xanthomonadaceae analyzed in the present study, the *Xanthomonas* genus, when taken as a whole, is pathogenic to a large number of monocot and dicot plant species. However, at the strain level, these bacteria often exhibit a high host specialization towards a limited number of genetically related plant species or genera with different tissue specificity (*i.e.*, parenchyma vs. vascular pathogens) (Leyns *et al.* 1984). Their natural spread capacity is often restricted to distances over which contaminated water can disperse (Potnis *et al.* 2015; Serizawa & Inoue

1975). These two characteristics cause *Xanthomonas* populations to often experience small-scale geographic isolation. In some cases, distinct *Xanthomonas* species have the ability to colonize a same plant niche, a feature known as pathological convergence. Indeed, community 2 grouped two species pathogenic to tomato and/or pepper parenchyma such as *X. cynarae* pv. *gardneri*, and *X. euvesicatoria* pvs. *euvesicatoria* and *perforans* (Figure 4). Consistent with our results, the latter two pathovars were recently found to dynamically exchange DNA, including genes involved in pathogenicity (Jibrin *et al.* 2018). Similarly, community 1 was shared by two bean pathogens, *X. phaseoli* pv. *phaseoli* and *X. citri* pv. *fuscans*, which share the ability to colonize both parenchyma and the vascular system of bean. Pathological convergence in these two pathovars was recently emphasized as a likely consequence of HGT involving plasmids hosting transcription activator-like effector (TALE) genes (Chen *et al.* 2018; Ruh *et al.* 2017). These two examples also emphasized the probable interaction between taxonomic and niche barriers (and possibly other factors) in the delineation of communities.

Finally, within Community 4, a link between commensal and pathogenic bacteria was inferred. In fact, several *Stenotrophomonas* strains shared genes with plant pathogenic bacteria. *Stenotrophomonas* is a commensal bacterial genus often found in association with plants and soil, and whose involvement as a reservoir of adaptive genes has been suggested in studies focusing on the adaptation of plant or human pathogenic bacteria (Crossman *et al.* 2008).

Some communities were mainly structured by one or a few gene functions. Indeed, a great proportion of shared genes amongst community 3 were avirulence genes, consistent with the plasmidic nature of *Xanthomonas citri*'s known pathogenicity determinants. The sharing of antimicrobial resistance genes also appeared to strongly structure the network. Indeed, community 5 gathered most of the Enterobacteriaceae and Pseudomonadaceae of the network and the second most shared gene among the community was associated to a streptomycin resistance function. Moreover, members of community 2, grouping *Xanthomonas* species pathogenic to tomato and/or pepper, mostly shared copper (and to a lesser extent other heavy metals) resistance genes, which is congruent with our previous findings in which we evidenced the sharing of copper resistance genes between these species (Richard *et al.* 2017).

Table 2: Inter-Genomic Compartment Gene transfer events details.

IGCG ID	Plasmid taxonomy	Plasmid prototype accession	Chromosome taxonomy	Maximum nb of shared	IGCG main function(s)
1	<i>X. citri</i> / <i>X. euvesicatoria</i>	CP018464	<i>S. maltophilia</i> / <i>S. sp</i>	38	Heavy metals resistance
2	<i>X. vesicatoria</i>	CP018471	<i>P. aeruginosa</i> / <i>P. mendocina</i> / <i>P. pseudoalcaligenes</i> / <i>P.</i>	11	Mercuric resistance
3	<i>X. euvesicatoria</i>	CP018472	<i>X. euvesicatoria</i>	117	Conjugation
4	<i>X. citri</i>	CM002265	<i>X. campestris</i>	14	Conjugation
5	<i>X. citri</i>	CM007623	<i>X. citri</i>	22	Restriction-modification
6	<i>X. citri</i> / <i>X. phaseoli</i>	CP020966	<i>X. citri</i> / <i>X. phaseoli</i>	24	Toxin-antitoxin and conjugation
7	<i>X. fastidiosa</i>	CM003750	<i>X. fastidiosa</i>	35	Toxin-antitoxin and conjugation
8	<i>X. translucens</i>	CM003053	<i>X. translucens</i>	11	Conjugation
9	<i>X. cassavae</i> / <i>X. citri</i> / <i>X. oryzae</i>	AE008925	<i>X. cassavae</i> / <i>X. citri</i> / <i>X. oryzae</i>	33	Avirulence, toxin-antitoxin and conjugation

Multiple integrations of plasmid-borne genes into chromosomes

While plasmids are efficient vehicles for genes conferring a selective advantage in response of a biotic or abiotic selection pressure, plasmid maintenance in hosts is metabolically costly, potentially leading to the integration of beneficial genes originally carried on plasmids into the bacterial chromosome (Harrison & Brockhurst 2012). The analysis of the copper resistance in *Xanthomonas citri* strains previously revealed the plasmid support of the resistance, encoded into a transposon (Richard *et al.* 2017). Gene similarity analysis revealed the presence of this very transposon integrated into a distinct plasmid backbone in *Xanthomonas euvesicatoria* but also into the chromosome of a strain of a commensal *Stenotrophomonas* species. Indeed, the transfer of adaptive genes from plasmid to chromosome can reduce the burden of plasmid replication (Bennett & Richmond 1976; Dobrindt *et al.* 2004; Ravatn *et al.* 1998). The loss of a plasmid by a bacterial lineage, which can be impaired by plasmid maintenance systems such as toxin-antitoxin (TA) systems, becomes possible once a copy of the TA system is encoded on the chromosome (Van Melderren & Saavedra De Bast 2009). On the contrary, the integration of chromosomal genes into a plasmid may facilitate their spread in bacterial communities, as illustrated by the success of the plasmid/transposon association (Bennett 2004; Davies 1997; Liebert *et al.* 1999; Okinaka *et al.* 1999). We sought, from our dataset, to identify such transfers among replicons to evaluate whether or not they represent a common process within the Xanthomomadaceae and evaluate their importance within the bacterial community.

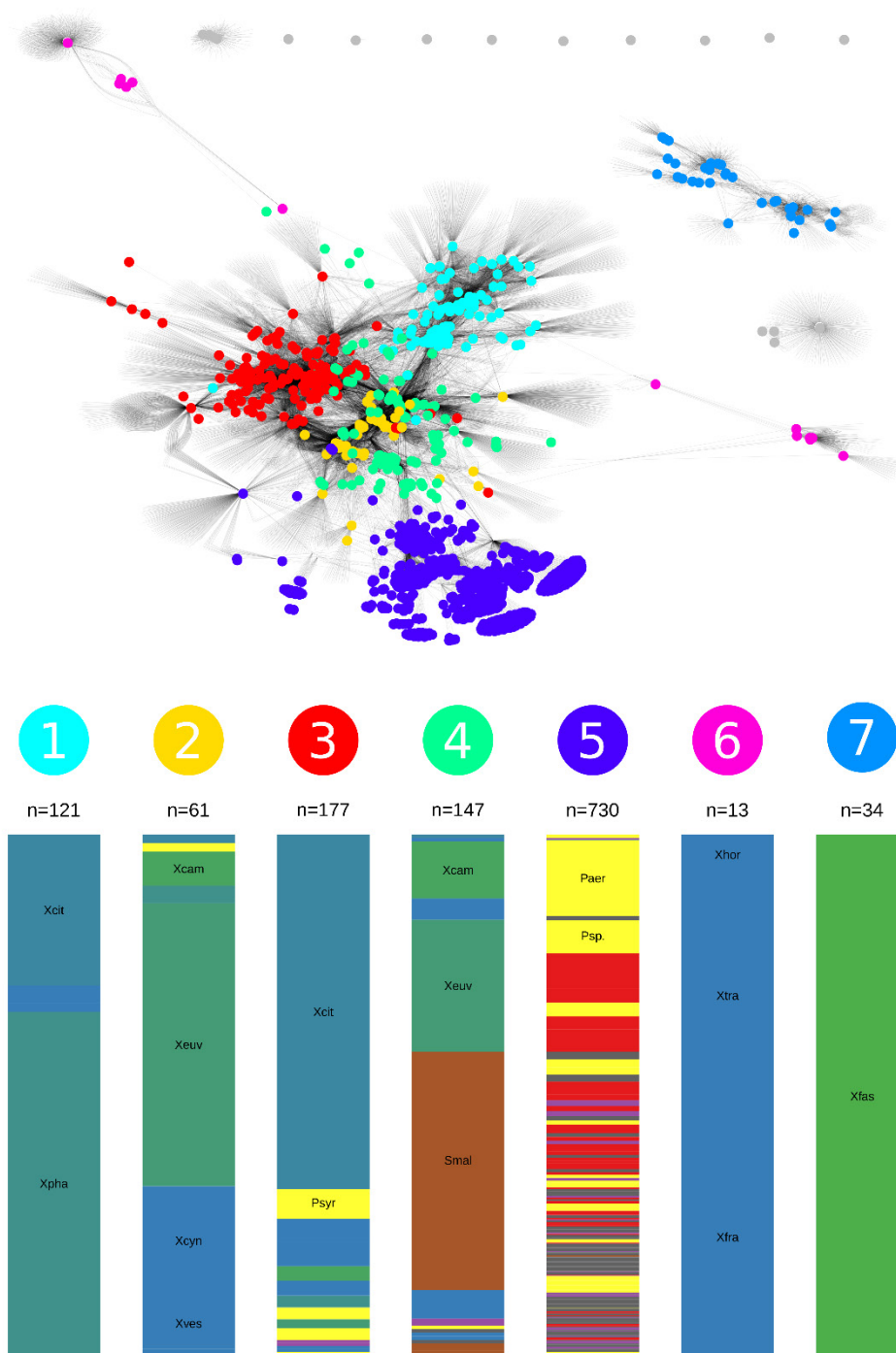


Figure 4: Representation of community membership as inferred using the Louvain algorithm over the bipartite network of plasmidic gene sharing. Bacterial nodes belonging to the seven main communities are colored according to their Louvain community membership as shown in the legend at the bottom half of the figure. The barplot graph indicates the taxonomy of the bacteria of each community. Numbers over each bar give the number of

individuals included in the community. Bars are divided into tracks whose size is proportional to the number of bacteria of each taxon level and are colored according to Figure 1. For tracks representing more than 5% of a given community, an abbreviation of the bacterial species name is given: Paer, *Pseudomonas aeruginosa*; Psp., *P. sp.*; Psyr, *P. syringae*; Smal, *Stenotrophomonas maltophilia*; Xcam, *Xanthomonas campestris*; Xcit, *X. citri*; Xcyn, *X. cynarae*; Xeuu, *X. euvesicatoria*; Xfra, *X. frageriae*; Xhor, *X. hortorum*; Xpha, *X. phaseoli*; Xtra, *X. translucens*; Xves, *X. vesicatoria*; Xfas, *Xylella fastidiosa*.

In order to detect integration events, we screened large contigs (>1,000 genes), that should represent chromosome fragments, for the presence of gene clusters shared with plasmidic query contigs (Supplementary Figure S2). Importantly, knowing the existence of such events for the copper resistance system *copLAB* offered the possibility to test the ability of our methodology to detect such events (Richard *et al.* 2017).

Despite using conservative criterion (integrations of at least ten genes, using a 99.8%id / 90%len detection threshold), the analysis revealed that at least nine distinct events occurred between query plasmids and chromosomes (Table 2). These were found in one to nine individuals and encompassed up to 117 genes. Notably, of the nine events, three involved toxin-antitoxin components, two metal resistance, one pathogenicity genes. More specifically, event 1 corresponded to that of the transposon sequence of pLH201.1 copper resistance plasmid that was found on the chromosome of several *Stenotrophomonas* strains (Richard *et al.* 2017). Interestingly, a mercury resistance locus, originally detected on the pLM159.2 plasmid in a *X. vesicatoria* tomato strain, was putatively found within the chromosome of four *Pseudomonas* species. A striking event that involved the gene encoding for TALE *pthAw2* involved in pathogenicity was detected on the chromosome of *X. citri* pv. *citri* strains from Texas assembled from sequence data produced by long-read technology (event 9). TALEs present in this bacterium have been originally solely detected on plasmids (Jalan *et al.* 2014). Genome assembly errors cannot be fully excluded and more thorough investigation with long-read sequencing or molecular biology tests may be required to confirm these integrations. Multiple toxin-antitoxin genes were found encoded on both genomic compartments. More TA systems are found in Xcc genome than in other species of the genus, with no evident explanation (Martins *et al.* 2016). These authors suggest that the chromosomal capture of TA systems could act as “anti-addiction” systems that could prevent foreign plasmids to stably establish in Xcc lineages. This hypothesis remains to be confirmed since TA systems functions extend beyond plasmid maintenance and can also be involved in dormancy or bacteriophage immunity (Harms *et al.* 2018).

One limitation of our approach was associated with the unfinished nature of many genome assemblies that were obtained after screening the NR and WGS databases for plasmid gene homologues. Indeed, many projects presented no or little contigs of sufficient size to confidently be flagged as chromosomes.

Concluding remarks

Gene sharing bipartite networks were used to characterize the sharing of plasmid genes of the Xanthomonadaceae, a bacterial family comprising (i) numerous plant pathogenic species having massive economic impact worldwide, and (ii) a genus (i.e., *Stenotrophomonas*) known to be a major reservoir for antibiotic and heavy metal resistance genes. Importantly, the studied bacterial family includes some species known to adapt through plasmid integration and/or whose pathogenicity depends on plasmid genes. This study evidenced the dense connection that plasmid genes create between the bacteria of the family. Indeed, most members of the family comprise plasmid genes and their plasmid gene content is dynamic, even connecting the plasmid and the chromosome genomic compartments. Rather than observing each strain encoding for its own private set of plasmid genes, we observed a grouping of almost every individual into a single structure. Taxonomy strongly structured the gene sharing network, while ecological niches could shape sub-structures of individuals sharing certain functions and types of genes. Among the most important genes in the cohesion of the network were genes associated to crop pathogenicity or antibiotic and metal resistance, highlighting once again the imprints of human-mediated selective pressure over pathogenic populations. Xanthomonadaceae communities therefore form a unit whose study as a whole makes sense, particularly considering the possible consequences that the spreading of adaptive genes amongst this pathogenic family can have on plant, animal and human health.

Acknowledgments

The European Union (ERDF contract GURDT I2016-1731-0006632), Conseil Régional de La Réunion, ANSES and CIRAD provided financial support. DR acknowledged a PhD grant from ANSES and CIRAD. This work was supported by Agropolis Fondation (Labex Agro – Montpellier, E-SPACE project number 1504-004). Analyses were performed using the SouthGreen HPC server.

References

- Avison MB, Higgins CS, von Heldreich CJ, Bennett PM, Walsh TR (2001) Plasmid location and molecular heterogeneity of the L1 and L2 β -lactamase genes of *Stenotrophomonas maltophilia*. *Antimicrobial Agents and Chemotherapy* **45**, 413-419.
- Baltrus DA, McCann HC, Guttman DS (2017) Evolution, genomics and epidemiology of *Pseudomonas syringae*: challenges in bacterial molecular plant pathology. *Molecular plant pathology* **18**, 152-168.
- Bennett PM (2004) Genome plasticity: insertion sequence elements, transposons and integrons, and DNA rearrangement. *Methods Mol Biol* **266**, 71-113.
- Bennett PM, Richmond MH (1976) Translocation of a discrete piece of deoxyribonucleic acid carrying an amp gene between replicons in *Escherichia coli*. *Journal of Bacteriology* **126**, 1-6.
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* **2008**, P10008.
- Cesbron S, Briand M, Essakhi S, *et al.* (2015) Comparative genomics of pathogenic and nonpathogenic strains of *Xanthomonas arboricola* unveil molecular and evolutionary events linked to pathoadaptation. *Front. Plant Sci.* **6**, 1126.
- Chen NW, Serres-Giardi L, Ruh M, *et al.* (2018) Horizontal gene transfer plays a major role in the pathological convergence of *Xanthomonas* lineages on common bean. *Bmc Genomics* **19**, 606.
- Cordero OX, Polz MF (2014) Explaining microbial genomic diversity in light of evolutionary ecology. *Nature Reviews Microbiology* **12**, 263.
- Corel E, Lopez P, Meheust R, Baptiste E (2016) Network-thinking: graphs to analyze microbial complexity and evolution. *Trends in Microbiology* **24**, 224-237.
- Crossman LC, Gould VC, Dow JM, *et al.* (2008) The complete genome, comparative and functional analysis of *Stenotrophomonas maltophilia* reveals an organism heavily shielded by drug resistance determinants. *Genome Biol* **9**, R74.
- Csardi G, Nepusz T (2006) The igraph software package for complex network research. *InterJournal, Complex Systems* **1695**, 1-9.
- Dagan T, Artzy-Randrup Y, Martin W (2008) Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proceedings of the National Academy of Sciences* **105**, 10039-10044.
- Davies JE (1997) Origins, acquisition and dissemination of antibiotic resistance determinants. *Ciba Found Symp* **207**, 15-27; discussion 27-35.
- De la Cruz F, Davies J (2000) Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends in Microbiology* **8**, 128-133.
- Dobrindt U, Hochhut B, Hentschel U, Hacker J (2004) Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* **2**, 414-424.
- Dorer MS, Fero J, Salama NR (2010) DNA damage triggers genetic exchange in *Helicobacter pylori*. *PLoS pathogens* **6**, e1001026.
- Fargier E, Manceau C (2007) Pathogenicity assays restrict the species *Xanthomonas campestris* into three pathovars and reveal nine races within *X. campestris* pv. *campestris*. *Plant Pathology* **56**, 805-818.
- Garita-Cambronero J, Palacio-Bielsa A, López MM, Cubero J (2016) Comparative genomic and phenotypic characterization of pathogenic and non-pathogenic strains of *Xanthomonas arboricola* reveals insights into the infection process of bacterial spot disease of stone fruits. *Plos One* **11**, e0161977.

- Gochez AM, Huguet-Tapia JC, Minsavage GV, *et al.* (2018) Pacbio sequencing of copper-tolerant *Xanthomonas citri* reveals presence of a chimeric plasmid structure and provides insights into reassortment and shuffling of transcription activator-like effectors among *X. citri* strains. *Bmc Genomics* **19**, 16.
- Gogarten JP, Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* **3**, 679-687.
- Gordon JL, Lefeuvre P, Escalon A, *et al.* (2015) Comparative genomics of 43 strains of *Xanthomonas citri* pv. *citri* reveals the evolutionary events giving rise to pathotypes with different host ranges. *Bmc Genomics* **16**, 1098.
- Gullberg E, Albrecht LM, Karlsson C, Sandegren L, Andersson DI (2014) Selection of a multidrug resistance plasmid by sublethal levels of antibiotics and heavy metals. *MBio* **5**, e01918-01914.
- Halary S, Leigh JW, Cheaib B, Lopez P, Baptiste E (2010) Network analyses structure genetic diversity in independent genetic worlds. *Proceedings of the National Academy of Sciences, USA* **107**, 127-132.
- Harms A, Brodersen DE, Mitarai N, Gerdes K (2018) Toxins, targets, and triggers: an overview of toxin-antitoxin biology. *Molecular cell*.
- Harrison E, Brockhurst MA (2012) Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends in Microbiology* **20**, 262-267.
- He YW, Xu M, Lin K, *et al.* (2006) Genome scale analysis of diffusible signal factor regulon in *Xanthomonas campestris* pv. *campestris*: identification of novel cell-cell communication-dependent genes and functions. *Molecular Microbiology* **59**, 610-622.
- Hobman JL, Crossman LC (2015) Bacterial antimicrobial metal ion resistance. *J Med Microbiol* **64**, 471-497.
- Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167-170.
- Hyatt D, Chen GL, Locascio PF, *et al.* (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *Bmc Bioinformatics* **11**, 119.
- Hynes AP, Shakya M, Mercer RG, *et al.* (2016) Functional and evolutionary characterization of a gene transfer agent's multilocus "genome". *Molecular Biology and Evolution* **33**, 2530-2543.
- Hyun J-W, Kim H-J, Yi P-H, Hwang R-Y, Park E-W (2012) Mode of action of streptomycin resistance in the citrus canker pathogen (*Xanthomonas smithii* subsp. *citri*) in Jeju Island. *The Plant Pathology Journal* **28**, 207-211.
- Jalan N, Yan Q, Kogenaru S, *et al.* (2014) Genomics of *Xanthomonas citri* and related species. In: *Genomics of Plant-Associated Bacteria*, pp. 151-176. Springer.
- Jibrin MO, Potnis N, Timilsina S, *et al.* (2018) Genomic inference of recombination-mediated evolution in *Xanthomonas euvesicatoria* and *X. perforans*. *Applied and Environmental Microbiology*, AEM. 00136-00118.
- Kanamori H, Yano H, Tanouchi A, *et al.* (2015) Prevalence of Smqnr and plasmid-mediated quinolone resistance determinants in clinical isolates of *Stenotrophomonas maltophilia* from Japan: novel variants of Smqnr. *New microbes and new infections* **7**, 8-14.
- Lang AS, Westbye AB, Beatty JT (2017) The distribution, evolution, and roles of gene transfer agents in prokaryotic genetic exchange. *Annual review of virology* **4**, 87-104.
- Leyns F, De Cleene M, Swings JG, De Ley J (1984) The host range of the genus *Xanthomonas*. *Bot. Rev.* **50**, 308-356.
- Liebert CA, Hall RM, Summers AO (1999) Transposon Tn21, flagship of the floating genome. *Microbiol Mol Biol Rev* **63**, 507-522.

- Martins PMM, Machado MA, Silva NV, Takita MA, de Souza AA (2016) Type II toxin-antitoxin distribution and adaptive aspects on *Xanthomonas* genomes: focus on *Xanthomonas citri*. *Frontiers in Microbiology* **7**, 652-652.
- McInerney JO (2013) More than tree dimensions: inter-lineage evolution's ecological importance. *Trends in ecology & evolution* **28**, 624-625.
- Minsavage GV, Canteros BI, Stall RE (1990) Plasmid-mediated resistance to streptomycin in *Xanthomonas campestris* pv. *vesicatoria*. *Phytopathology* **80**, 719-723.
- Nakato V, Mahuku G, Coutinho T (2018) *Xanthomonas campestris* pv. *musacearum*: a major constraint to banana, plantain and enset production in central and east Africa over the past decade. *Molecular plant pathology* **19**, 525-536.
- Naushad S, Adeolu M, Wong S, *et al.* (2015) A phylogenomic and molecular marker based taxonomic framework for the order Xanthomonadales: proposal to transfer the families Alginiphilaceae and Solimonadaceae to the order Nevskiales ord. nov. and to create a new family within the order Xanthomonadales, the family Rhodanobacteraceae fam. nov., containing the genus *Rhodanobacter* and its closest relatives. *Antonie Van Leeuwenhoek* **107**, 467-485.
- O'Donovan C, Martin MJ, Gattiker A, *et al.* (2002) High-quality protein knowledge resource: SWISS-PROT and TrEMBL. *Briefings in Bioinformatics* **3**, 275-284.
- Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299-304.
- Okinaka RT, Cloud K, Hampton O, *et al.* (1999) Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes. *Journal of Bacteriology* **181**, 6509-6515.
- Polz MF, Alm EJ, Hanage WP (2013) Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet* **29**, 170-175.
- Popa O, Dagan T (2011) Trends and barriers to lateral gene transfer in prokaryotes. *Curr Opin Microbiol* **14**, 615-623.
- Popa O, Hazkani-Covo E, Landan G, Martin W, Dagan T (2011) Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. *Genome Research* **21**, 599-609.
- Potnis N, Timilsina S, Strayer A, *et al.* (2015) Bacterial spot of tomato and pepper: diverse *Xanthomonas* species with a wide variety of virulence factors posing a worldwide challenge. *Mol Plant Pathol* **16**, 907-920.
- Pruvost O, Magne M, Boyer K, *et al.* (2014) A MLVA genotyping scheme for global surveillance of the citrus pathogen *Xanthomonas citri* pv. *citri* suggests a worldwide geographical expansion of a single genetic lineage. *Plos One* **9**, e98129.
- Punina N, Ignatov A, Pekhtereva ES, *et al.* (2007) Occurrence of *Xanthomonas campestris* pv. *raphani* on tomato plants in the Russian Federation, 287-290.
- Ravatn R, Studer S, Zehnder AJ, van der Meer JR (1998) Int-B13, an unusual site-specific recombinase of the bacteriophage P4 integrase family, is responsible for chromosomal insertion of the 105-kilobase *clc* element of *Pseudomonas* sp. strain B13. *Journal of Bacteriology* **180**, 5505-5514.
- Richard D, Ravigne V, Rieux A, *et al.* (2017) Adaptation of genetically monomorphic bacteria: evolution of copper resistance through multiple horizontal gene transfers of complex and versatile mobile genetic elements. *Mol Ecol* **26**, 2131-2149.
- Rogers EE, Stenger DC (2012) A conjugative 38 kB plasmid is present in multiple subspecies of *Xylella fastidiosa*. *Plos One* **7**, e52131.
- Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584.

- Roux B, Bolot S, Guy E, *et al.* (2015) Genomics and transcriptomics of *Xanthomonas campestris* species challenge the concept of core type III effectome. *Bmc Genomics* **16**, 975.
- Ruh M, Briand M, Bonneau S, Jacques M-A, Chen NW (2017) *Xanthomonas* adaptation to common bean is associated with horizontal transfers of genes encoding TAL effectors. *Bmc Genomics* **18**, 670.
- Seitz P, Blokesch M (2013) Cues and regulatory pathways involved in natural competence and transformation in pathogenic and environmental Gram-negative bacteria. *FEMS Microbiology Reviews* **37**, 336-363.
- Serizawa S, Inoue K (1975) Studies on citrus canker III. The influence of wind on infection. *Bull. Shizuoka Citrus Exp. Sta.* **11**, 54-67.
- Shapiro BJ, Friedman J, Cordero OX, *et al.* (2012) Population genomics of early events in the ecological differentiation of bacteria. *Science* **336**, 48-51.
- Skippington E, Ragan MA (2012) Phylogeny rather than ecology or lifestyle biases the construction of *Escherichia coli*–*Shigella* genetic exchange communities. *Open biology* **2**, 120112.
- Smillie C, Garcillan-Barcia MP, Francia MV, Rocha EPC, de la Cruz F (2010) Mobility of plasmids. *Microbiology and Molecular Biology Reviews* **74**, 434-+.
- Sørensen SJ, Bailey M, Hansen LH, Kroer N, Wuertz S (2005) Studying plasmid horizontal transfer in situ: a critical review. *Nature Reviews Microbiology* **3**, 700.
- Staehlin BM, Gibbons JG, Rokas A, O'Halloran TV, Slot JC (2016) Evolution of a heavy metal homeostasis/resistance island reflects increasing copper stress in *Enterobacteria*. *Genome Biology and Evolution* **8**, 811-826.
- Steinegger M, Söding J (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature biotechnology* **35**, 1026.
- Tenaillon O, Skurnik D, Picard B, Denamur E (2010) The population genetics of commensal *Escherichia coli*. *Nature Reviews Microbiology* **8**, 207.
- Thomas CM, Nielsen KM (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology* **3**, 711-721.
- Touchon M, de Sousa JAM, Rocha EP (2017) Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. *Current Opinion in Microbiology* **38**, 66-73.
- Van Melderden L, Saavedra De Bast M (2009) Bacterial toxin-antitoxin systems: more than selfish entities? *PLoS Genet* **5**, e1000437.
- Wiedenbeck J, Cohan FM (2011) Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiology Reviews* **35**, 957-976.
- Yang Y (1994) Host-specific symptoms and increased release of *Xanthomonas citri* and *X. campestris* pv. *malvacearum* from leaves are determined by the 102-bp tandem repeats of *pthA* and *avrb6*, respectively. *Mol. Plant-Microbe Interact.* **7**, 345-355.

Discussion générale

Discussion générale

Considérée comme une contrainte majeure aussi bien en santé humaine qu'en élevage ou en agriculture, la résistance aux antimicrobiens dont les antibiotiques mobilise les recherches publique et privée depuis plusieurs décennies. Des ressources considérables sont allouées à la compréhension des mécanismes conditionnant l'apparition et la diffusion de phénotypes résistants. C'est dans le contexte de l'apparition d'une résistance au cuivre chez la bactérie responsable du chancre asiatique des agrumes que ce travail a été initié. Le chancre Asiatique des agrumes est causé par *Xanthomonas citri* pv. *citri* et représente une forte contrainte pour l'agrumiculture mondiale. La détection sur deux territoires français d'un phénotype résistant au cuivre et sa diffusion rapide ont motivé une caractérisation approfondie des bases génétiques de cette adaptation.

En s'appuyant sur des technologies de séquençage à haut débit de seconde et de troisième génération, nous avons confirmé le support plasmidique des gènes causant le phénotype résistant au cuivre chez *Xanthomonas citri* pv. *citri*. L'annotation des séquences plasmidiques a permis de révéler la présence de plusieurs clusters de gènes de résistance au cuivre mais également de gènes de résistance à d'autres métaux lourds. Aussi, certains gènes associés au maintien et au transfert par conjugaison du plasmide ont été identifiés. Cette capacité de transfert intercellulaire du plasmide par conjugaison a été vérifiée expérimentalement dans plusieurs espèces du genre *Xanthomonas*, incluant plusieurs espèces commensales isolées d'agrumes. En plus de nous avoir motivé à caractériser l'importance de la contribution des plasmides et de leur échange dans l'adaptation chez les bactéries du genre *Xanthomonas*, ces premiers résultats aident à intégrer cette nouvelle contrainte aux mesures de lutte contre Xcc.

I La lutte contre les phyto bactérioses résistantes

Pour prévenir l'émergence de la maladie sur des territoires sains, la lutte contre le chancre asiatique des agrumes repose sur des mesures de contrôle drastiques des échanges de matériel végétal. La limitation de l'incidence¹ de la maladie dans les zones où la bactérie s'est installée de façon durable sur des superficies importantes est principalement basée sur les principes de la lutte intégrée. Cette méthode de lutte consiste en une combinaison de mesures chimiques, mécaniques, génétiques et parfois biologiques, avec comme objectif de maintenir les

¹ L'incidence d'une maladie est la proportion d'individus qu'elle touche, et n'informe pas sur la mortalité ou l'intensité des symptômes causés.

populations pathogènes à des seuils de nuisibilité économiquement acceptables. Dans cette optique, la plantation de plants certifiés indemnes de maladie ainsi que l'utilisation de porte-greffe de faible vigueur lors de la mise en place des vergers est primordiale. Ensuite, les mesures visent à minimiser : la quantité d'inoculum présent (taille des rameaux malades); la pénétration de la bactérie dans son hôte (lutte contre la mineuse des agrumes qui exacerbe les infections par création de blessures sur les feuilles) ; sa multiplication (résistance variétale, induction chimique de résistance systémique acquise) ; et sa transmission (pratiques culturales, haies brises vent, application de pesticides pendant les périodes à risque). Les pesticides utilisés sont majoritairement des produits cupriques. Si ceux-ci n'ont qu'un effet curatif réduit puisqu'ils sont inactifs sur les cellules bactériennes présentes à l'intérieur des chancres, ils limitent par contre la dispersion des bactéries au sein des vergers. La propagation de Xcc de proche en proche se fait par le vent, qui dissémine l'eau libre contaminée présente à la surface des organes végétaux. C'est durant cette phase de transmission que le cuivre a une action destructrice sur Xcc. L'acquisition d'une résistance au cuivre par Xcc représente ainsi une limitation forte du potentiel de lutte intégrée.

La connaissance des séquences des gènes *copLAB* et *copABCD* associés au phénotype de résistance au cuivre chez Xcc ouvre la possibilité de mise au point de méthodes de détection par amplification isothermale (en cours de développement dans l'équipe). Le caractère résistant peut ainsi être testé sur le terrain à partir de symptômes de la maladie. Au champ, de tels tests pourraient permettre aux producteurs de statuer sur la présence de gènes de résistance au cuivre au sein de leurs vergers et d'adapter la lutte en conséquence. Le support plasmidique de cette résistance présente un niveau de contrainte additionnel. En effet, la capacité de diffusion intercellulaire qu'il confère aux gènes de résistance au cuivre les rend ubiquistes. Nous les avons ainsi observés dans plusieurs espèces de *Xanthomonas* mais aussi dans des souches commensales. Au même titre que la recherche agronomique intègre la gestion des organismes phytopathogènes, elle doit aussi inclure la gestion des gènes de résistance. Cette dernière passe par une caractérisation fine de la dynamique spatiale et temporelle de diffusion et de persistance des gènes de résistance en général. Pour cela, l'intérêt des approches de génomique ou de métagénomique à haut débit est qu'elles permettent d'appréhender les résistances à l'échelle de la population et non plus de l'individu, et sans contraindre l'analyse par un a priori sur le gène recherché (Heuer & Smalla 2012).

De manière générale, le suivi de la résistance aux pesticides des populations pathogènes dans les agrosystèmes est nécessaire à la mise en place de méthodes de lutte pertinentes mais

nécessite le développement de méthodes de détection qui peuvent être biologiques, biochimiques ou moléculaires (R4PNetwork 2016). Pour les acteurs de la recherche, le suivi des résistances et des plasmides dans les lignées bactériennes va de manière générale profiter largement de l'utilisation de technologies telles que la cytométrie en flux (Loftie-Eaton *et al.* 2014) et le séquençage de troisième génération.

Les effets négatifs de l'utilisation massive d'antimicrobiens pour lutter contre les maladies des plantes sont bien connus en termes d'impact environnemental ainsi que de menace pour la santé publique (Cheng *et al.* 2016; Hu *et al.* 2016a; Jechalke *et al.* 2014; Li *et al.* 2015 ; Tasho & Cho 2016). Le développement de nouvelles molécules antibactériennes ne représente donc probablement pas la solution à une gestion durable du chancre Asiatique des agrumes. La diminution de l'utilisation de pesticides pour la lutte contre cette phyto bactériose nécessite une acceptation plus marquée de la lutte intégrée. Même si cette stratégie est préconisée depuis de nombreuses années, l'utilisation de la lutte chimique est encore très importante du fait (i) de l'impact sur les coûts de production et les rendements d'autres composantes de la lutte intégrée comme l'élimination physique de l'inoculum ou les modifications de pratiques culturales (irrigation localisée ou installation de haies brise-vent) et (ii) de l'absence de résistance variétale efficace dans plusieurs espèces d'agrumes à fort intérêt agronomique.

II Le caractère monomorphe : un taux de mutation faible?

Les bactéries monomorphes présentent des degrés de polymorphismes faibles. Cette caractéristique, historiquement détectée par l'uniformité des séquences de quelques gènes de ménage utilisés comme marqueurs durant l'ère pré-NGS (Maiden 2006) est aujourd'hui synonyme de faibles taux de substitution SNP dans les génomes (Achtman 2012). C'est ce taux de substitution que nous avons choisis de déterminer chez Xcc, constituant une des premières études du genre appliquées à une bactérie phytopathogène (Chapitre II). La bonne connaissance de notre modèle d'étude associée à un échantillonnage important et à un choix raisonné des souches à l'échelle d'étude ont rendu possible une caractérisation précise de la dynamique d'évolution du génome de Xcc.

Nous avons bénéficié d'une résolution bien supérieure à celle apportée par les marqueurs MLVA, historiquement utilisés en épidémiologie chez Xcc, sans que nos résultats soient contradictoires avec ce qui était connu. Associées au caractère hétérochrone de notre échantillonnage, les caractéristiques des marqueurs SNP ont en outre permis la datation de plusieurs événements clés de la dispersion de Xcc dans la zone SWIO.

A l'échelle de l'Océan Indien, Xcc présente un taux de substitution de 8.4×10^{-8} par site par an comparable à ceux établis pour d'autres bactéries monomorphes (Achtman 2012; Duchene *et al.* 2016). Une des origines possibles du caractère monomorphe d'un taxon est l'existence d'un effet de fondation relativement récent. Ce dernier peut être lié à un évènement évolutif tel que le gain de gènes par HGT qui permettent un changement de niche écologique d'un nombre d'individus très restreint et qui engendre donc un goulot d'étranglement (Achtman 2008). Ce scénario semble avoir eu lieu chez l'ancêtre de *Yersinia pestis*, responsable de la peste chez l'homme (Achtman *et al.* 1999), et est peut-être également à l'origine de certains pathovars chez *Xanthomonas* (Ferreira *et al.* 2015).

Xcc pathotype A illustre le succès épidémiologique d'une bactérie phytopathogène monomorphe causé par sa dissémination d'origine humaine dans la grande majorité des bassins de production d'agrumes. L'environnement de Xcc est très stable puisqu'il ne comprend quasiment que son hôte cultivé. Au vu de l'homogénéité génétique de celui-ci, nous pouvons imaginer que les génotypes de Xcc actuels ont une très bonne valeur sélective, améliorée au fil du temps. Dans le cas de bactéries clonales, l'acquisition d'une mutation avantageuse sur le chromosome va augmenter en fréquence dans la population et *in fine* parvenir à sa fixation, et du fait de l'absence d'indépendance entre loci, la diversité sera purgée sur des grandes portions (voire la totalité) du génome (balayage sélectif) entraînant l'homogénéisation de ceux-ci. Dans ce contexte, les transferts horizontaux de gènes peuvent avoir une grande importance puisqu'ils peuvent fournir du matériel pour faire face aux variations de contraintes environnementales.

Les résultats de cette thèse constituent une contribution supplémentaire illustrant le fait que les éléments génétiques mobiles jouent un rôle important dans l'évolution du genre *Xanthomonas*. Nos résultats et ceux de plusieurs auteurs montrent que le contenu en gènes de *Xanthomonas* est variable à des échelles phylogénétiques larges (Chen *et al.* 2018; Gordon *et al.* 2015 ; Merda *et al.* 2017). L'étude de la microévolution de Xcc suggère toutefois que l'importance de cette forme d'évolution génomique ne peut pas être appréciée pleinement à de larges échelles. En effet, les gains et les pertes de gènes apparaissent à des échelles fines et à des taux importants, occasionnant donc une sous-estimation de la fréquence de ces évènements lorsqu'ils sont observés à l'échelle d'une espèce ou d'un pathovar. Une majorité de ces évènements ne sont en effet pas conservés dans les lignées bactériennes. Nous avons ainsi mis en évidence (Chapitre II) que des variations de contenu en gènes apparaissent chez Xcc sur des pas de temps très courts, occasionnant parfois une quantité de gènes variables du même ordre de grandeur que

celle des SNP variables. Le nombre conséquent de nucléotides affectés par le gain d'un seul gène contraste avec l'idée d'un génome « monomorphe ».

Ces résultats préliminaires gagneraient à être confortés par des études basées sur des technologies de séquençage produisant des données génomiques de très grande qualité, permettant ainsi de caractériser précisément la temporalité d'apparition des variations de contenu en gènes chez *Xcc*. Les échelles géographiques et phylogénétiques étudiées dans le Chapitre II de cette thèse ne paraissent pas assez fines pour observer l'apparition de ces variations : nous n'avons par exemple observé que très peu de souches présentant un contenu en gènes strictement identique. Chez *Xcc*, la très forte diversité présente au sein d'un seul verger n'est pas surprenante au vu de sa biologie et de ses modes de dissémination. L'étude d'individus apparentés depuis un nombre de générations faible impliquerait donc probablement le séquençage d'individus à des échelles géographiques très restreintes: l'arbre, le rameau, la feuille voir le symptôme chancreux.

III Les plasmides, véhicules pour l'échange de gènes

La notion de promiscuité, qui désigne littéralement une pratique sexuelle fréquente avec différents partenaires quels qu'ils soient, illustre bien l'intensité du phénomène d'échanges de gènes chez les bactéries (Wiedenbeck & Cohan 2011). L'échange de matériel génétique entre souches bactériennes ne se fait pas pour autant de manière aléatoire. Plusieurs paramètres structurent en effet ces échanges. Le contact physique entre souche donneuse et souche receveuse, conditionné par le partage du même habitat, est par exemple souvent mentionné (McInerney 2013; Polz *et al.* 2013). Ce n'est en revanche pas toujours le facteur principal de structuration des transferts horizontaux de gènes chez les bactéries, puisque l'apparement entre souches semble fortement structurer le partage de gènes chez de nombreux groupes bactériens (Popa & Dagan 2011; Popa *et al.* 2011; Skippington & Ragan 2012). Nous avons ainsi montré que chez les Xanthomonadaceae, l'échange de gènes semble être taxonomiquement restreint à la famille (Chapitres I et III). Des contraintes biologiques et génétiques peuvent expliquer ce résultat. En effet, les systèmes CRISPR/Cas et les systèmes de restriction-modification, omniprésents chez les bactéries, les protègent d'une trop forte prévalence d'ADN étranger dans leur génome et constituent une barrière à l'échange de plasmides dont l'intensité augmente avec l'éloignement phylogénétique (Ando *et al.* 2000; Jinek *et al.* 2012; Marraffini & Sontheimer 2008; Waldron & Lindsay 2006). Egalement, les gènes acquis doivent présenter un usage de codon compatible avec la machinerie cellulaire de

l'hôte. En ce sens, certains éléments génétiques mobiles complexes et plus ou moins autonomes surmontent ces limites avec différents systèmes qu'ils encodent eux-mêmes. Certains plasmides comprennent par exemple des ARNt nécessaires à la traduction des gènes qu'ils portent. Certains autres possèdent des mécanismes de protection contre les enzymes de restriction (Tock & Dryden 2005), voire ont détourné ces systèmes pour favoriser leur propre stabilité (Kulakauskas *et al.* 1995). Le caractère très aléatoire de certains mécanismes de transfert horizontal de gènes tels que la transformation contraste avec les transferts horizontaux plasmidiques.

La conjugaison plasmidique est souvent reconnue comme le processus majeur de transmission de matériel génétique entre bactéries (Norman *et al.* 2009). Les plasmides permettent l'échange simultané d'un grand nombre de gènes à caractère adaptatif qui sont souvent organisés en opérons comprenant un promoteur et un ou plusieurs clusters de gènes. Ces voies métaboliques sont donc immédiatement réutilisables par la bactérie receveuse. En plus des gènes adaptatifs, les plasmides sont des éléments génétiques possédant souvent un squelette qui les rend autonomes. Ce dernier comprend généralement des gènes liés au maintien du plasmide dans l'hôte et dans sa lignée génétique ainsi qu'à son transfert vers d'autres hôtes (voir II.7). Ces caractéristiques valent aux plasmides d'être souvent qualifiés d'invasif voir de parasites, et d'imposer un coût en énergie aux cellules porteuses. La persistance des plasmides au sein de certaines lignées génétiques qui bénéficieraient à première vue d'une intégration chromosomique des gènes adaptatifs et de la perte du squelette plasmidique superflu est difficile à expliquer. En effet, les mécanismes génétiques d'addiction plasmidique (voir II.7.g) ne suffisent pas à expliquer cette persistance, ajouté au fait que certains plasmides très stables n'en possèdent pas (Modi & Adams 1991). De plus, les changements de phénotype liés aux plasmides peuvent être plus nombreux que le seul phénotype adaptatif, influençant potentiellement leur dynamique de persistance (Dougherty *et al.* 2014). Certaines théories expliquent les bénéfices de gènes adaptatifs plasmidiques plutôt que chromosomiques entre autres par les fluctuations temporelles et spatiales des pressions de sélections associées à ces plasmides (Heuer *et al.* 2008; Slater *et al.* 2008).

Les plasmides sont des éléments génétiques très étudiés, principalement du fait de leur forte implication dans l'apparition et la dissémination de phénotypes antibiorésistants en santé humaine. Si la bactériologie appliquée à la phytopathologie bénéficie de l'avancée des connaissances générales sur ce compartiment génomique, une grande quantité de caractéristiques sont spécifiques au type de réplicon plasmidique. Or, ce dernier est souvent

spécifique à un ou plusieurs groupes taxonomiques. La classification des plasmides a surtout été étudiée chez les Enterobacteriaceae médicales (Carattoli *et al.* 2014) et ne profite donc par exemple pas à celle de bactéries pathogènes de plantes appartenant à d'autres familles bactériennes. La connaissance réduite des groupes d'incompatibilité, des systèmes de restriction/modification des systèmes de conjugaison et plus largement des réplicons plasmidiques de la famille des Xanthomonadaceae freine la compréhension de leur dynamique évolutive. Plusieurs *Xanthomonas* dont Xcc apparaissent comme un bon modèle d'étude pour débiter une caractérisation approfondie des plasmides de bactéries phytopathogènes. En effet, la communauté scientifique dispose pour ces bactéries à la fois d'une grande diversité d'accessions microbiennes et d'une grande quantité de génomes d'ores et déjà séquencés. Pour finir, en plus de son rôle dans l'adaptation contemporaine de Xcc, le compartiment plasmidique de ce pathovar encode également certains des gènes responsables de son caractère pathogène, indispensables à son succès épidémiologique. Une caractérisation exhaustive du plasmidome de Xcc pourrait apporter des connaissances telles que la caractérisation de nouvelles fonctions adaptatives, la précision du rôle des plasmides en tant que véhicule de déterminants du caractère pathogène ou l'amélioration de la compréhension de leur transfert et de leur stabilité. Ces ressources pourraient également être utiles pour préciser le rôle des transposon Tn3, dont l'importance a été soulignée dans cette thèse et dans d'autres travaux (Ferreira *et al.* 2015).

L'utilisation conjointe de technologies de séquençage de deuxième et troisième génération permet dès aujourd'hui d'obtenir à haut débit des séquences plasmidiques à la fois circulaires et avec des taux d'erreur très faible. Cependant, l'exploitation de ces données ne peut être complète sans caractérisation fonctionnelle de certains éléments génétiques, qui repose sur des techniques comme la mutagenèse dirigée que l'on ne peut pas encore qualifier de haut débit.

IV Deux types de mobilité : focus sur les transposons Tn3 chez *Xanthomonas*

La découverte des premiers éléments génétiques mobiles a donné lieu à la création de familles d'éléments bien distinctes. Il est cependant vite apparu, au vu de leur diversité et complexité, que ces éléments formaient un continuum (Frost *et al.* 2005; Osborn & Böltner 2002). Des unités fonctionnelles conférant par exemple la capacité de copier, de modifier, d'exciser, d'excréter, d'incorporer ou encore de lier un ou plusieurs fragments d'ADN se combinent pour former une variété d'éléments génétiques mobiles (Toussaint & Merlin 2002). Replacés dans le contexte de la mobilité qu'ils confèrent à certains gènes adaptatifs, les éléments génétiques mobiles peuvent toutefois être séparés en deux groupes distincts selon que leur mobilité soit

intra ou intercellulaire. L'association entre éléments possédant des types de mobilités différentes est couramment observée et décrite dans la littérature, particulièrement entre transposon (à mobilité intracellulaire) et plasmide (à mobilité intercellulaire) (Bennett 2004; Davies 1997; Petrovski & Stanisich 2011). Les plasmides sont en effet des réplicons très recombinants et très plastiques, aussi bien au niveau du squelette plasmidique (Gochez *et al.* 2018; Smillie *et al.* 2010) que du contenu en gènes adaptatifs (Böltner *et al.* 2002; Pothier *et al.* 2011). Cette variabilité est en partie due à la présence, au sein des plasmides, d'éléments génétiques eux-mêmes mobiles et/ou très recombinants (Ferreira *et al.* 2015; Norman *et al.* 2009). Les gènes à caractère adaptatifs y sont d'ailleurs souvent inclus.

Les transposons de la famille Tn3 ont été identifiés dans le contexte de la résistance aux antibiotiques de la famille des beta-lactamines. Ils ont depuis été qualifiés de vecteurs majeurs de gènes "passagers" encodant pour des traits adaptatifs en réponse à des pressions de sélection anthropiques, notamment en lien avec la résistance à des molécules antimicrobiennes (antibiotiques, métaux lourds) ou des fonctions cataboliques permettant la dégradation de polluants dans différents environnements (Nicolas *et al.* 2015).

Chez *Xanthomonas*, des gènes adaptatifs conférant des phénotypes tels que la résistance aux métaux lourds (Kholodii *et al.* 2000; Niu *et al.* 2015; Richard *et al.* 2017c) ou aux antibiotiques (Sundin & Bender 1995) ont été observés sur des transposons de la famille des Tn3. Dans le cas de la résistance au cuivre par exemple, les gènes causatifs du phénotype sont toujours encodés sur un transposon de la famille Tn3, localisé sur un plasmide ou sur le chromosome (Chapitre I). La résistance à la streptomycine, encodée sur un Tn3 plasmidique chez plusieurs espèces de *Xanthomonas* (Hyun *et al.* 2012; Sundin & Bender 1995), est conférée par deux gènes retrouvés chez différentes espèces sur plusieurs types de plasmides (Chiou & Jones 1993) ainsi que le chromosome (McManus & Jones 1994). Chez *X. citri*, *X. oryzae*, *X. euvesicatoria* et *X. arboricola*, des gènes codant pour des effecteurs de type III dont certains de type TALE, impliqués voire indispensables au caractère pathogène de ces espèces, sont encodés sur des transposons de cette famille (Ferreira *et al.* 2015). Dans ce cas, le mode de répllication de ce type de transposon a pu jouer un rôle important dans la diversification et l'évolution du nombre de répétitions des gènes TALE (dont le fonctionnement est détaillé par Boch *et al.* (2009)), modifiant ainsi la relation hôte/pathogène et sa spécialisation chez les *Xanthomonas*. Ces résultats illustrent l'importance des transposons Tn3 pour l'évolution et l'adaptation de *Xanthomonas*. Ces éléments, encodés aussi bien sur des plasmides que sur des chromosomes, pourraient se déplacer de l'un à l'autre par transposition, conférant par la même cette mobilité

aux gènes adaptatifs (Ferreira *et al.* 2015). Le transposon Tn3 porteur de résistance au mercure chez un *Xanthomonas* a pu être transféré expérimentalement sur un plasmide à large spectre introduit dans cette souche (Kholodii *et al.* 2000).

Chez *Xanthomonas*, les transposons Tn3, de par leur ubiquité, confèrent une connectivité aux différents réplicons intracellulaires par transposition mais aussi par recombinaison (liée au mode de transposition des Tn3, détaillé en II.6.a). En conséquence, on peut imaginer que tous les gènes de ce genre bactérien puissent potentiellement acquérir une mobilité intercellulaire, mais également que tous les gènes du compartiment accessoire puissent s'intégrer durablement dans les génomes bactériens en étant transférés sur le chromosome par le biais d'un transposon Tn3. Cependant, contrairement à certains éléments mobiles comme les IS, les Tn3 n'envahissent pas les génomes. Un phénomène d'immunité, qui empêche l'insertion trop proche de plusieurs Tn3 similaires sur un même réplicon, explique cette caractéristique (Nicolas *et al.* 2015).

La plasticité des éléments génétiques mobiles bactériens et leur interconnexion semblent donc constituer des éléments clés de l'échange de gènes entre cellules bactériennes (Norman *et al.* 2009). La présence d'un gène sur un réplicon non mobile ne serait ainsi pas un frein à son échange intercellulaire par conjugaison, tant qu'un élément à mobilité intercellulaire est présent et que le taux de recombinaison inter-réplicon n'est pas nul. De tels transferts sont alors complexes puisqu'ils impliquent un changement de réplicon puis un changement d'hôte.

V Le rôle de réservoir de gènes des bactéries commensales

L'existence de plasmides identiques à quelques SNP près et encodant pour les mêmes clusters de gènes de résistance chez des espèces de *Xanthomonas* respectivement pathogènes de solanées maraîchères et d'agrumes, comme c'est le cas pour les versions de pLH201.1 présentes à la Réunion chez *Xcc* et *X. cynarae* pv. *gardneri* est surprenante (Chapitre I). Les échanges directs de gènes de résistance entre les bactéries phytopathogènes du genre *Xanthomonas* sont pourtant probablement rares. En effet, leur forte spécialisation parasitaire parfois associée à une capacité de survie réduite en dehors de leur hôte limite les possibilités de flux de gènes directs entre populations pathogènes de différentes espèces végétales. Si certaines pratiques culturales (*e.g.* agroécologie) ou conditions environnementales peuvent favoriser un contact physique temporaire entre *Xanthomonas* pathogènes de familles botaniques différentes, le contact entre ces bactéries et des déterminants de résistance peut en revanche être favorisé par la présence d'un intermédiaire capable de maintenir et de diffuser la résistance. Les bactéries commensales,

c'est à dire non pathogènes, semblent de très bons candidats pour ce rôle. Des phénotypes antibiorésistants ont ainsi été observés chez les bactéries commensales de différents habitats (Bailey *et al.* 2010; Marshall *et al.* 2009; Mongkolrattanothai *et al.* 2004; Séveno *et al.* 2002).

Si l'échange de matériel génétique entre les compartiments commensaux et pathogènes a été mis en évidence chez l'homme, ce lien reste très peu étudié dans le cas de bactéries associées aux plantes. Dans le cas de la résistance aux métaux lourds chez *Xanthomonas*, nos travaux ont mis en évidence (Chapitres I et III) la présence de transposons adaptatifs génétiquement proches chez Xcc et chez des souches commensales de *Stenotrophomonas*. Les bactéries commensales possèdent probablement toujours des gènes adaptés aux contraintes environnementales locales, constituant ainsi un réservoir de gènes potentiellement mobilisable par des bactéries pathogènes des agroécosystèmes. Ce phénomène sous-entend l'existence d'un génome « environnemental » (Norman *et al.* 2009; Polz *et al.* 2013).

Les résultats des Chapitres I et III montrent cependant que les transferts horizontaux de gènes chez *Xanthomonas* sont très majoritairement limités à la famille des Xanthomonadaceae. Nous suggérons donc que le réservoir de gènes adaptatifs disponible pour les *Xanthomonas* phytopathogènes est majoritairement présent dans cette famille, dont le compartiment commensal est peu connu. Le rôle de réservoir de gènes adaptatifs du genre *Stenotrophomonas* a été décrit en santé humaine (Crossman *et al.* 2008) et est suggéré par nos résultats. Les souches commensales du genre *Xanthomonas* sont globalement peu étudiées. Chez l'espèce *X. arboricola*, dans laquelle leur caractérisation est la plus aboutie, les souches non pathogènes couvrent une grande partie de la diversité de cette espèce et se différencient de quelques lignées responsables d'épidémies (*e.g.* les pathovars *corylina*, *juglandis* et *pruni*). Les souches non-pathogènes ont notamment des répertoires d'effecteurs de type III beaucoup plus restreints. Ces différences de contenu en effecteurs entre souches pathogènes et commensales sont souvent en lien avec des événements plus ou moins ancestraux de HGT (Merda *et al.* 2017). Les souches commensales de *X. arboricola* sont ubiquistes dans la phyllosphère : elles ont été observées en association avec une grande diversité de plantes cultivées, y compris les agrumes. Par ailleurs, plusieurs études ont mis en évidence des *Xanthomonas* non pathogènes génétiquement éloignés de la diversité connue, suggérant l'existence de plusieurs espèces commensales non encore caractérisées (Cottyn *et al.* 2009; Cottyn *et al.* 2001; Parkinson *et al.* 2009; Vauterin *et al.* 1996a; Vauterin *et al.* 1996b). La caractérisation génomique récente de quelques souches commensales a abouti à la création de nouvelles espèces, comme *X. maliensis* (Triplett *et al.* 2015) et *X. floridensis* (Vicente *et al.* 2017). Des études génomiques à l'échelle de la famille

permettraient de statuer sur l'importance de *Xanthomonas* spp., *Stenotrophomonas* spp. et d'autres genres commensaux en tant que constituants de la microflore associée à la phyllosphère et réservoirs d'éléments génétiques mobiles. L'essor des travaux de métabarcoding et de métagénomique ouvre des perspectives intéressantes pour aborder ces questions.

L'interconnexion entre les compartiments bactériens pathogène et commensal en agriculture semble ainsi être analogue à celle souvent décrite entre les bactéries environnementales et les bactéries pathogènes de l'humain (Walsh 2006). Les preuves de cette analogie sont multiples dans notre étude, et on peut donc raisonnablement penser que d'autres résultats obtenus en santé humaine sont transposables à la santé des végétaux. L'intérêt croissant pour l'antibiorésistance bactérienne en santé humaine a complété notre vision de la transmission des gènes de résistance. Une grande quantité de connections et de protagonistes ont ainsi été ajoutés à l'interaction bactérie pathogène/bactérie environnementale, formant une structure d'échange complexe (Davies and Davies, 2010).

Littérature citée

- Achtman M (2008) Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu. Rev. Microbiol.* **62**, 53-70.
- Achtman M (2012) Insights from genomic comparisons of genetically monomorphic bacterial pathogens. *Philos Trans R Soc Lond B Biol Sci* **367**, 860-867.
- Achtman M, Zurth K, Morelli G, *et al.* (1999) *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci U S A* **96**, 14043-14048.
- Al-Daoude A, Arabi M, Ammounieh H (2009) Studying *Erwinia amylovora* isolates from Syria for copper resistance and streptomycin sensitivity. *Journal of Plant Pathology*, 203-205.
- Alam K, Rolfe J (2006) Economics of plant disease outbreaks. *Agenda: A Journal of Policy Analysis and Reform*, 133-146.
- Alvarez-Martinez CE, Christie PJ (2009) Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev* **73**, 775-808.
- Alvarez-Ponce D, Lopez P, Baptiste E, McInerney JO (2013) Gene similarity networks provide tools for understanding eukaryote origins and evolution. *Proceedings of the National Academy of Sciences* **110**, E1594-E1603.
- Anderson PK, Cunningham AA, Patel NG, *et al.* (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in ecology & evolution* **19**, 535-544.
- Ando T, Xu Q, Torres M, *et al.* (2000) Restriction-modification system differences in *Helicobacter pylori* are a barrier to interstrain plasmid transfer. *Molecular Microbiology* **37**, 1052-1065.
- Anjem A, Imlay JA (2012) Mononuclear iron enzymes are primary targets of hydrogen peroxide stress. *Journal of Biological Chemistry*, jbc. M111. 330365.
- Arnaud D, Hwang I (2015) A sophisticated network of signaling pathways regulates stomatal defenses to bacterial pathogens. *Molecular Plant* **8**, 566-581.
- Bailey JK, Pinyon JL, Anantham S, Hall RM (2010) Commensal *Escherichia coli* of healthy humans: a reservoir for antibiotic-resistance determinants. *Journal of Medical Microbiology* **59**, 1331-1339.
- Balestra G, Sechler A, Schuenzel E, Schaad N (2008) First report of citrus canker caused by *Xanthomonas citri* in Somalia. *Plant Disease* **92**, 981-981.
- Balogh B, Canteros BI, Stall RE, Jones JB (2008) Control of citrus canker and citrus bacterial spot with bacteriophages. *Plant Disease* **92**, 1048-1052.
- Bar-Joseph M, Lee R (1989) Citrus tristeza virus. *AAB Descriptions of plant viruses* **353**.
- Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* **286**, 509-512.
- Barabasi A-L, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nature Reviews Genetics* **5**, 101.
- Barash I, Manulis-Sasson S (2009) Recent evolution of bacterial pathogens: the gall-forming *Pantoea agglomerans* case. *Annual Review of Phytopathology* **47**, 133-152.
- Barrett HC, Rhodes AM (1976) A numerical taxonomic study of affinity relationships in cultivated *Citrus* and its close relatives. *Systematic Botany* **1**, 105-136.
- Basim H, Minsavage GV, Stall RE, *et al.* (2005) Characterization of a unique chromosomal copper resistance gene cluster from *Xanthomonas campestris* pv. *vesicatoria*. *Appl. Environ. Microbiol.* **71**, 8284-8291.
- Baym M, Lieberman TD, Kelsic ED, *et al.* (2016) Spatiotemporal microbial evolution on antibiotic landscapes. *Science* **353**, 1147-1151.

- Behlau F, Canteros BI, Jones JB, Graham JH (2012) Copper resistance genes from different xanthomonads and citrus epiphytic bacteria confer resistance to *Xanthomonas citri* subsp. *citri*. *European Journal of Plant Pathology* **133**, 949-963.
- Behlau F, Canteros BI, Minsavage GV, Jones JB, Graham JH (2011) Molecular characterization of copper resistance genes from *Xanthomonas citri* subsp. *citri* and *Xanthomonas alfalfae* subsp. *citrumelonis*. *Applied and Environmental Microbiology* **77**, 4089-4096.
- Behlau F, Gochez AM, Lugo AJ, *et al.* (2017) Characterization of a unique copper resistance gene cluster in *Xanthomonas campestris* pv. *campestris* isolated in Trinidad, West Indies. *European Journal of Plant Pathology* **147**, 671-681.
- Behlau F, Hong JC, Jones JB, Graham JH (2013) Evidence for acquisition of copper resistance genes from different sources in citrus-associated xanthomonads. *Phytopathology* **103**, 409-418.
- Beiko RG, Harlow TJ, Ragan MA (2005) Highways of gene sharing in prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 14332-14337.
- Bender CL, Cooksey DA (1986) Indigenous plasmids in *Pseudomonas syringae* pv. *tomato*: conjugative transfer and role in copper resistance. *Journal of Bacteriology* **165**, 534-541.
- Bender CL, Malvick DK, Conway KE, George S, Pratt P (1990) Characterization of pXV10A, a copper resistance plasmid in *Xanthomonas campestris* pv. *vesicatoria*. *Applied and Environmental Microbiology* **56**, 170-175.
- Bennett PM (2004) Genome plasticity: insertion sequence elements, transposons and integrons, and DNA rearrangement. *Methods Mol Biol* **266**, 71-113.
- Bennett PM, Richmond MH (1976) Translocation of a discrete piece of deoxyribonucleic acid carrying an amp gene between replicons in *Escherichia coli*. *Journal of Bacteriology* **126**, 1-6.
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* **2008**, P10008.
- Boch J, Scholze H, Schornack S, *et al.* (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509-1512.
- Boerlin P, Reid-Smith RJ (2008) Antimicrobial resistance: its emergence and transmission. *Animal Health Research Reviews* **9**, 115-126.
- Bojer MS, Struve C, Ingmer H, Hansen DS, Krogfelt KA (2010) Heat resistance mediated by a new plasmid encoded Clp ATPase, ClpK, as a possible novel mechanism for nosocomial persistence of *Klebsiella pneumoniae*. *Plos One* **5**, e15467.
- Bolduc B, Jang HB, Doulcier G, *et al.* (2017) vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ* **5**, e3243.
- Bolotin E, Hershberg R (2015) Gene loss dominates as a source of genetic variation within clonal pathogenic bacterial species. *Genome Biology and Evolution* **7**, 2173-2187.
- Böltner D, MacMahon C, Pembroke JT, Strike P, Osborn AM (2002) R391: a conjugative integrating mosaic comprised of phage, plasmid, and transposon elements. *Journal of Bacteriology* **184**, 5158-5169.
- Bové JM (1986) Witches' broom disease of lime. *FAO Plant Protection Bulletin* **34**, 217-218.
- Bragg P, Rainnie D (1974) The effect of silver ions on the respiratory chain of *Escherichia coli*. *Canadian journal of microbiology* **20**, 883-889.
- Broadbent P (1995) Quarantine in relation to Australian citrus imports and exports. *Australasian Plant Pathology* **24**, 145-156.
- Brown NL, Barrett SR, Camakaris J, Lee BT, Rouch DA (1995) Molecular genetics and transport analysis of the copper-resistance determinant (pco) from *Escherichia coli* plasmid pRJ1004. *Mol Microbiol* **17**, 1153-1166.

- Brunings AM, Gabriel DW (2003) *Xanthomonas citri*: breaking the surface. *Mol Plant Pathol* **4**, 141-157.
- Bui Thi Ngoc L, Vernière C, Jouen E, *et al.* (2010) Amplified fragment length polymorphism and multilocus sequence analysis-based genotypic relatedness among pathogenic variants of *Xanthomonas citri* pv. *citri* and *Xanthomonas campestris* pv. *bilvae*. *International Journal of Systematic and Evolutionary Microbiology* **60**, 515-525.
- Bui Thi Ngoc L, Verniere C, Vital K, *et al.* (2009) Development of 14 minisatellite markers for the citrus canker bacterium, *Xanthomonas citri* pv. *citri*. *Molecular Ecology Resources* **9**, 125-127.
- Burrus V (2017) Mechanisms of stabilization of integrative and conjugative elements. *Current Opinion in Microbiology* **38**, 44-50.
- Burrus V, Pavlovic G, Decaris B, Guédon G (2002) Conjugative transposons: the tip of the iceberg. *Molecular Microbiology* **46**, 601-610.
- Büttner D, Bonas U (2010) Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiology Reviews* **34**, 107-133.
- Cai R, Lewis J, Yan S, *et al.* (2011) The plant pathogen *Pseudomonas syringae* pv. *tomato* is genetically monomorphic and under strong selection to evade tomato immunity. *PLoS Pathog* **7**, e1002130.
- Calavan E (1956) Citrus canker. A bacterial disease caused by *Xanthomonas citri*. *Bull Dept Agr* **45**, 259-262.
- Cambray G, Guerout AM, Mazel D (2010) Integrons. *Annu Rev Genet* **44**, 141-166.
- Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann M-L, Brüssow H (2003) Phage as agents of lateral gene transfer. *Current Opinion in Microbiology* **6**, 417-424.
- Canteros BI, Rybak M, Gochez A, *et al.* (2008) Occurrence of copper resistance in *Xanthomonas axonopodis* pv. *citri* in Argentina. *Phytopathology* **98**, S30-S31.
- Canteros BI, Gochez AM, Moschini RC (2017) Management of Citrus canker in Argentina, a success story. *The Plant Pathology Journal* **33**, 441-449.
- Carattoli A, Bertini A, Villa L, *et al.* (2005) Identification of plasmids by PCR-based replicon typing. *Journal of Microbiological Methods* **63**, 219-228.
- Carattoli A, Zankari E, García-Fernández A, *et al.* (2014) In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrobial Agents and Chemotherapy* **58**, 3895-3903.
- Casacuberta E, González J (2013) The impact of transposable elements in environmental adaptation. *Molecular Ecology* **22**, 1503-1517.
- Casewell M, Friis C, Marco E, McMullin P, Phillips I (2003) The European ban on growth-promoting antibiotics and emerging consequences for human and animal health. *Journal of Antimicrobial Chemotherapy* **52**, 159-161.
- Centner TJ, Ferreira S (2012) Ability of governments to take actions to confront incursions of diseases – a case study: citrus canker in Florida. *Plant Pathology* **61**, 821-828.
- Cesbron S, Briand M, Essakhi S, *et al.* (2015) Comparative genomics of pathogenic and nonpathogenic strains of *Xanthomonas arboricola* unveil molecular and evolutionary events linked to pathoadaptation. *Front. Plant Sci.* **6**, 1126.
- Chang CJ, Garnier M, Zreik L, Rossetti V, Bové JM (1993) Culture and serological detection of the xylem-limited bacterium causing citrus variegated chlorosis and its identification as a strain of *Xylella fastidiosa*. *Current Microbiology* **27**, 137-142.
- Chatterjee DK, Kellogg S, Hamada S, Chakrabarty A (1981) Plasmid specifying total degradation of 3-chlorobenzoate by a modified ortho pathway. *Journal of Bacteriology* **146**, 639-646.

- Chen NW, Serres-Giardi L, Ruh M, *et al.* (2018) Horizontal gene transfer plays a major role in the pathological convergence of *Xanthomonas* lineages on common bean. *Bmc Genomics* **19**, 606.
- Cheng W, Li J, Wu Y, *et al.* (2016) Behavior of antibiotics and antibiotic resistance genes in eco-agricultural system: a case study. *Journal of Hazardous Materials* **304**, 18-25.
- Chiou CS, Jones AL (1993) Nucleotide sequence analysis of a transposon (Tn5393) carrying streptomycin resistance genes in *Erwinia amylovora* and other gram-negative bacteria. *Journal of Bacteriology* **175**, 732-740.
- Chopin A, Chopin M-C, Moillo-Batt A, Langella P (1984) Two plasmid-determined restriction and modification systems in *Streptococcus lactis*. *Plasmid* **11**, 260-263.
- Collins RE, Higgs PG (2012) Testing the infinitely many genes model for the evolution of the bacterial core genome and pangenome. *Molecular Biology and Evolution* **29**, 3413-3425.
- Cooksey DA (1990) Genetics of bactericide resistance in plant pathogenic bacteria. *Annual Review of Phytopathology* **28**, 201-219.
- Cooksey DA (1993) Copper uptake and resistance in bacteria. *Mol Microbiol* **7**, 1-5.
- Cooksey DA, Azad HR, Cha JS, Lim CK (1990) Copper resistance gene homologs in pathogenic and saprophytic bacterial species from tomato. *Appl. Environ. Microbiol.* **56**, 431-435.
- Corel E, Lopez P, Meheust R, Baptiste E (2016) Network-thinking: graphs to analyze microbial complexity and evolution. *Trends in Microbiology* **24**, 224-237.
- Costerton JW, Stewart PS, Greenberg EP (1999) Bacterial biofilms: a common cause of persistent infections. *Science* **284**, 1318-1322.
- Cottyn B, Debode J, Regalado E, Mew T, Swings J (2009) Phenotypic and genetic diversity of rice seed-associated bacteria and their role in pathogenicity and biological control. *Journal of applied microbiology* **107**, 885-897.
- Cottyn B, Regalado E, Lanoot B, *et al.* (2001) Bacterial populations associated with rice seed in the tropical environment. *Phytopathology* **91**, 282-292.
- Couturier M, Bex F, Bergquist PL, Maas WK (1988) Identification and classification of bacterial plasmids. *Microbiological Reviews* **52**, 375-395.
- Crossman LC, Gould VC, Dow JM, *et al.* (2008) The complete genome, comparative and functional analysis of *Stenotrophomonas maltophilia* reveals an organism heavily shielded by drug resistance determinants. *Genome Biol* **9**, R74.
- Cubero J, Gell I, Johnson EG, Redondo A, Graham JH (2011) Unstable green fluorescent protein for study of *Xanthomonas citri* subsp. *citri* survival on citrus. *Plant Pathology* **60**, 977-985.
- Da Silva AC, Ferro JA, Reinach FC, *et al.* (2002) Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417**, 459-463.
- DAAF (2014) La production fruitière à la Réunion, recensement agricole 2010 (ed. agreste).
- Dagan T (2011) Phylogenomic networks. *Trends in Microbiology* **19**, 483-491.
- Dagan T, Artzy-Randrup Y, Martin W (2008) Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proceedings of the National Academy of Sciences* **105**, 10039-10044.
- Dahlberg C, Chao L (2003) Amelioration of the cost of conjugative plasmid carriage in *Escherichia coli* K12. *Genetics* **165**, 1641-1649.
- Davey MR, Anthony P, Power JB, Lowe KC (2005) Plant protoplasts: status and biotechnological perspectives. *Biotechnology Advances* **23**, 131-171.
- Davies J, Davies D (2010) Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews* **74**, 417-433.

- Davies JE (1997) Origins, acquisition and dissemination of antibiotic resistance determinants. *Ciba Found Symp* **207**, 15-27; discussion 27-35.
- de Boer TE, Taş N, Braster M, *et al.* (2012) The influence of long-term copper contaminated agricultural soil at different pH levels on microbial communities and springtail transcriptional regulation. *Environmental Science & Technology* **46**, 60-68.
- De la Cruz F, Davies J (2000) Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends in Microbiology* **8**, 128-133.
- De La Cruz F, Frost LS, Meyer RJ, Zechner EL (2010) Conjugative DNA metabolism in Gram-negative bacteria. *FEMS Microbiology Reviews* **34**, 18-40.
- Del Solar G, Giraldo R, Ruiz-Echevarría MJ, Espinosa M, Díaz-Orejas R (1998) Replication and control of circular bacterial plasmids. *Microbiology and Molecular Biology Reviews* **62**, 434-464.
- Denny TP (1999) Autoregulator-dependent control of extracellular polysaccharide production in phytopathogenic bacteria. *European Journal of Plant Pathology* **105**, 417-430.
- Derso E, Vernière C, Pruvost O (2009) First report of *Xanthomonas citri* pv. *citri*-A* causing citrus canker on lime in Ethiopia. *Plant Disease* **93**, 203-203.
- Dimijian GG (1999) Pathogens and parasites: insights from evolutionary biology **12**, 175-187.
- Dobrindt U, Hochhut B, Hentschel U, Hacker J (2004) Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* **2**, 414-424.
- Dostal L, Shao S, Schildbach JF (2011) Tracking F plasmid TraI relaxase processing reactions provides insight into F plasmid transfer. *Nucleic Acids Research* **39**, 2658-2670.
- Dougherty K, Smith BA, Moore AF, *et al.* (2014) Multiple phenotypic changes associated with large-scale horizontal gene transfer. *Plos One* **9**, e102170.
- Duan YP, Castañeda A, Zhao G, Erdos G, Gabriel DW (1999) Expression of a single, host-specific, bacterial pathogenicity gene in plant cells elicits division, enlargement, and cell death. *Molecular Plant-Microbe Interactions* **12**, 556-560.
- Dubey GP, Ben-Yehuda S (2011) Intercellular nanotubes mediate bacterial communication. *Cell* **144**, 590-600.
- Duchene S, Holt KE, Weill FX, *et al.* (2016) Genome-scale rates of evolutionary change in bacteria. *Microb Genom* **2**, e000094.
- Ducruet C (2010a) Les mesures globales d'un réseau.
- Ducruet C (2010b) Les mesures locales d'un réseau.
- Dupont CL, Grass G, Rensing C (2011) Copper toxicity and the origin of bacterial resistance--new insights and applications. *Metallomics* **3**, 1109-1118.
- Dye D (1969) Eradicating citrus canker from New Zealand. *New Zealand Journal of Agriculture* **118**, 20-&.
- Eberhard WG (1990) Evolution in bacterial plasmids and levels of selection. *The Quarterly review of biology* **65**, 3-22.
- Ebert D (1998) Experimental Evolution of Parasites. *Science* **282**, 1432-1436.
- Elisabeth G, J. CP, Gabriel W, Steffen B (2018) Type IV secretion in Gram-negative and Gram-positive bacteria. *Molecular Microbiology* **107**, 455-471.
- Essakhi S, Cesbron S, Fischer-Le Saux M, *et al.* (2015) Phylogenetic and variable-number tandem-repeat analyses identify nonpathogenic *Xanthomonas arboricola* lineages lacking the canonical type III secretion system. *Applied and Environmental Microbiology* **81**, 5395-5410.
- Ference CM, Gochez AM, Behlau F, *et al.* (2018) Recent advances in the understanding of *Xanthomonas citri* ssp. *citri* pathogenesis and citrus canker disease management. *Molecular plant pathology* **19**, 1302-1318.

- Ferreira RM, de Oliveira AC, Moreira LM, *et al.* (2015) A TALE of transposition: Tn3-like transposons play a major role in the spread of pathogenicity determinants of *Xanthomonas citri* and other xanthomonads. *MBio* **6**, e02505-02514.
- Fisher RA (1930) *The genetical theory of natural selection* Oxford Univ. Press.
- Fondi M, Fani R (2010) The horizontal flow of the plasmid resistome: clues from inter-generic similarity networks. *Environmental Microbiology* **12**, 3228-3242.
- Francis MI, Redondo A, Burns JK, Graham JH (2009) Soil application of imidacloprid and related SAR-inducing compounds produces effective and persistent control of citrus canker. *European Journal of Plant Pathology* **124**, 283-292.
- Franke S, Grass G, Rensing C, Nies DH (2003) Molecular analysis of the copper-transporting efflux system CusCFBA of *Escherichia coli*. *Journal of Bacteriology* **185**, 3804-3812.
- Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP (2009) The bacterial species challenge: making sense of genetic and ecological diversity. *Science* **323**, 741-746.
- Freeman LC (1977) A set of measures of centrality based on betweenness. *Sociometry*, 35-41.
- Frischer M, Stewart G, Paul J (1994) Plasmid transfer to indigenous marine bacterial populations by natural transformation. *Fems Microbiology Ecology* **15**, 127-135.
- Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology* **3**, 722.
- Gabaldón T, Koonin EV (2013) Functional and evolutionary implications of gene orthology. *Nature Reviews Genetics* **14**, 360.
- Galimand M, Guiyoule A, Gerbaud G, *et al.* (1997) Multidrug resistance in *Yersinia pestis* mediated by a transferable plasmid. *N Engl J Med* **337**, 677-680.
- Galyov EE, Håkansson S, Forsberg Å, Wolf-Watz H (1993) A secreted protein kinase of *Yersinia pseudotuberculosis* is an indispensable virulence determinant. *Nature* **361**, 730.
- Gambley C, Miles A, Ramsden M, *et al.* (2009) The distribution and spread of citrus canker in Emerald, Australia. *Australasian Plant Pathology* **38**, 547-557.
- Garcia-Vallvé S, Romeu A, Palau J (2000) Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Research* **10**, 1719-1725.
- Garcillan-Barcia MP, Alvarado A, de la Cruz F (2011) Identification of bacterial plasmids based on mobility and plasmid population biology. *FEMS Microbiol Rev* **35**, 936-956.
- Garcillan-Barcia MP, Francia MV, de la Cruz F (2009) The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol Rev* **33**, 657-687.
- Garita-Cambronero J, Palacio-Bielsa A, López MM, Cubero J (2016) Comparative genomic and phenotypic characterization of pathogenic and non-pathogenic strains of *Xanthomonas arboricola* reveals insights into the infection process of bacterial spot disease of stone fruits. *Plos One* **11**, e0161977.
- Gillings MR (2014) Integrins: past, present, and future. *Microbiol Mol Biol Rev* **78**, 257-277.
- Giovanardi D, Bonneau S, Gironde S, *et al.* (2016) Morphological and genotypic features of *Xanthomonas arboricola* pv. *juglandis* populations from walnut groves in Romagna region, Italy. *European Journal of Plant Pathology* **145**, 1-16.
- Giraud M, Prunet JP, Chevallier A, S R (2007) *La bactériose du noyer - Évolution des populations de Xanthomonas en verger* InfosCTIFL.
- Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 7821-7826.
- Glick BR (1995) Metabolic load and heterologous gene expression. *Biotechnology Advances* **13**, 247-261.
- Gmitter FG, Hu X (1990) The possible role of Yunnan, China, in the origin of contemporary *Citrus* species (Rutaceae). *Economic Botany* **44**, 267-277.

- Gochez AM, Huguet-Tapia JC, Minsavage GV, *et al.* (2018) Pacbio sequencing of copper-tolerant *Xanthomonas citri* reveals presence of a chimeric plasmid structure and provides insights into reassortment and shuffling of transcription activator-like effectors among *X. citri* strains. *Bmc Genomics* **19**, 16.
- Goessweiner-Mohr N, Arends K, Keller W, Grohmann E (2013) Conjugative type IV secretion systems in Gram-positive bacteria. *Plasmid* **70**, 289-302.
- Goessweiner-Mohr N, Arends K, Keller W, Grohmann E (2014a) Conjugation in Gram-Positive Bacteria. *Microbiol Spectr* **2**, PLAS-0004-2013.
- Goessweiner-Mohr N, Fercher C, Arends K, *et al.* (2014b) The type IV secretion protein TraK from the *Enterococcus* conjugative plasmid pIP501 exhibits a novel fold. *Acta Crystallographica Section D: Biological Crystallography* **70**, 1124-1135.
- Gogarten JP, Doolittle WF, Lawrence JG (2002) Prokaryotic evolution in light of gene transfer. *Molecular Biology and Evolution* **19**, 2226-2238.
- Gogarten JP, Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* **3**, 679-687.
- Golmohammadi M, Cubero J, López MM, Llop P (2013) The viable but non-culturable state in *Xanthomonas citri* subsp. *citri* is a reversible state induced by low nutrient availability and copper stress conditions. *Journal of Life Sciences* **7**, 1019.
- Gomis-Ruth FX, Coll M (2006) Cut and move: protein machinery for DNA processing in bacterial conjugation. *Curr Opin Struct Biol* **16**, 744-752.
- Gonzalez CF, Pettit EA, Valadez VA, Provin EM (1997) Mobilization, cloning, and sequence determination of a plasmid-encoded polygalacturonase from a phytopathogenic *Burkholderia* (*Pseudomonas*) *cepacia*. *Mol Plant Microbe Interact* **10**, 840-851.
- González JM, Brown BJ, Carlton BC (1982) Transfer of *Bacillus thuringiensis* plasmids coding for delta-endotoxin among strains of *B. thuringiensis* and *B. cereus*. *Proceedings of the National Academy of Sciences* **79**, 6951-6955.
- Gordon JL, Lefeuvre P, Escalon A, *et al.* (2015) Comparative genomics of 43 strains of *Xanthomonas citri* pv. *citri* reveals the evolutionary events giving rise to pathotypes with different host ranges. *Bmc Genomics* **16**, 1098.
- Gottwald TR (2010) Current epidemiological understanding of citrus huanglongbing. *Annual Review of Phytopathology* **48**, 119-139.
- Gottwald TR, Graham JH (1992) A device for precise and nondisruptive stomatal inoculation of leaf tissue with bacterial pathogens. *Phytopathology* **82**, 930-935.
- Gottwald TR, Hughes G, Graham JH, Sun X, Riley T (2001) The citrus canker epidemic in Florida: the scientific basis of regulatory eradication policy for an invasive species. *Phytopathology* **91**, 30-34.
- Gottwald TR, Irey M (2007) Post-hurricane analysis of citrus canker II: predictive model estimation of disease spread and area potentially impacted by various eradication protocols following catastrophic weather events. *Plant Health Progress* **10**.
- Graham J, Timmer L (1992) *Phytophthora* diseases of citrus. *Plant diseases of international importance* **3**, 250-269.
- Graham JH, Gottwald TR, Cubero J, Achor DS (2004) *Xanthomonas axonopodis* pv. *citri*: factors affecting successful eradication of citrus canker. *Mol. Plant Pathol.* **5**, 1-15.
- Graham JH, Myers ME (2011) Soil application of SAR inducers imidacloprid, thiamethoxam, and acibenzolar-S-methyl for citrus canker control in young grapefruit trees. *Plant Disease* **95**, 725-728.
- Graham JH, Myers ME (2013) Integration of soil applied neonicotinoid insecticides and acibenzolar-S-methyl for systemic acquired resistance (SAR) control of citrus canker on young citrus trees. *Crop Protection* **54**, 239-243.

- Grass G, Rensing C (2001) CueO is a multi-copper oxidase that confers copper tolerance in *Escherichia coli*. *Biochemical and biophysical research communications* **286**, 902-908.
- Grindley NDF (2002) The movement of Tn3-like elements: transposition and cointegrate resolution. In: *Mobile DNA II*. American Society of Microbiology.
- Guglielmini J, de la Cruz F, Rocha EP (2013) Evolution of conjugation and type IV secretion systems. *Molecular Biology and Evolution* **30**, 315-331.
- Guglielmini J, Néron B, Abby SS, *et al.* (2014) Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. *Nucleic Acids Research* **42**, 5715-5727.
- Guglielmini J, Quintais L, Garcillán-Barcia MP, de la Cruz F, Rocha EPC (2011) The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *PLOS Genetics* **7**, e1002222.
- Haegeman B, Weitz JS (2012) A neutral theory of genome evolution and the frequency distribution of genes. *Bmc Genomics* **13**, 196.
- Halary S, Leigh JW, Cheaib B, Lopez P, Baptiste E (2010) Network analyses structure genetic diversity in independent genetic worlds. *Proceedings of the National Academy of Sciences, USA* **107**, 127-132.
- Harrison E, Brockhurst MA (2012) Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends in Microbiology* **20**, 262-267.
- Hayward AC (1993) The hosts of *Xanthomonas*. In: *Xanthomonas* (eds. Swings JG, Civerolo EL), pp. 1-119. Springer Netherlands, Dordrecht.
- Heams T, Huneman P, Lecointre G, Silberstein M (2009) Les mondes darwiniens. Syllepse. Chapitre 4, p 88.
- Heppner JB (1993) *Citrus leafminer* University of Florida Cooperative Extension Service, Institute of Food and Agriculture Sciences, EDIS.
- Hernandez- Montes G, Argüello J, Valderrama B (2012) *Evolution and diversity of periplasmic proteins involved in copper homeostasis in gamma proteobacteria*.
- Heuer H, Abdo Z, Smalla K (2008) Patchy distribution of flexible genetic elements in bacterial populations mediates robustness to environmental uncertainty. *Fems Microbiology Ecology* **65**, 361-371.
- Heuer H, Schmitt H, Smalla K (2011) Antibiotic resistance gene spread due to manure application on agricultural fields. *Current Opinion in Microbiology* **14**, 236-243.
- Hobman JL, Crossman LC (2015) Bacterial antimicrobial metal ion resistance. *J Med Microbiol* **64**, 471-497.
- Holt KE, Thieu Nga TV, Thanh DP, *et al.* (2013) Tracking the establishment of local endemic populations of an emergent enteric pathogen. *Proc Natl Acad Sci U S A* **110**, 17522-17527.
- Hong R, Kang TY, Michels CA, Gadura N (2012) Membrane lipid peroxidation in copper alloy-mediated contact killing of *Escherichia coli*. *Appl Environ Microbiol* **78**, 1776-1784.
- Hooper SD, Mavromatis K, Kyrpides NC (2009) Microbial co-habitation and lateral gene transfer: what transposases can tell us. *Genome Biology* **10**, R45.
- Hsiao Y-M, Liu Y-F, Lee P-Y, *et al.* (2011) Functional characterization of *copA* gene encoding multicopper oxidase in *Xanthomonas campestris* pv. *campestris*. *Journal of Agricultural and Food Chemistry* **59**, 9290-9302.
- Hu H-W, Wang J-T, Li J, *et al.* (2016a) Field-based evidence for copper contamination induced changes of antibiotic resistance in agricultural soils. *Environmental Microbiology* **18**, 3896-3909.
- Hu Y, Duan S, Zhang Y, *et al.* (2016b) Temporal transcription profiling of sweet orange in response to PthA4-mediated *Xanthomonas citri* subsp. *citri* infection. *Phytopathology* **106**, 442-451.

- Hulme PE (2009) Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology* **46**, 10-18.
- Hurtle W, Lindler L, Fan W, *et al.* (2003) Detection and identification of ciprofloxacin-resistant *Yersinia pestis* by denaturing high-performance liquid chromatography. *J Clin Microbiol* **41**, 3273-3283.
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**, 254-267.
- Huson DH, Scornavacca C (2011) A survey of combinatorial methods for phylogenetic networks. *Genome Biology and Evolution* **3**, 23-35.
- Huxley J (1942) *Evolution the modern synthesis* G. Allen & Unwin Limited.
- Hyun J-W, Kim H-J, Yi P-H, Hwang R-Y, Park E-W (2012) Mode of action of streptomycin resistance in the citrus canker pathogen (*Xanthomonas smithii* subsp. *citri*) in Jeju Island. *The Plant Pathology Journal* **28**, 207-211.
- Irey M, Gottwald TR, Graham JH, Riley TD, Carlton G (2006) Post-hurricane analysis of citrus canker spread and progress towards the development of a predictive model to estimate disease spread due to catastrophic weather events. *Plant Health Prog.*, DOI: 10.1094/PHP-2006-0822-1001-RS.
- Istock CA, Duncan KE, Ferguson N, Zhou X (1992) Sexuality in a natural population of bacteria—*Bacillus subtilis* challenges the clonal paradigm. *Molecular Ecology* **1**, 95-103.
- Jalasvuori M, Friman V-P, Nieminen A, Bamford JK, Buckling A (2011) Bacteriophage selection against a plasmid-encoded sex apparatus leads to the loss of antibiotic-resistance plasmids. *Biology letters*, rsbl20110384.
- Jaramillo VDA, Sukno SA, Thon MR (2015) Identification of horizontally transferred genes in the genus *Colletotrichum* reveals a steady tempo of bacterial to fungal gene transfer. *Bmc Genomics* **16**, 2.
- Jechalke S, Heuer H, Siemens J, Amelung W, Smalla K (2014) Fate and effects of veterinary antibiotics in soil. *Trends in Microbiology* **22**, 536-545.
- Jenkins GM, Rambaut A, Pybus OG, Holmes EC (2002) Rates of molecular evolution in RNA viruses: A quantitative phylogenetic analysis. *Journal of Molecular Evolution* **54**, 156-165.
- Jetter KM, Civerolo EL, Sumner DA, Sumner D (2000) Ex ante economics of exotic disease policy: citrus canker in California. *Exotic Pests and Diseases: Biology, Economics and Public Policy for Biosecurity*, 121-149.
- Ji X, Shen Q, Liu F, *et al.* (2012) Antibiotic resistance gene abundances associated with antibiotics and heavy metals in animal manures and agricultural soils adjacent to feedlots in Shanghai; China. *J Hazard Mater* **235-236**, 178-185.
- Jia H, Orbovic V, Jones JB, Wang N (2016) Modification of the PthA4 effector binding elements in Type I CsLOB1 promoter using Cas9/sgRNA to produce transgenic Duncan grapefruit alleviating XccΔpthA4:dCsLOB1.3 infection. *Plant Biotechnology Journal* **14**, 1291-1301.
- Jia H, Zhang Y, Orbović V, *et al.* (2017) Genome editing of the disease susceptibility gene CsLOB1 in citrus confers resistance to citrus canker. *Plant Biotechnology Journal* **15**, 817-823.
- Jibrin MO, Potnis N, Timilsina S, *et al.* (2018) Genomic inference of recombination-mediated evolution in *Xanthomonas euvesicatoria* and *X. perforans*. *Applied and Environmental Microbiology*, AEM. 00136-00118.
- Jinek M, Chylinski K, Fonfara I, *et al.* (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*.
- Jones A (1982) Chemical control of phytopathogenic prokaryotes. *Phytopathogenic prokaryotes* **2**, 399-414.

- Jones JD, Dangl JL (2006a) The plant immune system. *Nature* **444**, 323.
- Jones JDG, Dangl JL (2006b) The plant immune system. *Nature* **444**, 323.
- Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics* **9**, 605.
- Keith LM, Sewake KT, Zee FT (2005) Isolation and characterization of *Burkholderia gladioli* from orchids in Hawaii. *Plant Disease* **89**, 1273-1278.
- Kettlewell HBD (1958) A survey of the frequencies of *Biston betularia* (L.) (Lep.) and its melanic forms in Great Britain. *Heredity* **12**, 51-72.
- Keyer K, Imlay JA (1996) Superoxide accelerates DNA damage by elevating free-iron levels. *Proceedings of the National Academy of Sciences* **93**, 13635-13640.
- Khalaf A, Moore GA, Jones JB, Gmitter FG (2007) New insights into the resistance of Nagami kumquat to canker disease. *Physiological and Molecular Plant Pathology* **71**, 240-250.
- Kholodii G, Yurieva O, Mindlin S, *et al.* (2000) Tn5044, a novel Tn3 family transposon coding for temperature-sensitive mercury resistance. *Research in Microbiology* **151**, 291-302.
- Kim E-H, Nies DH, McEvoy MM, Rensing C (2011) Switch or funnel: how RND-type transport systems control periplasmic metal homeostasis. *Journal of Bacteriology* **193**, 2381-2387.
- Kim J-G, Park BK, Kim S-U, *et al.* (2006) Bases of biocontrol: sequence predicts synthesis and mode of action of agrocin 84, the trojan horse antibiotic that controls crown gall. *Proceedings of the National Academy of Sciences* **103**, 8846-8851.
- Kirby R, Hopwood DA (1977) Genetic determination of methylenomycin synthesis by the SCP1 plasmid of *Streptomyces coelicolor* A3(2). *Microbiology* **98**, 239-252.
- Klaenhammer TR, Sanozky RB (1985) Conjugal transfer from *Streptococcus lactis* ME2 of plasmids encoding phage resistance, nisin resistance and lactose-fermenting ability: evidence for a high-frequency conjugative plasmid responsible for abortive infection of virulent bacteriophage. *Microbiology* **131**, 1531-1541.
- Kloesges T, Popa O, Martin W, Dagan T (2011) Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency at different phylogenetic depths. *Molecular Biology and Evolution* **28**, 1057-1074.
- Kogan M (1998) Integrated pest management: historical perspectives and contemporary developments. *Annual review of entomology* **43**, 243-270.
- Koizumi M, Kochinotsu B (1977) Relation of temperature to the development of citrus canker lesions in the spring **3**, 924-928.
- Koonin EV (2015) The turbulent network dynamics of microbial evolution and the statistical tree of life. *Journal of Molecular Evolution* **80**, 244-250.
- Koonin EV, Makarova KS, Aravind L (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* **55**, 709-742.
- Kulakauskas S, Lubys A, Ehrlich SD (1995) DNA restriction-modification systems mediate plasmid maintenance. *Journal of Bacteriology* **177**, 3451-3454.
- Kunin V, Goldovsky L, Darzentas N, Ouzounis CA (2005) The net of life: Reconstructing the microbial phylogenetic network. *Genome Research* **15**, 954-959.
- Lang AS, Zhaxybayeva O, Beatty JT (2012) Gene transfer agents: phage-like elements of genetic exchange. *Nature reviews. Microbiology* **10**, 472-482.
- Lapierre P, Gogarten JP (2009) Estimating the size of the bacterial pan-genome. *Trends in Genetics* **25**, 107-110.
- Lawrence JG, Ochman H (1997) Amelioration of bacterial genomes: rates of change and exchange. *Journal of Molecular Evolution* **44**, 383-397.
- Lecointre G (2012) L'adaptation et ses alternatives. In: *L'homme peut-il s'adapter à lui-même?*, pp. 60-65. Editions Quæ.

- Leduc A, Traore YN, Boyer K, *et al.* (2015) Bridgehead invasion of a monomorphic plant pathogenic bacterium: *Xanthomonas citri* pv. *citri*, an emerging citrus pathogen in Mali and Burkina Faso. *Environmental Microbiology* **17**, 4429-4442.
- Leduc A, Vernière C, Boyer C, *et al.* (2011) First report of *Xanthomonas citri* pv. *citri* pathotype A causing asiatic citrus canker on grapefruit and mexican lime in Senegal. *Plant Disease* **95**, 1311-1311.
- Lee Y, Schroth M, Henderson M, *et al.* (1993) Increased toxicity of iron-amended copper-containing bactericides to the walnut blight pathogen *Xanthomonas campestris* pv. *juglandis*. *Phytopathology* **83**, 1460-1465.
- Lee YA, Henderson M, Panopoulos NJ, Schroth MN (1994) Molecular-cloning, chromosomal mapping, and sequence analysis of copper resistance genes from *Xanthomonas campestris* pv. *juglandis* - homology with small blue copper proteins and multicopper oxidase. *Journal of Bacteriology* **176**, 173-188.
- Lemire JA, Harrison JJ, Turner RJ (2013) Antimicrobial activity of metals: mechanisms, molecular targets and applications. *Nature Reviews Microbiology* **11**, 371.
- Leroux P, Gardan L (2002) La lutte chimique en phytopathologie. In: *Traité de phytopathologie*. Semal, J. and Lepaire, P., Deboek, Bruxelles, Belgique.
- Li C, Chen J, Wang J, *et al.* (2015) Occurrence of antibiotics in soils and manures from greenhouse vegetable production bases of Beijing, China and an associated risk assessment. *Science of The Total Environment* **521-522**, 101-107.
- Li J, Wang N (2011) The wxacO gene of *Xanthomonas citri* ssp. *citri* encodes a protein with a role in lipopolysaccharide biosynthesis, biofilm formation, stress tolerance and virulence. *Mol Plant Pathol* **12**, 381-396.
- Li J, Wang N (2012) The gpsX gene encoding a glycosyltransferase is important for polysaccharide production and required for full virulence in *Xanthomonas citri* subsp. *citri*. *Bmc Microbiology* **12**, 31.
- Liebert CA, Hall RM, Summers AO (1999) Transposon Tn21, flagship of the floating genome. *Microbiol Mol Biol Rev* **63**, 507-522.
- Lipsitch M, Singer RS, Levin BR (2002) Antibiotics in agriculture: When is it time to close the barn door? *Proceedings of the National Academy of Sciences* **99**, 5752-5754.
- Llosa M, Gomis-Ruth FX, Coll M, de la Cruz Fd F (2002) Bacterial conjugation: a two-step mechanism for DNA transport. *Mol Microbiol* **45**, 1-8.
- Lobkovsky AE, Wolf YI, Koonin EV (2013) Gene frequency distributions reject a neutral model of genome evolution. *Genome Biol Evol* **5**, 233-242.
- Lodwick D, Owen D, Strike P (1990) DNA sequence analysis of the imp UV protection and mutation operon of the plasmid TP110: identification of a third gene. *Nucleic Acids Research* **18**, 5045-5050.
- Loftie-Eaton W, Tucker A, Norton A, Top EM (2014) Flow cytometry and real-time quantitative PCR as tools for assessing plasmid persistence. *Applied and Environmental Microbiology* **80**, 5439-5446.
- Ma W, Berkowitz GA (2007) The grateful dead: calcium and cell death in plant innate immunity. *Cell Microbiol* **9**, 2571-2585.
- Macomber L, Elsey SP, Hausinger RP (2011) Fructose-1, 6-bisphosphate aldolase (class II) is the primary site of nickel toxicity in *Escherichia coli*. *Molecular Microbiology* **82**, 1291-1300.
- Macomber L, Imlay JA (2009) The iron-sulfur clusters of dehydratases are primary intracellular targets of copper toxicity. *Proceedings of the National Academy of Sciences* **106**, 8344-8349.
- Macomber L, Rensing C, Imlay JA (2007) Intracellular copper does not catalyze the formation of oxidative DNA damage in *Escherichia coli*. *Journal of Bacteriology* **189**, 1616-1626.

- Madigan MT, Martinko JM, Parker J, Brock TD (1997) Biology of microorganisms. *Pernette Hall International Inc.*
- Maiden MC (2006) Multilocus sequence typing of bacteria. *Annu. Rev. Microbiol.* **60**, 561-588.
- Marraffini LA, Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in *Staphylococci* by targeting DNA. *Science* **322**, 1843-1845.
- Marshall BM, Ochieng DJ, Levy SB (2009) Commensals: underappreciated reservoir of antibiotic resistance. *Microbe* **4**, 231-238.
- Martinez-Freijo P, Fluit AC, Schmitz FJ, *et al.* (1998) Class I integrons in Gram-negative isolates from different European hospitals and association with decreased susceptibility to multiple antibiotic compounds. *J Antimicrob Chemother* **42**, 689-696.
- Maynard Smith J, Smith NH, O'Rourke M, Spratt BG (1993) How clonal are bacteria? *Proceedings of the National Academy of Sciences* **90**, 4384-4388.
- McCann HC, Li L, Liu Y, *et al.* (2017) Origin and evolution of the kiwifruit canker pandemic. *Genome Biology and Evolution* **9**, 932-944.
- McDaniel LD, Young E, Delaney J, *et al.* (2010) High frequency of horizontal gene transfer in the oceans. *Science* **330**, 50.
- McDonald BA, Linde C (2002) Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology* **40**, 349-379.
- McDonald BA, Stukenbrock EH (2016) Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. *Phil. Trans. R. Soc. B* **371**, 20160026.
- McGhee GC, Jones AL (2000) Complete nucleotide sequence of ubiquitous plasmid pEA29 from *Erwinia amylovora* strain Ea88: gene organization and intraspecies variation. *Applied and Environmental Microbiology* **66**, 4897-4907.
- McGowan J, Fitzpatrick DA (2017) Genomic, network, and phylogenetic analysis of the oomycete effector arsenal. *mSphere* **2**.
- McGrath BM, Pembroke JT (2004) Detailed analysis of the insertion site of the mobile elements R997, pMERPH, R392, R705 and R391 in *E. coli* K12. *Fems Microbiology Letters* **237**, 19-26.
- McInerney JO (2013) More than tree dimensions: inter-lineage evolution's ecological importance. *Trends in ecology & evolution* **28**, 624-625.
- McManus P, Jones A (1994) Epidemiology and genetic analysis of streptomycin-resistant *Erwinia amylovora* from Michigan and evaluation of oxytetracycline for control. *Phytopathology (USA)*.
- McManus PS, Stockwell VO, Sundin GW, Jones AL (2002) Antibiotic use in plant agriculture. *Annual Review of Phytopathology* **40**, 443-465.
- Meacock PA, Cohen SN (1980) Partitioning of bacterial plasmids during cell division: a cis-acting locus that accomplishes stable plasmid inheritance. *Cell* **20**, 529-542.
- Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R (2005) The microbial pan-genome. *Current opinion in genetics & development* **15**, 589-594.
- Mellano MA, Cooksey DA (1988) Nucleotide sequence and organization of copper resistance genes from *Pseudomonas syringae* pv. *tomato*. *Journal of Bacteriology* **170**, 2879-2883.
- Mellon M, Benbrook C, Benbrook K (2001) Hogging it: estimates of antimicrobial abuse in livestock. Cambridge, MA: Union of Concerned Scientists Publications.
- Merda D, Briand M, Bosis E, *et al.* (2017) Ancestral acquisitions, gene flow and multiple evolutionary trajectories of the type three secretion system and effectors in *Xanthomonas* plant pathogens. *Molecular Ecology* **26**, 5939-5952.
- Mhedbi-Hajri N, Darrasse A, Pigné S, *et al.* (2011) Sensing and adhesion are adaptive functions in the plant pathogenic xanthomonads. *BMC Evolutionary Biology* **11**, 67-67.

- Mochizuki S, Hiratsu K, Suwa M, *et al.* (2003) The large linear plasmid pSLA2-L of *Streptomyces rochei* has an unusually condensed gene organization for secondary metabolism. *Molecular Microbiology* **48**, 1501-1510.
- Modi RI, Adams J (1991) Coevolution in bacterial-plasmid populations. *Evolution* **45**, 656-667.
- Modi RI, Wilke CM, Rosenzweig RF, Adams J (1991) Plasmid macro-evolution: selection of deletions during adaptation in a nutrient-limited environment. *Genetica* **84**, 195-202.
- Mongkolrattanothai K, Boyle S, Murphy TV, Daum RS (2004) Novel Non-*mecA*-containing Staphylococcal chromosomal cassette composite island containing *pbp4* and *tagF* genes in a commensal Staphylococcal species: a possible reservoir for antibiotic resistance islands in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy* **48**, 1823-1836.
- Moreno P, Ambrós S, ALBIACH-MARTÍ MR, Guerri J, Pena L (2008) Citrus tristeza virus: a pathogen that changed the course of the citrus industry. *Molecular plant pathology* **9**, 251-268.
- Muller HJ (1927) Artificial transmutation of the gene. *Science* **66**, 84-87.
- Munson GP, Lam DL, Outten FW, O'Halloran TV (2000) Identification of a copper-responsive two-component system on the chromosome of *Escherichia coli* K-12. *Journal of Bacteriology* **182**, 5864-5871.
- Naas T, Cuzon G, Villegas MV, *et al.* (2008) Genetic structures at the origin of acquisition of the beta-lactamase bla KPC gene. *Antimicrob Agents Chemother* **52**, 1257-1263.
- Nakamura Y, Itoh T, Matsuda H, Gojobori T (2004) Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nature Genetics* **36**, 760.
- Nawapan S, Charoenlap N, Charoenwuttitarn A, *et al.* (2009) Functional and expression analyses of the cop operon, required for copper resistance in *Agrobacterium tumefaciens*. *Journal of Bacteriology* **191**, 5159-5168.
- Newman ME (2005) A measure of betweenness centrality based on random walks. *Social networks* **27**, 39-54.
- Nicolas E, Lambin M, Dandoy D, *et al.* (2015) The Tn3-family of replicative transposons. In: *Mobile DNA III*. American Society of Microbiology.
- Nies DH (1999) Microbial heavy-metal resistance. *Applied microbiology and biotechnology* **51**, 730-750.
- Niu XN, Wei ZQ, Zou HF, *et al.* (2015) Complete sequence and detailed analysis of the first indigenous plasmid from *Xanthomonas oryzae* pv. *oryzicola*. *Bmc Microbiology* **15**, 233.
- Nonaka L, Maruyama F, Miyamoto M, *et al.* (2012) Novel conjugative transferable multiple drug resistance plasmid pAQU1 from *Photobacterium damsela* subsp. *damsela* isolated from marine aquaculture environment. *Microbes Environ* **27**, 263-272.
- Nordstrom K, Gerdes K (2003) Clustering versus random segregation of plasmids lacking a partitioning function: a plasmid paradox? *Plasmid* **50**, 95-101.
- Norman A, Hansen LH, Sørensen SJ (2009) Conjugative plasmids: vessels of the communal gene pool. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**, 2275-2289.
- Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299-304.
- Ogiyama S, Sakamoto K, Suzuki H, *et al.* (2005) Accumulation of zinc and copper in an arable field after animal manure application. *Soil Science and Plant Nutrition* **51**, 801-808.
- Okinaka RT, Cloud K, Hampton O, *et al.* (1999) Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes. *Journal of Bacteriology* **181**, 6509-6515.

- Omar AA, Murata M, Yu Q, *et al.* (2017) Production of three new grapefruit cybrids with potential for improved citrus canker resistance. *In Vitro Cellular & Developmental Biology - Plant* **53**, 256-269.
- Oostendorp M, Kunz W, Dietrich B, Staub T (2001) Induced disease resistance in plants by chemicals. *European Journal of Plant Pathology* **107**, 19-28.
- Osborn AM, Böltner D (2002) When phage, plasmids, and transposons collide: genomic islands, and conjugative-and mobilizable-transposons as a mosaic continuum. *Plasmid* **48**, 202-212.
- Outten FW, Huffman DL, Hale JA, O'Halloran TV (2001) The independent cue and cus systems confer copper tolerance during aerobic and anaerobic growth in *Escherichia coli*. *Journal of Biological Chemistry* **276**, 30670-30677.
- Outten FW, Outten CE, Hale J, O'Halloran TV (2000) Transcriptional activation of an *Escherichia coli* copper efflux regulon by the chromosomal MerR homologue, cueR. *J Biol Chem* **275**, 31024-31029.
- Paauw A, Leverstein-van Hall MA, Verhoef J, Fluit AC (2010) Evolution in quantum leaps: multiple combinatorial transfers of HPI and other genetic modules in Enterobacteriaceae. *Plos One* **5**, e8662.
- Palmieri ACB, Amaral AMd, Homem RA, Machado MA (2010) Differential expression of pathogenicity-and virulence-related genes of *Xanthomonas axonopodis* pv. *citri* under copper stress. *Genetics and molecular biology* **33**, 348-353.
- Parker C, Meyer R (2005) Mechanisms of strand replacement synthesis for plasmid DNA transferred by conjugation. *Journal of Bacteriology* **187**, 3400-3406.
- Parkinson N, Cowie C, Heeney J, Stead D (2009) Phylogenetic structure of *Xanthomonas* determined by comparison of gyrB sequences. *International Journal of Systematic and Evolutionary Microbiology* **59**, 264-274.
- Partridge SR (2011) Analysis of antibiotic resistance regions in Gram-negative bacteria. *FEMS Microbiology Reviews* **35**, 820-855.
- Peng A, Chen S, Lei T, *et al.* (2017) Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene Cs LOB 1 promoter in citrus. *Plant Biotechnology Journal* **15**, 1509-1519.
- Pereira Y, Lagniel G, Godat E, *et al.* (2008) Chromate causes sulfur starvation in yeast. *Toxicol Sci* **106**, 400-412.
- Petrocelli S, Tondo ML, Daurelio LD, Orellano EG (2012) Modifications of *Xanthomonas axonopodis* pv. *citri* lipopolysaccharide affect the basal response and the virulence process during citrus canker. *Plos One* **7**, e40051.
- Petrovski S, Stanisich VA (2011) Embedded elements in the IncP β plasmids R772 and R906 can be mobilized and can serve as a source of diverse and novel elements. *Microbiology* **157**, 1714-1725.
- Platt TG, Morton ER, Barton IS, Bever JD, Fuqua C (2014) Ecological dynamics and complex interactions of *Agrobacterium* megaplasmids. *Frontiers in Plant Science* **5**, 635.
- Polz MF, Alm EJ, Hanage WP (2013) Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet* **29**, 170-175.
- Poole K (2012) Bacterial stress responses as determinants of antimicrobial resistance. *Journal of Antimicrobial Chemotherapy* **67**, 2069-2089.
- Poole K (2017) At the nexus of antibiotics and metals: the impact of Cu and Zn on antibiotic activity and resistance. *Trends in Microbiology* **25**, 820-832.
- Popa O, Dagan T (2011) Trends and barriers to lateral gene transfer in prokaryotes. *Curr Opin Microbiol* **14**, 615-623.

- Popa O, Hazkani-Covo E, Landan G, Martin W, Dagan T (2011) Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. *Genome Research* **21**, 599-609.
- Pothier JF, Vorhölter F-J, Blom J, *et al.* (2011) The ubiquitous plasmid pXap41 in the invasive phytopathogen *Xanthomonas arboricola* pv. *pruni*: complete sequence and comparative genomic analysis. *Fems Microbiology Letters* **323**, 52-60.
- Potnis N, Timilsina S, Strayer A, *et al.* (2015) Bacterial spot of tomato and pepper: diverse *Xanthomonas* species with a wide variety of virulence factors posing a worldwide challenge. *Mol Plant Pathol* **16**, 907-920.
- Pruvost O, Boher B, Brocherieux C, Nicole M, Chiroleu F (2002) Survival of *Xanthomonas axonopodis* pv. *citri* in leaf lesions under tropical environmental conditions and simulated splash dispersal of inoculum. *Phytopathology* **92**, 336-346.
- Pruvost O, Couteau A, Perrier X, Luisetti J (1998) Phenotypic diversity of *Xanthomonas* sp. *mangiferaeindicae*. *J Appl Microbiol* **84**, 115-124.
- Pruvost O, Magne M, Boyer K, *et al.* (2014) A MLVA genotyping scheme for global surveillance of the citrus pathogen *Xanthomonas citri* pv. *citri* suggests a worldwide geographical expansion of a single genetic lineage. *Plos One* **9**, e98129.
- Pruvost O, Robene I, Escalon A, *et al.* (2015) The dynamic world of the genus *Xanthomonas*. In: *Virulence mechanisms of plant pathogenic bacteria* (ed. Wang Nian JJBSGWWFFHSARCDLFLHJH), pp. 381-385. The American Phytopathological Society, Saint Paul, États-Unis.
- Qu Y, Daley AJ, Istivan TS, Rouch DA, Deighton MA (2010) Densely adherent growth mode, rather than extracellular polymer substance matrix build-up ability, contributes to high resistance of *Staphylococcus epidermidis* biofilms to antibiotics. *Journal of Antimicrobial Chemotherapy* **65**, 1405-1411.
- Ramón-Laca L (2003) The introduction of cultivated Citrus to Europe via northern Africa and the Iberian peninsula. *Economic Botany* **57**, 502-514.
- Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási A-L (2002) Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551-1555.
- Ravatn R, Studer S, Zehnder AJ, van der Meer JR (1998) Int-B13, an unusual site-specific recombinase of the bacteriophage P4 integrase family, is responsible for chromosomal insertion of the 105-kilobase *clc* element of *Pseudomonas* sp. strain B13. *Journal of Bacteriology* **180**, 5505-5514.
- Rensing C, Fan B, Sharma R, Mitra B, Rosen BP (2000) CopA: An *Escherichia coli* Cu(I)-translocating P-type ATPase. *Proceedings of the National Academy of Sciences* **97**, 652-656.
- Rensing C, Grass G (2003) *Escherichia coli* mechanisms of copper homeostasis in a changing environment. *FEMS Microbiology Reviews* **27**, 197-213.
- Richard D, Boyer C, Javegny S, *et al.* (2016a) First report of *Xanthomonas citri* pv. *citri* pathotype A causing asiatic citrus canker in Martinique, France. *Plant Disease* **100**, 1946-1946.
- Richard D, Boyer C, Lefeuvre P, *et al.* (2017a) Complete genome sequences of six copper-resistant *Xanthomonas* strains causing bacterial spot of solanaceous plants, belonging to *X. gardneri*, *X. euvesicatoria*, and *X. vesicatoria*, using long-read technology. *Genome Announc* **5**.
- Richard D, Boyer C, Lefeuvre P, Pruvost O (2016b) Complete genome sequence of a copper-resistant bacterium from the citrus phyllosphere, *Stenotrophomonas* sp. strain LM091, obtained using long-read technology. *Genome Announcements* **4**.

- Richard D, Boyer C, Vernière C, *et al.* (2017b) Complete genome sequences of six copper-resistant *Xanthomonas citri* pv. *citri* strains causing asiatic citrus canker, obtained using long-read technology. *Genome Announcements* **5**.
- Richard D, Ravigne V, Rieux A, *et al.* (2017c) Adaptation of genetically monomorphic bacteria: evolution of copper resistance through multiple horizontal gene transfers of complex and versatile mobile genetic elements. *Mol Ecol* **26**, 2131-2149.
- Richard D, Tribot N, Boyer C, *et al.* (2017d) First report of copper-resistant *Xanthomonas citri* pv. *citri* pathotype A causing asiatic citrus canker in Réunion, France. *Plant Disease*.
- Rigano LA, Siciliano F, Enrique R, *et al.* (2007) Biofilm formation, epiphytic fitness, and canker development in *Xanthomonas axonopodis* pv. *citri*. *Mol Plant Microbe Interact* **20**, 1222-1230.
- Rizzo L, Manaia C, Merlin C, *et al.* (2013) Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: A review. *Science of The Total Environment* **447**, 345-360.
- Rocha EPC, Cornet E, Michel B (2005) Comparative and evolutionary analysis of the bacterial homologous recombination systems. *PLOS Genetics* **1**, e15.
- Rouch D, Camakaris J, Lee BT, Luke RK (1985) Inducible plasmid-mediated copper resistance in *Escherichia coli*. *J Gen Microbiol* **131**, 939-943.
- Rozkov A, Avignone-Rossa CA, Ertl PF, *et al.* (2004) Characterization of the metabolic burden on *Escherichia coli* DH1 cells imposed by the presence of a plasmid containing a gene therapy sequence. *Biotechnology and Bioengineering* **88**, 909-915.
- Ryan RP, Vorhölter F-J, Potnis N, *et al.* (2011) Pathogenomics of *Xanthomonas*: understanding bacterium–plant interactions. *Nature Reviews Microbiology* **9**, 344.
- Rybak M, Minsavage GV, Stall RE, Jones JB (2009) Identification of *Xanthomonas citri* ssp. *citri* host specificity genes in a heterologous expression host. *Mol Plant Pathol* **10**, 249-262.
- Salyers AA, Gupta A, Wang Y (2004) Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends in Microbiology* **12**, 412-416.
- Santo CE, Lam EW, Elowsky CG, *et al.* (2011) Bacterial killing by dry metallic copper surfaces. *Applied and Environmental Microbiology* **77**, 794-802.
- Scally A, Durbin R (2012) Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet* **13**, 745-753.
- Scheck HJ, Pscheidt JW (1998) Effect of copper bactericides on copper-resistant and-sensitive strains of *Pseudomonas syringae* pv. *syringae*. *Plant Disease* **82**, 397-406.
- Schluter A, Heuer H, Szczepanowski R, *et al.* (2003) The 64 508 bp IncP-1beta antibiotic multiresistance plasmid pB10 isolated from a waste-water treatment plant provides evidence for recombination between members of different branches of the IncP-1beta group. *Microbiology* **149**, 3139-3153.
- Schoulties C, Civerolo E, Miller J, *et al.* (1987) Citrus canker in Florida. *Plant Disease* **71**, 388-395.
- Schubert TS, Rizvi SA, Sun X, *et al.* (2001) Meeting the challenge of eradicating citrus canker in Florida—again. *Plant Disease* **85**, 340-356.
- Scotland SM, Day NP, Rowe B (1983) Acquisition and maintenance of enterotoxin plasmids in wild-type strains of *Escherichia coli*. *Microbiology* **129**, 3111-3120.
- Seiler C, Berendonk TU (2012) Heavy metal driven co-selection of antibiotic resistance in soil and water bodies impacted by agriculture and aquaculture. *Front Microbiol* **3**, 399.
- Sena-Vélez M, Redondo C, Gell I, *et al.* (2015) Biofilm formation and motility of *Xanthomonas* strains with different citrus host range. *Plant Pathology* **64**, 767-775.
- Sengupta M, Austin S (2011) Prevalence and significance of plasmid maintenance functions in the virulence plasmids of pathogenic bacteria. *Infection and Immunity* **79**, 2502-2509.

- Serizawa S, Inoue K (1975) Studies on citrus canker III. The influence of wind on infection. *Bull. Shizuoka Citrus Exp. Sta.* **11**, 54-67.
- Séveno NA, Kallifidas D, Smalla K, *et al.* (2002) Occurrence and reservoirs of antibiotic resistance genes in the environment. *Reviews in Medical Microbiology* **13**, 15-27.
- Sharma SS, Dietz K-J (2009) The relationship between metal toxicity and cellular redox imbalance. *Trends in Plant Science* **14**, 43-50.
- Sheppard SK, Guttman DS, Fitzgerald JR (2018) Population genomics of bacterial host adaptation. *Nature Reviews Genetics*, 1.
- Shintani M, Sanchez ZK, Kimbara K (2015) Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. *Frontiers in Microbiology* **6**, 242.
- Siguier P, Gagnevin L, Chandler M (2009) The new IS1595 family, its relation to IS1 and the frontier between insertion sequences and transposons. *Research in Microbiology* **160**, 232-241.
- Siguier P, Gourbeyre E, Chandler M (2014) Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiology Reviews* **38**, 865-891.
- Siguier P, Gourbeyre E, Varani A, Ton-Hoang B, Chandler M (2015) Everyman's guide to bacterial insertion sequences. *Microbiology Spectrum* **3**.
- Singh SK, Grass G, Rensing C, Montfort WR (2004) Cuprous oxidase activity of CueO from *Escherichia coli*. *Journal of Bacteriology* **186**, 7815-7817.
- Skippington E, Ragan MA (2012) Phylogeny rather than ecology or lifestyle biases the construction of *Escherichia coli*–*Shigella* genetic exchange communities. *Open biology* **2**, 120112.
- Slater FR, Bailey MJ, Tett AJ, Turner SL (2008) Progress towards understanding the fate of plasmids in bacterial communities. *Fems Microbiology Ecology* **66**, 3-13.
- Smillie C, Garcillan-Barcia MP, Francia MV, Rocha EPC, de la Cruz F (2010) Mobility of plasmids. *Microbiology and Molecular Biology Reviews* **74**, 434-+.
- Smillie CS, Smith MB, Friedman J, *et al.* (2011) Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480**, 241.
- Smith IM (1985) *Fungicides for crop protection : 100 years of progress : proceedings of the Bordeaux Mixture Centenary Meeting : Bordeaux, France, 5th-7th September 1985*, Bordeaux, France.
- Smith JM, Feil EJ, Smith NH (2000) Population structure and evolutionary dynamics of pathogenic bacteria. *Bioessays* **22**, 1115-1122.
- Somerville C, Briscoe J (2001) Genetic engineering and water. *Science* **292**, 2217.
- Sommer MO, Dantas G, Church GM (2009) Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* **325**, 1128-1131.
- Spratt BG, Hanage WP, Feil EJ (2001) The relative contributions of recombination and point mutation to the diversification of bacterial clones. *Current Opinion in Microbiology* **4**, 602-606.
- Staehlin BM, Gibbons JG, Rokas A, O'Halloran TV, Slot JC (2016) Evolution of a heavy metal homeostasis/resistance island reflects increasing copper stress in *Enterobacteria*. *Genome Biology and Evolution* **8**, 811-826.
- Stall RE, Loschke DC, Jones JB (1986) Linkage of copper resistance and avirulence loci on a self-transmissible plasmid in *Xanthomonas campestris* pv. *vesicatoria*. **v. 76**.
- Stall RE, Miller J, Marco GM, de Echenique BC (1980) Population dynamics of *Xanthomonas citri* causing canker of citrus in Argentina **93**, 10-14.
- Stes E, Vandeputte OM, Jaziri ME, Holsters M, Vereecke D (2011) A successful bacterial coup d'état: how *Rhodococcus fascians* redirects plant development. *Annual Review of Phytopathology* **49**, 69-86.

- Stevenson J, Barwinska-Sendra A, Tarrant E, Waldron K (2013) Mechanism of action and applications of the antimicrobial properties of copper. *Microbial pathogens and strategies for combating them: Science, technology and education*, 468-479.
- Stockwell V, Duffy B (2012) Use of antibiotics in plant agriculture.
- Strange RN, Scott PR (2005) Plant disease: a threat to global food security. *Annual Review of Phytopathology* **43**, 83-116.
- Stukenbrock EH, McDonald BA (2008) The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology* **46**, 75-100.
- Sundin GW, Bender CL (1995) Expression of the strA-strB streptomycin resistance genes in *Pseudomonas syringae* and *Xanthomonas campestris* and characterization of IS6100 in *X. campestris*. *Applied and Environmental Microbiology* **61**, 2891-2897.
- Sundin GW, Jacobs JL, Murillo J (2000) Sequence diversity of rufA among natural isolates of *Pseudomonas syringae* and effect on function of rufAB-mediated UV radiation tolerance. *Appl Environ Microbiol* **66**, 5167-5173.
- Tamminen M, Virta M, Fani R, Fondi M (2012) Large-scale analysis of plasmid relationships through gene-sharing networks. *Molecular Biology and Evolution* **29**, 1225-1240.
- Tasho RP, Cho JY (2016) Veterinary antibiotics in animal waste, its distribution in soil and uptake by plants: A review. *Science of The Total Environment* **563-564**, 366-376.
- Tatum EL, Lederberg J (1947) Gene recombination in the bacterium *Escherichia coli*. *Journal of Bacteriology* **53**, 673-684.
- Teixeira EC, Franco de Oliveira JC, Marques Novo MT, Bertolini MC (2008) The copper resistance operon copAB from *Xanthomonas axonopodis* pathovar citri: gene inactivation results in copper sensitivity. *Microbiology* **154**, 402-412.
- Tetaz TJ, Luke R (1983) Plasmid-controlled resistance to copper in *Escherichia coli*. *Journal of Bacteriology* **154**, 1263-1268.
- Tettelin H, Masignani V, Cieslewicz MJ, et al. (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13950-13955.
- Thomas CM (2003) *Horizontal gene pool: bacterial plasmids and gene spread* CRC Press.
- Thrupp LA (1991) Long-term losses from accumulation of pesticide residues: a case of persistent copper toxicity in soils of Costa Rica. *Geoforum* **22**, 1-15.
- Timmer LW, Peever TL, Solel Z, Akimitsu K (2003) Alternaria diseases of citrus—novel pathosystems. *Phytopathologia Mediterranea* **42**, 99-112.
- Tock MR, Dryden DTF (2005) The biology of restriction and anti-restriction. *Current Opinion in Microbiology* **8**, 466-472.
- Toussaint A, Merlin C (2002) Mobile elements as a combination of functional modules. *Plasmid* **47**, 26-35.
- Triplett LR, Verdier V, Campillo T, et al. (2015) Characterization of a novel clade of *Xanthomonas* isolated from rice leaves in Mali and proposal of *Xanthomonas maliensis* sp. nov. *Antonie Van Leeuwenhoek* **107**, 869-881.
- Tsai JH (2008) Brown citrus aphid, *Toxoptera citricida* (Kirkaldy) (Hemiptera: Aphididae). In: *Encyclopedia of Entomology* (ed. Capinera JL), pp. 575-577. Springer Netherlands, Dordrecht.
- Valente TW, Coronges K, Lakon C, Costenbader E (2008) How correlated are network centrality measures? *Connections (Toronto, Ont.)* **28**, 16-26.
- Van-Zwieten L, Merrington G, Van-Zwieten M (2004) Review of impacts on soil biota caused by copper residues from fungicide application, 5-9.

- Van Belkum MJ, Kok J, Venema G (1992) Cloning, sequencing, and expression in *Escherichia coli* of lcnB, a third bacteriocin determinant from the lactococcal bacteriocin plasmid p9B4-6. *Applied and Environmental Microbiology* **58**, 572-577.
- Van Essen-Zandbergen A, Smith H, Veldman K, Mevius D (2007) Occurrence and characteristics of class 1, 2 and 3 integrons in *Escherichia coli*, *Salmonella* and *Campylobacter spp.* in the Netherlands. *Journal of Antimicrobial Chemotherapy* **59**, 746-750.
- Van Melderden L, Saavedra De Bast M (2009) Bacterial toxin-antitoxin systems: more than selfish entities? *PLoS Genet* **5**, e1000437.
- Vandroemme J, Cottyn B, Pothier JF, *et al.* (2013) *Xanthomonas arboricola* pv. *fragariae*: what's in a name? *Plant Pathology* **62**, 1123-1131.
- Vauterin L, Yang P, Alvarez A, *et al.* (1996a) Identification of non-pathogenic *Xanthomonas* strains associated with plants. *System. Appl. Microbiol.* **19**, 96-105.
- Vauterin L, Yang P, Swings J (1996b) Utilization of fatty acid methyl esters for the differentiation of new *Xanthomonas* species. *International Journal of Systematic and Evolutionary Microbiology* **46**, 298-304.
- Verniere C, Ngoc LBT, Jarne P, *et al.* (2014) Highly polymorphic markers reveal the establishment of an invasive lineage of the citrus bacterial pathogen *Xanthomonas citri* pv. *citri* in its area of origin. *Environmental Microbiology* **16**, 2226-2237.
- Vicente JG, Rothwell S, Holub EB, Studholme DJ (2017) Pathogenic, phenotypic and molecular characterisation of *Xanthomonas nasturtii* sp. nov. and *Xanthomonas floridensis* sp. nov., new species of *Xanthomonas* associated with watercress production in Florida. *International Journal of Systematic and Evolutionary Microbiology* **67**, 3645-3654
- Voloudakis AE, Reignier TM, Cooksey DA (2005) Regulation of resistance to copper in *Xanthomonas axonopodis* pv. *vesicatoria*. *Applied and Environmental Microbiology* **71**, 782-789.
- Wakelin SA, Chu G, Lardner R, Liang Y, McLaughlin M (2010) A single application of Cu to field soil has long-term effects on bacterial community structure, diversity, and soil processes. *Pedobiologia* **53**, 149-158.
- Waldron DE, Lindsay JA (2006) SauI: a novel lineage-specific type I restriction-modification system that blocks horizontal gene transfer into *Staphylococcus aureus* and between *S. aureus* isolates of different lineages. *Journal of Bacteriology* **188**, 5578-5585.
- Wales AD, Davies RH (2015) Co-selection of resistance to antibiotics, biocides and heavy metals, and its relevance to foodborne pathogens. *Antibiotics* **4**, 567-604.
- Walsh TR (2006) Combinatorial genetic evolution of multiresistance. *Current Opinion in Microbiology* **9**, 476-482.
- Wasukira A, Coulter M, Al-Sowayeh N, *et al.* (2014) Genome sequencing of *Xanthomonas vasicola* pathovar vasculorum reveals variation in plasmids and genes encoding lipopolysaccharide synthesis, type-IV pilus and type-III secretion effectors. *Pathogens (Basel, Switzerland)* **3**, 211-237.
- Welch RA, Burland V, Plunkett G, 3rd, *et al.* (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci USA* **99**, 17020-17024.
- Wiedenbeck J, Cohan FM (2011) Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiology Reviews* **35**, 957-976.
- Wightwick A, Walters R, Allinson G, Reichman S, Menzies N (2010) Environmental risks of fungicides used in horticultural production systems. In: *Fungicides*. InTech.
- Witte W (1998) Medical consequences of antibiotic use in agriculture. *Science* **279**, 996-997.

- Wozniak RA, Waldor MK (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol* **8**, 552-563.
- Wu GA, Terol J, Ibanez V, *et al.* (2018) Genomics of the origin and evolution of *Citrus*. *Nature* **554**, 311.
- Yan Q, Wang N (2012) High-throughput screening and analysis of genes of *Xanthomonas citri* subsp. *citri* involved in citrus canker symptom development. *Molecular Plant-Microbe Interactions* **25**, 69-84.
- Yano H, Genka H, Ohtsubo Y, *et al.* (2013) Cointegrate-resolution of toluene-catabolic transposon Tn4651: determination of crossover site and the segment required for full resolution activity. *Plasmid* **69**, 24-35.
- Yesuf M (2013) *Pseudocercospora* leaf and fruit spot disease of citrus: Achievements and challenges in the citrus industry: A review. *Agricultural Sciences* **4**, 324.
- Yonow T, Hattingh V, de Villiers M (2013) CLIMEX modelling of the potential global distribution of the citrus black spot disease caused by *Guignardia citricarpa* and the risk posed to Europe. *Crop Protection* **44**, 18-28.
- Zhou G, Shi Q-S, Huang X-M, Xie X-B (2015) The three bacterial lines of defense against antimicrobial agents. *International journal of molecular sciences* **16**, 21711-21733.
- Zhou Y, Sugiyama H, Nakano H, Johnson EA (1995) The genes for the *Clostridium botulinum* type G toxin complex are on a plasmid. *Infect Immun* **63**, 2087-2091.
- Zhou Z, McCann A, Litrup E, *et al.* (2013) Neutral genomic microevolution of a recently emerged pathogen, *Salmonella enterica* serovar Agona. *PLOS Genetics* **9**, e1003471.
- Zhou Z, McCann A, Weill F-X, *et al.* (2014) Transient Darwinian selection in *Salmonella enterica* serovar Paratyphi A during 450 years of global spread of enteric fever. *Proceedings of the National Academy of Sciences* **111**, 12199-12204.

Annexes

Annexe I : Pourcentage d'identité entre les séquences d'acides aminés encodées par les gènes des différents systèmes génétiques impliqués dans la résistance au cuivre et l'homéostasie du cuivre chez *Xanthomonas* (abrégié "X."). La localisation génomique et l'accension NCBI du génome encodant les différents systèmes sont précisés entre parenthèses.

			Résistance au cuivre																												Homéostasie du cuivre		
			copLAB														copABCD														cohLAB		
			X. citri (plasmide, HM362782)								X. citri (plasmide, CM007620)				X. arboricola (chromosome, CP0112251)				X. arboricola (chromosome, JZEG01000002)				E. coli (plasmide, DQ517526)				X. citri (chromosome, AE008923)						
			copL	copA	copB	copM	copG	copC	copD	copF	copA	copB	copC	copD	copA	copB	copC	copD	copA	copB	copC	copD	pcoA	pcoB	pcoC	pcoD	cohL	cohA	cohB				
Résistance au cuivre	copLAB	X. citri (plasmide)	copL	100	NA	27	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	33	NA	NA					
			copA	NA	100	NA	NA	NA	NA	NA	NA	80	23	NA	NA	80	23	NA	NA	51	NA	NA	NA	65	NA	NA	NA	67	NA				
			copB	27	NA	100	NA	NA	NA	NA	NA	NA	60	NA	NA	NA	60	NA	NA	NA	52	NA	NA	NA	45	NA	NA	NA	42				
			copM	NA	NA	NA	100	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	25	NA	NA				
			copG	NA	NA	NA	NA	100	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA				
			copC	NA	NA	NA	NA	NA	100	NA	NA	NA	NA	35	NA	NA	NA	35	NA	NA	40	NA	NA	NA	35	NA	NA	NA	NA				
			copD	NA	NA	NA	NA	NA	NA	100	NA	NA	NA	33	NA	NA	NA	34	NA	NA	NA	46	NA	NA	NA	34	NA	NA	NA				
			copF	NA	NA	NA	NA	NA	NA	100	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA				
	copABCD	X. citri (plasmide)	copA	NA	80	NA	NA	NA	NA	NA	100	23	NA	NA	97	23	NA	NA	52	NA	NA	NA	62	NA	NA	NA	NA	65	NA				
			copB	NA	23	60	NA	NA	NA	NA	23	100	NA	NA	24	98	NA	NA	NA	53	NA	NA	NA	49	NA	NA	NA	NA	46				
			copC	NA	NA	NA	NA	NA	35	NA	NA	NA	100	NA	NA	NA	98	NA	NA	51	NA	NA	54	NA	NA	NA	NA	NA					
			copD	NA	NA	NA	NA	NA	NA	33	NA	NA	NA	100	NA	NA	NA	97	NA	NA	NA	43	NA	NA	35	NA	NA	NA					
		X. arboricola (chromosome)	copA	NA	80	NA	NA	NA	NA	NA	97	24	NA	NA	100	23	NA	NA	51	NA	NA	NA	62	NA	NA	NA	NA	64	NA				
			copB	NA	23	60	NA	NA	NA	NA	23	98	NA	NA	23	100	NA	NA	NA	52	NA	NA	NA	49	NA	NA	NA	NA	46				
			copC	NA	NA	NA	NA	NA	35	NA	NA	NA	98	NA	NA	NA	100	NA	NA	52	NA	NA	NA	53	NA	NA	NA	NA					
			copD	NA	NA	NA	NA	NA	NA	34	NA	NA	NA	97	NA	NA	NA	100	NA	NA	NA	42	NA	NA	34	NA	NA	NA					
		E. coli (plasmide)	pcoA	NA	65	NA	NA	NA	NA	NA	62	NA	NA	NA	62	NA	NA	NA	54	NA	NA	NA	100	NA	NA	NA	NA	60	NA				
			pcoB	NA	NA	45	NA	NA	NA	NA	NA	49	NA	NA	NA	49	NA	NA	NA	46	NA	NA	NA	100	NA	NA	NA	NA	48				
			pcoC	NA	NA	NA	NA	NA	35	NA	NA	NA	54	NA	NA	NA	53	NA	NA	60	NA	NA	NA	100	NA	NA	NA	NA					
			pcoD	NA	NA	NA	NA	NA	NA	34	NA	NA	NA	35	NA	NA	NA	34	NA	NA	NA	42	NA	NA	NA	100	NA	NA	NA				
cohLAB	X. citri (chromosome)	cohL	33	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	100	NA	NA							
		cohA	NA	67	NA	25	NA	NA	NA	NA	65	NA	NA	NA	64	NA	NA	NA	53	NA	NA	NA	60	NA	NA	NA	100	NA					
		cohB	NA	NA	42	NA	NA	NA	NA	NA	NA	46	NA	NA	NA	46	NA	NA	49	NA	NA	NA	48	NA	NA	NA	100						

Annexe II Phénotype de résistance aux métaux lourds et systèmes génétiques associés chez Xcc

CMI à J+5 de différentes souches pour différents métaux lourds (mg/L). Les CMI sont évaluées comme indiqué par Pruvost *et al.* (1998). Lorsque l'analyse génomique a été effectuée, la présence totale (+), partielle (~) ou l'absence (-) des clusters de gènes correspondant au métal lourd étudié est représentée. S : *Stenotrophomonas* sp. ; Xcc : *X. citri* pv. *citri* (pathotype) ; Xcg : *X. cynarae* pv. *gardneri* ; Xee : *X. euvesicatoria* pv. *euvesicatoria* ; Xep : *X. euvesicatoria* pv. *perforans* ; Xv : *X. vesicatoria*

	Descriptif	Zinc sulfate chlorure	Sulfate de cadmium	Chlorure de cobalt	Czc (Cobalt, Zinc, Cadmium)	Sulfate de cuivre	Cop (Cuivre)	Cus (Cuivre)	Sodium arsenite	Ars (Arsenic)
C40	Xcc (A)	32	3,2	16		32			8	
LB302	Xcc (A ^w)	32	6,4	16		64			12	
LL74-4	Xcc (A)	32	6,4	16	+	128	+ (copLAB)	+	128	+
NCPPB3562	Xcc (A)	24	9,6	16		24			12	
LG97	Xcc (A)	32	12,8	16		32			8	
LM199	Xcc (A)	32	6,4	16	-	256	+ (copABCD)	~	12	-
LH201	Xcc (A)	32	6,4	16	+	256	+ (copLAB)	+	64	+
ICMP7383	Xcg	32	12,8	32	+	128	+ (copLAB)	+	192	+
LH3	Xep	32	12,8	16	+	128	~ (copLAB)	+	16	-
IAPAR306	Xcc (A)	32	12,8	24		40			40	
LMG911	Xv	24	12,8	24	+	256	+ (copLAB)	+	512	+
LE50-4	Xcc (A*)	32	12,8	16		32			16	
JS749-3	Xcg	32	12,8	32	+	256	+ (copLAB)	+	512	+

Annexe II

	Descriptif	Zinc sulfate chlorure	Sulfate de cadmium	Chlorure de cobalt	Czc (Cobalt, Zinc, Cadmium)	Sulfate de cuivre	Cop (Cuivre)	Cus (Cuivre)	Sodium arsenite	Ars (Arsenic)
LM91	S	48	12,8	32	+	128	+ (copLAB)	+	512	+
JK2-10	Xcc (A*)	32	12,8	16		48			12	
LMG930	Xee	32	12,8	16	+	128	+ (copLAB)	+	512	~
LM180	Xcc (A)	32	12,8	24	+	128	+ (copLAB)	+	192	+
LH276	Xcc (A)	32	12,8	16	+	256	+ (copLAB)	+	128	+
LM159	Xv	32	9,6	16	+	128	+ (copLAB)	+	256	+
LJ207-7	Xcc (A)	32	9,6	24	+	128	+ (copLAB)	+	192	+

Annexe III : Caractérisation *in vitro* du taux de conjugaison de pLH201.1

Matériels et méthodes

Le taux de conjugaison du plasmide entre souches peut être estimé à l'aide d'une souche donneuse qui possède pLH201.1 et d'une souche receveuse qui ne le possède pas mais qui est dotée d'un autre marqueur de résistance. Ainsi, la souche donneuse (LH201) est toujours la même et c'est l'aptitude du plasmide à être transféré par conjugaison depuis LH201 jusqu'à une bactérie receveuse qui est testée. Des mutants spontanés résistants à 100mg/l de rifamycine des souches dont on souhaite tester l'aptitude à recevoir pLH201.1 par conjugaison sont utilisés comme souches receveuses. Les deux types de bactéries (1×10^8 /mL ratio 1:1 ; spectrophotomètre 600nm puis dénombrement de dilutions en série à l'aide de l'ensemencement easySpiral) sont mis en co-culture à concentration égale pendant 4h à 28°C sur un filtre de nitrocellulose stérile placé à la surface d'une boîte de Pétri contenant un milieu LPGA additionné de 50mg/L de CuSO₄, 5 H₂O (pour quelques souches commensales ne poussant pas sur ce milieu, le LPGA a été utilisé). Nous estimons ensuite les effectifs bactériens des trois phénotypes possibles (CuR⁺, RifR⁺ et la combinaison de RifR⁺ et CuR⁺) par comptage des colonies sur milieux sélectifs (LPGA CuSO₄, 5H₂O 782,5mg/L (500ppm) ; LPGA Rifamycine 100mg/L ; LPGA CuSO₄, 5H₂O 782,5mg/L + Rifamycine 100mg/L, respectivement) après dilution en série et étalement des suspensions à l'aide de l'ensemencement easySpiral. Nous vérifions par génotypage des transconjugants à l'aide de 14 microsatellites (pour Xcc) (Verniere *et al.* 2014) ou par séquençage du gène de ménage *atpD* (pour les autres *Xanthomonas*) (Bui Thi Ngoc *et al.* 2010) que les colonies se développant sur milieu additionné de cuivre et de rifamycine ne sont pas des mutants spontanés de la souche donneuse devenus résistants à la rifamycine.

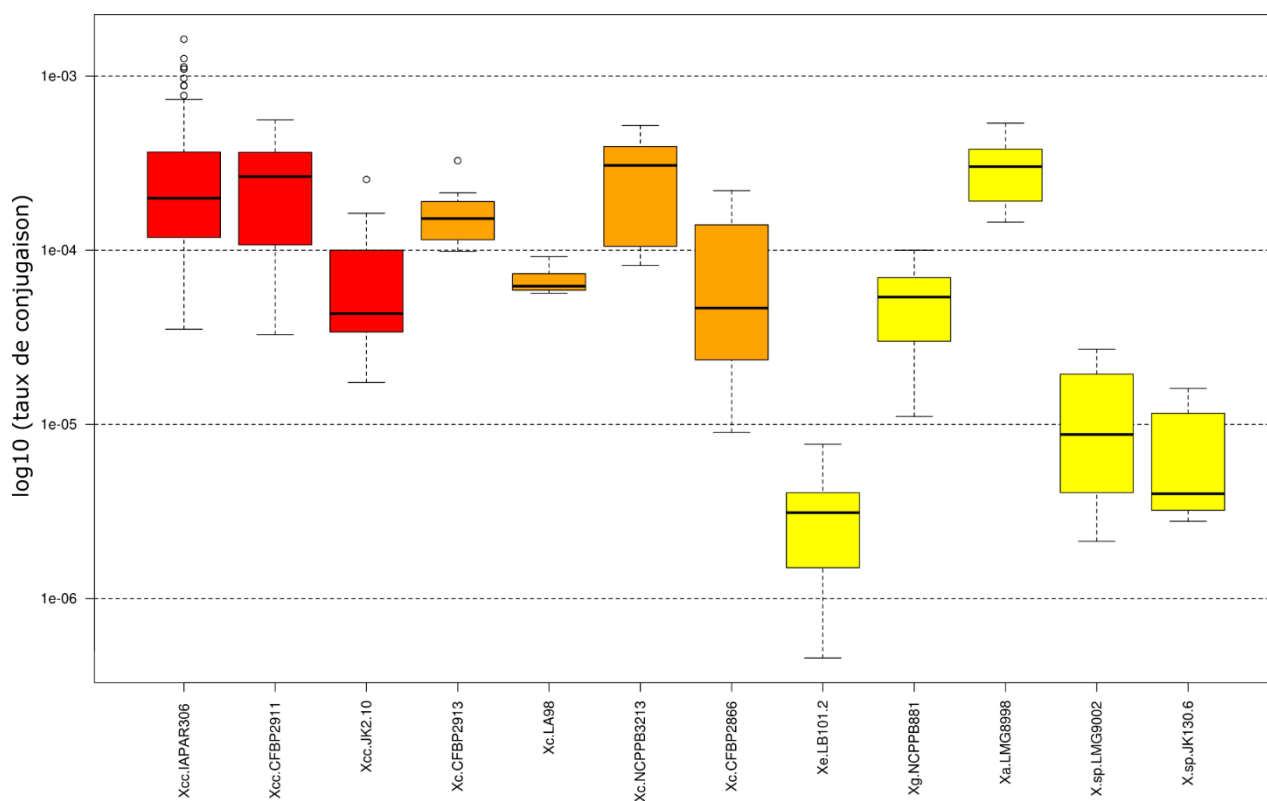
La méthode impose la sensibilité au cuivre des souches receveuses car ce phénotype est utilisé comme marqueur. Avant la conjugaison, le phénotype de résistance au cuivre des souches receveuses est donc testé. Un assez grand nombre de souches de la famille des Xanthomonadaceae autres que *Xanthomonas* présentaient déjà un phénotype résistant au cuivre (formation de colonies sur milieu supplémenté de 500 ppm de CuSO₄), suggérant que la résistance au cuivre est relativement répandue chez les souches commensales de cette famille. Cependant, les PCR *copLAB* effectuées sur ces dernières se sont révélées négatives, suggérant que les gènes leur conférant le phénotype résistant sont composés soit de *copLAB* dont la séquence nucléotidique diverge de *copLAB*-pLH201, soit de gènes différents.

Souches bactériennes dont l'aptitude à intégrer le plasmide pLH201.1 par conjugaison est testé. La souche donneuse est toujours Xcc LH201. Les nombres de répliquas pour lesquels la conjugaison a été mise en évidence ou non sont donnés en avant-dernière et dernière colonne, respectivement.

Souche	Classe	Famille	Genre et espèce	Pathovar	isolé de <i>Citrus</i>	Conjugaisons positives	Conjugaisons négatives
CFBP2866	Gammaproteobacteria	Xanthomonadaceae	<i>X. citri</i>	<i>aurantifolii</i>	oui	4	0
CFBP2905	Gammaproteobacteria	Xanthomonadaceae	<i>X. citri</i>	<i>aurantifolii</i>	oui	0	4
NCPPB3213	Gammaproteobacteria	Xanthomonadaceae	<i>X. citri</i>	<i>bilvae</i>	oui	3	1
IAPAR306	Gammaproteobacteria	Xanthomonadaceae	<i>X. citri</i>	<i>citri</i>	oui	36	0
CFBP2911	Gammaproteobacteria	Xanthomonadaceae	<i>X. citri</i>	<i>citri</i>	oui	5	0
JJ238-24	Gammaproteobacteria	Xanthomonadaceae	<i>X. citri</i>	<i>citri</i>	oui	0	5
JK2-10	Gammaproteobacteria	Xanthomonadaceae	<i>X. citri</i>	<i>citri</i>	oui	4	1
NCPPB3562	Gammaproteobacteria	Xanthomonadaceae	<i>X. citri</i>	<i>citri</i>	oui	0	5
X2003-3218	Gammaproteobacteria	Xanthomonadaceae	<i>X. citri</i>	<i>citri</i>	oui	0	5
CFBP2913	Gammaproteobacteria	Xanthomonadaceae	<i>X. citri</i>	<i>anarcadii</i>	non	2	0
LA98	Gammaproteobacteria	Xanthomonadaceae	<i>X. citri</i>	<i>anarcadii</i>	non	2	0
LMG7428	Gammaproteobacteria	Xanthomonadaceae	<i>X. citri</i>	<i>malvacearum</i>	non	0	3
JP757	Gammaproteobacteria	Xanthomonadaceae	<i>X. citri</i>	<i>mangiferaeindicae</i>	non	0	4
CFBP2940	Gammaproteobacteria	Xanthomonadaceae	<i>X. citri</i>	<i>mangiferaeindicae</i>	non	0	4
JB003-10	Gammaproteobacteria	Xanthomonadaceae	<i>X. citri</i>	<i>mangiferaeindicae</i>	non	0	4
LG081-27	Gammaproteobacteria	Xanthomonadaceae	<i>X. citri</i>	<i>mangiferaeindicae</i>	non	0	4
CFBP2916	Gammaproteobacteria	Xanthomonadaceae	<i>X. citri</i>	<i>mangiferaeindicae</i>	non	0	3
LM006	Gammaproteobacteria	Xanthomonadaceae	<i>X. citri</i>	<i>mangiferaeindicae</i>	non	0	3
NCPPB881	Gammaproteobacteria	Xanthomonadaceae	<i>X. cynarae</i>	<i>gardneri</i>	non	2	0

Suite du tableau précédent.

Souche	Classe	Famille	Genre et espèce	Pathovar	isolé de <i>Citrus</i>	Conjugaisons positives	Conjugaisons négatives
CFBP6369	Gammaproteobacteria	Xanthomonadaceae	<i>X. euvesicatoria</i>	<i>allii</i>	non	0	3
CFBP3138	Gammaproteobacteria	Xanthomonadaceae	<i>X. euvesicatoria</i>	<i>citrumelo</i>	oui	0	5
CFBP2910	Gammaproteobacteria	Xanthomonadaceae	<i>X. euvesicatoria</i>	<i>citrumelo</i>	oui	0	5
LB101-2	Gammaproteobacteria	Xanthomonadaceae	<i>X. euvesicatoria</i>	<i>perforans</i>	non	2	1
LMG876	Gammaproteobacteria	Xanthomonadaceae	<i>X. translucens</i>	<i>translucens</i>	non	0	4
LC160	Gammaproteobacteria	Xanthomonadaceae	<i>X. vesicatoria</i>		non	0	4
LMG8998	Gammaproteobacteria	Xanthomonadaceae	<i>X. arboricola</i>		oui (saprophyte)	2	0
JK130.6	Gammaproteobacteria	Xanthomonadaceae	<i>X. sp.</i>		oui (saprophyte)	4	0
LMG9002	Gammaproteobacteria	Xanthomonadaceae	<i>X. sp.</i>		oui (saprophyte)	2	0
DSM18010	Gammaproteobacteria	Xanthomonadaceae	<i>Arenimonas composti</i>		non	0	4
DSM18204	Gammaproteobacteria	Xanthomonadaceae	<i>Pseudoxanthomonas yeongjuensis</i>		non	0	3
DSM24757	Gammaproteobacteria	Xanthomonadaceae	<i>Stenotrophomonas ginsengisoli</i>		non	0	5
LMG23369	Gammaproteobacteria	Xanthomonadaceae	<i>Stenotrophomonas koreensis</i>		non	0	5
DSM6220	Gammaproteobacteria	Rhodanobacteraceae	<i>Frateuria aurantia</i>		non	0	3
DSM14263	Gammaproteobacteria	Rhodanobacteraceae	<i>Fulvimonas soli</i>		non	0	3
CFBP1787	Gammaproteobacteria	Pseudomonadaceae	<i>Pseudomonas syringae</i>	<i>syringae</i>	non	0	2
CFBP6699	Gammaproteobacteria	Enterobacteriaceae	<i>Pectobacterium carotovorum</i>		non	0	3
LMG2431	Gammaproteobacteria	Enterobacteriaceae	<i>Pectobacterium carotovorum</i> subsp. <i>carotovorum</i>		non	0	2
LMG2129	Betaproteobacteria	Burkholderiaceae	<i>Burkholderia andropogonis</i>		non	0	2



Distribution du log des taux de conjugaison pour les souches bactériennes dans lesquelles le plasmide pLH201.1 a pu être transféré par conjugaison. Les souches sont classées selon leur proximité taxonomique avec Xcc. Xcc : *X. citri* pv. *citri* (rouge) ; Xc : *X. citri* autres pathovars (orange) ; Xee : *X. euvesicatoria* pv. *euvesicatoria* , Xcg : *X. cynarae* pv. *gardneri*, Xa : *X. arboricola* et X.sp : *Xanthomonas* sp. (jaune).

Annexe IV : Informations supplémentaires publication 1 *Adaptation of genetically monomorphic bacteria: evolution of copper resistance through multiple horizontal gene transfers of complex and versatile mobile genetic elements*

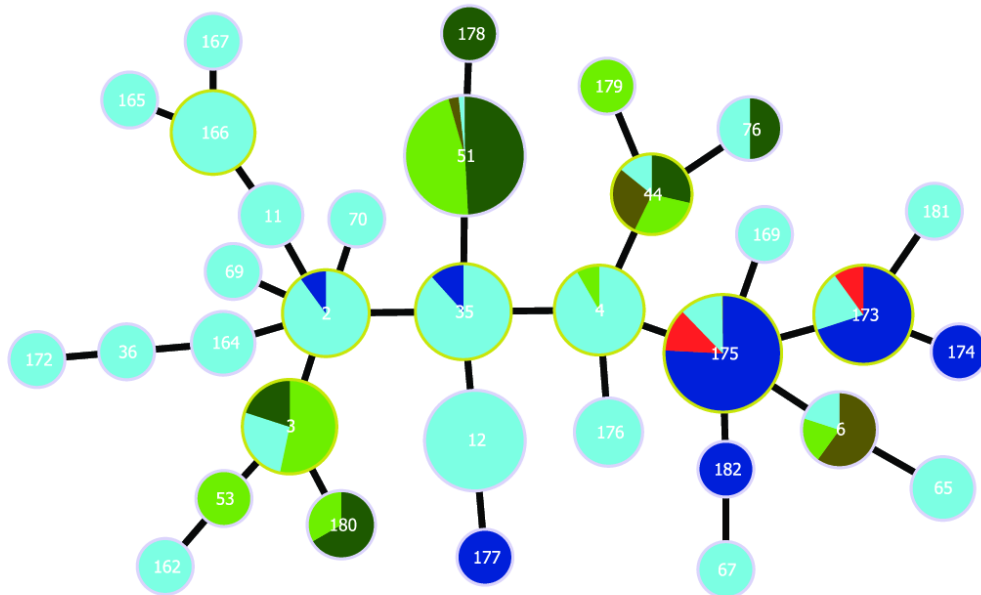


Fig. S1. Categorical minimum spanning tree of DAPC1 strains from Argentina, Martinique and Réunion, which differ in their susceptibility to copper (268 strains – 34 haplotypes), based on minisatellite data. These strains were organized as a single clonal complex (i.e. a network of haplotypes linked by single-locus variations). Dot diameter is representative of the number of strains per haplotype. Colour indicates the strain origin and copper phenotype: light green = copper-susceptible Argentina; dark green = copper-resistant Argentina; khaki = copper-resistant non-*copLAB* Argentina; red = copper-resistant Martinique; light blue = copper-susceptible Réunion; dark blue = copper-resistant Réunion.

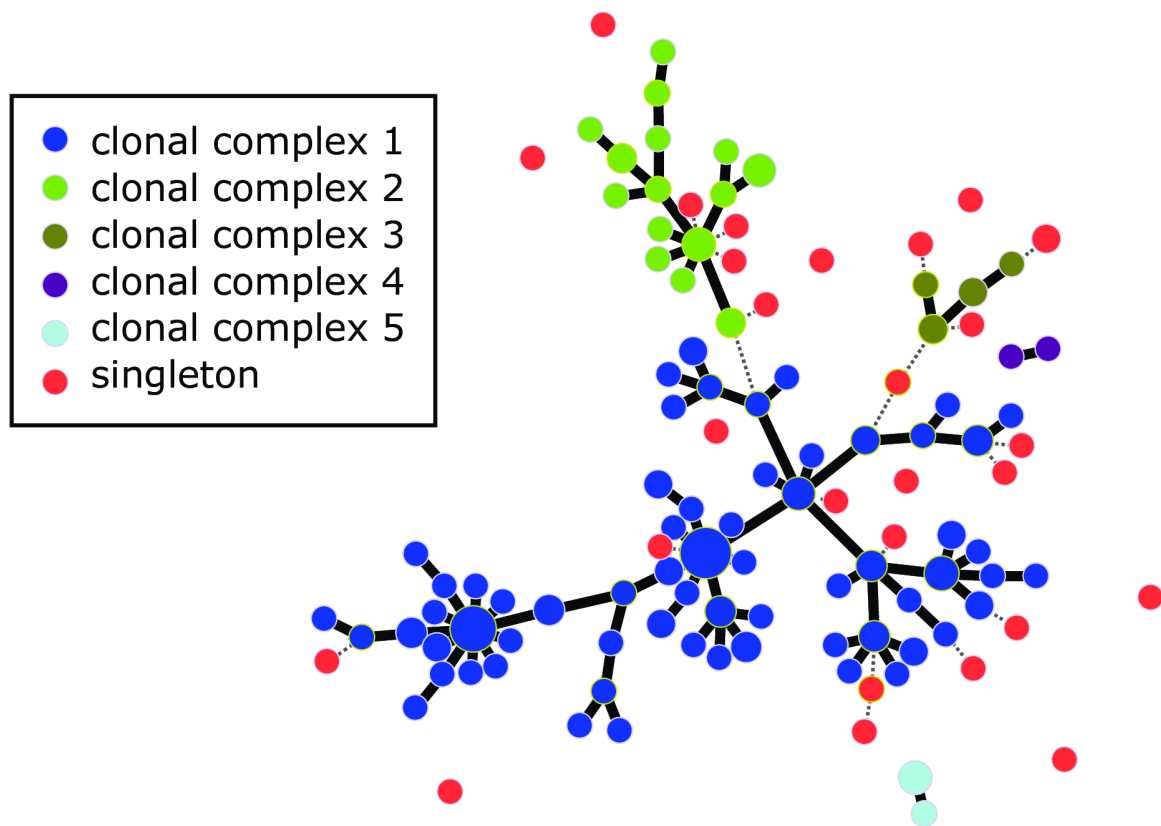
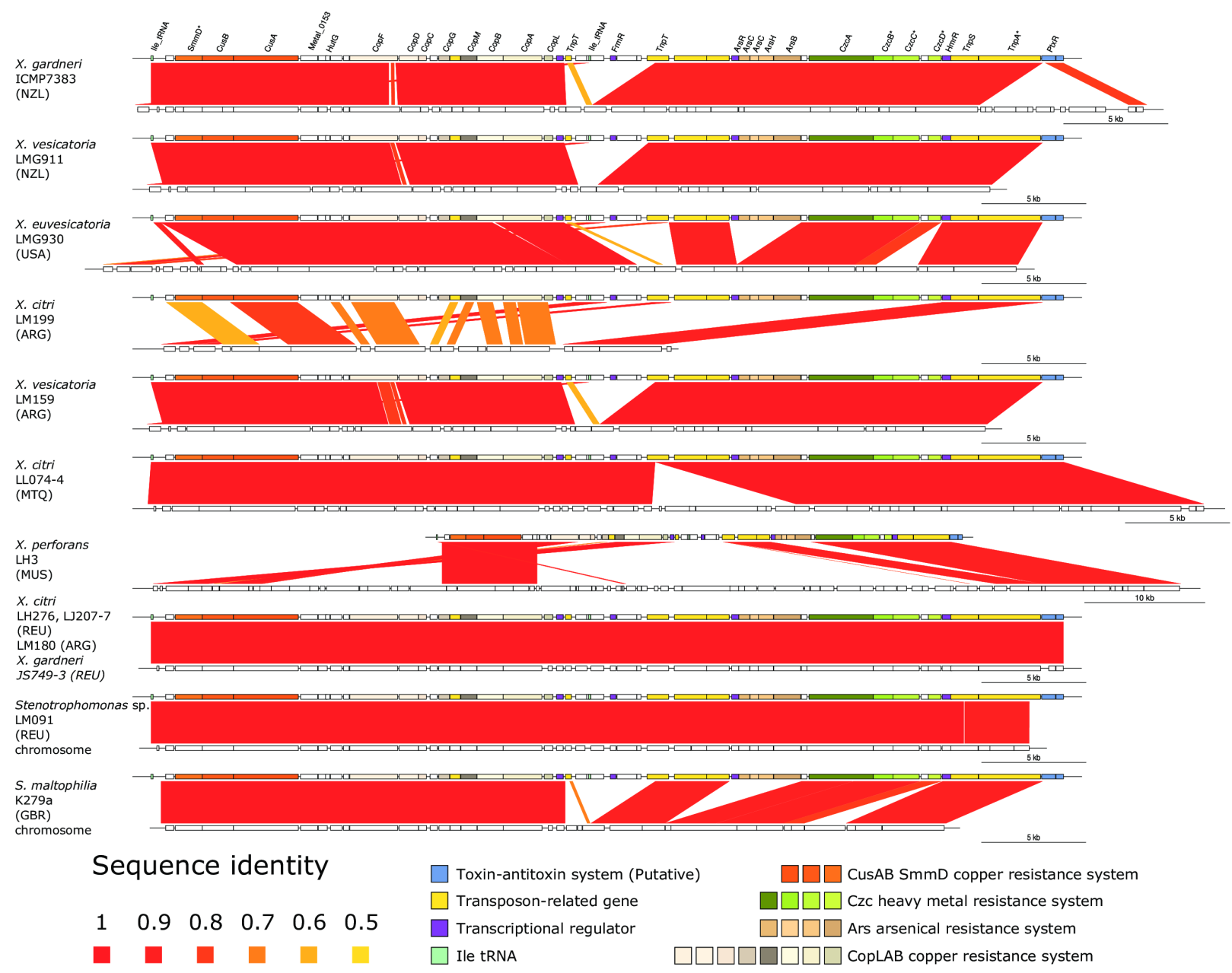


Fig. S2. Categorical minimum spanning tree of copper-resistant *Xanthomonas citri* pv. *citri* strains from Réunion based on microsatellite data. These strains were organized as five clonal complexes (i.e. networks of haplotypes linked by single-locus variations) and 28 singletons (i.e. haplotypes with no single-locus variants). Dot diameter represents the number of strains per haplotype. Single and double-locus variations were represented as solid and dotted lines joining haplotypes, respectively.



Page précédente : Fig. S3. Alignments between TnpLH201.1 and sequences carrying TnpLH201.1 homologues from strains sequenced in this study and *S. maltophilia* K279a. All sequences are plasmids, unless specified otherwise. On TnpLH201.1, blocks represent curated gene prediction (Mage), whereas on other genomes the blocks represent uncurated Prodigal gene prediction. The pLH201.1 genes are categorized into several groups according to the caption. The comparison zone shows homologous nucleotide sequences. Colour varies according to genetic distance between homologous blocks defined by a MAUVE alignment.

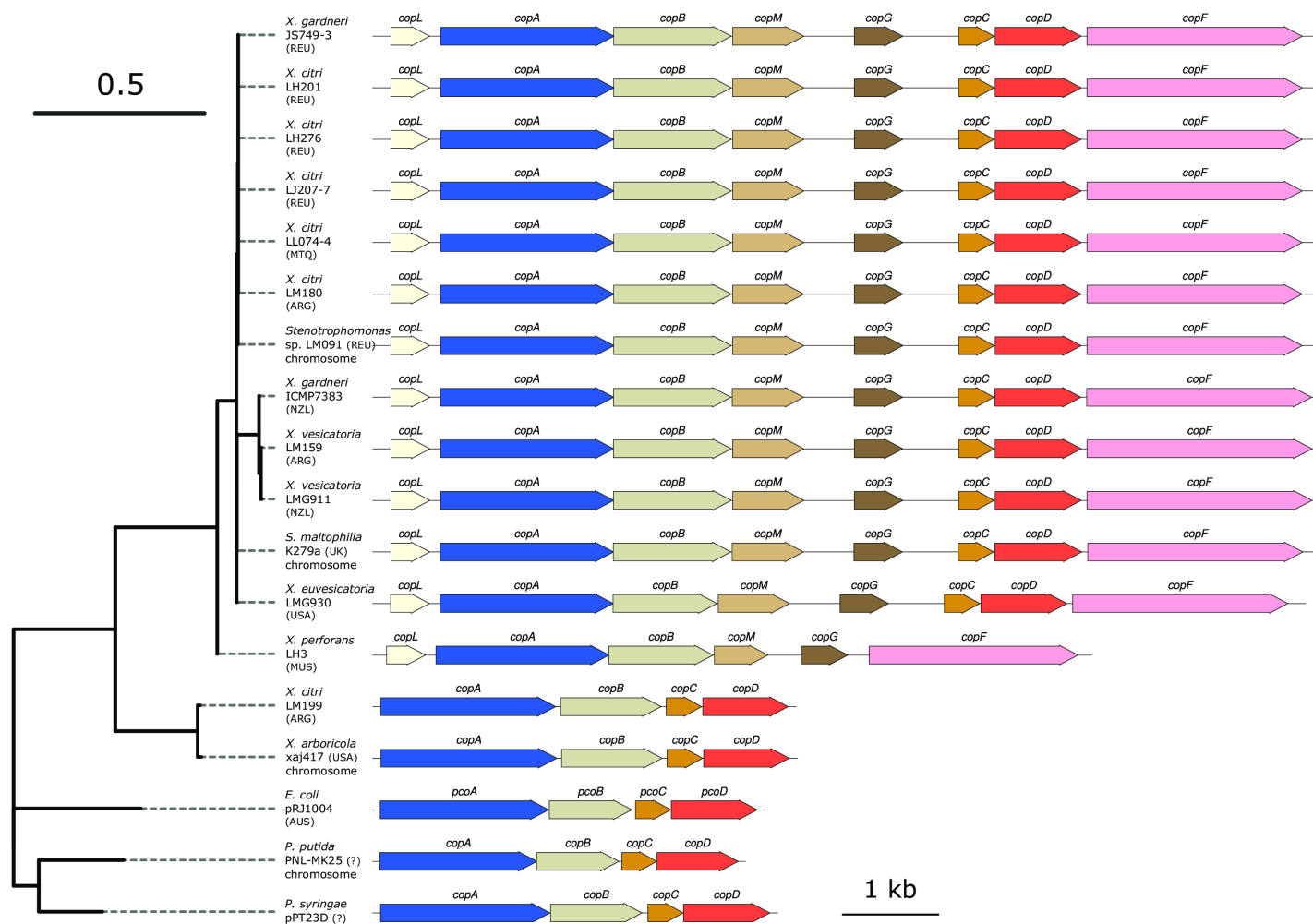


Fig. S4. Distance tree of the copper-resistance region of all the sequenced strains as well as some sequences coding for known copper-resistance systems extracted from public database along with graphical representations of the copper-resistance genes organisation. The alignment used for the tree computation was obtained after the concatenation of alignments of each gene.

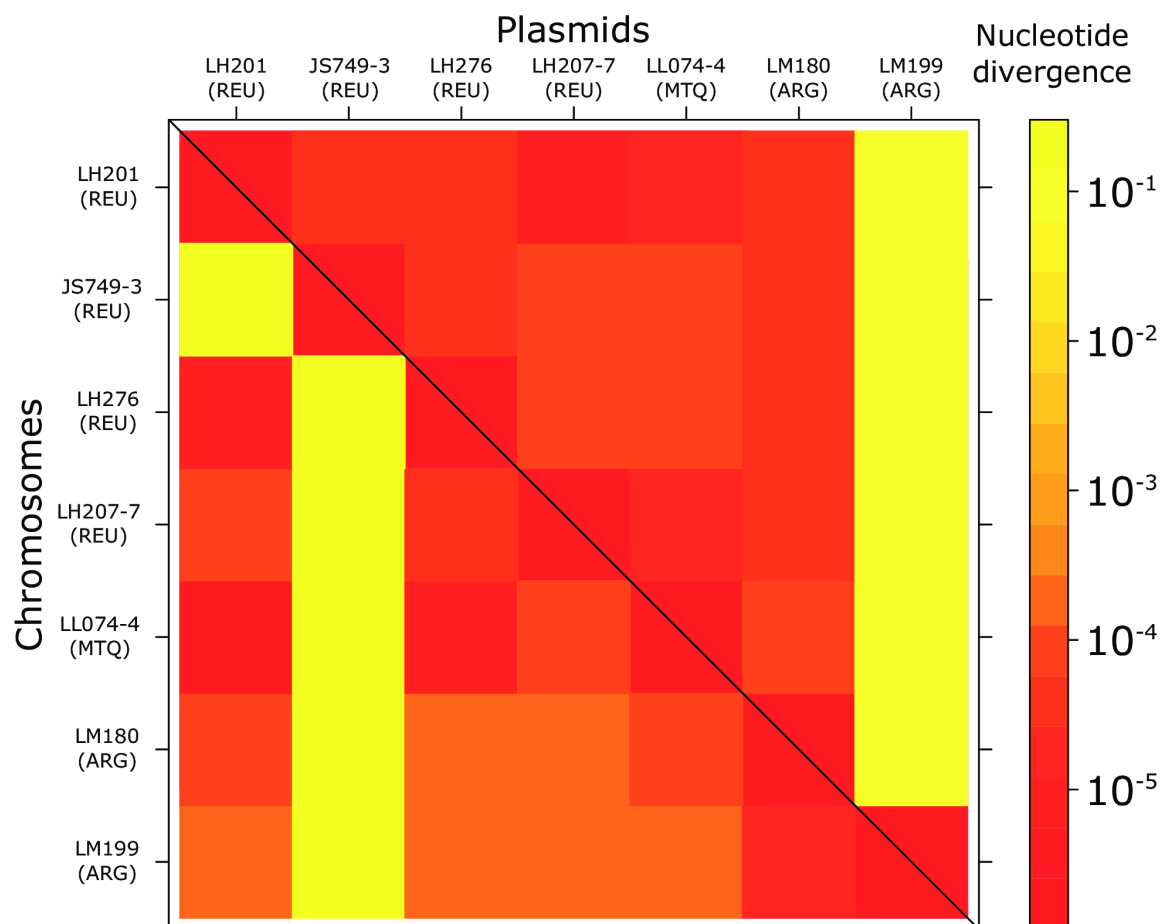
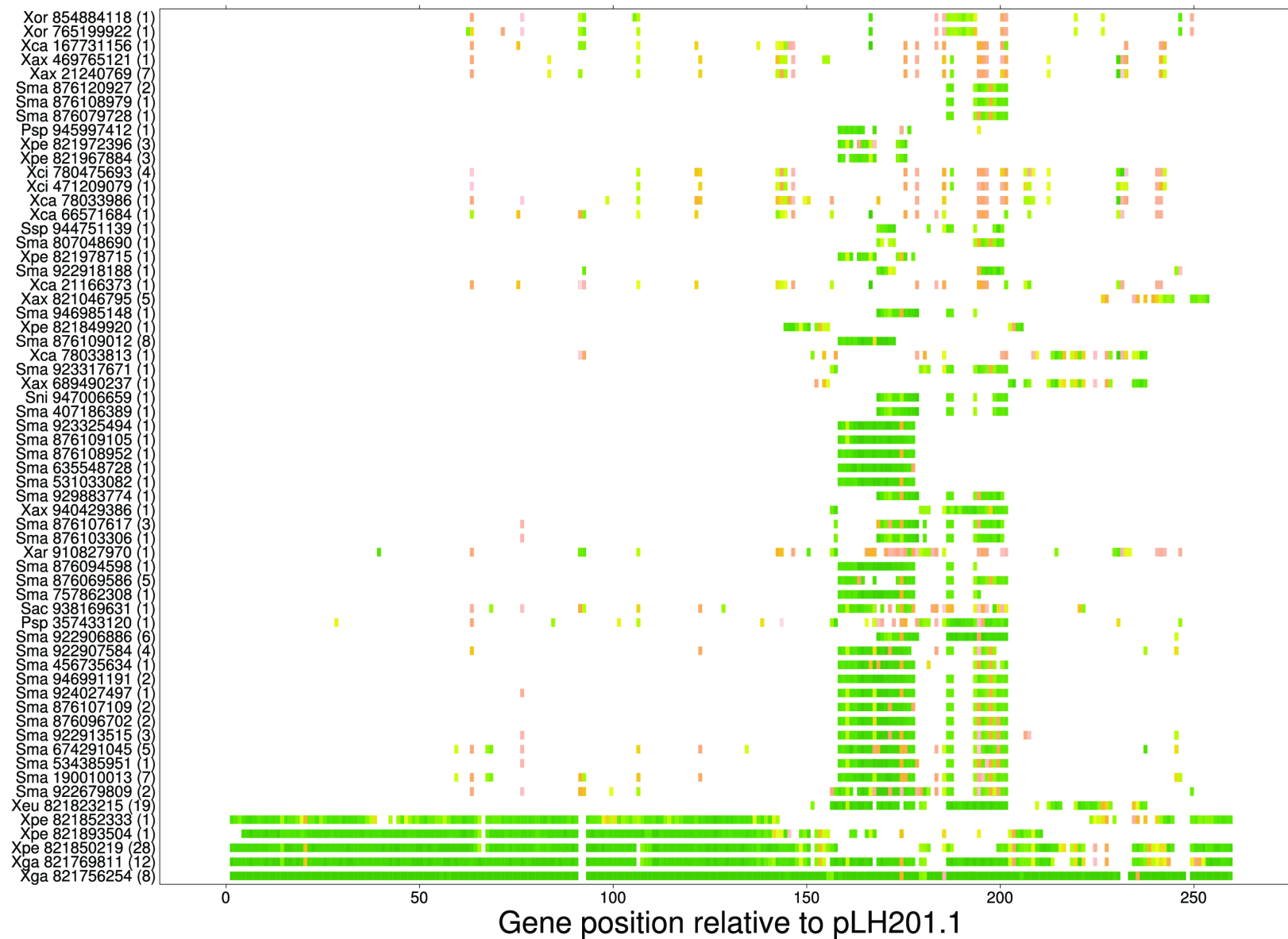


Fig. S5. Heatmap representation of the genetic divergence between six sequenced strains. Divergence values (proportions of variable nucleotides between two sequences) were obtained from the comparison of all homologous regions. Whereas the lower triangle represents the divergence between the chromosomes, the upper triangle represents divergence between the plasmids.



Page précédente : Fig. S6. Clusters of sequences with homologues to pLH201.1 genes. Each line represent a cluster of sequences obtained from GenBank that are homologous to at least 10 pLH201.1 genes. Squares represent homologous genes and are ordered as in the pLH201.1 sequence. Squares are coloured according to their level of nucleotide identity with pLH201.1, as indicated on the scale on the right. The names at the left of each line indicate the species of the parental sequence of the cluster, its GenBank identifier (gi) and the number of sequences in the cluster. Psp: *Pseudoxanthomonas* sp.; Sac: *Stenotrophomonas acidaminiphila*; Sma: *Stenotrophomonas maltophilia*; Sni: *Stenotrophomonas nitritireducens*; Ssp: *Streptomyces* sp.; Xar: *Xanthomonas arboricola*; Xax: *Xanthomonas axonopodis*; Xca: *Xanthomonas campestris*; Xci: *Xanthomonas citri*; Xeu: *Xanthomonas euvesicatoria*; Xga: *Xanthomonas gardneri*; Xor: *Xanthomonas oryzae*; Xpe: *Xanthomonas perforans*.

Table S1 : Size (in bp) of the contigs obtained for each of the 13 sequenced strains

LM199, Xcc, ARG	ICMP7383, Xga, NZL	LM180, Xcc, ARG	LMG911, Xve, NZL	JS749-3, Xga, REU	LMG930, Xeu, USA	LH3, Xpe, MUS	LH276, Xcc, REU	LJ207-7, Xcc, REU	LM159, Xve, ARG	LL074-4, Xcc, MTQ	LH201, Xcc, REU	LM091, Ssp, REU
5198476	5313102	5416746	5110163	5158913	5079107	4969893	5159828	5148274	5086726	5158276	5156885	4317450
197085	203385	223557	192558	211336	179490	222061	211337	213743	201433	220688	213737	
64936	61840	64978	47184	45952	167496	69764	73024	94139	62784	92709	92707	
31100	47122	49833	8743	8305	42763	45953	51192	49382	42967			
19636	9973	43340	3962	8128	31328	38746						
11194	6507	29660	3594									
7809	5401	28949	3519									
7304	4566	16267	1664									
6651	4031	14177										
5575	3644	3142										
5568	3082											
3494	2030											
2930	1760											
1653	1216											
1538												
1483												
1339												
1041												

Circularized contigs represented in bold

REU: Réunion; MTQ: Martinique; ARG: Argentina; NZL: New Zealand; USA: United States of America; MUS: Mauritius; Xcc: *Xanthomonas citri* pv. *citri* ; Xga: *Xanthomonas gardneri* ; Xve: *Xanthomonas vesicatoria* ; Xeu: *Xanthomonas euvesicatoria* ; Xpe: *Xanthomonas perforans* ; Ssp: *Stenotrophomonas* sp.

Table S2 plasmids and ICE having similar conjugative apparatus with pLH201.1

Database	database id	name	classification	size	class	family	genus	species
NCBI plasmid	NC_011838	pCAR1.2	IncP-7	200231	Gammaproteobacteria	Pseudomonadaceae	<i>Pseudomonas</i>	<i>putida</i>
NCBI plasmid	NC_012692	pAR060302	IncA/C	166530	Gammaproteobacteria	Enterobacteria	<i>Escherichia</i>	<i>coli</i>
NCBI plasmid	NC_016976	pR55	IncA/C	170810	Gammaproteobacteria	Enterobacteria	<i>Klebsiella</i>	<i>pneumoniae</i>
NCBI plasmid	NC_019153	pNDM-KN	IncA/C	162746	Gammaproteobacteria	Enterobacteria	<i>Klebsiella</i>	<i>pneumoniae</i>
NCBI plasmid	NC_009139	pYR1	IncA/C	158038	Gammaproteobacteria	Enterobacteria	<i>Yersinia</i>	<i>ruckeri</i>
NCBI plasmid	NC_022522	p1643_10	IncA/C	167779	Gammaproteobacteria	Enterobacteria	<i>Salmonella</i>	<i>enterica</i>
NCBI plasmid	NC_019069	pNDM10505	IncA/C	166744	Gammaproteobacteria	Enterobacteria	<i>Escherichia</i>	<i>coli</i>
NCBI plasmid	NC_023898	pRMH760	repA/C2	170613	Gammaproteobacteria	Enterobacteria	<i>Klebsiella</i>	<i>pneumoniae</i>
NCBI plasmid	NC_019375	pTC2	multireplicon	180184	Gammaproteobacteria	Enterobacteria	<i>Providencia</i>	<i>stuartii</i>
NCBI plasmid	NC_019158	pNDM10469	IncA/C	137813	Gammaproteobacteria	Enterobacteria	<i>Klebsiella</i>	<i>pneumoniae</i>
NCBI plasmid	NC_012885	pRA1	IncA/C	143963	Gammaproteobacteria	Aeromonadaceae	<i>Aeromonas</i>	<i>hydrophila</i>
NCBI plasmid	NC_014124	pDK1		128921	Gammaproteobacteria	Pseudomonadaceae	<i>Pseudomonas</i>	<i>putida</i>
NCBI plasmid	NC_019045	pNDM102337	IncA/C	165974	Gammaproteobacteria	Enterobacteria	<i>Escherichia</i>	<i>coli</i>
NCBI plasmid	NC_019380	pR148	IncA/C	165906	Gammaproteobacteria	Aeromonadaceae	<i>Aeromonas</i>	<i>hydrophila</i>
NCBI plasmid	NC_022995	pM7012		582142	Betaproteobacteria	Burkholderiaceae	<i>Burkholderia</i>	<i>sp.</i>
NCBI plasmid	NC_023908	pKP1-NDM-1	IncA/C	137538	Gammaproteobacteria	Enterobacteria	<i>Klebsiella</i>	<i>pneumoniae</i>
NCBI plasmid	NC_019121	pSH111_166		165791	Gammaproteobacteria	Enterobacteria	<i>Salmonella</i>	<i>enterica</i>
NCBI plasmid	NC_018994	pNDM-1_Dok01	IncA/C	195560	Gammaproteobacteria	Enterobacteria	<i>Escherichia</i>	<i>coli</i>
NCBI plasmid	NC_019065	pPG010208	IncA/C	135803	Gammaproteobacteria	Enterobacteria	<i>Escherichia</i>	<i>coli</i>
NCBI plasmid	NC_019107	pSD_174		173673	Gammaproteobacteria	Enterobacteria	<i>Salmonella</i>	<i>enterica</i>
NCBI plasmid	NC_012693	pAM04528	IncA/C	158213	Gammaproteobacteria	Enterobacteria	<i>Salmonella</i>	<i>enterica</i>
NCBI plasmid	NC_023291	pVCR94deltaX	IncA/C	120572	Gammaproteobacteria	Vibrionaceae	<i>Vibrio</i>	<i>cholerae</i>
NCBI plasmid	NC_008612	pP99-018	Putative IncI	150157	Gammaproteobacteria	Vibrionaceae	<i>Photobacterium</i>	<i>damselae</i>
NCBI plasmid	NC_004444	pCAR1	IncP-7	199035	Gammaproteobacteria	Pseudomonadaceae	<i>Pseudomonas</i>	<i>resinovorans</i>
NCBI plasmid	NC_022652	pIMP-PH114	IncA/C2	151885	Gammaproteobacteria	Enterobacteria	<i>Klebsiella</i>	<i>pneumoniae</i>
NCBI plasmid	NC_008613	pP91278	Putative IncI	131520	Gammaproteobacteria	Vibrionaceae	<i>Photobacterium</i>	<i>damselae</i>
NCBI plasmid	NC_003905	Rts1	IncT(1972)	217182	Gammaproteobacteria	Enterobacteria	<i>Proteus</i>	<i>vulgaris</i>
NCBI plasmid	NC_025127	pAQU2	IncA/C	160406	Gammaproteobacteria	Vibrionaceae	<i>Vibrio</i>	<i>sp.</i>
NCBI plasmid	NC_016983	pAQU1	IncA/C	204052	Gammaproteobacteria	Vibrionaceae	<i>Photobacterium</i>	<i>damselae</i>
NCBI plasmid	NC_019360	pNDM-CIT	IncHI1	288920	Gammaproteobacteria	Enterobacteria	<i>Citrobacter</i>	<i>freundii</i>
NCBI plasmid	NC_023289	pEQ1	IncHI1	239151	Gammaproteobacteria	Enterobacteria	<i>Escherichia</i>	<i>coli</i>
NCBI plasmid	NC_012555	pEC-IMP	IncHI2	318782	Gammaproteobacteria	Enterobacteria	<i>Enterobacter</i>	<i>cloacae</i>
NCBI plasmid	NC_019116	pSH163_135		135168	Gammaproteobacteria	Enterobacteria	<i>Salmonella</i>	<i>enterica</i>
NCBI plasmid	NC_022377	pSCEC2	IncA/C	135615	Gammaproteobacteria	Enterobacteria	<i>Escherichia</i>	<i>coli</i>
NCBI plasmid	NC_002305	R27	IncHI	180461	Gammaproteobacteria	Enterobacteria	<i>Salmonella</i>	<i>typhi</i>
NCBI plasmid	NC_023277	pEQ2	IncHI1	287616	Gammaproteobacteria	Enterobacteria	<i>Escherichia</i>	<i>coli</i>
NCBI plasmid	NC_016974	pMR0211	IncA/C	178277	Gammaproteobacteria	Enterobacteria	<i>Providencia</i>	<i>stuartii</i>
NCBI plasmid	NC_003384	pHCM1	IncHI1	218160	Gammaproteobacteria	Enterobacteria	<i>Salmonella</i>	<i>enterica</i>
NCBI plasmid	NC_024983	pSTm-A54650		309406	Gammaproteobacteria	Enterobacteria	<i>Salmonella</i>	<i>enterica</i>
NCBI plasmid	NC_023911	pKpNDM1	IncH	277682	Gammaproteobacteria	Enterobacteria	<i>Raoultella</i>	<i>planticola</i>
NCBI plasmid	NC_005211	R478	IncH	274762	Gammaproteobacteria	Enterobacteria	<i>Serratia</i>	<i>marcescens</i>
NCBI plasmid	NC_010870	pK29	IncHI2	269674	Gammaproteobacteria	Enterobacteria	<i>Klebsiella</i>	<i>pneumoniae</i>
NCBI plasmid	NC_021667	LS6	IncA/C	171925	Gammaproteobacteria	Enterobacteria	<i>Klebsiella</i>	<i>pneumoniae</i>
NCBI plasmid	NC_019118	pSH696_135		135423	Gammaproteobacteria	Enterobacteria	<i>Salmonella</i>	<i>enterica</i>
NCBI plasmid	NC_022589	pPrY2001	IncFIK	113295	Gammaproteobacteria	Enterobacteria	<i>Providencia</i>	<i>rettgeri</i>
NCBI plasmid	NC_012556	pEC-IMPQ	IncHI2	324503	Gammaproteobacteria	Enterobacteria	<i>Enterobacter</i>	<i>cloacae</i>
NCBI plasmid	NC_016980	pNDM-MAR	IncH	267242	Gammaproteobacteria	Enterobacteria	<i>Klebsiella</i>	<i>pneumoniae</i>
NCBI plasmid	NC_019066	pAPEC1990_61	IncA/C	161081	Gammaproteobacteria	Enterobacteria	<i>Escherichia</i>	<i>coli</i>
NCBI plasmid	NC_009981	pMAK1		208409	Gammaproteobacteria	Enterobacteria	<i>Salmonella</i>	<i>enterica</i>
NCBI plasmid	NC_019114	pSH111_227		227608	Gammaproteobacteria	Enterobacteria	<i>Salmonella</i>	<i>enterica</i>
NCBI plasmid	NC_019310	pEM65		65150	Gammaproteobacteria	Enterobacteria	<i>Erwinia</i>	<i>amylovora</i>
NCBI plasmid	NC_022372	pYT3	IncFIB	121723	Gammaproteobacteria	Enterobacteria	<i>Salmonella</i>	<i>enterica</i>
ICEberg database	22	ICEVchInd5	SXT/R391		Gammaproteobacteria	Vibrionaceae	<i>Vibrio</i>	<i>cholerae</i>
ICEberg database	16	ICEVchBan9	SXT/R391		Gammaproteobacteria	Vibrionaceae	<i>Vibrio</i>	<i>cholerae</i>
ICEberg database	38	ICEPdaSpa1	SXT/R391		Gammaproteobacteria	Vibrionaceae	<i>Photobacterium</i>	<i>damselae</i>
ICEberg database	23	ICEVchMex1	SXT/R391		Gammaproteobacteria	Vibrionaceae	<i>Vibrio</i>	<i>cholerae</i>
ICEberg database	115	ICESe3	Unclassed		Gammaproteobacteria	Enterobacteria	<i>Salmonella</i>	<i>enterica</i>
ICEberg database	375	ICEPm1	Unclassed		Gammaproteobacteria	Enterobacteria	<i>Proteus</i>	<i>mirabilis</i>
ICEberg database	44	SXT(MO10)	SXT/R391		Gammaproteobacteria	Vibrionaceae	<i>Vibrio</i>	<i>cholerae</i>
ICEberg database	40	ICEPmiUSA1	SXT/R391		Gammaproteobacteria	Enterobacteria	<i>Proteus</i>	<i>mirabilis</i>
ICEberg database	21	ICEVchInd4	SXT/R391		Gammaproteobacteria	Vibrionaceae	<i>Vibrio</i>	<i>cholerae</i>
ICEberg database	43	R391	SXT/R391		Gammaproteobacteria	Enterobacteria	<i>Providencia</i>	<i>rettgeri</i>
ICEberg database	37	ICEPalBan1	SXT/R391		Gammaproteobacteria	Enterobacteria	<i>Providencia</i>	<i>alcalifaciens</i>
ICEberg database	36	ICEVflInd1	SXT/R391		Gammaproteobacteria	Vibrionaceae	<i>Vibrio</i>	<i>fluvialis</i>
ICEberg database	12	ICEVchBan5	SXT/R391		Gammaproteobacteria	Vibrionaceae	<i>Vibrio</i>	<i>cholerae</i>
ICEberg database	39	ICESpuPO1	SXT/R391		Gammaproteobacteria	Shewanellaceae	<i>Shewanella</i>	<i>putrefaciens</i>
ICEberg database	114	ICESb2	SPI-7		Gammaproteobacteria	Enterobacteria	<i>Salmonella</i>	<i>bongori</i>
ICEberg database	437	ICEKpnHS11286-2	Unclassed		Gammaproteobacteria	Enterobacteria	<i>Klebsiella</i>	<i>pneumoniae</i>

Table S3. Primers and PCR conditions used for amplification of the copABCD system of *Xanthomonas citri* pv. *citri*.

Step	Temperature	Time
initial denaturation	95°C	5min
denaturation	95°C	30sec.
annealing	cop A copB copD 62°C	30sec.
	cop C 58°C	
extension	72°C	45sec.
final extension	72°C	10min
hold	12°C	∞

x30cycles

The GoTaq Flexi polymerase (Promega) was used as recommended by the manufacturer.

Primer	Sequence	Amplicon size (pb)
copA_F	GCCGTTTCGCCATAGTTCAATC	171
copA_R	CGGTACTGACCTACGCAATGCTC	
copB_F	TCAACACGCTCGGATTCGTCT	225
copB_R	ACTGCTGCTCACCAATCGTT	
copC_F	TACTTCACACTAAACGAGATG	300
copC_R	ACTTGTGGTTTCCTCGCCTGT	
copD_F	CGACACGGATCACCCACGTCA	282
copD_R	TCTCCATCCGTCTCGCGCTCT	

Table S4: Genetic divergence over the homologous sequence tracks defined in Fig. 2

Genetic divergence (proportion of variable nucleotides between two sequences) between plasmids						
	LH201, Xcc, REU	JS749-3, Xga, REU	LH276, Xcc, REU	LJ207-7, Xcc, REU	LL074-4, Xcc, MTQ	LM180, Xcc, ARG
JS749-3, Xga, REU	5.2e-05					
LH276, Xcc, REU	5.6e-05	5.2e-05				
LJ207-7, Xcc, REU	9.3e-06	6.1e-05	6.6e-05			
LL074-4, Xcc, MTQ	1.4e-05	8.0e-05	7.1e-05	2.3e-05		
LM180, Xcc, ARG	5.2e-05	4.7e-05	5.2e-05	4.2e-05	6.6e-05	
LM199, Xcc, ARG	0.099	0.099	0.099	0.099	0.099	0.092
Genetic divergence (proportion of variable nucleotides between two sequences) between chromosomes						
	LH201, Xcc, REU	JS749-3, Xga, REU	LH276, Xcc, REU	LJ207-7, Xcc, REU	LL074-4, Xcc, MTQ	LM180, Xcc, ARG
JS749-3, Xga, REU	0.14					
LH276, Xcc, REU	6.0e-06	0.14				
LJ207-7, Xcc, REU	5.9e-05	0.14	5.5e-05			
LL074-4, Xcc, MTQ	2.5e-06	0.14	8.3e-06	6.7e-05		
LM180, Xcc, ARG	1.1e-04	0.14	1.4e-04	2.0e-04	1.1e-04	
LM199, Xcc, ARG	2.0e-04	0.14	2.1e-04	2.3e-04	2.1e-04	2.0e-05

REU: Réunion; MTQ: Martinique; ARG: Argentina; Xcc: *Xanthomonas citri* pv. *citri*; Xga: *Xanthomonas gardneri*

Annexe V : Informations supplémentaires publication 2 *Population-genomic insights into the evolution of the crop pathogenic bacterium Xanthomonas citri pv. citri*

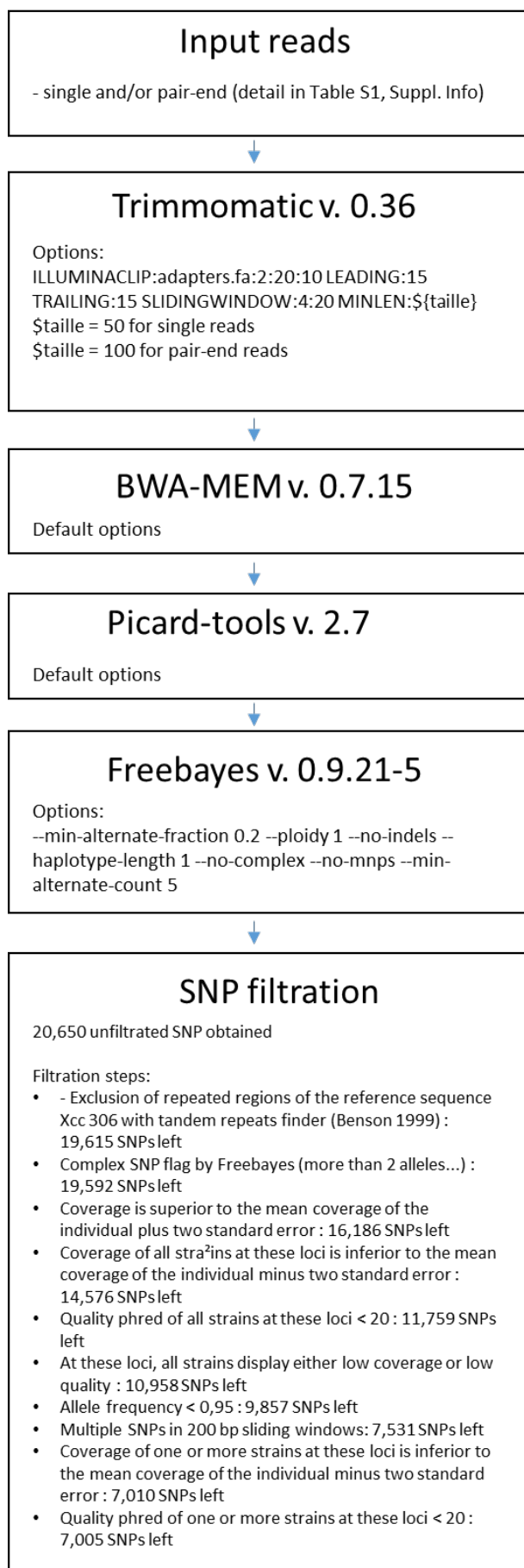


Fig. S1. Schematics of the bioinformatics pipeline steps used to obtain SNP from Illumina sequencing data.

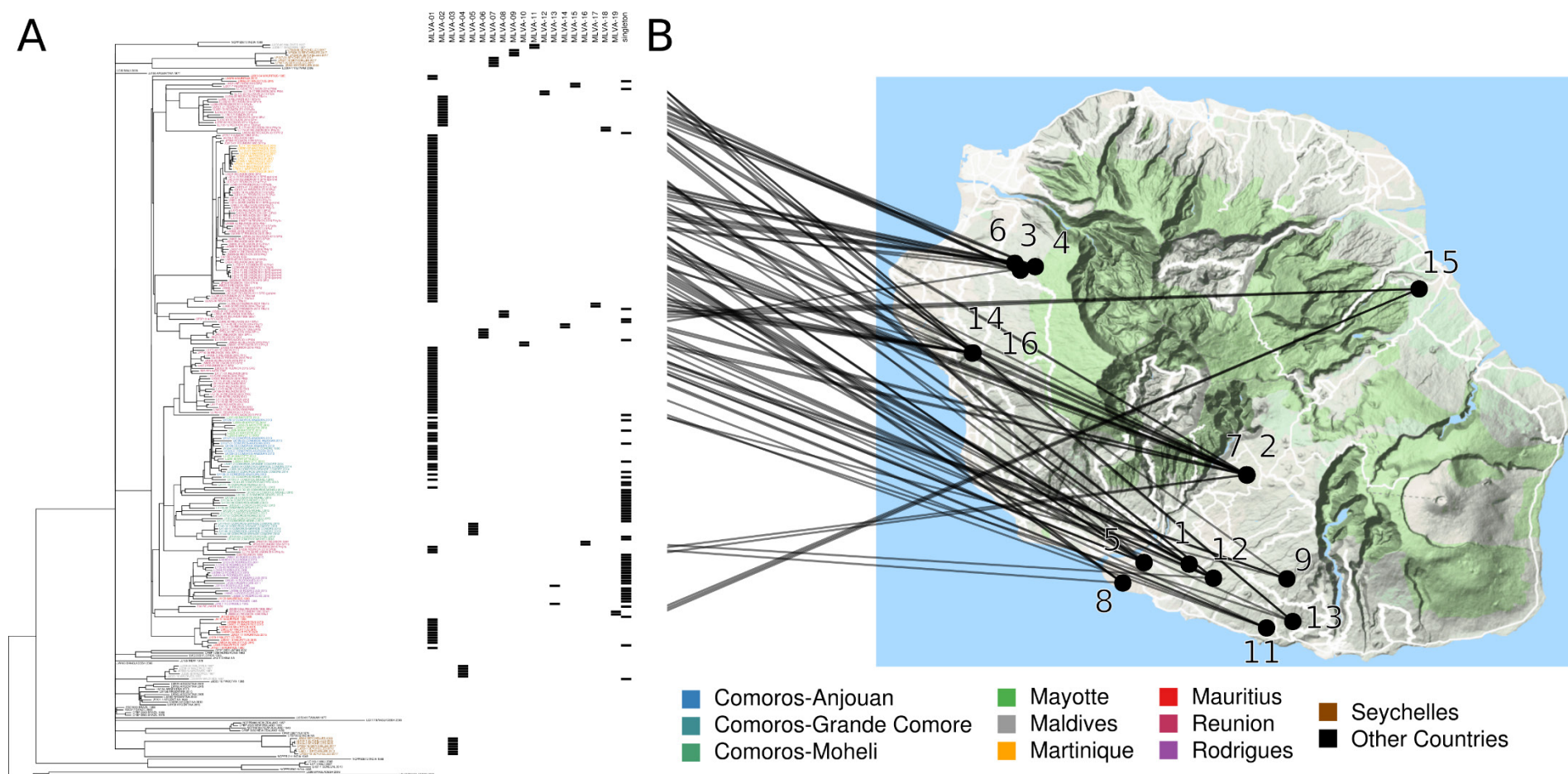


Fig. S2. (A) Maximum likelihood tree of the global dataset. Tip labels of the strains originating from the Indian Ocean are colorized according to their geographic origin according to the legend. Node annotated with the * symbol corresponds to the SWIO clade root. For each strain, the MLVA-14 typing scheme group membership is indicated with a black rectangle (B) Reunion island map indicating strains sampling location.

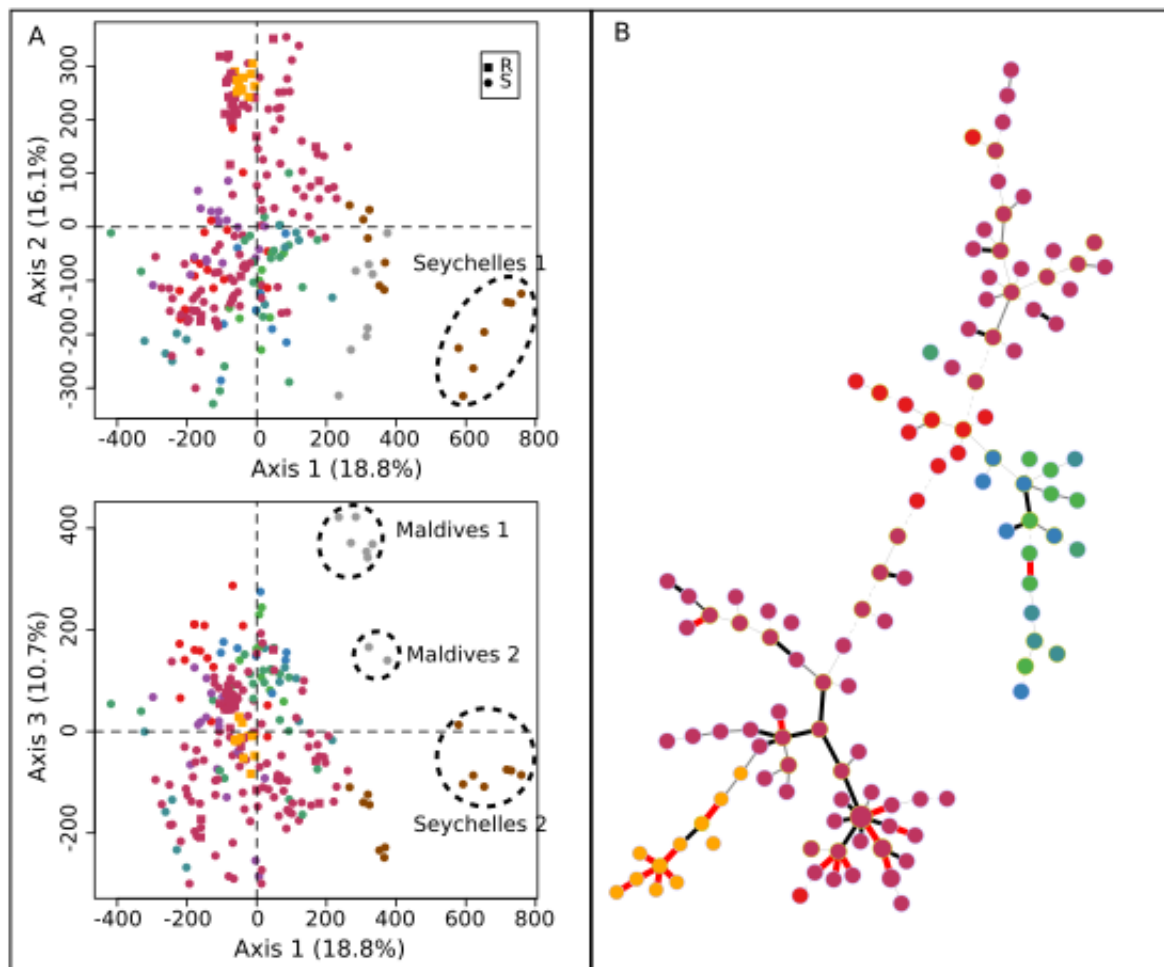


Fig. S3. Genetic relatedness based on microsatellite data (MLVA-14) among copper-susceptible (CuS) and copper-resistant (CuR) *Xanthomonas citri* pv. *citri* strains originating from the Indian Ocean region. Color code is for islands and is similar to that used in Fig. 1 and Fig. S2. A: Manhattan distances represented by multidimensional scaling (MDS). Note the remote placement of strains originating from the NIO region (*i.e.* Maldives and Seychelles archipelagos). B: Categorical minimum spanning tree representing the relatedness among strains originating from the Comoros and Mascarenes and assigned to cluster MLVA 1 (116 strains out of 211- see Fig. S2). Dot diameter is representative of the number of strains per haplotype. Thick red, thick black, regular grey, thin grey and thin dotted grey lines indicate single-, double-, triple-, quadruple- and quintuple-locus variations between haplotypes, respectively.

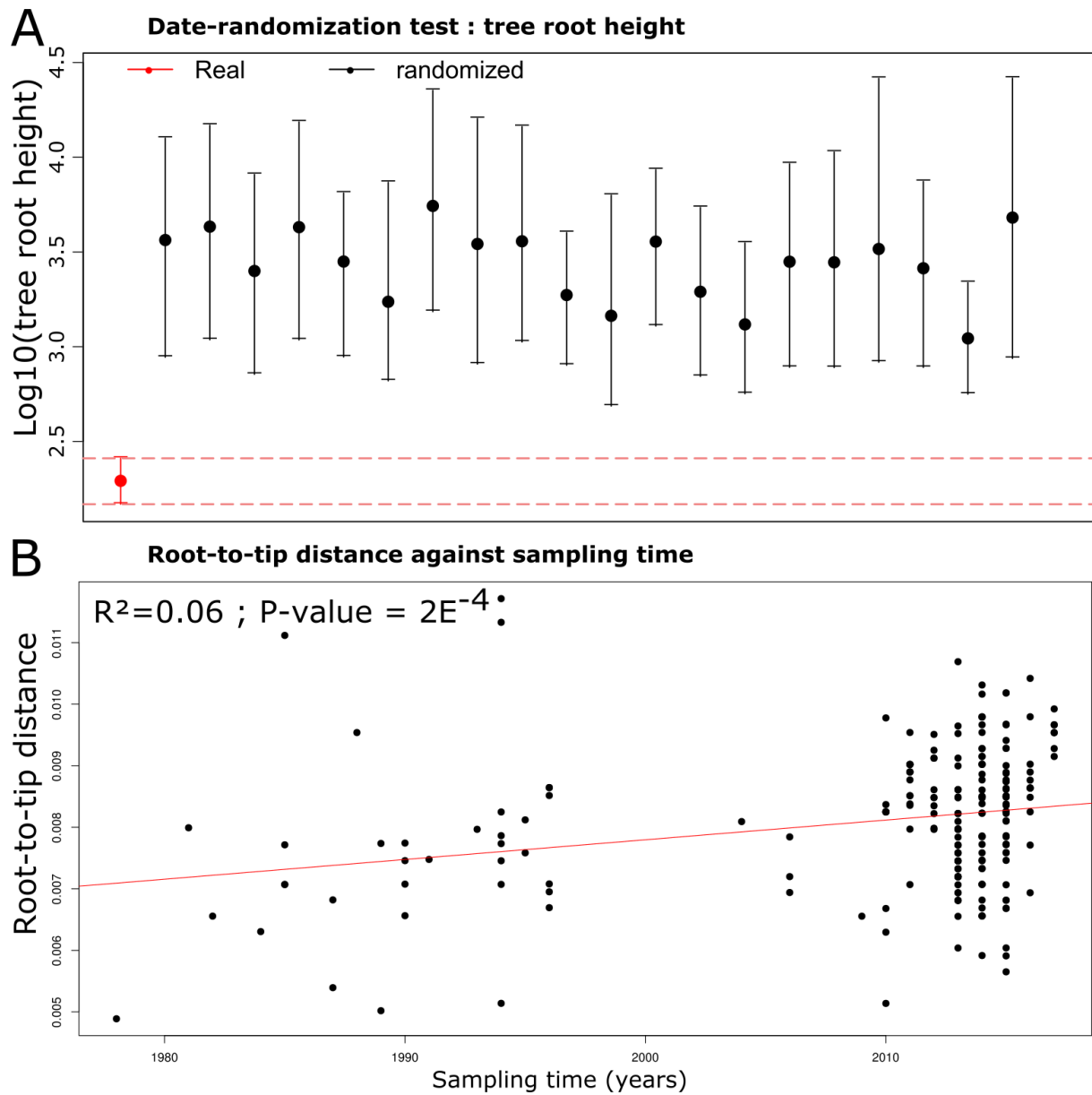


Fig. S4. (A) Estimated SWIO tree root age and associated 95% HPD obtained from 20 BEAST analysis ran with randomized sample dates and from the real dataset. (B) Root-to-tip regression between root-to-tip distance and tip dates.

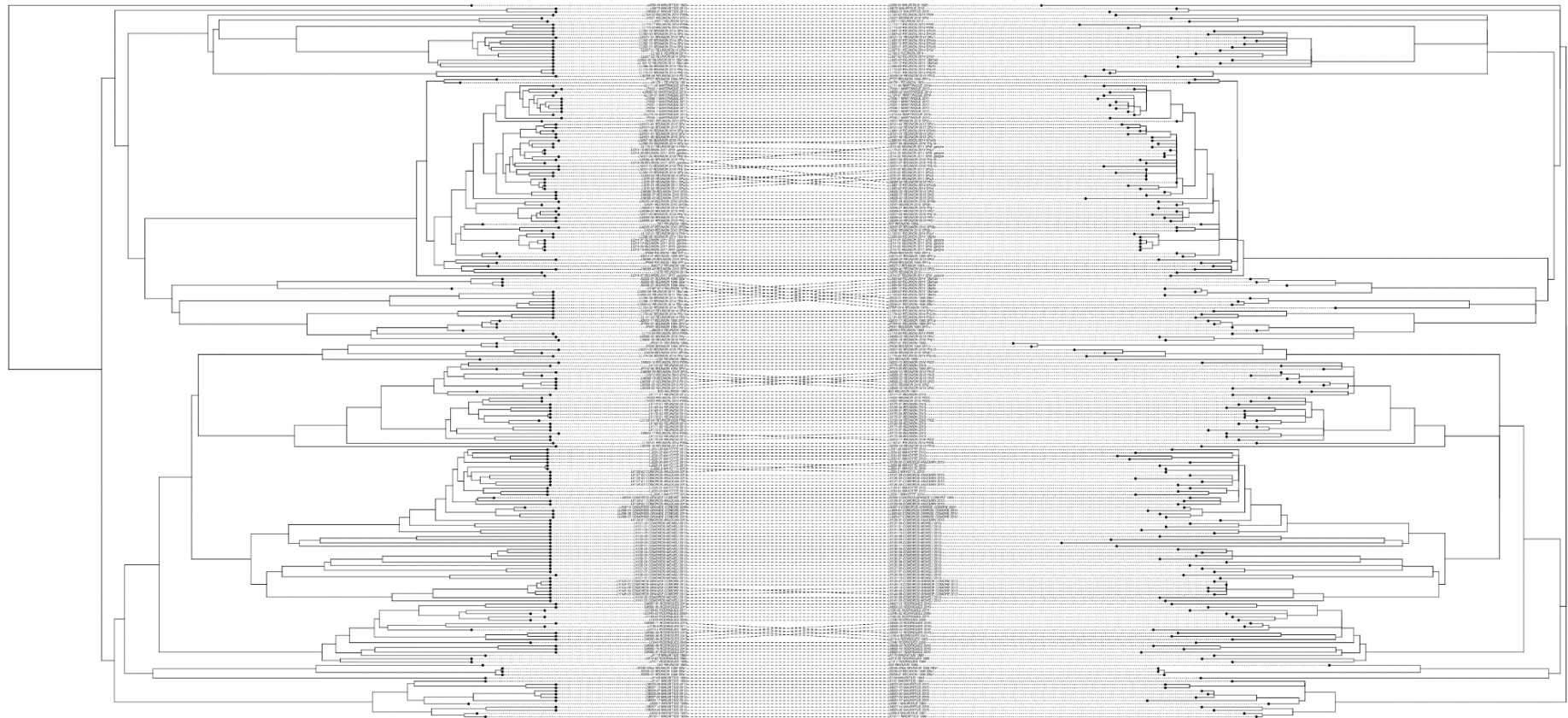


Fig. S5. Comparison between topologies of Maximum Likelihood phylogeny and Bayesian phylogeny.

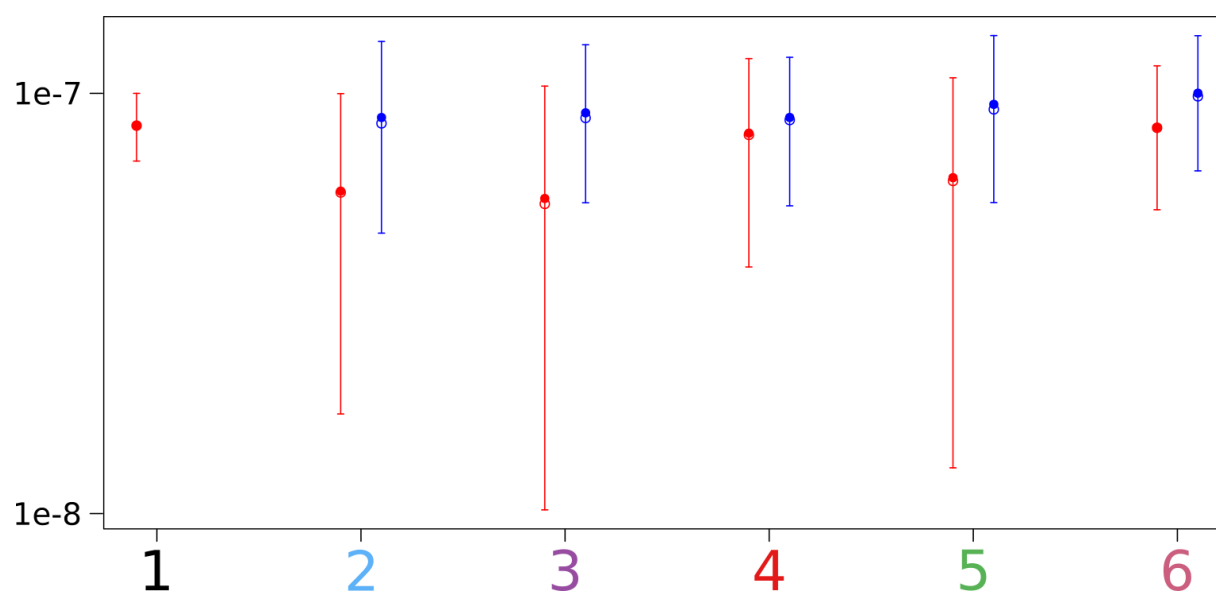


Fig. S6 Comparison of the 95% confidence intervals of the SNP substitution rates of the SWIO root and of five subclades as inferred using the total dataset (blue) or the subset of strains corresponding to each subclade (red). Filled circle: mean value; Empty circle: median.

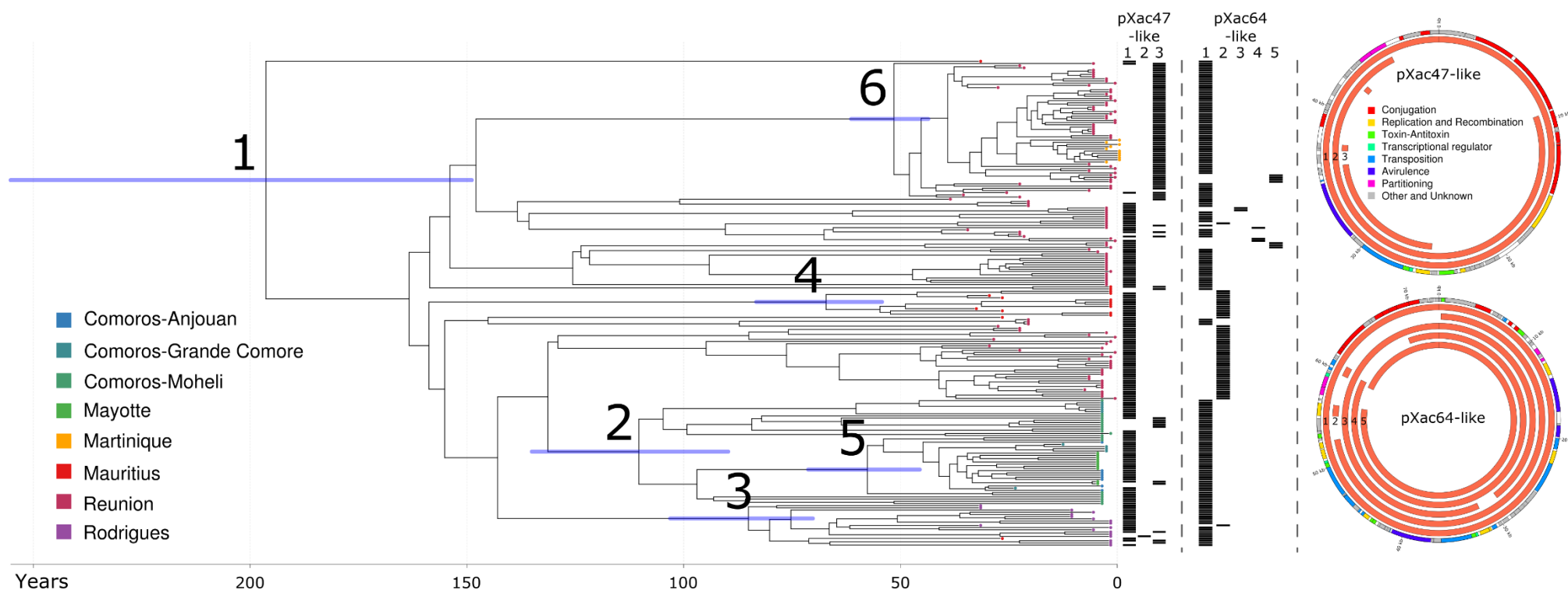


Fig. S7 A dated phylogeny of the SWIO clade of *Xanthomonas citri* pv. *citri*. Node bars represent 95% highest posterior density (HPD) for node ages estimated with tip-calibration. Tips are colored according to sampling location. Columns in the middle of the figure indicate which version of pXac47 and pXac64 each strain encode. Circular representations of the two plasmids pXac64-like (GenBank accession CP018856.1, from strain Xcc LH276) and pXac47-like (GenBank accession CP018853.1, from strain Xcc LJ207-7) indicate the positions of known gene functions according to the legend and plasmid regions (red: gene present in the strain; white: absent) present in the SWIO strains, numbered from 1 to 5 for pXac64-like and from 1 to 3 for pXac47-like.

Table S1. Host, geographic location, copper phenotype and MLVA profiles of the strains used in this study.

Strain	Host	Country	Site number	Date	Pathotype	Copper phenotype	seq	Haplotype	Population	clusterMST5
JJ155	Citrus sp.	Argentina		1977	A	S				
JK101-1	<i>C. x paradisi</i>	Argentina		1990	A	S				
LM158	<i>C. x sinensis</i>	Argentina		2013	A	Rabcd				
LM169	<i>C. x paradisi</i>	Argentina		2005	A	Rlab				
LM180	<i>C. x paradisi</i>	Argentina		2003	A	Rlab				
LM184	<i>C. x limon</i>	Argentina		2013	A	S				
LM198	<i>C. x sinensis</i>	Argentina		2015	A	S				
LM199	<i>C. x sinensis</i>	Argentina		2015	A	Rabcd				
LM205	<i>C. x paradisi</i>	Argentina		2010	A	Rlab				
LM229	<i>C. x sinensis</i>	Argentina		2012	A	Rabcd				
JW160	<i>C. x aurantiifolia</i>	Bangladesh		2000	A	S				
LG102	Citrus sp.	Bangladesh		2006	A	S				
LG117	<i>C. x aurantiifolia</i>	Bangladesh		2009	A	S				
LG97	<i>C. x limon</i>	Bangladesh		2006	A	S				
LG98	<i>C. x aurantiifolia</i>	Bangladesh		2006	A	S				
CFBP 2865	<i>C. x aurantiifolia</i>	Brazil		1976	A	S				
CFBP 2908	<i>C. reticulata</i>	Brazil		1980	A	S				
FDC1083	<i>C. reticulata</i>	Brazil		1980	A	S				
FDC217	<i>C. x sinensis</i>	Brazil		2003	A	S				
JK4-1	Citrus sp.	China		unknown	A	S				
LK126-01	<i>C. x aurantiifolia</i>	Comoros (Anjouan)		2013	A	S	O	4	1	MLVA-01
LK126-02	<i>C. x aurantiifolia</i>	Comoros (Anjouan)		2013	A	S	O	5	1	MLVA-01

LK126-03	<i>C. × aurantiifolia</i>	Comoros (Anjouan)	2013	A	S	O	6	1	MLVA-01
LK127-01	<i>C. x sinensis</i>	Comoros (Anjouan)	2013	A	S	O	10	1	singleton
LK127-03	<i>C. x sinensis</i>	Comoros (Anjouan)	2013	A	S	O	12	1	MLVA-01
LK128-01	<i>C. x sinensis</i>	Comoros (Anjouan)	2013	A	S	O	16	1	MLVA-01
LK128-02	<i>C. x sinensis</i>	Comoros (Anjouan)	2013	A	S	O	17	1	singleton
LK129-03	<i>C. × aurantiifolia</i>	Comoros (Anjouan)	2013	A	S	O	20	1	MLVA-01
JN564	<i>Citrus</i> sp.	Comoros (Grande Comore)	1993	A	S	O	1	2	MLVA-01
LA087-2	<i>Citrus</i> sp.	Comoros (Grande Comore)	2004	A	S	O	3	2	MLVA-01
LK144-08	<i>Citrus</i> sp.	Comoros (Grande Comore)	2013	A	S	O	109	2	MLVA-05
LK145-07	<i>Citrus</i> sp.	Comoros (Grande Comore)	2013	A	S	O	115	2	MLVA-05
LK145-10	<i>Citrus</i> sp.	Comoros (Grande Comore)	2013	A	S	O	114	2	MLVA-05
LK145-14	<i>Citrus</i> sp.	Comoros (Grande Comore)	2013	A	S	O	119	2	MLVA-05
LK148-03	<i>Citrus</i> sp.	Comoros (Grande Comore)	2013	A	S	O	127	2	MLVA-05
LL068-04	<i>Citrus</i> sp.	Comoros (Grande Comore)	2014	A	S	O	144	2	MLVA-01
LL068-06	<i>Citrus</i> sp.	Comoros (Grande Comore)	2014	A	S	O	146	2	MLVA-01
LL068-07	<i>Citrus</i> sp.	Comoros (Grande Comore)	2014	A	S	O	147	2	singleton
LK130-04	<i>C. × aurantiifolia</i>	Comoros (Moheli)	2013	A	S	O	24	6	singleton
LK130-06	<i>C. × aurantiifolia</i>	Comoros (Moheli)	2013	A	S	O	26	6	singleton
LK130-09	<i>C. × aurantiifolia</i>	Comoros (Moheli)	2013	A	S	O	29	6	singleton
LK130-11	<i>C. × aurantiifolia</i>	Comoros (Moheli)	2013	A	S	O	31	6	singleton
LK131-01	<i>C. x sinensis</i>	Comoros (Moheli)	2013	A	S	O	35	6	MLVA-01

LK131-04	<i>C. × aurantiifolia</i>	Comoros (Moheli)	2013	A	S	O	38	6	singleton
LK131-10	<i>C. × aurantiifolia</i>	Comoros (Moheli)	2013	A	S	O	44	6	singleton
LK132-03	<i>C. x sinensis</i>	Comoros (Moheli)	2013	A	S	O	47	6	singleton
LK132-08	<i>C. × aurantiifolia</i>	Comoros (Moheli)	2013	A	S	O	52	6	MLVA-01
LK135-03	<i>Citrus</i> sp.	Comoros (Moheli)	2015	A	S	O	59	6	singleton
LK136-01	<i>C. × aurantiifolia</i>	Comoros (Moheli)	2013	A	S	O	60	6	singleton
LK136-04	<i>C. × aurantiifolia</i>	Comoros (Moheli)	2013	A	S	O	63	6	singleton
LK136-05	<i>C. × aurantiifolia</i>	Comoros (Moheli)	2013	A	S	O	64	6	singleton
LK136-08	<i>C. × aurantiifolia</i>	Comoros (Moheli)	2013	A	S	O	66	6	singleton
LK137-01	<i>C. x sinensis</i>	Comoros (Moheli)	2013	A	S	O	68	6	singleton
LK137-02	<i>C. × aurantiifolia</i>	Comoros (Moheli)	2013	A	S	O	69	6	singleton
LK141-03	<i>C. x sinensis</i>	Comoros (Moheli)	2013	A	S	O	72	6	singleton
LK141-09	<i>C. x sinensis</i>	Comoros (Moheli)	2013	A	S	O	77	6	singleton
LK141-15	<i>C. x sinensis</i>	Comoros (Moheli)	2013	A	S	O	82	6	singleton
LK142-04	<i>C. × aurantiifolia</i>	Comoros (Moheli)	2013	A	S	O	86	6	singleton
CFBP 2857	<i>Citrus</i> sp.	Fiji	1976	A	S				
CFBP 1209	<i>C. maxima</i>	Hong Kong	1963	A	S				
CFBP2852	<i>Citrus</i> sp.	India		A	S				
JJ165	<i>C. maxima</i>	India	1978	A	S				
NCPPB 211	<i>Citrus</i> sp.	India	1948	A	S				
NCPPB3562	<i>C. x limon</i>	India	1988	A	S				
NCPPB3610	<i>P. trifoliata</i>	India	1988	A	S				
NCPPB3612	<i>C. × aurantiifolia</i>	India	1988	A	S				
CFBP 2855	<i>Citrus</i> sp.	Japan	1962	A	S				
JJ238-04	<i>C. × aurantiifolia</i>	Maldives	1987	A	S	O	153	3	MLVA-04
JJ238-06	<i>C. × aurantiifolia</i>	Maldives	1987	A	S	O	155	3	MLVA-11
JJ238-07	<i>C. × aurantiifolia</i>	Maldives	1987	A	S	O	156	3	MLVA-04
JJ238-08	<i>C. × aurantiifolia</i>	Maldives	1987	A	S	O	157	3	MLVA-04
JJ238-09	<i>C. × aurantiifolia</i>	Maldives	1987	A	S	O	158	3	singleton

JJ238-10	<i>C. × aurantiifolia</i>	Maldives	1987	A	S	O	159	3	MLVA-04
JJ238-11	<i>C. × aurantiifolia</i>	Maldives	1987	A	S	O	160	3	MLVA-11
JK004-04	<i>C. × aurantiifolia</i>	Maldives	1987	A	S	O	162	3	MLVA-04
LC80	<i>C. × nobilis</i>	Mali	2006	A	S				
LD7-1	<i>C. × aurantiifolia</i>	Mali	2007	A	S				
LE116-1	<i>C. × aurantiifolia</i>	Mali	2008	A	S				
LL074-04	<i>C. x paradisi</i>	Martinique	2014	A	Rlab	A	1182	10	MLVA-01
LL111-06	<i>C. x sinensis</i>	Martinique	2014	A	Rlab	A	1183	10	MLVA-01
LL124-01	<i>C. x sinensis</i>	Martinique	2014	A	Rlab	A	1184	10	MLVA-01
LM090-02	<i>Citrus</i> sp.	Martinique	2015	A	Rlab	A	1185	10	MLVA-01
LP030-1	<i>C. x latifolia</i>	Martinique	2017	A	Rlab	A	1186	10	MLVA-01
LP031-1	<i>C. x sinensis</i>	Martinique	2017	A	Rlab	A	1187	10	MLVA-01
LP032-1	<i>C. × aurantiifolia</i>	Martinique	2017	A	Rlab	A	1188	10	MLVA-01
LP033-1	<i>C. x paradisi</i>	Martinique	2017	A	Rlab	A	1189	10	MLVA-01
LP034-1	<i>C. x latifolia</i>	Martinique	2017	A	Rlab	A	1190	10	MLVA-01
LP035-1	<i>C. x latifolia</i>	Martinique	2017	A	Rlab	A	1191	10	MLVA-01
LP036-1	<i>C. x latifolia</i>	Martinique	2017	A	Rlab	A	1192	10	MLVA-01
JJ009-04	<i>C. x bergamia</i>	Mauritius	1985	A	S	O	166	4	MLVA-01
JJ009-1	<i>C. x insitorum</i>	Mauritius	1984	A	S	O	163	4	MLVA-01
JJ009-8	<i>C. x sinensis</i>	Mauritius	1987	A	S	O	169	4	singleton
JK161	<i>C. x sinensis</i>	Mauritius	1990	A	S	O	171	4	MLVA-01
JK167-1	<i>C. × aurantiifolia</i>	Mauritius	1990	A	S	O	169	4	MLVA-01
JK169	<i>Citrus</i> sp.	Mauritius	1990	A	S	O	177	4	singleton
JK170	<i>C. × aurantiifolia</i>	Mauritius	1990	A	S	O	178	4	singleton
LM053-06	<i>C. × aurantiifolia</i>	Mauritius	2015	A	S	O	190	4	MLVA-01
LM053-07	<i>C. × aurantiifolia</i>	Mauritius	2015	A	S	O	191	4	MLVA-01
LM054-06	<i>C. x sinensis</i>	Mauritius	2015	A	S	O	201	4	MLVA-01
LM054-17	<i>C. x sinensis</i>	Mauritius	2015	A	S	O	205	4	MLVA-01

<i>C. reticulata</i> × <i>C. ×</i>										
LM055-08	<i>paradisi</i>	Mauritius		2015	A	S	O	208	4	MLVA-01
LM057-04	<i>C. × meyeri</i>	Mauritius		2015	A	S	O	211	4	MLVA-01
LM057-14	<i>C. × meyeri</i>	Mauritius		2015	A	S	O	215	4	MLVA-01
LM057-15	<i>C. × meyeri</i>	Mauritius		2015	A	S	O	216	4	MLVA-01
LM069-01	<i>C. × aurantiifolia</i>	Mauritius		2015	A	S	O	220	4	singleton
LM070	<i>C. × aurantiifolia</i>	Mauritius		2015	A	S	O	221	4	MLVA-01
LJ125-01	<i>C. × sinensis</i>	Mayotte		2012	A	S	O	242	5	MLVA-01
LJ225-01	<i>C. × sinensis</i>	Mayotte		2012	A	S	O	246	5	MLVA-01
LJ225-2	<i>C. × sinensis</i>	Mayotte		2012	A	S	O	247	5	MLVA-01
LJ226-02	<i>C. × sinensis</i>	Mayotte		2012	A	S	O	250	5	MLVA-01
LJ228-1	<i>C. hystrix</i>	Mayotte		2012	A	S	O	259	5	singleton
LJ229-06	<i>C. reticulata</i>	Mayotte		2012	A	S	O	260	5	singleton
LJ230-03	<i>C. × sinensis</i>	Mayotte		2012	A	S	O	269	5	MLVA-01
LJ231-09	<i>Citrus</i> sp.	Mayotte		2012	A	S	O	278	5	MLVA-01
LJ232-01	<i>C. × sinensis</i>	Mayotte		2012	A	S	O	279	5	MLVA-01
LJ232-05	<i>C. × sinensis</i>	Mayotte		2012	A	S	O	283	5	MLVA-01
CFBP 2525	<i>C. × limon</i>	New Zealand		1956	A	S				
CFBP 2853	<i>Citrus</i> sp.	New Zealand		1958	A	S				
NCPBPB 226	<i>Citrus</i> sp.	New Zealand		1949	A	S				
NCPBPB406	<i>Citrus</i> sp.	New Zealand		1957	A	S				
JJ238-16	<i>C. × sinensis</i>	Pakistan		1988	S	S				
LI214-01	<i>C. × aurantiifolia</i>	Reunion	1	2011	A	S	O	612	7	MLVA-01
LI214-02	<i>C. × aurantiifolia</i>	Reunion	1	2011	A	Rlab	O	627	7	MLVA-01
LI214-05	<i>C. × aurantiifolia</i>	Reunion	1	2011	A	S	O	628	7	MLVA-01
LI214-07	<i>C. × aurantiifolia</i>	Reunion	1	2011	A	S	O	629	7	MLVA-01
LI214-09	<i>C. × aurantiifolia</i>	Reunion	1	2011	A	Rlab	O	630	7	MLVA-01
LI214-10	<i>C. × aurantiifolia</i>	Reunion	1	2011	A	S	O	631	7	MLVA-01
LI214-14	<i>C. × aurantiifolia</i>	Reunion	1	2011	A	S	O	632	7	MLVA-01

LI214-16	<i>C. × aurantiifolia</i>	Reunion	1	2011	A	Rlab	O	627	7	MLVA-01
LL131-03	<i>C. × nobilis</i>	Reunion	2	2014	A	S	O	803	7	MLVA-14
LL132-01	<i>C. × nobilis</i>	Reunion	2	2014	A	S	O	808	7	MLVA-01
LM089-02	<i>C. × nobilis</i>	Reunion	2	2015	A	Rlab	O	895	7	MLVA-10
LM089-06	<i>C. × nobilis</i>	Reunion	2	2015	A	S	O	896	7	MLVA-01
LM089-20	<i>C. × nobilis</i>	Reunion	2	2015	A	S	O	899	7	MLVA-01
LM089-41	<i>C. × nobilis</i>	Reunion	2	2015	A	S	O	907	7	MLVA-01
LM089-42	<i>C. × nobilis</i>	Reunion	2	2015	A	Rlab	O	908	7	MLVA-01
LL175-01	<i>C. × nobilis</i>	Reunion	2	2014	A	Rlab	O	611	7	MLVA-01
LN006-01	<i>C. × nobilis</i>	Reunion	2	2016	A	Rlab	O	906	7	MLVA-01
LN006-18	<i>C. × nobilis</i>	Reunion	2	2016	A	Rlab	O	1071	7	MLVA-10
LL082-03	<i>C. × nobilis</i>	Reunion	3	2014	A	S	O	645	7	MLVA-02
LL082-04	<i>C. × nobilis</i>	Reunion	3	2014	A	Rlab	O	646	7	MLVA-01
LL082-13	<i>C. × nobilis</i>	Reunion	3	2014	A	S	O	652	7	MLVA-02
LL082-15	<i>C. × nobilis</i>	Reunion	3	2014	A	Rlab	O	653	7	MLVA-01
LL082-16	<i>C. × nobilis</i>	Reunion	3	2014	A	S	O	654	7	MLVA-02
LL082-18	<i>C. × nobilis</i>	Reunion	3	2014	A	Rlab	O	656	7	MLVA-01
LL082-21	<i>C. × nobilis</i>	Reunion	3	2014	A	S	O	657	7	MLVA-02
LL082-25	<i>C. × nobilis</i>	Reunion	3	2014	A	S	O	661	7	MLVA-02
LL087-01	<i>C. x limon</i>	Reunion	4	2014	A	S	O	634	7	MLVA-02
LL087-02	<i>C. x limon</i>	Reunion	4	2014	A	S	O	679	7	MLVA-02
LL085-02	<i>C. x limon</i>	Reunion	4	2014	A	Rlab	O	653	7	MLVA-01
LL085-03	<i>C. x limon</i>	Reunion	4	2014	A	S	O	677	7	singleton
LM121-01	<i>C. x limon</i>	Reunion	4	2015	A	Rlab	O	909	7	MLVA-01
LM121-02	<i>C. x limon</i>	Reunion	4	2015	A	Rlab	O	673	7	MLVA-01
LM121-31	<i>C. x limon</i>	Reunion	4	2015	A	Rlab	O	920	7	MLVA-02
LM121-41	<i>C. x limon</i>	Reunion	4	2015	A	Rlab	O	925	7	MLVA-01
LM121-42	<i>C. x limon</i>	Reunion	4	2015	A	Rlab	O	923	7	MLVA-01
LJ207-7	<i>C. hystrix</i>	Reunion	5	2012	A	Rlab	O	1169	7	MLVA-15

LH201	<i>C. hystrix</i>	Reunion	5	2010	A	Rlab	O	433	7	MLVA-01
LH212	<i>C. hystrix</i>	Reunion	5	2010	A	S	O	442	7	MLVA-01
LH221	<i>C. hystrix</i>	Reunion	5	2010	A	S	O	448	7	MLVA-15
LM088-18	<i>C. hystrix</i>	Reunion	5	2015	A	S	O	876	7	MLVA-01
LM088-20	<i>C. hystrix</i>	Reunion	5	2015	A	Rlab	O	878	7	MLVA-01
LM088-25	<i>C. hystrix</i>	Reunion	5	2015	A	S	O	881	7	MLVA-01
LM088-28	<i>C. hystrix</i>	Reunion	5	2015	A	Rlab	O	882	7	MLVA-01
LM088-37	<i>C. hystrix</i>	Reunion	5	2015	A	Rlab	O	889	7	MLVA-01
LM088-40	<i>C. hystrix</i>	Reunion	5	2015	A	S	O	891	7	MLVA-01
LM088-42	<i>C. hystrix</i>	Reunion	5	2015	A	Rlab	O	893	7	MLVA-01
LI070-01	<i>C. hystrix</i>	Reunion	6	2011	A	Rlab	O	611	7	MLVA-01
LI070-02	<i>C. hystrix</i>	Reunion	6	2011	A	Rlab	O	611	7	MLVA-01
LI070-03	<i>C. hystrix</i>	Reunion	6	2011	A	Rlab	O	611	7	MLVA-01
LI070-05	<i>C. hystrix</i>	Reunion	6	2011	A	Rlab	O	611	7	MLVA-01
LN007-02	<i>C. x limon</i>	Reunion	7	2016	A	S	O	1072	7	MLVA-01
LN007-03	<i>C. x limon</i>	Reunion	7	2016	A	Rlab	O	1073	7	MLVA-01
LN007-05	<i>C. x limon</i>	Reunion	7	2016	A	Rlab	O	1074	7	MLVA-01
LN007-07	<i>C. x limon</i>	Reunion	7	2016	A	Rlab	O	1076	7	MLVA-01
LN007-09	<i>C. x limon</i>	Reunion	7	2016	A	Rlab	O	1077	7	MLVA-01
LN007-10	<i>C. x limon</i>	Reunion	7	2016	A	Rlab	O	1078	7	MLVA-01
LL174-01	<i>C. x limon</i>	Reunion	7	2014	A	S	O	842	7	MLVA-18
LL174-02	<i>C. x limon</i>	Reunion	7	2014	A	S	O	843	7	MLVA-14
LL174-04	<i>C. x limon</i>	Reunion	7	2014	A	S	O	844	7	MLVA-01
LL174-05	<i>C. x limon</i>	Reunion	7	2014	A	S	O	845	7	MLVA-18
LH238	<i>C. hystrix</i>	Reunion	8	2010	A	S	O	464	7	MLVA-01
LH240	<i>C. hystrix</i>	Reunion	8	2010	A	S	O	465	7	MLVA-01
LH241	<i>C. hystrix</i>	Reunion	8	2010	A	S	O	466	7	MLVA-01
LN005-04	<i>C. hystrix</i>	Reunion	8	2016	A	S	O	1054	7	MLVA-01
LN005-07	<i>C. hystrix</i>	Reunion	8	2016	A	S	O	1057	7	MLVA-01

LM358-02	<i>C. x limon</i>	Reunion	9	2015	A	S	O	994	7	MLVA-01
LM358-08	<i>C. x limon</i>	Reunion	9	2015	A	S	O	1000	7	singleton
LM358-13	<i>C. x limon</i>	Reunion	9	2015	A	S	O	1004	7	MLVA-01
LM358-18	<i>C. x limon</i>	Reunion	9	2015	A	S	O	1006	7	singleton
LM358-22	<i>C. x limon</i>	Reunion	9	2015	A	S	O	1009	7	MLVA-01
JP541	<i>C. x paradisi</i>	Reunion	10	1994	A	S	O	363	7	MLVA-06
JP608	<i>C. x paradisi</i>	Reunion	10	1994	A	S	O	371	7	MLVA-16
JP710-06	<i>C. x paradisi</i>	Reunion	10	1994	A	S	O	387	7	MLVA-01
JP727	<i>C. x paradisi</i>	Reunion	10	1994	A	S	O	390	7	MLVA-01
JP764-01	<i>C. x paradisi</i>	Reunion	10	1994	A	S	O	397	7	MLVA-06
LK169-01	<i>C. hystrix</i>	Reunion	11	2013	A	S	O	1170	7	MLVA-01
LK169-03	<i>C. hystrix</i>	Reunion	11	2013	A	S	O	1171	7	MLVA-01
LK169-04	<i>C. hystrix</i>	Reunion	11	2013	A	S	O	1172	7	MLVA-01
LK170-01	<i>C. hystrix</i>	Reunion	11	2013	A	S	O	1173	7	MLVA-01
LK170-03	<i>C. hystrix</i>	Reunion	11	2013	A	S	O	1174	7	MLVA-01
LK170-05	<i>C. hystrix</i>	Reunion	11	2013	A	S	O	1175	7	MLVA-01
LK171-01	<i>C. hystrix</i>	Reunion	11	2013	A	S	O	1176	7	MLVA-01
LK171-02	<i>C. hystrix</i>	Reunion	11	2013	A	S	O	1177	7	MLVA-01
LK172-01	<i>C. hystrix</i>	Reunion	11	2013	A	S	O	1178	7	MLVA-01
LK172-02	<i>C. hystrix</i>	Reunion	11	2013	A	S	O	1179	7	MLVA-01
LK173-01	<i>C. hystrix</i>	Reunion	11	2013	A	S	O	1180	7	MLVA-01
LK173-02	<i>C. hystrix</i>	Reunion	11	2013	A	S	O	1181	7	MLVA-01
LH320	<i>C. hystrix</i>	Reunion	11	2010	A	S	O	531	7	MLVA-01
LH322	<i>C. hystrix</i>	Reunion	11	2010	A	S	O	533	7	MLVA-01
LN003-10	<i>C. hystrix</i>	Reunion	11	2016	A	S	O	1021	7	MLVA-01
LN003-17	<i>C. hystrix</i>	Reunion	11	2016	A	S	O	1026	7	MLVA-01
LG136-04	<i>C. hystrix</i>	Reunion	11	2009	A	S	O	426	7	MLVA-01
JP668	<i>C. hystrix</i>	Reunion	12	1994	A	S	O	378	7	MLVA-01
JP669	<i>C. hystrix</i>	Reunion	12	1994	A	S	O	379	7	MLVA-01

JQ612-17	<i>C. hystrix</i>	Reunion	12	1995	A	S	O	368	7	MLVA-06
JQ613-01	<i>C. hystrix</i>	Reunion	12	1995	A	S	O	372	7	MLVA-01
LL115-09	<i>C. × nobilis</i>	Reunion	13	2014	A	S	O	778	7	singleton
LL115-17	<i>C. × nobilis</i>	Reunion	13	2014	A	S	O	783	7	MLVA-12
LL115-20	<i>C. × nobilis</i>	Reunion	13	2014	A	S	O	785	7	MLVA-12
LL162-01	<i>C. × nobilis</i>	Reunion	13	2014	A	S	O	823	7	MLVA-01
LL164-02	<i>C. × nobilis</i>	Reunion	13	2014	A	S	O	828	7	singleton
LL095-02	<i>C. × nobilis</i>	Reunion	14	2014	A	S	O	681	7	MLVA-01
LL095-08	<i>C. × nobilis</i>	Reunion	14	2014	A	S	O	686	7	MLVA-01
LL095-20	<i>C. × nobilis</i>	Reunion	14	2014	A	S	O	692	7	MLVA-02
LL098-02	<i>C. × nobilis</i>	Reunion	14	2014	A	S	O	687	7	MLVA-17
LL100-12	<i>C. × nobilis</i>	Reunion	14	2014	A	S	O	745	7	MLVA-02
JS502-01	<i>C. x paradisi</i>	Reunion	15	1996	A	S	O	407	7	MLVA-08
JS503-05	<i>C. x paradisi</i>	Reunion	15	1996	A	S	O	408	7	MLVA-08
JS536-01	<i>C. x paradisi</i>	Reunion	15	1996	A	S	O	413	7	MLVA-08
JS538-02	<i>C. x paradisi</i>	Reunion	15	1996	A	S	O	409	7	MLVA-19
JS858-01	<i>C. x paradisi</i>	Reunion	15	1996	A	S	O	418	7	MLVA-19
LL096-06	<i>C. x limon</i>	Reunion	16	2014	A	S	O	699	7	MLVA-01
LL096-08	<i>C. x limon</i>	Reunion	16	2014	A	S	O	701	7	MLVA-01
LL096-09	<i>C. x limon</i>	Reunion	16	2014	A	S	O	702	7	MLVA-02
LL096-13	<i>C. x limon</i>	Reunion	16	2014	A	S	O	703	7	MLVA-17
LL102-02	<i>C. x limon</i>	Reunion	16	2014	A	S	O	712	7	singleton
JP637-01	<i>C. reticulata</i>	Reunion		1994	A	S	O	327	7	MLVA-16
B25	<i>C. hystrix</i>	Reunion		1987	A	S	O	297	7	MLVA-01
C20	<i>C. x paradisi</i>	Reunion		1988	A	S	O	310	7	singleton
CFBP1814	<i>C. maxima</i>	Reunion		1978	A	S	O	332	7	singleton
D02	<i>C. × tangelo</i>	Reunion		1989	A	S	O	318	7	singleton
D07	<i>C. × unshiu</i>	Reunion		1989	A	S	O	336	7	MLVA-01
JA159-1	<i>C. × clementina</i>	Reunion		1981	A	S	O	342	7	MLVA-01

JB003-5	<i>C. × aurantiifolia</i>	Reunion	1982	A	S	O	346	7	MLVA-06
JM027-2	<i>C. x limon</i>	Reunion	1991	A	S	O	356	7	MLVA-01
LH276	<i>C. × nobilis</i>	Reunion	2010	A	Rlab	O	497	7	MLVA-01
LL186-5	<i>C. x sinensis</i>	Reunion	2014	A	S	O	862	7	MLVA-02
JJ010-02	<i>C. × aurantiifolia</i>	Rodrigues	1985	A	S	O	1080	8	singleton
JJ010-4	<i>C. × aurantiifolia</i>	Rodrigues	1985	A	S	O	1081	8	MLVA-13
JJ10-1	<i>C. × aurantiifolia</i>	Rodrigues	1985	A	S	O	1079	8	MLVA-13
LC045-02	<i>C. × aurantiifolia</i>	Rodrigues	2006	A	S	O	1087	8	singleton
LC046	<i>C. × aurantiifolia</i>	Rodrigues	2006	A	S	O	1088	8	singleton
LC048	<i>C. × aurantiifolia</i>	Rodrigues	2006	A	S	O	1089	8	singleton
LI139-03	<i>C. × aurantiifolia</i>	Rodrigues	2011	A	S	O	1095	8	singleton
LI139-05	<i>C. × aurantiifolia</i>	Rodrigues	2011	A	S	O	1097	8	singleton
LI139-6	<i>C. × aurantiifolia</i>	Rodrigues	2011	A	S	O	1098	8	singleton
LM095-04	<i>C. × aurantiifolia</i>	Rodrigues	2015	A	S	O	1105	8	singleton
LM095-05	<i>C. × aurantiifolia</i>	Rodrigues	2015	A	S	O	1106	8	singleton
LM095-07	<i>C. × aurantiifolia</i>	Rodrigues	2015	A	S	O	1107	8	singleton
LM095-11	<i>C. × aurantiifolia</i>	Rodrigues	2015	A	S	O	1110	8	singleton
LM095-14	<i>C. × aurantiifolia</i>	Rodrigues	2015	A	S	O	1112	8	singleton
LM096-03	<i>C. × aurantiifolia</i>	Rodrigues	2015	A	S	O	1120	8	singleton
LM096-08	<i>C. × aurantiifolia</i>	Rodrigues	2015	A	S	O	1124	8	singleton
LM096-09	<i>C. × aurantiifolia</i>	Rodrigues	2015	A	S	O	1125	8	singleton
LM097-01	<i>C. × aurantiifolia</i>	Rodrigues	2015	A	S	O	1127	8	singleton
JK002-10	<i>C. × aurantiifolia</i>	Saudi Arabia	1988	A*	S				
LH37-1	<i>C. x paradisi</i>	Senegal	2010	A	S				
JZ092	<i>C. x limon</i>	Seychelles	2003	A	S	O	1141	9	MLVA-03
JZ094	<i>C. x limon</i>	Seychelles	2003	A	S	O	1143	9	MLVA-07
LB100-1	<i>C. x insitorum</i>	Seychelles	2005	A	S	O	1144	9	MLVA-03
LB100-3	<i>C. x insitorum</i>	Seychelles	2005	A	S	O	1146	9	MLVA-03
LJ001-1	<i>Citrus</i> sp.	Seychelles	2012	A	S	O	1148	9	MLVA-03

LJ003-1	<i>C. x sinensis</i>	Seychelles	2012	A	S	O	1149	9	MLVA-03
LP027-03	<i>C. x aurantiifolia</i>	Seychelles	2017	A	S	O	1151	9	MLVA-07
LP027-05	<i>C. x aurantiifolia</i>	Seychelles	2017	A	S	O	1155	9	MLVA-07
LP027-13	<i>C. x aurantiifolia</i>	Seychelles	2017	A	S	O	1153	9	MLVA-07
LP028-03	<i>C. x aurantiifolia</i>	Seychelles	2017	A	S	O	1159	9	MLVA-09
LP028-05	<i>C. x aurantiifolia</i>	Seychelles	2017	A	S	O	1160	9	MLVA-09
LP028-06	<i>C. x aurantiifolia</i>	Seychelles	2017	A	S	O	1161	9	MLVA-09
LP029-13	<i>C. x sinensis</i>	Seychelles	2017	A	S	O	1164	9	MLVA-03
LP029-15	<i>C. x sinensis</i>	Seychelles	2017	A	S	O	1165	9	MLVA-03
JJ053-06	<i>C. reticulata</i> var. <i>sunki</i>	Taiwan	1977	A	S				
LMG9322	<i>C. x aurantiifolia</i>	USA (Florida)	1986	A	S				
LC004-1	<i>C. maxima</i>	Viet Nam	2006	A	S				

Table S2. Metrics at some sequencing and bioinformatics steps of the pipeline applied on the dataset.

Strain	SWIO	Mapping on Xcc 306 for SNP calling		De novo assembly using all reads		De novo assembly using reads unmapped on pacbio genomes	
		Mean coverage	Coverage std	De novo contigs	De novo contig length	De novo contigs	De novo contig length
JJ155	no	56	22	238	5195026		
JK101-1	no	77	31	179	5228129		
LM158	no	72	26	3257	6234579		
LM169	no	69	23	178	5448044		
LM180	no	76	28	188	5420695		
LM184	no	74	27	329	5272609		
LM198	no	57	21	259	5251083		
LM199	no	83	29	185	5421792		
LM205	no	73	28	172	5435765		
LM229	no	90	31	167	5415366		
JW160	no	753	212	384	5369368		
LG102	no	113	57	635	5357379		
LG117	no	122	38	284	5244134		
LG97	no	120	47	616	5271594		
LG98	no	105	33	334	5176559		
CFBP 2865	no	65	26	170	5160296		
CFBP 2908	no	73	30	194	5229104		
FDC1083	no	842	211	3182	5449210		
FDC217	no	866	203	2574	5423679		
JK4-1	no	110	37	252	5205526		
JN564	yes	77	24	239	5224619	0	0

LA087-2	yes	62	21	199	5217317	18	4425
LK145-10	yes	80	36	222	5245581	17	20494
LK126-01	yes	217	51	391	5283672	276	75286
LK126-02	yes	229	52	1553	5664588	1430	454048
LK126-03	yes	259	58	279	5249884	150	40535
LK127-01	yes	208	46	495	5313661	380	109623
LK127-03	yes	246	54	173	5213109	58	15551
LK128-01	yes	239	53	692	5376029	587	170401
LK128-02	yes	232	60	1199	5550712	1066	336817
LK129-03	yes	271	59	1073	5490599	949	283118
LK144-08	yes	314	66	151	5239308	17	20702
LK145-07	yes	268	58	132	5237470	15	21353
LK145-14	yes	292	63	139	5246157	18	21854
LK148-03	yes	239	53	139	5239190	12	20866
LL068-04	yes	282	60	133	5213205	19	5671
LL068-06	yes	300	64	119	5210319	17	4671
LL068-07	yes	299	60	149	5218204	54	14510
LK130-04	yes	246	51	118	5229954	16	19855
LK130-06	yes	272	58	227	5262575	39	25589
LK130-09	yes	287	64	129	5195252	11	17737
LK130-11	yes	596	127	1016	5504985	892	299825
LK131-01	yes	279	60	147	5216743	12	3864
LK131-04	yes	340	79	159	5214327	12	3649
LK131-10	yes	288	71	129	5215793	7	2323
LK132-03	yes	269	57	202	5249083	39	25638
LK132-08	yes	304	65	113	5172335	14	18462
LK135-03	yes	329	70	1084	5517186	953	299174
LK136-01	yes	290	73	133	5192502	11	18027
LK136-04	yes	291	63	137	5196697	19	20875
LK136-05	yes	272	63	190	5211363	24	21396

LK136-08	yes	283	62	131	5195940	9	17908
LK137-01	yes	332	68	158	5239681	12	18382
LK137-02	yes	242	53	202	5251473	23	21502
LK141-03	yes	247	53	141	5229478	13	18449
LK141-09	yes	477	107	1194	5516271	1062	304776
LK141-15	yes	235	52	142	5228596	15	19404
LK142-04	yes	343	85	1203	5537182	1078	319173
CFBP 2857	no	87	30	196	5204472		
LMG9322	no	746	201	8525	6092421		
CFBP 1209	no	99	38	211	5271668		
CFBP2852	no	397	103	216	5283443		
JJ165	no	72	26	203	5234107		
NCPPB 211	no	64	24	189	5211620		
NCPPB3562	no	803	240	3658	5853467		
NCPPB3610	no	93	35	591	5228916		
NCPPB3612	no	120	42	397	5348936		
CFBP 2855	no	53	24	244	5291240		
JJ238-04	no	637	136	1108	5563550		
JJ238-06	no	278	61	467	5382455		
JJ238-07	no	237	49	148	5281487		
JJ238-08	no	274	56	575	5395840		
JJ238-09	no	219	52	466	5423022		
JJ238-10	no	1031	236	2685	5472200		
JJ238-11	no	644	124	1170	5567469		
JK004-04	no	69	25	184	5268758		
LC80	no	855	296	3910	5849894		
LD7-1	no	67	27	812	5382142		
LE116-1	no	96	37	731	5405815		
LL074-04	yes	85	31	299	5436965	128	33406
LL111-06	yes	65	23	185	5402596	16	3663

LL124-01	yes	115	37	203	5404458	34	8126
LM090-02	yes	105	35	216	5411873	74	18043
LP030-1	yes	248	49	247	5436203	124	35052
LP031-1	yes	257	54	329	5461852	215	60355
LP032-1	yes	274	57	561	5521948	441	123131
LP033-1	yes	243	64	705	5563245	595	167373
LP034-1	yes	238	54	705	5572381	581	172143
LP035-1	yes	200	43	646	5549568	546	152498
LP036-1	yes	542	135	545	5533764	419	122358
JJ009-04	yes	221	35	406	5238734	274	76477
JJ009-1	yes	90	32	206	5251715	20	44379
JJ009-8	yes	80	28	194	5208656	14	3304
JK161	yes	235	55	289	5260281	141	55678
JK167-1	yes	69	27	193	5224073	12	2765
JK169	yes	199	40	427	5284908	334	92812
JK170	yes	196	46	290	5287304	150	54593
LM053-06	yes	217	46	515	5321550	379	115522
LM053-07	yes	592	127	1121	5495592	979	284959
LM054-06	yes	270	68	1213	5539226	1062	331817
LM054-17	yes	202	53	353	5267940	214	61867
LM055-08	yes	52	22	189	5202045	10	2301
LM057-04	yes	207	57	246	5238478	114	31613
LM057-14	yes	198	43	721	5370261	589	167623
LM057-15	yes	211	44	619	5331540	468	127995
LM069-01	yes	231	49	426	5270842	284	76459
LM070	yes	211	47	387	5255811	251	68102
LJ125-01	yes	277	60	134	5228117	12	18878
LJ225-01	yes	303	65	131	5188009	15	19602
LJ225-2	yes	317	69	142	5190808	21	21322
LJ226-02	yes	324	71	128	5212521	9	2876

LJ228-1	yes	264	55	129	5212442	14	4142
LJ229-06	yes	80	31	209	5216898	18	4252
LJ230-03	yes	457	99	1024	5488353	868	270243
LJ231-09	yes	317	67	190	5243977	21	21281
LJ232-01	yes	255	55	134	5212684	11	2768
LJ232-05	yes	307	62	141	5217379	18	6996
CFBP 2525	no	84	31	174	5194458		
CFBP 2853	no	75	28	196	5158506		
NCPPB 226	no	74	28	167	5148748		
NCPPB406	no	75	25	225	5210116		
JJ238-16	no	79	31	194	5323563		
B25	yes	60	28	229	5205798	17	3936
C20	yes	75	29	204	5209532	15	3568
CFBP1814	yes	46	51	214	5212783	6	15609
D02	yes	38	48	273	5171588	11	2470
D07	yes	73	27	191	5189763	15	3555
JA159-1	yes	88	33	154	5195276	14	3168
JB003-5	yes	59	23	249	5243081	65	16390
JM027-2	yes	60	45	237	5209457	27	6415
JP541	yes	337	77	2877	5906607	2737	710798
JP608	yes	192	46	147	5207582	7	1578
JP637-01	yes	69	28	223	5211993	13	2926
JP668	yes	211	49	129	5190659	3	582
JP669	yes	210	63	161	5197100	15	3993
JP710-06	yes	227	54	160	5205489	7	1631
JP727	yes	440	114	366	5262330	258	70117
JP764-01	yes	195	57	139	5217805	6	1785
JQ612-17	yes	385	88	436	5261194	282	77675
JQ613-01	yes	101	33	218	5207735	25	6029
JS502-01	yes	221	63	188	5242590	37	9644

JS503-05	yes	233	58	137	5234279	5	1306
JS536-01	yes	259	67	171	5238862	12	3117
JS538-02	yes	78	34	224	5247072	25	6005
JS538-02bis	yes	219	53	148	5233361	4	709
JS858-01	yes	92	35	194	5253786	13	3127
LG136-04	yes	73	29	228	5213087	37	9416
LH201	yes	111	38	164	5399155	16	3984
LH212	yes	248	65	160	5209003	9	2194
LH221	yes	198	46	130	5229435	3	423
LH238	yes	290	66	483	5297347	329	92988
LH240	yes	205	47	137	5196196	2	534
LH241	yes	68	24	155	5185990	11	2573
LH276	yes	60	22	179	5437012	8	1605
LH320	yes	207	46	144	5202231	7	1779
LH322	yes	305	68	545	5329635	374	114660
LI070-01	yes	401	87	1080	5705515	963	306718
LI070-02	yes	379	101	3852	6372866	3736	972520
LI070-03	yes	209	50	580	5523024	472	124614
LI070-05	yes	322	74	2210	6155145	2091	753393
LI214-01	yes	309	68	195	5206156	56	13930
LI214-02	yes	221	53	1069	5686375	945	285159
LI214-05	yes	250	59	1456	5574756	1322	382462
LI214-07	yes	234	58	889	5400632	759	209427
LI214-09	yes	266	61	534	5524635	413	118799
LI214-10	yes	280	63	740	5357852	612	168865
LI214-14	yes	384	83	2456	6010962	2368	822916
LI214-16	yes	370	77	2484	6269409	2368	861767
LJ207-7	yes	62	24	199	5442152	12	2753
LK169-01	yes	71	25	202	5208603	15	3438
LK169-03	yes	69	27	208	5147799	15	3663

LK169-04	yes	67	27	217	5153343	28	7201
LK170-01	yes	66	27	234	5158077	40	10053
LK170-03	yes	55	22	202	5160312	14	3434
LK170-05	yes	71	28	212	5209977	17	4073
LK171-01	yes	71	29	239	5061017	20	4786
LK171-02	yes	80	27	219	5207395	14	3217
LK172-01	yes	88	38	235	5213585	22	5375
LK172-02	yes	83	31	210	5205101	11	2715
LK173-01	yes	85	30	192	5201892	10	2193
LK173-02	yes	83	29	212	5206643	14	3223
LL082-03	yes	194	43	127	5229051	4	923
LL082-04	yes	200	42	264	5439692	155	41356
LL082-13	yes	195	48	139	5227384	4	930
LL082-15	yes	215	45	551	5522659	442	123768
LL082-16	yes	237	59	288	5270096	143	36961
LL082-18	yes	189	52	271	5438716	152	41208
LL082-21	yes	520	129	2015	5775146	1793	515737
LL082-25	yes	229	54	137	5234866	4	895
LL085-02	yes	191	47	288	5436182	164	43410
LL085-03	yes	252	56	156	5193774	12	2965
LL087-01	yes	567	140	373	5299175	222	61965
LL087-02	yes	213	51	135	5166857	5	1142
LL095-02	yes	204	47	141	5227889	6	1736
LL095-08	yes	527	115	1387	5673185	1214	434946
LL095-20	yes	257	58	650	5371260	484	129922
LL096-06	yes	216	50	126	5234896	2	392
LL096-08	yes	592	128	652	5335418	489	136065
LL096-09	yes	258	63	298	5266690	156	40067
LL096-13	yes	204	59	135	5230076	5	1158
LL098-02	yes	213	59	141	5231553	9	2122

LL100-12	yes	186	57	129	5238767	5	1107
LL102-02	yes	237	56	139	5235699	5	1001
LL115-09	yes	198	48	484	5332299	337	109003
LL115-17	yes	208	50	145	5227759	3	470
LL115-20	yes	226	54	149	5230944	2	400
LL131-03	yes	565	129	952	5426850	772	215902
LL132-01	yes	262	58	731	5382782	597	187570
LL162-01	yes	530	122	908	5383033	742	206627
LL164-02	yes	273	68	147	5232522	5	1262
LL174-01	yes	196	54	146	5172635	3	644
LL174-02	yes	331	75	724	5417932	613	182668
LL174-04	yes	223	57	139	5204336	6	1288
LL174-05	yes	211	62	149	5171257	6	1670
LL175-01	yes	265	65	1835	5981205	1713	579885
LL186-5	yes	88	31	205	5242604	15	3533
LM088-18	yes	239	64	151	5204691	4	923
LM088-20	yes	68	27	194	5417559	14	3432
LM088-25	yes	255	61	275	5233960	160	44026
LM088-28	yes	253	60	990	5637629	843	237888
LM088-37	yes	210	48	460	5487568	344	96468
LM088-40	yes	213	58	133	5189254	3	543
LM088-42	yes	243	57	774	5575343	637	169175
LM089-02	yes	63	24	281	5456315	96	24413
LM089-06	yes	216	52	132	5184415	2	314
LM089-20	yes	213	60	118	5183355	3	589
LM089-41	yes	223	64	158	5190325	5	1228
LM089-42	yes	201	50	438	5486643	315	86559
LM121-01	yes	87	33	182	5407337	15	3615
LM121-02	yes	312	76	657	5546556	539	148070
LM121-31	yes	234	60	169	5236240	8	2085

LM121-41	yes	327	72	747	5574100	612	170449
LM121-42	yes	184	42	651	5544017	531	144918
LM358-02	yes	576	129	2214	5821501	1973	584378
LM358-08	yes	333	81	197	5185210	51	13341
LM358-13	yes	231	57	142	5208738	6	1398
LM358-18	yes	206	53	138	5200121	4	906
LM358-22	yes	274	63	144	5202525	3	632
LN003-10	yes	225	57	155	5209813	7	1927
LN003-17	yes	197	50	150	5204738	6	1418
LN005-04	yes	232	54	133	5192513	3	644
LN005-07	yes	207	50	555	5317281	415	130517
LN006-01	yes	234	64	508	5511089	378	104564
LN006-18	yes	178	56	362	5501186	238	64355
LN007-02	yes	240	63	175	5202286	8	1819
LN007-03	yes	278	66	1698	5667230	1549	470298
LN007-05	yes	208	61	838	5606340	711	199934
LN007-07	yes	222	49	251	5434764	127	33710
LN007-09	yes	221	62	511	5511046	389	105919
LN007-10	yes	245	63	204	5429774	83	19419
JJ010-02	yes	255	52	140	5238295	14	4887
JJ010-4	yes	70	31	201	5159179	19	18236
JJ10-1	yes	171	54	237	5184891	0	0
LC045-02	yes	304	75	225	5261632	31	9119
LC046	yes	367	83	137	5230998	21	5578
LC048	yes	90	32	209	5244920	24	18928
LI139-03	yes	293	62	138	5247967	22	20541
LI139-05	yes	336	69	127	5232919	15	4295
LI139-6	yes	61	22	204	5251668	25	20011
LM095-04	yes	268	64	153	5169179	18	4844
LM095-05	yes	250	59	221	5264080	36	24724

LM095-07	yes	296	65	145	5250060	12	18210
LM095-11	yes	212	48	135	5190134	7	17030
LM095-14	yes	307	76	146	5201982	10	2466
LM096-03	yes	256	59	153	5183054	18	4619
LM096-08	yes	388	86	194	5159809	31	10654
LM096-09	yes	288	68	235	5217841	101	26312
LM097-01	yes	282	64	134	5241994	14	18757
LH37-1	no	120	38	351	5399521		
JZ092	no	90	35	213	5283006		
JZ094	no	231	51	213	5210607		
LB100-1	no	135	43	267	5265459		
LB100-3	no	98	41	212	5294050		
LJ001-1	no	209	50	435	5357904		
LJ003-1	no	58	27	209	5273014		
LP027-03	no	204	46	271	5224064		
LP027-05	no	253	56	596	5317100		
LP027-13	no	222	49	342	5239147		
LP028-03	no	211	48	525	5302493		
LP028-05	no	227	52	965	5423067		
LP028-06	no	224	50	1317	5559283		
LP029-13	no	659	155	683	5428207		
LP029-15	no	215	52	587	5406109		
JJ053-06	no	88	35	219	5222872		
LC004-1	no	90	54	265	5246630		

Annexe VI : Informations supplémentaires publication 3 *The structure of bacterial communities as seen through the sharing of Xanthomonadaceae plasmid genes*

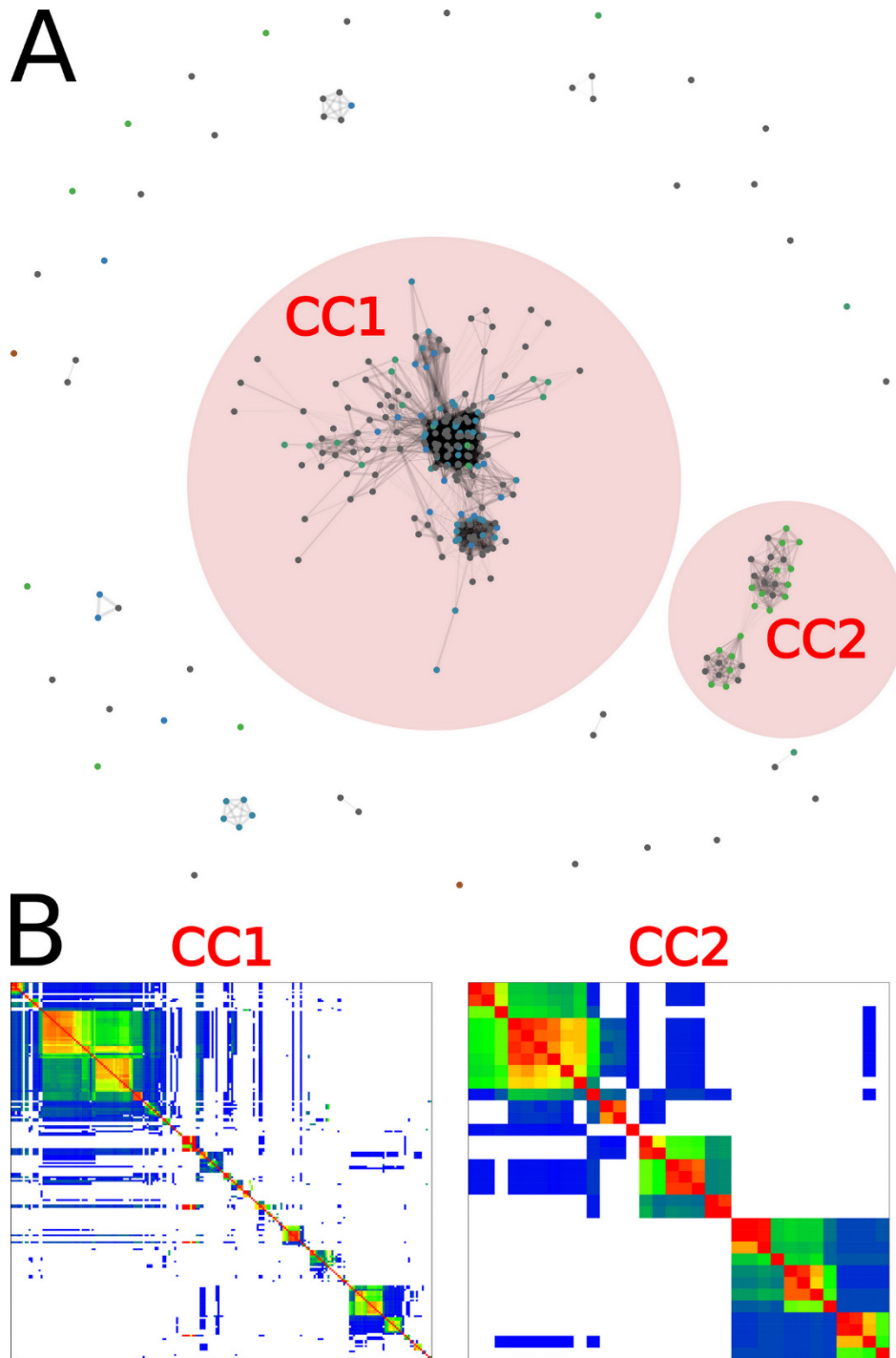


Fig. S1. (A) Network of the Xanthomonadaceae plasmid relatedness. Each node represents one of the 305 query plasmid and edges between nodes indicate the sharing of gene(s) cluster(s) between those plasmids. Nodes are colored according to the taxonomy of the bacteria from which the plasmid was obtained. Color key is depicted on Figure 1. (B) Matrices of the proportion of genes shared between plasmids are depicted for the two main connected component (CC) as indicated on panel A. Colors vary from blue to red according to the low to high proportion of genes shared, respectively. The white region indicates the absence of gene sharing.

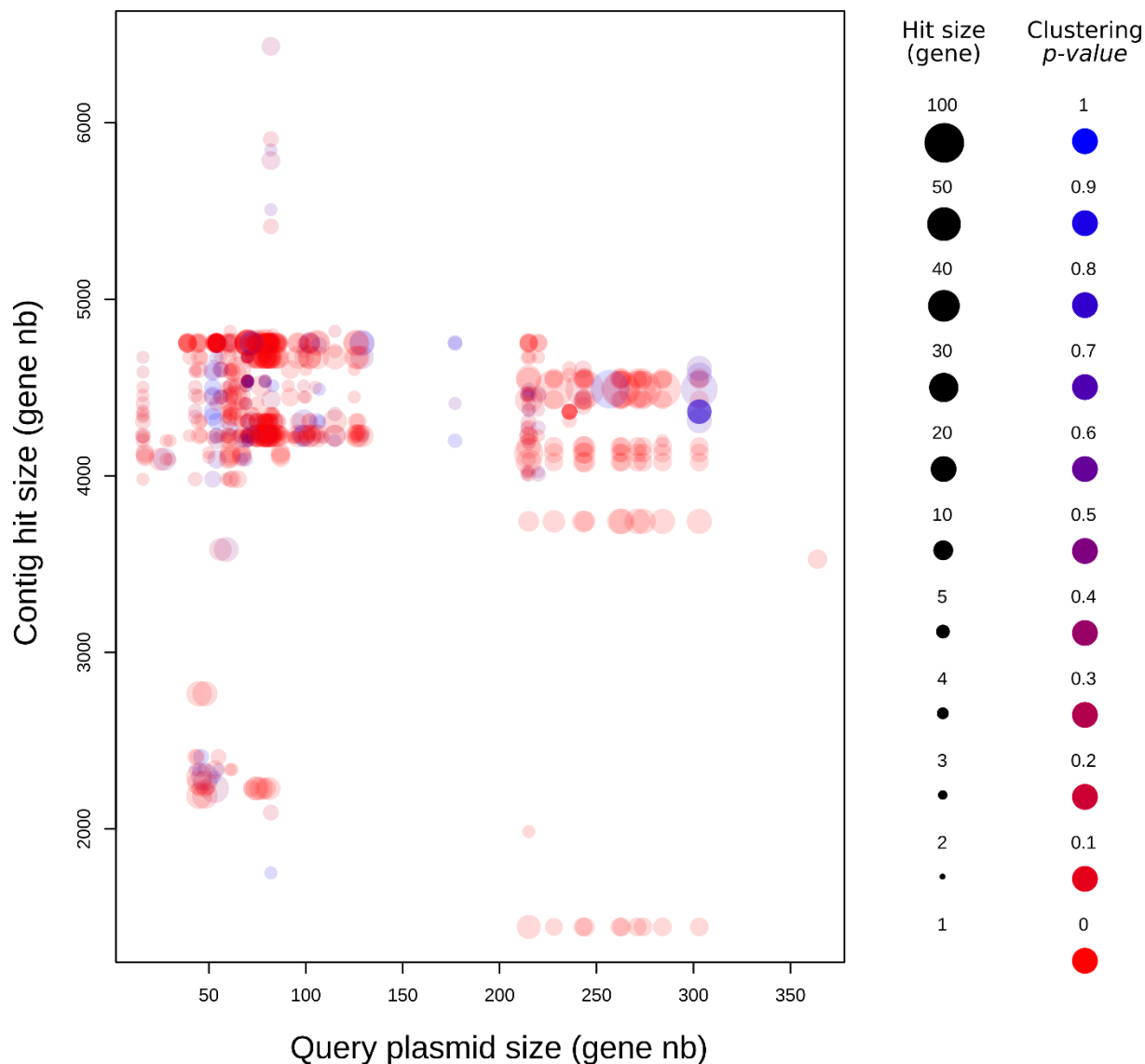


Fig. S2. Representation of all the putative gene sharing events between plasmid and chromosome. Each circle on the graph represents a putative event between plasmids (x-axis) and chromosomal contigs (*i.e.* contigs including more than 1,000 genes, y-axis). The circle size depends on the number of genes involved in the sharing event and its color depends on the p-value of the clustering test of these genes on the query plasmid. Reddish colors indicate gene sharing events involving genes whose proximity is conserved in both a plasmid and a chromosome, potentially indicative of the presence of a cassette.

Table S1 Description of the Xanthomonadaceae plasmid database

Cette table, qui comporte 17 colonnes et 307 lignes, n'est pas incluse dans le présent manuscrit.

Table S2 Gene sharing network statistics and composition depending on the clustering threshold.

Thresholds (% identity / % length)	90/90	95/90	95/95	98/90	98/98	99.8/90
Individuals	10112	9788	2762	9562	2229	9359
Individuals after merging	2063	1785	1286	1591	1055	1422
Genes	87421	73873	69991	61250	57518	54711
Contigs	14576	11107	9482	8475	6727	6526
Gene families	6399	6924	7484	7628	8775	8197
Xanthomonadaceae (%)	952 (9)	812 (8)	805 (29)	658 (7)	748 (34)	621 (7)
Xanthomonadaceae merged (%)	837 (41)	703 (39)	686 (53)	612 (38)	606 (57)	550 (39)
<i>Xanthomonas</i> (merged)	656 (596)	652 (564)	647 (549)	507 (481)	597 (479)	481 (435)
<i>Xanthomonas</i> (% nr and wgs)	84	83	83	65	76	62
<i>Xylella</i> (merged)	47 (43)	47 (43)	47 (43)	46 (43)	46 (41)	46 (39)
<i>Stenotrophomonas</i> (merged)	186 (144)	104 (87)	102 (85)	99 (82)	99 (80)	92 (74)
<i>Pseudoxanthomonas</i> (merged)	18 (18)	5 (5)	5 (5)	2 (2)	2 (2)	0 (0)
<i>Lysobacter</i> (merged)	31 (24)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Xanthomonadaceae edges	6801508	3678677	3256983	2212130	1928016	1689829
Total edges	8400972	5102744	3640932	3470358	2155648	2733968
% of Xanthomonadaceae edges	81	72	89	64	89	62

Résumé

Le cuivre, souvent utilisé pour gérer les bactérioses en agriculture, est largement utilisé dans la lutte contre *Xanthomonas citri* pv. *citri* (Xcc), agent responsable du chancre Asiatique des agrumes. La récente détection d'un phénotype résistant au cuivre (CuR) chez Xcc dans deux territoires ultramarins Français a motivé une étude génomique qui a révélé, dans les génomes de souches CuR, la présence d'un plasmide conjugatif portant un transposon adaptatif de type Tn3. Sa conservation chez plusieurs espèces de *Xanthomonas* phytopathogènes suggère le rôle des transferts horizontaux (HGT) dans l'adaptation de Xcc. Nous avons donc analysé, dans l'océan Indien, les relations phylogénétiques de souches sensibles et CuR en prenant en compte à la fois les SNP et les variations de contenu en gènes. La datation de la phylogénie a permis de formuler des scénarii d'introduction de la bactérie dans la région. La phylogénie a montré une structure géographique forte à l'échelle de l'Océan Indien, qui s'estompe à l'échelle de la Réunion et disparaît à l'échelle du verger. Au sein des vergers, l'admixture est un élément favorable aux HGT entre souches génétiquement différentes. Ils sont pourtant peu caractérisés chez les bactéries du genre *Xanthomonas*. Nous avons ainsi analysé la dispersion de l'ensemble des gènes plasmidiques connus de la famille des Xanthomonadaceae dans l'ensemble des génomes bactériens de NCBI, mettant en évidence à la fois la forte prévalence des gènes plasmidiques au sein des Xanthomonadaceae mais aussi la limite taxonomique forte à leur échange par conjugaison. L'importance du mosaïsme plasmidique, en partie lié aux éléments mobiles a aussi été illustrée. L'ensemble de nos résultats souligne l'importance des HGT dans l'évolution des bactéries du genre *Xanthomonas*, et la nécessité de caractériser finement le contenu et le fonctionnement du génome environnemental des Xanthomonadaceae pour appréhender au mieux l'adaptation de ces bactéries phytopathogènes.

Mots-clés : *Xanthomonas citri* pv. *citri* ; chancre asiatique des agrumes ; bactérie monomorphe ; adaptation bactérienne; résistance au cuivre ; transfert horizontal de gènes ; phylogénie datée ; plasmide

Abstract

Copper, frequently used in agriculture to control bacterial diseases, is commonly used against *Xanthomonas citri* pv. *citri* (Xcc), the bacterial agent of Asiatic citrus canker. The recent detection of a copper-resistant phenotype in two French overseas regions motivated a genomic study which revealed, in copper-resistant (CuR) strains, a conjugative plasmid encoding an adaptive transposon of the Tn3 family. Its conservation in several *Xanthomonas* species suggested the role of horizontal gene transfer (HGT) in Xcc adaptation. We therefore analyzed the evolutionary history of susceptible and CuR Xcc strains in the Indian Ocean using both SNP and gene content variations. The dating of the obtained phylogeny allowed us to hypothesize the history of Xcc introduction into the region. The phylogeny showed a strong geographic structure among islands of the Indian Ocean region, which faded at the Réunion scale and disappeared at the grove scale. Among the groves, admixture is a factor favoring HGT between genetically distinct strains. This form of evolution is however largely uncharacterized in the *Xanthomonas* genus. To fill this gap, we searched genetic homology between the whole known plasmid gene content of the Xanthomonadaceae family and the complete set of genomes hosted in NCBI databases. We highlighted both the ubiquity of plasmid genes in the Xanthomonadaceae family and the taxonomical barrier of their sharing by conjugation. The small fraction of genes that were exchanged through the complete sharing of plasmids also revealed the importance of plasmid mosaicism, partly due to mobile genetic elements.

Taken together, our results highlight the importance of bacterial communities in the evolution of phytopathogenic bacteria of the *Xanthomonas* genus, and the need for a precise characterization of the content and the functioning of the Xanthomonadaceae environmental genome in order to fully apprehend the adaptation of these phytopathogenic bacteria.

Keywords: *Xanthomonas citri* pv. *citri*; asiatic citrus canker; monomorphic bacteria; bacterial adaptation; copper resistance; horizontal gene transfer; dated phylogeny; plasmid

LETTRE D'ENGAGEMENT DE NON-PLAGIAT

Je, soussigné **Damien Richard** en ma qualité de doctorant de l'Université de La Réunion, déclare être conscient que le plagiat est un acte délictueux passible de sanctions disciplinaires. Aussi, dans le respect de la propriété intellectuelle et du droit d'auteur, je m'engage à systématiquement citer mes sources, quelle qu'en soit la forme (textes, images, audiovisuel, internet), dans le cadre de la rédaction de ma thèse et de toute autre production scientifique, sachant que l'établissement est susceptible de soumettre le texte de ma thèse à un logiciel anti-plagiat.

Fait à Saint-Denis le : **28.01.19**

Signature :



Extrait du Règlement intérieur de l'Université de La Réunion
(validé par le Conseil d'Administration en date du 11 décembre 2014)

Article 9. Protection de la propriété intellectuelle – Faux et usage de faux, contrefaçon, plagiat

L'utilisation des ressources informatiques de l'Université implique le respect de ses droits de propriété intellectuelle ainsi que ceux de ses partenaires et plus généralement, de tous tiers titulaires de ces droits.

En conséquence, chaque utilisateur doit :

- utiliser les logiciels dans les conditions de licences souscrites ;
- ne pas reproduire, copier, diffuser, modifier ou utiliser des logiciels, bases de données, pages Web, textes, images, photographies ou autres créations protégées par le droit d'auteur ou un droit privatif, sans avoir obtenu préalablement l'autorisation des titulaires de ces droits.

La contrefaçon et le faux

Conformément aux dispositions du code de la propriété intellectuelle, toute représentation ou reproduction intégrale ou partielle d'une œuvre de l'esprit faite sans le consentement de son auteur est illicite et constitue un délit pénal.

L'article 444-1 du code pénal dispose : « Constitue un faux toute altération frauduleuse de la vérité, de nature à causer un préjudice et accomplie par quelque moyen que ce soit, dans un écrit ou tout autre support d'expression de la pensée qui a pour objet ou qui peut avoir pour effet d'établir la preuve d'un droit ou d'un fait ayant des conséquences juridiques ».

L'article L335_3 du code de la propriété intellectuelle précise que : « Est également un délit de contrefaçon toute reproduction, représentation ou diffusion, par quelque moyen que ce soit, d'une œuvre de l'esprit en violation des droits de l'auteur, tels qu'ils sont définis et réglementés par la loi. Est également un délit de contrefaçon la violation de l'un des droits de l'auteur d'un logiciel (...) ».

Le plagiat est constitué par la copie, totale ou partielle d'un travail réalisé par autrui, lorsque la source empruntée n'est pas citée, quel que soit le moyen utilisé. Le plagiat constitue une violation du droit d'auteur (au sens des articles L 335-2 et L 335-3 du code de la propriété intellectuelle). Il peut être assimilé à un délit de contrefaçon. C'est aussi une faute disciplinaire, susceptible d'entraîner une sanction.

Les sources et les références utilisées dans le cadre des travaux (préparations, devoirs, mémoires, thèses, rapports de stage...) doivent être clairement citées. Des citations intégrales peuvent figurer dans les documents rendus, si elles sont assorties de leur référence (nom d'auteur, publication, date, éditeur...) et identifiées comme telles par des guillemets ou des italiques.

Les délits de contrefaçon, de plagiat et d'usage de faux peuvent donner lieu à une sanction disciplinaire indépendante de la mise en œuvre de poursuites pénales.