



HAL
open science

Évolution génomique au sein d'une population naturelle de *Streptomyces*

Abdoul-Razak Tidjani

► **To cite this version:**

Abdoul-Razak Tidjani. Évolution génomique au sein d'une population naturelle de *Streptomyces*. Ecosystèmes. Université de Lorraine, 2019. Français. NNT : 2019LORR0159 . tel-02461512

HAL Id: tel-02461512

<https://hal.univ-lorraine.fr/tel-02461512>

Submitted on 22 Jun 2020

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

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



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>



UNIVERSITÉ
DE LORRAINE

SIReNa



INRA
SCIENCE & IMPACT

Ecole Doctorale SIReNa - Sciences et Ingénierie des Ressources Naturelles

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de
DOCTEUR DE L'UNIVERSITE DE LORRAINE

Mention : « **Ecotoxicologie, Biodiversité, Ecosystèmes** »

Par **Abdoul-Razak TIDJANI**

Evolution génomique au sein d'une population naturelle de *Streptomyces*

Soutenance publique prévue le Mardi 03 décembre 2019

Membres du jury :

Rapporteurs :

Dr. Philippe NORMAND

Directeur de recherche CNRS, Lyon

Pr. Dr. Gregory VELICER

Professeur, *Eidgenössische Technische Hochschule (ETH), Zürich*

Examineurs :

Pr. Stéphanie BURY-MONE

Professeure, Université Paris-Sud

Dr. Catherine MASSON-BOIVIN Directrice de recherche INRA, Toulouse

Dr. Cyril BONTEMPS

Maître de conférences, Université de Lorraine, Co-directeur de thèse

Pr. Pierre LEBLOND

Professeur, Université de Lorraine, Directeur de thèse

UMR 1128 DynAMic

INRA-Université de Lorraine, Faculté des Sciences et Technologies,
Boulevard des Aiguillettes, BP 70239, 54506 Vandœuvre-lès-Nancy

« Le commencement de toutes les sciences, c'est l'étonnement de ce que les choses sont ce qu'elles sont »

Aristote

Dédicaces

A ma Mère Assia et à mon Père Mashoud,

Pour avoir toujours cru en moi, pour votre soutien indéfectible et pour tous les sacrifices consentis jusqu'à présent à mon égard.

A mes sœurs, Kadirra, Medina et à mes frères, Wassihou, Nazif, Rachid,

Loin des yeux près du cœur dit-on souvent. Mais vous, vous ne m'avez jamais perdu de vue. Et je comprends maintenant l'importance de tous ces messages quotidiens que nous échangeons.

A Maud,

Trois années de thèse c'est long, tout comme les 1018 km parcourus les week-ends pour nous voir. Merci pour ta patience, ta persévérance et ton indulgence au cours de ces presque 5 dernières années.

REMERCIEMENTS

I would like to thank Prof. Gregory Velicer, Dr. Philippe Normand, Dr. Catherine Masson and Prof. Stéphanie Bury-Moné for accepting to review this manuscript.

J'aimerais remercier Mr le Professeur Pierre Leblond et Mr le Docteur Cyril Bontemps. Le plus gratifiant est d'avoir eu l'honneur de vous avoir avant tout comme mentors, et puis comme directeur et co-directeur de thèse. Ces années passées à me guider sur ce chemin sinueux mais plein de promesses que représente le doctorat feront parties des plus riches et des plus intenses de ma vie. J'aimerais à travers ces modestes lignes vous dire merci(s) ; un pour chacun des 1126 jours passés à apprendre à vos côtés (week-ends compris).

Je voudrais remercier le Pr Bertrand Aigle, pour son accueil (chez lui alors que je venais juste de poser mes valises à Nancy) et au labo ainsi que pour les multiples échanges scientifiques ou non que nous avons eus.

Un grand merci à ma Lolo. Pour m'avoir initié à la grande messe que nous vouons à *Streptomyces* et pour ta disponibilité lorsque le besoin se faisait sentir. Merci à Emilie Pio pour tous ces moments intra et extra labo. Tu as été comme une bouffée d'air frais pendant cette thèse. Merci à mon frère Stéphane, pour ta présence, ta sollicitude et le fait d'être toi. Un grand merci également à Louise (pour les « on mange ? » de chaque midi), à Emilie Robert, à Johan et à Anthony. Ces 3 années n'auraient pas été pareilles sans vous tous. Merci à toi Layla (guide omniscient dans les méandres administratifs) pour ta sollicitude.

Je voudrais remercier mes amis ~~thésards~~ doctorants : Milena, Matthieu, Julie, Li, Haïfa, et Caroline (je viens de remarquer que vous êtes arrivés au labo inversement chronologiquement à l'ordre alphabétique de vos prénoms. Oui je sais ! on ne change pas un scientifique). Merci pour ces moments passés ensemble dans la quête de ce « précieux ». Bon courage à vous. Sans oublier mes amis Docteurs, Maxime pour les moments de partage au cours de ta dernière année, Charles, Narimane, Clàudia, Greg, Xavier (pour les moments d'évasion scientifique PS4, d'exploration de débit de boissons jus d'ananas à Nancy), Yaouba, Estefania, Yordenka et Sonia. Je n'oublie pas dans cette équation complexe de remerciements les doctorantes du « labo d'à côté », Katleen et Rania. Et je terminerai par les « djeuns » (par ordre alphabétique) : Abbas, Anissa, Anthony, Camille, Louise, Maha et Redouane.

J'aimerais remercier tous les membres du laboratoire, permanents et non, avec qui j'ai eu à partager une partie ou toutes ces 3 dernières années. Et puis ... allez la farandole pour tout le monde.... Merci à Claire pour ton écoute et ta disponibilité ; merci à toi virginie pour ces moments de partage tant au labo qu'en extra labo ; Merci à Gérard pour nos discussions scientifiques et non, ces moments qui font « qu'on ne sent pas la thèse passer » ; merci à Annabelle (eh oui Artémis ça se transmet), Nicolas, Sophie, Nathalie, Yvonne, Florence, Badreddine pour ces discussions enrichissantes lors des repas de midi et pendant mes déambulations quotidiennes (que j'ose appeler pauses) dans les bureaux.

LISTE DES ABBREVIATIONS

°C : Degrés Celsius

μl : microlitre

μM : micromolaire

A : Adenylation

ACP : Acyl Carrier Protein

ACV : L-aminoadipyl-L-cysteinyl-D-valine

ADN : Acide désoxyribonucléique

ADNg : Acide désoxyribonucléique génomique

AFLP : Amplified Fragment Length Polymorphism

AICE : Actinomycete Integrative and Conjugative Element

ANI : Average nucleotide identity

antiSMASH : antibiotics and Secondary Metabolite Analysis Shell

ARN : Acide ribonucléique

ARNr : Acide ribonucléique ribosomique

AT : Acyltransferase

attB : site d'attachement bactérien

attP : site d'attachement plasmidique

BET : Bromure d'Ethidium

BIR : Break Induced Replication

C : Condensation

CGSMS : Clusters de Gènes de Synthèse de Métabolites Spécialisés

cm : Centimètre

CRISPR : Clustered Regularly Interspaced Short Palindromic Repeats

dNTP : Désoxyribose nucléoside triphosphate

DO : Densité optique

Hfr : Haute fréquence de recombinaison

ICE : Integrative and Conjugative Element

kb : Kilobases

l : Litres

Mb : Mégabases

MEB : Microscope Electronique à Balayage

mg : Milligramme

ml : millilitre

MLST : MultiLocus Sequence Typing

mM : Millimolaire

NCBI : National Center of Biotechnology

NGS : Next Generation Sequencing technologies

NRP : Non-Ribosomal Peptide

NRPS : Non ribosomal Protein Synthase

nt : Nucléotide

ORF : Open Reading Frame

oriT : origine de Transfert

PCR : Polymerase Reaction Chain

pH : Potentiel d'hydrogène

PKS : Polyketide Synthase

PSAM2 : *Streptomyces ambofaciens* plamid 2

RAST : Rapid Annotation using Subsystem Technology

REP-PCR : Repetitive Element Polymerase Reaction Chain

RFLP : Restriction Fragment Length Polymorphism

RIPP : Ribosomally synthesized and Post-translationally modified Peptide

RPM : Rotations Par Minutes

SDS : Sodium Dodecyl Sulfate

SFM: Soya Floor Manitol

SNP : Single Nucleotide Polymorphism

ST : Sequence Type

T4SS : Type IV Secretion System

TAP : Telomere Associated Protein

Te : Thioesterase

TIR : Terminal Inverted Repeats

TP : Terminal Protein

UTO : Operational Taxonomic Unit

SOMMAIRE

SOMMAIRE	11
Préambule	15
INTRODUCTION	17
I. Un modèle d'étude : Le micro-habitat sol forestier	19
I.1 Importance de l'échelle dans l'étude des bactéries du sol.	21
I.2 Etude de la diversité et de la répartition des bactéries à la micro-échelle.....	22
I.2.1 L'étude des habitats à la micro-échelle et de leurs bactéries.....	23
I.2.2 Le microcosme.....	25
I.3 La diversité génétique et fonctionnelle bactérienne à la micro-échelle.....	26
II. Concept d'espèce bactérienne et notion de population.....	29
II.1 Classification des micro-organismes	29
II.2 Concept d'espèce bactérienne.....	30
II.2.1 Le concept d'écotype	33
III. Les mécanismes d'évolution et de dynamique des génomes chez les bactéries.....	37
III.1 Le génome bactérien.....	39
III.2 La dynamique du génome bactérien.....	39
III.2.1 Variations génétiques dues au transfert horizontal de gènes	40
III.2.1.1 La transduction.....	41
III.2.1.2 La conjugaison	42
III.2.1.3 La transformation naturelle	44
III.2.1.4 Fonctions acquises par transfert horizontal de gènes	45
III.2.2 Les barrières au transfert de gène	46
III.2.2.1 Les barrières au transfert de gène	46
III.2.2.2 Les barrières à l'intégration et au maintien.....	47
III.2.2.3 Les barrières à l'expression génique	48
III.2.3 Les réarrangements chromosomiques.....	49
III.2.4 La variation par mutation	50
III.2.4.1 Les mutations spontanées.....	51
III.2.4.2 Les mutations induites	51
III.2.4.3 Les hypermutateurs.....	51
III.2.5 Évolution des fréquences alléliques dans une population.....	53
III.2.6 La notion de pangéome, génome « core » et génome accessoire.....	53
IV. Le genre <i>Streptomyces</i>	56
IV.1 <i>Streptomyces</i> et son cycle de vie.....	56

IV.2	Une organisation chromosomique atypique.....	58
IV.3	Un mécanisme de conjugaison dépendant de TraB.....	58
IV.4	Une grande variabilité inter- et intra-spécifique.....	61
IV.5	Une distribution compartimentée des gènes le long du chromosome	62
IV.6	Une instabilité génétique liée aux réarrangements chromosomiques.....	62
IV.7	Une diversité de production de métabolites spécialisés au niveau intra-spécifique.....	65
Objectifs de la thèse :.....		67
RESULTATS.....		69
I.	Stratégie d'échantillonnage et séquençage des génomes de 11 souches sympatriques et conspécifiques de <i>Streptomyces</i>	73
II.	Etude de la diversité génétique au sein d'une population de 11 souches sympatriques et conspécifiques de <i>Streptomyces</i> par l'approche de la génomique comparative.....	83
III.	Etude de la variabilité chromosomique terminale des génomes de 11 souches sympatriques et conspécifiques de <i>Streptomyces</i>	99
IV.	Etude de la diversité des éléments conjugatifs et intégratifs au sein d'une population de 11 souches sympatriques et conspécifiques de <i>Streptomyces</i>	125
IV.1	Stratégie d'identification des éléments conjugatifs et intégratifs	127
IV.2	Identification, prévalence et classification des AICE et ICE.....	129
IV.3	Sites d'insertion des AICE et ICE.....	131
IV.4	Répartition chromosomique des positions d'insertion des AICE et ICE identifiés.....	133
IV.5	Mobilité et transfert des éléments conjugatifs au sein de la population	135
V.	Etude des mécanismes régissant l'évolution rapide des clusters de gènes de synthèse de métabolites spécialisés (CGSMS).....	140
V.1	Identification et classification des clusters de synthèse de métabolites spécialisés	145
V.2	Clusters de métabolites spécialisés <i>core</i> et accessoires de la population/ <i>Prédiction du pan-métabolome et évolution</i>	146
V.1	Distribution génomique des clusters de synthèse de métabolites spécialisés	149
V.2	Dynamique évolutive des clusters de synthèse de métabolites spécialisés	150
DISCUSSION		157
I.	Analyse du pan-génome : une ouverture représentative de la diversité génomique au sein du micro-habitat.....	159
II.	La compartimentation du génome de <i>Streptomyces</i>	162
III.	Des extrémités chromosomiques fortement plastiques.....	163
	La dynamique des génomes impacte également les télomères	164
IV.	Mécanismes de la diversité génomique.....	164
V.	Mécanismes de mobilisation des marqueurs chromosomiques.....	167
VI.	Existe-t-il des barrières au transfert au sein de la population de <i>Streptomyces</i> ?.....	168

VII. Les clusters de synthèse de métabolites spécialisés (CGSMS) comme proxy de l'impact fonctionnel de la diversité génétique.....	168
VIII. Concept de population : approche écologique ou génétique?.....	170
VIII.1 « L'écologie reverse » comme approche d'identification et de définition des populations.....	171
VIII.1.1 Utilisation d'une approche avec des caractères écologiques prédéfinis	172
VIII.1.2 Utilisation d'une approche sans <i>a priori</i> sur les caractères écologiques	173
CONCLUSION	174
REFERENCES BIBLIOGRAPHIQUES	175
Annexes : Figures supplémentaires de l'article du chapitre II	201

Préambule

Les bactéries sont présentes dans tous les écosystèmes et leur diversité dépasserait de plusieurs logs celle des végétaux et des animaux. Leur nombre est également très important (par exemple un gramme de sol peut en contenir entre 10^6 et 10^9) et elles sont au cœur de nombreuses interactions biotiques assurant le bon fonctionnement de l'écosystème et son homéostasie.

Toutefois, l'organisation, la structuration et le fonctionnement des communautés bactériennes reste mal connue. La répartition spatiale des bactéries a pendant longtemps été abordée selon le principe énoncé par Bass Becking qui stipulait que « *Everything is everywhere, but, the environment selects* ». Ce concept, l'un des plus unificateurs en écologie microbienne, a été utilisé comme assertion de départ dans de nombreuses études sur la biogéographie des microorganismes et dans l'assemblage des communautés bactériennes dans l'environnement.

Peu d'études toutefois se sont intéressées à l'impact des interactions intercellulaires, pouvant exister entre les populations bactériennes, et à leurs conséquences sur la genèse de la diversité génétique globale observée. Selon Cordero & Datta (2016), les interactions bactériennes à l'échelle cellulaire impacteraient les propriétés globales des systèmes étudiés et permettraient de combler le gap existant dans la compréhension des diversités fonctionnelles observées entre les individus appartenant à une même espèce, voir à une même population.

Au cours de ces deux dernières décennies, l'amélioration des techniques de séquençage a permis d'avoir accès à des milliers de génomes de procaryotes. La comparaison des génomes a permis déterminer qu'une partie des gènes était partagée par l'ensemble des individus d'une même espèce ou population et définit le core génome de cette espèce ou population. En plus du core génome, les bactéries disposent d'un répertoire de gènes présent uniquement dans une fraction de la population qui constitue le génome accessoire. Ce génome accessoire confère des avantages adaptatifs aux bactéries qui les possèdent. Au sein d'une même espèce, la taille du génome accessoire excède le plus souvent celle du génome d'un individu, soulignant l'extraordinaire collection de gènes dont disposent les bactéries. Le génome accessoire provient des événements de pertes et de gains d'informations génétique au sein d'un phylum. L'acquisition d'information génétique auprès des individus non ascendant est un phénomène maintenant bien connu appelé transfert horizontal de gènes. Les phénomènes d'acquisition et de perte d'informations génétiques constituent des facteurs déterminant l'évolution du contenu, de la structure et de l'architecture des chromosomes.

Les bactéries du genre *Streptomyces*, de par la structure atypique de leur génome, constituent à cet égard un modèle chromosomique intéressant pour l'étude des mécanismes aboutissant à la diversité génomique. En effet, l'information accessoire peut représenter jusqu'à 75 % du génome chez cette

bactérie. Cette information est majoritairement localisée au niveau des extrémités du chromosome linéaire de cette bactérie tandis que l'information conservée est localisée dans la partie centrale. L'étendu de la diversité génomique au niveau infra-spécifique, les mécanismes qui meuvent cette diversité aux étapes initiales de la divergence génétique ainsi que les conséquences sur l'écologie des populations bactériennes restent encore peu explorées. A travers ce travail de thèse, les résultats obtenus ont permis de mettre à jour une dynamique insoupçonnée des génomes au sein d'une population de souches de *Streptomyces* ayant divergé d'un ancêtre commun récent et partageant le même micro-habitât ; La diversité génétique observée au sein de ces souches suggère une évolution rapide des génomes ayant des impacts écologiques permettant d'assurer la cohésivité de la population étudiée. Nous avons proposé le mécanisme de conjugaison comme moteur de cette diversité génétique observée.

INTRODUCTION

I. Un modèle d'étude : Le micro-habitat sol forestier

Le sol fait partie des écosystèmes les plus diversifiés en terme de contenu microbien sur Terre avec entre 10^6 à 10^9 bactéries par gramme (Whitman *et al.*, 1998; Quince *et al.*, 2008). Cet environnement joue un rôle essentiel dans la régulation de divers facteurs biogéochimiques globaux grâce aux différentes communautés microbiennes qu'il renferme. Le sol est une matrice structurellement complexe qui rend difficile l'appréhension de la diversité et de la répartition des microorganismes qu'il contient. C'est un système hétérogène qui comprend des gaz, de l'eau et des matériaux solides qui fournissent des micro-habitats variés pour les microorganismes ((Daniel, 2005)). Le sol possède ainsi un arrangement hiérarchique allant des méta-communautés à grande échelle jusqu'aux populations bactériennes formées d'individus interagissant à l'échelle cellulaire (Cordero & Datta, 2016) (**Figure I1**). Ces individus interagissant tendent à se regrouper dans leurs habitats et forment ce que l'on appelle des « *hotspots* microbiens » (Juyal *et al.*, 2018) qui correspondent à des zones dans lesquelles l'activité biologique est beaucoup plus rapide et plus intense comparée à la moyenne (Kuzyakov & Blagodatskaya, 2015).

La distribution et la diversité des communautés bactériennes à l'échelle de la planète sont régies par une hypothèse communément connue qui stipule que « *everything is everywhere, but the environment selects* » (Baas Becking & Nicolai, 1934 ; de Wit & Bouvier, 2006). Cette affirmation suggère que les espèces microbiennes sont largement distribuées mais que la composition des communautés bactériennes est régie localement par la sélection et la dérive écologique, processus qui entraîne des fluctuations aléatoires de l'abondance des espèces, réduisant la diversité au sein des communautés et augmentant les différences entre communautés par ailleurs équivalentes (Stegen *et al.*, 2013).

La métagénomique a permis de grandes avancées dans l'étude de la diversité microbienne du sol en incluant notamment les organismes non cultivables. Elle a permis d'avoir une meilleure vision, quoiqu'encore partielle, de la distribution des bactéries dans différents environnements (Castro *et al.*, 2010; Kuramae *et al.*, 2012). En effet, la distribution spatiale non aléatoire des bactéries dans leur environnement est un phénomène bien décrit maintenant dans la littérature (Hanson *et al.*, 2012). Cette distribution étant conditionnée par la distance géographique et par l'habitat (Vos *et al.*, 2013) (**Figure I2**).

À l'échelle supérieure au mètre (comparaison de la composition entre deux points distants de plus d'un mètre), l'organisation et la distribution des bactéries sont régies par les facteurs biogéochimiques globaux comme la disponibilité en éléments nutritifs (Koyama *et al.*, 2014; Leff *et al.*, 2015), la couverture végétale (Berg & Smalla, 2009), le type de sol (Griffiths *et al.*, 2011), l'humidité et le pH du sol (Cruz-Martínez *et al.*, 2009; Lauber *et al.*, 2009; Angel *et al.*, 2010), ainsi que par des facteurs

anthropiques comme l'utilisation des fertilisants agricoles (Fierer *et al.*, 2012). À cette échelle, les bactéries contrairement aux espèces animales et végétales ne sont pas soumises à une barrière de dispersion et sont censées coloniser toutes les niches vacantes (Whitaker *et al.*, 2003; Bell, 2010). La composition apparemment stable des communautés bactériennes résulterait d'une homogénéisation des conditions biotiques et abiotiques d'une multitude de meso — et micro-environnements.

À l'échelle centimétrique, l'échantillonnage permet d'observer une organisation bactérienne en méta-communautés provenant de l'agrégation d'une multitude de plus petites communautés bactériennes interconnectées entre elles. Les gradients environnementaux comme l'oxygène déterminent la structure et la composition de ces communautés bactériennes à cette échelle.

Enfin, à l'échelle cellulaire, la structure et la distribution des populations bactériennes seraient régies par les multiples interactions intra et inter populationnelles, les interactions abiotiques et la physiologie de chaque bactérie qui auront un effet mesurable sur la composition et la dynamique de la population étudiée (Cordero & Datta, 2016).

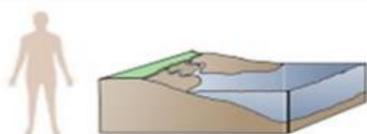
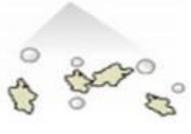
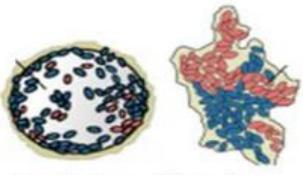
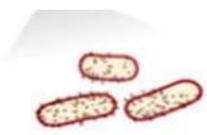
	Échelle	Unité écologique	Paramètres influant sur les communautés
	> m	Écosystème	Paramètres biogéochimiques
	cm — m	Méta-communautés et méta-populations	Gradients environnementaux (pH, O ₂ , etc.)
 Agrégats multi espèces	10 -10 ³ μm	Communautés locales et populations	Interactions écologiques
	1 -10 μm	Cellules individuelles	Physiologie cellulaire

Figure I1 : Paramètres gouvernant la distribution et la composition des communautés bactériennes à diverses échelles (adapté de Cordero & Datta 2016).

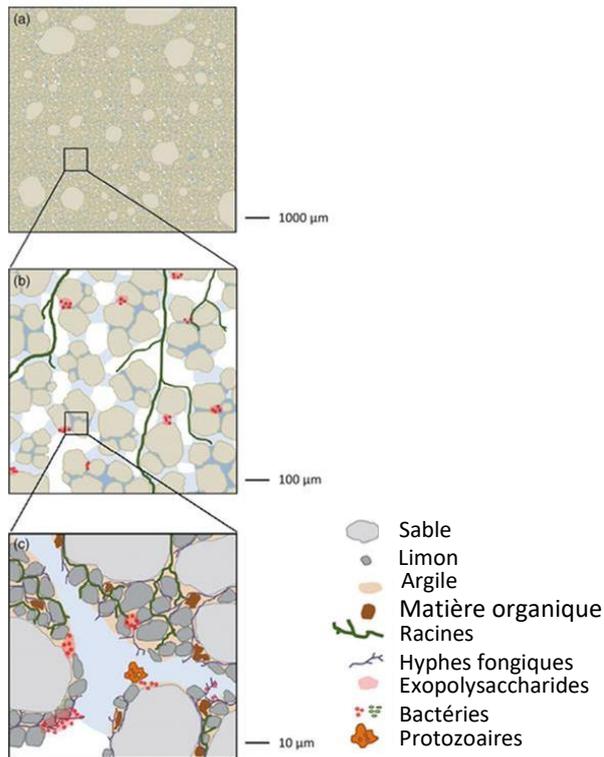


Figure 12 : Organisation de l'écosystème « sol » à différentes échelles. Le sol est un environnement qui apparaît homogène aux grandes échelles(a), mais qui affiche une hétérogénéité lorsque l'on s'approche de l'échelle cellulaire (b) et (c). À ces échelles, le sol correspond à une agrégation de micro- et macro-agrégats de compositions diverses [sable, limon, argile], formant une structure poreuse reliée par des canaux pouvant contenir de l'eau [bleu] ou de l'air [blanc]. Les différents pores peuvent être colonisés par des bactéries libres, des bactéries organisées en biofilm, des protozoaires, des champignons. Les racines, les hyphes mycéliens et les exopolysaccharides permettent de structurer cet environnement. [Adapté de Vos et al., 2013].

I.1 Importance de l'échelle dans l'étude des bactéries du sol.

Un gramme de sol contient plusieurs milliards de bactéries représentant plusieurs centaines d'espèces (Portillo *et al.*, 2013; Regan *et al.*, 2014), et ces bactéries représentent une biomasse humide de 300 à 3.000 kg par hectare (Sylvia, 2005). Cependant moins de 1% du volume du sol serait occupé par les microorganismes (Schmidt *et al.*, 2011). Une étude basée sur des centaines de coupes fines de sol prélevées à différentes profondeurs a permis d'estimer qu'à une concentration de 10^9 cellules par gramme de sol, le nombre moyen de cellules voisines entourant une cellule unique était de 1.040 à une distance de 50 µm et de 120 cellules (équivalent à 100 espèces bactériennes) à une distance de 20 µm. À une concentration de 10^{10} cellules par gramme de sol ce nombre moyen passerait de 5.500 cellules voisines à 50 µm à 550 (équivalent à 280 espèces bactériennes) à 20 µm (Raynaud & Nunan, 2014). Ces résultats démontrent que l'échelle spatiale est un paramètre important à prendre en compte dans l'étude des interactions bactériennes et que cette étude nécessite d'être menée à une échelle inférieure au millimètre (Nunan *et al.*, 2003). L'une des raisons est la grande hétérogénéité que l'on retrouve à la micro-échelle. En effet à cette échelle, on observe une succession assez hétéroclite de micro-habitats sur de très courtes distances de sols, des zones de poches d'air fournissant un environnement aérobie alternant avec des microenvironnements anaérobies, des micro-habitats saturés en eau succédant à des micro zones arides.

Dans cet ensemble assez hétérogène les bactéries ont tendance à former des communautés dans des pores et anfractuosités remplis d'eau ou couverts de pellicules d'eau (Vos *et al.*, 2013) (**Figure 12**). Cette

répartition bactérienne sera alors fortement influencée par la nature du sol (limoneux ou argileux) et sa granulométrie. Plus les grains formant le sol seront grossiers, plus on aura divers micro-habitats qui influenceront la diversité bactérienne retrouvée dans cet écosystème (Carson *et al.*, 2010; Ruamps *et al.*, 2011; Chau *et al.*, 2011).

En effet les caractéristiques géométriques des pores dans le sol comme le volume du pore, sa forme, ses connectivités et sa taille peuvent impacter la composition et l'activité des bactéries. Ces pores régulent l'accessibilité aux matières organiques, la diffusion de l'oxygène, la diffusion des matières solubles à travers la phase aqueuse ainsi que le mouvement des micro-organismes. La distribution des communautés bactériennes serait alors irrégulière aux petites échelles et dépendrait fortement de la structure physique de ces micro-habitats.

L'étude des bactéries à des échelles très fines (micro-habitats) pourrait aider à déterminer les facteurs contrôlant l'activité et la distribution des communautés bactériennes. Les données recueillies et la connaissance précise de ces facteurs pourraient contribuer au développement des modèles prédictifs qui favoriseraient la compréhension du rôle joué par les bactéries dans les fonctions du sol (Juyal *et al.*, 2018).

I.2 Etude de la diversité et de la répartition des bactéries à la micro-échelle

Le sol qui globalement paraît homogène lorsqu'il est observé à des échelles supérieures au mètre serait en fait très hétérogène à l'échelle de la cellule (échelle micrométrique). À cette échelle les bactéries protégées par les structures complexes du sol sont à l'abri des changements globaux rapides de l'environnement. Cet environnement avec des structures poreuses définit des niches écologiques distinctes par l'ensemble des caractéristiques physico-chimiques et structurales (facteurs abiotiques) au sein desquelles évoluent des communautés bactériennes qui auront tendance à interagir (facteurs biotiques) par des relations de prédateurs, de compétitions ou de coopération pour l'accès aux différentes ressources nécessaires à leur développement (van de Mortel & Halverson, 2004; Chowdhury *et al.*, 2011).

Pendant des années, la distribution spatiale des bactéries environnementales a été étudiée soit à partir de l'analyse des prélèvements de sols remaniés ou à partir de la reconstitution des modèles de sols dans des environnements maîtrisés (microcosme). Cependant, les stratégies d'échantillonnage classique du sol ne sont pas adaptées à l'étude des facteurs qui meuvent et maintiennent la distribution bactérienne dans le sol. L'échantillonnage des sols remaniés provoque la dégradation de la structure de ces derniers (facteur important dans la distribution des communautés bactériennes) et les microcosmes ne permettent d'avoir qu'un aperçu de la complexité de la diversité et des interactions bactériennes pouvant exister dans le sol.

I.2.1 L'étude des habitats à la micro-échelle et de leurs bactéries

Plusieurs techniques comme le tamisage (Ranjard & Richaume, 2001) et la microdissection (Grundmann & Normand, 2000; Dechesne *et al.*, 2003) permettant le fractionnement des échantillons de sols ont permis de réaliser des échantillonnages qui se rapprochent le plus de l'échelle du micro-habitat. Le fractionnement des sols peut être réalisé par tamisage dans des conditions humides ou sèches. Chaque condition ayant des avantages et des inconvénients. Les conditions sèches entraînent lors du tamisage l'abrasion des agrégats formant le sol; ce qui peut affecter leur taille et leur stabilité ((Kemper & Rosenau, 1986) et casser la structure du sol. Dans ces conditions on peut également avoir une altération de l'activité des communautés bactériennes ainsi que de leur composition (Hattori, 1988). La taille des tamis utilisés peut également limiter l'échantillonnage à des fragments uniquement visibles à l'œil (Ranjard & Richaume, 2001). Le tamisage dans des conditions humides entraîne aussi une dislocation des agrégats de grande taille, la stabilité des agrégats dans l'eau étant inversement corrélée à leur taille (Haynes & Swift, 1990). Dans ces conditions on observe également une altération de structure des communautés bactériennes, ces dernières étant très dispersibles en milieu aqueux.

Les techniques de fractionnement ne permettent pas de préserver l'architecture des sols et de prendre en compte dans l'analyse la distribution des populations bactériennes. Cependant malgré la perte de ces informations, ces techniques ont le mérite d'être accessibles et à moindre coût. Les échantillons récupérés par ce genre de techniques n'ayant pas ou peu subi de traitement sont plus adaptés aux analyses moléculaires en aval.

La difficulté dans l'étude de la structure et la distribution des communautés du sol réside dans l'opacité physique de ce dernier. Ainsi les approches d'études envisagées reposent sur l'utilisation des sources d'énergie ou des techniques capables de traverser cette opacité.

Plusieurs techniques permettant la conservation de la structure des sols ont ainsi été développées et utilisées pour caractériser la structure des communautés microbiennes à l'échelle du micro-habitat. Parmi celles-ci on peut citer la technique de microscopie électronique à balayage (MEB); la micro-tomographie par ordinateur permettant la visualisation de la structure en trois dimensions à l'échelle micrométrique (Tippkötter *et al.*, 2009); l'utilisation des microélectrodes permettant de détecter certains ions résultant du métabolisme de certaines espèces bactériennes (azote (Okabe *et al.*, 1999), phosphore (Wang & Bishop, 2005), et pH (Bezbaruah & Zhang, 2002)); la résonance magnétique nucléaire (RMN) permettant la recherche des noyaux hydrogène (contenu dans l'eau) dans les échantillons (Shi *et al.*, 2017) ainsi que les techniques de spectrométrie (Rossel *et al.*, 2007; Wetterlind *et al.*, 2010; Heister *et al.*, 2012) qui ont permis d'affiner la caractérisation des paramètres physico-chimiques des particules de sol étudiées. Cette évolution des techniques de caractérisation des

particules a été renforcée par une amélioration des techniques de visualisation des bactéries comme l'utilisation de la fluorescence, couplée à l'*hybridation in situ* et à la microscopie, permettant ainsi l'obtention des cartes en trois dimensions de la localisation des bactéries au sein des échantillons de sols.

Les techniques d'échantillonnage à la micro-échelle ont permis grâce au marquage des bactéries et à l'utilisation de la microscopie épi-fluorescente de démontrer que la répartition bactérienne est hétérogène dans des échantillons de sols observés en trois dimensions (Raynaud & Nunan, 2014) (**Figure I3**). L'utilisation des coordonnées des fragments d'échantillons de sols a également permis de créer des cartes en trois dimensions de la diversité tant génétique que fonctionnelle des populations bactériennes qui les occupent (Vogel *et al.*, 2003; Gonod *et al.*, 2006). Grundmann *et al.*, en 2001 avaient utilisé la technique de la microdissection couplée à une simulation et avaient déterminé que la répartition des bactéries oxydatrices de NO_2^- et de NH_4^+ n'était pas aléatoire au sein de l'échantillon et qu'il existait des zones qui concentraient de fortes densités bactériennes. Ces zones de forte concentration bactériennes étant interconnectées grâce aux pores qui permettaient leur dispersion (**Figure I4**).

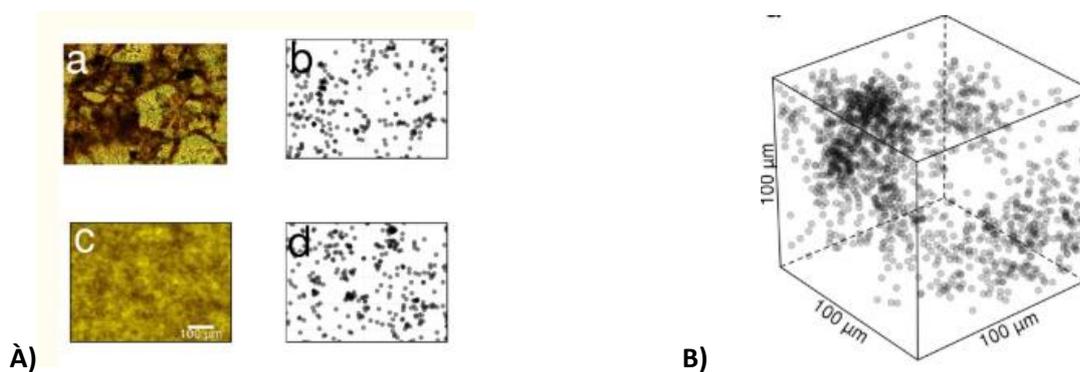


Figure I3 : Distribution et simulation en trois dimensions de la répartition bactérienne au sein d'un échantillon de sol de $100 \mu\text{m}^3$. (A) La visualisation de la répartition en trois dimensions des bactéries à l'aide d'un microscope à épi-fluorescence dans des couches minces d'agrégats de sol a permis d'estimer que la distance intercellules est en moyenne de $12,46 \mu\text{m}$ et que cette distance augmente avec la profondeur d'échantillonnage. (B) Modélisation de la répartition théorique en 3D d'une population bactérienne de 10^9 cellules/g au sein d'un échantillon de sol de $100 \mu\text{m}^3$. (Extrait de Raynaud & Nunan, 2014).

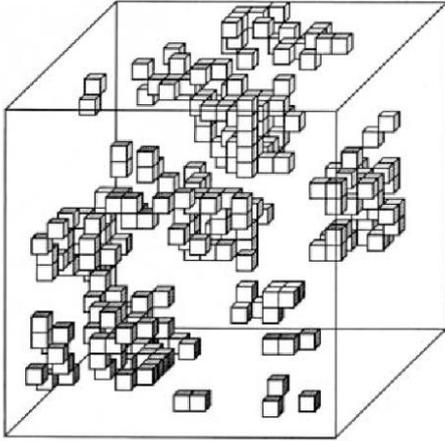


Figure 14 : Modélisation spatiale de la répartition de bactéries oxydatrice du NO₂. Chaque cube de 50 µm contient au moins une bactérie. La simulation a été réalisée sur un échantillon de 1mm³. (Adapté de Grundman *et al.*, 2001).

Juyal *et al.*, (2018) ont quant à eux démontré grâce à une technique de micro-tomographie à rayons X combinée à la microscopie à fluorescence dans des sections de sols imprégnées de résines, l'influence de la géométrie des pores sur la distribution de *Pseudomonas fluorescens*. Cette distribution bactérienne étant analysée à la macro — (5,2 mm x 5,2 mm), méso — (1 mm x 1 mm) et micro-échelle (0,2 mm x 0,2 mm) avec des montages expérimentaux simulant différentes architectures de sol. Cette étude a permis de déterminer que la porosité et la connectivité pouvaient influencer la distribution bactérienne à l'échelle millimétrique.

Ces différentes études montrent l'importance de la prise en compte de l'architecture et de l'échelle d'échantillonnage dans les analyses de la distribution bactérienne au sein des échantillons provenant du sol.

1.2.2 Le microcosme

Pour comprendre certains mécanismes écologiques, il est important d'aller au-delà des études purement descriptives du micro-habitat. Cela implique d'élaborer des systèmes expérimentaux dans lesquels tous les paramètres du sol sont sous contrôle de façon à examiner leurs impacts sur les communautés bactériennes. Plusieurs systèmes mimant des écosystèmes sont ainsi disponibles, systèmes dans lesquels les paramètres biotiques peuvent être contrôlés (Guenet *et al.*, 2011). Les systèmes expérimentaux destinés à tester les hypothèses sur l'écologie des communautés bactériennes ont le plus souvent utilisé des microcosmes contenant des échantillons de sols prélevés en carottes (Ruamps *et al.*, 2011), de sol tamisés et stérilisés (Wright *et al.*, 1995), non stérilisés (Chowdhury *et al.*, 2011) ou carrément de sols artificiels (Ellis, 2004). Ces systèmes permettent la manipulation des paramètres comme l'humidité et la composition minérale et l'observation de leurs impacts, par différentes techniques moléculaires, sur les communautés bactériennes natives ou introduites.

L'utilisation de matériaux poreux et transparents à la place des sols permet également de contourner le problème causé par l'opacité des sols dans l'étude des communautés bactériennes qu'ils renferment. Ainsi en utilisant un matériau permettant une transmission de la lumière à travers un microscope confocal, le Nafion (copolymère fluoropolymère à base de tétrafluoroéthylène sulfoné), Leis *et al.*, en 2015 ont réussi à observer la croissance d'un biofilm de *Pseudomonas aeruginosa* SG81.

Même si une boîte de pétri peut permettre d'apporter des réponses aux interrogations concernant l'écologie microbienne (Dechesne *et al.*, 2003), certaines questions nécessitent l'utilisation des techniques plus élaborées. L'utilisation des systèmes mimant l'environnement peut améliorer notre compréhension de nombreux processus écologiques. Il a ainsi été démontré que processus évolutif de *Pseudomonas* et de l'un de ses phages était très différent lorsqu'ils évoluent dans un microcosme non stérile que dans un bouillon à haute valeur nutritive (Gómez & Buckling, 2011). L'utilisation des approches de micro-fluidique complètement artificielles permettent également l'observation en temps réel du comportement des communautés bactériennes même à l'échelle de la cellule (Keymer *et al.*, 2006; Männik *et al.*, 2009). Cependant ces techniques d'observation étant couplées à la microscopie restent limitées car une saturation en eau est nécessaire pour une meilleure visualisation au microscope.

1.3 La diversité génétique et fonctionnelle bactérienne à la micro-échelle

Les investigations pour caractériser les communautés bactériennes ont été historiquement faites par les méthodes basées sur la culture. Toutefois, malgré des améliorations des milieux de culture standards par des apports d'éléments mimant l'environnement, plus de 99 % des bactéries restent non cultivables (Hugenholtz, 2002; Oliver, 2005). Pour contourner ces limites des techniques de culture, des méthodes moléculaires qui s'affranchissent de tout processus de culture ont été proposées pour caractériser la diversité des communautés microbiennes.

L'étude de la diversité bactérienne sur des petits échantillons de sol en utilisant comme marqueur taxonomique la séquence nucléotidique codant l'ARN ribosomique 16S a permis de démontrer qu'il existait une importante diversité au sein de ces échantillons (Vogel *et al.*, 2003; Grundmann, 2004). Cette approche combinée à la métagénomique a permis de grandes avancées dans l'étude de la diversité microbienne du sol et notamment concernant la phase non cultivable. Elle a permis d'avoir une meilleure vision de la distribution de ces bactéries dans différents environnements et dans certains cas de corréler la répartition des grands groupes bactériens (phylotypes définis à 97% d'identité sur une portion du gène codant l'ARNr 16S) avec certains facteurs environnementaux, chimiques, physiques ou biotiques (Lauber *et al.*, 2009; Fierer *et al.*, 2012).

Peu d'études en revanche se sont intéressées à la distribution des bactéries à des échelles inférieures au mètre (Vos *et al.*, 2013). À ces échelles, seules quelques communautés bactériennes issues de sédiments marins ont été caractérisées par l'utilisation de la métagénomique (Probandt *et al.*, 2018). Cette dernière couplée à l'utilisation des techniques de séquençage de nouvelle génération a permis d'avoir une autre aperçue de la diversité bactérienne. Le séquençage de l'ADN total extrait des échantillons de sols a permis de mettre en évidence une diversité des fonctions biologiques codées par les gènes prédits sur ces ADN et en les comparant à des bases de données de déterminer par homologie de séquences les groupes taxonomiques auxquels elles se réfèrent (identification limitée au genre bactérien) (Delmont *et al.*, 2012); Une autre étude combinant l'utilisation des séquences des gènes codants l'ARN 16s et le gène *rpoB* (OTU définis à 97,3% d'identité avec le gène *rpoB* et 1% pour le gène de l'ARNr 16S) a démontré une diversité taxonomique au sein d'un échantillon de sol (Vos *et al.*, 2012). L'ensemble de ces études concordent à dire qu'il existe une importante diversité microbienne au sein des micro-habitats. Paradoxalement, le problème lié à la plupart des études de métagénomique provient de l'utilisation du gène codant l'ARN 16S. En effet, si cela permet de définir des unités taxonomiques opérationnelles (OTU) (avec des seuils fixés à 97% d'identité), ce sont néanmoins des regroupements classificatoires très larges. L'utilisation de cette technique devient moins adaptée lorsque l'on s'intéresse à l'étude des communautés bactériennes phylogénétiquement proches. Par le fait, les souches étroitement apparentées, désignée comme étant génétiquement cohésives (Shapiro and Polz, 2014), présentent des séquences du gène codant l'ARN 16S extrêmement semblables, voire identiques.

Pour caractériser la diversité génétique des populations bactériennes, il devient alors indispensable d'utiliser des substrats d'horloge moléculaire plus précis, fournissant beaucoup plus de sites informatifs (séquences renfermant les mutations et/ou polymorphismes plus grandes) à analyser comme la MLST (*Multi Locus Sequence Typing*), ou la phylogénomique (prise en compte du génome pour inférer la phylogénie au sein de la population). Certaines études ayant utilisé les techniques de la MLST ou la séquence ITS 16-23S (analyse des séquences hypervariables à évolution rapide) ont montré qu'à l'échelle du mètre, voire du centimètre, il existerait des souches proches de différentes espèces possédant une variabilité génétique et fonctionnelle importante (Antony-Babu *et al.*, 2008; Stefanic & Mandic-Mulec, 2009). L'une de ces études portant sur l'analyse de 78 isolats de *Myxococcus xanthus* prélevés sur une parcelle de terrain de 16*16 cm, indifférenciables sur la base de la séquence codant l'ARN 16S a montré l'existence de 21 génotypes distincts, sur la base de trois gènes de ménage concaténés.

La caractérisation fonctionnelle de cette population a permis de démontrer qu'il existait une incompatibilité d'essaimage et de sporulation (Vos & Velicer, 2009), une différence de mobilité (Vos &

Velicer, 2008), de développement de leur organe de fructification (Kraemer *et al.*, 2010), de la capacité de prédation (Morgan *et al.*, 2010), et de la production de métabolites secondaires (Krug *et al.*, 2008) au sein des différents groupes conspécifiques. L'analyse de la capacité à sporuler des souches en interaction bipartite a par exemple montré que certaines souches inhibaient la capacité de sporulations des souches n'appartenant pas au même sous-clade MLST qu'elles. Des études similaires réalisées sur des isolats génétiquement différents provenant de différentes zones géographiques avaient également conduit aux mêmes conclusions (Fiegna & Velicer, 2005). Les travaux de Vos et Velicer permettent alors de démontrer qu'une structuration fonctionnelle observée entre les individus appartenant à des populations différentes pouvait s'appliquer aux individus appartenant à une population d'individus conspécifiques et sympatriques.

Vetsigian *et al.*, en 2011 avait également démontré en réalisant des interactions bipartites qu'entre 64 souches de *Streptomyces* isolées à partir d'un même grain de sol et des grains distants de 1 à 20 centimètres de distance, il existait des phénomènes d'inhibition ainsi que de promotion de la croissance et de sporulation. Cette étude a également permis d'établir qu'il existait beaucoup plus d'interactions biotiques entre les souches provenant d'un même grain de sol qu'entre les souches provenant de deux grains différents. Les phénomènes d'interactions (biotiques) au sein d'un micro-habitat seraient donc plus nombreux et divers qu'entre deux micro-habitats (**Figure 15**).

Les techniques de séquençage à haut débit ont également été utilisées pour accéder à la diversité des populations bactériennes cultivables issues du micro-habitat (Polz *et al.*, 2013) et *in fine* améliorer notre compréhension de cette diversité. De telles approches ont commencé à être utilisées chez les actinobactéries et protéobactéries (Garcia *et al.*, 2018). Cette approche est particulièrement puissante lorsqu'elle est appliquée (i) sur des bactéries provenant d'échantillons environnementaux et (ii) combinée à la méta-génomique réalisée sur le même environnement. Cela permet alors de prendre en compte les bactéries non cultivables dans l'analyse de la diversité (Garcia *et al.*, 2018).

Les résultats de tous ces travaux ont permis de montrer qu'une identité de la séquence du gène codant l'ARNr 16S entre des souches isolées sur des échelles spatiales courtes (inférieures au mètre) pouvait cacher une grande diversité tant au niveau génétique que fonctionnelle. La prise en compte de l'échelle dans l'étude des communautés bactériennes est donc primordiale ceci dans le but de comprendre les processus écologiques qui régissent et maintiennent ces populations dans leurs écosystèmes.

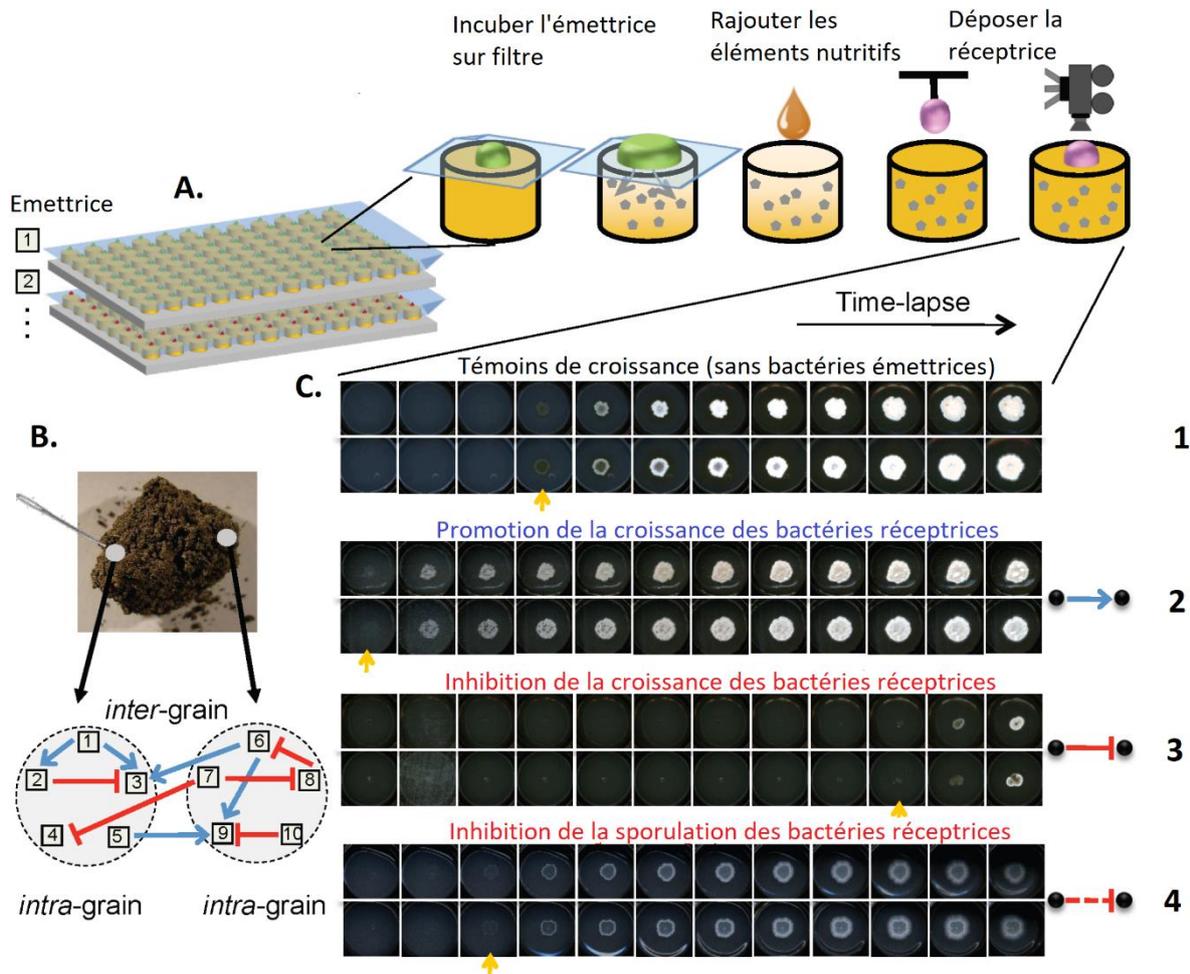


Figure 15 : Plateforme d'étude à haut débit d'interactions bipartites entre des souches de *Streptomyces*. Une collection de 64 souches de *Streptomyces* provenant de quatre grains de sol échantillonnés à l'aide d'une aiguille (B) à une distance de 1 cm pour trois d'entre eux et de 20 centimètres pour le quatrième ont été étudiées pour leur capacité à agir sur le développement phénotypique de souches provenant du même grain ou de grains différents. Pour cela (A) un filtre à pores fins est placé sur une microplaque contenant de l'agar. Un tapis de la bactérie testée comme émettrice de composés est réalisé. Le filtre est ensuite enlevé, laissant ainsi une gélose contenant les molécules excrétées par la souche donatrice. Le plug d'agar est ensuite complété en nutriments afin de compenser la consommation des donatrices. La souche réceptrice est ensuite déposée sous la forme d'une goutte sur la gélose et sa croissance est suivie en *time-lapse* afin de voir l'effet des molécules produites sur son développement. Les résultats obtenus sont présentés en (C). La série (1) montre le time laps pour une réceptrice qui a poussé sur une gélose où aucune bactérie donatrice n'a été déposée et sert donc de témoin de croissance. La série (2) présente un cas de promotion de la croissance de la bactérie réceptrice par les composés libérés par la bactérie donatrice. La série (3) présente un cas d'inhibition de la bactérie réceptrice par les composés libérés par la bactérie donatrice. La série (4) présente un cas d'inhibition de la sporulation de la bactérie réceptrice par les composés libérés par la bactérie donatrice. (Adapté de Vetsigian *et al.*, 2011)

II. Concept d'espèce bactérienne et notion de population

II.1 Classification des micro-organismes

Les tentatives de classification et de catégorisation des organismes avec lesquelles nous partageons notre planète est une des tâches les plus ardues. Pour les organismes supérieurs, une certaine classification avait été amorcée par Aristote dès l'antiquité où il divisait les êtres vivants en quatre

grands groupes sur la base de son observation visuelle : minéral, végétal, animal et l'Homme. Au 18^{ème} siècle le naturaliste Carl von Linnée créa la systématique fixiste qui permit de classer et d'organiser les groupes suivant une base logique. La classification proposée par Linnée repose sur une nomenclature composée d'un nom de genre dont la première lettre est en majuscule suivie d'une épithète en minuscule désignant l'espèce. Le concept d'espèce permettait alors de regrouper des individus possédant en commun des critères morphologiques, biologiques et physiologiques. L'espèce est alors considérée comme l'unité de base dans la classification des êtres vivants.

La notion d'espèce a par la suite évolué avec l'introduction du terme « reproduction sexuée » dans sa définition. Dans son ouvrage paru en 1749 intitulé Histoire générale des animaux, George-Louis Leclerc de Buffon écrivait : « On doit regarder comme la même espèce celle qui, au moyen de la copulation, se perpétue et conserve la similitude de cette espèce, et comme des espèces différentes celles qui, par les mêmes moyens, ne peuvent rien produire ensemble ».

On dit souvent que l'espèce est l'unité fondamentale en écologie parce qu'elle regroupe des individus qui sont phénotypiquement et écologiquement similaires entre eux comparés à ceux d'une autre espèce (Gevers *et al.*, 2005). Cette notion fait écho au concept biologique défini par Ernst Mayr qui stipule que : « *Species are groups of actually and potentially interbreeding natural populations, which are isolated from other such groups* » (Mayr, 1942). La définition de E. Mayr repose sur l'interfécondité des organismes qui constituent une espèce et ne peut donc s'appliquer ni aux organismes parthénogénétiques, ni aux microorganismes (Claridge *et al.*, 1997). La notion d'espèce chez les procaryotes continue de faire polémique.

II.2 Concept d'espèce bactérienne

La définition actuelle de l'espèce bactérienne se base sur l'homogénéité génomique permettant ainsi de définir des espèces appelées « génomiques », qui reflètent l'abondance naturelle de la diversité que l'on peut observer lors d'un échantillonnage (Lassalle *et al.*, 2015). Bien que cette définition de l'espèce bactérienne reste axée sur le génome, le défi reste de comprendre les mécanismes sous-jacents conduisant à l'émergence d'une telle diversité bactérienne (Stackebrandt *et al.*, 2002). La difficulté inhérente à la définition de l'espèce chez les procaryotes est intimement liée à l'incapacité actuelle des évolutionnistes en microbiologie de trouver un modèle commun dans la réalisation de la classification des micro-organismes. Contrairement au concept classique de l'espèce biologique définie par Mayr chez les animaux, qui place la sexualité au centre de la définition des espèces ou des clades, chez les procaryotes la reproduction est quant à elle asexuée et se fait par scissiparité. Il existe également une notion de parasexualité chez ces derniers qui intervient au niveau intra ou interspécifique à travers le transfert horizontal des gènes et la recombinaison homologue. Dans ce

contexte il devient alors impossible de définir l'espèce bactérienne par le concept proposé par Mayr (Doolittle & Zhaxybayeva, 2009), sauf d'élargir cette définition en considérant à la même enseigne le brassage génétique et la fusion des gamètes lors de la méiose chez les organismes sexués et l'internalisation et recombinaison lors des transferts génétiques chez ceux asexués ; donc *in fine* de considérer la possibilité de recombinaison comme paramètre dans la définition des espèces bactériennes (Riley & Lizotte-Waniewski, 2009).

Le concept d'espèce bactérienne a cependant évolué depuis les premières tentatives de regroupement et de classification des organismes vivants. On note au fil du temps des changements notoires dans les paramètres pris en compte dans la définition de l'espèce bactérienne. Le regroupement phénotypique utilisé par Sokal & Crovello en 1970 permettait de regrouper les bactéries en fonction de certains caractères physiques ou métaboliques. Dans cette approche, des capacités comme la dégradation ou le métabolisme de certains composés, l'aptitude à produire ou à résister aux antibiotiques, la capacité d'utilisation des sources de carbone ou d'azote étaient prises en compte. Le caractère morphologique des bactéries que ce soit sur des milieux gélosés (en colonie) ou en microscopie à travers la rétention ou non de colorants (coloration de Gram) a également servi dans l'approche phénotypique (Goodfellow *et al.*, 1997). Les bactéries étaient alors classées en espèces suivant le nombre de caractères qu'ils avaient en communs sans prendre en compte une quelconque ascendance commune. Ces méthodes ont été mises à profit dans l'identification des souches pathogènes en bactériologie médicale. Seulement ces caractères variables voire rares ne peuvent évidemment pas servir de base pour des regroupements (sauf dans des contextes très restreints, ex. coliformes commensaux humains).

L'avènement des techniques moléculaires a permis d'aborder des critères supplémentaires dans la classification des bactéries surtout avec la prise en compte du concept d'horloge moléculaire. La systématique bactérienne s'est ainsi basée au cours de ces dernières décennies sur l'analyse des variations au niveau moléculaire tant nucléotidique que protéique. Dans les années 1970 la pierre angulaire de la caractérisation génotypique a été l'utilisation de la similarité génétique globale entre les bactéries qui est évaluée par le degré d'hybridation de leurs ADN dans des conditions standards. Les bactéries sont assignées à une même espèce si le taux d'hybridation de leur ADN est supérieur ou égal à 70 % (Wayne *et al.*, 1987). Dans cette approche, la similarité en contenu génétique et l'identité nucléotidique des gènes partagés sont prises en compte dans le regroupement des bactéries en espèce. Cette technique a permis une standardisation des moyens d'identification et de classification des procaryotes qui ne disposaient pas de caractéristiques phénotypiques bien définies.

L'avènement des techniques de séquençage bas débit dans les années 1980 a permis l'émergence de la classification des espèces bactériennes basée sur le degré de divergence nucléotidique entre les séquences codant l'ARN ribosomique 16S (Fox *et al.*, 1977). Ce gène, universel chez les êtres vivants, est peu variable et est utilisé comme une horloge moléculaire (Woese, 1987). Une identité nucléotidique inférieure ou égale à 98,7 % permettait d'identifier deux espèces bactériennes. Ce seuil était corrélé à celui établi pour l'hybridation d'ADN (< 70 %). Cependant l'inverse n'est pas forcément vrai. Certaines souches classées comme appartenant à la même espèce en utilisant l'identité de séquence de l'ARN ribosomique 16S possèdent un taux d'hybridation ADN-ADN inférieur à 70 %. L'utilisation du seuil d'identité nucléotidique de l'ARN ribosomique 16S étant discutée pour la définition des espèces bactériennes (Stackebrandt *et al.*, 2002; Brown, 2003), certains auteurs ont préféré le regroupement des bactéries en Unité Taxonomique Opérationnelle (OTU) à partir d'un seuil de 97 % d'identité de séquence de l'ARN ribosomique 16S (DeSantis *et al.*, 2006). Cette notion d'OTU est beaucoup utilisée en écologie microbienne et reste dans le cas des bactéries non cultivables l'un des outils d'appréciation de la diversité.

L'utilisation d'un seul marqueur taxonomique pouvant être soumise à de simples variations stochastiques et à l'influence de la recombinaison ou du transfert horizontal de gènes (Gogarten *et al.*, 2002), la définition des espèces sur la base des séquences nécessite alors l'utilisation de *loci* qui évoluent plus rapidement que l'ARN ribosomique 16S. La combinaison de plusieurs de ces *loci* permettrait de minimiser le biais éventuel produit par l'effet d'une recombinaison au niveau d'un seul *locus*. Ainsi la MLST (*Multi Locus Sequences typing*), une méthode de caractérisation génotypique des procaryotes à l'échelle infra spécifique en utilisant les variations nucléotidiques de plusieurs gènes de ménages, a été développée (Maiden *et al.*, 1998). Des isolats possédant des profils alléliques identiques en MLST permettent de définir des souches ou clones et forment une base pour le système de classification génotypique des bactéries. Cependant le problème de cette méthode, comme d'autres comme l'AFLP (Amplified Fragment-Length Polymorphism), la RFLP (Restriction Fragment Length Polymorphism) et la RepPCR (Repetitive element palindromic-Polymerase Chain Reaction) est la définition du seuil de démarcation (identité nucléotidique pour la MLST et coefficient de similarité pour les autres) à partir duquel on définit une espèce.

Plusieurs concepts et définitions ont été suggérés pour l'espèce bactérienne. Cependant, aucun n'a été unanimement accepté. L'espèce bactérienne est actuellement définie par une approche pragmatique et polyphasique basée sur des règles portant sur les propriétés génotypiques que phénotypiques (Vandamme *et al.*, 1996). Cette approche est en train d'être adaptée à l'ère du séquençage de nouvelle génération (NGS) et de la biologie à haut débit. Ainsi l'utilisation des techniques de phénotypage (ex Maldi TOF MS, composition en acides gras, Biolog) et des données

issues du séquençage des génomes entiers (e.g *Average Nucleotide Identity test* (ANI), (Konstantinidis *et al.*, 2006), phylogénomique) ont permis d'améliorer la résolution taxonomique des groupes précédemment formés (Huse *et al.*, 2008; Roh *et al.*, 2010; Claesson *et al.*, 2010). L'ANI comme son nom l'indique est une méthode qui permet de comparer des génomes et de déterminer les identités nucléotidiques moyennes entre les séquences comparées. Cette technique offre une robustesse de regroupement avec une discrimination des souches possédant une séquence de l'ARN 16S identique (identité de la séquence du gène de l'ARNr 16S correspondant à des ANI >98.5% ; Konstantinidis *et al.*, 2006). Les techniques de séquençage de nouvelle génération appliquées à la systématique bactérienne ne suffisent pas cependant à elles seules à résoudre le problème de consensus dans la définition de l'espèce (Gevers *et al.*, 2005; Koeppel *et al.*, 2008; Achtman & Wagner, 2008; Fraser *et al.*, 2009) car ces approches ne prennent pas en compte l'écologie de l'organisme dans cette définition de l'espèce. Cohan et ses collaborateurs (2002a) dans leur concept d'écotype ont ajouté cette notion d'écologie dans le regroupement des micro-organismes.

II.2.1 Le concept d'écotype

Le modèle et concept d'écotype a été défini et décrit par Cohan dans ses différents travaux (Cohan, 2002a; Cohan, 2002b; Cohan & Perry, 2007; Wiedenbeck & Cohan, 2011). Cohan propose que l'espèce, comme définie chez les bactéries, devrait être subdivisée en de plus petites unités en y incorporant le concept d'écotype. Selon lui un « modèle d'écotype » fournirait des bases rationnelles dans la séparation des taxons bactériens. Les écotypes sont définis comme étant des populations génétiquement cohésives et écologiquement distinctes. La cohésivité dans le cas de l'écotype résulterait alors de la sélection périodique des individus au sein de la population ayant pour but de purger la diversité génétique qui s'y accumule. Dans ce modèle, l'acquisition d'un gène par transfert ou d'une mutation conférant l'adaptation et la colonisation d'une nouvelle niche imposerait une forte sélection pour le seul mutant adapté à la niche. La multiplication de cet individu fondateur et sa rapide expansion par scissiparité entraîneraient la formation d'une population adaptée à la niche. Les écotypes étant supposés être irréversiblement séparés les uns des autres, ce concept détiendrait alors les propriétés spécifiques à la définition de l'espèce comme énoncés dans les systématiques autres que microbiennes (Queiroz, 1998).

Selon Cohan et ses collaborateurs, différents écotypes peuvent coexister indéfiniment uniquement en raison de leur divergence écologique (utilisation de différentes ressources ou différentes conditions de croissance). Les écotypes seraient alors définis de manière à ce que les divergences écologiques au sein des lignées dans une même niche soient suffisantes pour leur permettre de cohabiter indéfiniment. Ces lignées écologiques distinctes dans un écotype seraient donc des « écotypes éphémères ». La scission d'un écotype en deux nouveaux serait le mécanisme fondamental sous-jacent

de la création de la diversité aboutissant à la spéciation bactérienne (Koeppel *et al.*, 2008; Cohan, 2011). Plusieurs modèles de spéciation bactérienne au sein des écotypes ont alors été proposés (**Figure 16**). Le modèle d'écotype stable dans lequel l'écotype subit une sélection périodique menant au maintien et à l'expansion du clone le plus adapté à la niche. Dans cette situation, les écotypes ne peuvent coexister que s'ils n'exploitent pas les mêmes ressources ; la spéciation se fait de manière lente. Le modèle de spéciation rapide est similaire à celui d'écotype stable excepté que la spéciation se fait si rapidement que la plupart des nouveaux écotypes divergents ne peuvent être détectés par analyses de séquences des loci. Le modèle nano-niche de spéciation bactérienne correspond à la coexistence de plusieurs écotypes dans une même niche globale et exploitant une même ressource de cette niche à différente proportion (Cohan, 2017).

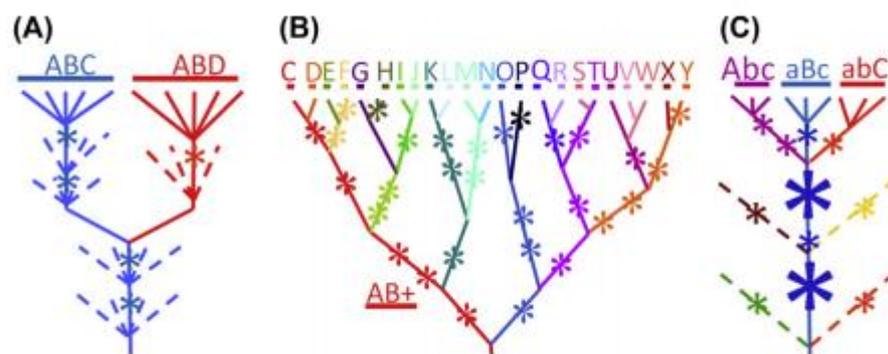


Figure 16 : Modèle de spéciation bactérienne. Les écotypes sont représentés par différentes nuances de couleurs; les événements de sélection périodiques sont indiqués par des astérisques, et les lignées éteintes sont représentées par des lignes interrompues. Les lettres au-dessus représentent les ressources que chaque groupe d'organismes peut utiliser. Dans les cas où les écotypes utilisent le même ensemble de ressources, mais dans des proportions différentes, la ressource prédominante de chaque écotype est indiquée par une lettre majuscule. A) Le modèle d'écotype stable. Dans ce modèle, chaque écotype est soumis à de nombreux événements de sélection périodiques; chaque écotype correspondant à un génotype; la spéciation se faisant lentement. (B) Le modèle de spéciation rapide. Ce modèle ressemble beaucoup au modèle d'écotype stable, sauf que la spéciation se produit si rapidement que la plupart des nouveaux écotypes divergents ne peuvent être détectés par analyses de séquences des loci. (C) Le modèle nano-niche de spéciation bactérienne dans lequel 3 écotypes occupant des nano-niches au sein de la niche principale utilisent la même ressource à différentes proportions. (Extrait de Cohan, 2017).

Cependant ces modèles d'écotype proposés par Cohan ne permettent pas d'expliquer l'occurrence dans un écotype d'un mélange de bactéries ne possédant pas de génotypes identiques. En effet dans certaines conditions, un écotype pourrait renfermer plusieurs génotypes. Ce phénomène a été repris par Lassalle *et al.*, (2015) où ils définissaient un nouveau concept d'écotype « *sensu lato* » (s.l.) en appui au concept d'écotype « *sensu stricto* » (s.s.) décrit par Cohan. Dans cette définition, l'écotype s.l. serait formé par une population d'individus résultant d'une série d'épisodes au cours desquelles une mutation adaptative est suivie par une spéciation réussie d'un écotype s.s.. Les écotypes s.s. émergeraient constamment dans l'écotype s.l. et pour la plupart disparaîtraient sans fonder une

population écologiquement différente; ce processus maintient la diversité génotypique et phénotypique au sein de l'écotype s.l.. Les variations apparaissant au sein de l'écotype s.l. peuvent être dispatchées au sein de la population par recombinaison homologue entraînant la formation d'une population génétiquement non homogène; ceci expliquerait l'hétérogénéité observée au sein de certains écotypes. L'écotype s.l. serait donc une métapopulation formée d'écotypes s.s., chacun ayant sa propre niche écologique définie dérivant de la niche principale de l'écotype s.l. (Lassalle *et al.*, 2015).

Les écotypes s.s. sont difficiles à repérer sur le terrain en raison de leur nature éphémère et de très rares sinon uniques caractéristiques génomiques et écologiques qui les différencient des populations sœurs. Ils peuvent cependant être identifiés et leur évolution caractérisée par des études d'évolution expérimentale, dont les principes sont décrits en détail par Barrick & Lenski (2013). D'autre part, l'observation de la spéciation des écotypes s.l. dans l'environnement est plus probable, comme le démontre l'étude de Shapiro *et al.* (2012) sur des souches de *Vibrio vulnificus*. Dans cette étude, il est montré qu'au sein de deux populations bactériennes qui ont récemment divergé, la différenciation écologique de chaque population par rapport à sa niche, réduisait le taux de recombinaison entre ces populations. Donc les barrières aux échanges génétiques surviendraient dès les premières étapes de spéciation, avec des flux de gènes plus importants au sein d'une population occupant une même niche, qu'entre deux populations ayant récemment divergé et adaptées à deux niches différentes (**Figure 17**).

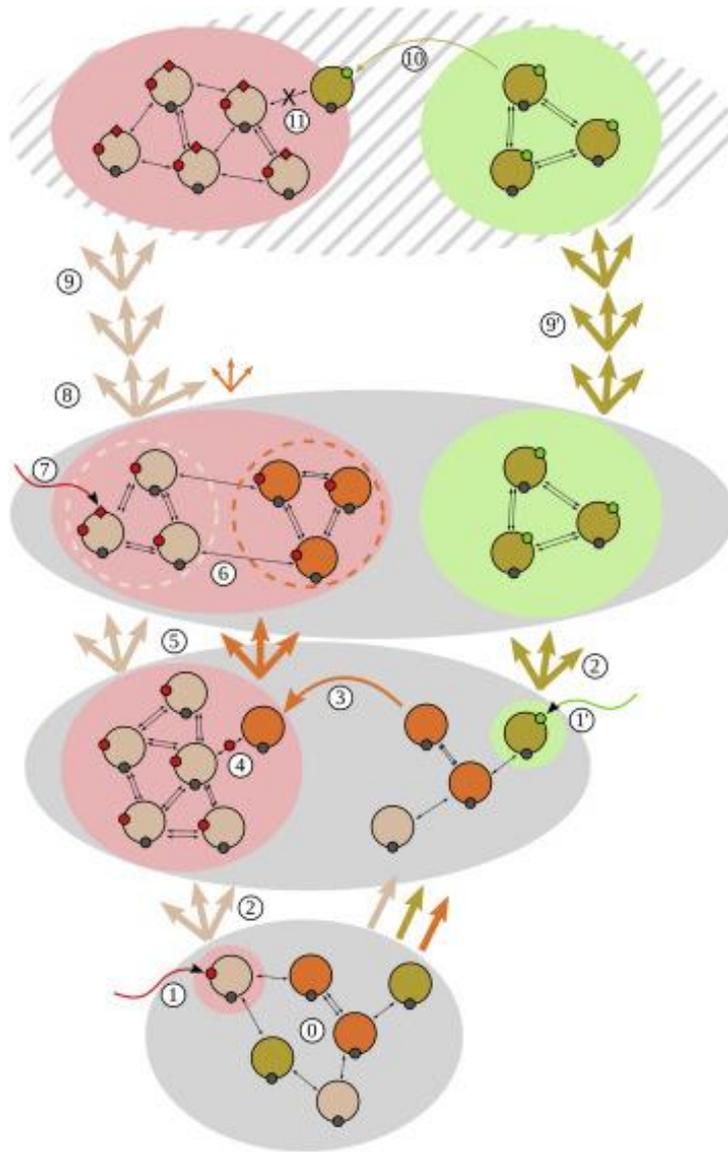


Figure 17 : Émergence, maintien et diversification des espèces écologiques.

Les cercles de différentes couleurs représentent des génotypes bactériens et les perles en périphérie de ces cercles symbolisent des gènes d'adaptation aux différentes niches. Les fonds colorés représentent différents micro-habitats; la couleur du micro-habitat étant la même que celle du gène y permettant l'adaptation. (0) Une population hétérogène occupant une niche écologique (fond gris), avec plusieurs génotypes partageant un gène commun codant pour une adaptation à la niche commune. L'échange du matériel génétique est proportionnel à la relation phylogénétique (deux flèches à deux sens pour des échanges plus importants). (1 et 1') Les individus reçoivent un nouveau gène (perle rouge ou verte) par transfert horizontal provenant de taxons différents, fournissant de nouvelles fonctions qui modifient la définition de la niche écologique. Les nouvelles niches écologiques (fond rouge ou vert, respectivement) dérivent de la niche ancestrale (fond gris), avec une dimension supplémentaire et/ou des modifications des dimensions préexistantes de la définition de la niche. La superposition des fonds gris et rouge ou vert représente donc le partage d'une niche centrale (grise) par des populations adaptées aux niches rouges ou vertes. La

coexistence de plusieurs écotypes est conditionnée par un compromis d'exploitation de la niche ancestrale tout en s'adaptant à la nouvelle. (2) L'exploitation réussie de la nouvelle niche déclenche la multiplication clonale de l'individu fondateur pour former un nouvel écotype. (3) le partage d'une niche ancestrale permet aux membres de la population initiale de se rencontrer et d'échanger du matériel génétique (4) avec la possibilité de transfert du nouveau gène d'adaptation permettant au mutant de génotype distinct (cercle orange) de concurrencer de manière égale pour la nouvelle niche. (5) Avec la dérive génétique, la fréquence du génotype migrant peut augmenter (6) et la différence de métabolisme peut permettre aux deux génotypes d'exploiter des nano-niches (cercles discontinus) légèrement différentes, leur permettant de cohabiter temporairement avec une faible compétition et des possibilités de flux de gènes. (7) Un événement secondaire de transfert horizontal introduit un autre nouveau gène (losange rouge) dans l'un des fonds génomiques, conférant une meilleure adaptation à la niche (8), conduisant à l'extinction de la diversité des écotypes spécialisés dans les nano-niches. L'invasion ultérieure de l'écotype devient plus difficile avec un nombre croissant de gènes d'adaptation à la niche à transférer afin de maintenir une sélection périodique. (9 et 9') Les épisodes de sélection périodique maintiennent l'homogénéité génomique au sein des écotypes, tandis que l'accumulation de substitutions neutres entraîne la diversité du génome entier de différents écotypes. Le flux de gènes entre écotypes ayant divergé est limité (10) par l'importance accrue des gènes impliqués dans l'adaptation à chaque niche écologique par rapport à la niche principale, ce qui limite la présence de niches partagées et donc le taux de rencontre, ainsi qu'une apparition progressive de barrière à l'échange d'ADN (11), notamment en raison de la forte diminution de l'efficacité de la recombinaison homologue avec la divergence des génomes. (Extrait de Lassalle *et al.*, 2015).

III. Les mécanismes d'évolution et de dynamique des génomes chez les bactéries

Le concept de sélection naturelle comme moteur de l'évolution a été proposé indépendamment par Charles Darwin (1859) et Alfred Russel Wallace (Wallace, 1870). Cette théorie majoritairement construite par Darwin reposait sur l'observation des macroorganismes (animaux et végétaux) au niveau phénotypique. Les mécanismes sous-jacents de ce concept reposent sur le principe de variation, le principe d'adaptation, et le principe d'hérédité. Dans ce concept de sélection naturelle, il faut préciser que le but de la sélection n'est pas l'évolution, mais le maintien des espèces les plus adaptées à un écosystème donné, à un moment donné. Le biotope étant lui-même sujet au changement, modifie constamment les conditions auxquelles les espèces doivent s'adapter. De plus une espèce ne devrait pas seulement s'adapter aux changements survenant dans son biotope, mais aussi aux changements que subissent les autres espèces partageant le même écosystème. Darwin était alors conscient que les interactions entre les espèces partageant un même biotope jouaient un rôle important dans la sélection naturelle.

Le concept de sélection naturelle darwinienne a été appliqué aux microorganismes, plus précisément aux procaryotes. En effet, au sein des procaryotes, à chaque fois qu'un « variant » de la population acquiert un avantage lui permettant de survivre, que ce soit pour l'acquisition d'une ressource ou l'adaptation à une modification du biotope, ce variant se maintient et se propage au sein de la population. La nuance entre l'évolution décrite par Darwin chez les macroorganismes et celle chez les procaryotes est le temps. Le temps de génération chez les procaryotes étant de quelques heures à quelques jours. Darwin a développé le concept de la sélection naturelle en se basant sur les observations phénotypiques sans pour autant connaître les mécanismes sous-jacents de cette évolution. L'étude de l'évolution des procaryotes a joué un rôle capital dans la compréhension de ces mécanismes. L'étude de l'évolution chez les procaryotes a permis d'apporter des preuves expérimentales de l'action de la sélection naturelle et d'autres mécanismes impliqués dans l'évolution (van Ditmarsch & Xavier, 2014).

Darwin cependant reconnaissait à son époque la limite de son modèle. Selon ses propres conclusions, la sélection naturelle à elle seule ne suffisait pas à expliquer l'évolution. Comme il le disait « *I am convinced that natural selection has been the main but not exclusive means of modification* ». D'autres théories ont été ainsi développées par la suite pour expliquer la notion d'évolution. Parmi celles-ci on peut citer la théorie synthétique ou néo-darwiniste. Celle-ci actualise le concept de sélection naturelle de Darwin en y apportant la touche qui manquait à ce dernier à son époque, à savoir la génétique. Ici le terme « variant » est remplacé par « mutant ». La théorie neutraliste de l'évolution moléculaire proposée par Motoo Kimura (Kimura & Ohta, 1974; Kimura, 1986) stipule quant à elle que les mutations génétiques échappent à la sélection naturelle et peuvent se diffuser indépendamment de

celle-ci au sein des populations. En effet selon Kimura, dans une population, l'apparition, la diffusion et la fixation des mutations (qui pour lui sont à majorité neutres) surviennent de façon aléatoire à un rythme régulier au cours du temps ; d'où la notion d'horloge moléculaire. La théorie de l'équilibre ponctuée proposée par Eldredge & Gould (1972) viens bouleverser celle de Darwin. En effet selon celle-ci, le changement, au cours des temps géologiques, ne s'était pas fait de manière graduelle comme l'avait proposé Darwin, mais par des phases de stabilité suivies de phases de changement rapides, permettant l'apparition de nouvelles espèces. Enfin la théorie du gène égoïste a été vulgarisée par Dawkins (1981). Cette théorie propose que l'échelle pertinente à laquelle s'applique la sélection naturelle (ou sélection des plus aptes) est celle du gène en qualité de « répliqueur », et non à l'échelle de l'individu ou de l'espèce. Toutes ces théories hormis celle vulgarisée par Dawkins qui repose sur la perpétuation des éléments génétiques mobiles, ne sont que des adaptations du concept de sélection naturelle proposé par Darwin au regard des moyens techniques disponibles à chaque époque pour comprendre et expliquer les mécanismes sous-jacents de la diversité phénotypique observée.

L'évolution phénotypique peut survenir graduellement ou par « saut brutal » qui dépend en fin de compte du tempo et du mode de modification génotypique sous-jacente. Se pose alors la question du lien qui existe entre le phénotype et l'évolution du génome. Les études sur l'évolution expérimentale ont été mises au point pour comprendre tous les mécanismes qui gouvernent les modèles évolutifs dans les populations en les cultivant sur plusieurs générations dans des conditions expérimentales contrôlées. L'évolution expérimentale se définirait alors comme l'étude des processus évolutifs se produisant dans des populations expérimentales en réponse à des conditions imposées par l'expérimentateur. Cette approche de recherche est de plus en plus utilisée pour étudier l'adaptation, estimer des paramètres évolutifs et tester diverses hypothèses évolutives (Kawecki *et al.*, 2012). L'approche de l'évolution expérimentale est complémentaire aux moyens classiques utilisés par les évolutionnistes qui travaillent sur des modèles formés dans des conditions écologiques et à des échelles de temps beaucoup plus long. Cependant les études sur l'évolution expérimentale ont permis d'observer le processus évolutif comme il survient et ont ouvert des possibilités d'inférer directement sur les mécanismes moteurs de l'évolution. L'évolution expérimentale a été, ces dernières années, boostée par la réduction des coûts de séquençage et la croissance de la biologie des systèmes qui tend à démontrer que le comportement d'une population n'est que la somme des interactions établies entre les individus qui la compose.

III.1 Le génome bactérien

L'évolution des techniques de séquençage a permis l'accès aux génomes complets de deux bactéries, *Haemophilus influenzae* et *Mycoplasma genitalium*, il y'a de cela un peu plus de deux décennies (Fleischmann *et al.*, 1995; Fraser *et al.*, 1995) marquant ainsi le début de l'ère de la génomique bactérienne. L'originalité des projets de séquençage de ces procaryotes résidait dans la stratégie utilisée qui était innovante pour l'époque. Aujourd'hui, près de 191.411 bactériens sont disponibles dans les bases de données génomiques et ce nombre ne cesse de croître de façon exponentielle, ceci grâce aux coûts de plus en plus faibles du séquençage. Le plus petit génome séquencé appartient à la bactérie endosymbiote *Candidatus Hodgkinia cicadicola* avec une taille de 105,76 kb. Le plus grand chromosome bactérien jamais séquencé appartient à la bactérie *Minicystis rosea* (Garcia *et al.*, 2014) avec une taille d'environ 16 Mb. La taille des génomes bactériens varie en fonction du mode de vie des bactéries, de leurs conditions de croissance et de la stratégie de nutrition qu'elles possèdent.

Le génome des bactéries a longtemps été décrit comme constitué d'un chromosome circulaire et unique à l'instar de la bactérie modèle *Escherichia coli* (Cairns, 1963). Dans les années 1990, le séquençage du génome de la bactérie *Borrelia burgdorferi* (Baril *et al.*, 1989) a permis de découvrir l'existence de formes linéaires chromosomiques chez les procaryotes. Certaines bactéries comme *Rhodobacter sphaeroides* possèdent quant à elles deux chromosomes. Les génomes des bactéries hormis le chromosome sont souvent complétés par la présence d'éléments génétiques extra chromosomiques qui peuvent être eux aussi circulaires ou linéaires. La comparaison de la distribution des gènes au sein des génomes a permis d'identifier une location préférentielle le long des chromosomes selon leurs fonctions. Les gènes impliqués dans le métabolisme central (réplication, traduction...) essentiels à la survie de l'organisme sont localisés autour de l'origine de réplication (Rocha & Danchin, 2003; Arakawa & Tomita, 2007; Christen *et al.*, 2011). La comparaison des génomes provenant des souches appartenant au même genre ou à la même espèce a permis d'identifier une conservation de l'ordre des gènes sur le chromosome. Cette conservation de l'ordre des gènes au sein de plusieurs génomes comparés est appelée synténie (Charlebois, 1999). Cette synténie diminue avec l'augmentation de la distance phylogénétique séparant les génomes comparés démontrant une dynamique au sein des génomes bactériens (Charlebois, 1999).

III.2 La dynamique du génome bactérien

La dynamique des génomes bactériens résulte des mécanismes qui peuvent être de natures et d'origines différentes.

III.2.1 Variations génétiques dues au transfert horizontal de gènes

Les bactéries sont des organismes asexués, c'est-à-dire qui se multiplient par scissiparité. Une cellule mère se divise en deux cellules filles qui sont génétiquement identiques à cette dernière aux mutations près. Il s'agit du transfert vertical de l'information génétique. Ces bactéries vivent et évoluent dans une large variété de biotopes où les conditions physiques, chimiques et biologiques changent rapidement dans le temps et dans l'espace. Au-delà de la régulation physiologique (adaptation des niveaux d'expression des gènes en réponse aux stimuli environnementaux) et de l'activité des enzymes qui leur permettent de s'adapter à ces changements, les mutations qui sont fréquentes chez les procaryotes y jouent également un rôle déterminant. Cependant, les aptitudes d'adaptations de procaryotes ne peuvent pas s'expliquer majoritairement par les mutations qui sont cumulatives et lentes (très peu d'exemples d'acquisition de nouvelles fonctions ont été décrits par ce mécanisme). En effet le succès évolutif des procaryotes s'explique principalement par leur capacité à échanger du matériel génétique par transfert horizontal. Ce transfert correspond à l'acquisition d'informations provenant d'individus non ascendants. Ce mode de transfert d'information génétique, indépendant de la division cellulaire, est une cause particulièrement efficace de dissémination de gènes conférant des avantages d'adaptation. L'aptitude d'acquisition de gènes par transfert horizontal varie en fonction des organismes et est fortement influencé par les contraintes environnementales. Une analyse complète du génome de la bactérie *E. coli* MG1665 a permis d'estimer que cette dernière avait acquis près de 18% de ses gènes depuis sa divergence avec la lignée *Shigella* (Lawrence & Ochman, 1998). En revanche, chez la bactérie *Buchnera*, symbiote obligatoire des aphides, aucun évènement de transfert horizontal n'as pu être détecté (Tamas *et al.*, 2002). Ainsi contrairement à *E. coli* qui se développe dans l'intestin qui est un écosystème très dense ou de nombreux évènements de transfert horizontal ont été rapportés (Liu *et al.*, 2012), la bactérie intracellulaire *Buchnera* est protégé par son habitat clos de toute acquisition de matériel génétique exogène. Dans certaines conditions, le matériel génétique acquis peut s'intégrer dans le génome de la cellule réceptrice et devenir un élément que cette dernière pourra transmettre à sa descendance. Le transfert horizontal de gènes est donc l'un des moyens les plus efficaces d'adaptation des procaryotes aux changements environnementaux qu'ils rencontrent dans leurs habitats ; il joue également un rôle majeur dans leur évolution (Ochman *et al.*, 2000; Gogarten *et al.*, 2002; Pál *et al.*, 2005; Koonin, 2016).

Les trois principaux mécanismes de transfert horizontal d'ADN sont la transduction, qui consiste à un transfert de matériel génétique d'une bactérie à une autre par l'intermédiaire d'un phage (Lederberg *et al.*, 1951), la transformation définie comme l'acquisition par un organisme d'ADN se trouvant sous forme libre dans l'environnement (Griffith, 1928; Avery *et al.*, 1944), et la conjugaison qui se traduit

par le transfert d'ADN par contact direct entre deux bactéries après formation d'un pore de conjugaison entre les deux cellules (Lederberg & Tatum, 1946) (**Figure 18**).

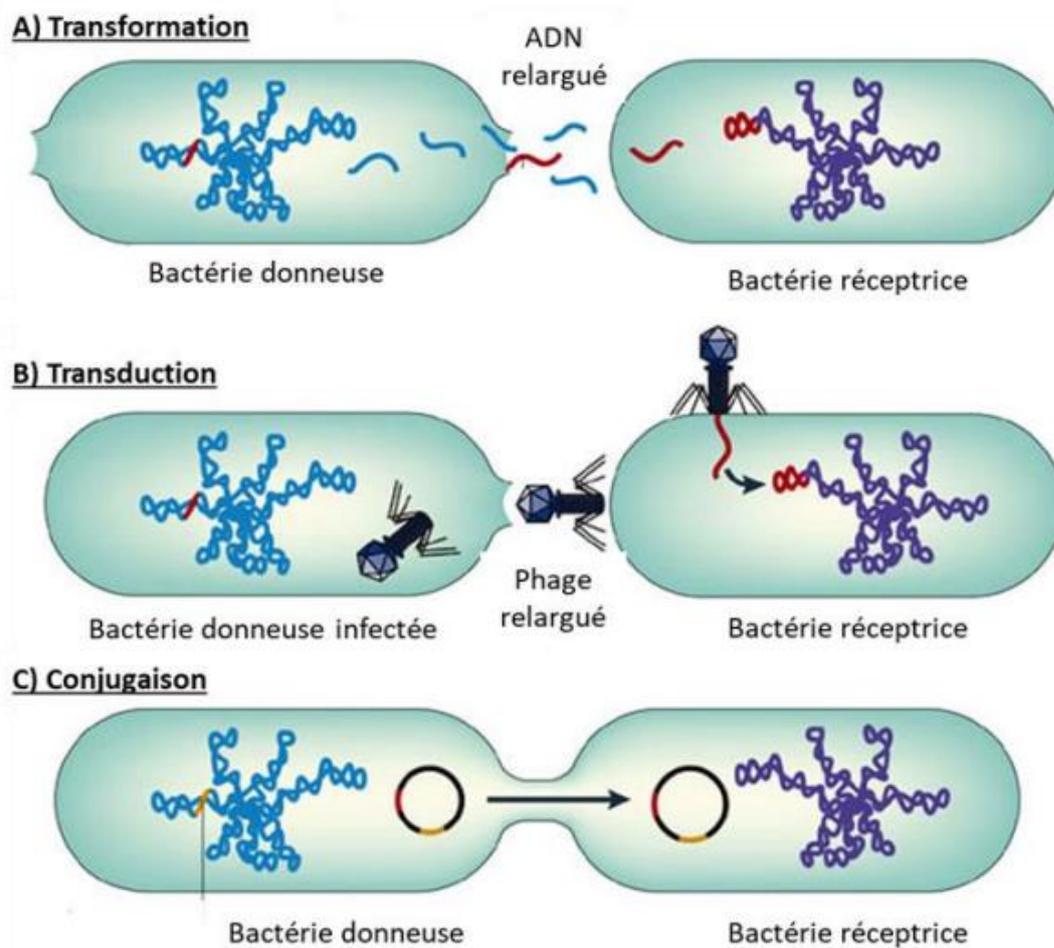


Figure 18 : Représentation des différents mécanismes de transferts horizontaux de gènes : a) la transformation naturelle b) la transduction bactérienne c) la conjugaison bactérienne (Furuya & Lowy, 2006)

III.2.1.1 La transduction

La transduction est le mécanisme par lequel un bactériophage (virus bactériens) délivre de façon unidirectionnelle, du matériel génétique (autre que son propre génome) d'une cellule à une autre. Durant une infection virale classique, le bactériophage injecte son matériel génétique dans une bactérie et détourne la machinerie cellulaire de cette dernière au profit de la fabrication de nouvelles particules virales. Ces dernières sont ensuite libérées dans l'environnement suite à la lyse de la bactérie infectée. Dans le cas de la transduction, le bactériophage va transférer du matériel génétique d'une première bactérie infectée vers une seconde. Il existe deux types de transduction : la transduction généralisée et la transduction spécialisée. Dans la transduction généralisée, le bactériophage injecte son ADN génomique dans la bactérie hôte où il est répliqué par la machinerie cellulaire de cette

dernière. L'ADN de la bactérie hôte étant ensuite fragmenté, ces fragments sont (en fonction de leur taille et à de faibles fréquences) empaquetés dans les particules virales néoformées et délivrés à de nouveaux hôtes infectés par ces bactériophages. Dans la transduction spécialisée dite également localisée, les phages dits « tempérés » intègrent leur matériel génétique dans le génome de leurs hôtes dans un site d'insertion spécifique. Lorsqu'un cycle productif succède à une phase de lysogénie (donc lorsque le répresseur déterminant la lysogénie est inactivé), l'ADN phagique est excisé dans son intégralité ; cependant à de faibles fréquences également, l'excision est anormale et on obtient une molécule d'ADN « hybride » constituée d'un fragment d'ADN phagique et d'un fragment d'ADN bactérien adjacent à la zone d'intégration du prophage. Ce fragment hybride est empaqueté dans des particules virales néoformées et délivrés à de nouveaux hôtes. Une fois intégré, l'ADN du phage est répliqué en même temps que celui de l'hôte et transmis à la descendance sans qu'il n'y ait la lyse de la cellule hôte.

Dans les mécanismes de conjugaison et de transformation, le transfert horizontal de gènes survient entre des individus partageant le même biotope (nécessité de contact entre les différents acteurs) ; l'obligation de présence concomitant de la donatrice et de la réceptrice réduirait le spectre de la transduction comparée à la transformation. Le spectre de transfert de l'information génétique est également plus ou moins large en fonction du mécanisme. Ainsi la transduction nécessitant la présence de récepteurs spécifiques, aura un spectre moins large que la transformation.

III.2.1.2 La conjugaison

La conjugaison est certainement le mécanisme de transfert d'information génétique le plus répandu chez les bactéries (Low, 2001). Elle permet l'émergence de la diversité entre bactéries (Bordeleau *et al.*, 2012). Pendant la conjugaison, deux cellules entrent en contact et l'ADN est transféré de l'une vers l'autre à travers un pore de conjugaison formé entre elles. Plusieurs types d'éléments génétiques mobiles dont les plasmides conjugatifs et les ICE (Integrating Conjugative Elements) utilisent ce mécanisme pour leur transfert. Mais d'un point de vue mécanistique, il existe une différence dans ces processus de conjugaison dépendant des modèles bactériens.

- La conjugaison bactérienne « classique » T4SS dépendante

La conjugaison bactérienne dite « classique », car la plus répandue (chez les bactéries Gram – et les firmicutes) et la plus étudiée implique généralement le transfert de plasmides ou d'ICE (*integrative and conjugative element*). Comme l'indique leur nom, les ICE sont des unités d'ADN présentes au sein du chromosome, qui sont capables de s'exciser et de se transférer vers une cellule receveuse puis de se réintégrer dans le chromosome de la cellule réceptrice (Pour revue : Wozniak & Waldor, 2010). Les mécanismes du processus de conjugaison « classique » ont pour la plupart, été mis en évidence chez

E. coli, mais demeurent transposables à la majorité des espèces bactériennes connues. Le processus commence par l'établissement d'un contact entre deux cellules par la formation d'un pore de conjugaison assemblé grâce à la machinerie de sécrétion de type IV (T4SS) (pour revue : Alvarez-Martinez & Christie, 2009; Christie *et al.*, 2014). Une fois le pore établi, un complexe enzymatique permet la translocation de la molécule d'ADN sous forme simple brin de la cellule donatrice à la cellule réceptrice. Parmi les molécules intervenant dans la translocation, la relaxase induit une coupure simple-brin au niveau de *oriT* et se lie à l'extrémité 5' de ce brin. Ce complexe va alors être reconnu par la protéine de couplage qui va piloter le passage de l'ADN à travers le pore. Lorsque le transfert est achevé, la relaxase reconnaît l'extrémité 3' de *oriT* et recircularise le simple brin transféré par un processus reverse à celle de la coupure. La synthèse du brin complémentaire au brin transféré dans la cellule réceptrice résulte alors en un nouveau transconjugant capable de transférer à son tour cet élément conjugatif.

Dans certaines souches appelées Hfr (haute fréquence de recombinaison), la translocation du brin d'ADN s'accompagne de la mobilisation et du transfert de l'ADN chromosomique en *cis* (à travers la reconnaissance du *cis-acting oriT*). Ce mécanisme de transfert, similaire à celui décrit précédemment, a cependant quelques particularités. En effet l'ADN chromosomique nécessite plus de temps de transfert et n'est pas sujet à la coupure par la relaxase avant la formation du complexe relaxase-*oriT*. Ce transfert chromosomique est souvent incomplet, en raison de la dissociation physique des deux brins d'ADN ou des cassures survenant dans le brin d'ADN à transférer. En conséquence, seul le segment du chromosome donneur adjacent à *oriT* est transféré de manière fiable. Ce transfert partiel d'ADN chromosomique qui empêche la reconstitution de *oriT* n'aboutit pas à la recircularisation du brin d'ADN transféré. Ce dernier ne pourrait s'intégrer et se maintenir dans le chromosome de la réceptrice que par recombinaison homologue. Le transfert chromosomique contraste ainsi du transfert plasmidique par : (i) le fait que les fonctions de recombinaison de l'hôte soient nécessaires pour intégrer l'ADN transféré dans le chromosome de la réceptrice, (ii) le fait que *oriT* n'est pas reconstitué chez la réceptrice suite au transfert (cette dernière ne peut donc pas transférer de nouveau ce fragment d'ADN) et (iii) le fait que les gènes proches de l'extrémité 3' de *oriT* sont plus fréquemment transférés que les autres ; l'efficacité de transfert des gènes chromosomiques dépend donc de leur localisation.

- Conjugaison distributive chez les mycobactéries

Chez les mycobactéries un mécanisme de conjugaison différent de celui initié au niveau de *oriT* a également été décrit (pour review : Gray & Derbyshire, 2018). Cette nouvelle forme de transfert horizontal de gènes a été appelée transfert conjugatif distributif (DCT). Tout en satisfaisant aux critères

de conjugaison dite « classique », le DCT se produit par un mécanisme distinct de la conjugaison médiée par l'*oriT* qu'il pourrait être considéré comme une quatrième catégorie de HGT. Le DCT implique le transfert de l'ADN chromosomique entre les mycobactéries et génère surtout des transconjugants possédants un génome formé d'une mosaïque entre le génome de la donatrice et celui de la réceptrice. Plusieurs segments de l'ADN chromosomique du donneur peuvent être co-transférés indépendamment de leur localisation. En conséquence, le génome de la réceptrice contient de nombreux segments provenant du donneur; d'où le nom DCT. Cette particularité caractéristique du DCT le distingue des autres mécanismes connus de transfert horizontal de gènes. De plus, les transconjugants en mosaïque sont générés à partir d'un seul et unique événement de conjugaison, ce qui fournit une énorme capacité d'adaptation et d'évolution.

- Conjugaison dépendant de TraB

Ce mécanisme de conjugaison décrit chez les actinomycètes, contrairement aux autres, permet la translocation de l'ADN double brin à travers un pore formé par un hexamère de protéines TraB.

Cette partie sera plus détaillée en IV.3 en parlant du modèle d'étude (*Streptomyces*).

III.2.1.3 La transformation naturelle

L'ADN sous forme libre est abondant dans l'environnement. Cet ADN provient essentiellement des cellules lysées ou est excrété par des cellules encore vivantes. L'ADN libre peut garder son intégrité dans l'environnement dans des substrats comme le sol (argile) pendant plusieurs années (Paget & Simonet, 1994). Au cours de la transformation naturelle, l'ADN libre pénètre dans la cellule réceptrice sous forme simple brin dans la mesure où cette dernière est compétente (présence de protéines spécialisées à la surface de la cellule qui lui permettent d'internaliser des fragments d'ADN) (Lorenz & Wackernagel, 1994; Chen & Dubnau, 2004; Thomas & Nielsen, 2005; Johnsborg *et al.*, 2007).

Les gènes transmis de façon horizontale codent généralement pour des fonctions métaboliques autres que les processus moléculaires fondamentaux comme la réplication d'ADN, la transcription et la traduction. Cependant, il existe des exceptions à cette règle et certains gènes ribosomiques montrent une évidence de traces de transfert horizontal (Brochier *et al.*, 2000). L'acquisition des gènes par transfert horizontal peut favoriser l'adaptation des bactéries à une niche écologique si ces gènes sont fonctionnels (Bertolla & Simonet, 1999; Syvanen, 2012). Parmi les avantages acquis par transfert horizontal, on peut citer la résistance aux antibiotiques, la transmission de facteurs de virulence, la conquête de nouvelles niches, et la capacité de dégrader des composés complexes. Hiller *et al.* (2010) ont ainsi montré dans leur étude qu'il existait un fort taux de recombinaison dû au transfert intensif de matériel génétique entre les génomes des souches bactériennes responsables d'une infection

polyclonale des voies aériennes chez un enfant ; et qu'un unique évènement de compétence chez les souches de *Streptococcus pneumoniae* pouvait résulter en un remplacement d'ADN dans plusieurs *loci* non adjacents. Ce transfert de gènes pourrait servir de contrepoint à la réponse immunitaire adaptative de l'hôte et aider à expliquer le phénomène de persistance bactérienne ayant pour but de submerger l'immunité de l'hôte.

III.2.1.4 Fonctions acquises par transfert horizontal de gènes

L'avantage le plus connu conféré par le transfert horizontal de gènes est la résistance aux antibiotiques (Dobrindt *et al.*, 2004; Juhas *et al.*, 2009) qui peut se faire en fonction des espèces par conjugaison, par transformation naturelle et ou par transduction. La caractéristique majeure de l'acquisition de la résistance aux antibiotiques par transfert horizontal est sa vitesse de dissémination dans une population. Cette vitesse de dissémination peut s'expliquer par le regroupement des gènes de résistance en clusters le plus souvent portés par des intégrons. Les intégrons sont des éléments génétiques susceptibles d'acquérir ou de perdre des gènes de résistance aux antibiotiques. Ils peuvent héberger des gènes de résistance insérés sous forme d'éléments mobiles, les cassettes. Celles-ci sont en effet intégrées ou excisées par un système de recombinaison site spécifique qui fait intervenir une intégrase (Stokes & Hall, 1989). Les intégrons constituent donc un système original de capture de gènes capable de promouvoir efficacement la dissémination de la résistance aux antibiotiques au sein du monde bactérien.

Les gènes acquis de façon horizontale ne sont pas répartis de façon homogène le long du chromosome, mais sont regroupés sur les plasmides et dans des régions particulières du chromosome appelé îlots génomiques (Langille *et al.*, 2010). De ce fait à travers le transfert d'un îlot génomique, l'acquisition de plusieurs marqueurs de résistance peut se faire en une seule étape transformant la cellule réceptrice en bactérie multirésistante aux antibiotiques. En outre, comme elles sont transmises d'un hôte à l'autre, les îlots génomiques peuvent accroître le nombre de gènes de résistance qu'ils contiennent. L'autre problème de transfert horizontal de gènes concerne le transfert des facteurs de virulence par les bactéries pathogènes (adhésines, toxines, sidérophores, modulines, etc.). Ces facteurs de virulences sont également localisés, tout comme les résistances aux antibiotiques, à des endroits précis du chromosome ou des plasmides appelés îlots de pathogénicité (Hacker & Kaper, 2000).

Le transfert horizontal joue un rôle important dans la conquête de nouvelles niches à travers l'acquisition de nouvelles capacités métaboliques. Il a en effet été démontré chez une bactérie appartenant au genre *Thermus*, qui pousse en condition d'aérobic stricte, l'acquisition de la capacité à pousser en anaérobiose suite à l'acquisition d'un cluster de gènes ayant une fonction nitrate réductase, permettant à cette bactérie de coloniser les biotopes anoxiques (Ramírez-Arcos *et al.*,

1998). Certaines bactéries ont quant à elles acquis la capacité à résister à certaines substances toxiques de fabrication humaines. Parmi ces capacités on peut citer la précipitation des métaux lourds et radionucléides qui sont des fonctions qui sont disséminées dans des biotopes contaminés (Moënnelocoz *et al.*, 2015).

Cependant cette acquisition d'information génétique peut avoir des effets délétères sur les bactéries réceptrices. Par exemple, la modification d'un réseau de régulation ou la perturbation du métabolisme primaire de la bactérie réceptrice peut s'avérer défavorable (Thomas & Nielsen, 2005; Baltrus, 2013). Il existe pour contrer ces effets délétères dus au transfert horizontal de gènes chez les bactéries, un ensemble de mécanismes de protection permettant de limiter le transfert, le maintien et l'expression des gènes exogènes.

III.2.2 Les barrières au transfert de gène

Plusieurs mécanismes peuvent limiter le transfert, le maintien et l'expression de l'ADN exogène chez la bactérie réceptrice.

III.2.2.1 Les barrières au transfert de gène

Le succès du processus de transfert d'ADN entre deux organismes dépend fortement de leur situation spatiale et temporelle, l'une par rapport à l'autre dans leur habitat, ainsi que de leur capacité à pouvoir interagir entre eux et avec leur environnement (Vos *et al.*, 2013). Le contact physique de la donatrice et de la réceptrice est nécessaire dans le processus de conjugaison. Dans le cas de la transformation, la condition *sine qua non* reste la persistance dans l'environnement de l'ADN sous forme libre (qui peut également provenir de la lyse des cellules sœurs non compétentes par celles compétentes (Claverys *et al.*, 2007)) et de la concordance entre cette persistance et l'état physiologique de compétence de la bactérie (Nielsen *et al.*, 1998; Cohan, 2002b).

Des barrières mécanistiques telle que l'exclusion membranaire peuvent également empêcher le transfert du matériel génétique entre les bactéries. Ainsi, chez la bactérie *E. coli*, la synthèse des protéines issues de l'expression des gènes *traS* et *traT* empêchent la fixation du pilus nécessaire au transfert du matériel génétique par conjugaison (Achtman *et al.*, 1980). L'expression de ces deux protéines transforme les cellules en mauvaises réceptrices d'ADN vis-à-vis des autres cellules exprimant les mêmes gènes tout en restant bonnes donatrices. La protéine TraT produit l'exclusion de surface en inhibant la formation des agrégats de conjugaison et TraS inhibe quant à elle le transfert d'ADN au sein d'un agrégat de conjugaison déjà formé.

Dans le cas de la transduction, l'étape initiale et nécessaire au transfert de l'ADN viral dans la cellule hôte reste l'adsorption du bactériophage à la surface de cette dernière. Le bactériophage a donc

besoin de reconnaître spécifiquement des récepteurs, sur la surface de l'hôte, à ses protéines de liaison (RBPs) (Achtman *et al.*, 1980). L'expression stochastique des récepteurs, la mutation des gènes responsables de leur synthèse ou l'expression des protéines de surface chez l'hôte camouflant ces récepteurs sont autant de facteurs pouvant inhiber la fixation du bactériophage et donc limiter le transfert du matériel génétique (Samson *et al.*, 2013).

III.2.2.2 Les barrières à l'intégration et au maintien

Les barrières à l'intégration et au maintien de l'ADN dans la cellule réceptrice peuvent être multiples. Les systèmes de « restriction-modification » (RM) ont émergé au cours de l'évolution à cause du risque très élevé que représentent les phages pour les bactéries (Oliveira *et al.*, 2016). Les enzymes de restriction réduisent considérablement l'absorption d'ADN étranger par les bactéries. Cela réduit considérablement l'acquisition de gènes étrangers à des niveaux tolérables. Ils coupent en fragments des molécules d'ADN étrangères lors de leur pénétration dans les bactéries réceptrices (Arber, 2000). Les systèmes RM reconnaissent une séquence donnée de 4 nucléotides ou plus au niveau de l'ADN entrant. L'état de méthylation de l'ADN entrant permet également à ce système de réaliser une discrimination entre ce dernier et l'ADN de l'hôte. Dans cette lutte perpétuelle entre les bactéries et les phages, ces derniers vont muter la séquence reconnue par les systèmes RM, échappant ainsi à la dégradation lors de l'infection. Les bactéries quant à elles vont aussi muter en augmentant le niveau de reconnaissance de leur système élargissant ainsi le panel de séquences reconnues par le système RM qu'elles portent.

Le système CRISPR-Cas (*clustered regularly interspaced short palindromic repeats-CRISPR-associated proteins*) est un autre moyen de défense des bactéries contre l'intégration de l'ADN exogène dans leur génome (Barrangou *et al.*, 2007; Marraffini & Sontheimer, 2008). Le système CRISPR-Cas est composé de courtes répétitions (24 à 48 nucléotides) séparés par des séquences uniques de même taille appelées « spacer » (Grissa *et al.*, 2007; Sorek *et al.*, 2008), le tout flanqué des gènes codant pour des protéines (*CRISPR-associated (cas) proteins*) (Haft *et al.*, 2005; Barrangou *et al.*, 2007) qui fournissent le mécanisme enzymatique nécessaire à l'acquisition de nouveaux « spacers » à partir de l'ADN entrant ainsi qu'à la dégradation de ce dernier (Richter *et al.*, 2012). C'est un système qui agit comme une mémoire proto-immunologique de la bactérie suite à un premier contact avec un phage ou avec de l'ADN exogène.

L'intégration d'ADN exogène ou sa répllication par le génome de la bactérie réceptrice sont les moyens pérennes de maintien de cet ADN. L'une des limites à l'intégration d'ADN exogène est le niveau d'identité nucléotidique entre l'ADN exogène et le génome de la bactérie réceptrice. L'identité nucléotidique étant requise par le mécanisme de la recombinaison homologue (Shen & Huang, 1986;

Majewski *et al.*, 2000; de Vries & Wackernagel, 2002). Des études ont en effet démontré chez *E. coli* qu'une réduction de l'homologie de séquences de 100% à 90% réduisait de 40 fois le taux de recombinaison observé (Shen & Huang, 1986). Cette barrière limite donc le transfert inter-espèces et serait donc à l'origine de l'isolement génétique entre les bactéries. Dans le cas des éléments génétiques mobiles, l'insertion site-spécifique requiert un mécanisme d'intégration propre et permet de contourner cette barrière (Alberts *et al.*, 2002). En effet, contrairement à la recombinaison homologue qui nécessite une forte homologie de séquence pour se produire, la recombinaison site-spécifique est guidée par une enzyme qui reconnaît une courte séquence nucléotidique présente sur une ou les deux séquences qui recombinent. Chaque type d'élément génétique mobile code généralement pour l'enzyme qui va médié son intégration, et contient les sites spécifiques sur lesquels agit cette enzyme (Alberts *et al.*, 2002).

III.2.2.3 Les barrières à l'expression génique

L'expression des gènes portés par l'ADN exogène ayant réussi à s'intégrer dans le génome de l'hôte peut être limité par divers facteurs dont la différence de pourcentage en G+C. Une étude réalisée sur 2.670 génomes de procaryotes dont le pourcentage G+C varie de 13,5 à 74,9% a permis de déterminer que, quel que soit la distance phylogénétique séparant des souches, un contenu similaire en G+C des génomes les obligeait à utiliser la même base pour les mêmes triplets de codons. Cette étude a démontré que dans le génome des procaryotes, l'utilisation des bases, des codons et des acides aminés change en fonction du contenu en G+C (Zhou *et al.*, 2014).

L'expression des gènes acquis par transfert horizontal peut aussi être limité par la toxicité du produit d'expression de ces gènes qui peut être nocif et délétère pour la bactérie réceptrice (Avni & Snir, 2017). Il faut noter que seule une faible proportion de l'ADN acquis par transfert horizontal se maintient de façon durable dans le génome des cellules réceptrices (Thomas & Nielsen, 2005); et ce maintien de l'ADN exogène ne sera pérenne dans l'hôte que s'il demeure neutre ou confère un avantage à ce dernier.

Si le transfert horizontal de gènes survient fréquemment entre les espèces bactériennes apparentées, il est aussi possible entre les individus distants sur le plan évolutif (entre les bactéries à Gram positif et les bactéries à Gram négatif) et parfois entre les individus appartenant à différents domaines ; c'est le cas de la bactérie *Agrobacterium* qui utilise comme stratégie d'infections des plantes, un transfert horizontal de matériel génétique résultant en une intégration stable de ce dernier dans le génome de l'hôte (Bourras *et al.*, 2015). De cette même manière, des protéines caractéristiques des bactéries sont codés par les *Archaea* et vice-versa. Le génome de la bactérie hyperthermophile *Thermotoga maritima* contient entre 8 et 11% de gènes de l'archaea hyperthermophile (Zhaxybayeva *et al.*, 2009). Cette

présence de gènes de archaea dans le génome de la bactérie *T. maritima* résulterait du transfert horizontal entre les deux organismes qui partagent le même biotope que représente les sources thermales chaudes. Donc à travers le transfert horizontal de gènes, en une seule étape, des barrières correspondantes à des millions d'années d'évolution pourraient être franchies.

III.2.3 Les réarrangements chromosomiques

La réplication et les mécanismes de réparation des lésions de l'ADN sont des processus pouvant engendrer des réarrangements chromosomiques. Ces réarrangements peuvent être des inversions, des duplications ou des délétions. Ils ont plusieurs origines. Ils peuvent survenir suite à des événements de recombinaison homologue entre des séquences répétées de tailles variables, directes ou inversées (Roth *et al.*, 1996) (**Figure 19**). Ces remaniements chromosomiques peuvent également provenir de la recombinaison illégitime entre des séquences non homologues, notamment suite à une cassure double brin de l'ADN. Les réarrangements chromosomiques peuvent avoir diverses conséquences sur le plan fonctionnel. Les événements de délétion entraînent la perte irréversible d'information génétique. Les événements de duplication quant à eux n'entraînent pas de perte d'informations génétiques tant qu'ils n'impliquent pas l'interruption d'un élément fonctionnel comme un gène ou un opéron. La duplication peut cependant engendrer la colocalisation de deux copies d'un gène ou d'une région génomique (en tandem) ou ces dernières peuvent être localisées à distance l'une de l'autre sur le chromosome. Les événements de duplication peuvent aboutir à une surexpression des régions codantes situées dans la zone dupliquée par effet de surdosage. Les inversions n'entraînent pas de gain ou de perte d'informations génétiques. Cependant, tout comme avec les événements de délétion et de duplication, les séquences au point de jonction peuvent perdre ou conserver leur caractère fonctionnel. Les inversions de grandes tailles peuvent ou pas perturber de façon importante l'architecture du chromosome.

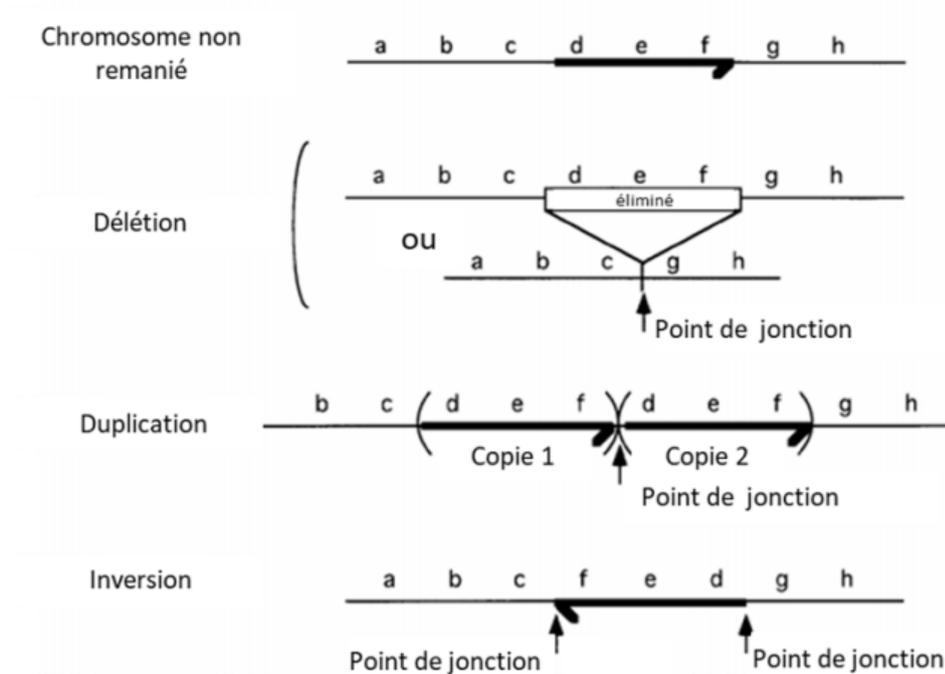


Figure 19 : Illustration des différents types de remaniements génomiques. La ligne horizontale représente le chromosome bactérien. La demi-flèche en noir foncé et en gras délimite une portion du chromosome sujet à des remaniements. Les lettres en minuscules représentent la succession d'éléments génétiques dans leur ordre de succession sur le chromosome. (Adaptée de Roth *et al.*, 1996).

III.2.4 La variation par mutation

Une mutation, tout comme un réarrangement, correspond à une altération du matériel génétique (substitution, insertion, délétion), survenant à l'échelle génique (affectant un ou plusieurs nucléotides) de façon aléatoire. Les mutations peuvent être causées par un mauvais fonctionnement de l'une des ADN polymérases ou par des mécanismes de tautomérisation qui engendrent des *mismatches* sans erreurs répliquatives ni dommages. La première évidence du caractère aléatoire des mutations a été apportée par Luria & Delbrück (1943) en utilisant comme modèle la souche d'*E. coli*. En effet ces auteurs ont démontré qu'*E. coli* était capable de résister à l'infection virale grâce à des mutations préexistantes indépendantes du phage; ces mutations apparaissant spontanément au cours de la phase exponentielle de croissance chez cette bactérie avant tout contact avec le phage. L'une des conséquences de la rapidité de reproduction des procaryotes est leur capacité à produire un grand nombre de mutants qui seront soumis à l'action de la « sélection naturelle ». Ces mutations pouvant être favorables, défavorables voir létales ou neutres en termes de survie et de croissance d'un organisme dans un environnement biotique ou abiotique donné. Les mutations neutres n'auront aucun effet sur la croissance de l'organisme et celles délétères seront éliminées soit par la sélection

naturelle ou par la recombinaison (élimination physique) avec acquisition de mutations non délétères, ceci sans contre sélection. Les mutations favorables conférant une adaptation à un nouvel environnement par exemple permettront à l'individu les portant de prospérer et de coloniser cet environnement

Les mutations peuvent être spontanées ou induites.

III.2.4.1 Les mutations spontanées

Les mutations, qu'elles soient spontanées ou induites, résultent d'un processus naturel. En effet, au cours de la réplication de l'ADN chez les bactéries et certains eucaryotes, les ADN polymérase I, II et III font des erreurs d'incorporation de bases qu'elles peuvent spontanément corriger grâce à leur activité de relecture (« proofreading »). Au cas où ces erreurs d'incorporation de bases ne sont pas corrigées par le système de relecture, elles pourront l'être ultérieurement par le système de correction de mésappariement qui va assurer la validité des bases incorporées dans le brin néoformé par contrôle de l'appariement. Toutefois, des erreurs peuvent subsister à un taux estimé entre $1.5 \cdot 10^{-11}$ et $5.4 \cdot 10^{-10}$ par base et par génération (Ochman *et al.*, 1999; Wielgoss *et al.*, 2011). Ces fréquences d'erreur très faibles attestent de la stabilité relative de l'information codée. Ces mutations qui font partie du processus d'évolution naturelle et qui surviennent de façon aléatoire sur les séquences nucléotidiques sont essentielles à l'évolution des organismes (Radman, 1999).

III.2.4.2 Les mutations induites

Le second type de mutations qu'on rencontre est celui qui est induit par l'action d'agents exogènes et ou endogènes (mutagènes). Qu'il soit d'origine physique (rayonnements ultra-violetts, radiations ionisantes), chimique (ex. agents alkylants) ou biologique (mitomycine C, aflatoxine, bléomycine), les agents mutagènes agissent tout comme les mutations spontanées de façon aléatoire sur le génome. Le système SOS est l'exemple typique d'un système d'induction de mutations. Ce système coordonne la réponse à un dommage de l'ADN. Il comprend non seulement des mécanismes de réparation correcte de l'ADN (excision) dont le but est d'assurer la survie de la bactérie suite à un stress environnemental ayant altéré le génome (d'Ari, 1985), mais aussi un système mutagène (*UmuCD*, *dinB*). Le système SOS peut être déclenché par les dommages générés par les rayonnements ultra-violetts, les composés chimiques ou oxydatives, les acides, les mutagènes organiques, certains antibiotiques et les espèces réactives d'oxygène (ROS). Ce système a été décrit pour la première fois chez *E. coli* (Radman, 1975).

III.2.4.3 Les hypermutateurs

L'une des propriétés expliquant le succès évolutif des procaryotes dans l'environnement est la présence des bactéries ayant un taux de mutations spontanées 100 à 1000 fois supérieurs aux

bactéries dont elles dérivent. Ces bactéries sont appelées hypermutatrices (Taddei *et al.*, 1997). Cette hypermutabilité peut être observable et quantifiable en temps réel dans les expériences d'évolution expérimentales (Barrick & Lenski, 2013; Raynes & Sniegowski, 2014 pour revue). Le taux de mutation étant déterminé par la fidélité avec laquelle l'information génétique est répliquée et réparée. Les gènes *mutT* et *mutY* codent respectivement pour une hydrolase qui élimine de la cellule les guanine oxydés (8-oxo-dGTP) qui sont capables de s'apparier avec des adénines, entraînant une réversion A:T -> G:C après la réplication de l'ADN, et pour une glycosylase de réparation d'ADN, qui excise les mésappariements dans la molécule d'ADN. Une mutation dans chacune de ces gènes individuellement a précédemment été décrit comme responsable de l'apparition de forts taux de mutations chez *E. coli* (Bhatnagar & Bessman, 1988; Rotman & Kuzminov, 2007). Cependant, une mutation du gène *MutY* dans une souche mutant pour le gène *mutT* permet de réduire de 60% le taux de mutation comparé aux mutants uniques (Rotman & Kuzminov, 2007). Une déficience de l'un des systèmes de réparation (ex. *mutT* ou *mutY*) (Bhatnagar & Bessman, 1988; Wielgoss *et al.*, 2013) et ou une mutation de la sous-unité de l'ADN polymérase responsable de l'activité de relecture de l'ADN (Degnen & Cox, 1974; Michaels *et al.*, 1990) peuvent conduire à l'apparition de ce phénotype hypermutateur (Bridges, 2001).

L'apparition des hypermutateurs est moindre dans des environnements stables ; mais elle augmente dans des conditions de stress où seules les bactéries capables d'acquérir beaucoup plus de mutations pour contrer l'effet du stress (par apparition d'allèles adaptés) sont capables de survivre. Les bactéries hypermutatrices ont été décrites chez diverses espèces (*E. coli*, *Salmonella enterica*, *Neisseria meningitidis*, *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Pseudomonas aeruginosa*, etc.). L'hypermutabilité capable de créer la diversité est un avantage sélectif considérable pour les bactéries lorsque la population rencontre un changement dans son environnement (Thomas *et al.*, 2016). Les hypermutateurs ont une vitesse de croissance supérieure à celle de la souche sauvage qui possède un faible taux de mutation. Cependant l'augmentation du taux de mutation chez les bactéries hypermutatrices impliquant aussi les gènes essentiels à la survie de ces bactéries, le nombre de mutations létales chez ces dernières serait supérieure à celles de la souche sauvage. Les hypermutateurs sont donc capables de s'adapter rapidement, mais sont beaucoup plus sujets à des mutations létales. Cependant malgré le nombre important de mutations délétères survenant chez les hypermutateurs, la faible portion des mutations favorables confère un avantage indéniable à ces derniers, leur permettant ainsi une meilleure adaptation à leur environnement. La perte du caractère mutateur qui intervient à des fréquences également élevée permet de stabiliser un génotype adapté.

Le succès évolutif des bactéries dépend donc du réglage constant du taux de mutations survenant au sein de leur génome ; ce qui optimise leur adaptation à des conditions environnementales en permanente évolution. L'adaptation des bactéries à leurs environnements a longtemps été considérée

comme graduelle et résultante d'une sélection constante de mutations ayant peu d'effets (Fisher, 1931). Cependant un modèle d'évolution alternatif stipule que des périodes rapides d'évolutions phénotypiques se succèderaient avec des périodes d'arrêts (Wright, 1932; Eldredge, 1972; Eldredge *et al.*, 2005). Les causes de ces variations restent mal connues et sont probablement diverses : changement de paramètres environnementaux (ex. colonisation de nouveaux habitats), modulation des flux de gènes (Reznick *et al.*, 1997; Schluter, 2000; Blount *et al.*, 2008).

III.2.5 Évolution des fréquences alléliques dans une population

La quantité des variants alléliques au sein d'une population est régulée par deux mécanismes que sont la sélection naturelle et la dérive génétique. La sélection naturelle résultant de la pression de sélection imposée par un environnement sur la diversité génétique présente, retient uniquement les allèles qui confèrent un meilleur succès reproductif et par association dans certains cas, des allèles non adaptés; ces allèles sont donc propagés au sein de la population. Cette adaptation d'individus à leur environnement est appelée « fitness ». Quant à la dérive génétique, elle correspond à la variation de fréquences alléliques due au hasard, ceci indépendamment de l'avantage que pourrait conférer un allèle et de l'intervention d'autres forces évolutives. La dérive génétique ferait donc varier les fréquences de mutations sans tenir compte du fitness qu'elles pourraient conférer aux individus.

La fréquence des allèles au sein d'une population va donc dépendre de l'équilibre entre la sélection naturelle (conférant des mutations adaptatives) et la dérive génétique (mutations neutres) ainsi que de la balance entre l'effet physiologique des mutations sur le fitness des bactéries et leurs coûts énergétique (coûts énergétique à maintenir bas le taux de mutations) (Drake *et al.*, 1998). Les travaux de Sniegowski *et al.* (1997), Mao *et al.* (1997) et Oliver *et al.* (2000) ont permis de déterminer qu'au sein d'une population naturelle, la fréquence des mutations était beaucoup plus élevée que celles prédites en laboratoire sur la base de l'équilibre entre les mutations et la sélection. Ces études suggèrent donc qu'il existe dans l'environnement des situations telles que la densité des cellules présentes et la multiplicité des défis environnementaux (changements des conditions environnementales entre autres) à relever qui favoriseraient la fréquence élevée des mutations.

III.2.6 La notion de pangénome, génome « core » et génome accessoire

L'avènement du séquençage à haut débit a permis des avancées considérables dans le domaine de la génomique comparative. Les études de génomiques comparatives ont montré que le nombre de gènes présents dans les souches bactériennes appartenant à la même espèce est très variable. Cette situation a conduit à distinguer au sein du génome bactérien, le pangénome qui comprend l'ensemble des gènes présents au sein de tous les individus d'une population ou d'une espèce (McInerney *et al.*, 2017), le génome 'core' qui comprend les gènes Partagés par l'ensemble des membres d'une population ou

d'une espèce, et le génome accessoire qui est défini par l'ensemble des gènes présents au sein d'un ou d'une partie des génomes d'un groupe ou d'une espèce.

La disponibilité des milliers de génomes de procaryotes a permis de révéler l'omniprésence des introgressions de toutes sortes (Baptiste *et al.*, 2012). Certaines espèces bactériennes possèdent un pan génome ouvert alors que d'autres possèdent des génomes possédants peu de différences en terme de contenu en gènes (pan génome fermé). Le plus grand pan génome jusqu'alors analysé pour une espèce bactérienne est celui de *E. coli* avec près de 2.085 génomes comparés. Cette comparaison a estimé à 3.188 gènes ceux appartenant au core génome (déterminés comme présents dans 95% des génomes) et approximativement 90.000 gènes accessoires (Land *et al.*, 2015). En contraste, le pathogène intracellulaire *Chlamydia trachomatis* a un pan génome dont la taille est à peine plus grand que celui du core génome (974 gènes dans le pan génome contre 821 dans le core génome) avec 67 génomes séquencés (<http://pangenome.tuebingen.mpg.de>). Cela permet d'avoir une idée de l'échelle de variation du core génome par rapport au pan génome qui peut aller de 3 à 84% pour les génomes bien séquencés. La taille du core génome varie inversement proportionnellement au nombre de génomes séquencés. Plus ce nombre augmente plus la taille du pan génome aussi augmente. De façon intéressante, une analyse de la présence et absence de gènes dans un échantillon de 573 génomes et son extrapolation à un nombre plus grand de génomes a permis de déterminer que la taille du pan génome des bactéries était infinie (Lapierre & Gogarten, 2009). Cela s'expliquerait par une « constante pluie de matériels génétiques sur les génomes » (Lapierre & Gogarten, 2009) et impliquerait que les génomes possèdent une réserve illimitée de gènes dans laquelle elles pourraient piocher.

L'origine du pan génome au sein d'une espèce bactérienne a fait débat ces derniers temps entre des équipes travaillant sur la génétique des populations bactériennes et sur les processus évolutifs. Selon McInerney *et al.*, (2017), le processus menant à la génération du pan génome nécessite des explications théoriques qui prennent en compte le fait qu'il existe une grande distribution de la taille du pan génome allant des plus petits aux plus ouverts. Selon lui les transferts horizontaux sont des formes de mutations qui doivent être considérés comme telle dans les modèles d'évolution de pan génome et que ces modèles devraient également tenir compte de la variation de la taille effective de la population, du taux de mutations, des coefficients de sélection, l'influence de la dérive génétique, le type de spéciation et la particularité de certains procaryotes à avoir des pan génomes très ouverts. L'entrée de nouveaux gènes ou allèles dans une cellule (taux d'acquisition de matériel génétique) ne suffit pas à assurer leur intégration dans le génome des cellules réceptrices (taux d'efficacité d'intégration) (Shapiro, 2016). La forte proportion des éléments génétiques mobiles et des ADN exogènes font penser que la délivrance d'ADN aux bactéries est fréquente, la question reste alors de

savoir qu'est ce qui promeut la rétention de cet ADN et pourquoi n'existe-t-il alors pas un génome typique pour chaque espèce bactérienne ?

McInerney *et al.*, (2017) ont proposé un modèle de genèse et de maintien de pan génomes basé sur l'existence de nombreuses niches écologiques combinée à la sélection naturelle. Dans ce modèle, ils suggèrent que les gènes acquis par transfert horizontal sont capables de conférer pour la majorité (mais pas tous) un avantage adaptatif et que la présence du pan génome ne serait alors qu'un phénomène adaptatif mais pas dans le sens de la purge sélective communément nommée « selective sweep ». Selon McInerney même si la théorie évolutive standard qui stipule que l'acquisition d'un allèle conférant un avantage sélectif et sa fixation par la sélection naturelle (selective sweep) tend à réduire la diversité dans une population, serait en contradiction a première vue avec le modèle qu'il propose, la conciliation entre les deux résiderait dans le modèle de nouveaux compartiments de Niehus *et al.*, (2015). Le modèle de nouveau compartiment de Niehus *et al.*, (2015), qui explicite le transfert horizontal de gènes et la migration des bactéries, montre l'implication de la sélection naturelle sur les gènes issus du transfert horizontal et son rôle dans la différenciation des populations. En utilisant un modèle mathématique, Niehus a montré qu'en cas d'avantage sélectif conféré par un gène issu du transfert horizontal, la diversité génétique est éliminée de la population lorsqu'il n'existe pas de migration dans ou en dehors du compartiment ou niche occupé par cette population bactérienne. Dans le cas échéant, une combinaison de la migration et de l'acquisition d'un gène conférant un avantage sélection n'aboutirait pas théoriquement à l'élimination de la diversité génétique au sein de la population. Même si l'objectif de ce modèle n'est pas d'expliquer le pan génome, il permet cependant de montrer que la diversité au sein d'une espèce pouvait se maintenir même en cas d'acquisition d'avantage sélectif par un variant d'une population. L'exemple serait l'espèce *E. coli* qui est capable de bouger et d'aller d'un tractus digestif à l'autre et qui possède un pan génome ouvert alors que chez la bactérie intracellulaire obligatoire *Chlamydia trachomatis*, les variant doivent obligatoirement rentrer en compétition locale avec les souches sauvages. Donc pour McInerney *et al.*, la taille du pan génome d'une population serait déterminée par la taille de cette population et l'aptitude des membres de celle-ci à migrer vers de nouvelles niches.

Ce modèle proposé par McInerney *et al.*, ne trouve pas écho auprès de Andreani *et al.*, (2017) qui jugent plutôt que le « *turn over* » (changement) du génome accessoire (partie variable du pan génome déterminant la taille de cette dernière) serait plutôt régis par l'évolution neutre du génome (*neutral evolution*). Cependant selon Shapiro (2017)), les deux points de vues peuvent être conciliés en acceptant que les populations de grande taille renferment un taux élevé de diversité génétique neutre et répondraient de façon plus efficace à la sélection naturelle. Donc certaines corrélations entre la

taille de la population et la variation du pan génome observé par Andreani *et al.* (2017) pourraient s'expliquer pour certaines par l'évolution neutre et pour d'autre part l'adaptation à la niche.

IV. Le genre *Streptomyces*

IV.1 *Streptomyces* et son cycle de vie

Les bactéries du genre *Streptomyces* sont des bactéries à coloration Gram positive appartenant à l'embranchement des actinobactéries et au groupe des actinomycètes. Ce sont de bactéries filamenteuses et sporulantes, ubiquistes des sols. Les *Streptomyces* jouent un rôle important dans l'écosystème forestier (Uroz *et al.*, 2012; Bontemps *et al.*, 2013) et participent au cycle biogéochimique comme le cycle du carbone grâce à leur puissant arsenal enzymatique. Elles sont pour la plupart saprophytes à l'exception de *Streptomyces scabies* et *S. somaliensis* qui sont respectivement pathogène de la pomme de terre (gale commune de la pomme de terre) et pathogène opportuniste de l'homme (actinomycetome). Ce genre bactérien est connu pour sa production d'un grand panel de métabolites spécialisés ayant des propriétés antibiotiques, antifongiques, antiparasitaires et anticancéreux. En 2005 une étude estimait à 53 % (8.700 sur 16.500) la proportion d'antibiotiques connus produits par les actinomycètes dont 40 % (6.550) sont produits par le seul genre *Streptomyces* (Bérdy, 2005). Contrairement à la majeure partie des bactéries qui se multiplient par fission binaire, *Streptomyces* possède un cycle de développement complexe (Chater, 1993; Champness, 2000; Elliot *et al.*, 2008) (**Figure I10**). Son cycle cellulaire commence par la germination d'une spore qui va donner des hyphes. Le passage de l'étape de spore aux hyphes est concomitant à la réplication de l'ADN génomique qui aboutit à la formation des hyphes pluri-génomiques. Ces hyphes par une croissance radiale vont former un réseau ramifié appelé mycélium végétatif. Lorsque le milieu devient pauvre en éléments nutritifs, le mycélium végétatif va se différencier pour donner le mycélium aérien qui est composé d'hyphes non ramifiés. La différenciation du mycélium aérien s'achève par une septation asynchrone et par la formation de chainettes de spores qui sont la forme de résistance et de dissémination de ces bactéries. Dans des conditions favorables, ces spores vont initier un nouveau cycle. La différenciation morphologique s'accompagne d'une forte activité du métabolisme secondaire et notamment, la production d'antibiotiques durant la différenciation du mycélium végétatif en mycélium aérien (Chater, 2006) (**Figure I11**)

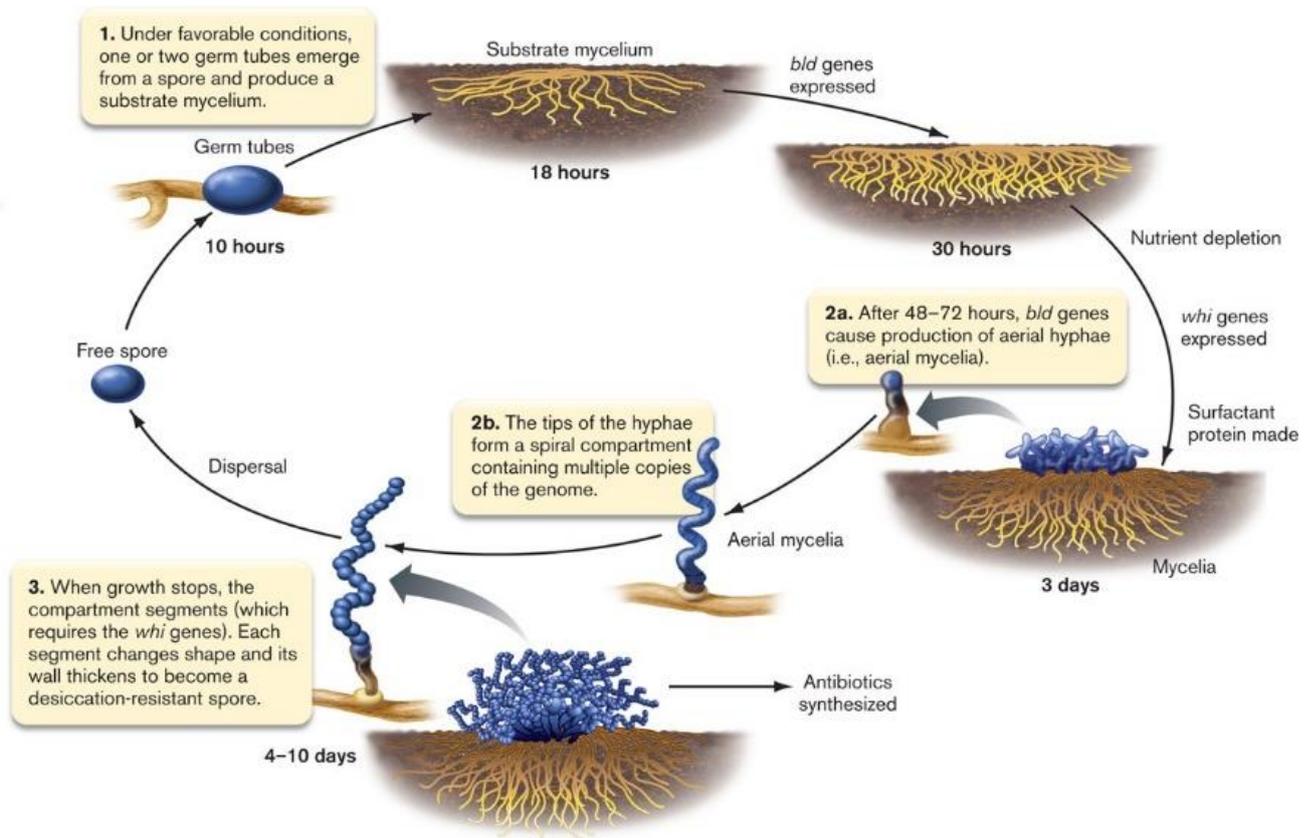


Figure I10 : Cycle de développement des bactéries du genre *Streptomyces* : <https://cellcode.us/quotes/Streptomyces-scabies-life-cycle.html>

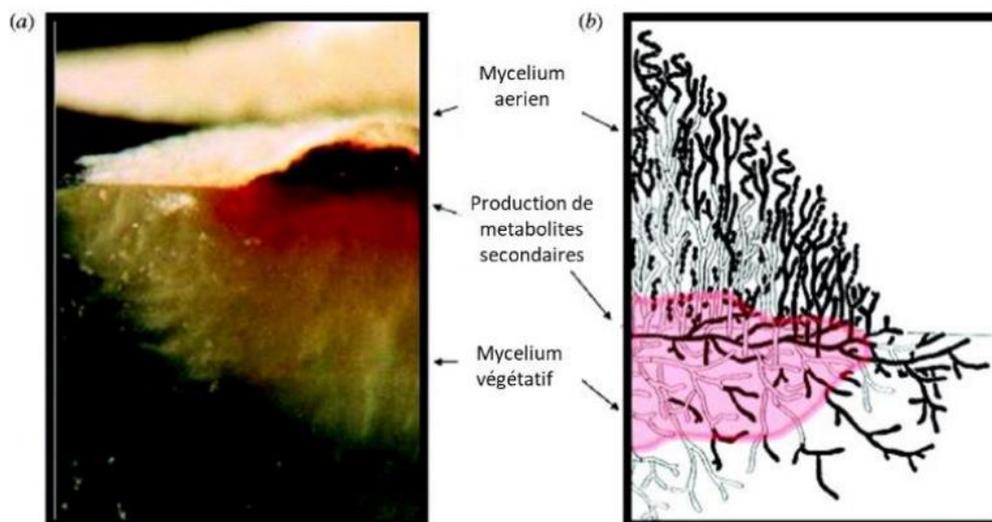


Figure I11 : Illustration de la morphologie d'une colonie du genre *Streptomyces* produisant des métabolites secondaires sur une gélose (Chater, 2006).

IV.2 Une organisation chromosomique atypique

Au-delà du cycle de développement atypique, les bactéries du genre *Streptomyces* ont une organisation chromosomique remarquable. Elles possèdent un chromosome linéaire dont la taille varie de 6 à 12 Mb et riche en G+C (71 à 73 %) (Lin *et al.*, 1993). Ce génome est souvent complété par la présence de plasmides de grande taille. La linéarité du chromosome de *Streptomyces* résulterait de la recombinaison entre un chromosome circulaire ancestral et un plasmide linéaire (Volff & Altenbuchner, 2000). L'origine de réplication (*oriC*) du chromosome de *Streptomyces* se situe approximativement au centre et chaque extrémité du chromosome est souvent constituée d'une séquence terminale répétée et inversée (TIR pour terminal inverted repeats) de composition et de taille variable en fonction des espèces. Les TIR possèdent à leur extrémité des séquences, appelés télomères, contenant des répétitions inversées palindromiques capables former des structures secondaires. Ces télomères sont coiffés par des protéines terminales appelées TP (pour Terminal Proteins) qui leur sont liées de façon covalente (Bao & Cohen, 2001; Chen *et al.*, 2002). Les TP jouent un rôle important dans la synthèse du brin complémentaire au brin sens ayant l'extrémité 3' libre, lors de la réplication du chromosome ou du plasmide linéaire de *Streptomyces* (Chen *et al.*, 2002). Cette protéine terminale assure donc la conservation de la linéarité du chromosome et des plasmides de *Streptomyces* au cours de la réplication (Bao & Cohen, 2003) ainsi que la protection des extrémités contre la dégradation par les exonucléases (Chen *et al.*, 1993). Chez *Streptomyces*, les TP sont très conservés en séquence et en taille (environ 185 acides aminés) et sont codés par les gènes *tpg* qui se trouve en aval du gène *tap* dans le même opéron (Bao & Cohen, 2001; Yang *et al.*, 2002). La protéine Tap est également essentielle à la réplication des réplicons linéaires de *Streptomyces*. Cette protéine s'attache de façon spécifique à une structure secondaire formée l'extrémité simple brin 3' libre durant la réplication et serait présumée recruter la protéine Tpg au niveau des télomères pour réaction finale de synthèse du brin complémentaire au brin sens (Bao & Cohen, 2003). L'opéron *tap-tpg* se retrouve le plus souvent au niveau des régions terminales du chromosome de *Streptomyces*. Quelques réplicons linéaires ont également été décrits comme possédant l'opéron *tap-tpg* ou des homologues de gènes *tpg* seuls. En plus des opérons *tap-tpg*, des pseudogènes de gènes *tpg* ont été retrouvés dans les génomes séquencés des souches de *Streptomyces* (Spatz *et al.*, 2002; Ikeda *et al.*, 2003; Choulet *et al.*, 2006b; Ohnishi *et al.*, 2008; Suzuki *et al.*, 2008).

IV.3 Un mécanisme de conjugaison dépendant de TraB

Outre le mécanisme classique de conjugaison dépendant de système de sécrétion de type IV (T4SS) décrit précédemment, il existe chez les actinomycètes dont *Streptomyces*, un autre mécanisme de conjugaison dépendant uniquement du gène *traB* dont la protéine reconnaît une petite séquence non codante appelée *clt* (*cis acting locus of transfer*) (Thoma & Muth, 2012; Thoma & Muth, 2015) (**Figure**

I12). La protéine TraB homologue de FtsK, l'enzyme clé de la ségrégation des chromosomes lors de la division cellulaire, est une ATPase qui par multimérisation (hexamère) forme un anneau. Cet anneau constitue le pore de conjugaison par lequel est transloqué l'ADN après reconnaissance de la séquence *clt* du plasmide conjugatif (Thoma & Muth, 2012). La séquence *clt* ne représente pas une origine « classique » de transfert (*oriT*) reconnue par la relaxase et sa reconnaissance par TraB n'entraîne pas son clivage. Cette séquence varie en fonction des éléments conjugatifs, mais reste toujours constituée de séquences répétées capables de former des structures secondaires (Vogelmann *et al.*, 2011) (**Figure I13**). Le transfert de l'élément conjugatif dépendant de TraB se fait sous forme double brin et circulaire (Possoz *et al.*, 2001). Suite au transfert primaire de l'élément conjugatif (Passage de l'élément conjugatif de la donatrice à la réceptrice via le pore formé par l'hexamère de TraB), une machinerie multiprotéique formée par TraB et les protéines de dispersion Spd (Spd pour *spread*) catalyse la dispersion intra-mycélienne de l'élément conjugatif (Thoma & Muth, 2015). La dispersion intra-mycélienne de l'élément conjugatif est associée à un retard de croissance des cellules réceptrices et se manifeste par la formation de zones d'inhibition circulaires appelées pocks (Thoma & Muth, 2015).

La conjugaison dépendante de TraB est également capable de mobiliser l'ADN chromosomique. En fonction de la forme linéaire ou circulaire de l'élément conjugatif, le mécanisme de la mobilisation des marqueurs chromosomiques va différer. Dans le cas d'un plasmide conjugatif circulaire, TraB va reconnaître des séquences *clt-like* présents au sein du chromosome (Thoma & Muth, 2015). 25 séquences *clt-like* ont été identifiées en analysant le génome de *S. coelicolor*. Des analyses *in vitro* ont permis de mettre en évidence la reconnaissance de ces séquences par TraB confortant ainsi leur rôle dans la conjugaison (Vogelmann *et al.*, 2011). Dans le cas des plasmides conjugatifs linéaires, la mobilisation de l'ADN chromosomique pourrait se faire suivant un modèle nommé « *end-first* » proposé par Chen (1996) dans lequel les télomères joueraient le rôle d'origine de transfert et se déroulerait à partir de l'une des extrémités.

Chez *Streptomyces* il a été démontré qu'un plasmide linéaire conjugatif est capable de mobiliser un chromosome linéaire, mais pas un chromosome circularisé artificiellement, tandis qu'un plasmide conjugatif circulaire est capable de mobiliser les deux (circulaire et linéaire) (Lee *et al.*, 2011).

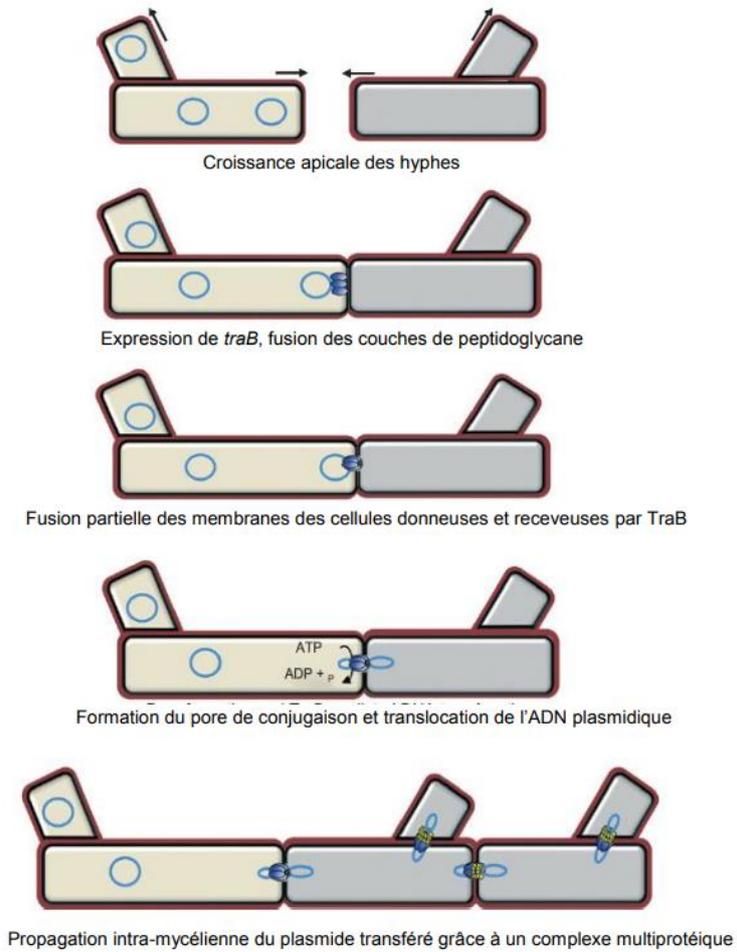


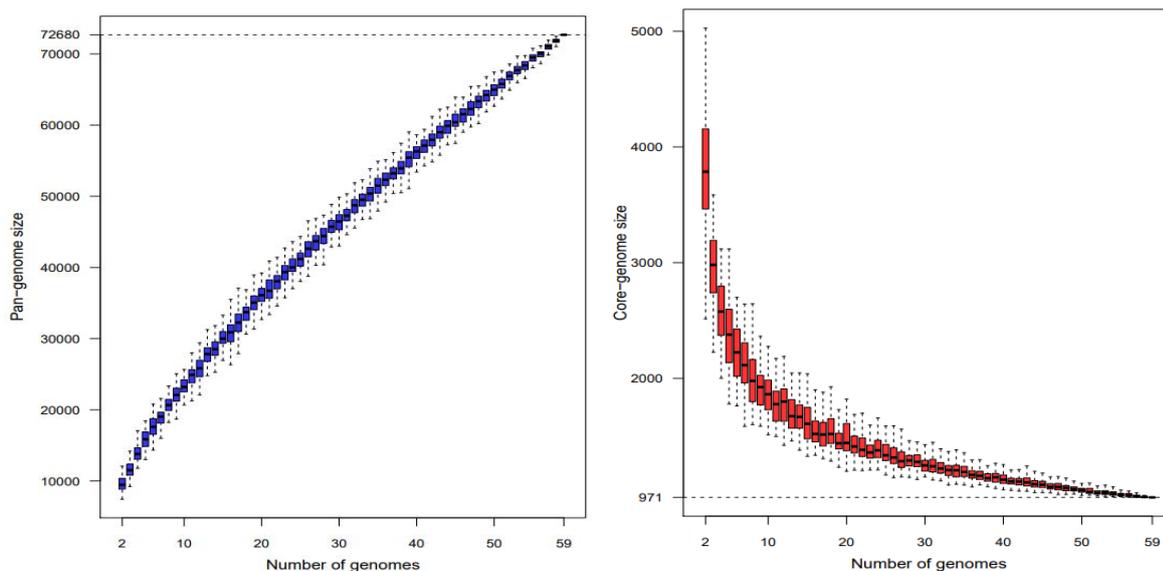
Figure I12 : Fusion partielle membranes lors de la conjugaison plasmidique chez les *Streptomyces* (adapté de Thoma & Muth, 2012).

pJV1	p1119	pSLS	pSN22	SLP1	pSG2
CCCGCACA	CCCGCACA	CCCGCACA	CCCGCACA	CGTCACAC	CCTGACAC
CCCGCACA	GCCGCACA	GCCGCACA	GCCGCACA	CGTCACAC	CCTGACAC
CCCGCACA	GCCGCACA	CCCGCACA	GCCGCACA	CCTCACAG	CCTGACAG
GCCGCACA	GCCTCACA	CCCGCACA	GCCGCACA	CGTCACAC	CCTGACAC
CCCGCACA	CCCGCACA	GCCGCACA	GCCGGACC	CGTCACAC	TCTGACAC
CCCGCACA	CTCGCACA	CCCGCACA	ACCGCACA	CGTCACAC	CCTGACAC
GCCGCACA	GCCGCACA	CCCGGACG	CCCGCACA		CCTGACAC
CCCGCACA	CCCGCACA				
CCCGCACA					
CCCGCACA*	SCCGCACA*	CCCGCACA*	NCCGCACA*	CGTCACAC*	CCTGACAC*
pIJ101	p1424	pSVH1	pSG5	pFP11	pMEA300
TGGAACC	TCAACTCG	GACCCGGC	GCCCGGAC	GACCCGGA	CTCCCTCC
CGCAACCT	GCAACTTC	GACCCGGA	CCCGGAC	ACCCTGGA	CTCCCTCA
TGCAACTT	TCAACCGG	CGCCCGGA	GCCCGCCC	CCCCCGGA	CTCCATTC
GTCAACCG	GCAACCGG	GACCCGGA	GCCCGGTC	GACCTGGA	TTCCCTAC
TGCAACGG	GCAACCGG	GACCCGGC	TCCCGGAC	GACCTGGA	TTCCCTCG
CTCAACCG	GCAACCGG	GACCCGGA	CCCGCAG	AACCTGA	CTCCCTGC
GGCAACCC		GACCCGGA	GCCCTGAC	GACCCCGA	TTCCCTAC
GGAAACCA		GGCCCGGA			
NGCAACCN*	GCAACCGG*	GACCCGGA*	SCCGGAC*	GACCYGGA*	YTCCCTNN*

Figure I13 : Répétitions imparfaites de 8 paires de bases des séquences clt (cis acting locus of transfert) des plasmides conjugatifs de *Streptomyces*. Les régions clt confirmées (pJV1, p1119, pIJ101, pSVH1) et les loci putatifs de clt en aval du gène *traB* contiennent plusieurs répétitions imparfaites de 8 pb qui pourraient représenter les séquences reconnues par TraB (TRS : TraB Recognition Sequences) (adapté de Vogelmann *et al.*, 2011).

IV.4 Une grande variabilité inter- et intra-spécifique

La comparaison des génomes des espèces appartenant au genre *Streptomyces* a permis de révéler de fortes variabilités pouvant exister entre deux espèces. La taille des chromosomes varie du simple au double allant de 5,96 Mb pour *Streptomyces xiamenensis* à 11,94 Mb pour *Streptomyces bingchenggensis*. Une étude de génomique comparative sur 59 génomes a révélé un pangénome important (72.680 gènes) au niveau de ce genre (Lorenzi *et al.*, 2019 en préparation) (**figure I14**). La taille du pangénome de *Streptomyces* s'accroît avec le nombre de génomes séquencés (**Figure I14**), ce qui permet d'affirmer que ce genre bactérien possède un pangénome ouvert où le génome accessoire est plus important que le génome core. Le génome core du genre *Streptomyces* est en effet constitué de 971 gènes, ce qui représente 9,7 à 19 % de la taille des génomes analysés. La variabilité génétique se distingue également à l'échelle intra-spécifique chez cette bactérie. La comparaison de deux génomes de *Streptomyces ambofaciens* (ATCC23877 et DSM 40697) a montré qu'il existait une différence globale de 200 kb au niveau des génomes et que certains éléments génétiques mobiles présents chez ATCC23877 ne le sont pas chez DSM40697 et réciproquement (Thibessard & Leblond, 2016).



Figures I14 : évolution du pan génome et du core génome du genre *Streptomyces* en fonction du nombre de génomes séquencés. Le nombre total de gènes non orthologues (à gauche) et le nombre de gènes communs (à droite) pour un nombre donné de génomes analysés pour les différentes souches de *Streptomyces*. Les bords supérieur et inférieur des boîtes indiquent le premier (25e centile des données) et le troisième quartile (75e centile), respectivement, de 59 génomes pris de façon aléatoire. La ligne horizontale centrale indique la médiane de l'échantillon (50e centile). Sur 59 génomes séquencés, le core génome renferme 971 gènes, et le pan-génome avait 72.680 gènes au total. (Extrait de Lorenzi *et al.*, 2019 ; en préparation).

IV.5 Une distribution compartimentée des gènes le long du chromosome

La comparaison des génomes de *S. coelicolor* et de *S. avermitilis* a permis de révéler une disposition génétique particulière : l'ensemble des gènes supposés essentiels sont localisés dans la partie centrale du chromosome et forment la région « core » tandis que les gènes variables dits de contingence se situent majoritairement dans les bras qui se trouvent de part et d'autre de la région « core » (Bentley *et al.*, 2002; Ikeda *et al.*, 2003). Une analyse de synténie (conservation de l'ordre des gènes) entre les génomes des souches de *S. ambofaciens*, *S. coelicolor* et de *S. avermitilis* a permis de déterminer qu'il existait une diminution progressive de la synténie au fur et à mesure que l'on s'éloigne de la partie centrale du chromosome ; cette diminution de la synténie serait provoquée par les nombreux événements d'insertions et délétions (indels), témoins des transferts horizontaux ainsi que des réarrangements génomiques (Choulet *et al.*, 2006a).

IV.6 Une instabilité génétique liée aux réarrangements chromosomiques

L'instabilité génétique est un phénomène que l'on retrouve dans l'ensemble du monde vivant. Cependant cette instabilité est particulièrement forte chez les bactéries du genre *Streptomyces*. Chez la souche *S. ambofaciens*, l'instabilité génétique apparaît à une fréquence de 1% lorsqu'elle est cultivée sur gélose solide en conditions de laboratoire. Cette instabilité génétique affecte le processus de différenciation cellulaire (dépigmentation des spores), l'expression du métabolisme secondaire ou la réponse aux changements environnementaux. Une augmentation drastique de la fréquence de

l'instabilité génétique chez la descendance des mutants issues des souches sauvages a également été rapportée. Cette instabilité passe à 25% chez la descendance des mutants issus de la souche sauvage de *S. lividans* (Altenbuchner & Cullum, 1985) et à 87% chez celles des mutants issus de la souche sauvage de *S. ambofaciens* (Leblond *et al.*, 1989). Seuls 13% des mutants instables donnent une descendance ayant un phénotype homogène. Cette instabilité également décrite chez d'autres espèces du genre *Streptomyces* (Birch *et al.*, 1989), est directement liée à l'observation des événements de réarrangements génomiques (Leblond *et al.*, 1990; Birch *et al.*, 1991). Ces réarrangements sont soit des délétions, des translocations, des circularisations ou des amplifications et surviennent généralement au niveau des régions subterminales.

L'analyse des événements de réarrangements survenant au niveau subterminal a permis de démontrer que les réarrangements entraînant la circularisation du chromosome intervenaient de façon spontanée chez plusieurs espèces de *Streptomyces* (Fischer *et al.*, 1997; Kameoka *et al.*, 1999; Inoue *et al.*, 2003). Les mutants ayant subi ces réarrangements se caractérisaient par de grandes délétions au niveau des deux bras englobant les TIR pouvant atteindre 2,1 Mb ; ce qui représente environ un quart du génome de ces bactéries. La circularisation observée résulte d'une recombinaison illégitime entre les deux extrémités du chromosome délété. De même des réarrangements observés chez deux mutants spontanés de *S. ambofaciens* a permis de mettre en évidence une taille de TIR plus élevée (480 kb et 850kb) (Fischer *et al.*, 1998) comparée à la souche sauvage (210 kb) (Leblond *et al.*, 1996). Cette diversité dans la taille des TIR résulte de la délétion d'un des bras entraînant son remplacement par le second bras. Cet événement a pu se produire grâce à la recombinaison homologue entre deux gènes codant pour le facteur sigma (HasR et HasL) situé chacun sur un de bras du chromosome et partageant 99 % d'identité au niveau nucléotidique (Fischer *et al.*, 1998; **Figure I15**). Des événements similaires ont également été décrits chez *Streptomyces griseus* impliquant une recombinaison entre deux gènes codant pour des lipoprotéines (Uchida *et al.*, 2003) ainsi que chez *Streptomyces coelicolor* A3(2) entre deux transposons (Widenbrant *et al.*, 2007)). Wenner *et al.*, (2003) ont également rapporté un autre type de réarrangement chromosomique dans lequel la perte d'un bras du chromosome est réparée par une fusion avec une autre copie du chromosome ayant perdu le même bras. Cette structure dupliquée entre dans un cycle de transmission de l'instabilité génétique à sa descendance. L'héritage de cette instabilité confère à la descendance un état mutateur conduisant à la génération de la diversité génétique et phénotypique au sein de la population. La perte de la duplication par contre constitue une sortie de cet état d'instabilité avec la production de chromosome possédant une grande variabilité dans la longueur des TIR (Wenner *et al.*, 2003).

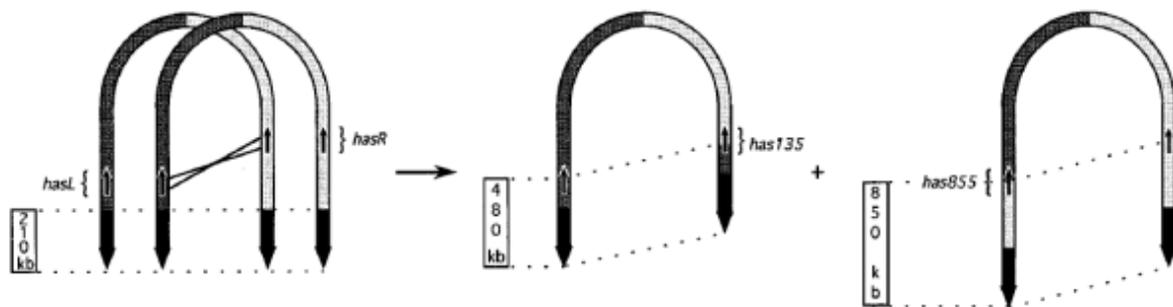


Figure I15 : Représentation schématique de la recombinaison ectopique entre les *loci hasL* et *hasR* de deux copies sauvages du chromosome de *S. ambofaciens*, engendrant une première version du chromosome avec des TIR de 480 kb et une seconde version avec des TIR de 850 kb (Fischer *et al.*, 1998).

L'amplification d'ADN chromosomique a pu être observée chez plusieurs espèces de *Streptomyces*. Il s'agit de mécanismes qui engendrent des répétitions en tandem d'un locus donné en plusieurs dizaines de copies. Ce phénomène largement répandu dans le monde bactérien n'est pas que l'apanage des *Streptomyces*. Cependant dans les génomes des souches de *Streptomyces* la fréquence d'amplification est beaucoup plus importante, le nombre de copies générées y est également beaucoup plus élevé (pouvant atteindre des centaines) et l'amplification s'accompagne souvent de la formation d'une grande délétion (Young & Cullum, 1987). L'amplification n'implique dans ces souches que certains *loci* particuliers appelés unités amplifiables d'ADN (AUD) (Demuyter *et al.*, 1988). Il s'agit généralement de structures contenant des répétitions ; Les gènes de type PolyKetide Synthase (PKS) sont particulièrement favorables à ce type de recombinaison de par leur nature redondante avec des domaines codant des activités enzymatiques fortement homologues dans les gènes de biosynthèse. Pour exemple, on peut citer le cluster de synthèse de la kanamycine chez *S. kanamycetius* (Yanai *et al.*, 2006), celui de l'actinorhodine chez *S. coelicolor* (Widenbrant *et al.*, 2008) et celui des caroténoïdes chez *S. albus* (Myronovskiy *et al.*, 2014).

Dans une récente étude, Hoff *et al.*, (2018) ont démontré l'implication de la cassure double brin dans la plasticité du génome chez *Streptomyces*. En effet, alors que la réparation de la cassure double brin dans la partie centrale du chromosome est mutagène et conserve la structure globale du chromosome, elle induit par contre la perte du bras correspondant ainsi que des délétions extensives au-delà du point de cassure lorsque cette dernière intervient au niveau de l'un des bras. Alors que cette étude pointe le rôle central joué par la recombinaison homologue dans la réparation des cassures double brin, elle démontre également l'implication de la recombinaison non homologue (*NHEJ* : *Non Homologous End Joining*) dans les réparations mutagènes ainsi que dans de grands réarrangements chromosomiques comme la circularisation du génome. Cette réparation des cassures double brin par la recombinaison non homologue étant concomitante à l'intégration de nouveaux matériels géniques

au point de réparation de la cassure. La réparation de la cassure double brin au sein du chromosome favoriserait alors la plasticité du génome et serait impliqué dans l'évolution chez *Streptomyces* ; la recombinaison non homologue jouerait aussi un rôle important dans le transfert horizontal (Hoff *et al.*, 2018).

IV.7 Une diversité de production de métabolites spécialisés au niveau intra-spécifique

Les bactéries du genre *Streptomyces* sont très connues pour leur production de métabolites spécialisés. Ces métabolites jouent un rôle important dans l'écologie de ces bactéries particulièrement dans l'acquisition d'éléments essentiels, comme le fer grâce aux sidérophores dans leurs milieux, dans la communication chimique et dans leur défense (Wietz *et al.*, 2013).

Le séquençage des génomes des bactéries du genre *Streptomyces* a permis d'affirmer que ces bactéries sont des réservoirs de gènes impliqués dans la synthèse de métabolites spécialisés et de déceler qu'elles exprimaient juste une partie de leur potentiel de production de métabolites en condition de culture au laboratoire. Dix-huit clusters supplémentaires de production de métabolites secondaires ont été identifiés chez *S. coelicolor* alors qu'il était connu pour produire seulement quatre métabolites secondaires (Bentley *et al.*, 2002). Ces clusters de biosynthèse de métabolites secondaires, qui sont présents au sein du génome, mais non exprimés sont dits cryptiques.

Plusieurs études se sont intéressées à la diversité métabolique entre les espèces bactériennes. Doroghazi & Metcalf (2013) ont fourni la première analyse comparative du métabolisme secondaire au sein des organismes appartenant au phylum des Actinobactéries incluant sept souches de *Streptomyces* (Doroghazi & Metcalf, 2013). Ziemert *et al.* ont également réalisé une analyse du métabolisme secondaire de 75 souches de *Salinispora* et y ont déterminé 124 voies métaboliques codées par les génomes de ces souches ; ils ont également démontré l'existence de l'échange de ces voies métaboliques au sein de la population dans l'environnement marin (Ziemert *et al.*, 2014).

Une récente étude portant sur l'analyse des génomes de 6 souches de *S. albus* provenant de divers environnements a permis une identification des clusters de métabolites secondaires qui sont souches spécifiques (Seipke, 2015). La comparaison des six génomes de *S. albus* ainsi étudiés a révélé qu'il existait au sein d'une même espèce une diversité du point de vue métabolique. Des études similaires entreprises sur des souches de *Streptomyces*, cette fois isolées de lichens côtiers, et possédant une séquence identique du gène codant l'ARNr 16S ont permis d'identifier des profils métaboliques variés d'une souche à l'autre (Antony-Babu *et al.*, 2017). Ces différentes études permettent de démontrer que les clusters de métabolites secondaires peuvent varier d'un individu à un autre au niveau inter ou intra-spécifique. Les clusters de synthèse de métabolites secondaires figurent parmi les éléments génétiques les plus diversifiés et qui évoluent le plus rapidement (Fischbach *et al.*, 2008). L'étude des

génomomes des souches de *Streptomyces* génétiquement cohésives et sympatriques permettrait de mieux comprendre l'évolution des clusters de gènes de métabolisme secondaire et expliquer ainsi la diversité qui en découlent au niveau inter et intra spécifique.

Objectifs de la thèse :

Bien que la plasticité du génome de *Streptomyces* soit avérée par des études de génomique comparative entre différentes espèces de cette bactérie, l'étendue de la diversité génomique et son impact sur l'écologie des souches appartenant à une même population définie comme un ensemble d'individus dérivant d'un ancêtre commun récent et partageant le même micro-habitât, reste encore à élucider. Les objectifs du projet de thèse étaient :

- D'analyser la diversité génomique et la vitesse d'évolution des génomes au sein d'une population (formée de souches ayant des liens phylogéniques très étroits parce que divergeant d'un ancêtre commun récent) sympatrique (partageant le même habitat) de *Streptomyces* dans le but d'estimer la fréquence des événements de recombinaison au sein de l'écosystème sol forestier.
- D'analyser l'organisation et la répartition de la diversité observée le long du chromosome linéaire de ces bactéries ainsi que son impact sur l'évolution des répétitions terminales inversées (TIR).
- De tester le caractère fonctionnel de la diversité génétique dans le sol à travers la production des métabolites spécialisés.
- D'étudier les mécanismes à l'origine de cette diversité génomique au sein de cette population.

Pour la suite de ce manuscrit, les différents résultats sont détaillés en cinq chapitres distincts. Dans Les trois premiers chapitres, les résultats sont présentés sous forme d'articles scientifiques. Un résumé en français des résultats exposés, ainsi que certains résultats complémentaires (pour la partie 2) seront également fournis. Pour les chapitres 4 et 5 les résultats seront exposés en français (articles en cours de rédaction).

Les cinq parties s'intituleront comme suit :

- I. Stratégie d'échantillonnage et séquençage des génomes de 11 souches sympatriques et conspécifiques de *Streptomyces*.
- II. Etude de la diversité génétique au sein d'une population de 11 souches sympatriques et conspécifiques de *Streptomyces* par l'approche de la génomique comparative.
- III. Etude de la variabilité chromosomique terminale des génomes de 11 souches sympatriques et conspécifiques de *Streptomyces*
- IV. Etude de la diversité des éléments conjugatifs et intégratifs au sein d'une population de 11 souches sympatriques et conspécifiques de *Streptomyces*

V. Etude des mécanismes régissant l'évolution rapide des clusters de gènes de synthèse de métabolites spécialisés (CGSMS).

Par la suite, ces résultats seront discutés dans une discussion générale et les perspectives à apporter seront détaillées.

RESULTATS

CHAPITRE I

I. Stratégie d'échantillonnage et séquençage des génomes de 11 souches sympatriques et conspécifiques de *Streptomyces*.

Genome Sequences of 11 Conspecific *Streptomyces* sp. Strains

Tidjani A.-R., Lorenzi J.-N., Toussaint M., van Dijk E., Naquin D., Lespinet O., Bontemps C., Leblond P.

Article N°2 : Publié dans le journal Microbiology Resource Announcements le 19/09/2019

Résumé des résultats de l'article :

Dans ce chapitre, les approches, stratégies et techniques utilisées pour le séquençage de 11 souches de *Streptomyces* d'une même population seront détaillées. Notre approche trouve sa justification dans le fait que les approches de génomique comparative entre bactéries permettent de visualiser les différences génétiques qui résultent de leur évolution. Or, ces dernières sont généralement entreprises en étudiant des souches appartenant à la même espèce ou au même genre bactérien mais avec généralement une histoire évolutive relativement longue. Pour comprendre les événements évolutifs récents à l'œuvre dans la diversification de souches, il apparaît donc important de comparer le génome de souches qui i) interagissent ensemble dans leur niche et donc peuvent réaliser physiquement des transferts de gènes et ii) qui soient suffisamment proches taxonomiquement pour que leur divergence nucléotidique ne soit pas un frein à la recombinaison.

Dans cette étude, pour accéder à des populations bactériennes, une collection de souches issues de micro-habitats (grain de solde l'ordre du mm³) a été réalisée à partir d'une motte de terre entourant les racines d'un jeune plan de hêtre provenant de la forêt de Montier-Sur-Saulx en France. Cette collection a été réalisée et caractérisée taxonomiquement au cours de la thèse de Maxime Toussaint. Neuf micro-agrégats de sol y ont été prélevés soit en contact de la racine du jeune plan de hêtre, soit au niveau du sol nu. Ces échantillons étaient distants de 8 cm au maximum. La nomenclature et la localisation spatiale des échantillons prélevés par rapport au jeune plan de hêtre sont présentées dans la **figure R1**. Une première collection de 129 souches a été obtenue et réduite à 46 souches sur la base de leur capacité à sporuler (pour des raisons de stockage et de conservation des souches).

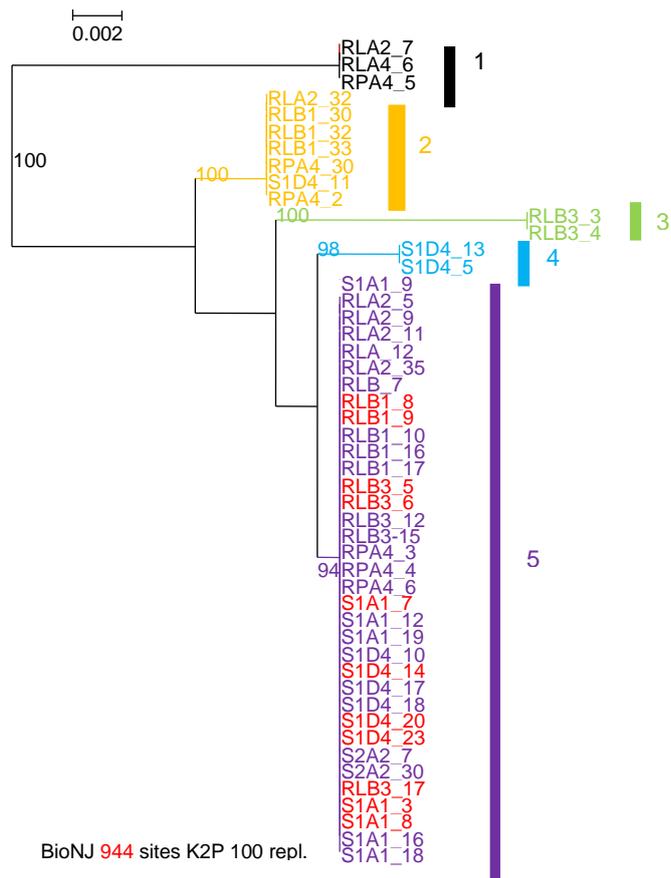


Figure R2 : Arbre phylogéniques basé sur la séquence partielle du gène codant l'ARNr 16S des souches de la collection de Montier-Sur-Saulx. L'arbre 16S est basé sur les alignements des séquences partielles (944 nt) du gène codant l'ARNr 16S. l'arbre a été réalisé avec le logiciel Seaview selon la méthode des distances avec l'algorithme BioNJ et le modèle K2P.

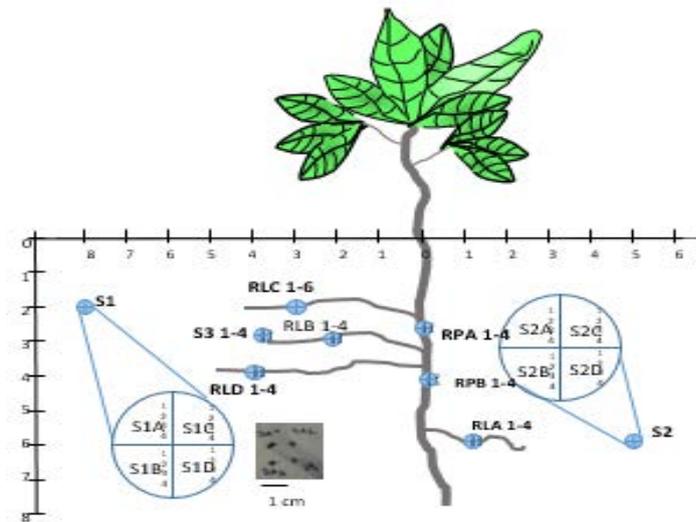


Figure R1 : Stratégie de réalisation de la collection de souches de la Forêt de Montier-Sur-Saulx. Neuf micro agrégats ont été prélevés au niveau de la racine d'un hêtre comme indiqué sur la figure. Chaque micro agrégat a ensuite été divisé en 4 fragments. Les fragments provenant de sol nu sont à leur tour subdivisés en 4 micro fragments. Ces micro fragments ont ensuite été dilués et étalés sur un milieu favorisant la croissance des *Streptomyces*. Sur 126 isolats mis en collection, 49 ont été sélectionnés sur leur capacité à sporuler au laboratoire et utilisés au cours de cette étude.

Un arbre phylogénique basé sur une séquence partielle de l'ARNr 16S de ces souches a révélé 5 clusters différents chacun constitué de souches ayant 100% d'identité (**Figure R2**). Le groupe 5 renferme 35 des 46 souches de la collection. Une étude plus précise de ces 46 souches a été effectuée par une analyse de *Multi Locus Sequence analysis* (MLSA) en amplifiant 5 gènes de ménage (*atpD*, *gyrB*, *recA*, *rpoB* et *trpB*) (Guo *et al.* 2008). Vingt-trois séquences types (ST) au total ont été définies sur les 46 souches analysées. Sur les 32 souches du groupe 5, onze ST ont été définies. La distance moyenne entre les séquences concaténées était de 0.001, ce qui est inférieur au seuil de 0.007 établi pour la délimitation d'espèces pour les *Streptomyces* (Rong *et al.*, 2010), suggérant que le groupe 5 est constitué de souches fortement apparentées appartenant à la même espèce. Onze souches de ce groupe ont été choisies de manière à couvrir toute la diversité observée au sein de ce groupe et séquencées.

Les bactéries du genre *Streptomyces* possèdent des génomes de grande taille et riches en GC (environ 70%). Ces caractéristiques rendent difficile le séquençage des génomes de cette bactérie. Dans ce chapitre nous décrivons une approche combinant deux technologies de séquençage : Oxford Nanopore et Illumina. Un assemblage hybride réalisé à partir des données de séquençage issues des deux technologies a permis d'obtenir tous les chromosomes en un seul contig et de détecter *in silico* la présence des réplicons extrachromosomiques. La présence des réplicons extrachromosomiques linéaires fut confirmée par des analyses de PFGE (*Pulsed-field gel electrophoresis*) (**Figure S2, Annexes**). La taille totale des génomes varie de 11,76 à 12,45 MB.



Genome Sequences of 11 Conspecific *Streptomyces* sp. Strains

Abdoul-Razak Tidjani,^a Jean-Noël Lorenzi,^{a,b} Maxime Toussaint,^a Erwin van Dijk,^b Delphine Naquin,^b Olivier Lespinet,^b Cyril Bontemps,^a Pierre Leblond^a

^aUniversité de Lorraine, INRA, DynAMic, Nancy, France

^bInstitute for Integrative Biology of the Cell (I2BC), CEA, CNRS, University Paris-Sud, University Paris-Saclay, Gif-sur-Yvette, France

ABSTRACT The genomes of 11 conspecific *Streptomyces* strains, i.e., from the same species and inhabiting the same ecological niche, were sequenced and assembled. This data set offers an ideal framework to assess the genome evolution of *Streptomyces* species in their ecological context.

Streptomyces species are soil-dwelling bacteria that harbor large linear chromosomes (1). We report here the genome sequences of 11 sympatric *Streptomyces* strains belonging to the same species. In order to select conspecifics, we sampled soil grains at the cubic milliliter scale from a French forest (maximal distance of 8 cm from each other). After dissolution in sterile water and spreading of serial dilutions on *Streptomyces* isolation medium (SIM) (2), the 16S rRNA sequences of the strains were determined and analyzed using BLAST (3), and their phylogenetic relationships were characterized by multilocus sequence analysis (MLSA) using the Molecular Evolutionary Genetics Analysis version 7 (MEGA7) software (4). After growth in liquid Hickey-Tresner medium at 30°C for 30 h, DNA purification was performed using the salting-out method (5), followed by chloroform extraction. The targeted genes (rRNA gene and the MLSA genes) were amplified using universal (16S rRNA gene [6]) and specific (MLSA [7]) primers. The 11 strains selected showed identical 16S rRNA gene sequences and minimal MLSA divergence. These strains are related to *Streptomyces olivochromogenes* (strain DSM40451), with an average identity of 99.93% for the 16S rRNA gene sequences (8). A hybrid assembly using Oxford Nanopore technology for scaffolding and Illumina technology for sequence improvement was performed (Table 1). Base calling of these sequences was performed using the Oxford Nanopore base callers Albacore (v0.8.4 or v2.0.2) or Guppy (v0.3.0). Nanopore reads (minimum quality mean, 7) were generated on minION or gridION systems. When strains were multiplexed, Porechop (v0.2.4, using default settings) was used for demultiplexing (and adaptor trimming). The coverage ranged from 41× to 344×. The Illumina paired-end libraries were created using the Illumina Nextera kit, except for RLB1-8 and RLB1-9, for which sonication (Covaris) and adaptor ligation (Illumina TruSeq) were used instead. Paired reads were generated using a MiSeq reagent kit v3 (150 cycles) and the Genome Analyzer system (Illumina). The minimum read size was set to 10 bp, and adaptor trimming was performed using Cutadapt (v1.15, using default settings). The coverages of the paired-end reads (length, 75 to 300 bp) ranged from 58× to 320×. The hybrid assembly was performed using Unicycler (9) v0.4.2 or v0.4.3 (using default settings) to assemble 1 to 19 large contigs covering the whole genome of each strain, enabling the acquisition of each linear chromosome in one scaffold and the identification of extrachromosomal elements when present. One or two extrachromosomal linear or circular replicons were identified in 5 of the 11 strains by *in silico* prediction or pulsed-field gel electrophoresis experiments (10). The total genome sizes ranged from 11.76 to 12.45 Mb, positioning these strains among the largest bacterial genomes (Table 1).

Citation Tidjani A-R, Lorenzi J-N, Toussaint M, van Dijk E, Naquin D, Lespinet O, Bontemps C, Leblond P. 2019. Genome sequences of 11 conspecific *Streptomyces* sp. strains. *Microbiol Resour Announc* 8:e00863-19. <https://doi.org/10.1128/MRA.00863-19>.

Editor Christina A. Cuomo, Broad Institute

Copyright © 2019 Tidjani 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 Pierre Leblond, pierre.leblond@univ-lorraine.fr.

Received 24 July 2019

Accepted 28 August 2019

Published 19 September 2019

TABLE 1 Genome features, sequencing statistics, and accession numbers of the 11 conspecific *Streptomyces* strains

Strain or replicon	Illumina sequencing information			Oxford Nanopore sequencing information							GenBank accession no.			
	No. of reads (approx coverage [x])	SRA accession no.	No. of reads (approx coverage [x])	<i>N</i> ₅₀ of raw reads (kb)	Flow cell type(s)	Sequencing kit(s)	Base caller	SRA accession no.	Replicon size (bp) ^c	Genome size (bp)		Total no. of CDS ^a	G+C content (%)	TIR ^b (kb)
RLB1-8	15,381,622 (320)	SRR9661592	655,482 (150)	4.3	FAH118893 (9.5), FAH18988 (9.5), FAH24488 (9.5)	sqk-lsk308	albacore_2.0.2	SRR9710048	11,765,340	11,765,340	10,635	70.2	357	CP041650
RLB1-9	18,329,970 (115)	SRR9661591	88,694 (41)	7.6	FAF19789 (9.4)	sqk-lsk308	albacore_0.8.4	SRR9710047	11,940,408	12,200,709	10,838	70.2	311	CP041654
pRLB1-9.1									154,158 ^c		175	69.0		CP041653
pRLB1-9.2									106,143 ^c		111	68.7	24	CP041652
RLB3-5	3,196,108 (67)	SRR9661590	144,521 (56)	7.1	FAH24352 (9.5), FAH29240 (9.4)	sqk-lsk308, sqk-lsk108	albacore_2.0.2	SRR9710050	11,898,970	11,898,970	10,731	70.2	365	CP041651
RLB3-6	3,274,272 (68)	SRR9661589	299,155 (52)	4.6	FAF19789 (9.4)	sqk-lsk308	guppy_0.3.0	SRR9710049	12,338,263	12,448,281	11,255	70.1	587	CP041602
pRLB3-6.1									110,314 ^c		101	70.6		CP041601
RLB3-17	3,976,622 (83)	SRR9661596	202,455 (50)	5.6	FAF19789 (9.4)	sqk-lsk308	guppy_0.3.0	SRR9710052	12,023,175	12,023,175	10,934	70.2	451	CP041610
S1A1-3	3,243,184 (68)	SRR9661595	198,567 (51)	5.2	FAF19789 (9.4)	sqk-lsk308	guppy_0.3.0	SRR9710051	12,042,091	12,042,091	10,920	70.2	393	CP041611
S1A1-7	3,504,210 (73)	SRR9661594	533,299 (73)	2.8	FAF19789 (9.4)	sqk-lsk308	guppy_0.3.0	SRR9710054	11,713,151	12,005,504	10,580	70.3	513	CP041604
pS1A1-7.1									292,353 ^c		252	69.7		CP041603
S1A1-8	3,191,318 (66)	SRR9661593	780,962 (53)	1.1	FAF19789 (9.4)	sqk-lsk308	guppy_0.3.0	SRR9710053	12,036,971	12,036,971	10,918	70.2	394	CP041612
S1D4-14	2,794,454 (58)	SRR9661598	2,066,754 (344)	3.2	FAF19789 (9.4)	sqk-lsk308	guppy_0.3.0	SRR9710056	11,723,487	11,934,498	10,591	70.2	369	CP041607
pS1D4-14.1									112,196 ^c		118	68.7	0	CP041605
S1D4-20	3,327,172 (69)	SRR9661597	1,424,402 (255)	4.1	FAF19789 (9.4)	sqk-lsk308	guppy_0.3.0	SRR9710055	11,851,257	12,245,276	10,742	70.2	373	CP041609
pS1D4-20.1									98,815 ^c		138	69.1		CP041606
S1D4-23	3,543,760 (74)	SRR9661599	400,348 (61)	3.5	FAF19789 (9.4)	sqk-lsk308	guppy_0.3.0	SRR9710057	394,019 ^c	12,057,712	329	69.1	68	CP041608
									12,057,712	12,057,712	10,971	70.2	421	CP041613

^a As determined through automatic annotation by the NCBI Prokaryotic Genome Annotation Pipeline. CDS, coding sequences.^b TIR, terminal inverted repeat.^c Linear (L) or circular (C) replicon configuration, as predicted by the assembler and tested by pulsed-field gel electrophoresis (not shown).

Data availability. Genome sequences and raw sequence reads are available from GenBank and the NCBI Sequence Read Archive under the accession numbers shown in Table 1.

ACKNOWLEDGMENTS

This work was funded by the French National Research Agency (grants ANR LABEX ARBRE and ANR-11-LABX-0002-01), by the French National Institute for Agricultural Research (INRA), and by Région Lorraine (now called Région Grand Est).

REFERENCES

1. Thibessard A, Leblond P. 2014. Subtelomere plasticity in the bacterium *Streptomyces*, p 243–258. In Louis EJ, Becker MM (ed), *Subtelomeres*. Springer, Berlin, Germany.
2. D'Costa VM, McGrann KM, Hughes DW, Wright GD. 2006. Sampling the antibiotic resistome. *Science* 311:374–377. <https://doi.org/10.1126/science.1120800>.
3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic Local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
4. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874. <https://doi.org/10.1093/molbev/msw054>.
5. Kieser T, Bibb M, Buttner M, Chater K, Hopwood D. 2000. *Practical Streptomyces genetics*. John Innes Foundation, Norwich, United Kingdom.
6. Rintala H, Nevalainen A, Rönkä E, Suutari M. 2001. PCR primers targeting the 16S rRNA gene for the specific detection of streptomycetes. *Mol Cell Probes* 15:337–347. <https://doi.org/10.1006/mcpr.2001.0379>.
7. Guo Y, Zheng W, Rong X, Huang Y. 2008. A multilocus phylogeny of the *Streptomyces griseus* 16S rRNA gene clade: use of multilocus sequence analysis for streptomycete systematics. *Int J Syst Evol Microbiol* 58:149–159. <https://doi.org/10.1099/ijs.0.65224-0>.
8. Tidjani AR, Lorenzi JN, Toussaint M, van Dijk E, Naquin D, Lespinet O, Bontemps C, Leblond P. 2019. Massive gene flux drives genome diversity between sympatric *Streptomyces* conspecifics. *mBio* 10:e01533-19. <https://doi.org/10.1128/mBio.01533-19>.
9. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
10. Leblond P, Francou FX, Simonet JM, Decaris B. 1990. Pulsed-field gel electrophoresis analysis of the genome of *Streptomyces ambofaciens* strains. *FEMS Microbiol Lett* 60:79–88. <https://doi.org/10.1111/j.1574-6968.1990.tb03866.x>.

CHAPITRE II

- II. Etude de la diversité génétique au sein d'une population de 11 souches sympatriques et conspécifiques de *Streptomyces* par l'approche de la génomique comparative.

Massive Gene Flux Drives Genome Diversity between Sympatric *Streptomyces* Conspécifics

Tidjani A.-R., Lorenzi J.-N., Toussaint M., van Dijk E., Naquin D., Lespinet O., Bontemps C., Leblond P.

Article N°1 : Publié dans le journal mBio le 03/09/2019 Résumé des résultats de l'article :

Résumé des résultats de l'article :

Les bactéries du genre *Streptomyces* possèdent un chromosome linéaire de grande taille (6 Mb à 12 Mb) très compartimenté, avec une partie centrale conservée (génome core) et des régions terminales variables. L'acquisition des gènes par HGT participe à la dynamique des génomes et la diversité génétique observé à l'échelle du genre chez *Streptomyces*. Cependant peu d'études se sont intéressées à l'intensité et à l'impact à court terme de l'HGT à l'échelle d'une population. Le sol est un écosystème très hétérogène qui impose diverses conditions abiotiques aux communautés bactériennes qui y vivent. Ces dernières établissent des interactions biotiques entre elles et le plus souvent entre des individus issus de la division binaire vivant sur le même grain de sol. Cette proximité physique et génétique maximise la fréquence des échanges génétiques entre ces dernière brisant ainsi la clonalité issue de la reproduction binaire. Cependant l'étendu des flux de gènes qui engendre la diversité et l'impact de cette dernière sur l'écologie des populations bactériennes reste à étudier.

Dans cette étude, nous avons démontré que bien que les souches de la population partageaient un ancêtre commun récent, toutes possèdent un répertoire de gènes différent. Cette différence illustre un flux de gènes rapide et important entre ces souches. Le pan-génome de la population est constitué de 13.814 gènes dont 8.778 sont conservés au sein des 11 souches. Les gènes composant le génome accessoire de ces souches (partie variable non conservée du génome) étaient distribués dans des îlots génomiques issus des événements d'insertions ou de délétions (indels). Les indels ne sont pas distribués de façon homogène le long du chromosome linéaire de ces bactéries. Ils sont plus importants dans les régions terminales où ils forment des points chauds. Au total, 452 événements uniques ont été identifiés depuis la diversification de la population.

Certains des indels renferment des clusters de gènes de synthèse de métabolites spécialisés. L'analyse de l'un de ces clusters a permis de relier sa présence dans une partie des souches de la population à l'activité inhibitrice de ces dernières vis-à-vis d'une souche indicatrice de *Bacillus* provenant du même écosystème. Le métabolite produit par ces souches n'entrave pas la croissance des autres souches de

Streptomyces non productrices. Ce métabolite servirait alors de bien commun à l'ensemble de la population et participerait à la cohésivité des individus la composant d'un point de vue écologique.

Une partie des gènes variables portent des signatures d'éléments génétiques mobiles. La recherche des éléments conjugatifs et intégratifs a permis d'identifier au sein de la population un type d'ICE (Integrative and Conjugative Element) et 25 type d'AICE (Actinomycetes Integrative and Conjugative Element). Ces éléments sont présumés fonctionnels (capable de s'exciser et de transférer de façon autonome), car les gènes codant pour les fonctions clés nécessaires à leurs activités (intégrases, répliques et translocases) y ont été identifiés. Ces éléments conjugatifs situés dans les indels attestent qu'ils font partie du génome accessoire et témoignent de leur acquisition ou perte récente. Le répertoire unique des AICE/ICE de chaque génome de cette population suggère une forte intensité des échanges génétiques dans l'écosystème sol forestier.

Massive Gene Flux Drives Genome Diversity between Sympatric *Streptomyces* Conspecifics

Abdoul-Razak Tidjani,^a Jean-Noël Lorenzi,^{a,b} Maxime Toussaint,^a Erwin van Dijk,^b Delphine Naquin,^b Olivier Lespinet,^b Cyril Bontemps,^a Pierre Leblond^a

^aUniversité de Lorraine, INRA, DynAMic, Nancy, France

^bInstitute for Integrative Biology of the Cell (I2BC), CEA, CNRS, University Paris-Sud, University Paris-Saclay, Gif-sur-Yvette, France

ABSTRACT In this work, by comparing genomes of closely related individuals of *Streptomyces* isolated at a spatial microscale (millimeters or centimeters), we investigated the extent and impact of horizontal gene transfer in the diversification of a natural *Streptomyces* population. We show that despite these conspecific strains sharing a recent common ancestor, all harbored significantly different gene contents, implying massive and rapid gene flux. The accessory genome of the strains was distributed across insertion/deletion events (indels) ranging from one to several hundreds of genes. Indels were preferentially located in the arms of the linear chromosomes (ca. 12 Mb) and appeared to form recombination hot spots. Some of them harbored biosynthetic gene clusters (BGCs) whose products confer an inhibitory capacity and may constitute public goods that can favor the cohesiveness of the bacterial population. Moreover, a significant proportion of these variable genes were either plasmid borne or harbored signatures of actinomycete integrative and conjugative elements (AICEs). We propose that conjugation is the main driver for the indel flux and diversity in *Streptomyces* populations.

IMPORTANCE Horizontal gene transfer is a rapid and efficient way to diversify bacterial gene pools. Currently, little is known about this gene flux within natural soil populations. Using comparative genomics of *Streptomyces* strains belonging to the same species and isolated at microscale, we reveal frequent transfer of a significant fraction of the pangenome. We show that it occurs at a time scale enabling the population to diversify and to cope with its changing environment, notably, through the production of public goods.

KEYWORDS *Streptomyces*, conjugation, gene transfer, plasticity, population

Streptomyces organisms are prominent soil-dwelling bacteria found in all terrestrial ecosystems, typified by their complex differentiation cycle and their central roles in biogeochemical cycles and the homeostasis of the soil. They are known as the most prolific bacterial genus for the production of specialized metabolites (e.g., antibiotics and antifungals) and enzymes of high biotechnological and medical interest (1). In soil, these compounds are involved in interactions with the surrounding microbial communities and contribute to plant health and growth (2). Each species or strain was reported to produce one or only a few compounds, but access to the whole-genome sequences revealed a much more important biosynthetic gene reservoir (3, 4).

The genome of *Streptomyces* is remarkably large (6 to 12 Mb) and possesses one of the few linear bacterial chromosomes ever described. The genome is highly compartmentalized, with genes conserved across the species distributed at the center of the chromosome and variable genes in the terminal parts (5, 6). The biosynthetic gene clusters (BGCs) are consistently enriched in the variable part of the chromosome. The proportion of the flexible genome is impressive and is reflected at the genus level by

Citation Tidjani A-R, Lorenzi J-N, Toussaint M, van Dijk E, Naquin D, Lespinet O, Bontemps C, Leblond P. 2019. Massive gene flux drives genome diversity between sympatric *Streptomyces* conspecifics. *mBio* 10:e01533-19. <https://doi.org/10.1128/mBio.01533-19>.

Editor Julian E. Davies, University of British Columbia

Copyright © 2019 Tidjani 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 Cyril Bontemps, cyril.bontemps@univ-lorraine.fr, or Pierre Leblond, pierre.leblond@univ-lorraine.fr.

Received 13 June 2019

Accepted 25 July 2019

Published 3 September 2019

a wide-open pangenome, i.e., the entire gene set of all species of the genus (7–9). The *Streptomyces* genomes can be completed by large linear and circular plasmids. Linear chromosomes and plasmids share the same invertron structure that consists of long terminal inverted repeats (TIRs) ending with telomeric sequences (10–12). In addition, the *Streptomyces* genomes are recombinogenic, as revealed either by the high frequency of spontaneous rearrangements of large DNA fragments (13) or by multilocus sequence analysis-based studies at the intra- and interspecific levels showing that recombination rates exceed those observed within many bacterial species (14).

Gene flux, i.e., the gain and loss of genetic material, crucially depends first on the access to exogenous information that can be achieved thanks to horizontal gene transfer (HGT) and second on the recombination capacities of the recipient genome that promotes the integration of genetic material in the genome (15). Regarding HGT, *Streptomyces* exhibits little talent for natural competence, and transducing phages have remained elusive (16, 17) despite extensive research. Thus, conjugation remains the only gene exchange mechanism described in *Streptomyces*, and the genomes of these bacteria are known to be rich in conjugative elements (integrated in the chromosome or of plasmid origin). Two types of conjugative elements coexist: integrated and conjugative element (ICE) depending on a type IV secretion system, and actinomycete ICE (AICE) whose transfer depends on a DNA translocase (TraB) (18, 19). The latter is the most widespread in *Streptomyces* and is capable of mobilizing chromosomal DNA in *trans*, i.e., not physically linked to the element (20).

These gene fluxes contribute to increase the flexibility of the genome and explain the diversity observed at the genus level. McDonald and Currie (21) recently showed that the *Streptomyces* genus is ancient (380 million years old) and that acquisition and retention of genes through HGT seem rare among nonclosely related lineages. Yet, nothing is known regarding the intensity and short-term impact of HGT at the population scale. Only some recent examples have shown that *Streptomyces* sharing an identical 16S rRNA gene sequences can exhibit differential phenotypes, including antimicrobial activities (22–24), suggesting rapid genome diversification.

Genome sequencing of closely related strains locally cooccurring has already revealed the existence of a large diversity in terms of gene content within different bacterial species (25–29). However, little is known regarding bacteria living in soil. Only a few examples are documented, and similarly, they seemed to indicate that variability in gene or allelic content can impact the ecology of soil bacteria. For instance, the polymorphism in quorum-sensing genes in *Bacillus subtilis* living in the same cubic centimeter impacted the ability to communicate and led to kin differentiation (30). In the same way, genomic comparison of two closely related groups of *Myxococcus* strains sampled at a centimeter scale enabled the identification of a 150-kb region involved in their sexual isolation (31). It seems thus relevant to assess genome diversification among soil inhabiting congenics, i.e., closely related strains, to unravel the genetic bases underlying the phenotypic diversity that supports niche differentiation and adaptation. The deeper the phylogenetic relationships (infraspecific level) are, the more recent the revealed events will be. Beyond analysis of 16S rRNA gene sequences, multilocus sequence analysis (MLSA), average nucleotide identity (ANI) across the genome, or phylogenomic tree reconstruction enable us to reach close phylogenetic relationships and to infer recent molecular events.

In soil ecosystems, the heterogeneity of the matrix constitutes a powerful driver of microbial growth and dynamics (32, 33). Biotic interactions in soil are most likely to occur between genetically related cells, produced by cell division, living close together on the same soil particle (34). Thus, bacteria in soil do not live as individual cells but are more likely structured in populations composed of closely related strains deriving from a recent common ancestor. The physical and phylogenetic proximities within these populations maximize the potentiality of gene exchanges. These latter reduce genetic isolation, disrupting the confinement of bacteria resulting from their reproduction scheme. However, the true extent of the gene flux that breaks clonality, and the impact of this on the function of natural bacterial populations, remains to be explored.

For that purpose, we undertook the first comparative genomics study of *Streptomyces* conspecifics. We favored the isolation of such closely related strains by sampling *Streptomyces* at the microscale and sequenced and compared their genomes. Next, we identified a high level of divergence in terms of the presence/absence of genes between them and mapped the events along the linear chromosome. This allowed us to identify recombination hot spots, some of which include BGCs. Next, we revealed the high prevalence of actinomycete integrative and conjugative elements (AICEs). We then consider the role of AICEs as motor of genome diversification and how this can later contribute to population functioning.

RESULTS

High levels of genetic diversity within a *Streptomyces* soil population. To uncover microevolutionary processes, we isolated and sequenced the genomes of strains from a *Streptomyces* population inhabiting soil aggregates. For that purpose, four grains of soil, on the order of cubic millimeters in size and distant from each other from 2 cm to a maximum of 8 cm, were sampled from a clod of soil from which we isolated 129 strains. Based on identical 16S rDNA-coding gene sequences and a high degree of identity (>99.8%, nucleotide [nt]) measured with a five-gene multilocus sequence typing (MLST) scheme (see Fig. S1A in the supplemental material), we identified 32 highly related strains at the infraspecific level. The closest species to our group of strains was *Streptomyces olivochromogenes*, with in average 99.93% identity in the 16S rRNA gene and 94.5% ANIblast with the strain DSM40451. The genomes of 11 strains, representative of the different MLSTs identified in the cluster 6 of the population (Fig. S1A), were fully sequenced by a combination of Nanopore (Oxford Nanopore Technologies) and Illumina technologies (see Table S1). For each strain, a single scaffold for each replicon (i.e., chromosome and plasmids) was obtained with high-quality standards enabling full genome comparison. Phylogenomic tree construction (Fig. S1B) and calculation of ANIb percentages (ranging from 98.68% to 99.99%) between all isolates confirmed that they belonged to the same species and diverged recently from a common ancestor. The genomes ranged from 11.75 Mb to 12.44 Mb in size, positioning our strains among the biggest bacterial sequenced genomes (35). The chromosomes are linear with terminal inverted repeats (TIRs) of various sizes (from 311 kb to 587 kb) (see Table S2), as previously reported at the intra- and interspecific levels (12, 36). Following genome annotation (RAST), we determined that the pangenome of our population consisted of 13,814 genes (Fig. 1A). By extrapolating, we found that it reaches a maximum of 23,772 genes by 1,000 genomes. In contrast, for 11 strains randomly chosen in the genus, the pangenome was found to be approximately twice that of the population (23,672 versus 13,814) (Fig. 1B). Regarding the population pangenome, approximately one-third (5,036 genes) was not shared across the population and constituted the accessory genome (Fig. 1A). No two strains shared the same gene content; the closest differed by 12 predicted genes, while the two more distant isolates possessed 1,014 and 1,393 specific genes. Approximately one-quarter (27.6%; 1,392 genes) of the accessory genes were harbored on circular or linear extrachromosomal elements (98.8 kb to 394 kb), which were predicted *in silico* or experimentally observed by pulsed-field gel electrophoresis (PFGE) (see Fig. S2 and Table S2). These elements were found in single strains, and their mobile nature highlights the role of HGT as a driver of specificity, enabling the transfer of numerous genes simultaneously.

Genome diversity is unevenly distributed in numerous indels along the linear chromosome. Pairwise genome comparisons ($n = 65$) among the 11 sequenced genomes enabled us to identify disruptions in chromosomal synteny with insertion or deletion events (indels) of a minimum of one gene. The numbers of breaks differing between two strains ranged from 1 to 124 and increased linearly with the phylogenetic distance, indicating that they accumulate over time (see Fig. S3). When a break was shared by at least two strains, we considered it a single genetic event that may have been vertically inherited. This allowed consideration of 452 independent events that occurred during the population diversification (see Table S3). In extreme cases, the

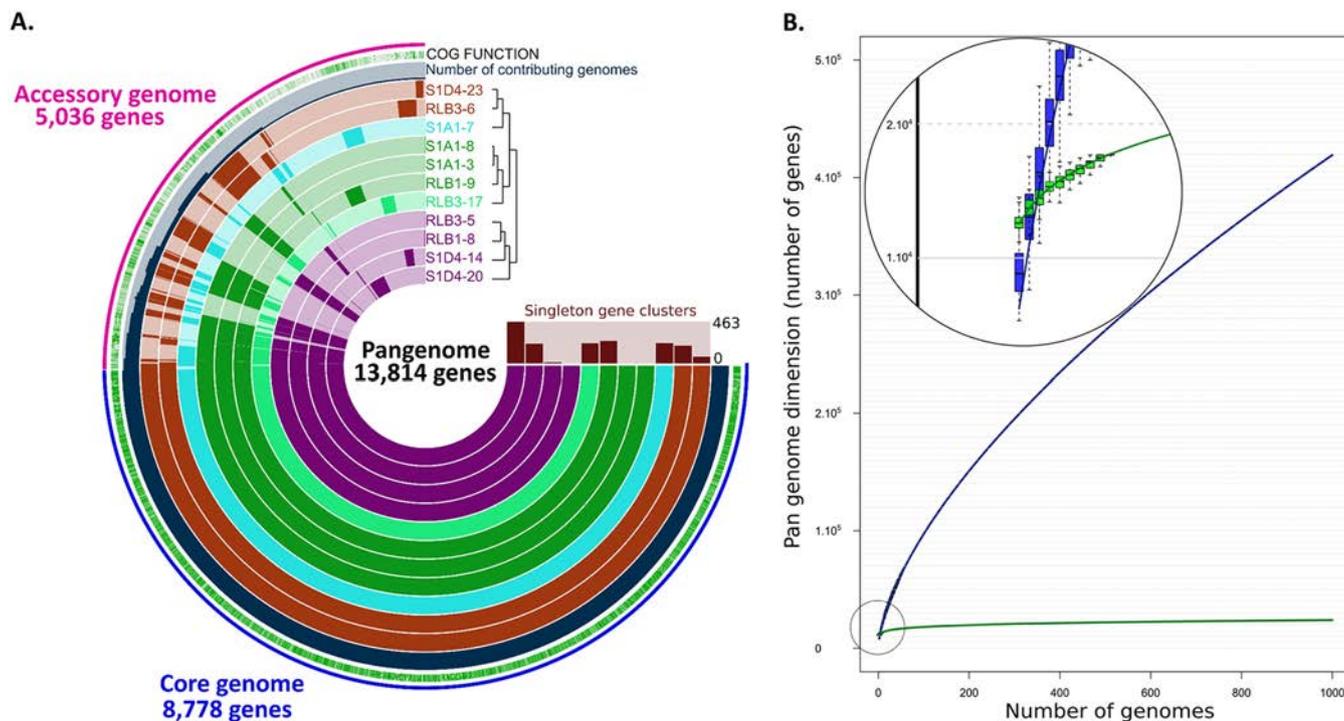


FIG 1 Pan genome analysis of the *Streptomyces* population. (A) Comparative Anvi'o genomic analysis of the 11 conspecific strains isolated in this study. The inner layers represent individual genomes organized regarding their phylogenetic relationships as indicated by the dendrogram. In the layers, dark colors indicate the presence of a gene group and light color its absence. The core (8,778 genes) and the accessory (5,036 genes) genomes are indicated in blue and pink, respectively, in the outmost layer. The blue layer represents the number of genomes among the population contributing to each gene group, and the green layer describes the gene groups in which at least one gene was functionally annotated using cluster of orthologous genes (COGs). (B) Comparison of the *Streptomyces* pan genome evolution at the genus and the population levels. The evolution of the pan genome of 59 complete *Streptomyces* genomes at the genus level is represented in blue. Its extrapolation to 1,000 genomes did not show any shift in the trend of the curve, indicating an open pan genome. The evolution of the population pan genome is represented in green. Its extrapolation rapidly reaches a plateau (see zoomed section).

closest strains differed by only two events, while the most distinct strains differed by 235.

Comparative genomics of the 11 strains shows that indels were scattered along the chromosome but tended to form hot spots (Fig. 2A) with locations where at least three distinct gene assemblies were encountered (Fig. 3 and 4). Remarkably, indel distribution was heterogeneous along the chromosome, with increased density in the terminal regions of the chromosome, and was exaggerated as the phylogenetic distance of strains increased (e.g., RLB1-8 versus RLB3-6). Mapping the core genome of the *Streptomyces* genus (971 genes [J.N. Lorenzi, A. Thibessard, O. Lespinet, P. Leblond, unpublished]) identified the two distal positions, which delineate the core region from chromosomal arms of approximately 2.4 Mb and 1.4 Mb which are devoid of conserved genes on the left and right arms, respectively. The indel density was approximately 5-fold higher in the arms than in the core region (Fig. 2B). Moreover, the compilation of all the pairwise comparisons shows that the density of indels increases from these positions toward the ends of the chromosome, forming a gradient of events. Although the size of the indel in the core region appears larger than in the arms, the total numbers of variable genes follow the same trend, i.e., a strong increase in the arms. The highest variability was noticed at the very ends of the chromosome, as no two strains were identical in terms of the size of their terminal inverted repeats (Table S2). As previously reported, this variability might result from the formation of indels in the TIRs followed by chromosome arm replacement (36–38).

Conjugative elements as the motor of rapid genome diversification. Genome mining for the presence of integrative and conjugative elements in our population allowed the identification of one ICE and 25 AICEs. These 26 elements are likely to be functional (i.e., excisable and self-transmissible), since all the key functions (integrase

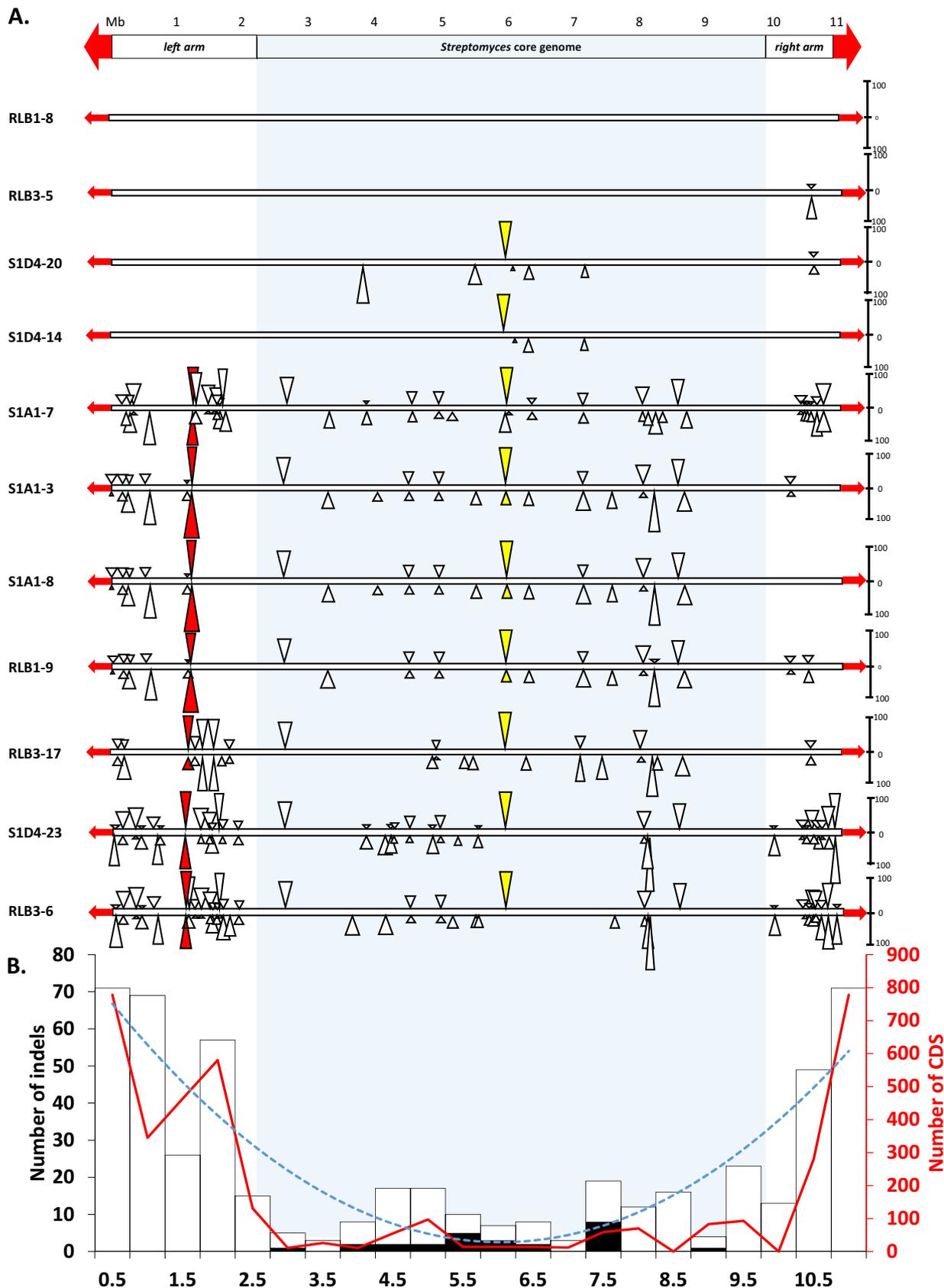


FIG 2 Distribution of insertion and deletion events along the linear chromosome of *Streptomyces*. (A) The scheme at the top represents the *Streptomyces* chromosome with a megabase scale. The position of the core genome of the *Streptomyces* genus is highlighted by a light gray (Continued on next page)

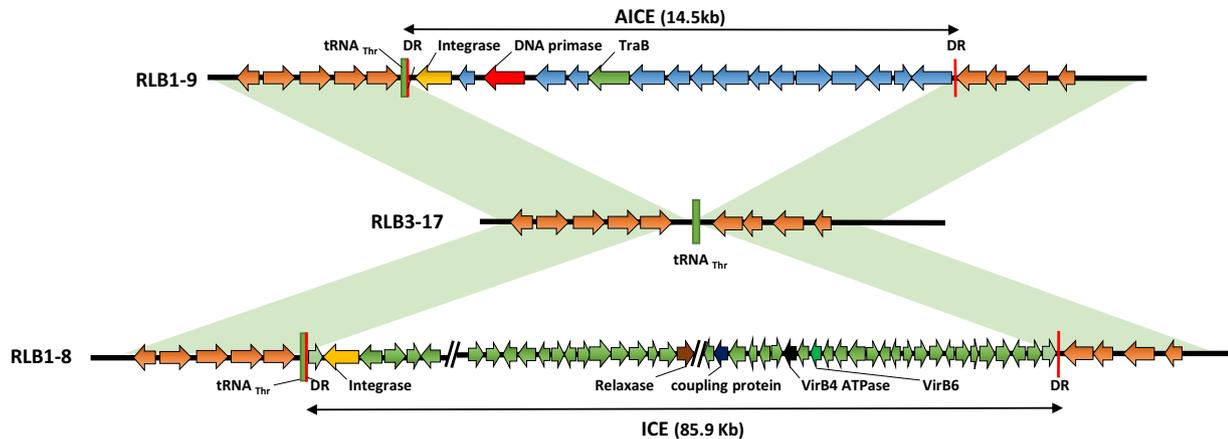


FIG 3 ICE/AICE insertion hot spot. The scheme is illustrating a hot spot for conjugative and integrative element insertion. Two different elements, AICE and ICE, are inserted in the $tRNA_{Thr}$ genes of strains RLB1-9 and RLB1-8, respectively. The $tRNA_{Thr}$ insertion site remains empty in RLB3-17. Direct repeats (DR) flanking the mobile element are represented in red (43 nucleotides for RLB1-9 and 42 nucleotides for RLB1-8). Only the key genes used to identify the elements are shown. Regions in light green correspond to syntenic regions.

and Tra translocase) and *cis* sequences (flanking *att* sites) are present in each of them. Each genome has on average five elements, which make our population particularly rich compared to two elements found on average in a previous study on *Actinobacteria* (18). All the identified elements in our population were present in indels, meaning that they were not part of the core genome, highlighting their remarkable dynamics among the population. They were mainly inserted within the central region of the chromosome (Fig. 2B) and tended to form hot spots with several distinct elements being aggregated in the same target sequence. As illustrated in Fig. 3, the same $tRNA$ target site was occupied by two distinct elements; while an AICE was present in RLB1-9, an ICE was observed in RLB1-8. In the third strain (RLB3-17), the target site was empty, which suggests that at least two of these strains experienced different conjugative events. While conjugative transfer is acknowledged as a powerful driver of diversity in telluric bacteria, their mobility in the soil remains difficult to demonstrate (39, 40). Here, the fact that each strain of the population is unique in terms of AICE/ICE content suggests that the exchanges are intense in the soil.

HGT fosters social cohesiveness of bacterial population. *Streptomyces* organisms are prolific antibiotic producers, and specialized metabolite (SM) production has already been reported to be highly variable among related strains of *Streptomyces* (22–24). We tested our *Streptomyces* isolates in pairwise inhibitory and resistance bioassays and found that none was able to inhibit its conspecifics. On the other hand, some but not all produced antimicrobial activity against a *Bacillus* strain isolated from the same soil, named “killer strains.” To correlate the inhibitory phenotype with the presence/absence of biosynthetic gene clusters (BGCs), we performed comparative genomic analyses using antiSMASH (41). We identified 51 unique BGCs within the population, each strain possessing around 35 clusters (not shown). Interestingly, some

FIG 2 Legend (Continued)

frame. The terminal inverted repeats (TIRs) are shown as red arrows. The bottom portion shows illustrations of pairwise genome comparisons (among the 10 possible pairs) within the population using strain RLB1-8 as a reference. Rectangles represent the linear chromosome of each strain. The strains are ordered from top to bottom relative to their phylogenetic distance to the reference. Triangles above the chromosome represent insertions in the reference strain, while triangles below correspond to insertions in the compared strain. For the sake of clarity, only the insertions of at least 10 predicted genes are shown. The height of a triangle reflects the number of genes involved in the insertion (the scale on the right of the chromosome indicates the number of genes). The colored triangles correspond to examples of insertion hot spots (the yellow and red triangles correspond to the hot spots depicted in Fig. 3 and 4, respectively). (B) The scheme represents the distribution of all the indel events identified within the population by genome pairwise comparisons. Each histogram bar corresponds to the number of indels within a 0.5-Mb window. The proportion of ICE/AICE insertions within a window is shown in black. The dotted blue line (“smiley” curve) corresponds to the polynomial trend curve (order 2, R^2 value) of the indel distribution. The red curve shows the number of genes involved in indels within a window.

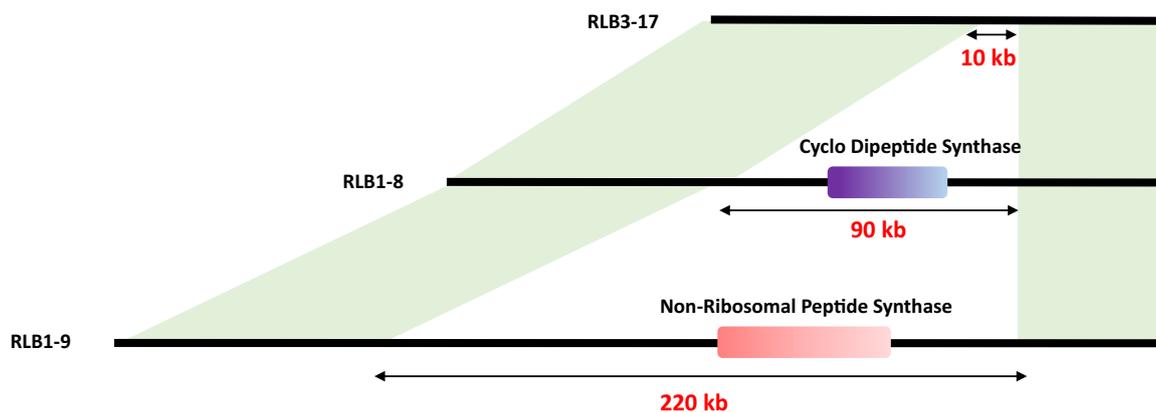


FIG 4 Identification of biosynthetic gene clusters in a variability hot spot. The scheme is illustrating a hot spot of variability observed between three strains (RLB3-17, RLB1-8, and RLB1-9). In two of them, different biosynthetic gene clusters were predicted by antiSMASH: one including a tRNA-dependent cyclodipeptide synthase in RLB1-8 and one including a nonribosomal peptide synthase in RLB1-9. Regions in light green correspond to syntenic regions.

of these BGCs were present in a variability hot spot (Fig. 4), with three strains exhibiting different gene content at these locations. The strain RLB3-17 contained no BGC, while RLB1-8 harbored a cyclodipeptide synthase BGC and RLB1-9 a nonribosomal peptide synthetase (NRPS) BGC. The disruption of a key gene belonging to the NRPS BGC (Fig. 4) only present in the *Streptomyces* killer strains (including RLB1-9) abolished the inhibitory activity against *Bacillus* (see Fig. S4), proving that this BGC is responsible for the antimicrobial activity of the killer strains under these conditions.

DISCUSSION

In this work, we isolated *Streptomyces* conspecifics at a soil microscale, and we revealed by genome comparison the extent of the massive gene flux that occurs in this natural population. This is reflected by a large pangenome, with almost clonal strains harboring strain-specific gene content. The size of the pangenome is influenced by three parameters: (i) the capacity to capture new DNA sequences, (ii) ecological selection of adaptive gene sets, and (iii) the evolutionary time enabling the accumulation of genome divergence (42). The fact that the population pangenome appears reduced in comparison with that of the genus (Fig. 1B) might reflect a constrained flux of genetic information within the microhabitat that is dependent of the diversity of the incoming information mostly arising from sister strains. However, the genetic diversity remained high in our population, implying an intense rate of gain and loss of genome regions over a short evolutionary time.

In *Streptomyces*, we have shown recently that repair of chromosomal double-strand breaks (DSB) experimentally induced in the terminal regions triggers the formation of chromosomal rearrangements (43). The formation of short insertions was also associated with DSB repair depending of nonhomologous end joining (NHEJ). Although the underlying mechanisms are mostly unknown, the plasticity of the subtelomeric regions appears as a hallmark of chromosome linearity in many organisms, from the limited range of bacterial taxa harboring linear chromosomes (44) to eukaryotes (45), including humans (46). In *Streptomyces*, tolerance to DNA rearrangements in the terminal regions could find its origins in the ancestral chromosomal linearization event. It is assumed that the ancestral actinomycete chromosome (i.e., harboring the ancestral core genes) was circular and linearized by recombination with an incoming linear replicon, leading to the addition of chromosomal arms harboring mostly contingency genes (47). However, if the ancestral gain of chromosomal arms can explain the tolerance to DNA rearrangements, it does not shed light on the evolutionary mechanisms that led to the contemporary variability of the *Streptomyces* genome (8, 9, 48). With respect to this, previous work has shown that chromosomal regions separating the core region from chromosomal arms were riddled with short indels (1 to 10 genes) forming a gradient

toward the terminal ends (5). We speculate that the indels, observed at the population level, when cumulated over evolutionary times, gave rise to the genetic compartmentalization observed within the genus (5).

Considering the high prevalence and plasticity of AICEs, we further propose that conjugative mechanisms are responsible for the massive gene flux observed in indels. To mobilize DNA, the translocase TraB binds to specific 8-mer repeats (named *clt* for *cis* acting locus of transfer) present on the AICE and assumed to be also present on the chromosome (*clt* like). Thus, beyond promoting self-transfer of the elements, TraB is able to transfer chromosomal markers in *trans* (20). This property was characterized and exploited in early studies on *Streptomyces* genetics (49). While conjugative transfer of chromosomal DNA was proved to be efficient under laboratory conditions (49, 50), the extent (frequency and size of transferred DNA stretches) remains to be elucidated. The indels ranging from 1 to 241 genes give an indication of the size of the chromosomal DNA stretches that may be transferred through conjugational processes. This is reminiscent of the distributive conjugal transfer in mycobacteria that enables the transfer of chromosomal fragments creating chimeric chromosomes in the recipient cells (51, 52). It is tempting to speculate that several DNA fragments could be transferred at once and generate multiple and diverse recombination events in the recipient *Streptomyces* hyphae.

At the ecological level, genome plasticity increases the functional potential at the strain but also the population level. For instance, it can enable the specific expression of specialized metabolites that can constitute public goods. Bacterial populations (i.e., sympatric closely related strains) can act as social units, meaning that their competitive interactions are more intense toward individuals not belonging to the population rather than between members of the population. Variability of SM production and resistance has already been reported as a driver of ecological cohesive units in *Vibrio* (53) and *Streptomyces* (23). Accordingly, differential SM production and resistance can be used as a proxy to understand the ecological dynamics of the bacterial population (53–55). Here, we confirm that the intense gene flux that *Streptomyces* conspecifics experienced promotes this variability, as previously proposed by Vetsigian et al. (23).

The production of SM within a bacterial population may underpin two principal ecological outcomes, bearing in mind that actual interactions in the soil may be more complex or different than under laboratory conditions. The production of antimicrobial activity and its respective resistance may benefit the sole carrier of the antibiotic production gene and mediate competition with conspecifics as well as with other species. Alternatively, an antibiotic produced by only some individuals of the population where the rest are resistant supports the hypothesis that antibiotics can constitute public goods, benefiting nonproducing but resistant conspecifics. Since none of our isolates was able to inhibit its conspecifics, it supports the hypothesis that they formed a cooperative population. The acquisition of the capacity to produce antimicrobial activity together with the capacity to resist its toxicity can result from a single HGT event thanks to the clustering of biosynthetic and resistance genes in the same genetic region, which is a typical trait of BGCs in actinomycetes. Therefore, the emergence of individuals able to overcome their conspecifics by producing a lethal activity is easily achieved by HGT. However, as we show in this work, our *Streptomyces* population experienced a massive and intense gene flux, which, in addition to favoring the acquisition of BGCs by individuals can also promote a rapid dissemination of all or parts of BGCs to the rest of the population. The replacement of a complete BGC obviously diversifies the antimicrobial arsenal, while its loss accompanied or not by that of the resistant determinant can result in a high intrapopulation SM variability. Further, recombination within or between BGCs may be enhanced thanks to their highly redundant structure (i.e., modular organization of NRPS and polyketide synthase [PKS]) and foster the emergence of new BGCs and the production of new antimicrobial

compounds. The intense turnover of the BGC and resistance gene sets would constantly modulate the interactions within the population and, through this dynamic state, favor the social cohesiveness of the population.

MATERIALS AND METHODS

Strain isolation and maintenance. Four grains of soil on the order of cubic millimeters in size and distant from each other from 2 cm to a maximum of 8 cm were sampled from a clod of soil collected in the Montiers-sur-Saulx forest in France (GPS coordinates: 48°32'37.248''N, 5°18'21.946''E) and stored for 48 h at 4°C before processing. For bacterial isolation, each soil aggregate was dissolved 1:100 (wt/vol) in sterile water by vortexing for 15 min. Serial dilutions from 10^{-1} to 10^{-3} were spread on *Streptomyces* isolation medium (SIM) (56), and bacterial colonies showing a typical *Streptomyces* phenotype were randomly picked after 7 days of incubation at 30°C. After three consecutive subcultures on mannitol soy agar plates (57), strains were stored as spore suspensions in 20% glycerol (57).

Genome sequencing, annotation, and DNA manipulation. Total genomic DNA was isolated as previously described by Kieser et al. (57). Each genome was sequenced by one-dimension MinION (Oxford Nanopore Technologies, UK) (mean coverage of 108×) and further corrected with MiSeq sequencing (Illumina, CA, USA) (mean coverage 50×), enabling the acquisition of each chromosome in one scaffold and the identification of extrachromosomal elements when present. Sequencing and assembling were performed via the I2BC NGS platform (France). Genome sequencing data and genome accession numbers are listed in Table S2 in the supplemental material. Coding sequence prediction and annotation were performed using the NCBI Prokaryotic Genome Annotation Pipeline (58). Genes involved in SM biosynthesis were predicted with antiSMASH (41). To disrupt a gene cluster putatively involved in the biosynthesis of a NRPS in strain RLB1-9, a 897-bp sequence internal to the peptide synthetase gene (open reading frame [ORF]-07845) was amplified and cloned into the suicide vector pJ8668 to give pJ8668-*intORF-07845* (59). Intergeneric *Escherichia coli*/*Streptomyces* conjugation (60) followed by selection for apramycin resistance (pJ8668) allowed us to select NRPS-encoding gene disruption by homologous recombination. Independent mutant clones (i.e., from distinct transformation assays) were tested for their antimicrobial activities. Pulsed-field gel electrophoresis (PFGE) was performed on whole genomic DNA in agarose plugs as described previously (61).

Phylogenetic analyses. 16S rRNA genes and MLST genes (*atpD*, *gyrB*, *recA*, *rpoB*, and *trpB*) were amplified using primers and PCR conditions described by Rintala et al. (62) and Guo et al. (63), respectively. Neighbor-joining phylogenetic trees were built using single or concatenated MLST gene (3,178 positions) sequences with a Kimura's two-parameter distance correction and 100 bootstrap replicates (MEGA7 software [64]). Phylogenomic analyses were performed with the nucleotide sequences of the core genome identified by a Best BLAST Hit analysis (65), subsequently aligned with Muscle (v3.8.31 [66]), and corrected with Gblocks (67). From the 8,778 genes of the population core genome, we selected the 5,213 genes (5,149,602 nucleotide positions) shared with *Streptomyces avermitilis*, the reference strain used as a root in the phylogenetic reconstruction. Maximum likelihood phylogenomic tree (GTR, 100 bootstrap replicates) was built with RaxML (68). Python scripts (unpublished) were used to calculate the average nucleotide identity (ANIb) (69) and to identify and to extrapolate pangenomes. Pangenome visualization was performed using Anvi'o (70), and extrapolation graphs were computed with the UpsetR package in script R. Functional AICEs and ICEs were identified first by searching signature genes (replication, transfer, excision, and integration) using BLASTP (65) as described by Bordeleau et al. (18) and then by identifying direct repeats flanking the element. Breaks of synteny were identified using an in-lab developed software (unpublished and available upon request).

Antibacterial activity assays. Drops of *Streptomyces* spore suspensions (5 μ l of spore stock, 10^7 spores per ml) were spotted and incubated for 5 days at 30°C on GA medium plates (54). Plates were overlaid with a Luria-Bertani top (0.4% agar), inoculated with *Bacillus* (optical density at 600 nm [OD₆₀₀] of 0.1), and incubated at 4°C for 2 h and then at 30°C overnight. Growth inhibition of *Bacillus* was assessed by the observation of a zone of clearance. To test *Streptomyces-Streptomyces* inhibitory capacities, spores of a first strain were spotted on mannitol soy medium plates for a 3-day incubation at 30°C. Then, a similar deposit of spores of the second strain (indicator) was adjacency spotted (5 mm) and incubated for 5 days at 30°C. The absence of phenotypic variation in comparison with control strains inoculated alone was considered a neutral interaction, and growth or sporulation alterations were recorded as antagonism.

Accession number(s). The genome accession numbers for the strains in this study are CP041650 to CP041654 and CP041601 to CP041613.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mBio.01533-19>.

FIG S1, PDF file, 0.6 MB.

FIG S2, PDF file, 0.4 MB.

FIG S3, PDF file, 0.4 MB.

FIG S4, PDF file, 0.4 MB.

TABLE S1, PDF file, 0.4 MB.

TABLE S2, PDF file, 0.5 MB.

TABLE S3, PDF file, 0.8 MB.

ACKNOWLEDGMENTS

This work was supported by a grant overseen by the French National Research Agency (ANR) as part of the Investissements d'Avenir program (ANR-11-LABX-0002-01, Lab of Excellence ARBRE).

We thank the research system SOERE-OPE (Système d'Observation et d'Experimentation au long terme pour la Recherche en Environnement, Observatoire Pérenne de l'Environnement) for access to the sampling site.

REFERENCES

- Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, Meier-Kolthoff JP, Klenk H-P, Clément C, Ouhdouch Y, van Wezel GP. 2016. Taxonomy, physiology, and natural products of *Actinobacteria*. *Microbiol Mol Biol Rev* 80:1–43. <https://doi.org/10.1128/MMBR.00019-15>.
- Olanrewaju OS, Babalola OO. 2019. *Streptomyces*: implications and interactions in plant growth promotion. *Appl Microbiol Biotechnol* 103:1179–1188. <https://doi.org/10.1007/s00253-018-09577-y>.
- Aigle B, Lautre S, Spitteller D, Dickschat JS, Challis GL, Leblond P, Pernodet J-L. 2014. Genome mining of *Streptomyces ambofaciens*. *J Ind Microbiol Biotechnol* 41:251–263. <https://doi.org/10.1007/s10295-013-1379-y>.
- Baltz RH. 2017. Gifted microbes for genome mining and natural product discovery. *J Ind Microbiol Biotechnol* 44:573–588. <https://doi.org/10.1007/s10295-016-1815-x>.
- Choulet F, Aigle B, Gallois A, Mangenot S, Gerbaud C, Truong C, Francou F-X, Fourrier C, Guérineau M, Decaris B, Barbe V, Pernodet J-L, Leblond P. 2006. Evolution of the terminal regions of the *Streptomyces* linear chromosome. *Mol Biol Evol* 23:2361–2369. <https://doi.org/10.1093/molbev/msl108>.
- Redenbach M, Kieser HM, Denapate D, Eichner A, Cullum J, Kinashi H, Hopwood DA. 1996. A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. *Mol Microbiol* 21:77–96. <https://doi.org/10.1046/j.1365-2958.1996.6191336.x>.
- Tian X, Zhang Z, Yang T, Chen M, Li J, Chen F, Yang J, Li W, Zhang B, Zhang Z, Wu J, Zhang C, Long L, Xiao J. 2016. Comparative genomics analysis of *Streptomyces* species reveals their adaptation to the marine environment and their diversity at the genomic level. *Front Microbiol* 7:998. <https://doi.org/10.3389/fmicb.2016.00998>.
- Kim J-N, Kim Y, Jeong Y, Roe J-H, Kim B-G, Cho B-K. 2015. Comparative genomics reveals the core and accessory genomes of *Streptomyces* species. *J Microbiol Biotechnol* 25:1599–1605. <https://doi.org/10.4014/jmb.1504.04008>.
- Zhou Z, Gu J, Li Y-Q, Wang Y. 2012. Genome plasticity and systems evolution in *Streptomyces*. *BMC Bioinformatics* 13 Suppl 10:S8. <https://doi.org/10.1186/1471-2105-13-S10-S8>.
- Lin YS, Kieser HM, Hopwood DA, Chen CW. 1993. The chromosomal DNA of *Streptomyces lividans* 66 is linear. *Mol Microbiol* 10:923–933. <https://doi.org/10.1111/j.1365-2958.1993.tb00964.x>.
- Huang CH, Lin YS, Yang YL, Huang SW, Chen CW. 1998. The telomeres of *Streptomyces* chromosomes contain conserved palindromic sequences with potential to form complex secondary structures. *Mol Microbiol* 28:905–916. <https://doi.org/10.1046/j.1365-2958.1998.00856.x>.
- Chen CW, Huang C-H, Lee H-H, Tsai H-H, Kirby R. 2002. Once the circle has been broken: dynamics and evolution of *Streptomyces* chromosomes. *Trends Genet* 18:522–529. [https://doi.org/10.1016/S0168-9525\(02\)02752-X](https://doi.org/10.1016/S0168-9525(02)02752-X).
- Thibessard A, Leblond P. 2014. Subtelomere plasticity in the bacterium *Streptomyces*, p 243–258. *In* Louis EJ, Becker MM (ed), *Subtelomeres*. Springer, Berlin, Germany.
- Doroghazi JR, Buckley DH. 2010. Widespread homologous recombination within and between *Streptomyces* species. *ISME J* 4:1136–1143. <https://doi.org/10.1038/ismej.2010.45>.
- Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299–304. <https://doi.org/10.1038/35012500>.
- Stuttard C. 1982. Temperate phages of *Streptomyces venezuelae*: lysogeny and host specificity shown by phages SV1 and SV2. *Microbiology* 128:115–121. <https://doi.org/10.1099/00221287-128-1-115>.
- Burke J, Schneider D, Westpheling J. 2001. Generalized transduction in *Streptomyces coelicolor*. *Proc Natl Acad Sci U S A* 98:6289–6294. <https://doi.org/10.1073/pnas.101589398>.
- Bordeleau E, Ghinet MG, Burrus V. 2012. Diversity of integrating conjugative elements in actinobacteria: coexistence of two mechanistically different DNA-translocation systems. *Mob Genet Elements* 2:119–124. <https://doi.org/10.4161/mge.20498>.
- Thoma L, Muth G. 2016. Conjugative DNA-transfer in *Streptomyces*, a mycelial organism. *Plasmid* 87–88:1–9. <https://doi.org/10.1016/j.plasmid.2016.09.004>.
- Pettis GS. 2018. Spreading the news about the novel conjugation mechanism in *Streptomyces* bacteria. *Environ Microbiol Rep* 10:503–510. <https://doi.org/10.1111/1758-2229.12659>.
- McDonald BR, Currie CR. 2017. Lateral gene transfer dynamics in the ancient bacterial genus *Streptomyces*. *mBio* 8:e00644-17. <https://doi.org/10.1128/mBio.00644-17>.
- Antony-Babu S, Stien D, Eparvier V, Parrot D, Tomasi S, Suzuki MT. 2017. Multiple *Streptomyces* species with distinct secondary metabolomes have identical 16S rRNA gene sequences. *Sci Rep* 7:11089. <https://doi.org/10.1038/s41598-017-11363-1>.
- Vetsigian K, Jajoo R, Kishony R. 2011. Structure and evolution of *Streptomyces* interaction networks in soil and *in silico*. *PLoS Biol* 9:e1001184. <https://doi.org/10.1371/journal.pbio.1001184>.
- Vicente CM, Thibessard A, Lorenzi J-N, Benhadj M, Hôtel L, Gacemi-Kirane D, Lespinet O, Leblond P, Aigle B. 2018. Comparative genomics among closely related *Streptomyces* strains revealed specialized metabolite biosynthetic gene cluster diversity. *Antibiotics (Basel)* 7:e86. <https://doi.org/10.3390/antibiotics7040086>.
- Harris HMB, Bourin MJB, Claesson MJ, O'Toole PW. 2017. Phylogenomics and comparative genomics of *Lactobacillus salivarius*, a mammalian gut commensal. *Microb Genom* 3:e000115. <https://doi.org/10.1099/mgen.0.000115>.
- Park CJ, Andam CP. 2019. Within-species genomic variation and variable patterns of recombination in the tetracycline producer *Streptomyces rimosus*. *Front Microbiol* 10:552. <https://doi.org/10.3389/fmicb.2019.00552>.
- Croucher NJ, Coupland PG, Stevenson AE, Callendrello A, Bentley SD, Hanage WP. 2014. Diversification of bacterial genome content through distinct mechanisms over different timescales. *Nat Commun* 5:5471. <https://doi.org/10.1038/ncomms6471>.
- Levade I, Terrat Y, Leducq J-B, Weil AA, Mayo-Smith LM, Chowdhury F, Khan AI, Boncy J, Buteau J, Ivers LC, Ryan ET, Charles RC, Calderwood SB, Qadri F, Harris JB, LaRocque RC, Shapiro BJ. 2017. *Vibrio cholerae* genomic diversity within and between patients. *Microb Genom* 3:e000142. <https://doi.org/10.1099/mgen.0.000142>.
- Bruns H, Crüsemann M, Letzel A-C, Alanjary M, McClrnerney JO, Jensen PR, Schulz S, Moore BS, Ziemert N. 2018. Function-related replacement of bacterial siderophore pathways. *ISME J* 12:320–329. <https://doi.org/10.1038/ismej.2017.137>.
- Stefanic P, Kraigher B, Lyons NA, Kolter R, Mandic-Mulec I. 2015. Kin discrimination between sympatric *Bacillus subtilis* isolates. *Proc Natl Acad Sci U S A* 112:14042–14047. <https://doi.org/10.1073/pnas.1512671112>.

31. Wielgoss S, Didelot X, Chaudhuri RR, Liu X, Weedall GD, Velicer GJ, Vos M. 2016. A barrier to homologous recombination between sympatric strains of the cooperative soil bacterium *Myxococcus xanthus*. *ISME J* 10:2468–2477. <https://doi.org/10.1038/ismej.2016.34>.
32. Vos M, Wolf AB, Jennings SJ, Kowalchuk GA. 2013. Micro-scale determinants of bacterial diversity in soil. *FEMS Microbiol Rev* 37:936–954. <https://doi.org/10.1111/1574-6976.12023>.
33. Wolf AB, Vos M, de Boer W, Kowalchuk GA. 2013. Impact of matric potential and pore size distribution on growth dynamics of filamentous and non-filamentous soil bacteria. *PLoS One* 8:e83661. <https://doi.org/10.1371/journal.pone.0083661>.
34. Raynaud X, Nunan N. 2014. Spatial ecology of bacteria at the microscale in soil. *PLoS One* 9:e87217. <https://doi.org/10.1371/journal.pone.0087217>.
35. Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, Connor R, Fiorini N, Funk K, Hefferon T, Holmes JB, Kim S, Kimchi A, Kitts PA, Lathrop S, Lu Z, Madden TL, Marchler-Bauer A, Phan L, Schneider VA, Schoch CL, Pruitt KD, Ostell J. 2019. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 47:D23–D28. <https://doi.org/10.1093/nar/gky1069>.
36. Wenner T, Roth V, Fischer G, Fourrier C, Aigle B, Decaris B, Leblond P. 2003. End-to-end fusion of linear deleted chromosomes initiates a cycle of genome instability in *Streptomyces ambofaciens*. *Mol Microbiol* 50: 411–425. <https://doi.org/10.1046/j.1365-2958.2003.03698.x>.
37. Fischer G, Wenner T, Decaris B, Leblond P. 1998. Chromosomal arm replacement generates a high level of intraspecific polymorphism in the terminal inverted repeats of the linear chromosomal DNA of *Streptomyces ambofaciens*. *Proc Natl Acad Sci U S A* 95:14296–14301. <https://doi.org/10.1073/pnas.95.24.14296>.
38. Uchida T, Miyawaki M, Kinashi H. 2003. Chromosomal arm replacement in *Streptomyces griseus*. *J Bacteriol* 185:1120–1124. <https://doi.org/10.1128/jb.185.3.1120-1124.2003>.
39. Clerc S, Simonet P. 1996. Efficiency of the transfer of a pSAM2-derivative plasmid between two strains of *Streptomyces lividans* in conditions ranging from agar slants to non-sterile soil microcosms. *FEMS Microbiol Ecol* 21:157–165. <https://doi.org/10.1111/j.1574-6941.1996.tb00343.x>.
40. Sullivan JT, Ronson CW. 1998. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proc Natl Acad Sci U S A* 95:5145–5149. <https://doi.org/10.1073/pnas.95.9.5145>.
41. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH. 2017. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* 45:W36–W41. <https://doi.org/10.1093/nar/gkx319>.
42. McInerney JO, McNally A, O'Connell MJ. 2017. Why prokaryotes have pangenomes. *Nat Microbiol* 2:17040. <https://doi.org/10.1038/nmicrobiol.2017.40>.
43. Hoff G, Bertrand C, Piotrowski E, Thibessard A, Leblond P. 2018. Genome plasticity is governed by double strand break DNA repair in *Streptomyces*. *Sci Rep* 8:5272. <https://doi.org/10.1038/s41598-018-23622-w>.
44. Casjens S, Palmer N, van Vugt R, Huang WM, Stevenson B, Rosa P, Lathigra R, Sutton G, Peterson J, Dodson RJ, Haft D, Hickey E, Gwinn M, White O, Fraser CM. 2002. A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. *Mol Microbiol* 35:490–516. <https://doi.org/10.1046/j.1365-2958.2000.01698.x>.
45. Ricchetti M, Dujon B, Fairhead C. 2003. Distance from the chromosome end determines the efficiency of double strand break repair in subtelomeres of haploid yeast. *J Mol Biol* 328:847–862. [https://doi.org/10.1016/S0022-2836\(03\)00315-2](https://doi.org/10.1016/S0022-2836(03)00315-2).
46. Linardopoulou EV, Williams EM, Fan Y, Friedman C, Young JM, Trask BJ. 2005. Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature* 437:94–100. <https://doi.org/10.1038/nature04029>.
47. Bentley SD, Chater KF, Cerdeño-Tárraga A-M, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang C-H, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabinowitsch E, Rajandream M-A, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417:141–147. <https://doi.org/10.1038/417141a>.
48. Hsiao N-H, Kirby R. 2008. Comparative genomics of *Streptomyces avermitilis*, *Streptomyces cattleya*, *Streptomyces maritimus* and *Kitasatospora aureofaciens* using a *Streptomyces coelicolor* microarray system. *Antonie Van Leeuwenhoek* 93:1–25. <https://doi.org/10.1007/s10482-007-9175-1>.
49. Hopwood DA. 2006. Soil to genomics: the *Streptomyces* chromosome. *Annu Rev Genet* 40:1–23. <https://doi.org/10.1146/annurev.genet.40.110405.090639>.
50. Hopwood DA, Kieser T, Wright HM, Bibb MJ. 1983. Plasmids, recombination and chromosome mapping in *Streptomyces lividans* 66. *J Gen Microbiol* 129:2257–2269. <https://doi.org/10.1099/00221287-129-7-2257>.
51. Gray TA, Krywy JA, Harold J, Palumbo MJ, Derbyshire KM. 2013. Distributive conjugal transfer in mycobacteria generates progeny with meiotic-like genome-wide mosaicism, allowing mapping of a mating identity locus. *PLoS Biol* 11:e1001602. <https://doi.org/10.1371/journal.pbio.1001602>.
52. Derbyshire KM, Gray TA. 2014. Distributive conjugal transfer: new insights into horizontal gene transfer and genetic exchange in mycobacteria. *Microbiol Spectr* 2:MGM2-0022-2013. <https://doi.org/10.1128/microbiolspec.MGM2-0022-2013>.
53. Cordero OX, Wildschutte H, Kirkup B, Proehl S, Ngo L, Hussain F, Le Roux F, Mincer T, Polz MF. 2012. Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. *Science* 337:1228–1231. <https://doi.org/10.1126/science.1219385>.
54. Abrudan MI, Smakman F, Grimbergen AJ, Westhoff S, Miller EL, van Wezel GP, Rozen DE. 2015. Socially mediated induction and suppression of antibiosis during bacterial coexistence. *Proc Natl Acad Sci U S A* 112:11054–11059. <https://doi.org/10.1073/pnas.1504076112>.
55. Butaitė E, Baumgartner M, Wyder S, Kümmerli R. 2017. Siderophore cheating and cheating resistance shape competition for iron in soil and freshwater *Pseudomonas* communities. *Nat Commun* 8:414. <https://doi.org/10.1038/s41467-017-00509-4>.
56. D'Costa VM, McGrann KM, Hughes DW, Wright GD. 2006. Sampling the antibiotic resistome. *Science* 311:374–377. <https://doi.org/10.1126/science.1120800>.
57. Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA. 2000. *Practical Streptomyces genetics*. John Innes Foundation, Norwich, United Kingdom.
58. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 44: 6614–6624. <https://doi.org/10.1093/nar/gkw569>.
59. Sun J, Kelemen GH, Fernández-Abalos JM, Bibb MJ. 1999. Green fluorescent protein as a reporter for spatial and temporal gene expression in *Streptomyces coelicolor* A3(2). *Microbiology* 145:2221–2227. <https://doi.org/10.1099/00221287-145-9-2221>.
60. Kim M-K, Ha H-S, Choi S-U. 2008. Conjugal transfer using the bacteriophage phiC31 att/int system and properties of the attB site in *Streptomyces ambofaciens*. *Biotechnol Lett* 30:695–699. <https://doi.org/10.1007/s10529-007-9586-0>.
61. Leblond P, Francou FX, Simonet J-M, Decaris B. 1990. Pulsed-field gel electrophoresis analysis of the genome of *Streptomyces ambofaciens* strains. *FEMS Microbiol Lett* 72:79–88. <https://doi.org/10.1111/j.1574-6968.1990.tb03866.x>.
62. Rintala H, Nevalainen A, Rönkä E, Suutari M. 2001. PCR primers targeting the 16S rRNA gene for the specific detection of streptomycetes. *Mol Cell Probes* 15:337–347. <https://doi.org/10.1006/mcpr.2001.0379>.
63. Guo Y, Zheng W, Rong X, Huang Y. 2008. A multilocus phylogeny of the *Streptomyces griseus* 16S rRNA gene clade: use of multilocus sequence analysis for streptomycete systematics. *Int J Syst Evol Microbiol* 58: 149–159. <https://doi.org/10.1099/ijs.0.65224-0>.
64. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33: 1870–1874. <https://doi.org/10.1093/molbev/msw054>.
65. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
66. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. <https://doi.org/10.1186/1471-2105-5-113>.
67. Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein se-

- quence alignments. *Syst Biol* 56:564–577. <https://doi.org/10.1080/10635150701472164>.
68. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
69. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 57:81–91. <https://doi.org/10.1099/ijs.0.64483-0>.
70. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3:e1319. <https://doi.org/10.7717/peerj.1319>.

CHAPITRE III

III. Etude de la variabilité chromosomique terminale des génomes de 11 souches sympatriques et conspécifiques de *Streptomyces*.

Telomeric and sub-telomeric regions undergo rapid turnover within a *Streptomyces* population

Tidjani A.-R., Bontemps C., Leblond P.

Article N°2 : Soumis au journal Scientific report le xx/10/16:

Résumé des résultats de l'article :

Les génomes des bactéries du genre *Streptomyces* sont composés d'un chromosome linéaire ; ce chromosome est complété souvent par la présence des réplicons extrachromosomiques qui peuvent être linéaires ou circulaires. Les réplicons linéaires des génomes de *Streptomyces* partagent une structure formée de répétitions terminales inversées (TIR) à chacune de leur extrémité. La taille de ces TIR, qui dans certains réplicons comme le chromosome de *S. avermitilis* se réduisent à quelques dizaines de nucléotides peut atteindre 1 Mb à 1,4 Mb dans les souches de *S. avermitilis* et *S. ambofaciens* respectivement. Cette variabilité de la taille des TIR a été reportée au niveau inter et intra-spécifique et pourrait être une conséquence de la réparation des cassures double-brin survenant dans les régions terminales.

Comme chez d'autres bactéries, la réplication du chromosome linéaire de *Streptomyces* débute au niveau de l'*oriC* ; cependant cette réplication n'est pas continue jusqu'à l'extrémité 3' du brin parental qui nécessite une amorce pour la synthèse du brin complémentaire. Ce rôle est joué par un couple de protéines Tap et Tpg codé par un opéron de deux gènes et par une structure terminale typique en forme de tiges et boucles qu'on retrouve aux extrémités des TIR. Ces structures sont formées par les séquences des télomères. A la fin de la réplication terminale, la protéine Tpg demeure attachée de façon covalente à l'extrémité 5' du chromosome, le protégeant ainsi de la dégradation par les exonucléases. Des études précédentes ont démontré une conservation des séquences télomériques dont le rôle serait de conserver les structures typiques nécessaires au recrutement des protéines terminales. Dans une étude précédente, nous avons montré qu'il existe une grande diversité génétique au sein des génomes des souches étroitement apparentés d'une population de *Streptomyces* (Tidjani *et al.*, 2019a) et que cette diversité concernant particulièrement les régions terminales. Nous nous sommes demandé si les télomères, associés aux protéines terminales ainsi que les TIR pouvaient être impactés par cette dynamique de génomes observée dans les régions terminales.

Pour ce faire, les TIR des réplicons linéaires retrouvés au sein de la population ont été identifiés *in silico* en utilisant les lectures Illumina issus du séquençage qu'on réalignait sur les réplicons linéaires. Cette approche a permis d'identifier et de comparer la taille et les séquences de ces TIR. Une diversité de la taille des TIR a été déterminée au niveau des chromosomes. Seuls deux des 3 réplicons extrachromosomiques linéaires possèdent des TIRs, même si des télomères ont été identifiés pour les 3 réplicons. La variabilité de la taille des TIR confirme la plasticité des régions terminales décrite précédemment. Les réarrangements survenant dans ces régions impliquent des duplications qui suggèrent l'intervention des mécanismes d'homogénéisation des régions terminales.

L'analyse des séquences télomériques a également permis de mettre à jour une diversité de ces dernières au sein de la population même si elles demeurent conservées au sein des sous-clades formées par certaines souches de la population. Les différents télomères identifiés au sein de cette population suggèrent que les événements de recombinaisons homologues survenues au cours de la courte histoire évolutive de ces souches ont également impacté les extrémités des réplicons.

Nous nous sommes également intéressé à la co-occurrence des télomères et des protéines terminales avec lesquelles ils interagissent. La co-occurrence des télomères et des protéines terminales spécifiques aux sous-clades formés par les souches de la population a permis de reconstruire l'histoire évolutive des régions terminales faite d'évènements d'acquisition et de remplacement de ces régions ayant aboutis à la diversité constatée au sein de cette population.

Telomeric and sub-telomeric regions undergo rapid turnover within a *Streptomyces* population

Abdoul-Razak Tidjani¹, Cyril Bontemps^{1*}, Pierre Leblond^{1*}

¹Université de Lorraine, INRA, DynAMic, F-54000 Nancy, France

*co-corresponding authors

Cyril Bontemps cyril.bontemps@univ-lorraine.fr

Pierre Leblond pierre.leblond@univ-lorraine.fr

Tel +33 (0)3 72 74 51 43

Abstract (200 wds)

Genome dynamics was investigated within natural populations of the soil bacterium *Streptomyces*. The exploration of a set of closely related strains isolated from micro-habitats of a forest soil revealed a strong diversity of the terminal structures of the linear chromosome, i.e. terminal inverted repeats (TIR). Insertions, deletions and translocations of large sizes were observed and putative transfer events between strains could be inferred. In addition, the telomere and its cognate terminal protein complex which is required for terminal replication and chromosome maintenance was shown to be variable within the population probably reflecting telomere exchanges between chromosome and other linear replicon (i.e., plasmids). Considering the close genetic relatedness of the strains, these data suggest the terminal regions are prone to a high turnover thanks to a high recombination rate favoring horizontal gene transfer.

Introduction

Streptomyces are filamentous soil bacteria producing a large array of extracellular enzymatic activities involved in biogeochemical cycles (i.e. degradation of organic matter, mineral weathering...) (1, 2) and a large chemical diversity of metabolites involved in the establishment of complex biotic interactions within the microbial community as well as with plants or insects. Their arsenal of enzymes and of specialized metabolites is largely exploited for industry (3, 4) and medical applications. The latter include the production of antibiotics, herbicides, or antitumor compounds.

Streptomyces possess genomes among the largest ever reported for bacteria with a 6 Mb to 12 Mb chromosome. They share with *Borrelia burgdorferi*,⁽⁵⁾ *Agrobacterium tumefaciens* (6) and other actinobacteria such as *Rhodococcus* (7) the rare characteristic of chromosome linearity. The linear chromosomal DNA is completed by the presence of linear as well as circular plasmids. *Streptomyces* linear replicons share the same invertron structure, i.e., the presence of terminal redundancies of variable sizes (TIRs for Terminal Inverted Repeats). Hence, while TIRs can be as short as several tens of nucleotides and serve as telomere (see below) such as in *S. avermitilis* (Lorenzi *et al.*, unpublished), they can reach important sizes with for example up to 1 Mb in *S. coelicolor* A3(2) (8) or 1.4 Mb in *S. ambofaciens* (9). Variability of the TIR size was reported at the interspecific (10) and intraspecific levels (8, 9, 11, 12). However, to date, no study investigated the TIR variability at the population level, that is to say, at a very short evolutionary scale.

Like in other bacteria, the chromosomal replication is initiated at *oriC* towards both extremities and duplicates the most part of the DNA with the exception of the very ends of the parental 3' DNA strands (failure to prime the discontinuous replication process of the lagging strand). This classical end-of-replication problem of linear chromosomes is overcome thanks to a complex of terminal proteins called TP-TAP and a *cis*-acting sequence of about 160 nucleotides which form critical stem-loops involved in terminal structures known as 'rabbit ears' or 'clover leaves' models (13). These constitute a functional analogue of eukaryotic telomeres by preventing chromosomal ends from progressive shortening and loss of genetic information through successive replication rounds. The TAP/TPG

complex (through TAP) allows priming DNA replication from the 3' end and filling of the replication gap. At the end of the 'end-patching' (14) the TPG protein remains covalently attached to the 5' end of the DNA strand ensuring the protection against nucleases and ready to recruit TAP for the next replication round.

The telomere sequence is highly conserved, but mostly to preserve the formation of the secondary structures needed for terminal protein recruitment and activities. Hence, several telomere sequences have been described for plasmid as well as chromosomal DNA (15–18). In the same way, several types of terminal protein complexes have been identified including those of the models *S. coelicolor* A3(2) and *S. griseus* (19–22).

In a previous work, we isolated and fully sequenced the genomes of a *Streptomyces* population consisting of eleven conspecific and sympatric strains isolated at micro-scale (23). They share close phylogenetic relationships revealed by identical 16S rRNA gene sequences and weak MLSA divergence, and thus form a population deriving from a recent common ancestor. In consequence, genome divergence within the population derives from genome rearrangements and gene acquisitions that occurred over a rather short evolutionary period. We also revealed a gradient of insertion and deletion events toward the chromosomal extremities supporting the hypothesis that the ends could correspond to hotspot of recombination and/or tolerated better DNA rearrangements. The chromosome extremities including the terminal inverted repeats (TIRs) and the telomeres from our strains provide an ideal material to observe chromosome plasticity in an ecological context, *i.e.*, from a natural population. Moreover, it gives access to the more recent molecular events (recombination) impacting the chromosome structure. In this work, we show that the ends of the linear replicons are highly variable, implying a high recombination activity between chromosomal ends as well as acquisition and loss of terminal replication machineries including telomeres and the TAP/TP partner complex.

Materials and Methods

Terminal inverted repeats (TIRs) identification

Genome sequencing of the *Streptomyces* chromosomes was described earlier (24). It consists in one scaffold per chromosome and the identification of extra-chromosomal elements (accession numbers CP 41601 to CP 416013 and CP 41650 to CP 41654). The terminal inverted repeats of the chromosomes and linear plasmids were determined *in silico* (CLC Genomics Workbench 6.0, Qiagen) by mapping the Illumina sequencing reads against the reference sequence of the corresponding replicon. The TIRs were determined as the terminal regions with a number of reads twice than the central region of each replicon.

Annotation, DNA fold prediction and phylogenetic analyses

The CDS prediction and functional annotation in the TIR sequences were performed using RAST (25). The secondary structures of the telomeres were constructed using the mfold Web Server ((26) <http://unafold.rna.albany.edu/?q=mfold/DNA-Folding-Form>) with a folding temperature of 30°C in 1M NaCl. Sequence alignments were performed with ClustalW Multiple alignment algorithm (27) and the software BioEdit (27) . The telomere phylogenetic tree was constructed and edited with MEGA X (28) using a Neighbour Joining (NJ) method with a K2P correction model. All positions with < 80% site coverage were eliminated and support for the tree branches was estimated with 100 bootstrap replicates.

To identify *tap-tpg* and *gtpB-gtpA* homologues, the amino-acids sequences of Tap and Tpg of *Streptomyces coelicolor* A3(2), *Streptomyces lidycus* A02, *S. lidycus* 103, *Streptomyces lividans* (accession number AAL05040.1), *Streptomyces avermitilis* (WP_010988960.1, WP_010988961.1) and those of the GtpB and GtpA of *Streptomyces griseus* (SGR_RS00475, SGR_RS00470) were used as a bait using Blast (29) in BioEdit. Homologues were identified with a cutoff of 50% of sequence identity and a minimum coverage of 98% between the sequences.

Comparative genomics and homologous recombination events detection

The comparison of the 11 TIRs was performed with the software Progressive Mauve tool (multiple alignment of conserved genomic sequence with rearrangements) (30) with default parameters. Recombination events were detected with the RDP4 v4.97 program (31) in a two-step procedure as described in González-Torres *et al.* (2019) (32). First, a full exploratory recombination scan using the nine methods available was performed. The detected events were then rescanned with the RDP, GENECONV, MaxCHI, Chimera and 3Seq algorithms and only events positively detected by three of these five methods were considered.

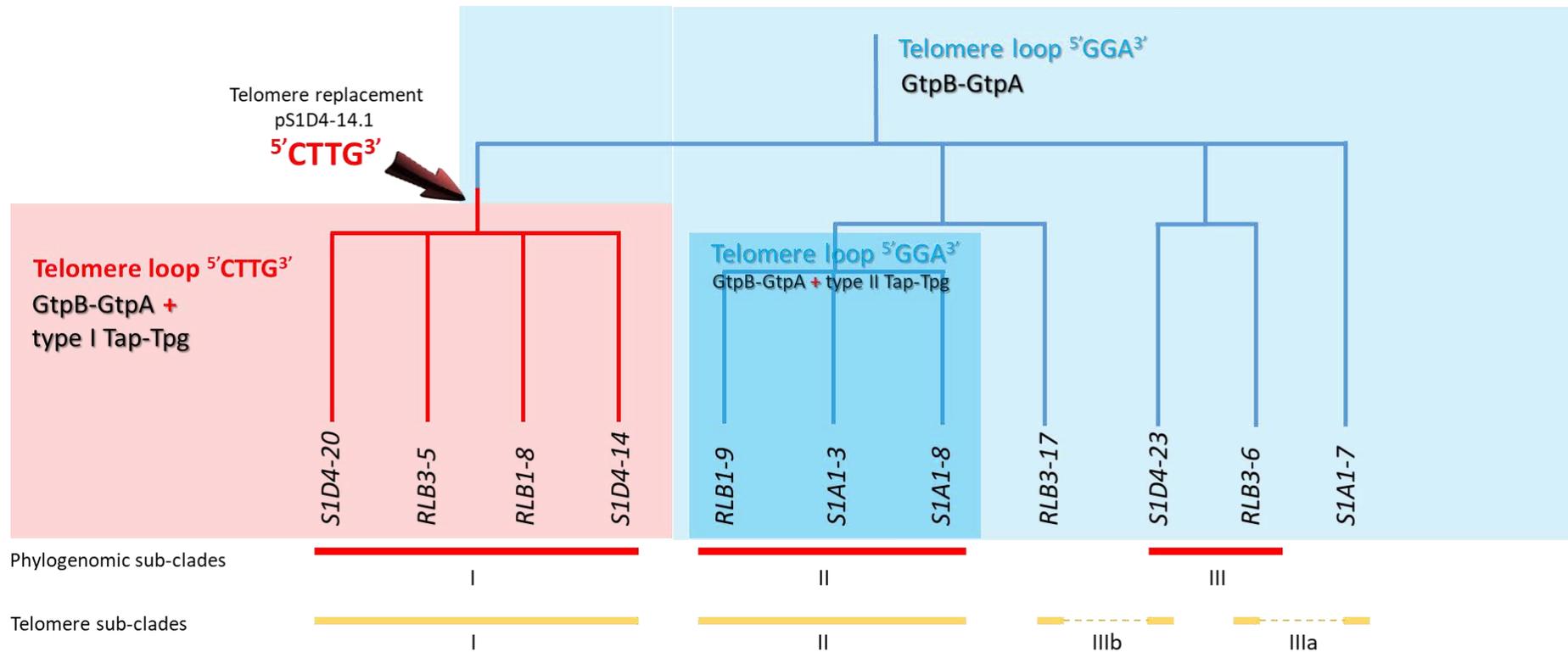


Fig. 1 Telomere and terminal protein distribution among the *Streptomyces* population. A maximum likelihood phylogenomic tree of the 11 strains was built based on 5,149,602 nucleotide positions. All branches are supported by bootstrap values of 100% (100 replicates). Branchings with bootstrap values below 70 are represented by polytomies. According to the tree topology and bootstrap values, three phylogenomic sub-clades (I, II and III) were defined. The strains belonging to a same phylogenomic sub-clade are connected with a red bar. The yellow bars connect stains harbouring identical or highly closely related telomere sequences (more than 92% sequence identity). The dashed yellow lines indicate incongruences between the phylogenomic tree and the distribution of similar telomeres. The coloured backgrounds represent the distribution of the chromosomal telomere loops and of the different *tap-tpg* /*gtpB-gtpA* systems. The light blue background encompasses strains for which only the *gtpB-gtpA* system was identified. Their chromosomal telomeres are typified with a $5'GGA^3'$ loop. A parsimonious evolutionary scenario suggest that the ancestor of the population shared the same combination. The turquoise background encompass strains having also telomeres with a $5'GGA^3'$ loop and a *gtpB-gtpA* operon, but with an additional type II *tap-tpg* operon. The light red background represent strains having the ancestral *gtpB-gtpA* operon, but also an additional type I *tap-tpg* operon and telomeres typified by a $5'CTTG^3'$ loop. The red arrow indicates a potential telomere replacement with a plasmid at the root of this sub-cluster.

Results and Discussion

Intraspecific variability at the chromosomal ends

The first step of our investigation was the detection and delimitation of the TIRs. *De novo* assemblies performed after next generation sequencing approaches (NGS) do not allow identifying such large intragenomic duplications (*i.e.*, duplications larger than the read size). However, the size of TIRs can be deduced from the analysis of read coverage (33). Hence, a twofold quantity of sequencing reads can be aligned to regions of the assembly that are readily duplicated in the genome. Mapping of the bulk reads onto the genomic sequence obtained at the assembler output is thus efficient to detect precisely the limits and extent of such terminal duplications (see Material and Methods section).

The size of the TIRs ranged from 303 kb to 579 kb for chromosomes (Table 1). In addition, among the 7 extrachromosomal elements detected in individuals of the population, three (pRLB1-9.2, pS1D4-20.1 and pS1D4-14.1) were linear. The two first possessed TIRs of 24 kb and 68 kb respectively, but no TIR was detected for pS1D4-14.1. Telomeres were found at the extremities of all three (see below).

This size variability reveals an intense plasticity of the terminal regions and could be correlated to the different phylogenomic sub-clades (Fig. 1). Hence, while the size of TIRs is conserved between strains belonging to the sub-clade I (*e.g.*, S1D4-20, RLB3-5, RLB1-8, and S1D4-14), it is highly variable within or between the other sub-clades (Table 1). For example, TIRs of strain RLB1-9 were shorter by 90 kb compared to those of its sister strains (*i.e.*, S1A1-3 and S1A1-8) and lack the distal regions of the chromosome, but without loss of linearity (not shown). Further, different types of recombination events (translocation, inversion and indel as labelled respectively A, B, C in Fig. 2) were revealed by global comparison of the TIRs of the 11 genomes using MAUVE (Fig. 2). Some regions are unique to one strain, suggesting that an insertion occurred (*e.g.* Fig. 2, region C). Since this DNA region (16.7 kb) is not present elsewhere in the other genomes, this probably results from acquisition by horizontal gene transfer.

Whatever the nature of the recombination event is, it modifies a single chromosomal end and consequently disrupt the terminal duplication. However, since we identify duplicated sequences, a

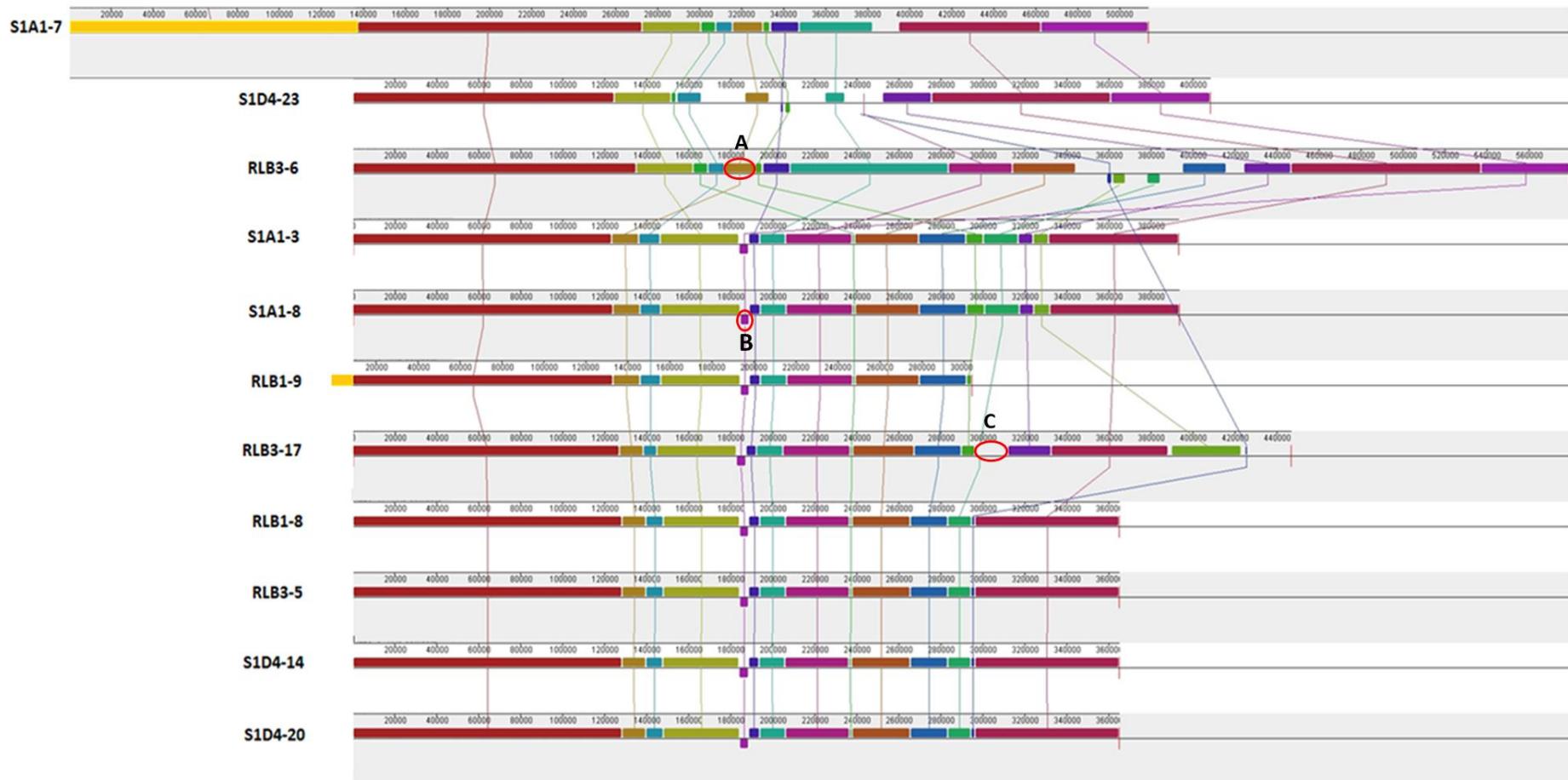


Fig. 2 Identification of rearrangements events within the Terminal Inverted Repeats. The TIR sequences of the 11 strains were compared and visualized using the progressive Mauve algorithm for Windows. The alignment was performed on collinear sequences in which S1A1-7 TIR was used as reference sequence. Boxes with identical colors represent local collinear blocks (LCBs), indicating homologous DNA regions shared by two or more chromosomes without sequence rearrangements. Lines are drawn between LCBs in two adjacent species to indicate homologous regions. The placement of a block below the axis indicates an inversion event. Examples of rearrangement events occurring in the TIRs are highlighted by red circles; A. translocation event, B. inversion event, C. indel event.

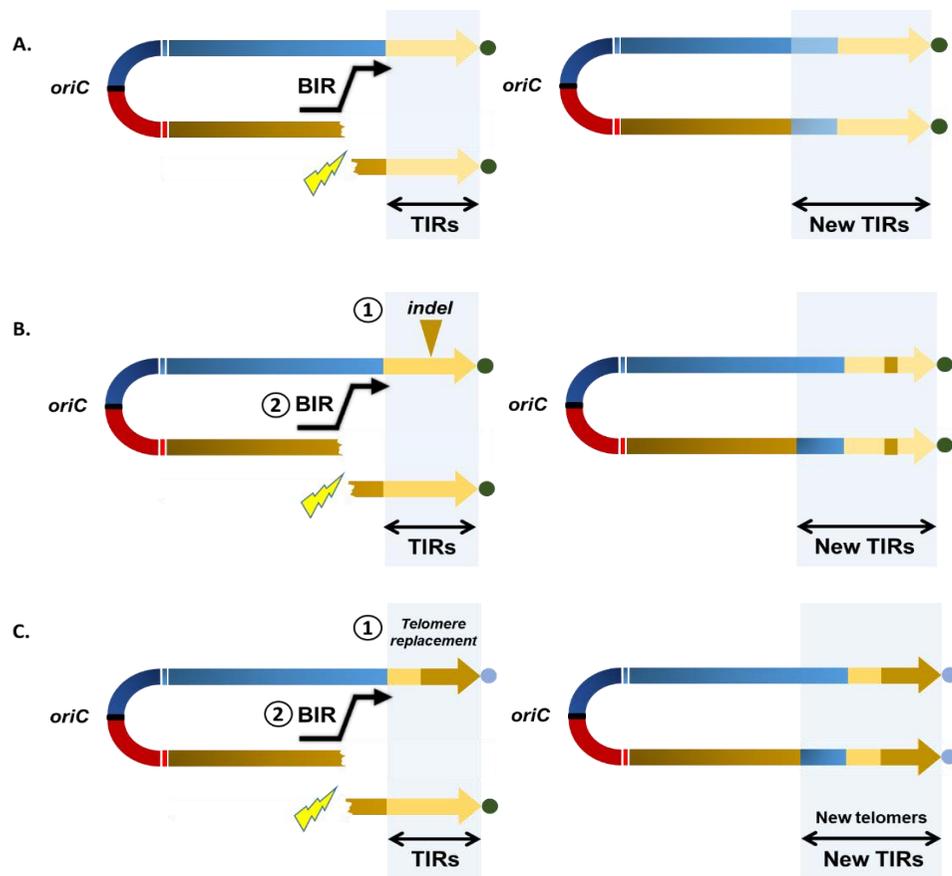


Fig. 3 Inter-chromosomal arm recombination scenarios leading to homogenization of TIRs. Double-headed arrows represent *Streptomyces* linear replicons. The TIRs are highlighted with a light gray frame and the telomeres are represented by colored circles. The yellow flash symbols represent a DNA double strand break (DSB) requiring an upstream Break Induced Replication (BIR) event to rescue the broken replicon and to keep linearity. The BIR event uses the other arm as a matrix and recopy it until the end, including the telomere. If the DSB occurs within a TIR, the BIR will not change its size or sequence (not represented). **A.** If the DSB occurs upstream the TIR, the BIR event will lead to the homogenization and extension in size of the TIRs. **B.** An insertion/deletion event (indel) occurred (1) in one TIR (yellow triangle) before a DSB event. Like before, the BIR event (2) upstream the TIR will lead to a size increase, but will also propagate the indel event by homogenization. **C.** In this case, a telomere replacement occurred in one arm (1) before the DSB. Following the BIR event (2), the arm homogenization will lead to a change of the telomere sequence.

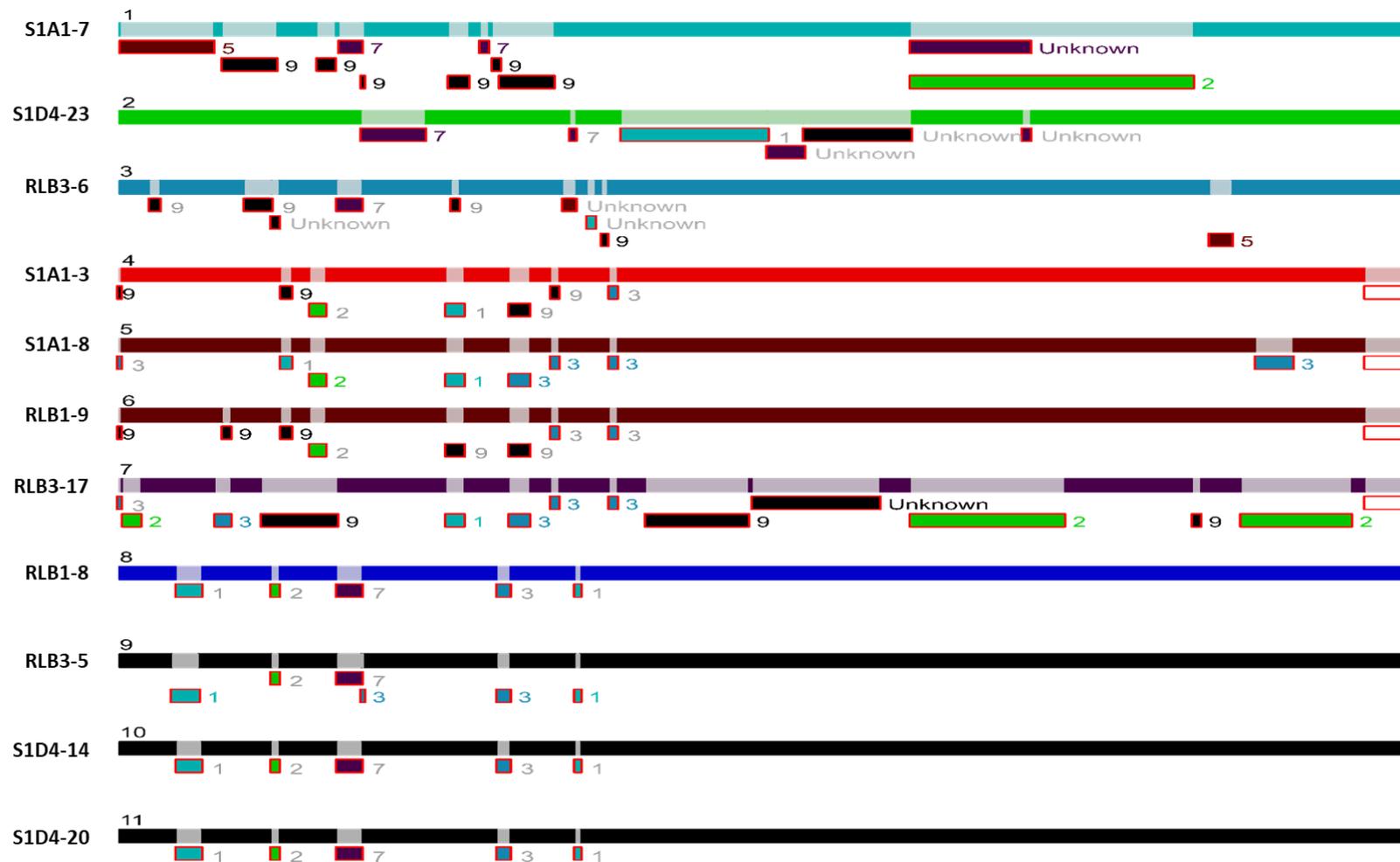


Fig. 4 RDP4 analysis for DNA recombination and transfer within TIRs of the *Streptomyces* population. Each bar represents the TIR sequence of one strain identified by its name and a number from 1 to 11. Regions of recombination events on each TIR are represented by a light shade of the corresponding TIR color. Sequence fragments from potential donors (parents from which recombinant sequence derived from) is depicted under each recombination region with the number corresponding to the donor strain. If the potential donor does not belong to the 11 strains, the fragment is assigned as unknown.

homogenization mechanism should take place frequently to keep in a chromosome the two copies of the TIR identical. Figure 3 depicts the recombination events required to homogenize the chromosomal ends. This mechanism is reminiscent of the Break Induced Replication (BIR), which rescue the broken chromosome by recopying the intact arm until the end (including the telomere). This is the mechanism likely operating between the TIRs and keeping the TIR sequence identical. It also participates to the shortening or increasing of the TIR size variability; hence, if the broken point is located upstream or downstream the TIR border, then the size of the TIRs is increasing or decreasing respectively (Fig. 3A depicts a TIR increase). It was shown to be a powerful mechanism generating a high variability of the TIR sizes in lab conditions in *S. ambofaciens*. When an insertion takes place in one duplicated region (Fig. 3B), the same mechanism leads to conversion of the original TIRs.

Given these results, the presence of TIRs at the ends of the *Streptomyces* chromosome appears as a consequence of terminal recombination activity. Reciprocally, their presence could help rescuing double-strand breaks occurring in the terminal part of the chromosome by providing an intact substrate for recombination repair (34). Furthermore, terminal duplication can have functional consequences such as expression of specific gene function (*e.g.* specialized metabolite biosynthetic genes, (35) or could help maintaining a terminal cohesive structure such as that called the 'racket-frame' structure (36).

Identification of homologous recombination in the TIRs

We inferred homologous recombination (HR) events by scanning the aligned sequences of the TIRs of the *Streptomyces* population with the RDP4 program. In total for the 11 genomes, 45 unique events were detected (Fig. 4). Strains of a same sub-clade mostly share the same HR events, when other strains exhibited ones that are more specific. It is the case for RLB3-17, S1A1-7 and S1D4-23 that altogether represented 30 out of 45 unique HR events. These results convey the evolutionary history of the population. Common HR events within a sub-clade probably occurred in a common recent ancestor and spread vertically in these strains, when other strains could have cumulated more recombination

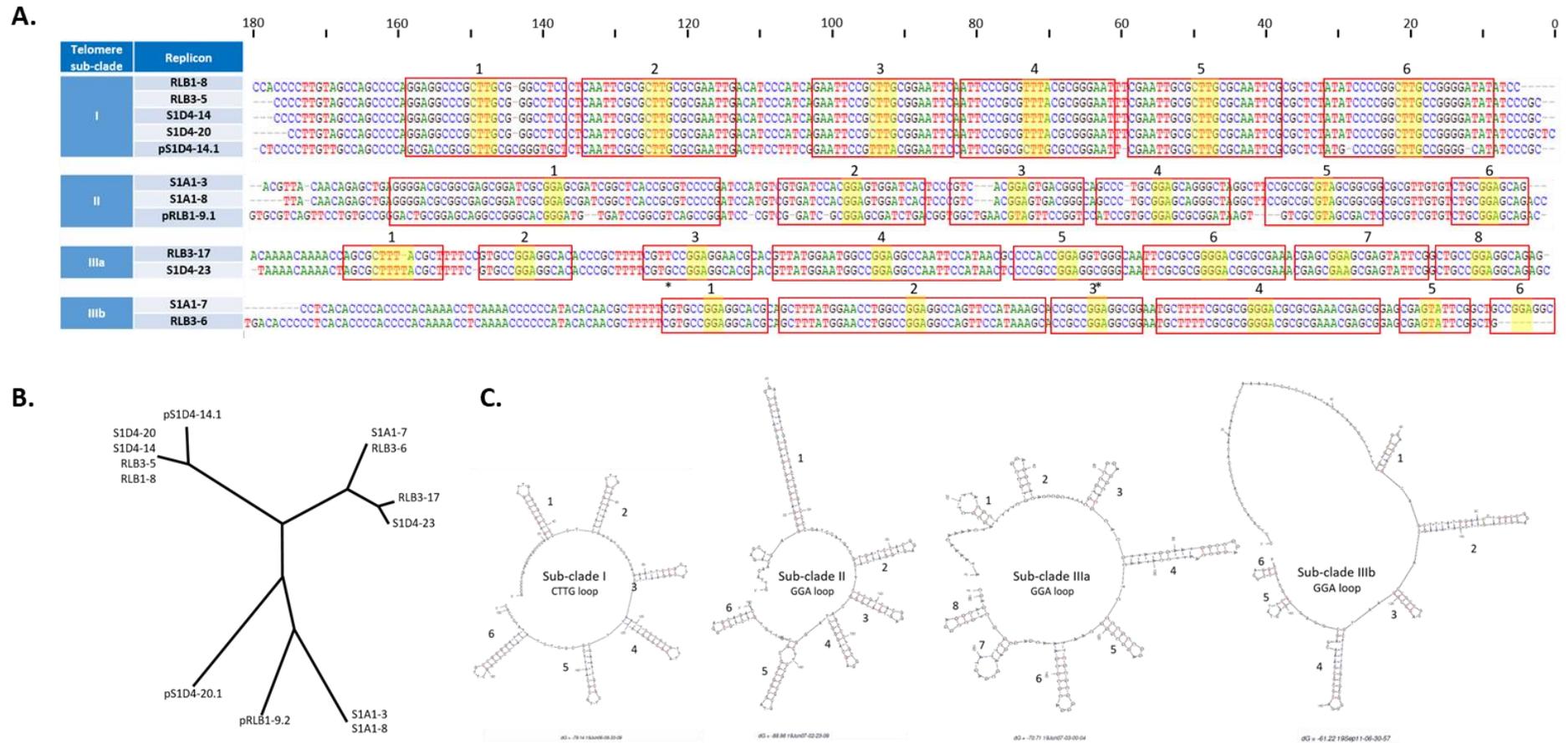


Fig. 5 Comparison of the different chromosomal and plasmidic telomeres. **A.** The terminal 180 nucleotides sequences from the 10 chromosomes and 2 linear plasmids were aligned. Four different groups designed in the table as telomere sub-clades were defined according to their sequence identities. Palindromic sequences are boxed and numbered in arabic numerals above the sequences and the telomere loops are highlighted in yellow. Within each palindrome, asymmetrical nucleotide pairs are identified with asterisks. **B.** Unrooted NJ phylogenetic tree built with the different telomere sequences. Positions with < 80% site coverage in the alignment were eliminated enabling to have a total of 164 nucleotide positions in the final dataset. Bootstrap percentages are indicated on the branches. **C.** The predicted secondary structures for a representative sequence of each telomere sub-cluster are represented. Two different loops (CTTG and GGA) can be observed at the top of the hairpins structures.

events since the origin of the population. *Streptomyces* were already shown to be recombinogenic, either at the genus (Cheng 2016) or at the population levels (Doroghazi). These former studies were performed in calculating recombination frequencies ($\rho/\theta\omega$ ratio) with seven housekeeping genes (3910 bp) scattered in the core genome. Here, because our strains are highly similar, it enabled to visualize HR events between their collinear TIRs (circa 408.938 bp). RDP suggests when possible the most probable donor of a recombining DNA sequence, thus here the potential donor within our population. For instance, in one of the most extreme cases, the TIR of RLB3-17 seemed to have recombined several times with strains S1A1-7, S1D4-23, RLB3-6 and strains of sub-cluster I. This almost mosaic structure confirmed that terminal regions are highly recombinogenic. It also highlights the massive gene flux previously described at the population level (23) and that our population strains have experienced, altogether, many gene transfers.

Telomere switching

Considering the high frequency of insertion and deletion events in the TIRs, we questioned the variability of the DNA extremities themselves, *i.e.*, the telomere motifs, within our population. While insertions/deletions in TIRs require at least two recombination events to take place, the replacement of the most distal regions may take a single cross-over event. This terminal exchange result in the formation of a hybrid chromosome (*i.e.* with two different telomeres) further homogenized by interchromosomal arm recombination as depicted on Fig. 3C.

The 180 last nucleotides of the sequence of each replicon, that are assumed to include telomere regions, were aligned and folded (DNA fold) for their *in silico* detection (Fig. 5). Telomeric sequences were identified for all chromosomes at the exception of RLB1-9 for which the telomeres were absent from the assembly. Regarding the other ten strains, four different telomere sequences were identified (Fig. 5A) which from 5 to 8 typical palindromic stems of variable length capped with conserved loop sequences. Two telomere loop sequences that are crucial for the Tap recognition and specificity were identified among the different strains: sub-clade I strains possess a $5'CTTG^3'$ loop, while the others have

a 5'GGA^{3'} sequence. Within the phylogenomic sub-clades I and II, respective strains shared identical telomere sequences (Fig. 5B) while their sequence identities fell to about 30% between sub-clades and were barely possible to align. In contrast, the telomeres of strains S1D4-23 and RLB3-6 forming the sub-clade III only shared a weak identity (65%), while they were more closely related to the telomere sequences of RLB3-17 and S1A1-7 respectively that do not belong to sub-clade III (Fig. 5C). For instance, S1D4-23 and RLB3-17 telomere sequences aligned almost perfectly (93% sequence identity) and exhibited only two mismatches that were compensatory mutations keeping the stem structures. Thus, telomere sequences defined two new telomere sub-divisions: IIIa with strains S1A1-7 and RLB3-6, and IIIb with strains RLB3-17 and S1D4-23 that were incongruent with the phylogenomic analysis (Fig. 5B). This data strongly supports the hypothesis of telomere exchange within the population, and especially that either of the two strains RLB3-6 and S1D4-23 acquired a new telomere, possibly from strain S1A1-7 or RLB3-17.

In addition to the chromosomes, telomere structures were identified for the three linear plasmids (pRLB1-9.1., pS1D4-20.1, pS1D4-14.1). The telomere of pRLB1-9.1 possesses 5'GGA^{3'} loops (5 of 8 stem-loops, all sharing the classical G-A sheared pairing), the one of pS1D4-14.1 a 5'CTTG^{3'} loop at the top of 5 of the 6 last stem-loops and the one of pS1D4-20.1 is typified by an original 5'GCA^{3'} loop sequence (at the top of the last 3 of the 5 stem-loops). The novelty of this telomere was confirmed by that fact that no identity would be found with any sequence of the nr database.

The different telomere sequences in the population suggest that various recombination events occurred during the recent evolutionary history of the population. Hence, telomeres of strains of sub-clade I are typified by 5'CTTG^{3'} loops when other strains harbor 5'GCA^{3'} ones. Interestingly, the telomere of pS1D4-14.1 shared a strong nucleotide identity (82% to 84%) and an identical 5'CTTG^{3'} loop at the top of 5 of the 6 last stem-loops with the telomere sequences of chromosomes of sub-clade I. The plasmid pS1D4-14.1 showed a strong identity (97%, $p < 5 \cdot 10^{-88}$) with telomeres of a plasmid (named 'unnamed 1') of *Streptomyces dengpaensis* strain XZHG99 (GenBank accession number CP026653.1). The alignment between the complete sequences of this plasmid (72 kb) and pS1D4-20.1 (112 kb)

revealed a conserved organization over several tenths of kilobases, interrupted by long insertions and deletions. It suggests a common origin to both plasmids and that pS1D4-20.1 might have been acquired by lateral transfer. According to the strain phylogeny (Fig. 1), a chromosome/plasmid replacement of the ancestral telomere loop $5'GGA^3'$ at the root of sub-clade I could explain the emergence of this telomere in the population.

Although terminal recombination appear highly efficient to homogenise the terminal sequences and eliminate hybrid replicons, their presence was reported previously. Hence, in *S. coelicolor* A3(2), both the chromosome (7.2-Mb) and a SCP1' linear plasmid (1.85 Mb) are chimeric generated by a single crossover between the wild-type chromosome and SCP1(37). Similarly, in *S. cattleya* NRLL 8057, the linear chromosome and a megaplasmid appear to have exchanged telomere leading to coexisting hybrid replicons (38). Telomere plasticity seems to be also the rule in *Borrelia* (spirochetes), the other main bacterial groups(38) possessing linear replicon (39). This may result from telomere exchange as well as from telomere fusion, which result from reversal of the telomere resolution reaction at the end of the replication process (40). At the functional level, telomere recombination triggered by the formation of double strand breaks was associated to antigenic variation in *Trypanosoma brucei* (41) .

Co-occurrence of telomere and terminal protein genes

Since terminal proteins (TP) interact in a specific way with the telomere in order to achieve terminal replication (21), the turnover of telomeres should be accompanied by that of the cognate terminal protein machineries. Therefore, we searched for homologues of the classical Tap/Tpg genes described in *Streptomyces coelicolor* A3(2) (20) as well as of GtpB-GtpA of *Streptomyces griseus* (19) in the chromosomes and plasmids of our population. We found that all the strains of the population possess a chromosomal homologue of the GtpB-GtpA encoding operon (c. 50% of amino acid identity with the *S. griseus* product). Among the population, the conservation is high with amino acid identities higher

Strain	TIR size (bp)	Phylogenomic sub-clade ^a	Telomere sub-clade ^a	Type I Tap-Tpg	Type II Tap-Tpg	GtpB-GtpA
S1D4-20	364 918*	I	I			
RLB3-5	364 918*					
RLB1-8	364 918*					
S1D4-14	364 918*					
S1A1-7	513 563	-	IIIb			
RLB1-9	303 829	II	II			
S1A1-3	393 282**					
S1A1-8	393 282**					
RLB3-17	446 660	-	IIIa			
S1D4-23	408 223	III	IIIb			
RLB3-6	579 809					

Plasmid		Host / sub-clade	Telomere sub-clade	Tap-Tpg
pS1D4-14.1	-	S1D4-1/I	I	
pS1D4-20.1	68 997	S1D4-20/I	other	
pRLB1-9.1	24 912	RLB1-9/II	II	-

Table 1. Size of the Terminal Inverted Repeats (TIRs) and distribution of the telomere types and terminal protein encoding genes on the different replicons.

^a Phylogenomic and telomere sub-clades defined in fig.1. The different Tap-Tpg and GtpB-GtpA operons are depicted with arrows. The type I Tap-Tpg and the GtpB-GtpA systems had two open reading frames (ORFs) and the type II Tap-Tpg had three. Within a model, similar colours of the ORFs between strains indicated a sequence identity superior at 98% aa. Differences in colour of plasmid ORFs in type I Tap-Tpg indicated that the sequences are related, but more distant with a sequence identity of 80 %.

– symbol indicated that no Tap-Tpg system was identified

*, **: the assessment uncertainty (circa 300 bp, i.e., a read length) did not allow to conclude to a size difference.

than 98% for both gene products. This operon was most probably inherited from the ancestor of the population (Fig. 1).

Using the *S. coelicolor* Tap and Tpg sequences, we identified and distinguished two additional types of operons (Table 1). Their distribution was variable and followed the sub-clade delimitation. Hence, sub-clade I strains possess a specific *tap-tpg* operon (called type I Tap-Tpg) encoding products with amino acid identities of 49% with Tap/Tpg of *S. coelicolor* A3 (2). Their sequences are highly conserved within the sub-clade with 100% aa identity. The two plasmids pS1D4-14.1 and pS1D4-20.1 belonging to individuals of the sub-clade I also harbour a *tap-tpg* operon. While that of pS1D4-14.1 clustered with type I Tap-Tpg with 82% of aa identity with the chromosomal homologues, that of pS1D4-20.1 showed only 46% of identity with the *S. coelicolor* A3(2) product and the chromosomal homologues. The presence of this original Tap-Tpg operon on the plasmid is co-occurring with the unique telomere sequence in our population having 5'GCA³ loops. It is likely that a specific interaction takes place between these Tap/Tpg proteins (type Ib) and the unique telomere.

Sub-clade II possesses a specific Tap-Tpg locus with a specific organization (called type II Tap-Tpg locus). Hence a truncated Tap alongside a full-length Tpg gene was found; however, an additional coding sequence just upstream of the truncated copy encodes a gene product including the Tap functional domains (pfam07926), suggesting that a Tap-Tpg locus could be rebuilt (the new Tap protein could complement the truncated one). Again, the sequence of the gene products is highly conserved with 100% within the sub-clade. In contrast to linear plasmids of sub-clade I (pS1D4-14.1 and pS1D4-20.1), pRLB1-9.1 which belonged to strain RLB1-9 (sub-clade II) do not encode any Tap or Tpg homologue, and should benefit from host functions (type II Tap-Tpg or GtpA-GtpB).

In addition to their clustering, the two Tap-Tpg operon types are located in genomic regions, which are specific of the genomic organisation of each genome. It is located in a terminal location, which belongs to the TIRs in sub-clade I while it is in a specific genome island in sub-clade II.

When the Tap/Tpg gene distribution is put in perspective with that of the type of telomere, it is possible to identify some correlation and to draw a scenario of telomere turnover within the natural population (Fig. 1). The presence of a type I Tap-Tpg locus is associated to the $5'CTTG^3'$ loop at the top of the stems of the telomere. This locus was identified in sub-clade I and on the plasmid pS1D4-14.1. Sub-clade II possesses a type II cluster while IIIa and IIIb possess only the 'ancestral' GtpA-GtpB cluster. Considering that the telomere sequences of the sub-clade I chromosomes and the telomeres of pS1D4-14.1 plasmid showing strong identities, it is tempting to speculate that a telomere replacement took place at the origin of this sub-clade and substituted the ancestral telomere (loop $5'GGA^3'$) by incoming plasmid-born one ($5'CTTG^3'$). Type II Tap-Tpg is found in sub-clade II and co-occur with telomere that typify this sub-clade. This result suggests that type II Tap-Tpg interacts with the specific sub-clade II telomere. Consequently, in RLB1-9, the telomere should belong to this type, and this therefore reinforces the hypothesis that the genome sequence is incomplete, *i.e.*, do not include the telomere for unknown reasons.

In the sub-clade III, the inconsistency between the phylogenomic tree and the telomere-based one suggests itself that some terminal DNA exchange has occurred. All these strains share the same set of terminal proteins (ancestral GtpB-GtpA), and so this telomere variability should be tolerated to achieve terminal replication.

Acknowledgments

This work was supported by a grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'Avenir" program (ANR-11-LABX-0002-01, Lab of Excellence ARBRE).

Competing interest

The authors declare that they have no competing interests.

References

1. Colin Y, Nicolitch O, Turpault M-P, Uroz S. 2017. Mineral Types and Tree Species Determine the Functional and Taxonomic Structures of Forest Soil Bacterial Communities. *Appl Environ Microbiol* 83.
2. Adams AS, Jordan MS, Adams SM, Suen G, Goodwin LA, Davenport KW, Currie CR, Raffa KF. 2011. Cellulose-degrading bacteria associated with the invasive woodwasp *Sirex noctilio*. *ISME J* 5:1323–1331.
3. Baltz RH. 2019. Natural product drug discovery in the genomic era: realities, conjectures, misconceptions, and opportunities. *J Ind Microbiol Biotechnol* 46:281–299.
4. Lewin GR, Carlos C, Chevrette MG, Horn HA, McDonald BR, Stankey RJ, Fox BG, Currie CR. 2016. Evolution and Ecology of Actinobacteria and Their Bioenergy Applications. *Annu Rev Microbiol* 70:235–254.
5. Casjens S, Huang WM. 1993. Linear chromosomal physical and genetic map of *Borrelia burgdorferi*, the Lyme disease agent. *Mol Microbiol* 8:967–980.
6. Wood DW, Setubal JC, Kaul R, Monks DE, Kitajima JP, Okura VK, Zhou Y, Chen L, Wood GE, Almeida NF, Woo L, Chen Y, Paulsen IT, Eisen JA, Karp PD, Bovee D, Chapman P, Clendenning J, Deatherage G, Gillet W, Grant C, Kutayin T, Levy R, Li MJ, McClelland E, Palmieri A, Raymond C, Rouse G, Saenphimmachak C, Wu Z, Romero P, Gordon D, Zhang S, Yoo H, Tao Y, Biddle P, Jung M, Krespan W, Perry M, Gordon-Kamm B, Liao L, Kim S, Hendrick C, Zhao ZY, Dolan M, Chumley F, Tingey SV, Tomb JF, Gordon MP, Olson MV, Nester EW. 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science* 294:2317–2323.
7. McLeod MP, Warren RL, Hsiao WWL, Araki N, Myhre M, Fernandes C, Miyazawa D, Wong W, Lillquist AL, Wang D, Dosanjh M, Hara H, Petrescu A, Morin RD, Yang G, Stott JM, Schein JE, Shin H, Smailus D, Siddiqui AS, Marra MA, Jones SJM, Holt R, Brinkman FSL, Miyauchi K, Fukuda M, Davies JE, Mohn WW, Eltis LD. 2006. The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc Natl Acad Sci U S A* 103:15582–15587.
8. Weaver D, Karoonuthaisiri N, Tsai H-H, Huang C-H, Ho M-L, Gai S, Patel KG, Huang J, Cohen SN, Hopwood DA, Chen CW, Kao CM. 2004. Genome plasticity in *Streptomyces*: identification of 1 Mb TIRs in the *S. coelicolor* A3(2) chromosome. *Mol Microbiol* 51:1535–1550.
9. Wenner T, Roth V, Fischer G, Fourrier C, Aigle B, Decaris B, Leblond P. 2003. End-to-end fusion of linear deleted chromosomes initiates a cycle of genome instability in *Streptomyces ambofaciens*. *Mol Microbiol* 50:411–425.
10. Hopwood DA. 2006. Soil to genomics: the *Streptomyces* chromosome. *Annu Rev Genet* 40:1–23.
11. Chen W, He F, Zhang X, Chen Z, Wen Y, Li J. 2010. Chromosomal instability in *Streptomyces avermitilis*: major deletion in the central region and stable circularized chromosome. *BMC Microbiol* 10:198.
12. Choulet F, Gallois A, Aigle B, Mangenot S, Gerbaud C, Truong C, Francou F-X, Borges F, Fourrier C, Guérineau M, Decaris B, Barbe V, Pernodet J-L, Leblond P. 2006. Intraspecific variability of the terminal inverted repeats of the linear chromosome of *Streptomyces ambofaciens*. *J Bacteriol* 188:6599–6610.

13. Yang C-C, Tseng S-M, Pan H-Y, Huang C-H, Chen CW. 2017. Telomere associated primase Tap repairs truncated telomeres of *Streptomyces*. *Nucleic Acids Res* 45:5838–5849.
14. Chang PC, Cohen SN. 1994. Bidirectional replication from an internal origin in a linear *Streptomyces* plasmid. *Science* 265:952–954.
15. Chen CW, Yu TW, Lin YS, Kieser HM, Hopwood DA. 1993. The conjugative plasmid SLP2 of *Streptomyces lividans* is a 50 kb linear molecule. *Mol Microbiol* 7:925–932.
16. Lin YS, Kieser HM, Hopwood DA, Chen CW. 1993. The chromosomal DNA of *Streptomyces lividans* 66 is linear. *Mol Microbiol* 10:923–933.
17. Huang CH, Lin YS, Yang YL, Huang SW, Chen CW. 1998. The telomeres of *Streptomyces* chromosomes contain conserved palindromic sequences with potential to form complex secondary structures. *Mol Microbiol* 28:905–916.
18. Ohnishi Y, Ishikawa J, Hara H, Suzuki H, Ikenoya M, Ikeda H, Yamashita A, Hattori M, Horinouchi S. 2008. Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J Bacteriol* 190:4050–4060.
19. Suzuki H, Marushima K, Ohnishi Y, Horinouchi S. 2008. A novel pair of terminal protein and telomere-associated protein for replication of the linear chromosome of *Streptomyces griseus* IFO13350. *Biosci Biotechnol Biochem* 72:2973–2980.
20. Yang C-C, Huang C-H, Li C-Y, Tsay Y-G, Lee S-C, Chen CW. 2002. The terminal proteins of linear *Streptomyces* chromosomes and plasmids: a novel class of replication priming proteins. *Mol Microbiol* 43:297–305.
21. Bao K, Cohen SN. 2001. Terminal proteins essential for the replication of linear plasmids and chromosomes in *Streptomyces*. *Genes Dev* 15:1518–1527.
22. Bao K, Cohen SN. 2003. Recruitment of terminal protein to the ends of *Streptomyces* linear plasmids and chromosomes by a novel telomere-binding protein essential for linear DNA replication. *Genes Dev* 17:774–785.
23. Tidjani A-R, Lorenzi J-N, Toussaint M, van Dijk E, Naquin D, Lespinet O, Bontemps C, Leblond P. 2019. Massive Gene Flux Drives Genome Diversity between Sympatric *Streptomyces* Conspecifics. *mBio* 10.
24. Tidjani A-R, Lorenzi J-N, Toussaint M, van Dijk E, Naquin D, Lespinet O, Bontemps C, Leblond P. 2019. Genome Sequences of 11 Conspecific *Streptomyces* sp. Strains. *Microbiol Resour Announc* 8.
25. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75.
26. Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406–3415.
27. Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.

28. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* 35:1547–1549.
29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410.
30. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403.
31. Martin DP, Murrell B, Khoosal A, Muhire B. 2017. Detecting and Analyzing Genetic Recombination Using RDP4. *Methods Mol Biol Clifton NJ* 1525:433–460.
32. González-Torres P, Rodríguez-Mateos F, Antón J, Gabaldón T. 2019. Impact of Homologous Recombination on the Evolution of Prokaryotic Core Genomes. *mBio* 10.
33. Skovgaard O, Bak M, Løbner-Olesen A, Tommerup N. 2011. Genome-wide detection of chromosomal rearrangements, indels, and mutations in circular chromosomes by short read sequencing. *Genome Res* 21:1388–1393.
34. Hoff G, Bertrand C, Zhang L, Piotrowski E, Chipot L, Bontemps C, Confalonieri F, McGovern S, Lecointe F, Thibessard A, Leblond P. 2016. Multiple and Variable NHEJ-Like Genes Are Involved in Resistance to DNA Damage in *Streptomyces ambofaciens*. *Front Microbiol* 7:1901.
35. Pang X, Aigle B, Girardet J-M, Mangenot S, Pernodet J-L, Decaris B, Leblond P. 2004. Functional angucycline-like antibiotic gene cluster in the terminal inverted repeats of the *Streptomyces ambofaciens* linear chromosome. *Antimicrob Agents Chemother* 48:575–588.
36. Kinashi H, Shimaji-Murayama M, Hanafusa T. 1992. Integration of SCP1, a giant linear plasmid, into the *Streptomyces coelicolor* chromosome. *Gene* 115:35–41.
37. Yamasaki M, Kinashi H. 2004. Two chimeric chromosomes of *Streptomyces coelicolor* A3(2) generated by single crossover of the wild-type chromosome and linear plasmid scp1. *J Bacteriol* 186:6553–6559.
38. Barbe V, Bouzon M, Mangenot S, Badet B, Poulain J, Segurens B, Vallenet D, Marlière P, Weissenbach J. 2011. Complete genome sequence of *Streptomyces cattleya* NRRL 8057, a producer of antibiotics and fluorometabolites. *J Bacteriol* 193:5055–5056.
39. Huang WM, Robertson M, Aron J, Casjens S. 2004. Telomere exchange between linear replicons of *Borrelia burgdorferi*. *J Bacteriol* 186:4134–4141.
40. Chaconas G, Kobryn K. 2010. Structure, function, and evolution of linear replicons in *Borrelia*. *Annu Rev Microbiol* 64:185–202.
41. Li B. 2015. DNA double-strand breaks and telomeres play important roles in *trypanosoma brucei* antigenic variation. *Eukaryot Cell* 14:196–205.

CHAPITRE IV

IV. Etude de la diversité des éléments conjugatifs et intégratifs au sein d'une population de 11 souches sympatriques et conspécifiques de *Streptomyces*

La conjugaison chez les bactéries du genre *Streptomyces* se fait par l'intermédiaire de deux mécanismes de transfert distincts : l'un dépendant des protéines homologues du système de sécrétion de type IV et l'autre dépendant de la translocase TraB (**Figure A1**).

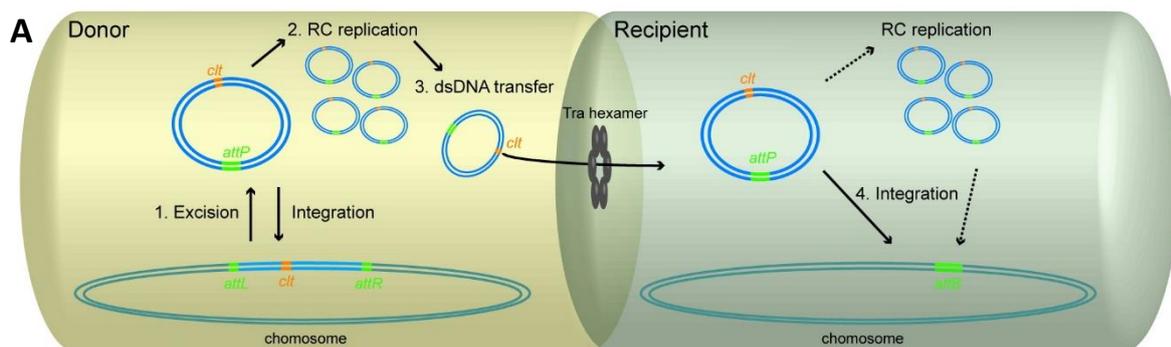
Le premier mécanisme de transfert conjugatif décrit chez les actinobactéries est celui des ICE (*integrative and conjugative elements*). Très répandue chez les bactéries à Gram négatif et chez les Firmicutes, la conjugaison des ICE est médiée par le système de sécrétion de type IV (T4SS). Ce dernier permet le transfert d'ADN simple brin et requiert l'assemblage d'un pore de conjugaison très complexe qui implique la sécrétion et l'assemblage d'un pilus extracellulaire (Fronzes *et al.*, 2009; Alvarez-Martinez & Christie, 2009). L'un des composants clé du T4SS est la sous-unité Virb4-like qui a une activité ATPase qui fournit de l'énergie pour l'assemblage et/ou le fonctionnement du pore de conjugaison. Le transfert de la molécule d'ADN via ce système est initié au niveau de l'origine de transfert (*oriT*), qui est reconnue par une relaxase d'ADN (protéine MOB) ainsi que d'autres protéines auxiliaires (**Figure A1.B**). Ce complexe nucléoprotéique ainsi formé, appelé relaxosome, sert de substrat au T4SS (Alvarez-Martinez & Christie, 2009; de la Cruz *et al.*, 2010; Smillie *et al.*, 2010). L'activité phosphodiesterase de la relaxase induit une coupure au niveau de l'*oriT* entraînant le déroulement de l'ADN et un transfert 5' vers 3' de la molécule simple brin à la cellule réceptrice. Les autres composants clés associés au T4SS sont la protéine de couplage (T4CP), qui fournit l'énergie nécessaire à la translocation de l'ADN simple brin, et la sous unité VirD4 qui sert de site d'accueil aux substrats du T4SS.

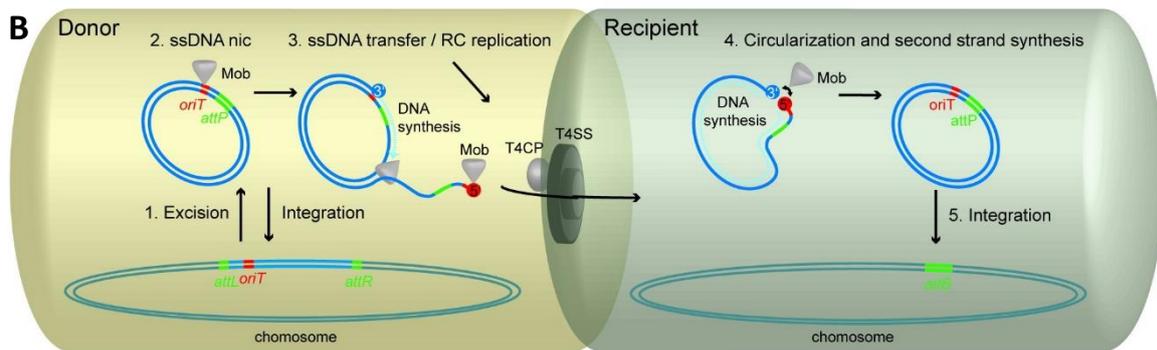
Le transfert conjugatif d'ADN double brin chez les actinobactéries incluant la bactérie filamenteuse du genre *Streptomyces* est un phénomène unique chez les procaryotes. A ce jour, de nombreux éléments génétiques utilisant ce mode de transfert ont été décrits chez cette bactérie ainsi que chez certains actinomycètes. Ces éléments incluent des plasmides linéaires (Kinashi *et al.*, 1987; Chen *et al.*, 1993), des éléments conjugatifs et intégratifs des actinomycètes (AICE) (Bibb *et al.*, 1981; Pernodet *et al.*, 1984; te Poele *et al.*, 2008), des plasmides circulaires à faible nombre de copie qui se répliquent en mode θ comme SCP2 (Bibb & Hopwood, 1981; Haug *et al.*, 2003) et les plasmides à grand nombre de copies qui utilisent un mode de réplication en cercle roulant comme pIJ101, pSN22, pJV1, pSG5, et pSVH1 (Kieser *et al.*, 1982; Kataoka *et al.*, 1991; Muth *et al.*, 1995; Servín-González *et al.*, 1995; Reuther *et al.*, 2006). En outre certains plasmides-phages ont été également décrits. Ces éléments possèdent un répertoire génétique de plasmide conjugatif leur permettant de se répliquer et de se transférer de façon autonome, en plus des gènes codant pour des protéines phagiques leur permettant de

s'empaquer dans des particules infectieuses (Chen *et al.*, 2012). Le transfert de ces différents éléments conjugatifs dépend de trois *loci* principaux (**Figure A1.A**). Le premier code pour la translocase d'ADN TraB (aussi appelée KilA pour 'kill', tuer), dont l'expression est létale en absence du régulateur KorA (aussi appelé TraR) (Kendall & Cohen, 1987; Hagège *et al.*, 1993). Le troisième élément est la séquence « *cis*-acting locus of transfert » (*clt*) (Pettis & Cohen, 1994; Servín-González, 1996) qui, reconnue par TraB, permet la mobilisation de l'élément génétique. Cette séquence *cis* ne serait pas indispensable à la mobilisation des marqueurs chromosomiques (Servín-González, 1996). En revanche, insérée dans un plasmide non conjugatif, elle permet sa mobilisation lorsque les protéines KorA et KilA sont fournies en *trans* (Fuchs & Muth (unpublished)).

Certains éléments conjugatifs comme le plasmide pZL1 (128 kb) de *Streptomyces* sp. 14R-10 comportent deux régions distinctes impliquées dans leur transfert. La première région code pour une translocase de la famille FtsK/SpoIIIE (TraB) et la seconde région code pour certains éléments du système de sécrétion de type IV (VirB6, VirB4, VirD4 et TraA). Il a été démontré que ces deux régions sont capables de médier le transfert de l'élément conjugatif l'une indépendamment de l'autre avec une efficacité égale (Zhao *et al.*, 2014).

Les AICE et ICE sont intégrés dans le chromosome hôte par un mécanisme de recombinaison site-spécifique. L'intégration site-spécifique est souvent médiée par une recombinase à tyrosine (plus occasionnellement à sérine) codée par le gène *int*. La recombinaison site-spécifique intervient entre deux courtes séquences identiques ou quasi-identiques dont l'une est située sur l'AICE (*attP*) et l'autre sur le chromosome de l'hôte (*attB*). Lors de l'induction de la conjugaison, Les AICE sont capables de s'exciser, de se répliquer de façon autonome sous forme circulaire (Sezonov *et al.*, 1995; Sezonov *et al.*, 1998), de se transférer et de s'intégrer dans le génome du nouvel hôte. La mutation de l'un des déterminants de la réplication (protéine Rep ou origine de réplication) abolit le transfert de l'élément conjugatif comme cela a été démontré chez l'AICE pSAM2 de *S. ambifaciens*. La réplication autonome est donc nécessaire au transfert conjugatif de l'AICE (Smokvina *et al.*, 1991; Hagège *et al.*, 1993a; Hagège *et al.*, 1994).





Adapté de Bordeleau et al., Mobile Genetic Elements 2:2, 119–124; March/April 2012

Figure A1 : Schéma des deux mécanismes de transfert des éléments conjugatifs AICE (A) et ICE (B) chez les actinomycètes. A : la recombinaison des sites *attL* et *attR* entraîne l'excision de l'AICE du chromosome, suivie d'une série de répllication en cercle roulant et d'une réintégration dans le chromosome de l'hôte ou du transfert à une cellule réceptrice par conjugaison via le pore formé par un hexamère de la protéine Tra qui reconnaît la séquence *clt* (*cis-acting locus of transfer*) de l'AICE. Une fois dans la cellule réceptrice, l'AICE circulaire s'intègre dans le chromosome de l'hôte directement ou après une série de répllications. **B :** suite à l'excision par recombinaison site-spécifique, la relaxase Mob reconnaît l'origine de transfert (*oriT*) où elle génère une cassure sur un brin et se fixe de façon covalente à l'extrémité 5' du brin coupé. Pendant que le complexe nucléoprotéique se déplace en réalisant une répllication en cercle roulant à partir de l'extrémité 3'OH libre, il interagit avec la protéine de couplage T4CP qui génère de l'énergie pour sa translocation à travers le système de sécrétion de type IV. Une fois dans la cellule réceptrice, la protéine Mob relie la molécule d'ADN simple brin suivie de la synthèse du brin complémentaire et de l'intégration dans le chromosome hôte par recombinaison sit-spécifique.

L'analyse de la diversité génomique dans le chapitre II a permis l'identification des évènements d'insertions et délétions (indels). Une grande partie de ces indels renferme des gènes codants pour des protéines signatures des éléments génétiques mobiles suggérant des acquisitions par transfert horizontal. Cependant, Les bactéries du genre *Streptomyces* ne semblent pas être naturellement transformables (Chater & Hopwood, 1984). En effet aucun homologue des gènes impliqués dans la compétence naturelle n'est prédit chez cette bactérie. Par ailleurs, une seule étude a rapporté la transduction généralisée chez *Streptomyces coelicolor* (Burke *et al.*, 2001) ; cette analyse n'a pourtant pu être reproduite car personne n'a pu avoir accès à ces phages transducteurs. En revanche de nombreux éléments conjugatifs ont été identifiés chez ces bactéries faisant de la conjugaison le mécanisme principal d'acquisition ou de transfert latéral de gènes chez les bactéries du genre *Streptomyces*. Nous avons pour cela entrepris dans cette étude d'analyser la prévalence et la diversité des éléments conjugatifs et intégratifs de type AICE (Actinomycete Integrative and Conjugative Element) ainsi que les éléments conjugatifs de type ICE (Integrative and Conjugative Element) dans les génomes de notre population issue du sol forestier.

IV.1 Stratégie d'identification des éléments conjugatifs et intégratifs

La première étape était d'identifier les éléments AICE et ICE dans les génomes. Pour valider la présence d'un élément conjugatif, nous avons recherché les gènes codant pour les fonctions d'excision (intégrase, excisionase), de conjugaison (VirB4, VirD4 et Mob pour les ICE et TraB pour le AICE) et de répllication (primase-polymérase, réplcicase). Les CDS prédites (annotation RAST) (Aziz *et al.*, 2008) ont

été analysées en utilisant, pour l'identification des ICE, le logiciel Conjscan-T4SSscan (<https://research.pasteur.fr/fr/software/conjscan-t4ssscan/>) développé à l'Institut Pasteur par Eduardo Rocha (Guglielmini *et al.*, 2014; Abby *et al.*, 2016) et pour l'identification des AICE, une combinaison de l'utilisation de BLASTP (Altschul *et al.*, 1990) pour la recherche des protéines canoniques et le visualiseur Artemis Comparison Tool (ACT)(Carver *et al.*, 2005) pour analyser la synténie entre les souches et identifier les sites d'insertion des divers éléments conjugatifs (**Figure A2**).

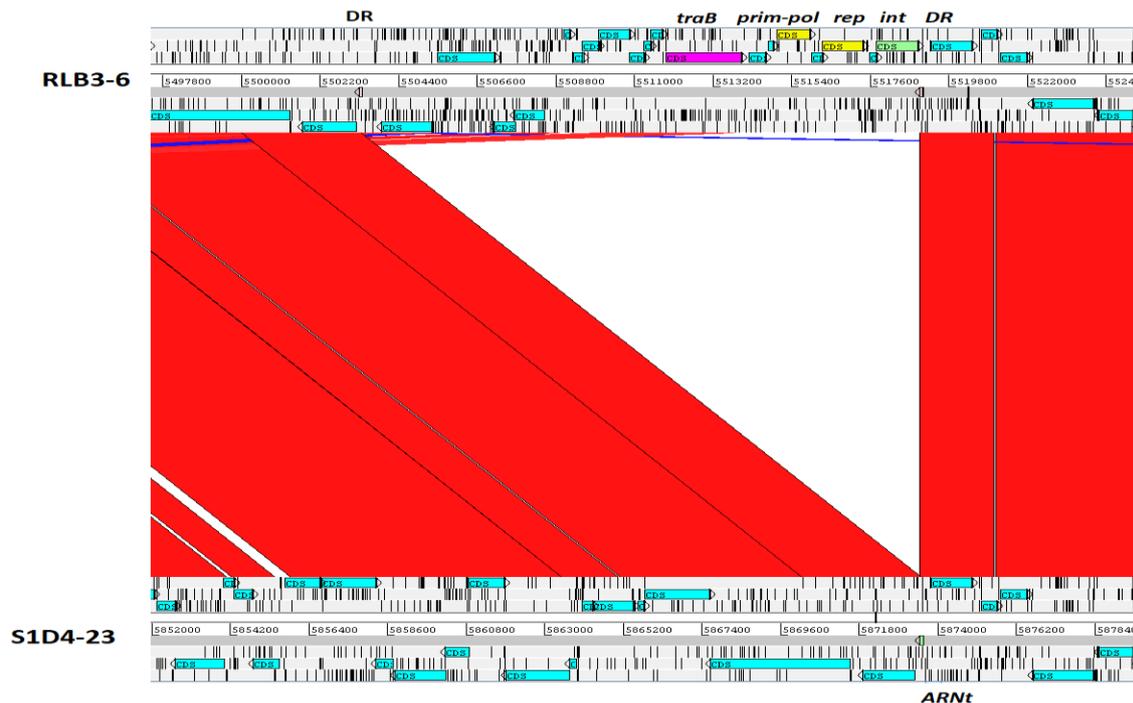


Figure A2 : Visualisation de l'insertion des éléments génétiques (AICE et ICE) par rupture de synténie en utilisant Artemis Comparison Tool (ACT)(Carver *et al.*, 2005). La capture d'écran présentée ci-dessus montre l'identité des séquences dans une fenêtre de 26 kb entre deux souches de la population. L'annotation (RAST) de chacun des génomes est présentée dans les 6 phases de lectures. Les boîtes rouges identifient les régions de fortes identité de séquences (supérieure à 99%). La souche RLB3-6 présente un îlot génomique de 15 kb codant potentiellement l'ensemble de protéines signature d'un AICE. La région est flanquée de répétitions directes qui correspondent à la duplication de la région 3' d'un gène codant un tRNA. Le site d'insertion est vide chez la souche S1D4-23.

Le logiciel Conjscan-T4SSscan permet, grâce à des modèles HMM spécialisés, de comparer les CDS extraites des génomes à une base de données contenant les protéines canoniques de la machinerie de sécrétion de type IV. La proximité géographique des différentes protéines nécessaires à la formation du T4SS permet de valider ou non la présence d'un élément conjugatif complet (Abby *et al.*, 2016). Pour l'identification des AICE, nous avons recherché les séquences codant les protéines TraB potentielles, puis recherché dans une fenêtre de 20 gènes en amont et aval la présence des autres fonctions signature (intégrase, excisionase, primase-polymérase et réplécase). Cela nous a permis d'éviter l'identification d'autres éléments mobiles (transposons, prophages ou plasmides intégrés). L'AICE est validé lorsqu'on a la présence simultanée d'un gène codant une intégrase (Int), une réplécase

(Rep) et une translocase (TraB). Pour les deux types d'éléments conjugatifs (AICE et ICE), l'identification des bornes constituées de répétitions directes (DR) est également indispensable à la validation finale de l'élément.

Diversité génomique au sein d'une population *Streptomyces*

Dans le chapitre II, nous avons utilisé la génomique comparative pour analyser la diversité génétique au sein de 11 souches conspécifiques de *Streptomyces*, entièrement séquencées, isolées à micro-échelle (Tidjani *et al.*, 2019b). Trois sous-clades, en phylogénomique, ont été identifiés au sein de cette population : le sous-clade I formé par les souches RLB1-8, RLB3-5, S1D4-14 et S1D4-20 ; le sous-clade II formé par les souches RLB1-9, S1A1-3 et S1A1-8 ; et le sous-clade III formé par les souches RLB3-6 et S1D4-23.

IV.2 Identification, prévalence et classification des AICE et ICE

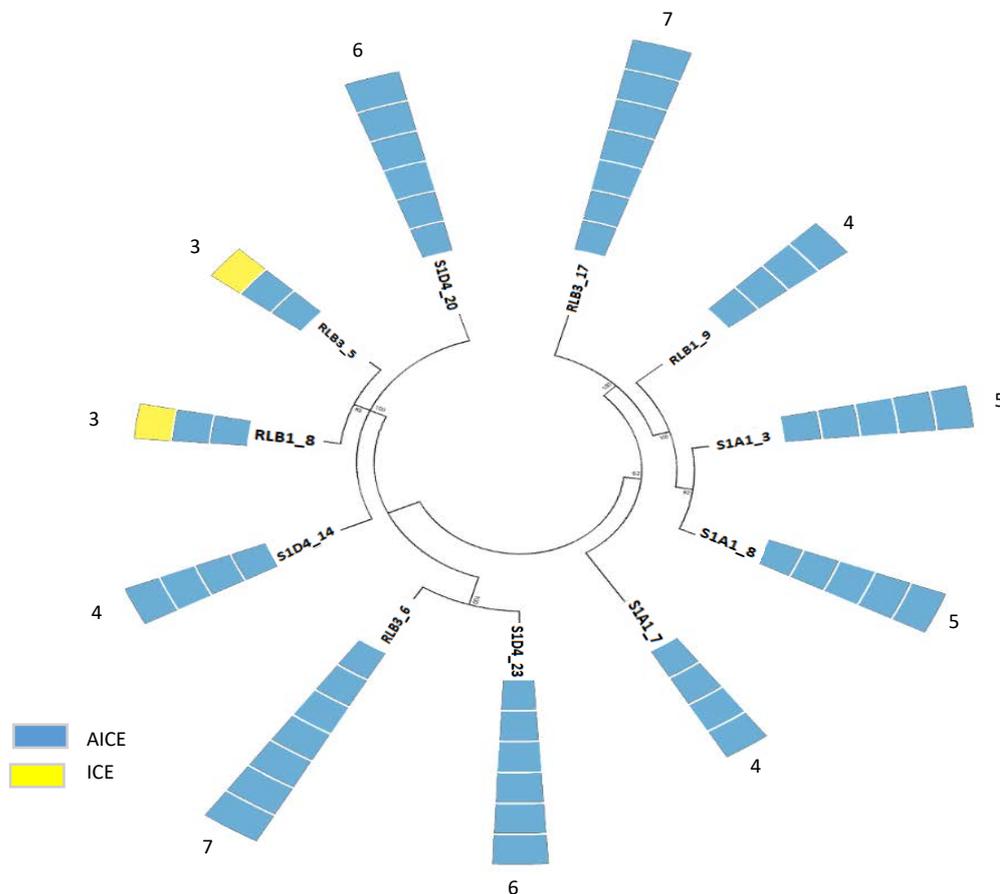


Figure A3 : Prévalence des éléments conjugatifs (AICE et ICE) au sein des 11 génomes séquencés. En bleu le nombre d'AICE par souche et en jaune celui des ICE. Le nombre total d'éléments conjugatifs pour chaque souche est indiqué au sommet de chaque diagramme. L'arbre phylogénomique a été réalisée avec 5 213 gènes (5 149 602 positions de nucléotide) du core génome de la population en utilisant *Streptomyces avermitilis* comme racine. L'arbre a été construit avec un maximum de vraisemblance avec un modèle GTR et en 100 répliques de bootstraps.

Tous les génomes des 11 souches analysées renferment au moins trois éléments conjugatifs. Le nombre d'AICE et ICE par souche est variable au sein de la population: de trois (2 AICE et 1 ICE) pour les souches RLB1-8 et RLB3-5 à sept AICE pour les souches RLB3-6 et RLB3-17 (Tableau 1). L'analyse a permis au total l'identification de 52 AICE et 2 ICE dans l'ensemble des 11 génomes (**Figure A3**). Seules deux souches, RLB1-8 et RLB3-5, renferment un ICE chacun qui s'avère être identique entre les 2 souches. La taille des éléments conjugatifs varie de 11 908 pb à 43 763 pb pour les AICE et est de 85 990 pb pour l'ICE. L'analyse et la comparaison des séquences nucléotidiques et du contenu génétique des éléments conjugatifs (AICE et ICE) a permis de les regrouper en 26 AICE différents et un seul ICE (1 seul élément identifié). Parmi les 26 AICE identifiés, 15 ne sont présents que dans une seule souche. Les 11 autres (AICE et ICE) sont partagés par au moins deux souches sans qu'aucun ne soit présent simultanément chez tous les individus de la population (**Figure A4 et A5**). Les éléments conjugatifs identifiés ne font pas partie du génome core de la population.

IV.3 Sites d'insertion des AICE et ICE

Le système d'intégration des AICE et ICE est analogue à celui des bactériophages lysogéniques (Boccard *et al.*, 1989). Les AICE et ICE s'intègrent de manière site spécifique et préférentiellement dans des sites chromosomiques conservés (comme les gènes de tRNA ou autres gènes conservés) (Ghinet *et al.*, 2011; Ambroset *et al.*, 2015). Ce site d'intégration situé vers l'extrémité 3' du gène de tRNA n'abolit pas le caractère fonctionnel de ce dernier en cas d'insertion ou d'excision (la séquence est reformée par la recombinaison des sites *att* en cas d'intégration ou d'excision). Cette affinité pour les sites conservés permettrait d'élargir le nombre d'hôtes potentiels dans lesquels ils pourraient s'intégrer (Sentchilo *et al.*, 2009). Certains AICE sont cependant capables de s'intégrer dans différents sites d'un même chromosome (Brown *et al.*, 1988a, b). Dans le but d'identifier les sites d'insertion des AICE et ICE, nous avons recherché et identifié les répétitions directes (DR) correspondant aux motifs *attL* et *attR* flanquant les éléments. Ces sites ont été déterminés par comparaison de deux génomes, l'un possédant l'élément inséré, l'autre non (**Figure A2**). Une recherche d'homologie entre les deux bornes de l'élément (correspondant à la rupture de synténie) permet d'identifier les DR, et donc également le site d'insertion dans le génome dépourvu de l'élément (tRNA, CDS, région intergénique).

Les ARNt représentent le site d'insertion privilégié de AICE et ICE identifiés. En effet, plus de 70% (38 sur 54) de ces éléments sont insérés dans un gène codant un ARNt. Il faut également noter l'insertion de certains AICE dans des gènes conservés comme le gène codant une méthylase de la grande sous-unité de l'ARN ribosomique (Tableau A.1).

La taille des DR varie de 2 pb à 66 pb. Les DR flanquant les éléments conjugatifs distincts mais s'insérant dans le même locus sont identiques en séquence mais pas en taille. C'est le cas des DR de 42 pb

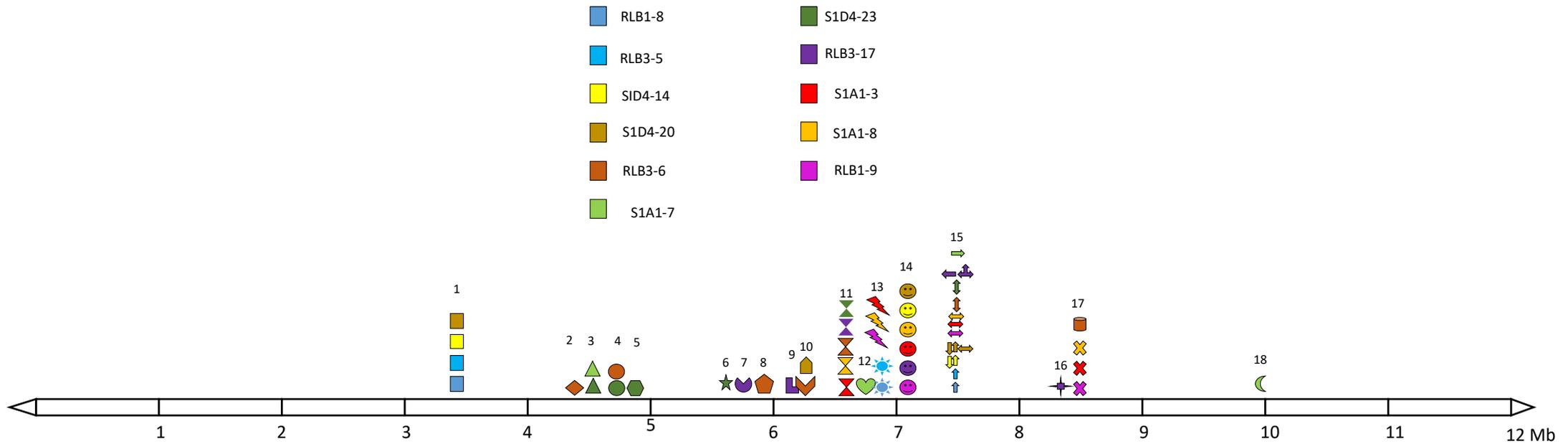


Figure A4 : Distribution des éléments conjugatifs (AICE et ICE) le long du chromosome. Le chromosome de référence est représenté avec une échelle de taille en méga base. Les différents types d'éléments conjugatifs sont représentés par les formes géométriques. Les couleurs distinguent chacune des 11 souches séquencées. Le numero au-dessus de chaque élément conjugatif représente le numéro du site d'insertion de l'élément.

flanquant l'ICE partagé par les souche RLB1-8 et RLB3-5 qui sont identiques en séquence à ceux des AICEs occupant la même position dans les souches RLB1-9, S1A1-3 et S1A1-8 qui ont juste une paire de base en sus en 5'. Ces deux types d'éléments ne possèdent cependant pas la même intégrase.

IV.4 Répartition chromosomique des positions d'insertion des AICE et ICE identifiés

Nous avons recherché la distribution des AICE et ICE le long des génomes des souches de notre population. Pour ce faire les AICE et ICE identifiés ont été reportés sur un chromosome de référence en utilisant la synténie existante entre les 11 chromosomes (**Figure A4**). Les AICE et ICE identifiés occupent 18 positions d'insertion uniques. La moitié de ces positions est propre et spécifique à une seule souche (ces positions ne sont pas retrouvées comme site d'insertion chez leurs consœurs) et les éléments conjugatifs qui y sont insérés sont également spécifiques (Ils ne sont pas retrouvés dans d'autres sites d'insertion). Il s'agit des positions d'insertion 2, 5, 6, 7, 8, 9, 12, 16 et 18. Cinq autres positions sur les 18 identifiés (1, 3, 4, 11, 14) sont en revanche retrouvées chez au moins deux souches et renferment à chaque fois un même AICE. Les 4 positions restantes (10, 13, 15, 17) sont retrouvées comme sites d'insertions chez différentes souches, en revanche elles ne contiennent pas forcément le même AICE. Par exemple, le site 15 qui est le seul site contenant au moins un élément conjugatif chez toutes les souches permet l'insertion de 7 AICE différents au total. Ces positions correspondraient à des points chauds d'insertion. Dans certaines souches, on observe également à la même position une insertion de plusieurs éléments conjugatifs en tandem. On parle ici d'accrétion, c'est-à-dire d'éléments qui s'accumulent au même locus en reconnaissant les mêmes séquences d'insertion (Pavlovic *et al.*, 2004). C'est le cas de la position d'insertion 15 dans la souche S1D4-20 qui renferme trois AICE insérés dans un gène codant un ARNt arginine. La séquence du gène codant l'ARNt arginine a précédemment été décrite comme site d'insertion en accrétion de deux AICE (Sav3708 et Sav3728) de *Streptomyces avermitilis* bien qu'ils codent pour des intégrases appartenant à deux différentes sous-familles (Ghinet *et al.*, 2011). La position 13 quant à elle contient un ICE dans les souches RLB1-8 et RLB3-5 alors qu'un AICE est présent à la même position chez les souches RLB1-9, S1A1-3 et S1A1-8. Cette position reste vide dans les autres souches ce qui démontre la mobilité de ces éléments.

Le chromosome de *Streptomyces* est compartimenté, c'est-à-dire possède une région centrale conservée au sein du genre et des régions variables renfermant des gènes non essentiels formant des 'bras chromosomiques' (Bentley *et al.*, 2002). Cette analyse montre que la totalité des AICE et ICE identifiés s'intègrent dans la partie centrale du chromosome. Les bras chromosomiques en sont dépourvus (**Figure A4**). Cela s'expliquerait par l'insertion préférentielle des AICE et ICE dans des sites chromosomiques conservés (comme les gènes de tRNA) (Ghinet *et al.*, 2011; Ambroset *et al.*, 2015) qui sont situés dans le cas des souches de *Streptomyces* dans la partie centrale du chromosome.

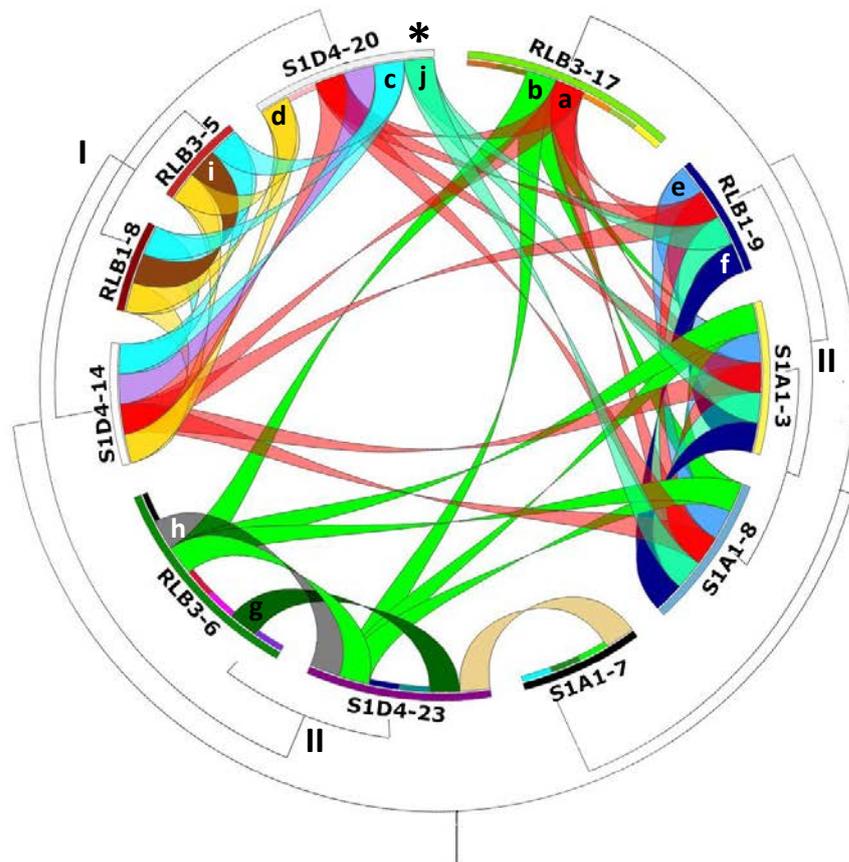


Figure A5 : Distribution des éléments conjuguatifs (AICE et ICE) au sein des 11 génomes séquencés. Le chromosome de chaque souche est représenté par un arc de cercle de différente couleur. Les différents éléments conjuguatifs identifiés ont été comparés les uns aux autres. Chaque élément conjuguatif est représenté par une portion de l'arc de cercle du chromosome dans lequel il est identifié. Les éléments identiques entre deux ou plusieurs souches sont reliés par un ruban de couleur identique et notés de (a) à (j) pour faciliter l'interprétation. Les éléments uniques à chaque souche sont représentés sans ruban. L'arbre phylogénomique a été réalisée avec 5 213 gènes (5 149 602 positions de nucléotide) du core génome de la population en utilisant *Streptomyces avermitilis* comme racine. L'arbre a été construit avec un maximum de vraisemblance avec un modèle GTR et en 100 répliques de *bootstraps*. Les sous-clades formés par les souches de la population sont indiqués par les chiffres romains I, II et III.

IV.5 Mobilité et transfert des éléments conjugatifs au sein de la population

Nous avons analysé la distribution des différents éléments conjugatifs caractérisés en terme de mobilité au sein de la population (**Figure A5**). Aucun AICE ou ICE n'est commun à toutes les souches. L'histoire des éléments conjugatifs ne suit donc pas une ascendance verticale au niveau de la population et peut être interprétée en terme de pertes et gains. L'analyse parcimonieuse de la distribution suggère que les éléments retrouvés dans peu de souches ou ceux spécifiques à une seule souche résulteraient d'une acquisition récente plutôt que d'une perte dans la majeure partie de la population. Réciproquement, les éléments qui seraient partagés par une majeure partie voire la totalité des individus reflèteraient une hérédité verticale ou une perte chez certains individus. La **figure A5** est un CIRCOS (Krzywinski *et al.*, 2009) qui est un outil de visualisation permettant de faciliter l'identification et l'analyse des similitudes et des différences résultant des comparaisons de génomes. Dans cette figure les rubans de même couleur permettent de relier les éléments conjugatifs identiques (notés de (a) à (j)) partagés par les différentes souches au sein de la population. Hormis deux AICE (a) et (b), aucun des éléments conjugatifs identifiés au sein de la population ne semble avoir été acquis par l'ancêtre commun récent et transféré de façon verticale. L'AICE (a) est présent dans le sous-clade II, dans la souche RLB3-17 ainsi que dans deux souches (S1D4-20 et S1D4-14) du sous-clade I. Cet AICE aurait donc été perdu par l'ancêtre commun récent du sous-clade III, par les souches RLB1-8, RLB3-5 du sous-clade I ainsi que par la souche S1A1-7. L'AICE (b) aurait subi des évènements similaires au précédent: acquis par l'ancêtre de la population, perdu par l'ancêtre commun du sous-clade I et les souches S1A1-7 et RLB1-9. Certains éléments, partagés par les souches appartenant au même sous-clade auraient été acquis par un transfert horizontal par l'ancêtre commun récent du sous-clade qui l'aurait ensuite transmis verticalement. Il s'agit des AICE (c) et (d) partagés par les souches du sous-clade I ; des AICE (e) et (f) partagés par les 3 souches du sous-clade II et des AICE (g) et (h) partagés par les deux souches du sous-clade III. Les éléments conjugatifs présents chez une partie des souches d'un même sous-clade comme l'ICE (i) auraient été acquis par l'ancêtre commun du sous-clade mais perdu par certaines souches (S1D4-14 et S1D4-20) au sein du sous-clade. La souche RLB3-17 est la plus riche en AICE souche-spécifiques avec 5 éléments conjugatifs uniques. Un événement de transfert horizontal potentiel entre les souches de la population est marqué par un astérisque. En effet AICE (j) que l'on retrouve dans toutes les souches du sous-clade II n'est présent que chez une souche (S1D4-20) du sous-clade I. Cet AICE acquis par l'ancêtre commun du clade II aurait été transmis verticalement à la descendance. Une des souches de cette descendance l'aurait transmis horizontalement (directement ou indirectement) à la souche S1D4-20 appartenant à un autre sous-clade.

L'analyse de la distribution révèle une grande plasticité et dynamique de ces éléments avec des histoires complexes au sein de la population sur un temps évolutif relativement court. Le fait qu'ils ne

soient pas partagés par toutes les souches lève aussi la question des barrières aux échanges. En effet, on constate qu'il existe des éléments partagés entre les sous-clades I et II, entre les sous-clades II et III mais pas entre les clades I et III. Deux hypothèses peuvent être émises ; soit ces souches n'ont pas été en contact dans leur environnement, soit il existe des barrières aux échanges. Au vue de l'échelle d'échantillonnage de la population, la seconde hypothèse serait plus probable. Un phénomène similaire a été décrit précédemment chez *Micrococcus xanthus* chez qui le transfert serait limité à des groupes ayant une compatibilité des protéines de surface. La base moléculaire de cette barrière serait associée à des toxines transmises lors de la fusion des membranes entre souches incompatibles (Wielgoss *et al.*, 2018).

L'analyse des éléments conjugatifs et intégratifs a permis de mettre en évidence une diversité et une prévalence assez importante de ces éléments au sein de la population. Ces éléments conjugatifs identifiés semblent fonctionnels (capables de s'exciser et de se transférer de manière autonome), car les gènes codant les protéines clés (intégrase et translocase) ainsi que les sites d'insertion (séquences *att*) ont été identifiés pour chaque élément. Tous les éléments conjugatifs identifiés sont situés dans les îlots génomiques impliquant qu'ils ne font pas partie du core génome ; ce qui suggère l'extrême mobilité de ces éléments au sein de la population. Nous avons proposé au vu de ces résultats que le mécanisme de conjugaison serait à l'origine de la forte diversité génétique observée au sein de la population. La tailles des indels identifiés au sein des génomes (pouvant atteindre 241 gènes) donne un aperçu de la quantité d'ADN pouvant être mobilisé par ce mécanisme.

Souche	Site	Nature de l'élément	Position sur le génome		Taille (pb)	Séquence du site Att (DR)	site d'insertion
			Début	Fin			
RLB1-8	1	Aice	3247920	3276331	28411	CACTGGTACTTCCAGTGGTTCACGGTCTGTTCTCTG	gène (Urea carboxylase-related amino acid permease)
	13	Ice	6667962	6753952	85990	CTCGTAATGCGTAGGTCTCGGGTTCGAATCCCGAAGGCGGCT	tRNA-Thr-CGT
	15	Aice	7909779	7930078	20299	CTCCGGAGCCGTGTGCGCAGGTTTGAATCCTGCCGGGGGCAC	tRNA-Arg-CCG
RLB1-9	13	Aice	6493066	6507621	14555	ACTCGTAATGCGTAGGTCTCGGGTTCGAATCCCGAAGGCGGCT	tRNA-Thr-CGT
	14	Aice	6832290	6847754	15464	CCTTCCAGGTTTGAATCCTGGCGCCGCCA	tRNA-Tyr-GTA
	15	Aice	7674234	7702743	28509	CTCCGGAGCCGTGTGCGCAGGTTTGAATCCTGCCGGGGGCAC	tRNA-Arg-CCG
	17	Aice	8116778	8139805	23027	GGGCCGTCCCTGGGCCGT	Espace intergenique
RLB3-5	1	Aice	2784770	2813181	28411	CACTGGTACTTCCAGTGGTTCACGGTCTGTTCTCTG	gène (Urea carboxylase-related amino acid permease)
	13	Ice	6204672	6290662	85990	CTCGTAATGCGTAGGTCTCGGGTTCGAATCCCGAAGGCGGCT	tRNA-Thr-CGT
	15	Aice	7446349	7466649	20300	GTGCCCCCGCAGGATTGAACTGCGCACACGGCTCCGAGG	tRNA-Arg-CCG
RLB3-6	2	Aice	3887605	3902832	15227	GCTGACGTG	en 5' du gene N-acetylmuramoyl-L-alanine amidase
	4	Aice	4331275	4356360	25085	CTGACGGTCAAGAACGCAGGTCCCACGCCG	en 3' du gene Fe-S oxidoreductase
	8	Aice	5503311	5519045	15734	CTGTAAAGCGAAGGTCGTCGGTTCGAATCCGACAGGGGGCTCC	tRNA-Thr-TGT
	10	Aice	5759377	5772160	12783	CTACGGATCAGAAGGTTGCAGGTTTGAATCCTGCCGAGTGCA	tRNA-Arg-ACG
	11	Aice	5788186	5803046	14860	CTACGGATCAGAAGGTTGCAGGTTTGAATCCTGCCGAGTGCA	tRNA-Arg-ACG
	15	Aice	7549115	7572470	23355	CCTCCGGAGCCGTGTGCGCAGGTTTGAATCCTGCCGGGGGCA	tRNA-Arg-CCG
	17	Aice	7984589	7996497	11908	CCCGGAGAACT	5' du gene Ribosomal RNA large subunit methyltransferase N
RLB3-17	7	Aice	5039506	5066784	27278	AG	putative integral membrane protein
	9	Aice	5626251	5639324	13073	CCTCGAAAGCGAGTGTGGCGCAAGTACCGAGGTTCAAATCCCTCCGCTACCG	tRNA-Ser-CGA
	11	Aice	5680101	5701081	20980	GGTGGGTTGCCGAGCGGCTAAGGGAACGGTCTTGAAAA	tRNA-Ser-TGA
	14	Aice	6614794	6630257	15463	CCTTCCAGGTTTGAATCCTGGCGCCGCCA	tRNA-Tyr-GTA
	15	Aice	7457875	7494116	36241	CCTCCGGAGCCGTGTGCGCAGGTTTGAATCCTGCCGGGGGCA	tRNA-Arg-CCG
	15	Aice	7494116	7512546	18430	CCTCCGGAGCCGTGTGCGCAGGTTTGAATCCTGCCGGGGGCA	
	16	Aice	7915981	7959744	43763	GGCGGCCAGGGA	3' du gène Mg chelatase-related protein;
S1A1-3	11	Aice	5889688	5906719	17031	GGTGGGTTGCCGAGCGGCTAAGGGAACGGTCTTGAAAA	tRNA-Ser-TGA
	13	Aice	6492923	6507478	14555	ACTCGTAATGCGTAGGTCTCGGGTTCGAATCCCGAAGGCGGCT	tRNA-Thr-CGT
	14	Aice	6832141	6847604	15463	CCTTCCAGGTTTGAATCCTGGCGCCGCCA	tRNA-Tyr-GTA

Souche	Site	Nature de l'élément	Position sur le génome		Taille (pb)	Séquence du site Att (DR)	site d'insertion
			Début	Fin			
	15	Aice	7673713	7703852	30139	CTCCGGAGCCGTGTGCGCAGGTTCAATCCTGCCGGGGGCA	tRNA-Arg-CCG
	17	Aice	8117749	8141018	23269	CTCCGGGCTTAGGCATGGCCATACCAGTGTGCGAGATCGAATTGCGTGACCTGGGCCGCGAGCCGG	5' du gene Ribosomal RNA large subunit methyltransferase N
S1A1-7	3	Aice	4276386	4293233	16847	CTTACAAGGCGGATGTGCGCGGTTTCAAACCGTCCGCGCCCACCA	tRNA-Val-TAC
	12	Aice	6567709	6601470	33761	TTCGGGACGAAGAGGTCGTGGTTCAAATCCCGCCACCCCGA	tRNA-Pro-CGG
	15	Aice	7828265	7851462	23197	CCTCCGGAGCCGTGTGCGCAGGTTCAATCCTGCCGGGGGCA	tRNA-Arg-CCG
	18	Aice	9744424	9759478	15054	GCAGATCCATGCTTCA	5' du gene ATP-dependent DNA ligase (EC 6.5.1.1) LigC
S1A1-8	11	Aice	5885243	5902274	17031	GGTGGGTTGCCGAGCGGCTAAGGGAACGGTCTTGAAAA	tRNA-Ser-TGA
	13	Aice	6488475	6503030	14555	ACTCGTAATGCGTAGGTCTCGGGTTCGAATCCCGAAGGCGGCT	tRNA-Thr-CGT
	14	Aice	6827692	6843155	15463	CCTTCCAGGTTTCAATCCTGGCGCCGCCA	tRNA-Tyr-GTA
	15	Aice	7669249	7699388	30139	CTCCGGAGCCGTGTGCGCAGGTTCAATCCTGCCGGGGGCA	tRNA-Arg-CCG
	17	Aice	8113165	8136192	23027	CTCCGGGCTTAGGCATGGCCATACCAGTGTGCGAGATCGAATTGCGTGACCTGGGCCGCGAGCCGG	5' du gene Ribosomal RNA large subunit methyltransferase N
S1D4-14	1	Aice	2866949	2895360	28411	CACTGGTACTTCCAGTGGTTCACGGTCTGTTCTCG	3' du gène (Urea carboxylase-related amino acid permease)
	14	Aice	6613202	6634074	20872	GACTGTAAATCTGCCGGCTCAGCCTTCCAGGTTTCAATCCTGGCGCCGCCA	tRNA-Tyr-GTA
	15	Aice	7461875	7481516	19641	CCTCCGGAGCCGTGTGCGCAGGTTCAATCCTGCCGGGGGCA	tRNA-Arg-CCG
	15	Aice	7481473	7501776	20303	CCTCCGGAGCCGTGTGCGCAGGTTCAATCCTGCCGGGGGCA	
S1D4-20	1	Aice	2872843	2901254	28411	CACTGGTACTTCCAGTGGTTCACGGTCTGTTCTCG	3' du gène (Urea carboxylase-related amino acid permease)
	10	Aice	5756086	5781632	25546	CACTGTAATGCGGTTTGGGCCTTAAAGCCATCGAGGTTCAAATCCCTCCGCTCCG	tRNA-Ser-GCT
	14	Aice	6705197	6726069	20872	GACTGTAAATCTGCCGGCTCAGCCTTCCAGGTTTCAATCCTGGCGCCGCCA	tRNA-Tyr-GTA
	15	Aice	7555357	7574997	19640	GGCCTCCGGAGCCGTGTGCGCAGGTTCAATCCTGCCGGGGGCA	
	15	Aice	7574997	7595255	20258	CCTCCGGAGCCGTGTGCGCAGGTTCAATCCTGCCGGGGGCA	tRNA-Arg-CCG
	15	Aice	7595214	7629563	34349	GGCCTCCGGAGCCGTGTGCGCAGGTTCAATCCTGCCGGGGGCA	
S1D4-23	3	Aice	4306378	4320007	13629	CTTACAAGGCGGATGTGCGCGGTTTCAAACCGTCCGCGCCCACCA	tRNA-Val-TAC
	4	Aice	4670476	4696042	25566	GCGTTCTTGACCGTCAGAGGCAGCAGTCTTCTGCC	en 3' du gène Fe-S oxidoreductase
	5	Aice	4813646	4825725	12079	CCAGGTACAGGCCCCCGGCCGAACGGGTTTCC	5' du gène hypothétique (ABC transporter)
	6	Aice	5489109	5511057	21948	TGCCCGTGTACAGGGCCAGAAGTAGATGAAGT	3' du gène phospholipid carrier-dependent glycosyltransferase
	11	Aice	6145900	6160760	14860	GGTGGGTTGCCGAGCGGCTAAGGGAACGGTCTTGAAAA	tRNA-Ser-TGA
	15	Aice	7908784	7932132	23348	CCTCCGGAGCCGTGTGCGCAGGTTCAATCCTGCCGGGGGCA	tRNA-Arg-CCG

Tableau A.1 : récapitulatif de l'identification des AICE et ICE et de leurs caractéristiques.

CHAPITRE V

V. Etude des mécanismes régissant l'évolution rapide des clusters de gènes de synthèse de métabolites spécialisés (CGSMS).

Les métabolites spécialisés sont de petites molécules organiques produites par les plantes, certains invertébrés et des microorganismes qui peuvent conférer un avantage sélectif à l'organisme qui les produit, mais qui ne sont pas essentielles à la croissance et à la reproduction (Hanson, 2003). Il existe une grande variété de métabolites spécialisés ayant des structures chimiques complexes et variées et de très nombreuses activités biologiques. Parmi ces dernières on peut citer la capacité à tuer ou à inhiber la croissance des bactéries (antibiotiques)(Sosio *et al.*, 2004), les champignons (antifongiques)(Brautaset *et al.*, 2000), des organismes eucaryotes telles que les protistes ou les helminthes (antiparasitaires) et la capacité à inhiber la multiplication des cellules cancéreuses (anticancéreux)(Tang *et al.*, 2000). Ces molécules peuvent également servir à la communication entre les cellules (*quorum sensing*), à la nutrition par la dégradation des composés complexes et à la promotion de la croissance des plantes par la production des phytohormones (Davies, 2013). Le rôle le plus connu joué par ces métabolites est sans concession celui d'antibiotique depuis la découverte de la pénicilline par Fleming en 1928 chez le champignon *Penicillium notatum*. L'âge d'or des antibiotiques commence à partir de 1944 avec la découverte de la streptomycine par Waksman chez *Streptomyces griseus* (Waksman *et al.*, 1946). Si l'utilisation des métabolites spécialisés par les Hommes pour le traitement de diverses pathologies reste la principale application connue, le rôle biologique à l'origine de sa synthèse chez les organismes les produisant demeure encore peu connu et exploré. Répondre à la question « quel est le rôle biologique des métabolites spécialisés ? » revient le plus souvent à entrer dans un domaine où la spéculation est facile et les vérifications très difficiles (Davies, 2013).

Il n'existe pas de schéma rigide de classification des métabolites spécialisés. Leur immense diversité structurale, fonctionnelle et de biosynthèse ne leur permet pas d'être classés dans quelques simples catégories (McMurry, 2015). Les enzymes nécessaires à la biosynthèse des métabolites spécialisés sont codées par des loci appelés clusters de gènes de synthèse de métabolites spécialisés (CGSMS). Les CGSMS sont classiquement organisés en groupes de 10 à 30 gènes qui coordonnent la biosynthèse des métabolites, leur régulation et l'expression des gènes de résistance (Zotchev, 2014). Traditionnellement, ces clusters de gènes incluent les synthèses des peptides non ribosomiaux (*non – ribosomal peptide synthase NRPS*), les synthèses des polycétides (*polyketide synthase PKS*) et les clusters de synthèse des RiPP (*ribosomally synthesized and post-translationally modified peptide*).

Les peptides non ribosomiaux (NRP) constituent l'une des principales familles de métabolites spécialisés. Ces peptides sont assemblés par les NRPS qui sont constitués d'un ou plusieurs modules chargés de l'incorporation des acides aminés dans la chaîne d'assemblage du peptide. Chaque module

est subdivisé en domaines. Un module typique de NRPS contient trois domaines essentiels: un domaine d'adénylation (A) responsable de la sélection et de l'activation des monomères, un domaine porteur de peptide (*peptidyl carrier protein* : PCP) également appelé domaine de thiolation (T), qui est nécessaire à l'attachement et au transfert des monomères adénylés et un domaine de condensation (C) qui catalyse la formation de liaisons amides et l'élongation de la chaîne peptidique (Challis & Naismith, 2004). Le module d'initiation ne possède pas de domaine C. En plus de ces 3 domaines classiques, il existe également des domaines dits accessoires, qui ne sont pas nécessairement présents, mais qui permettent la modification du peptide en cours de synthèse. Ce sont, par exemple, les domaines méthyltransférase (MT) ou épimérase (E). Enfin, le dernier module contient un domaine thioesthérase (Te) qui permet la libération du peptide (Finking & Marahiel, 2004). De nombreux métabolites sont synthétisés par la voie des NRPS. Parmi ceux-ci on peut citer le groupe ACV (L-aminoadipyl-L-cysteinyl-D-valine) qui est un tripeptide impliqué dans la synthèse des antibiotiques de la famille des beta-lactames, les antibiotiques vancomycines (glycopeptide) et daptomycine (lipopeptide).

Les polycétides, une autre famille de métabolites spécialisés sont synthétisés par les PKS. Ces dernières partagent des similitudes avec les synthèses des acides gras (Dutta *et al.*, 2014). Les PKS peuvent être divisés en trois types: Les PKS de type I qui contiennent de grandes enzymes modulaires successives renfermant des domaines multifonctionnels (l'érythromycine A, un antibiotique produite par *Saccharopolyspora erythraea* est produite par un PKS de type I); les PKS de type II constitués d'enzymes simples ou bifonctionnelles (la famille des tétracyclines est produite par ce type de PKS); et les PKS de type III qui utilisent des substrats liés à la co-enzyme A plutôt que le domaine porteur d'acyle (*Acyl Carrier Protein* : ACP) pour attacher et transférer des monomères comme les PKS de type I et de type II (Shen, 2003). Tout comme les NRPS, le PKS de type I classique contient également 3 domaines essentiels : un domaine acyltransférase (AT), qui permet le transfert du groupe malonyl de la malonyl-CoA vers le groupe 40-phosphopantetheineprosthétique d'un domaine ACP, et un domaine céto-synthase (KS) responsable de l'élongation de la chaîne de polycétides (Masschelein *et al.*, 2017). De plus, les similitudes entre les voies de biosynthèse des NRPS et des PKS conduisent à la synthèse des peptides hybrides NRP-polycétides, qui élargissent la diversité des métabolites spécialisés bactériens (Miller *et al.*, 2002; Liu *et al.*, 2004; Shou *et al.*, 2016).

Les RiPP représentent une famille de métabolites spécialisés considérée comme des agents antimicrobiens alternatifs prometteurs en raison de leur grande activité bactéricide (Lázár *et al.*, 2018). Les RiPP peuvent être regroupés en différentes sous-familles en fonction de leurs mécanismes de biosynthèse et de leurs caractéristiques structurelles. Parmi ces sous-familles, on distingue les lanthipeptides, les lassopeptides, les thiopeptides et les bactériocines (Ortega & van der Donk, 2016).

Au-delà de cette classification, presque toutes les RiPP contiennent un peptide leader N-terminal et un peptide cœur C-terminal (Ortega & van der Donk, 2016). De plus, contrairement aux deux autres familles de métabolites spécialisés (NRP et Polycétide), les RiPP suivent toujours une voie de biosynthèse simple dans laquelle le peptide leader guide la modification post-traductionnelle du précurseur du peptide et est éliminé par une ou plusieurs peptidases à l'étape finale de la maturation du peptide cœur (Letzel *et al.*, 2014; Ortega & van der Donk, 2016).

En évolution, un caractère est sélectionné s'il apporte un avantage sélectif à un organisme par rapport à ses compétiteurs. Les gènes du métabolisme spécialisé sont organisés en cluster occupant des dizaines de kilobases du génome des organismes producteurs. Un tel poids dans le génome des organismes doit être contrebalancé par les effets bénéfiques qu'il apporte, auquel cas il peut être éliminé ou perdu. Etudier l'évolution des gènes du métabolisme spécialisé revient à comprendre non pas comment évolue un gène, mais plutôt un groupe de gènes fonctionnant de façon collective. Les streptogamines qui sont formées par une paire de molécule A et B ayant une action synergique en sont un exemple. Prises séparément, ces molécules possèdent des activités bactériostatiques alors qu'ensemble elles sont bactéricides ((Di Giambattista *et al.*, 1989). Chez *S. pristinaespiralis*, les gènes de biosynthèse de streptogamines sont organisés en un cluster de 200 kb (Bamas-Jacques *et al.*, 1999) laissant penser que le regroupement de ces gènes aurait été sélectionné afin de renforcer leur activité antibiotique.

Les CGSMS sont considérés comme faisant partie d'une famille d'éléments génétiques parmi les plus diversifiés et subissant une évolution rapide (Fischbach *et al.*, 2008). Cette rapidité d'évolution serait due à leur structure modulaire permettant une recombinaison facile d'une part et aux avantages sélectifs et écologiques qu'ils confèrent leur permettant une sélection rapide d'autre part. Ainsi ces clusters fournissent une opportunité d'étudier des éléments génétiques qui évoluent pendant des échelles de temps très courts. Etudier l'évolution des CGSMS est également intéressant parce que leur phénotype (métabolites spécialisés) est mesurable. Les métabolites synthétisés par des clusters de biosynthèse peuvent être isolés, caractérisés et leur activité biologique étudiée fournissant ainsi un lien direct entre le génotype et le phénotype. En parallèle, l'amélioration des capacités à modifier de façon génétique les microorganismes a rendu possible l'élaboration de nouvelles molécules en modifiant les voies de biosynthèse de molécules déjà connues (Walsh, 2002). Pour bien réaliser ces modifications de façon efficace ou en construire de nouvelles à partir de rien, il est important de comprendre les règles qui régissent l'évolution des clusters de gènes de synthèse de métabolites spécialisés. De cette manière, il serait possible de les mimer dans le but de produire des molécules capables d'avoir une activité biologique (Arya *et al.*, 2002).

La modularité des NRPS et PKS les rend sujets à une évolution très rapide (Fischbach *et al.*, 2008). L'apparition de nouvelles molécules peut être le résultat des événements de mutation, de substitutions, de délétions et ou d'insertions survenant dans la séquence d'un CGSMS pouvant permettre ou non une meilleure adaptation de l'organisme qui la porte. Ces processus évolutifs présentent de nombreux défis car les CGSMS non fonctionnels ont une histoire évolutive très courte (Ochman & Moran, 2001; Lawrence *et al.*, 2001). Les événements de réarrangement d'ADN peuvent impliquer un seul gène dans le cluster ou l'ensemble des gènes modifiant ainsi le nombre ou l'identité des gènes formant le cluster; mais l'un des moyens de modification direct des gènes de biosynthèse passe par la mutation (Fischbach *et al.*, 2008). C'est le cas lorsque le métabolite à produire a pour précurseur un peptide synthétisé par voie ribosomale. Les patellamides produites par les bactéries du genre *Prochloron* sont synthétisées à partir d'un cluster de 7 gènes ; une région hypervariable (renfermant de nombreuses mutations) présente dans l'un des gènes entraîne la formation de plusieurs familles de la molécule (Donia *et al.*, 2006). Les mutations ponctuelles peuvent également affecter la structure des métabolites assemblés par la voie des NRPS et des PKS, ceci par des mutations affectant les domaines précis des modules responsables de l'assemblage des blocks formant la molécule finale (Wu *et al.*, 2000). Les gènes cœur de synthèse de métabolites spécialisés peuvent eux aussi subir des réarrangements entraînant le changement de position des domaines impliqués dans la synthèse du métabolite. C'est le cas des NRPS bacillomycine et mucosubtiline qui suite à un réarrangement de la position du domaine d'adénylation produisent des *building-blocks* enchaînés dans un ordre différent de la molécule native (Jenke-Kodama *et al.*, 2006).

Les CGSMS peuvent subir des modifications plus importants entraînant l'apparition de nouveaux gènes ou la suppression des gènes préexistants. Ces gènes peuvent jouer un rôle déterminant dans la modification post-assemblage du squelette de la molécule. C'est le cas des aminoglycosides qui conservent les gènes de biosynthèse de la partie 2-doexystreptamine alors que les gènes synthétisant les sucres branchés varient (Flatt & Mahmud, 2007).

L'évolution des métabolites spécialisés peut également aboutir à la duplication des clusters au sein des génomes. C'est le cas du cluster de synthèse de la kinamycine qui se retrouve dans les deux TIR de *S. ambifaciens*. Ces événements ont des conséquences qui sont beaucoup plus fonctionnelles. En effet, la présence de deux copies d'un même cluster rend possible l'accumulation des mutations dans l'un sans impacter l'apport du cluster natif à l'adaptation de l'organisme hôte (Fischbach *et al.*, 2008). Le cluster subissant les modifications peut alors poursuivre un processus évolutif propre indépendamment de la pression de sélection subit par le cluster intact. Cette divergence liée à la duplication pourrait conduire à la synthèse des molécules apparentées à la molécule native. C'est le cas des clusters de gènes à l'origine de la synthèse des lantibiotiques (Patton & van der Donk, 2005), qui

forment une grande classe de peptides ribosomiaux qui auraient un ancêtre commun (Fischbach *et al.*, 2008).

La transmission des gènes de synthèse de métabolites spécialisés se fait par transfert vertical, mais aussi par transfert horizontal (Fischbach *et al.*, 2008). Cette transmission se fait souvent à travers des éléments génétiques mobiles comme les ilots de pathogénicité (Hacker *et al.*, 1997) ou les plasmides (Szczepanowski *et al.*, 2005). De nombreux exemples de clusters de gènes acquis par transferts horizontaux ont été décrits chez les bactéries. On peut citer le cluster de biosynthèse du sidérophore yersiniabactine, initialement identifié chez l'agent responsable de la peste *Yersinia pestis* (Pelludat *et al.*, 1998), et qui est retrouvé chez de nombreuses espèces parfois éloignées. Ce cluster a été retrouvé chez le symbionte du nématode *Photorhabdus luminescens* (Duchaud *et al.*, 2003) et la bactérie marine *Salinispora tropica* (Udwary *et al.*, 2007) entre autre. Cette propagation horizontale des CGSMS permet de les retrouver chez de nombreuses espèces, ayant des niches écologiques différentes et variées.

Ces dernières années, la diminution drastique des coûts et l'amélioration des techniques de séquençage des génomes ont permis une augmentation sans cesse croissante du nombre de génomes bactériens dans les bases de données. C'est ce qui a permis de développer les stratégies de « *genome mining* » destinées à l'identification de nouveaux CGSMS.

Plusieurs études se sont intéressées à l'analyse de la diversité des CGSMS chez les bactéries. Doroghazi *et al.* (2014) dans leur étude portant sur 830 actinobactéries ont développé une méthode de classification de molécules en familles, basée sur le nombre de gènes homologues partagés, sur la proportion de nucléotides impliqués dans l'alignement et sur le pourcentage d'identité des acides aminés formant les domaines des différents modules. Cette approche a permis de regrouper les 11.422 clusters identifiés dans ces souches en 4.122 familles. Des études similaires ont été réalisées sur les bactéries du genre *Salinispora* et ont permis d'identifier 124 CGSMS (Ziemert *et al.*, 2014). Des études récentes portant sur la diversité des CGSMS au sein des souches bactériennes appartenant à la même espèce et échantillonnées dans divers environnements n'ont fait que confirmer la grande diversité qui avait été précédemment évoquée par les études antérieures (Antony-Babu *et al.*, 2017; Vicente *et al.*, 2018). Parmi celles-ci, l'étude portant sur 98 souches de *Myxococcus xanthus* a révélé que les génomes de ces dernières renfermaient 37 clusters de synthèse de métabolites uniques (Krug *et al.*, 2008). Les analyses bio-informatiques portant sur 6 souches de *Streptomyces albus* (Seipke, 2015) a déterminé 48 clusters uniques de synthèse de métabolites spécialisés dont 16 coderaient pour la production de métabolites qui sont souches spécifiques. Park & Andam (2019) ont également observé une variation de la distribution et de l'abondance des CGSMS au sein de 32 souches de *S. rimosus* collectées dans les bases de données, dont chacune porte un répertoire unique de clusters en plus de celui partagé.

Toutefois, si le profil métabolique reste presque identique entre les souches étroitement apparentées, les métabolites réellement produits peuvent différer d'une souche à l'autre. Au cours de sa thèse, Maxime toussaint avait testé les capacités d'inhibition de croissance des souches de la population vis-à-vis de plusieurs bactéries utilisées comme indicatrices. Ces souches ont démontré des profils d'inhibition différents les uns des autres. Le séquençage des génomes entiers des souches bactériennes étroitement apparentées et partageant un même habitat a permis de révéler l'existence d'une grande diversité ayant pour genèse la différence en contenu de gènes (Croucher *et al.*, 2014; Levade *et al.*, 2017; Tidjani *et al.*, 2019a). Cette différence pourrait impacter également les CGSMS. Les processus évolutifs et écologiques à la base de la diversité des CGSMS restent largement inexplorés. La diversité de ces clusters dans le sol, par exemple, est liée à des facteurs édaphiques et biotiques (Charlop-Powers *et al.*, 2014). Donc, pour mieux comprendre les mécanismes qui régissent la diversité et l'évolution des CGSMS, il apparaissait intéressant d'étudier cette diversité au sein des souches dérivant d'un ancêtre commun récent et formant une population cohésive (Shapiro & Polz, 2014). Cette dernière étant définie comme un ensemble de souches sympatriques de la même espèce. L'échantillonnage de ces souches partageant la même niche à petite échelle pourrait permettre de minimiser voire d'éliminer l'influence des facteurs d'adaptation à l'écosystème sur la diversité des CGSMS.

Diversité génomique au sein d'une population de *Streptomyces*

Dans ce chapitre, nous avons utilisé les génomes séquencés des souches de *Streptomyces* appartenant à une population (Tidjani *et al.*, 2019b). Pour rappel, ces 11 souches possèdent une séquence identique du gène codant pour l'ARNr 16S. L'analyse phylogénique de ces souches en utilisant le core génome a permis de confirmer les liens phylogénétiques étroits entre ces souches et le fait qu'elles ont divergé à partir d'un ancêtre commun récent (Tidjani *et al.*, 2019a). Trois sous-clades ont été néanmoins identifiés au sein de cette population : le sous-clade I formé par les souches RLB1-8, RLB3-5, S1D4-14 et S1D4-20 ; le sous-clade II formé par les souches RLB1-9, S1A1-3 et S1A1-8 ; et le sous-clade III formé par les souches RLB3-6 et S1D4-23 (**Figure B2**)

V.1 Identification et classification des clusters de synthèse de métabolites spécialisés

Nous avons utilisé antiSMASH version 5.0 (Blin *et al.*, 2019) pour identifier les clusters de synthèse de métabolites présents dans chaque souche de la population. Le pipeline antiSMASH a permis l'annotation de 383 clusters de synthèse de métabolites au sein des 11 souches de la population. Chaque génome contenait entre 34 et 37 CGSMS (**Figure B1**) représentant de 9,7 à 10,4% de la séquence génomique selon les souches (tableau S1). En fonction de leur identité nucléotidique et de leur contenu en gènes (analyse de la conservation des CDS prédits dans les clusters), ces clusters ont

été regroupés en 42 clusters uniques dont les tailles allaient de 7.625 pb à 120.895 pb. Après annotation et comparaison avec antiSMASH, 22 classes de métabolites ont été prédites (**Figure B1**). Près de la moitié (n = 20) des 42 CGSMS codent des PKS (regroupant 3 classes), des NRPS (regroupant 4 classes) ou des hybrides des deux. Les CGSMS les plus abondants dans la population sont ceux codant pour les NRPS et leurs analogues (n = 6 à 7 par génome), ainsi que les terpènes (n = 5 par génome). En utilisant un seuil de plus de 70 % des gènes du cluster présentant une similarité significative avec des gènes de référence de la base de données MIBiG (Minimum Information about a Biosynthetic Gene), 23,8% (n = 10) des clusters prédits dans notre population ont pu être plus précisément reliés à la production de molécules connues telles que le 2-méthylisobornéol (terpène), l'alkylrésorcinol (polycétide), l'ectoïne, la desferrioxamine B, l'albaflavenone (terpène), le pigment de spores (polycétide), le sapB (RiPP), le MS-271 (RiPP), la géosmine (terpène) et le hopène (terpène).

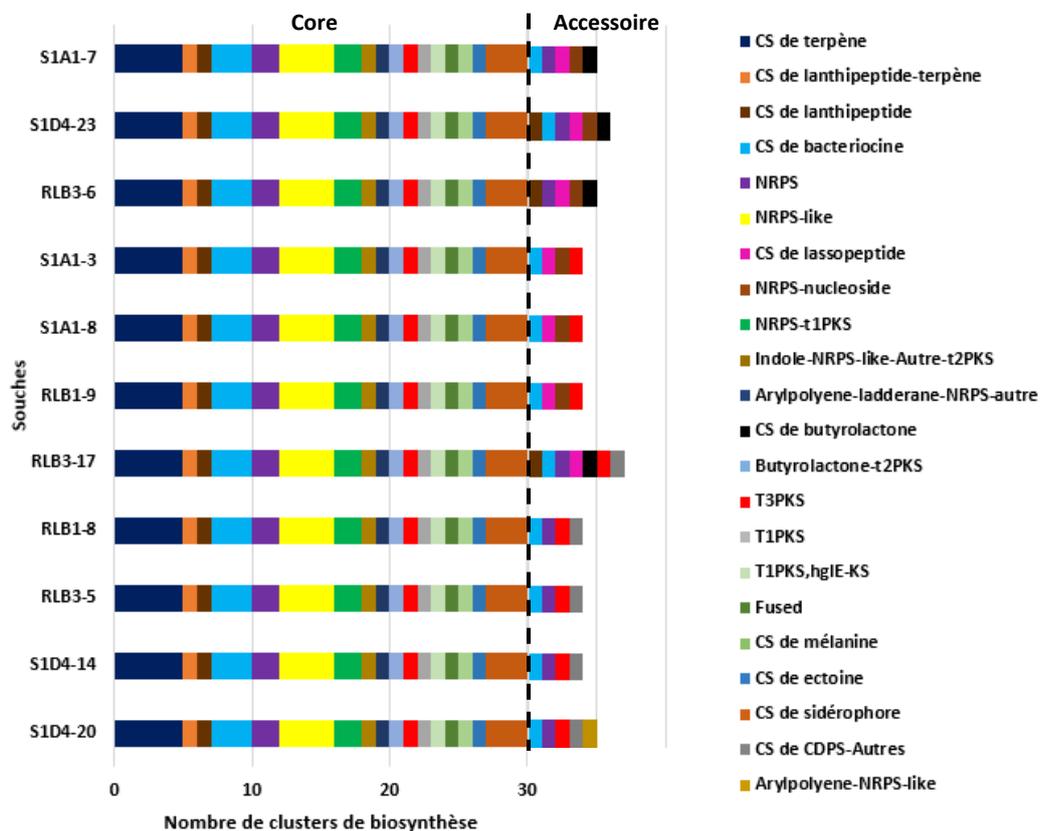


Figure B1 : Identification et classification des clusters de synthèse de métabolites spécialisés. 22 types de clusters de synthèse de métabolites spécialisés ont été prédits par l'outil antiSMASH. Les barres d'échelle représentent le nombre de clusters appartenant à chaque type de clusters identifiés par génome. Les couleurs illustrent le type du cluster de synthèse comme annoté dans la légende. Avec CS = « cluster de synthèse de ».

V.2 Clusters de métabolites spécialisés *core* et accessoires de la population/

Prédiction du pan-métabolome et évolution

Sur les 42 CGSMS uniques présents dans la population, 30 sont conservés au sein des 11 souches et représentent les CGSMS core (**Figure B1**), tandis que les 12 restants représentent les CGSMS

accessoires. L'ensemble de ces 42 CGSMS représentent les clusters codant pour la synthèse du pan-métabolome de la population. Les CGSMS core sont prédits pour coder des peptides non ribosomiaux (NRP), des polycétides, des hybrides de NRP-polycétides, des terpènes, de la mélanine, de l'ectoïne, des lanthipeptide-terpènes, des bactériocines et des sidérophores. Pour ces derniers, les analyses antiSMASH ont permis de prédire 3 clusters de biosynthèse de sidérophores par génome dont seule celui codant pour la synthèse de la desferrioxamine B est référencée dans la base de données MIBiG. Les classes de métabolites codés par les CGSMS core avaient précédemment été identifiés comme conservés chez le genre *Streptomyces* (Vicente *et al.*, 2018). Parmi les GCSMS accessoires figurent des NRP, des polycétides, des CDPS (*tRNA-dependent cyclodipeptide synthases*), des clusters de synthèse des lanthipeptides, des bactériocines, des lassopeptides et des butyrolactones (**Figure B1**). Certaines classes de métabolites comme les NRP, les polycétides et les bactériocines sont à la fois présentes dans les métabolites spécialisés core et accessoire. D'autres ne sont en revanche retrouvées uniquement que dans les clusters de métabolites spécialisés core de la population; il s'agit des clusters codant pour la synthèse des terpènes, de la mélanine, de l'ectoïne et des sidérophores. Ces derniers pourraient être considérés comme nécessaires à la survie des souches dans leur écosystème. D'autres sont enfin spécifiques aux clusters de métabolites spécialisés accessoires (NRPS-nucléoside, CDPS, cluster de synthèse des butyrolactones et des lassopeptide) (**Figure B1**).

Pour les CGSMS variables, 10 sont partagés par au moins deux souches de la population et 5 se sont révélés spécifiques à des sous-clades phylogénétiques formés par les souches de la population. Les deux derniers CGSMS variables sont souches spécifiques et sont retrouvés sur le chromosome de la souche S1D4-23 et sur un réplicon extrachromosomique de la souche S1D4-20. Ces CGSMS codent respectivement pour une bactériocine et un NRP-arylopène. Les souches du sous-clade I ne possèdent pas ce CGSMS (N° 43, **Figure B2**). Son obtention par la souche S1D4-20 pourrait donc être envisagée par transfert horizontal, tout comme pour S1D4-23 via l'acquisition de son réplicon extrachromosomique. La comparaison de la présence de CGSMS et de la phylogénie des souches de la population montre une congruence forte. Toutefois, un transfert d'un CGSMS codant un NRP (n°34 **Figure B2**) des souches du sous-clade I vers RLB3-17 pourrait également être envisagé. Enfin, si certains événements similaires pourraient également être inférés pour expliquer la présence des clusters 1, 5 et 31 (**Figure B2**) chez S1A1-7, ceux-ci ne peuvent être confirmés à cause du placement hasardeux de cette souche dans la phylogénie de la population.

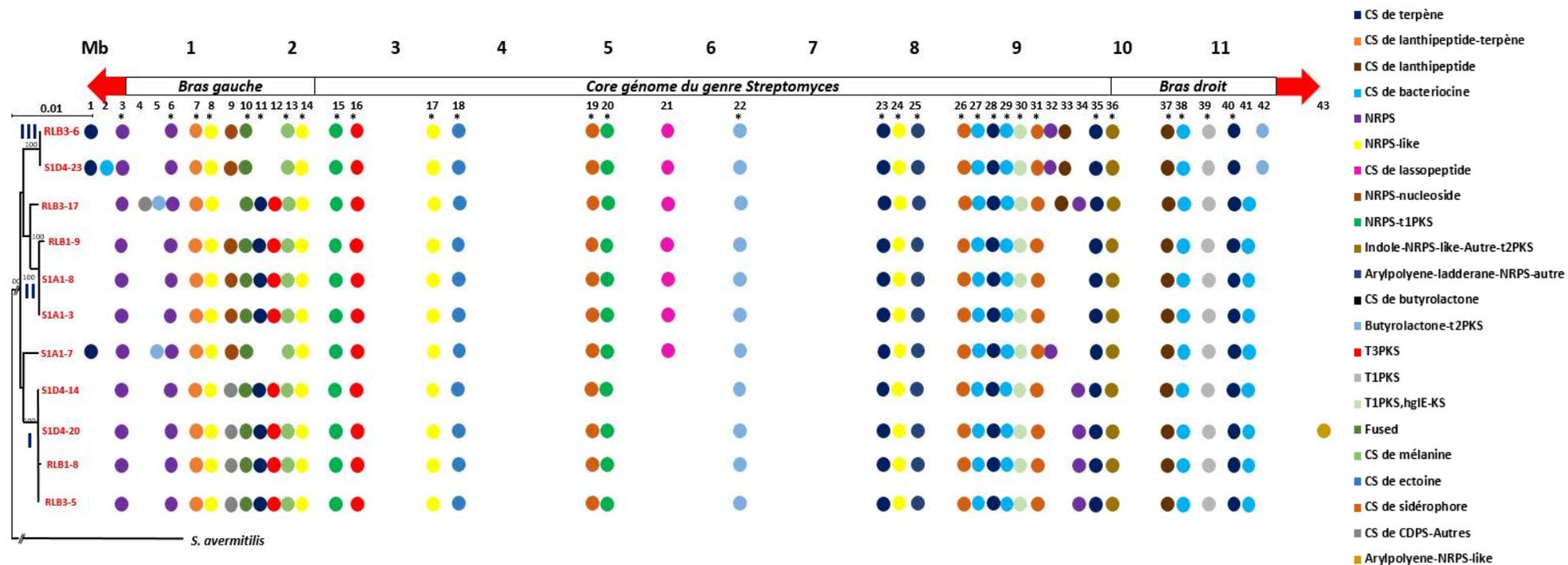


Figure B2 : Distribution génomique des clusters de synthèse de métabolites spécialisés le long du chromosome linéaire de *Streptomyces*. Le schéma représente le chromosome linéaire de *Streptomyces* avec l'échelle en mégabase. La position du core génome du genre *Streptomyces* est indiquée. Les répétitions terminales inversées (TIR) sont représentées par des flèches rouges. Les différentes boules de couleur représentent les clusters de synthèse de métabolites spécialisés prédits au niveau de chaque souche. Les couleurs illustrent le type de famille de molécule du cluster comme annoté dans la légende. Les souches sont ordonnées en fonction l'arbre phylogénétique présenté à gauche. Les chiffres romains (I, II et III) indiquent les sous clades formés par les souches au sein de la population.

V.1 Distribution génomique des clusters de synthèse de métabolites spécialisés

La comparaison génomique des souches de la population a permis de déterminer que la diversité génomique n'était pas distribuée de façon homogène le long des chromosomes. Cette diversité augmentait vers les régions terminales (Tidjani *et al.*, 2019a). Nous nous sommes alors demandé si cette diversité pouvait également impacter la distribution des clusters de synthèse de métabolites spécialisés le long des chromosomes. Pour ce faire, nous avons analysé et compilé la distribution des clusters de synthèse de métabolites le long d'un chromosome de référence (**Figure B2**). Les clusters de synthèse de métabolites spécialisés ne sont pas uniformément distribués le long du chromosome de référence. La densité de ces clusters augmente vers les régions terminales de même que la variabilité du nombre souche contenant chaque cluster (**Figure B2**). Les limites du core génome du genre *Streptomyces* (971 gènes, Lorenzi *et al.*, Non publié) au niveau du génome de référence ont été déterminés (en positionnant les deux gènes situés aux extrémités du core génome du genre sur le chromosome de référence de notre population) à 2,4 Mb du début du chromosome et à 1,4 Mb de la fin. De part et d'autre de ce core génome se situent les bras chromosomiques qui constituent les régions variables au sein du genre (Tidjani *et al.*, 2019a). Sur les 41 CGSMS chromosomiques, 22 (53,7%) sont situés dans le génome central et 19 (46,3%) dans les bras. Les bras représentent 3.8 Mb contre 7.9 Mb pour le génome central. Proportionnellement, cela implique que les bras concentrent environ deux fois plus de CGSMS que la région core avec respectivement 5 contre 2,8 CGSMS par Mb. Sur les 22 CGSMS situés dans le génome core, n=18 soit 80% appartiennent au CGSMS core; dans les bras seule n=12 sur les 19 CGSMS (63 %) appartiennent à ce dernier. Ces résultats permettent de conclure que la région centrale du chromosome est plus enrichie en CGSMS core par rapport aux bras et que ces derniers en revanche concentrent deux fois plus de CGSMS accessoires par rapport à la région centrale (36% contre 18%). Les résultats similaires ont été obtenus par Choulet *et al.* (2006a) en comparant la proportion des gènes appartenant au core génome à ceux appartenant au génome accessoire dans la partie centrale et dans les bras chromosomiques de plusieurs souches appartenant à différentes espèces de *Streptomyces*.

V.2 Dynamique évolutive des clusters de synthèse de métabolites spécialisés

Suite à l'observation de la grande diversité de présence/absence de CGSMS entre les différentes souches formant notre population, nous nous sommes intéressés aux mécanismes y ayant conduit. Les clusters de métabolites secondaire core n'ont pas été pris en compte dans cette analyse car ils reflètent l'histoire évolutive commune de la population et l'héritage vertical depuis un ancêtre commun de ces CGSMS. En revanche, les CGSMS accessoires reflètent des processus évolutifs (transferts, pertes, translocations) récents qui ont pu se produire dans une souche ou au niveau de l'ancêtre d'un sous-clade au sein de la population. La présence/absence des CGSMS a préalablement été établie par antiSMASH. Via son pipeline d'annotation antiSMASH vise à reconnaître des gènes clés au sein du génome annoté préalablement par RAST, nécessaires à la biosynthèse des molécules et étend ensuite le CGSMS par homologie des gènes environnants avec d'autres gènes connus, impliqués dans la synthèse de métabolites. Cette approche traduit une certaine réalité biologique car l'absence d'un gène de biosynthèse sera très probablement corrélée à l'absence de la production d'un métabolite. Toutefois, nous avons voulu aller plus loin dans ce type d'analyse afin de mieux comprendre les mécanismes évolutifs ayant conduit à la diversité de CGSMS prédits par antiSMASH. Pour ce faire, nous avons procédé par génomique comparative en utilisant l'outil *Artemis Comparison tool* et tenté d'identifier si les CGSMS prédits dans une partie des souches par antiSMASH étaient partiellement présents chez les autres souches prédites comme ne les possédant pas. Les gènes formant 12 CGSMS accessoires prédits ont ainsi été recherchés sur l'ensemble des souches.

Cette analyse a permis de déterminer que sur les 12 CGSMS accessoires 8 se retrouvaient dans des îlots génomiques. La désignation "îlot génomique" correspond aux événements d'insertions et/ou délétions de grande taille (plus de 10 CDS) entraînant une rupture de synténie lors de la comparaison de deux génomes. L'un de ces 8 CGSMS correspond à celui retrouvé sur le réplicon extrachromosomique de la souche S1D4-20. Les 7 autres sont localisés sur le chromosome et correspondent aux clusters de synthèse 2, 12, 32, 33, 34 et aux deux clusters (NRPS-nucléoside et CDPS) à la position 9. Les clusters 32, 33 et 34 et ceux situés à la position 9 sont spécifiques des sous-clades (sous-clade III pour les CGSMS 32 et 33, sous-clade I pour le CGSMS 34, sous-clade II et III pour une partie des CGSMS de la position 9 et sous-clade I pour le reste des CGSMS de cette position) et sont probablement issus du HGT avec acquisition par l'ancêtre commun des sous-clades correspondants. En se référant à l'arbre phylogénétique de la population, nous avons par parcimonie pu inférer l'histoire évolutive de ces clusters. Le cluster 2 (CS de bactériocine) a été soit acquis par l'ancêtre commun récent du sous-clade III et perdu par la souche RLB3-6, soit directement acquis par la souche S1D4-23. Les clusters 32 et 34 (NRPS) ont quant à eux ont été acquis respectivement par l'ancêtre commun du sous-clade III et celui du sous-clade I. ces derniers l'auraient ensuite transféré

latéralement respectivement aux souches S1A1-7 et RLB3-17. Le cluster 33 (CS de lanthipeptide) aurait quant à lui été acquis latéralement par l'ancêtre commun des souches du sous-clade II et III et par la souche RLB3-17 et perdu par l'ancêtre commun des souches du sous-clade II (**Figure B2**). L'ADN acquis par HGT est inséré par recombinaison dans le génome des bactéries réceptrices. Certains de ces points représentent des points chauds d'insertion d'ADN où différentes séquences peuvent s'intégrer. La position 9 (**Figure B5**) illustre un exemple de point chaud de recombinaison.

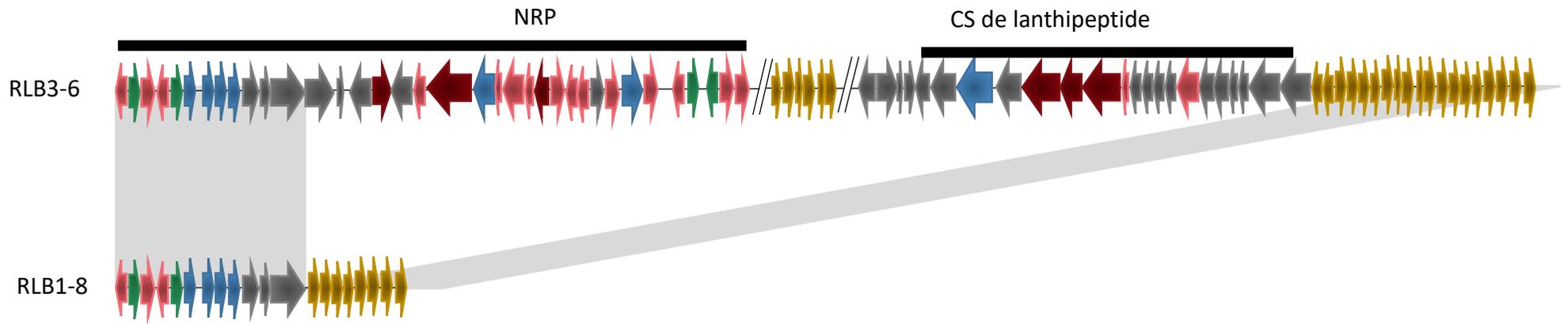


Figure B4 : Acquisition des clusters de synthèse de métabolites spécialisés par transfert horizontal. Les différentes flèches représentent les CDS (en gris foncé les CDS de fonctions inconnues, en bleu les CDS participant au transport de la molécule synthétisée, en vert les gènes de régulation, en rouge les gènes cœur de synthèse du métabolite et en rose les gènes additionnels de synthèse de métabolites. Les régions en gris clair correspondent à la conservation de la synténie et en blanc la rupture de la synténie. L'îlot génomique présent dans la souche RLB3-6 renferme deux clusters de synthèse (un NRPS et un cluster de synthèse de lanthipeptide) absents du génome de la souche RLB1-8.

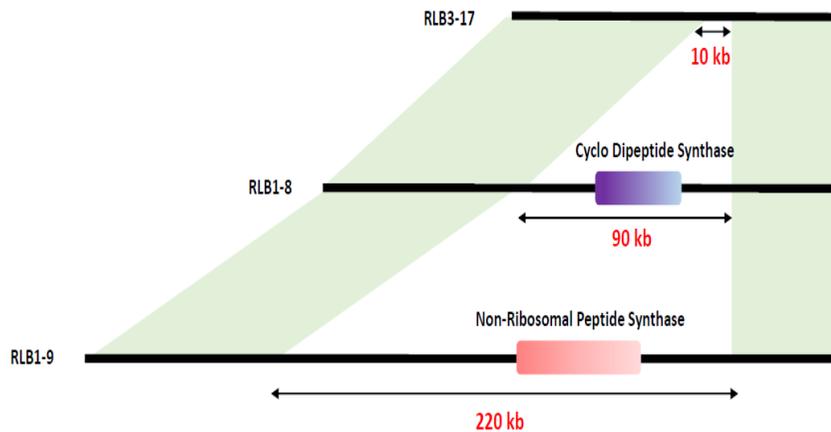


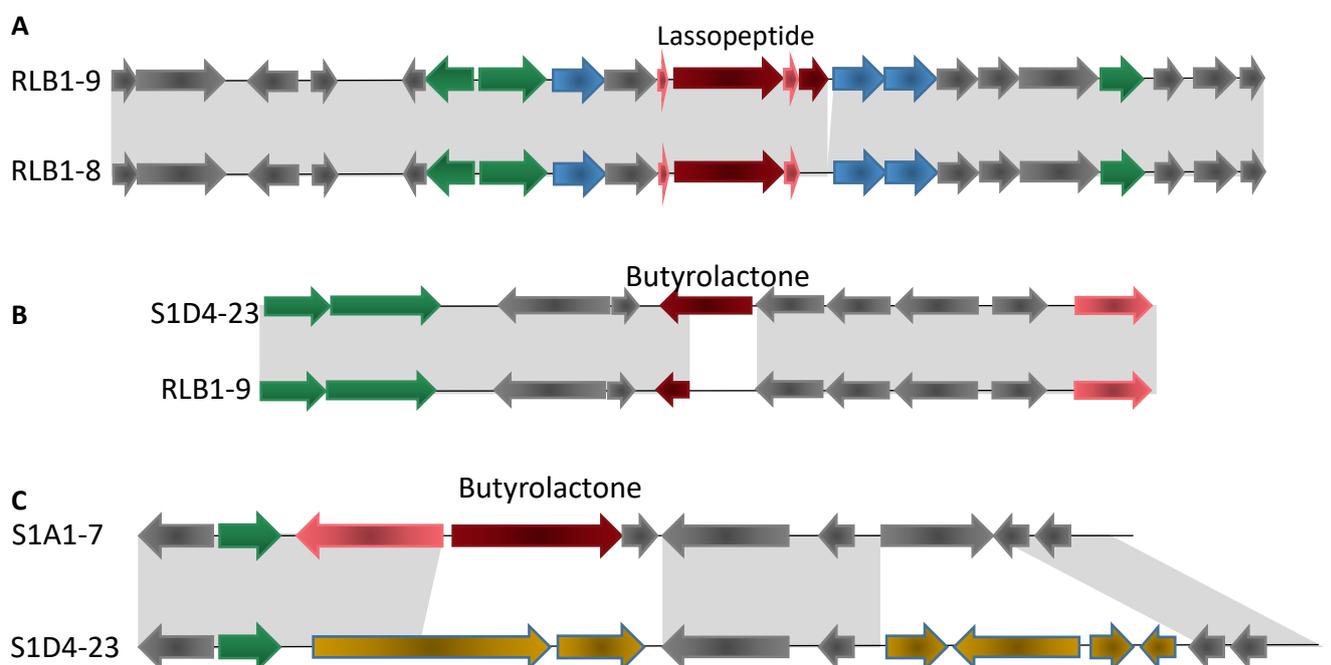
Figure B5 : Identification des clusters de gènes de synthèse de métabolite dans un point chaud de variabilité. Le schéma illustre un point chaud de variabilité observé entre trois souches (RLB3-17, RLB1-8 et RLB1-9). antiSMASH a permis de prédire différents clusters de gènes de synthèse de métabolite : l'un comprenant une cyclodipeptide synthase dépendante de l'ARNt dans RLB1-8 et l'autre comprenant une peptide synthase non ribosomale dans RLB1-9. Les régions en vert clair correspondent aux régions synténiques (Tidjani *et al.*, 2019a).

A cette position, il existe un groupe de gènes codant pour un CDPS dans les souches appartenant au sous-clade de la souche RLB1-8; ces derniers sont remplacés par un cluster codant pour un NRPS-nucléoside dans les autres souches de la population, à l'exception de la souche RLB3-17 dans laquelle aucun cluster de synthèse de métabolite n'est prédit à cette position dans les 10 kb d'ADN présent (Tidjani *et al.*, 2019a). Cette diversité observée en ces points pourrait dépendre soit (i) des propriétés recombinogènes de l'ADN à cette position du chromosome faisant de cette région une zone de tolérance pour l'apparition de la diversité (ii) et/ou de l'importance des fonctions codées par les voies de métabolismes spécialisés présents. En effet si les clusters responsables de la production d'un métabolite donné ne subissent pas une pression de sélection assez forte, ils sont voués à être perdus.

Les réarrangements génomiques comme la translocation d'ADN peuvent également affecter les voies métaboliques par interruption et/ou la séparation des gènes codants pour la voie métabolique (démembrement des voies métaboliques). Deux clusters prédits, l'un à la position 4 dans les souches RLB3-17 et l'autre à la position 1 dans les souches RLB3-6, S1D4-23 et S1A17 sont respectivement identiques aux clusters prédits à la position 9 dans les souches appartenant au sous-clade I et à la position 11 dans toutes les autres souches de la population (**Figure B2**). Ces clusters sont probablement passés de leurs positions initiales (9 et 11) aux positions auxquelles ils sont maintenant prédits (4 et 1 respectivement) suite à des remaniements chromosomiques. Ce phénomène est souvent observé suite à des cassures double brins survenant dans les parties terminales du chromosome. Ces cassures sont généralement réparées par recombinaison homologue où le bras intact est utilisé comme matrice pour la réparation de la cassure aboutissant souvent au rallongement ou au raccourcissement du bras portant la cassure (Hoff *et al.*, 2018).. Ces voies métaboliques spécialisés, par le biais d'événements de translocations (réparation de cassure double brin par recombinaison homologue et l'uniformisation des parties terminales du chromosome), enrichissent les régions terminales du chromosome. Le cluster 1 est maintenant situé dans les répétitions inversées terminales (TIR) des souches RLB3-6, S1D4-23 et S1A17 ; par conséquent, les événements de réarrangements génomiques peuvent entraîner la duplication des voies de synthèse de métabolites spécialisés lorsque ces dernières se retrouvent transloquées dans les TIR. Ceci peut avoir des impacts physiologiques comme l'augmentation du niveau de synthèse d'un métabolite donné due à la présence d'un double cluster. L'acquisition de gènes peut stimuler l'évolution rapide des génomes (Ochman *et al.*, 2000), notamment lorsqu'elle permet l'acquisition simultanée de clusters de gènes par transfert horizontal (HGT). De nombreux clusters de synthèse de métabolisme secondaire présents chez les actinomycètes seraient le résultat de tels événements (Jensen *et al.*, 2007; Penn *et al.*, 2009; Ziemert *et al.*, 2014).

Sur les 12 CGSMS accessoires analysées, 4 CGSMS partiels, non prédits par antiSMASH ont pu être identifiés par l'approche blast. Dans les 4 cas, l'altération d'un gène de biosynthèse était impliquée

dans la non détection par antiSMASH et différentes causes ont pu être inférées afin d'expliquer cette dernière. Le premier cas concerne le CGSMS n°21 prédit pour coder un lassopeptide présent dans l'ensemble des souches à l'exception des souches du sous-clade I (**Figure B2**). La comparaison des génomes de ces souches a révélé qu'elles possédaient l'ensemble des gènes prédits du CGSMS mais que l'un de deux gènes de biosynthèse avait subi une délétion de 13 pb ayant entraîné la perte de son codon stop (**Figure B3.A**). Ce type de perte d'une courte région nucléotidique pourrait être le résultat d'un événement de recombinaison. Pour le CGSMS n°42, présent uniquement dans les souches du sous-clade III, des événements similaires se sont produits avec une recombinaison ayant entraîné la troncature et l'introduction prématurée d'un codon stop dans la séquence du gène responsable de la synthèse du butyrolactone (**Figure B3.B**). Le CGSMS N°5 codant pour la synthèse d'une autre butyrolactone est prédit dans deux souches (RLB3-17 et S1A1-7) uniquement. Les gènes composant ce cluster ont été retrouvés chez les 9 autres souches à l'exception du gène de biosynthèse du métabolite et d'un gène additionnel adjacent. La présence d'autres gènes à leur place dans les souches dans lesquelles ce CGSMS n'est pas prédit a pu être le résultat d'un remplacement de séquence (**Figure B3.C**). Les événements de remplacement peuvent également affecter une voie de biosynthèse sur un nombre plus important de gènes. C'est le cas du CGSMS 41 où le remplacement de l'ADN conduit à la perte du cluster de synthèse bactériocine dans les souches RLB3-6 et S1D4-23 (**Figure B3.D**).



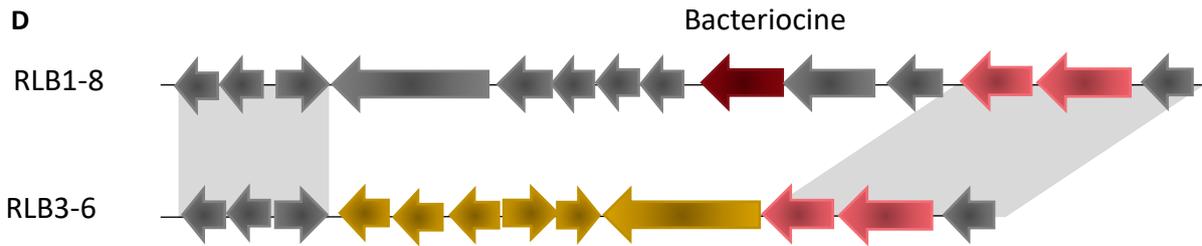


Figure B3 : Evolution des clusters de synthèse de métabolites spécialisés suite à des événements de recombinaison. Les différentes flèches représentent les CDS (en gris foncé les CDS de fonction inconnues, en bleu les CDS participant au transport de la molécule synthétisée, en vert les gènes de régulation, en rouge les gènes cœur de synthèse du métabolite et en rose les gènes additionnels de synthèse de métabolites. Les régions en gris claire correspondent à la conservation de la synténie et en blanc la rupture de la synténie. **A.** Absence de prédiction d'un des gènes cœur de synthèse d'un lassopeptide dans la souche RLB1-8 suite à une recombinaison ayant fait disparaître le codon stop dudit gène. **B.** Le gène cœur de synthèse du butyrolactone a été tronqué suite à une recombinaison ayant introduit un stop prématuré dans la séquence dudit gène dans la souche RLB1-9. **C.** Délétion complète du gène cœur du butyrolactone suite à une recombinaison ayant entraîné le remplacement de ce gène par d'autres chez S1D4-23. **D.** remplacement du cluster de synthèse de la bactériocine suite à une recombinaison ayant entraîné la disparition de ce cluster de la souche RLB3-6.

DISCUSSION

I. Analyse du pan-génomique : une ouverture représentative de la diversité génomique au sein du micro-habitat.

L'analyse du pan-génomique de la population de *Streptomyces* a montré que ce dernier évolue moins vite que celui du genre (**Figure 1**, Chapitre II). En effet alors que l'extrapolation (à 1.000 génomes analysés) du pan-génomique de la population atteint rapidement un plateau avec un maximum de 23.772 gènes, celui du genre continue d'évoluer (420.000 gènes). L'histoire évolutive de ces souches étant très récente (elles partagent une séquence identique du gène codant l'ARN 16S), comparée à celle des souches appartenant à différentes espèces, leur accès à la diversité génétique aurait été plus réduit. Deux facteurs peuvent influencer l'accès d'une population à la diversité génétique. Le premier facteur est temporel et le second est spatial. En effet selon McInerney *et al.*, (2017), la taille du pan-génomique serait influencée par trois paramètres principaux : (i) la capacité des souches de la population à capter de nouvelles séquences d'ADN, (ii) la sélection et (iii) le temps évolutif nécessaire à l'accumulation de la divergence au sein des génomes.

Le facteur temporel concerne le temps évolutif s'étant écoulé depuis la séparation des souches analysées depuis leur ancêtre commun. Au cours de ce temps, les diverses lignées ou espèces composant le genre ont expérimenté des événements génétiques de toutes natures : acquisition et/ou perte de gènes et sélection par l'environnement au sein duquel ces lignées et espèces évoluent. Au sein d'une population formée par des souches ayant récemment divergé, ces dernières auraient eu moins de temps pour accumuler ces événements génétiques (comparé au genre) et la plus grande partie de leurs génomes serait acquise par ascendance verticale. Ainsi, dans une étude portant sur 616 souches de *Streptococcus pneumoniae*, l'analyse de la divergence du génome accessoire par rapport à celle du génome core a permis de déterminer que la diversité du contenu du génome accessoire augmentait proportionnellement à celui du génome core, donc en fonction du temps (**Figure C1**) (Croucher *et al.*, 2014). L'analyse du nombre d'indels différenciant les souches de la population de *Streptomyces* en fonction de la divergence sur leur génome core corroborent ces observations (**Figure S3**, Annexes). Il existe en effet une proportion importante du génome accessoire partagée au sein d'une même sous-clade de la population (**Figure C2**).

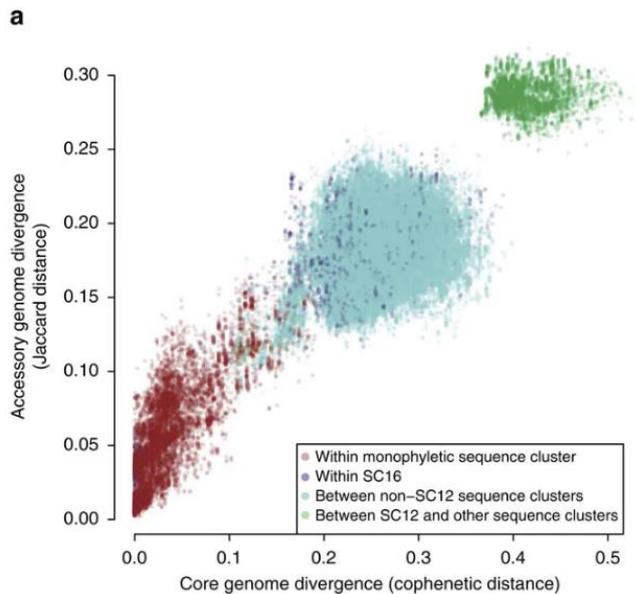


Figure C1 : divergence du génome accessoire en fonction de celle du génome core au sein de 616 souches de *Streptococcus pneumoniae*. SC (pour *Séquence Clusters*) représente des groupes phylogénétiques définis à partir du génome core (Croucher *et al.*, 2014).

Le second facteur qui restreint l'accès à la diversité génétique est d'ordre spatial. La diversité génétique contenue au sein d'un même écosystème est moindre par rapport à la diversité globale dans plusieurs écosystèmes différents. En effet, la diversité génétique est soumise à la pression de sélection de l'écosystème, ce qui réduit le répertoire de gènes disponibles à ceux sélectionnés par la niche (Cohan, 2002b; 2002a). Cela a été démontré chez *Streptococcus pneumoniae* où la fermeture du pan-génome reflète l'adaptation à la niche écologique actuelle que représente la sphère ORL (voies aérodigestives supérieures) (Donati *et al.*, 2010). Dans notre étude, notre échantillon de génomes serait donc représentatif de la diversité génétique du micro-habitat (Vernikos *et al.*, 2015).

De plus, ces deux facteurs, spatial et temporel, sont liés ; les individus génétiquement proches ont une probabilité importante d'avoir évolué au sein du même écosystème contrairement aux espèces séparées par une histoire évolutive longue.

Deux études de génétique de la population portant sur le pan-génome procaryote proposent une explication divergente quant à l'origine de ce dernier. Alors, qu'Andreani *et al.* (2017) suggèrent qu'une grande partie de la diversité génétique au sein d'une population serait neutre (non sélectionnée par l'environnement), McInerney *et al.* (2017) suggèrent que le pan-génome est adaptatif. Notre étude chez *Streptomyces* permettrait de concilier ces deux points de vue. Les *Streptomyces* sont retrouvées dans de nombreux écosystèmes (sol, océan, insectes etc...). Cette adaptation a imposé alors une dynamique forte du génome (acquisitions et pertes de gènes fréquentes). L'ouverture du pan-génome aurait participé au succès évolutif de cette bactérie. À l'échelle de la population, *Streptomyces* démontre également une ouverture de son pan-génome relative à la pression de sélection liée à la niche occupée. Ainsi, une partie des gènes du génome accessoire pourrait être neutre.

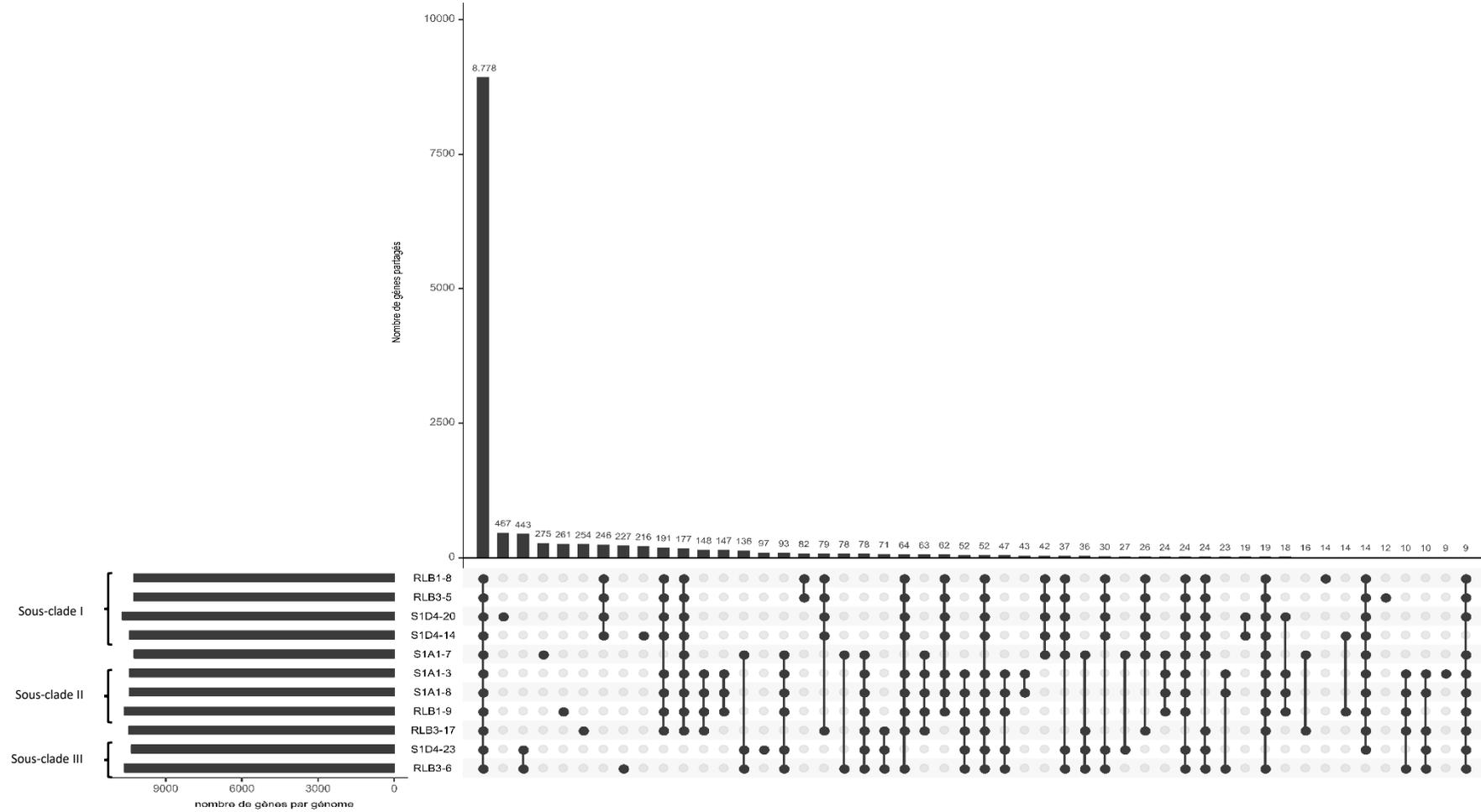


Figure C2 : Répartition des gènes identifiés entre les souches de la population. Le diagramme en bas à gauche représente la contribution de chaque génome en nombre de gènes et définissent les différents sous clades phylogénétiques auxquels appartiennent les souches de la population. Les histogrammes représentent le nombre (écrit au-dessus) de chaque groupe de gènes partagés entre les individus reliés par les boules en dessous. Ainsi le premier histogramme correspond au core génome (8 778 gènes) partagé par l'ensemble des souches, le second à 467 gènes spécifiques de S1D4-20 et le troisième à 443 gènes retrouvés uniquement chez S1D4-23 et RLB3-6.

II. La compartimentation du génome de *Streptomyces*

La comparaison génomique de plusieurs souches de *Streptomyces* (*S. ambofaciens*, *S. coelicolor* et *S. avermitilis*) a permis de révéler l'existence d'une compartimentation des chromosomes avec une partie centrale conservée et une variabilité confinée au niveau des régions terminales constituant les bras chromosomiques (Choulet *et al.*, 2006). Ces régions terminales représentent plusieurs centaines de kilobases ; soit jusqu'à 20 % des chromosomes comparés (Choulet *et al.*, 2006a). Au laboratoire, cette compartimentation a été observée au travers des études sur l'instabilité chromosomique (Leblond *et al.*, 1989 ; Fischer *et al.*, 1997 ; Wenner *et al.*, 2003 ; Hoff *et al.*, 2018). Les régions terminales pourraient donc être sujettes à des remaniements chromosomiques à l'origine de la diversification de l'information génétique.

La comparaison deux à deux des génomes de la population a révélé que les insertions/délétions (indels) sont 5 fois plus fréquentes dans les régions terminales que dans la région centrale (**Figure C3**). Ces résultats corroborent ceux obtenus par Choulet *et al.* (2006a). De façon complémentaire à cette étude portant sur la comparaison d'espèces séparées par un temps évolutif long, l'analyse des indels dans notre population révélerait des événements moléculaires récents représentant les premières étapes de diversification dont l'accumulation sur des temps évolutifs plus importants finirait par saturer les régions terminales qui seraient devenues spécifiques. En contraste, les régions centrales auraient conservé la synténie ancestrale.

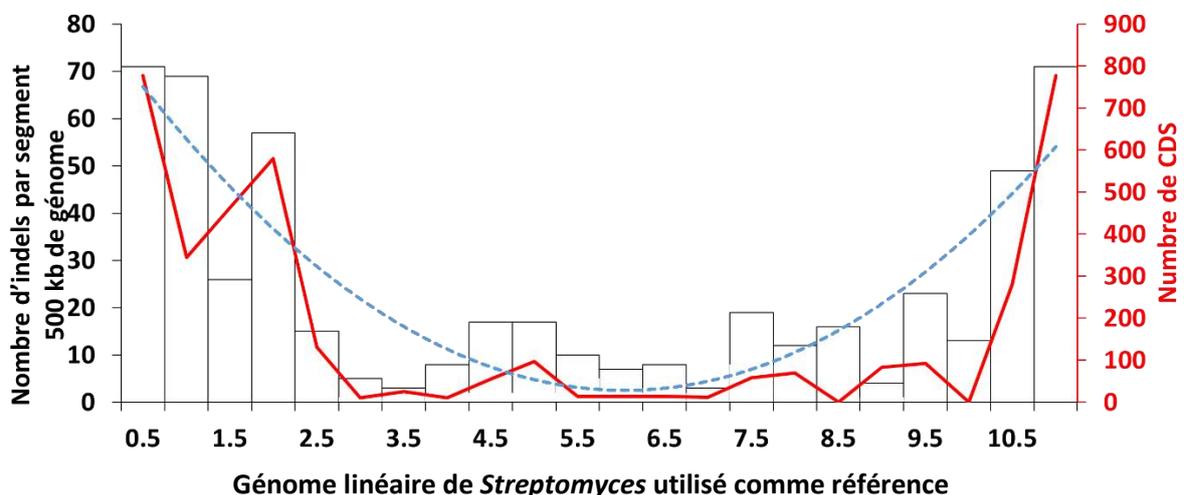


Figure C3 : Distribution des indels le long du chromosome linéaire de *Streptomyces* : les indels identifiés (par fenêtre de 0,5 Mb) par comparaison des 11 génomes de la population sont répartis le long d'un chromosome de référence. La courbe bleue correspond à la courbe de distribution polynomiale d'ordre 2 des indels. La courbe rouge correspond au nombre de gènes impliqués dans les indels dans chaque fenêtre de 0,5 Mb.

Nous proposons que la plasticité génomique mise en évidence dans notre étude soit à l'origine de la diversification rapide des génomes chez les *Streptomyces*.

La linéarité du chromosome n'est pas que l'apanage des bactéries du genre *Streptomyces* chez les procaryotes. D'autres bactéries comme *Borrelia* et *Agrobacterium* possèdent également des chromosomes linéaires. La comparaison des génomes des souches de *Borrelia burgdorferi* a permis d'identifier que ces souches possédaient un chromosome stable sauf sur les derniers 20 kb à l'extrémité (Casjens *et al.*, 2012). Les séquences de cette région semblent héritées suite à des échanges avec les nombreux plasmides linéaires retrouvés chez cette bactérie (Casjens *et al.*, 1997). Dans cette étude il a été démontré que presque tous les réplicons linéaires ont subi des réarrangements depuis leur ancêtre commun.

La plasticité des régions subtélomériques des chromosomes linéaires a également été décrite chez les eucaryotes de la levure (pour revue Dujon *et al.*, 2004) aux chromosomes humains. Alors que 98 % des gènes présentent une organisation identique entre les génomes de *Saccharomyces cerevisiae* et de *Saccharomyces bayanus* var. *uvarum*, un biais dans la localisation des ruptures de synténie a été démontré : 46 % des ruptures observées concernent les régions subtélomériques qui ne représentent que 10 % du génome (Fischer *et al.*, 2001). Des caractéristiques communes à celles décrites pour les génomes de levures ressortent de l'analyse des 14 chromosomes du parasite humain *Plasmodium falciparum* (Gardner *et al.*, 2002) chez qui les extrémités chromosomiques concentrent des familles de gènes hautement variables.

L'ensemble de ces résultats suggère qu'il existerait un lien entre la linéarité et la plasticité des régions subtélomériques.

III. Des extrémités chromosomiques fortement plastiques

Les bras chromosomiques se terminent par des répétitions terminales inversées (TIR) dont la taille varie selon les souches étudiées. Chez *S. avermitilis* les TIR sont restreintes aux séquences télomériques (169 pb). Chez *S. coelicolor*, la variation de la taille des TIR a été associée à l'évolution des différentes lignées de cette souche. En effet, alors que la souche originelle isolée du sol possède une taille de TIR de 1,06 Mb (Weaver *et al.*, 2004), la souche de laboratoire possède des TIR réduites à 22 kb (Bentley *et al.*, 2002). Les TIR peuvent également subir des événements de recombinaison entraînant leur expansion. Le cas le plus spectaculaire a été décrit chez *S. ambofaciens* dont la taille varie de 5 kb à 1,4 Mb (Wenner *et al.*, 2003) alors que la souche sauvage possède des TIR de 210 kb (Leblond *et al.*, 1996). L'ensemble de ces réarrangements survenus dans les conditions de laboratoire ont été inférés à partir de la comparaison des souches de la même espèce (Wenner *et al.*, 2003).

La comparaison globale des TIR des 11 génomes par alignement multiples a permis d'identifier plusieurs types d'évènements de réarrangement : insertions, délétions, inversions et translocations. Certaines régions des TIR sont souche-spécifiques et témoigneraient d'une acquisition récente par

transfert horizontal. L'ensemble de ces évènements seraient à l'origine de la diversité, en taille et en séquence, des TIR constatée au sein de la population.

La dynamique des génomes impacte également les télomères

Les télomères des *Streptomyces* (180 derniers nucléotides du réplicon) sont capables de former des structures secondaires de type tige-boucle (Huang *et al.*, 1998) qui jouent un rôle important dans la réplication terminale (Chang & Cohen, 1994) et le maintien de la linéarité des réplicons. Des études précédentes avaient montré que ces séquences étaient conservées au niveau intra et interspécifique chez *Streptomyces* (Huang *et al.*, 1998). Au niveau des chromosomes de la population, les séquences télomériques sont uniques et spécifiques aux sous-clades I et II avec une identité nucléotidique intra-clade allant de 92 à 97 %. En revanche, l'identité nucléotidique en inter-clade chute à près de 30 %. Ces résultats contrastent de ceux obtenus par Huang *et al.* (1998). Ces derniers avaient relevé une forte conservation de la séquence (170 derniers nucléotides), et donc de la structure, des télomères au niveau interspécifique. L'identité nucléotidique des séquences télomériques entre les chromosomes et les plasmides était cependant moindre que celle observée entre les chromosomes (Huang *et al.*, 1998).

Les souches du sous-clade I partagent des télomères similaires à ceux du plasmide pS1D4-14.1 identifié dans l'une de ces souches. Ce résultat suggère un échange de télomères entre chromosome et plasmide. Ce phénomène a précédemment été décrit chez *S. coelicolor* où un chromosome chimérique a été obtenu par échange télomérique avec le plasmide linéaire SCP1 (Yamasaki & Kinashi, 2004). L'interaction entre le plasmide pZG101 et le chromosome de *S. rimosus* a également abouti à un échange d'extrémités provoquant également le transfert du cluster de biosynthèse de l'antibiotique oxytétracycline du chromosome vers le plasmide (Pandza *et al.*, 1998). Chez *S. lividans*, l'analyse des régions terminales du chromosome et du plasmide linéaire SLP2 suggère également un échange d'extrémités. L'extrémité droite du plasmide étant identique à celle du chromosome sur environ 15,4 kb (Huang *et al.*, 2003).

L'échange des extrémités n'est donc pas un phénomène rare chez les *Streptomyces*. Ce type d'échange serait plus facilité, puisqu'il ne nécessite qu'un simple crossing-over, contrairement aux évènements de remplacement de séquences internes au chromosome, qui eux nécessitent un double événement de recombinaison. Les échanges terminaux incluant le télomère seraient favorisés, car ils permettent le maintien de la linéarité chromosomique.

IV. Mécanismes de la diversité génomique

Les bactéries du genre *Streptomyces* ne sont pas naturellement transformables (Chater & Hopwood, 1984). Par ailleurs, une seule étude a décrit la transduction généralisée chez *Streptomyces coelicolor*

(Burke *et al.*, 2001). En revanche de nombreux éléments conjugatifs ont été identifiés chez ces bactéries faisant de la conjugaison le mécanisme de transfert horizontal le plus probable chez les *Streptomyces*. Ces éléments s'intègrent dans le génome par recombinaison site-spécifique en reconnaissant des séquences cibles spécifiques (sites *attB*). Cette insertion a lieu majoritairement dans les séquences des gènes conservés comme les gènes codant les ARNt. Le plasmide conjugatif pSAM2, par exemple, s'intègre de façon site-spécifique dans le gène codant ARNt-proline. Cette prédilection pour les gènes conservés leur permet d'élargir le nombre des hôtes potentiels dans lesquels ils pourraient s'intégrer (Ghinet *et al.*, 2011). Ainsi pSAM2 est capable par exemple de s'intégrer dans le génome de *M. smegmatis* dans le gène codant pour l'ARNt-proline (Martin *et al.*, 1991). Contrairement au phage lambda, qui nécessite le facteur d'intégration dans l'hôte (*IHF* pour *Integration Host Factor*) pour son intégration, les AICE n'en nécessitent aucun (Katz *et al.*, 1991 ; Raynal *et al.*, 1998), les rendant alors moins dépendants de l'hôte pour leur intégration.

L'analyse des génomes de la population a permis d'identifier au total 2 ICE (*Integrative and Conjugative Element*) et 52 AICE (*Actinomycete Integrative and Conjugative Element*). Les AICE et ICE sont intégrés exclusivement dans la partie centrale conservée du chromosome. Ces éléments conjugatifs ont été identifiés par la présence des gènes codant les protéines clés (intégrase et translocase). Ils sont donc potentiellement capables d'excision et de transfert autonome. Les souches de la population possèdent en moyenne 5 éléments conjugatifs, ce qui est compatible compte tenu de la taille des génomes avec l'estimation de Bordeleau *et al.* (2012) chez les actinomycètes.

Outre les AICE et ICE, des études ont montré que les plasmides circulaires pSN22 (Kataoka *et al.*, 1991), pSG5 (Muth *et al.*, 1988) et SCP2 (Brolle *et al.*, 1993) étaient capables de réaliser une conjugaison *via* TraB.

Les génomes des souches de la population renferment des plasmides qui sont soit linéaires ou circulaires. Ces réplicons étant souche-spécifiques, ils auraient alors été acquis par transfert horizontal. Alors que le mécanisme de transfert des réplicon linéaires a été évoqué, celui des réplicons linéaires chez *Streptomyces* soulève des interrogations. En effet, les réplicons linéaires possèdent à leurs extrémités 5' des protéines terminales attachées covalamment. Ces protéines avec leur taille d'environ 20 kDa passeraient difficilement dans le pore formé par la translocase TraB dont le diamètre atteint 3,1 nm (Vogelmann *et al.*, 2011). Chez le plasmide linéaire SLP2 de *S. lividans*, il a été démontré que l'inactivation du gène *traB* abolissait son transfert (Hsu & Chen, 2010), ce qui suggère que la translocase participe au transfert des réplicons linéaires. L'hypothèse d'un mécanisme impliquant un système enzymatique, lors du transfert d'un réplicon linéaire *via* TraB, capable de détacher les

protéines terminales avant le transfert et les rattacher après celui-ci a été proposé (Vogelmann *et al.*, 2011).

CW Chen en 1996 avait proposé un mécanisme de transfert conjugatif des réplicons linéaires appelé « end-first », dans lequel le télomère jouerait le rôle d'origine de transfert. Ce mécanisme permettrait le transfert des réplicons linéaires sous forme simple brin. Le transfert par les extrémités pourrait alors faire le lien avec la grande variabilité des régions terminales chez *Streptomyces*. La compartimentation du chromosome chez *Streptomyces* pourrait alors refléter une organisation optimale des gènes vis-à-vis de leur dissémination par transfert horizontal à partir de l'extrémité. Chez les souches Hfr d'*E. coli*, la mobilisation du chromosome depuis l'origine de transfert induit un gradient de fréquence de transfert des marqueurs chromosomiques en fonction de leur localisation par rapport à l'*oriT* (Reimann & Haas, 1993). Par analogie, chez *Streptomyces* il pourrait exister un gradient de fréquence de mobilisation des marqueurs en fonction de leur distance par rapport aux télomères.

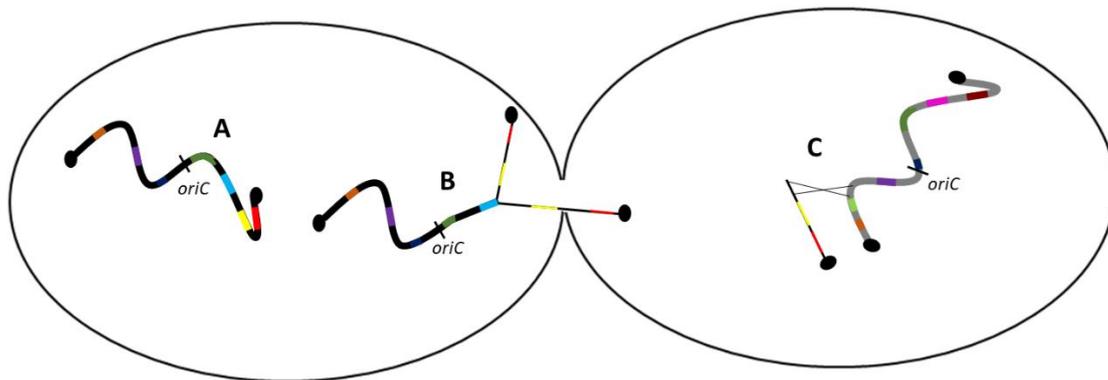


Figure C4 : Modèle de conjugaison « End First » chez *Streptomyces*. A. La conjugaison est initiée à partir des télomères du chromosome de *Streptomyces*. B. Les marqueurs chromosomiques des régions terminales (rouge et jaune) seront transférés en premier. C. Par recombinaison, le fragment transféré sera incorporé au niveau des régions terminales du chromosome de la cellule réceptrice.

Un argument en faveur de ce modèle est la présence du gène *ttrA* localisé dans les régions télomériques de SLP2 et d'autres réplicons linéaires et dont le produit présente une activité hélicase. L'implication de ce gène dans le transfert avait été suggérée (Chen, 1996 ; Huang *et al.*, 2003). Le rôle proposé pour cette hélicase codée par *ttrA* est le déroulement de l'ADN lors de l'initiation du transfert conjugatif. Un homologue de ce gène a été identifié sur certains réplicons linéaires des souches de notre population. Sa localisation conservée dans une région variable suggère une pression sélective forte.

Ce mécanisme proposé par CW Chen, ne permet cependant pas d'expliquer la compartimentation par accumulation des indels dans les régions terminales au niveau du chromosome chez *Streptomyces*. En effet par le transfert des régions terminales, le mécanisme « End First » permettrait plutôt une homogénéisation de ces régions par des événements de recombinaisons entre l'extrémité acquise par transfert horizontal et celle de la souche réceptrice (Figure C4). Les deux mécanismes suggérés pourraient ne pas être exclusifs.

V. Mécanismes de mobilisation des marqueurs chromosomiques

La mobilisation des marqueurs chromosomiques lors de la conjugaison des ICE (transfert de molécules simple brin) a été décrite chez les bactéries à coloration gram négative et chez les firmicutes. Chez la souche de *E. coli* K-12 (contenant le facteur F), ce mécanisme impliqué dans la mobilisation des marqueurs chromosomiques (Lloyd & Buckman, 1995) nécessite l'intégration préalable du plasmide (avec son origine de transfert *oriT*) dans le chromosome (Thomas & Nielsen, 2005). Chez *Mycoplasma agalactiae*, une étude récente a montré l'échange conjugatif de grands fragments chromosomiques, qui mobilise de façon équiprobable toutes les parties du chromosome par un mécanisme non conventionnel, lors de l'échange d'un ICE. Cette étude a montré que l'ensemble du génome est potentiellement mobilisable au cours de la conjugaison (Faucher *et al.*, 2019). L'exemple le plus documenté de mobilisation des marqueurs chromosomique, chez les actinobactéries, a été décrit lors du transfert conjugatif distributif chez *Mycobacterium smegmatis* (Distributive Conjugative Transfer, DCT). Chez *M. smegmatis*, le génome des transconjugants est une mosaïque des génomes parentaux. Plus du quart des génomes des transconjugants provient de la bactérie donatrice et est reparti en une trentaine de segments sur les chromosomes des réceptrices. Ces fragments varient en taille de 50 pb à 225.000 pb. Cette mosaïque génomique générée en seul événement de transfert horizontal ressemblerait beaucoup plus à la recombinaison méiotique (brassage de l'information génétique des informations parentales)(Derbyshire & Gray, 2014). Chez *Streptomyces*, des études ont démontré que la translocase TraB était également capable de mobiliser l'ADN chromosomique *in trans* en reconnaissant des séquences ressemblant aux *clt* (*cis-acting locus of transfer*), appelés *clc* (*clt-like chromosomal sequence*) qui sont disséminées le long du chromosome (Kieser *et al.*, 1982). TraB qui est une ADN translocase de la famille FtsK/SpoIIIE agirait alors comme une pompe permettant non seulement le passage des éléments conjugatifs, mais aussi de fragments de chromosomes. L'ampleur et la taille des marqueurs chromosomiques mobilisés lors du transfert des AICE chez *Streptomyces* restent cependant inconnues. Il serait intéressant d'analyser la fréquence de mobilisation *in trans* des différentes régions chromosomiques (par leur marquage) lors de la conjugaison au sein des souches de notre population. Le séquençage des transconjugants permettrait de déterminer la quantité d'information transférée au cours du transfert de l'AICE. La taille des indels identifiés au sein des

génomomes (pouvant atteindre 241 gènes) donne un aperçu de la quantité d'ADN pouvant être mobilisé par ce mécanisme.

VI. Existe-t-il des barrières au transfert au sein de la population de *Streptomyces*?

Un mécanisme empêchant le transfert d'éléments conjugatifs apparentés aux éléments résidents, appelé mécanisme d'immunité, a été décrit chez plusieurs bactéries (Dunny, 1990 ; Weidlich *et al.*, 1992 ; Frost *et al.*, 1994). Chez l'élément pSAM2 de *S. ambofaciens*, un gène, *pif* pour pSAM2 immunity factor, empêche le transfert entre individus possédant l'élément, ceci en le maintenant intégré dans le chromosome sous forme non conjugatif (Possoz *et al.*, 2003). D'autres mécanismes comme le système de « restriction-modification » RM (Oliveira *et al.*, 2016) le système CRISPR-Cas (Barrangou *et al.*, 2007 ; Marraffini & Sontheimer, 2008) peuvent également empêcher l'échange de matériel génétique. Au laboratoire (thèse C. Choufa), nous nous intéressons à la cartographie des échanges des AICE au sein de la population. Les résultats préliminaires de cette analyse montrent que la population ne semble pas panmictique (toutes les souches analysées ne sont pas capables de donner ou de recevoir). Ces résultats suggèrent qu'il existerait des barrières au transfert des AICE au sein de la population. Si les barrières du type isolement spatial (Linz *et al.*, 2007) et génétique (Ansari & Didelot, 2014) sont peu vraisemblables au sein d'une population génétiquement cohésive et sympatrique, la présence de système de restriction-modification ou de CRISPR pourraient restreindre les échanges entre les souches.

La forte prévalence des ICEs/AICEs au sein de la population suggère que les mécanismes conjugatifs sont les moteurs principaux de la diversité génomique. Isolées de la rhizosphère d'un jeune hêtre, certaines souches étaient donc en contact avec les racines, d'autres non. Est-ce que certains paramètres biotiques (micro-organismes ou racine) sont-ils de nature à influencer les transferts de gènes ? Est-ce que les exsudats racinaires stimulent les échanges des éléments conjugatifs et intégratifs ?

VII. Les clusters de synthèse de métabolites spécialisés (CGSMS) comme proxy de l'impact fonctionnel de la diversité génétique

Les clusters de synthèse de métabolites spécialisés (CGSMS) représentent 6,7 % du génome accessoire de la population ; aussi l'impact fonctionnel des transferts de gènes dépasse celui de la capacité de synthèse de métabolites spécialisés. Cependant comme ceux-ci sont faciles à prédire grâce à des algorithmes et des bases de données dédiées (ex. antiSMASH, Blin *et al.*, 2019), aussi ces clusters de gènes peuvent servir de proxy pour estimer l'impact fonctionnel.

Les métabolites spécialisés jouent un rôle prépondérant dans l'adaptation des micro-organismes à leur environnement. L'évolution des voies de synthèse de ces métabolites contribuerait à la capacité d'adaptation et de colonisation d'une nouvelle niche. Les gènes de biosynthèse de métabolites spécialisés sont en général surreprésentés dans le génome accessoire (Nakamura *et al.*, 2004 ; Cordero & Hogeweg, 2009 ; Boucher *et al.*, 2011). Le « turnover » des gènes au sein des bactéries pourrait résulter de l'adaptation aux différents changements provoqués par les interactions biotiques (interactions entre organismes vivants).

L'analyse des clusters impliqués dans la synthèse des métabolites spécialisés au sein de la population (n = 42) a permis d'identifier un groupe de clusters « core » (n= 30) partagés par l'ensemble des individus de la population. Nous avons également pu identifier des clusters accessoires qui sont soit uniques, ou partagés par au moins deux souches de la population (n = 12). Les CGSMS core étant issus d'un héritage vertical, l'analyse des CGSMS accessoires nous a permis d'identifier et d'inférer les mécanismes récents de diversification impactant les voies de biosynthèse des métabolites spécialisés. Certains clusters entiers ont été perdus par délétion (exemple du CGSMS 12 ; **Figure A2**, chapitre IV). La perte d'un CGSMS peut ne pas être contre-sélectionnée ; les individus peuvent même être favorisés en ne supportant pas le coût de la maintenance du cluster de gènes et de la synthèse du métabolite tout en bénéficiant de ceux produits par les congénères. Ces individus peuvent être qualifiés de tricheurs. En contrepartie, ils peuvent contribuer sur un autre trait physiologique. Cela aboutirait à une répartition des tâches entre les individus de la population. Cette répartition des tâches assurerait la cohésion de la population.

Malgré la diversité (présence/absence) des CGSMS au sein de la population, les interactions bipartites réalisées au laboratoire ont permis de démontrer qu'aucune des souches n'était capable d'inhiber la croissance de ses consœurs. Cependant, une partie des souches inhibe la croissance d'une souche de *Bacillus* isolée du même environnement. La corrélation entre la présence de CGSMS et l'activité d'inhibition de la croissance de *Bacillus* par interruption nous a permis d'identifier un cluster de NRPS-nucléoside comme potentiel candidat (Cluster 9, **Figure A2**, chapitre IV). Le produit de ce cluster peut être considéré comme un bien commun de la population. En effet, tous les individus bénéficieraient de la production du métabolite, par son activité inhibitrice, mais seules certaines en paieraient le coût (maintien, production). De telles stratégies ont été précédemment décrites par Cordero *et al.*, (2012) chez les Vibrionaceae. Dans ce cas, des sidérophores (chélateurs de fer) sont produits par certains individus au bénéfice du reste de la population qui a perdu la capacité de biosynthèse tout en conservant celle du transport du sidérophore. Ces résultats montrent que la proportion de « profiteurs » dépend des interactions biotiques entre les individus de la population.

La population représenterait alors une unité écologique au sein de laquelle l'optimisation des apports de chaque individu tendrait à spécialiser ce dernier avec comme impact l'accroissement de la diversité génétique et fonctionnelle. Cette hypothèse dite de « la Reine Noire » formulée par Morris *et al.* (Morris *et al.*, 2012) propose que la perte d'une fonction est sélectivement favorable à l'échelle des individus dans une population. La proportion des individus profiteurs peut augmenter dans la population tant que la production du bien commun est suffisante pour la communauté. Cette hypothèse suggère alors que les espèces qui se spécialisent dans la production d'un bien commun échappent à la compétition et favorisent donc la coexistence entre les individus.

Cordero & Polz (2014) ont proposé de classer les gènes en fonction de leur fréquence dans les génomes d'une population : (i) les gènes de fréquence élevée, (ii) les gènes de fréquence intermédiaire et (iii) les gènes de basse fréquence. La fréquence d'un gène traduirait sa valeur sélective ; ce gène est conservé s'il est soumis à une sélection forte, et sporadique lorsqu'il ne subit pas de pression ou lorsque cette dernière est faible. Par analogie, on peut distinguer les CGSMS selon leur fréquence dans les individus de la population. Les CGSMS conservés assureraient le métabolisme spécialisé « core » (ex. synthèse des sidérophores). Ils seraient hérités verticalement et sont soumis à une forte pression de sélection. Les CGSMS trouvés dans une partie de la population pourraient assurer la synthèse de biens communs et leur maintien résulterait de la sélection pour l'établissement des interactions biotiques entre les organismes au sein de la niche. Ces CGSMS pourraient être préférentiellement perdus par certains individus de la population. Enfin, les clusters souches spécifiques sont localisés dans les régions variables du génome et ne confèreraient généralement pas de différenciation déterminante aux souches qui les contiennent. Ces clusters seraient dans certains cas impliqués dans la réponse transitoire à un changement local. Ils seraient issus d'un transfert récent. Cela pourrait être le cas par exemple du cluster de aryloène-NRPS-*like* localisé sur le réplicon extrachromosomique de la souche S1D4-20 et qui est spécifique à cette souche.

VIII. Concept de population : approche écologique ou génétique ?

L'espèce est souvent présentée comme l'unité fondamentale en écologie parce qu'elle rassemble des individus interféconds qui sont phénotypiquement donc écologiquement plus similaires entre eux qu'avec les individus appartenant à une autre espèce. Pour les bactéries et les archées, l'identification des groupes était difficile avant l'avènement des techniques moléculaires parce que les caractères phénotypiques utilisés par la taxonomie bactérienne traditionnelle étaient peu nombreux et arbitrairement définis. Même si la classification traditionnelle des bactéries permettait une identification fiable (Rosselló-Mora & Amann, 2001), il est évident que les caractères phénotypiques (ex. utilisation des symptômes causés par des bactéries infectieuses) ne définissent pas des individus appartenant à une même unité naturelle. L'accès aux séquences d'ADN a permis de définir des liens

génétiques plus ou moins forts entre groupes d'individus. Ces groupes dérivant d'un ancêtre commun récent et partageant le même micro-habitat sont qualifiés de « populations génotypiques » ou « groupes génotypiques » selon Shapiro et Polz (2014). Reste à trouver des consensus sur les seuils d'identité génétique délimitant les groupes taxonomiques opérationnels.

Mieux intégrer les approches génétiques et écologiques est une voie qui semble pertinente dans l'identification et la définition des populations.

VIII.1 « L'écologie reverse » comme approche d'identification et de définition des populations

L'écologie reverse est l'utilisation des données issues de la génomique et de l'écologie pour déterminer si des souches appartiennent ou non à une même population. Ce terme fut proposé pour la première fois par Matthew Rockman en 2007 (Li *et al.*, 2008). Elle a pour but de déterminer si des souches étroitement apparentées et sympatriques constituent des unités génotypiques (populations) et de tester les différences écologiques pouvant exister entre ces populations en utilisant les métadonnées écologiques recueillies ou par des expérimentations au laboratoire (Shapiro & Polz, 2014).

Pour réaliser une approche d'écologie réverse, il est important :

- i) de réaliser un échantillonnage de souches très apparentées appartenant à la même niche, à deux niches différentes ou encore possédant des caractéristiques phénotypiques qui les distinguent, dans le but de tester si elles appartiennent ou non à la même sous-population. Les isolats doivent être échantillonnés à la micro-échelle pour réduire les effets de la divergence allopatriques (divergence due à la séparation de niche), et permettre de se focaliser sur l'effet de la sélection et recombinaison locale.
- ii) De choisir une approche de génomique. Le séquençage des génomes permettra d'inférer les événements de recombinaison entre les souches.
- iii) De disposer d'un assemblage des génomes de qualité pour permettre une analyse des variations nucléotidiques (pouvant être biaisée par les erreurs de séquençage).
- iv) De réaliser un alignement de génomes permettant une comparaison génomique fine.
- v) D'évaluer les signatures phylogénétiques dans les Single-Nucleotide Polymorphism (SNP) trouvés dans le génome core. Des méthodes phylogénétiques standards peuvent être utilisées pour construire une phylogénie basée sur les gènes du génome core et l'impact de la recombinaison peut être mesuré en évaluant le déséquilibre de liaison entre les SNP.
- vi) De pouvoir évaluer si il existe une corrélation entre les groupes phylogénétiques formés et leur écologie (Li *et al.*, 2008 ; Levy & Borenstein, 2012 ; Shapiro & Polz, 2014).

VIII.1.1 Utilisation d'une approche avec des caractères écologiques prédéfinis

La population analysée dans cette étude est composée de souches dont certaines ont été échantillonnées au niveau des racines et d'autres au niveau du sol nu. Ces deux micro-niches déterminent deux conditions biotiques différentes. La plante par ses exsudats racinaires pourrait sélectionner des génotypes moins fréquents, voire absents du sol nu. Parallèlement, d'autres facteurs comme des interactions biotiques (*i.e.* présence de bactéries compétitrices) au niveau du sol nu pourraient également sélectionner des génotypes qui seraient absents au niveau racinaire. Deux sous-populations pourraient donc coexister. En appliquant le principe de l'écologie reverse qui consiste à utiliser les données de la génomique et de l'écologie, pour qu'on ait sélection par la racine ou une adaptation au niveau du sol nu, tous les individus appartenant à l'une ou à l'autre niche devraient posséder en commun une proportion de gènes accessoires. Si cette hypothèse est vérifiée, les souches de la population formeraient alors deux populations génétiquement et écologiquement différenciées et stables issues d'une sélection par la racine ; sinon la racine ne sélectionnerait pas de génotypes particuliers et les souches formeraient une seule et même population.

L'analyse de la répartition des gènes du pan-génome au sein des individus de la population (**Figure C2**), montre que ni les souches provenant de la racine (dont les noms commencent par l'initial R *i.e.* RLB1-8), ni celles provenant du sol nu (*i.e.* S1A1-3) ne possèdent en commun des gènes faisant partie du génome accessoire. Ces résultats suggéreraient que ni la racine, ni le sol nu ne sélectionneraient préférentiellement pas de génotypes particuliers.

Le choix des caractères écologiques à utiliser dans une approche de reverse écologie doit cependant être parcimonieux. En effet, les caractères écologiques, principalement ceux présents chez une partie de la population (fréquence intermédiaire), qui servent à maintenir les interactions sociales (*i.e.* biens communs ou interaction prédateur-proie) ne devraient pas être utilisés comme critères sélectifs. Au sein d'une population, ce type de caractère ne confère pas davantage qu'à un seul individu, mais plutôt à l'ensemble du groupe. Contrairement aux caractères adaptatifs (*vis-à-vis* d'une niche) qui sont capables d'engendrer une spéciation avec formation de sous-populations différenciées, les gènes participant à la cohésion sociale au sein d'une population ne pourraient pas servir à borner des sous-populations au sein d'une population cohésive de souches conspécifiques et sympatriques. C'est peut-être le cas du CGSMS responsable de l'activité inhibitrice identifiée chez une partie de la cette population.

En utilisant ce caractère écologique, selon l'approche de la reverse écologie, les souches de la population formeraient deux sous-populations distinctes (celles qui produisent l'activité antibiotique et celles ne la produisant pas) si les souches productrices de l'activité appartiennent au même sous-

clade phylogénique et que les non-productrices appartiennent également à un autre même sous-clade. L'analyse phylogénétique basée sur les gènes du génome core a permis d'identifier trois sous-clades au sein de cette population. Les souches non productrices de l'activité antibiotique appartiennent au même sous-clade phylogénique (sous-clade I); celles qui produisent une activité appartiennent également à deux sous-clades (II et III) différents de celui des non-productrices. Cela suggérerait alors que ces souches analysées appartiennent à deux sous-populations constituées d'individus phénotypiquement distincts. Cependant en analysant la répartition des gènes du génome accessoire, les souches productrices (souches du sous-clade II, II et la souche S1A1-7) ne partagent pas un groupe de gènes qui leur est spécifique par rapport aux souches non productrices (Sous-clade I et la souche RLB3-17) (**Figure C2**). Donc les souches productrices ne formeraient pas une sous-population en processus de différenciation par rapport aux souches non productrices.

VIII.1.2 Utilisation d'une approche sans *à priori* sur les caractères écologiques

L'approche sans prédéfinition de caractère écologique peut également être utilisée sur nos souches. Dans ce cas l'écologie reverse stipule d'identifier les clades phylogéniques au sein de la population (trois dans notre cas). Si la répartition des gènes (accessoire) au sein du génome est congruente avec la phylogénie, les individus étudiés forment des sous-populations différenciées, sinon ces individus appartiendraient à une même population. Dans notre échantillonnage la répartition des gènes accessoires est congruente avec la phylogénie (**Figure C2**); les souches appartenant au même sous-clades phylogénétique possèdent de nombreux gènes en commun spécifiques de chaque sous-clade. Ces souches pourraient donc former différentes sous-populations.

L'analyse du rapport du taux de recombinaison sur la mutation (r/s) calculé sur un alignement des séquences de 5 gènes de ménage (**Figure S1, Annexes**), a permis d'obtenir une valeur supérieure à 1 ($r/s > 2,3$). Dans leur étude, Fraser *et al.*, (2009) ont fixé le point de rupture de r/s à 0,53. En dessous de cette valeur, deux sous-populations auront tendance à diverger et devenir clonales suite à l'interruption de la recombinaison homologue (courbe verte **Figure C5**); cette faible recombinaison entre les souches maintient la divergence. Lorsque le taux de recombinaison augmente, les échanges génétiques entre les individus permettent d'homogénéiser les génomes. Cela permet de diminuer la divergence entre les populations. Au-dessus du point de rupture, les populations pourraient converger et la recombinaison pourra les faire fusionner de nouveau pour reformer une seule et même population. Le taux de $r/s > 2,3$ suggère que malgré la différenciation qui survient au sein de cette population, les différentes sous-populations formées par ces souches sont capables de recombiner entre elles (populations sexuelles) et quelles pourraient toujours fusionner pour former une seule et même population (**Figure C5**).

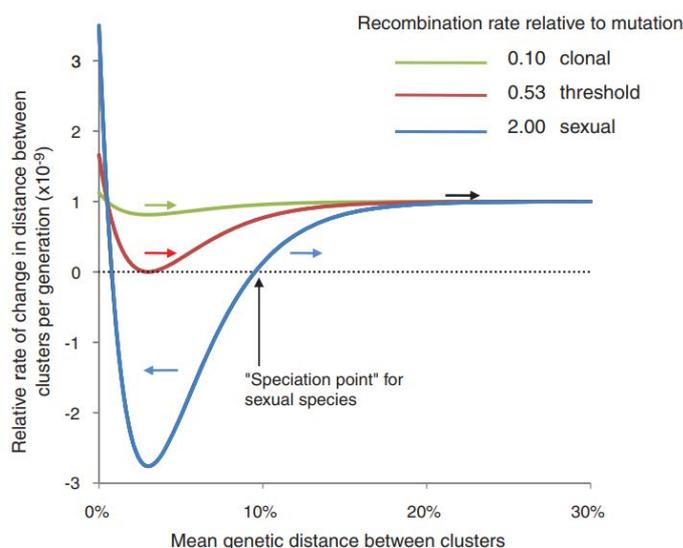


Figure C5 : dynamique de divergence des populations. La figure résume certains résultats clés de Fraser *et al.*, (2007) montrant une dynamique génétique de deux populations au sein desquelles une simulation de la variation du taux de recombinaison a été réalisée.

Le phénomène de convergence des populations différenciées pour reformer une seule et même population avait précédemment été décrit chez les bactéries du genre *Campylobacter*. Deux espèces, *Campylobacter jejuni* et *Campylobacter coli*, suite à des échanges de gènes ayant entraîné une homogénéisation de leurs génomes, ne représenteraient plus qu'une même population (Sheppard *et al.*, 2008).

CONCLUSION

Dans cette étude, nous avons montré qu'il existait une grande diversité génétique au sein d'une population bactérienne. L'apparition de cette diversité est rapide compte tenu du temps évolutif très court séparant les souches et elle serait largement dirigée par le transfert d'éléments conjuguatifs. En utilisant comme indicateur du fonctionnement écologique de la population la production différentielle des métabolites spécialisés engendrée par cette diversité, nous avons également montré qu'elle permettait de renforcer la cohésion sociale des souches via des biens communs. Ce travail suggère que les interactions biotiques assurent, par la pression exercée, la cohésion écologique des populations bactériennes.

L'écologie reverse est une approche qui permet d'inférer les processus de différenciation et ou de spéciation des souches dans un environnement en recherchant et en analysant les gènes ayant un intérêt écologique potentiel. Son utilisation pourrait permettre de répondre à certaines questions que soulève l'analyse de cette population. Quel serait le devenir de cette population sur un temps évolutif plus long ? Le processus d'évolution rapide du génome observé chez *Streptomyces* est-il l'apanage de cette bactérie ou est-il commun à toutes les bactéries partageant l'écosystème sol et le mode de vie saprophyte ? et si c'est le cas, les mécanismes à l'œuvre sont-ils les mêmes ?

REFERENCES BIBLIOGRAPHIQUES

- Abby, S.S., Cury, J., Guglielmini, J., Néron, B., Touchon, M., and Rocha, E.P.C. (2016) Identification of protein secretion systems in bacterial genomes. *Scientific Reports* **6**: 23080.
- Achtman, M., Manning, P.A., Kusecek, B., Schwuchow, S., and Willetts, N. (1980) A genetic analysis of F sex factor cistrons needed for surface exclusion in *Escherichia coli*. *J Mol Biol* **138**: 779–795.
- Achtman, M., and Wagner, M. (2008) Microbial diversity and the genetic nature of microbial species. *Nat Rev Microbiol* **6**: 431–440.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002) Site-Specific Recombination. *Molecular Biology of the Cell 4th edition* <https://www.ncbi.nlm.nih.gov/books/NBK26845/>. Accessed May 14, 2019.
- Altenbuchner, J., and Cullum, J. (1985) Structure of an amplifiable DNA sequence in *Streptomyces lividans* 66. *Mol Gen Genet* **201**: 192–197.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410.
- Alvarez-Martinez, C.E., and Christie, P.J. (2009) Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev* **73**: 775–808.
- Ambroset, C., Coluzzi, C., Guédon, G., Devignes, M.-D., Loux, V., Lacroix, T., *et al.* (2015) New Insights into the Classification and Integration Specificity of Streptococcus Integrative Conjugative Elements through Extensive Genome Exploration. *Front Microbiol* **6**: 1483.
- Andreani, N.A., Hesse, E., and Vos, M. (2017) Prokaryote genome fluidity is dependent on effective population size. *The ISME Journal* <http://www.nature.com/doi/10.1038/ismej.2017.36>. Accessed April 10, 2017.
- Angel, R., Soares, M.I.M., Ungar, E.D., and Gillor, O. (2010) Biogeography of soil archaea and bacteria along a steep precipitation gradient. *ISME J* **4**: 553–563.
- Ansari, M.A., and Didelot, X. (2014) Inference of the Properties of the Recombination Process from Whole Bacterial Genomes. *Genetics* **196**: 253–265.
- Antony-Babu, S., Stach, J.E.M., and Goodfellow, M. (2008) Genetic and phenotypic evidence for *Streptomyces griseus* ecovars isolated from a beach and dune sand system. *Antonie Van Leeuwenhoek* **94**: 63–74.
- Antony-Babu, S., Stien, D., Eparvier, V., Parrot, D., Tomasi, S., and Suzuki, M.T. (2017) Multiple *Streptomyces* species with distinct secondary metabolomes have identical 16S rRNA gene sequences. *Scientific Reports* **7**: 11089.
- Arakawa, K., and Tomita, M. (2007) Selection Effects on the Positioning of Genes and Gene Structures from the Interplay of Replication and Transcription in Bacterial Genomes. *Evol Bioinform Online* **3**: 279–286.
- Arber, W. (2000) Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS Microbiol Rev* **24**: 1–7.
- Ari, R. d' (1985) The SOS system. *Biochimie* **67**: 343–347.
- Arya, P., Joseph, R., and Chou, D.T.H. (2002) Toward high-throughput synthesis of complex natural product-like compounds in the genomics and proteomics age. *Chem Biol* **9**: 145–156.

- Avery, O.T., MacLeod, C.M., and McCarty, M. (1944) STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES. *J Exp Med* **79**: 137–158.
- Avni, E., and Snir, S. (2017) Toxic genes present a unique phylogenetic signature. *Mol Phylogenet Evol* **116**: 141–148.
- Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., *et al.* (2008) The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* **9**: 75–75.
- Baas Becking, L.G.M., and Nicolai, E. (1934) On the ecology of a Sphagnum Bog. *Blumea - Biodiversity, Evolution and Biogeography of Plants* **1**: 10–45.
- Baltrus, D.A. (2013) Exploring the costs of horizontal gene transfer. *Trends Ecol Evol (Amst)* **28**: 489–495.
- Bamas-Jacques, N., Lorenzon, S., Lacroix, P., De Swetschin, C., and Crouzet, J. (1999) Cluster organization of the genes of *Streptomyces pristinaespiralis* involved in pristinamycin biosynthesis and resistance elucidated by pulsed-field gel electrophoresis. *Journal of Applied Microbiology* <http://agritrop.cirad.fr/508213/>. Accessed October 6, 2019.
- Bao, K., and Cohen, S.N. (2001) Terminal proteins essential for the replication of linear plasmids and chromosomes in *Streptomyces*. *Genes Dev* **15**: 1518–1527.
- Bao, K., and Cohen, S.N. (2003) Recruitment of terminal protein to the ends of *Streptomyces* linear plasmids and chromosomes by a novel telomere-binding protein essential for linear DNA replication. *Genes Dev* **17**: 774–785.
- Baptiste, E., Lopez, P., Bouchard, F., Baquero, F., McInerney, J.O., and Burian, R.M. (2012) Evolutionary analyses of non-genealogical bonds produced by introgressive descent. *PNAS* **109**: 18266–18272.
- Baril, C., Richaud, C., Baranton, G., and Saint Girons, I.S. (1989) Linear chromosome of *Borrelia burgdorferi*. *Res Microbiol* **140**: 507–516.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., *et al.* (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709–1712.
- Barrick, J.E., and Lenski, R.E. (2013) Genome dynamics during experimental evolution. *Nat Rev Genet* **14**: 827–839.
- Bell, T. (2010) Experimental tests of the bacterial distance–decay relationship. *The ISME Journal* **4**: 1357–1365.
- Bentley, S.D., Chater, K.F., Cerdeño-Tárraga, A.-M., Challis, G.L., Thomson, N.R., James, K.D., *et al.* (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**: 141–147.
- Bérdy, J. (2005) Bioactive microbial metabolites. *J Antibiot* **58**: 1–26.
- Berg, G., and Smalla, K. (2009) Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiol Ecol* **68**: 1–13.
- Bertolla, F., and Simonet, P. (1999) Horizontal gene transfers in the environment: natural transformation as a putative process for gene transfers between transgenic plants and microorganisms. *Res Microbiol* **150**: 375–384.
- Bezbaruah, A.N., and Zhang, T.C. (2002) An innovative electro-corrosion recess creation technique for improved microelectrode fabrication. *Water Research* **36**: 4428–4432.

- Bhatnagar, S.K., and Bessman, M.J. (1988) Studies on the mutator gene, mutT of Escherichia coli. Molecular cloning of the gene, purification of the gene product, and identification of a novel nucleoside triphosphatase. *J Biol Chem* **263**: 8953–8957.
- Bibb, M.J., and Hopwood, D.A. (1981) Genetic Studies of the Fertility Plasmid SCP2 and its SCP2* Variants in *Streptomyces coelicolor* A3(2). *Microbiology* **126**: 427–442.
- Bibb, M.J., Ward, J.M., Kieser, T., Cohen, S.N., and Hopwood, D.A. (1981) Excision of chromosomal DNA sequences from *Streptomyces coelicolor* forms a novel family of plasmids detectable in *Streptomyces lividans*. *Mol Gen Genet* **184**: 230–240.
- Birch, A., Häusler, A., Rüttener, C., and Hütter, R. (1991) Chromosomal deletion and rearrangement in *Streptomyces glaucescens*. *J Bacteriol* **173**: 3531–3538.
- Birch, A., Häusler, A., Vögtli, M., Krek, W., and Hütter, R. (1989) Extremely large chromosomal deletions are intimately involved in genetic instability and genomic rearrangements in *Streptomyces glaucescens*. *Mol Gen Genet* **217**: 447–458.
- Blackburn, E.H. (2001) Switching and Signaling at the Telomere. *Cell* **106**: 661–673.
- Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., et al. (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* **47**: W81–W87.
- Blount, Z.D., Borland, C.Z., and Lenski, R.E. (2008) Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. *Proc Natl Acad Sci USA* **105**: 7899–7906.
- Boccard, F., Smokvina, T., Pernodet, J.-L., Friedmann, A., and Guérineau, M. (1989) Structural analysis of loci involved in pSAM2 site-specific integration in *Streptomyces*. *Plasmid* **21**: 59–70.
- Bontemps, C., Toussaint, M., Revol, P.-V., Hotel, L., Jeanbille, M., Uroz, S., et al. (2013) Taxonomic and functional diversity of *Streptomyces* in a forest soil. *FEMS Microbiol Lett* **342**: 157–167.
- Bordeleau, E., Ghinet, M.G., and Burrus, V. (2012) Diversity of integrating conjugative elements in actinobacteria. *Mob Genet Elements* **2**: 119–124.
- Boucher, Y., Cordero, O.X., Takemura, A., Hunt, D.E., Schliep, K., Baptiste, E., et al. (2011) Local Mobile Gene Pools Rapidly Cross Species Boundaries To Create Endemicity within Global *Vibrio cholerae* Populations. *mBio* **2**: e00335-10.
- Bourras, S., Rouxel, T., and Meyer, M. (2015) Agrobacterium tumefaciens Gene Transfer: How a Plant Pathogen Hacks the Nuclei of Plant and Nonplant Organisms. *Phytopathology* **105**: 1288–1301.
- Brautaset, T., Sekurova, O.N., Sletta, H., Ellingsen, T.E., Strøm, A.R., Valla, S., and Zotchev, S.B. (2000) Biosynthesis of the polyene antifungal antibiotic nystatin in *Streptomyces noursei* ATCC 11455: analysis of the gene cluster and deduction of the biosynthetic pathway. *Chem Biol* **7**: 395–403.
- Bridges, B.A. (2001) Hypermutation in bacteria and other cellular systems. *Philos Trans R Soc Lond, B, Biol Sci* **356**: 29–39.
- Brochier, C., Philippe, H., and Moreira, D. (2000) The evolutionary history of ribosomal protein RpS14: horizontal gene transfer at the heart of the ribosome. *Trends Genet* **16**: 529–533.
- Brolle, D.F., Pape, H., Hopwood, D.A., and Kieser, T. (1993) Analysis of the transfer region of the *Streptomyces* plasmid SCP2. *Mol Microbiol* **10**: 157–170.

- Brown, D.P., Tuan J.S., Boris K.A., DeWitt J.P., Idler K.B., Chiang S.J., and Katz L. (1988b) Plasmid-chromosome interactions in *Saccharopolyspora erythraea* and *Streptomyces lividans*. *Dev Ind Microbiol* **29**:97–105
- Brown, J.R. (2003) Ancient horizontal gene transfer. *Nat Rev Genet* **4**: 121–132.
- Burke, J., Schneider, D., and Westpheling, J. (2001) Generalized transduction in *Streptomyces coelicolor*. *Proc Natl Acad Sci USA* **98**: 6289–6294.
- Cairns, J. (1963) The Chromosome of *Escherichia coli*. *Cold Spring Harb Symp Quant Biol* **28**: 43–46.
- Carson, J.K., Gonzalez-Quiñones, V., Murphy, D.V., Hinz, C., Shaw, J.A., and Gleeson, D.B. (2010) Low pore connectivity increases bacterial diversity in soil. *Appl Environ Microbiol* **76**: 3936–3942.
- Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.-A., Barrell, B.G., and Parkhill, J. (2005) ACT: the Artemis comparison tool. *Bioinformatics* **21**: 3422–3423.
- Casjens, S., Murphy, M., DeLange, M., Sampson, L., Vugt, R. van, and Huang, W.M. (1997) Telomeres of the linear chromosomes of Lyme disease spirochaetes: nucleotide sequence and possible exchange with linear plasmid telomeres. *Mol Microbiol* **26**: 581–596.
- Casjens, S.R., Mongodin, E.F., Qiu, W.-G., Luft, B.J., Schutzer, S.E., Gilcrease, E.B., *et al.* (2012) Genome Stability of Lyme Disease Spirochetes: Comparative Genomics of *Borrelia burgdorferi* Plasmids. *PLoS One* **7** <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3303823/>. Accessed October 13, 2019.
- Castro, H.F., Classen, A.T., Austin, E.E., Norby, R.J., and Schadt, C.W. (2010) Soil Microbial Community Responses to Multiple Experimental Climate Change Drivers. *Appl Environ Microbiol* **76**: 999–1007.
- Challis, G.L., and Naismith, J.H. (2004) Structural aspects of non-ribosomal peptide biosynthesis. *Curr Opin Struct Biol* **14**: 748–756.
- Champness, W. (2000) Actinomycete Development, Antibiotic Production, and Phylogeny: Questions and Challenges. *Prokaryotic Development* 11–31.
- Charlebois, R.L. (ed.) (1999) *Organization of the prokaryotic genome*. ASM Press, Washington, D.C.
- Charlop-Powers, Z., Owen, J.G., Reddy, B.V.B., Ternei, M.A., and Brady, S.F. (2014) Chemical-biogeographic survey of secondary metabolism in soil. *Proc Natl Acad Sci USA* **111**: 3757.
- Chater, K.F. (1993) Genetics of differentiation in *Streptomyces*. *Annu Rev Microbiol* **47**: 685–713.
- Chater, K.F. (2006) *Streptomyces* inside-out: a new perspective on the bacteria that provide us with antibiotics. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**: 761.
- Chater, K.F., Hopwood, D.A. (1984) *Streptomyces* genetics. In: Goodfellow M, Mordarski M, Williams ST (eds) *Biology of the Actinomycetes*. Academic Press, London, pp 229-286
- Chau, J.F., Bagtzoglou, A.C., and Willig, M.R. (2011) The Effect of Soil Texture on Richness and Diversity of Bacterial Communities. *Environmental Forensics* **12**: 333–341.
- Chen, C.W. (1996) Complications and implications of linear bacterial chromosomes. *Trends in Genetics* **12**: 192–196.
- Chen, C.W., Huang, C.-H., Lee, H.-H., Tsai, H.-H., and Kirby, R. (2002) Once the circle has been broken: dynamics and evolution of *Streptomyces* chromosomes. *Trends in Genetics* **18**: 522–529.

- Chen, C.W., Yu, T.-W., Lin, Y.-S., Kieser, H.M., and Hopwood, D.A. (1993) The conjugative plasmid SLP2 of *Streptomyces lividans* is a 50 kb linear molecule. *Molecular Microbiology* **7**: 925–932.
- Chen, I., and Dubnau, D. (2004) DNA uptake during bacterial transformation. *Nat Rev Microbiol* **2**: 241–249.
- Chen, Z., Zhong, L., Shen, M., Fang, P., and Qin, Z. (2012) Characterization of *Streptomyces* plasmid-phage pFP4 and its evolutionary implications. *Plasmid* **68**: 170–178.
- Choulet, F., Aigle, B., Gallois, A., Mangenot, S., Gerbaud, C., Truong, C., *et al.* (2006a) Evolution of the Terminal Regions of the *Streptomyces* Linear Chromosome. *Mol Biol Evol* **23**: 2361–2369.
- Choulet, F., Gallois, A., Aigle, B., Mangenot, S., Gerbaud, C., Truong, C., *et al.* (2006b) Intraspecific Variability of the Terminal Inverted Repeats of the Linear Chromosome of *Streptomyces ambofaciens*. *J Bacteriol* **188**: 6599–6610.
- Chowdhury, N., Marschner, P., and Burns, R. (2011) Response of microbial activity and community structure to decreasing soil osmotic and matric potential. <https://digital.library.adelaide.edu.au/dspace/handle/2440/87959>. Accessed October 4, 2019.
- Christen, B., Abeliuk, E., Collier, J.M., Kalogeraki, V.S., Passarelli, B., Collier, J.A., *et al.* (2011) The essential genome of a bacterium. *Mol Syst Biol* **7**: 528.
- Christie, P.J., Whitaker, N., and González-Rivera, C. (2014) Mechanism and structure of the bacterial type IV secretion systems. *Biochim Biophys Acta* **1843**: 1578–1591.
- Claesson, M.J., Wang, Q., O’Sullivan, O., Greene-Diniz, R., Cole, J.R., Ross, R.P., and O’Toole, P.W. (2010) Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res* **38**: e200.
- Claridge, M.F., Dawah, A.H., and Wilson, M.R. (eds) (1997) *Species: The units of biodiversity*. Springer Netherlands, . <https://www.springer.com/gp/book/9780412631207>. Accessed October 5, 2019.
- Claverys, J.-P., Martin, B., and Håvarstein, L.S. (2007) Competence-induced fratricide in streptococci. *Mol Microbiol* **64**: 1423–1433.
- Cohan, F. (2011) Chapter 5 Are Species Cohesive ? — A View from Bacteriology. .
- Cohan, F.M. (2002a) What are Bacterial Species? *Annual Review of Microbiology* **56**: 457–487.
- Cohan, F.M. (2002b) Sexual isolation and speciation in bacteria. In *Genetics of Mate Choice: From Sexual Selection to Sexual Isolation*. Etges, W.J., and Noor, M.A.F. (eds). Springer Netherlands, Dordrecht. pp. 359–370 https://doi.org/10.1007/978-94-010-0265-3_17. Accessed October 5, 2019.
- Cohan, F.M. (2017) Transmission in the Origins of Bacterial Diversity, From Ecotypes to Phyla. *Microbiol Spectr* **5**.
- Cohan, F.M., and Perry, E.B. (2007) A systematics for discovering the fundamental units of bacterial diversity. *Curr Biol* **17**: R373-386.
- Cordero, O.X., and Datta, M.S. (2016) Microbial interactions and community assembly at microscales. *Current Opinion in Microbiology* **31**: 227–234.
- Cordero, O.X., and Hogeweg, P. (2009) The impact of long-distance horizontal gene transfer on prokaryotic genome size. *PNAS* **106**: 21748–21753.

- Cordero, O.X., and Polz, M.F. (2014) Explaining microbial genomic diversity in light of evolutionary ecology. *Nature Reviews Microbiology* **12**: 263–273.
- Cordero, O.X., Ventouras, L.-A., DeLong, E.F., and Polz, M.F. (2012) Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. *Proc Natl Acad Sci USA* **109**: 20059.
- Croucher, N.J., Coupland, P.G., Stevenson, A.E., Callendrello, A., Bentley, S.D., and Hanage, W.P. (2014) Diversification of bacterial genome content through distinct mechanisms over different timescales. *Nat Commun* **5**: 1–12.
- Cruz, F. de la, Frost, L.S., Meyer, R.J., and Zechner, E.L. (2010) Conjugative DNA metabolism in Gram-negative bacteria. *FEMS Microbiol Rev* **34**: 18–40.
- Cruz-Martínez, K., Suttle, K.B., Brodie, E.L., Power, M.E., Andersen, G.L., and Banfield, J.F. (2009) Despite strong seasonal responses, soil microbial consortia are more resilient to long-term changes in rainfall than overlying grassland. *ISME J* **3**: 738–744.
- Daniel, R. (2005) The metagenomics of soil. *Nat Rev Microbiol* **3**: 470–478.
- Darwin (1809-1882), C. (1859) *On the origin of the species by means of natural selection*. https://data.bnf.fr/fr/12287084/charles_darwin_on_the_origin_of_the_species_by_means_of_natural_selection/. Accessed October 5, 2019.
- Davies, J. (2013) Specialized microbial metabolites: functions and origins. *J Antibiot* **66**: 361–364.
- Dawkins, R. (1981) In Defence of Selfish Genes. *Philosophy* **56**: 556–573.
- Dechesne, A., Pallud, C., Debouzie, D., Flandrois, J.P., Vogel, T.M., Gaudet, J.P., and Grundmann, G.L. (2003) A novel method for characterizing the microscale 3D spatial distribution of bacteria in soil. *Soil Biology and Biochemistry* **35**: 1537–1546.
- Degnen, G.E., and Cox, E.C. (1974) Conditional Mutator Gene in *Escherichia coli*: Isolation, Mapping, and Effector Studies. *J Bacteriol* **117**: 477–487.
- Delmont, T.O., Prestat, E., Keegan, K.P., Faubladier, M., Robe, P., Clark, I.M., *et al.* (2012) Structure, fluctuation and magnitude of a natural grassland soil metagenome. *ISME J* **6**: 1677–1687.
- Demuyter, P., Leblond, P., Decaris, B., and Simonet, J.M. (1988) Characterization of two families of spontaneously amplifiable units of DNA in *Streptomyces ambofaciens*. *J Gen Microbiol* **134**: 2001–2007.
- Derbyshire, K.M., and Gray, T.A. (2014) Distributive Conjugal Transfer: New Insights into Horizontal Gene Transfer and Genetic Exchange in Mycobacteria. *Microbiol Spectr* **2** <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4259119/>. Accessed September 29, 2019.
- DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., *et al.* (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069–5072.
- Di Giambattista, M., Chinali, G., and Cocito, C. (1989) The molecular basis of the inhibitory activities of type A and type B synergimycins and related antibiotics on ribosomes. *J Antimicrob Chemother* **24**: 485–507.
- Ditmarsch, D. van, and Xavier, J.B. (2014) Seeing is believing: what experiments with microbes reveal about evolution. *Trends in Microbiology* **22**: 2–4.

- Dobrindt, U., Hochhut, B., Hentschel, U., and Hacker, J. (2004) Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* **2**: 414–424.
- Donia, M.S., Hathaway, B.J., Sudek, S., Haygood, M.G., Rosovitz, M.J., Ravel, J., and Schmidt, E.W. (2006) Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nat Chem Biol* **2**: 729–735.
- Doolittle, W.F., and Zhaxybayeva, O. (2009) On the origin of prokaryotic species. *Genome Res* **19**: 744–756.
- Doroghazi, J.R., Albright, J.C., Goering, A.W., Ju, K.-S., Haines, R.R., Tchalukov, K.A., *et al.* (2014) A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nature Chemical Biology* **10**: 963–968.
- Doroghazi, J.R., and Metcalf, W.W. (2013) Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics* **14**: 611.
- Duchaud, E., Rusniok, C., Frangeul, L., Buchrieser, C., Givaudan, A., Taourit, S., *et al.* (2003) The genome sequence of the entomopathogenic bacterium *Photobacterium luminescens*. *Nat Biotechnol* **21**: 1307–1313.
- Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., *et al.* (2004) Genome evolution in yeasts. *Nature* **430**: 35–44.
- Dunny, G.M. (1990) Genetic functions and cell-cell interactions in the pheromone-inducible plasmid transfer system of *Enterococcus faecalis*. *Mol Microbiol* **4**: 689–696.
- Dutta, S., Whicher, J.R., Hansen, D.A., Hale, W.A., Chemler, J.A., Congdon, G.R., *et al.* (2014) Structure of a modular polyketide synthase. *Nature* **510**: 512–517.
- Eldredge, N. & G. (1972) *Punctuated equilibria: an alternative to phyletic gradualism* // (1972) pp 82-115 in "Models in paleobiology", edited by Schopf, TJM Freeman, Cooper & Co, San Francisco. <http://archive.org/details/B-001-004-118>. Accessed October 5, 2019.
- Eldredge, N., Thompson, J.N., Brakefield, P.M., Gavrillets, S., Jablonski, D., Jackson, J.B.C., *et al.* (2005) The dynamics of evolutionary stasis. *Paleobiology* **31**: 133–145.
- Elliot, M.A., Buttner, M.J., and Nodwell, J.R. (2008) 24 Multicellular Development in *Streptomyces*. *Myxobacteria* 419–438.
- Ellis, R.J. (2004) Artificial soil microcosms: a tool for studying microbial autecology under controlled conditions. *J Microbiol Methods* **56**: 287–290.
- Faucher, M., Nouvel, L.-X., Dordet-Frisoni, E., Sagné, E., Baranowski, E., Hygonenq, M.-C., *et al.* (2019) Mycoplasmas under experimental antimicrobial selection: The unpredicted contribution of horizontal chromosomal transfer. *PLoS Genet* **15**: e1007910.
- Fiegna, F., and Velicer, G.J. (2005) Exploitative and Hierarchical Antagonism in a Cooperative Bacterium. *PLOS Biology* **3**: e370.
- Fierer, N., Lauber, C.L., Ramirez, K.S., Zaneveld, J., Bradford, M.A., and Knight, R. (2012) Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *ISME J* **6**: 1007–1017.

- Finking, R., and Marahiel, M.A. (2004) Biosynthesis of nonribosomal peptides¹. *Annu Rev Microbiol* **58**: 453–488.
- Fischbach, M.A., Walsh, C.T., and Clardy, J. (2008) The evolution of gene collectives: How natural selection drives chemical innovation. *Proc Natl Acad Sci USA* **105**: 4601–4608.
- Fischer, G., Decaris, B., and Leblond, P. (1997) Occurrence of deletions, associated with genetic instability in *Streptomyces ambofaciens*, is independent of the linearity of the chromosomal DNA. *J Bacteriol* **179**: 4553–4558.
- Fischer, G., Neuvéglise, C., Durrens, P., Gaillardin, C., and Dujon, B. (2001) Evolution of gene order in the genomes of two related yeast species. *Genome Res* **11**: 2009–2019.
- Fischer, G., Wenner, T., Decaris, B., and Leblond, P. (1998) Chromosomal arm replacement generates a high level of intraspecific polymorphism in the terminal inverted repeats of the linear chromosomal DNA of *Streptomyces ambofaciens*. *PNAS* **95**: 14296–14301.
- Fisher, R.A. (1931) XVII.—The Distribution of Gene Ratios for Rare Mutations. *Proceedings of the Royal Society of Edinburgh* **50**: 204–219.
- Flatt, P.M., and Mahmud, T. (2007) Biosynthesis of aminocyclitol-aminoglycoside antibiotics and related compounds. *Nat Prod Rep* **24**: 358–392.
- Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496–512.
- Fox, G.E., Magrum, L.J., Balch, W.E., Wolfe, R.S., and Woese, C.R. (1977) Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proc Natl Acad Sci USA* **74**: 4537–4541.
- Fraser, C., Alm, E.J., Polz, M.F., Spratt, B.G., and Hanage, W.P. (2009) The bacterial species challenge: making sense of genetic and ecological diversity. *Science* **323**: 741–746.
- Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., *et al.* (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**: 397–403.
- Fronzes, R., Christie, P.J., and Waksman, G. (2009) The structural biology of type IV secretion systems. *Nat Rev Microbiol* **7**: 703–714.
- Frost, L.S., Ippen-Ihler, K., and Skurray, R.A. (1994) Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol Rev* **58**: 162–210.
- Furuya, E.Y., and Lowy, F.D. (2006) Antimicrobial-resistant bacteria in the community setting. *Nature Reviews Microbiology* **4**: 36–45.
- Garcia, R., Gemperlein, K., and Müller, R. (2014) *Minicystis rosea* gen. nov., sp. nov., a polyunsaturated fatty acid-rich and steroid-producing soil myxobacterium. *International Journal of Systematic and Evolutionary Microbiology* **64**: 3733–3742.
- Garcia, S.L., Stevens, S.L.R., Crary, B., Martinez-Garcia, M., Stepanauskas, R., Woyke, T., *et al.* (2018) Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations. *The ISME Journal* **12**: 742–755.
- Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., *et al.* (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**: 498–511.

- Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., *et al.* (2005) Opinion: Re-evaluating prokaryotic species. *Nat Rev Microbiol* **3**: 733–739.
- Ghinet, M.G., Bordeleau, E., Beaudin, J., Brzezinski, R., Roy, S., and Burrus, V. (2011) Uncovering the Prevalence and Diversity of Integrating Conjugative Elements in Actinobacteria. *PLOS ONE* **6**: e27846.
- Gogarten, J.P., Doolittle, W.F., and Lawrence, J.G. (2002) Prokaryotic Evolution in Light of Gene Transfer. *Mol Biol Evol* **19**: 2226–2238.
- Gómez, P., and Buckling, A. (2011) Bacteria-Phage Antagonistic Coevolution in Soil. *Science* **332**: 106–109.
- Gonod, L.V., Martin-Laurent, F., and Chenu, C. (2006) 2,4-D impact on bacterial communities, and the activity and genetic potential of 2,4-D degrading communities in soil. *FEMS Microbiology Ecology* **58**: 529–537.
- Goodfellow, M., Freeman, R., and Sisson, P.R. (1997) Curie-point pyrolysis mass spectrometry as a tool in clinical microbiology. *Zentralbl Bakteriol* **285**: 133–156.
- Gray, T.A., and Derbyshire, K.M. (2018) Blending genomes: distributive conjugal transfer in mycobacteria, a sexier form of HGT. *Molecular Microbiology* **108**: 601–613.
- Griffith, Fred. (1928) The Significance of Pneumococcal Types. *J Hyg (Lond)* **27**: 113–159.
- Griffiths, R.I., Thomson, B.C., James, P., Bell, T., Bailey, M., and Whiteley, A.S. (2011) The bacterial biogeography of British soils. *Environ Microbiol* **13**: 1642–1654.
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**: 172.
- Grundmann, G.L. (2004) Spatial scales of soil bacterial diversity--the size of a clone. *FEMS Microbiol Ecol* **48**: 119–127.
- Grundmann, G.L., Dechesne, A., Bartoli, F., Flandrois, J.P., Chassé, J.L., and Kizungu, R. (2001) Spatial Modeling of Nitrifier Microhabitats in Soil. *Soil Science Society of America Journal* **65**: 1709–1716.
- Grundmann, G.L., and Normand, P. (2000) Microscale Diversity of the Genus *Nitrobacter* in Soil on the Basis of Analysis of Genes Encoding rRNA. *Appl Environ Microbiol* **66**: 4543–4546.
- Guenet, B., Leloup, J., Hartmann, C., Barot, S., and Abbadie, L. (2011) A new protocol for an artificial soil to analyse soil microbiological processes. *Applied Soil Ecology* **48**: 243–246.
- Guglielmini, J., Néron, B., Abby, S.S., Garcillán-Barcia, M.P., Cruz, F. de la, and Rocha, E.P.C. (2014) Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. *Nucleic Acids Res* **42**: 5715–5727.
- Hacker, J., Blum-Oehler, G., Mühldorfer, I., and Tschäpe, H. (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* **23**: 1089–1097.
- Hacker, J., and Kaper, J.B. (2000) Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* **54**: 641–679.
- Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* **1**: e60.

- Hagège, J., Boccard, F., Smokvina, T., Pernodet, J.L., Friedmann, A., and Guérineau, M. (1994) Identification of a gene encoding the replication initiator protein of the *Streptomyces* integrating element, pSAM2. *Plasmid* **31**: 166–183.
- Hagège, J., Pernodet, J.L., Friedmann, A., and Guérineau, M. (1993a) Mode and origin of replication of pSAM2, a conjugative integrating element of *Streptomyces ambofaciens*. *Mol Microbiol* **10**: 799–812.
- Hagège, J., Pernodet, J.L., Sezonov, G., Gerbaud, C., Friedmann, A., and Guérineau, M. (1993b) Transfer functions of the conjugative integrating element pSAM2 from *Streptomyces ambofaciens*: characterization of a kil-kor system associated with transfer. *J Bacteriol* **175**: 5529–5538.
- Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B.H. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506.
- Hanson, J.R. (2003) Natural products: the secondary metabolites. *Natural products: the secondary metabolites*. <https://www.cabdirect.org/cabdirect/abstract/20033159876>. Accessed October 6, 2019.
- Hattori, T. (Tohoku U. (1988) Soil aggregates as microhabitats of microorganisms. *Reports of the Institute for Agricultural Research - Tohoku University (Japan)*. <http://agris.fao.org/agris-search/search.do?recordID=JP8903709>. Accessed October 4, 2019.
- Haug, I., Weissenborn, A., Brolle, D., Bentley, S., Kieser, T., and Altenbuchner, J. (2003) *Streptomyces coelicolor* A3(2) plasmid SCP2*: deductions from the complete sequence. *Microbiology (Reading, Engl)* **149**: 505–513.
- Haynes, R.J., and Swift, R.S. (1990) Stability of soil aggregates in relation to organic constituents and soil water content. *Journal of Soil Science* **41**: 73–83.
- Heister, K., Höschel, C., Pronk, G.J., Mueller, C.W., and Kögel-Knabner, I. (2012) NanoSIMS as a tool for characterizing soil model compounds and organomineral associations in artificial soils. *J Soils Sediments* **12**: 35–47.
- Herbig, U., Jobling, W.A., Chen, B.P.C., Chen, D.J., and Sedivy, J.M. (2004) Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21(CIP1), but not p16(INK4a). *Mol Cell* **14**: 501–513.
- Hieter, P., Mann, C., Snyder, M., and Davis, R.W. (1985) Mitotic stability of yeast chromosomes: a colony color assay that measures nondisjunction and chromosome loss. *Cell* **40**: 381–392.
- Hiller, N.L., Ahmed, A., Powell, E., Martin, D.P., Eutsey, R., Earl, J., *et al.* (2010) Generation of genic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal pediatric infection. *PLoS Pathog* **6**: e1001108.
- Hoff, G., Bertrand, C., Piotrowski, E., Thibessard, A., and Leblond, P. (2018) Genome plasticity is governed by double strand break DNA repair in *Streptomyces*. *Scientific Reports* **8**: 5272.
- Hsu, C.-C., and Chen, C.W. (2010) Linear Plasmid SLP2 Is Maintained by Partitioning, Intrahyphal Spread, and Conjugal Transfer in *Streptomyces*. *Journal of Bacteriology* **192**: 307–315.
- Huang, C.-H., Chen, C.-Y., Tsai, H.-H., Chen, C., Lin, Y.-S., and Chen, C.W. (2003) Linear plasmid SLP2 of *Streptomyces lividans* is a composite replicon. *Mol Microbiol* **47**: 1563–1576.
- Huang, C.H., Lin, Y.S., Yang, Y.L., Huang, S.W., and Chen, C.W. (1998) The telomeres of *Streptomyces* chromosomes contain conserved palindromic sequences with potential to form complex secondary structures. *Mol Microbiol* **28**: 905–916.

- Hugenholtz, P. (2002) Exploring prokaryotic diversity in the genomic era. *Genome Biol* **3**: reviews0003.1-reviews0003.8.
- Huse, S.M., Dethlefsen, L., Huber, J.A., Welch, D.M., Relman, D.A., and Sogin, M.L. (2008) Exploring Microbial Diversity and Taxonomy Using SSU rRNA Hypervariable Tag Sequencing. *PLoS Genetics* **4**: e1000255.
- Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., *et al.* (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol* **21**: 526–531.
- Inoue, S., Higashiyama, K., Uchida, T., Hiratsu, K., and Kinashi, H. (2003) Chromosomal circularization in *Streptomyces griseus* by nonhomologous recombination of deletion ends. *Biosci Biotechnol Biochem* **67**: 1101–1108.
- Itahana, K., Campisi, J., and Dimri, G.P. (2004) Mechanisms of cellular senescence in human and mouse cells. *Biogerontology* **5**: 1–10.
- Jenke-Kodama, H., Börner, T., and Dittmann, E. (2006) Natural biocombinatorics in the polyketide synthase genes of the actinobacterium *Streptomyces avermitilis*. *PLoS Computational Biology* **2**: 1210–1218.
- Jensen, P.R., Williams, P.G., Oh, D.-C., Zeigler, L., and Fenical, W. (2007) Species-specific secondary metabolite production in marine actinomycetes of the genus *Salinispora*. *Appl Environ Microbiol* **73**: 1146–1152.
- Johnsborg, O., Eldholm, V., and Håvarstein, L.S. (2007) Natural genetic transformation: prevalence, mechanisms and function. *Res Microbiol* **158**: 767–778.
- Juhas, M., Meer, J.R. van der, Gaillard, M., Harding, R.M., Hood, D.W., and Crook, D.W. (2009) Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* **33**: 376–393.
- Juyal, A., Eickhorst, T., Falconer, R., Baveye, P.C., Spiers, A., and Otten, W. (2018) Control of Pore Geometry in Soil Microcosms and Its Effect on the Growth and Spread of *Pseudomonas* and *Bacillus* sp. *Front Environ Sci* **6** <https://www.frontiersin.org/articles/10.3389/fenvs.2018.00073/full>. Accessed October 4, 2019.
- Kameoka, D., Lezhava, A., Zenitani, H., Hiratsu, K., Kawamoto, M., Goshi, K., *et al.* (1999) Analysis of Fusion Junctions of Circularized Chromosomes in *Streptomyces griseus*. *J Bacteriol* **181**: 5711–5717.
- Kataoka, M., Seki, T., and Yoshida, T. (1991) Five genes involved in self-transmission of pSN22, a *Streptomyces* plasmid. *J Bacteriol* **173**: 4220–4228.
- Katz, L., Brown, D.P., and Donadio, S. (1991) Site-specific recombination in *Escherichia coli* between theatt sites of plasmid pSE211 from *Saccharopolyspora erythraea*. *Molec Gen Genet* **227**: 155–159.
- Kawecki, T.J., Lenski, R.E., Ebert, D., Hollis, B., Olivieri, I., and Whitlock, M.C. (2012) Experimental evolution. *Trends Ecol Evol (Amst)* **27**: 547–560.
- Kemper, W.D., and Rosenau, R.C. (1986) Aggregate Stability and Size Distribution 1. *Methods of Soil Analysis: Part 1—Physical and Mineralogical Methods sssabookseries*: 425–442.
- Kendall, K.J., and Cohen, S.N. (1987) Plasmid transfer in *Streptomyces lividans*: identification of a kil-kor system associated with the transfer region of pIJ101. *J Bacteriol* **169**: 4177–4183.

- Keymer, J.E., Galajda, P., Muldoon, C., Park, S., and Austin, R.H. (2006) Bacterial metapopulations in nanofabricated landscapes. *PNAS* **103**: 17290–17295.
- Kieser, T., Hopwood, D.A., Wright, H.M., and Thompson, C.J. (1982) pIJ101, a multi-copy broad host-range *Streptomyces* plasmid: functional analysis and development of DNA cloning vectors. *Mol Gen Genet* **185**: 223–228.
- Kimura, M. (1986) DNA and the neutral theory. *Philos Trans R Soc Lond, B, Biol Sci* **312**: 343–354.
- Kimura, M., and Ohta, T. (1974) On Some Principles Governing Molecular Evolution*. *Proc Natl Acad Sci U S A* **71**: 2848–2852.
- Kinashi, H., Shimaji, M., and Sakai, A. (1987) Giant linear plasmids in *Streptomyces* which code for antibiotic biosynthesis genes. *Nature* **328**: 454–456.
- Koeppel, A., Perry, E.B., Sikorski, J., Krizanc, D., Warner, A., Ward, D.M., *et al.* (2008) Identifying the fundamental units of bacterial diversity: A paradigm shift to incorporate ecology into bacterial systematics. *Proc Natl Acad Sci U S A* **105**: 2504–2509.
- Konstantinidis, K.T., Ramette, A., and Tiedje, J.M. (2006) The bacterial species definition in the genomic era. *Philos Trans R Soc Lond B Biol Sci* **361**: 1929–1940.
- Koonin, E.V. (2016) Horizontal gene transfer: essentiality and evolvability in prokaryotes, and roles in evolutionary transitions. *F1000Res* **5**.
- Koyama, A., Wallenstein, M.D., Simpson, R.T., and Moore, J.C. (2014) Soil bacterial community composition altered by increased nutrient availability in Arctic tundra soils. *Front Microbiol* **5**: 516.
- Kraemer, S.A., Toups, M.A., and Velicer, G.J. (2010) Natural variation in developmental life-history traits of the bacterium *Myxococcus xanthus*. *FEMS Microbiol Ecol* **73**: 226–233.
- Krug, D., Zurek, G., Revermann, O., Vos, M., Velicer, G.J., and Müller, R. (2008) Discovering the Hidden Secondary Metabolome of *Myxococcus xanthus*: a Study of Intraspecific Diversity. *Appl Environ Microbiol* **74**: 3058–3068.
- Krzywinski, M.I., Schein, J.E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., *et al.* (2009) Circos: An information aesthetic for comparative genomics. *Genome Res.* <http://genome.cshlp.org/content/early/2009/06/15/gr.092759.109>. Accessed September 20, 2019.
- Kuramae, E.E., Yergeau, E., Wong, L.C., Pijl, A.S., Veen, J.A. van, and Kowalchuk, G.A. (2012) Soil characteristics more strongly influence soil bacterial communities than land-use type. *FEMS Microbiol Ecol* **79**: 12–24.
- Kuzyakov, Y., and Blagodatskaya, E. (2015) Microbial hotspots and hot moments in soil: Concept & review. *Soil Biology and Biochemistry* **83**: 184–199.
- Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M.R., Ahn, T.-H., *et al.* (2015) Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics* **15**: 141–161.
- Langille, M.G.I., Hsiao, W.W.L., and Brinkman, F.S.L. (2010) Detecting genomic islands using bioinformatics approaches. *Nat Rev Microbiol* **8**: 373–382.
- Lapierre, P., and Gogarten, J.P. (2009) Estimating the size of the bacterial pan-genome. *Trends Genet* **25**: 107–110.

- Lassalle, F., Muller, D., and Nesme, X. (2015) Ecological speciation in bacteria: reverse ecology approaches reveal the adaptive part of bacterial cladogenesis. *Research in Microbiology* **166**: 729–741.
- Lauber, C.L., Hamady, M., Knight, R., and Fierer, N. (2009) Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale. *Appl Environ Microbiol* **75**: 5111–5120.
- Lawrence, J.G., Hendrix, R.W., and Casjens, S. (2001) Where are the pseudogenes in bacterial genomes? *Trends Microbiol* **9**: 535–540.
- Lawrence, J.G., and Ochman, H. (1998) Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci USA* **95**: 9413–9417.
- Lázár, V., Martins, A., Spohn, R., Daruka, L., Grézal, G., Fekete, G., *et al.* (2018) Antibiotic-resistant bacteria show widespread collateral sensitivity to antimicrobial peptides. *Nat Microbiol* **3**: 718–731.
- Leblond, P., Demuyter, P., Moutier, L., Laakel, M., Decaris, B., and Simonet, J.M. (1989) Hypervariability, a new phenomenon of genetic instability, related to DNA amplification in *Streptomyces ambofaciens*. *Journal of Bacteriology* **171**: 419–423.
- Leblond, P., Demuyter, P., Simonet, J.M., and Decaris, B. (1990) Genetic instability and hypervariability in *Streptomyces ambofaciens*: towards an understanding of a mechanism of genome plasticity. *Molecular Microbiology* **4**: 707–714.
- Leblond, P., Fischer, G., Francou, F.X., Berger, F., Guérineau, M., and Decaris, B. (1996) The unstable region of *Streptomyces ambofaciens* includes 210 kb terminal inverted repeats flanking the extremities of the linear chromosomal DNA. *Mol Microbiol* **19**: 261–271.
- Lederberg, J., Lederberg, E.M., Zinder, N.D., and Lively, E.R. (1951) Recombination analysis of bacterial heredity. *Cold Spring Harb Symp Quant Biol* **16**: 413–443.
- Lederberg, J., and Tatum, E.L. (1946) Gene Recombination in *Escherichia Coli*. *Nature* **158**: 558.
- Lee, H.-H., Hsu, C.-C., Lin, Y.-L., and Chen, C.W. (2011) Linear plasmids mobilize linear but not circular chromosomes in *Streptomyces*: support for the ‘end first’ model of conjugal transfer. *Microbiology* **157**: 2556–2568.
- Leff, J.W., Jones, S.E., Prober, S.M., Barberán, A., Borer, E.T., Firn, J.L., *et al.* (2015) Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. *PNAS* **112**: 10967–10972.
- Letzel, A.-C., Pidot, S.J., and Hertweck, C. (2014) Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) in anaerobic bacteria. *BMC Genomics* **15**: 983.
- Levade, I., Terrat, Y., Leducq, J.-B., Weil, A.A., Mayo-Smith, L.M., Chowdhury, F., *et al.* (2017) *Vibrio cholerae* genomic diversity within and between patients. *Microb Genom* **3** <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5761273/>. Accessed October 6, 2019.
- Levy, R., and Borenstein, E. (2012) Reverse Ecology: from systems to environments and back. *Adv Exp Med Biol* **751**: 329–345.
- Li, Y.F., Costello, J.C., Holloway, A.K., and Hahn, M.W. (2008) “REVERSE ECOLOGY” AND THE POWER OF POPULATION GENOMICS. *Evolution* **62**: 2984–2994.

- Lin, Y.S., Kieser, H.M., Hopwood, D.A., and Chen, C.W. (1993) The chromosomal DNA of *Streptomyces lividans* 66 is linear. *Mol Microbiol* **10**: 923–933.
- Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., *et al.* (2007) An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* **445**: 915–918.
- Liu, F., Garneau, S., and Walsh, C.T. (2004) Hybrid nonribosomal peptide-polyketide interfaces in epothilone biosynthesis: minimal requirements at N and C termini of EpoB for elongation. *Chem Biol* **11**: 1533–1542.
- Liu, L., Johnson, H.L., Cousens, S., Perin, J., Scott, S., Lawn, J.E., *et al.* (2012) Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet* **379**: 2151–2161.
- Lloyd, R.G., and Buckman, C. (1995) Conjugational recombination in *Escherichia coli*: genetic analysis of recombinant formation in Hfr x F- crosses. *Genetics* **139**: 1123–1148.
- Lorenz, M.G., and Wackernagel, W. (1994) Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol Rev* **58**: 563–602.
- Low, K.B. (2001) Conjugation. In *Encyclopedia of Genetics*. Brenner, S., and Miller, J.H. (eds). Academic Press, New York. pp. 449–453 <http://www.sciencedirect.com/science/article/pii/B0122270800002676>. Accessed May 13, 2019.
- Luria, S.E., and Delbrück, M. (1943) Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics* **28**: 491–511.
- McMurry, J.E. (2015) Organic chemistry with biological applications. In: *Secondary Metabolites: An Introduction to Natural Products Chemistry*. Stamford, USA: Cengage Learning Ltd; pp. 1016–1046.
- Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., *et al.* (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* **95**: 3140–3145.
- Majewski, J., Zawadzki, P., Pickerill, P., Cohan, F.M., and Dowson, C.G. (2000) Barriers to Genetic Exchange between Bacterial Species: *Streptococcus pneumoniae* Transformation. *Journal of Bacteriology* **182**: 1016–1023.
- Männik, J., Driessen, R., Galajda, P., Keymer, J.E., and Dekker, C. (2009) Bacterial growth and motility in sub-micron constrictions. *PNAS* **106**: 14861–14866.
- Mao, E.F., Lane, L., Lee, J., and Miller, J.H. (1997) Proliferation of mutators in A cell population. *J Bacteriol* **179**: 417–422.
- Marraffini, L.A., and Sontheimer, E.J. (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**: 1843–1845.
- Martin, C., Mazodier, P., Mediola, M.V., Gicquel, B., Smokvina, T., Thompson, C.J., and Davies, J. (1991) Site-specific integration of the *Streptomyces* plasmid pSAM2 in *Mycobacterium smegmatis*. *Molecular Microbiology* **5**: 2499–2502.
- Masschelein, J., Jenner, M., and Challis, G.L. (2017) Antibiotics from Gram-negative bacteria: a comprehensive overview and selected biosynthetic highlights. *Nat Prod Rep* **34**: 712–783.

- Mayr, E. (1999) *Systematics and the Origin of Species, from the Viewpoint of a Zoologist*. Harvard University Press, .
- McInerney, J.O., McNally, A., and O'Connell, M.J. (2017) Why prokaryotes have pangenomes. *Nat Microbiol* **2**: 17040.
- Michaels, M.L., Pham, L., Nghiem, Y., Cruz, C., and Miller, J.H. (1990) MutY, an adenine glycosylase active on G-A mispairs, has homology to endonuclease III. *Nucleic Acids Res* **18**: 3841–3845.
- Miller, D.A., Luo, L., Hillson, N., Keating, T.A., and Walsh, C.T. (2002) Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of *Yersinia pestis*. *Chem Biol* **9**: 333–344.
- Moëgne-Loccoz, Y., Mavingui, P., Combes, C., Normand, P., and Steinberg, C. (2015) Microorganisms and Biotic Interactions. In *Environmental Microbiology: Fundamentals and Applications: Microbial Ecology*. Bertrand, J.-C., Caumette, P., Lebaron, P., Matheron, R., Normand, P., and Sime-Ngando, T. (eds). Springer Netherlands, Dordrecht. pp. 395–444 https://doi.org/10.1007/978-94-017-9118-2_11. Accessed May 13, 2019.
- Morgan, A.D., MacLean, R.C., Hillesland, K.L., and Velicer, G.J. (2010) Comparative Analysis of Myxococcus Predation on Soil Bacteria. *Appl Environ Microbiol* **76**: 6920–6927.
- Morris, J.J., Lenski, R.E., and Zinser, E.R. (2012) The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss. *mBio* **3**: e00036-12.
- Mortel, M. van de, and Halverson, L.J. (2004) Cell envelope components contributing to biofilm growth and survival of *Pseudomonas putida* in low-water-content habitats. *Mol Microbiol* **52**: 735–750.
- Muth, G., Farr, M., Hartmann, V., and Wohlleben, W. (1995) *Streptomyces ghanaensis* plasmid pSG5: nucleotide sequence analysis of the self-transmissible minimal replicon and characterization of the replication mode. *Plasmid* **33**: 113–126.
- Muth, G., Wohlleben, W., and Pühler, A. (1988) The minimal replicon of the *Streptomyces ghanaensis* plasmid pSG5 identified by subcloning and Tn5 mutagenesis. *Molec Gen Genet* **211**: 424–429.**
- Myronovskyi, M., Tokovenko, B., Brötz, E., Rückert, C., Kalinowski, J., and Luzhetskyy, A. (2014) Genome rearrangements of *Streptomyces albus* J1074 lead to the carotenoid gene cluster activation. *Appl Microbiol Biotechnol* **98**: 795–806.
- Nakamura, N.H., Rosell, D.R., Akama, K.T., and McEwen, B.S. (2004) Estrogen and ovariectomy regulate mRNA and protein of glutamic acid decarboxylases and cation-chloride cotransporters in the adult rat hippocampus. *Neuroendocrinology* **80**: 308–323.
- Niehus, R., Mitri, S., Fletcher, A.G., and Foster, K.R. (2015) Migration and horizontal gene transfer divide microbial genomes into multiple niches. *Nat Commun* **6**: 8924.
- Nielsen, K.M., Bones, A.M., Smalla, K., and Elsas, J.D. van (1998) Horizontal gene transfer from transgenic plants to terrestrial bacteria--a rare event? *FEMS Microbiol Rev* **22**: 79–103.
- Nunan, N., Wu, K., Young, I.M., Crawford, J.W., and Ritz, K. (2003) Spatial distribution of bacterial communities and their relationships with the micro-architecture of soil. *FEMS Microbiol Ecol* **44**: 203–215.
- Ochman, H., Elwyn, S., and Moran, N.A. (1999) Calibrating bacterial evolution. *PNAS* **96**: 12638–12643.

- Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299–304.
- Ochman, H., and Moran, N.A. (2001) Genes Lost and Genes Found: Evolution of Bacterial Pathogenesis and Symbiosis. *Science* **292**: 1096–1099.
- Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., *et al.* (2008) Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J Bacteriol* **190**: 4050–4060.
- Okabe, S., Satoh, H., and Watanabe, Y. (1999) In Situ Analysis of Nitrifying Biofilms as Determined by In Situ Hybridization and the Use of Microelectrodes. *Appl Environ Microbiol* **65**: 3182–3191.
- Oliveira, H., Vilas Boas, D., Mesnage, S., Kluskens, L.D., Lavigne, R., Sillankorva, S., *et al.* (2016) Structural and Enzymatic Characterization of ABgp46, a Novel Phage Endolysin with Broad Anti-Gram-Negative Bacterial Activity. *Front Microbiol* **7** <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4768612/>. Accessed May 14, 2019.
- Oliver, A., Cantón, R., Campo, P., Baquero, F., and Blázquez, J. (2000) High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science* **288**: 1251–1254.
- Oliver, J.D. (2005) The viable but nonculturable state in bacteria. *Journal of microbiology* **43**: 93–100.
- Ortega, M.A., and Donk, W.A. van der (2016) New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products. *Cell Chem Biol* **23**: 31–44.
- Paget, E., and Simonet, P. (1994) On the track of natural transformation in soil. *FEMS Microbiol Ecol* **15**: 109–117.
- Pál, C., Papp, B., and Lercher, M.J. (2005) Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nature Genetics* **37**: 1372.
- Pandza, S., Biuković, G., Paravić, A., Dadbin, A., Cullum, J., and Hranueli, D. (1998) Recombination between the linear plasmid pPZG101 and the linear chromosome of *Streptomyces rimosus* can lead to exchange of ends. *Mol Microbiol* **28**: 1165–1176.
- Park, C.J., and Andam, C.P. (2019) Within-Species Genomic Variation and Variable Patterns of Recombination in the Tetracycline Producer *Streptomyces rimosus*. *Front Microbiol* **10** <https://www.frontiersin.org/articles/10.3389/fmicb.2019.00552/full>. Accessed October 6, 2019.
- Patton, G.C., and Donk, W.A. van der (2005) New developments in lantibiotic biosynthesis and mode of action. *Curr Opin Microbiol* **8**: 543–551.
- Pavlovic, G., Burrus, V., Gintz, B., Decaris, B., and Guédon, G. (2004) Evolution of genomic islands by deletion and tandem accretion by site-specific recombination: ICES1-related elements from *Streptococcus thermophilus*. *Microbiology (Reading, Engl)* **150**: 759–774.
- Penn, K., Jenkins, C., Nett, M., Udworthy, D.W., Gontang, E.A., McGlinchey, R.P., *et al.* (2009) Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *ISME J* **3**: 1193–1203.
- Pernodet, J.L., Simonet, J.M., and Guérineau, M. (1984) Plasmids in different strains of *Streptomyces ambofaciens*: free and integrated form of plasmid pSAM2. *Mol Gen Genet* **198**: 35–41.

- Pettis, G.S., and Cohen, S.N. (1994) Transfer of the pIJ101 plasmid in *Streptomyces lividans* requires a cis-acting function dispensable for chromosomal gene transfer. *Molecular Microbiology* **13**: 955–964.
- Poele, E.M. te, Bolhuis, H., and Dijkhuizen, L. (2008) Actinomycete integrative and conjugative elements. *Antonie Van Leeuwenhoek* **94**: 127–143.
- Polz, M.F., Alm, E.J., and Hanage, W.P. (2013) Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends in Genetics* **29**: 170–175.
- Portillo, M.C., Leff, J.W., Lauber, C.L., and Fierer, N. (2013) Cell Size Distributions of Soil Bacterial and Archaeal Taxa. *Appl Environ Microbiol* **79**: 7610–7617.
- Possoz, C., Gagnat, J., Sezonov, G., Guérineau, M., and Pernodet, J. (2003) Conjugal immunity of *Streptomyces* strains carrying the integrative element pSAM2 is due to the pif gene (pSAM2 immunity factor). *Molecular Microbiology* **47**: 1385–1393.
- Possoz, C., Ribard, C., Gagnat, J., Pernodet, J.L., and Guérineau, M. (2001) The integrative element pSAM2 from *Streptomyces*: kinetics and mode of conjugal transfer. *Mol Microbiol* **42**: 159–166.
- Probandt, D., Eickhorst, T., Ellrott, A., Amann, R., and Knittel, K. (2018) Microbial life on a sand grain: from bulk sediment to single grains. *ISME J* **12**: 623–633.
- Queiroz, K. de (1998) The General Lineage Concept of Species, Species Criteria, and the Process of Speciation. In *Endless Forms: Species and Speciation*. Howard, D.J., and Berlocher, S.H. (eds). Oxford University Press, pp. 57–75.
- Quince, C., Curtis, T.P., and Sloan, W.T. (2008) The rational exploration of microbial diversity. *ISME J* **2**: 997–1006.
- Radman, M. (1975) SOS repair hypothesis: phenomenology of an inducible DNA repair which is accompanied by mutagenesis. *Basic Life Sci* **5A**: 355–367.
- Radman, M. (1999) Enzymes of evolutionary change. *Nature* **401**: 866–867, 869.
- Ramírez-Arcos, S., Fernández-Herrero, L.A., Marín, I., and Berenguer, J. (1998) Anaerobic Growth, a Property Horizontally Transferred by an Hfr-Like Mechanism among Extreme Thermophiles. *J Bacteriol* **180**: 3137–3143.
- Ranjard, L., and Richaume, A. (2001) Quantitative and qualitative microscale distribution of bacteria in soil. *Research in Microbiology* **152**: 707–716.
- Raynal, A., Tuphile, K., Gerbaud, C., Luther, T., Guérineau, M., and Pernodet, J.-L. (1998) Structure of the chromosomal insertion site for pSAM2: functional analysis in *Escherichia coli*. *Molecular Microbiology* **28**: 333–342.
- Raynaud, X., and Nunan, N. (2014) Spatial Ecology of Bacteria at the Microscale in Soil. *PLoS One* **9** <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3905020/>. Accessed April 23, 2019.
- Raynes, Y., and Sniegowski, P.D. (2014) Experimental evolution and the dynamics of genomic mutation rate modifiers. *Heredity (Edinb)* **113**: 375–380.
- Regan, K.M., Nunan, N., Boeddinghaus, R.S., Baumgartner, V., Berner, D., Boch, S., *et al.* (2014) Seasonal controls on grassland microbial biogeography: Are they governed by plants, abiotic properties or both? *Soil Biology and Biochemistry* **71**: 21–30.

- Reimmann, C., and Haas, D. (1993) Mobilization of Chromosomes and Nonconjugative Plasmids by Cointegrative Mechanisms. In *Bacterial Conjugation*. Clewell, D.B. (ed.). Springer US, Boston, MA. pp. 137–188 https://doi.org/10.1007/978-1-4757-9357-4_6. Accessed September 29, 2019.
- Reuther, J., Wohlleben, W., and Muth, G. (2006) Modular architecture of the conjugative plasmid pSVH1 from *Streptomyces venezuelae*. *Plasmid* **55**: 201–209.
- Reznick, D.N., Shaw, F.H., Rodd, F.H., and Shaw, R.G. (1997) Evaluation of the Rate of Evolution in Natural Populations of Guppies (*Poecilia reticulata*). *Science* **275**: 1934–1937.
- Richter, C., Chang, J.T., and Fineran, P.C. (2012) Function and Regulation of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) / CRISPR Associated (Cas) Systems. *Viruses* **4**: 2291–2311.
- Riley, M.A., and Lizotte-Waniewski, M. (2009) Population Genomics and the Bacterial Species Concept. *Methods Mol Biol* **532**: 367–377.
- Rocha, E.P.C., and Danchin, A. (2003) Gene essentiality determines chromosome organisation in bacteria. *Nucleic Acids Res* **31**: 6570–6577.
- Roh, S.W., Abell, G.C.J., Kim, K.-H., Nam, Y.-D., and Bae, J.-W. (2010) Comparing microarrays and next-generation sequencing technologies for microbial ecology research. *Trends Biotechnol* **28**: 291–299.
- Rossel, R.A.V., Taylor, H.J., and McBratney, A.B. (2007) Multivariate calibration of hyperspectral γ -ray energy spectra for proximal soil sensing. *European Journal of Soil Science* **58**: 343–353.
- Rosselló-Mora, R., and Amann, R. (2001) The species concept for prokaryotes. *FEMS Microbiol Rev* **25**: 39–67.
- Roth, J.R., Benson, N., Galitski, T., Haack, K., Lawrence, J.G., and Miesel, L. Rearrangements of the Bacterial Chromosome: Formation and Applications. 37.
- Rotman, E., and Kuzminov, A. (2007) The mutT Defect Does Not Elevate Chromosomal Fragmentation in *Escherichia coli* Because of the Surprisingly Low Levels of MutM/MutY-Recognized DNA Modifications. *J Bacteriol* **189**: 6976–6988.
- Ruamps, L.S., Nunan, N., and Chenu, C. (2011) Microbial biogeography at the soil pore scale. *Soil Biology and Biochemistry* **43**: 280–286.
- Samson, J.E., Magadán, A.H., Sabri, M., and Moineau, S. (2013) Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol* **11**: 675–687.
- Schluter, D. (2000) Ecological Character Displacement in Adaptive Radiation. *The American Naturalist* **156**: S4–S16.
- Schmidt, M.W.I., Torn, M.S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I.A., *et al.* (2011) Persistence of soil organic matter as an ecosystem property. *Nature* **478**: 49–56.
- Seipke, R.F. (2015) Strain-Level Diversity of Secondary Metabolism in *Streptomyces albus*. *PLOS ONE* **10**: e0116457.
- Sentchilo, V., Czechowska, K., Pradervand, N., Minoia, M., Miyazaki, R., and Meer, J.R. van der (2009) Intracellular excision and reintegration dynamics of the ICE_{clc} genomic island of *Pseudomonas knackmussii* sp. strain B13. *Mol Microbiol* **72**: 1293–1306.

- Servín-González, L. (1996) Identification and properties of a novel *clt* locus in the *Streptomyces* phaeochromogenes plasmid pJV1. *Journal of Bacteriology* **178**: 4323–4326.
- Servín-González, L., Sampieri, A.I., Cabello, J., Galván, L., Juárez, V., and Castro, C. (1995) Sequence and functional analysis of the *Streptomyces* phaeochromogenes plasmid pJV1 reveals a modular organization of *Streptomyces* plasmids that replicate by rolling circle. *Microbiology (Reading, Engl)* **141** (Pt 10): 2499–2510.
- Sezonov, G., Duchêne, A.M., Friedmann, A., Guérineau, M., and Pernodet, J.L. (1998) Replicase, excisionase, and integrase genes of the *Streptomyces* element pSAM2 constitute an operon positively regulated by the *pra* gene. *J Bacteriol* **180**: 3056–3061.
- Sezonov, G., Hagege, J., Pernodet, J.L., Friedmann, A., and Guerineau, M. (1995) Characterization of *pra*, a gene for replication control in pSAM2, the integrating element of *Streptomyces* ambofaciens. *Molecular microbiology*. <http://agris.fao.org/agris-search/search.do?recordID=US201301510990>. Accessed August 11, 2019.
- Shapiro, B.J. (2016) How clonal are bacteria over time? *Curr Opin Microbiol* **31**: 116–123.
- Shapiro, B.J. (2017) The population genetics of pangenomes. *Nat Microbiol* **2**: 1574–1574.
- Shapiro, B.J., Friedman, J., Cordero, O.X., Preheim, S.P., Timberlake, S.C., Szabó, G., *et al.* (2012) Population genomics of early events in the ecological differentiation of bacteria. *Science* **336**: 48–51.
- Shapiro, B.J., and Polz, M.F. (2014) Ordering microbial diversity into ecologically and genetically cohesive units. *Trends in Microbiology* **22**: 235–247.
- Shen, B. (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr Opin Chem Biol* **7**: 285–295.
- Shen, P., and Huang, H.V. (1986) Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics* **112**: 441–457.
- Sheppard, S.K., McCarthy, N.D., Falush, D., and Maiden, M.C.J. (2008) Convergence of *Campylobacter* species: implications for bacterial evolution. *Science* **320**: 237–239.
- Shou, Q., Feng, L., Long, Y., Han, J., Nunnery, J.K., Powell, D.H., and Butcher, R.A. (2016) A hybrid polyketide-nonribosomal peptide in nematodes that promotes larval survival. *Nat Chem Biol* **12**: 770–772.
- Smillie, C., Garcillán-Barcia, M.P., Francia, M.V., Rocha, E.P.C., and Cruz, F. de la (2010) Mobility of plasmids. *Microbiol Mol Biol Rev* **74**: 434–452.
- Smokvina, T., Boccard, F., Pernodet, J.-L., Friedmann, A., and Guérineau, M. (1991) Functional analysis of the *Streptomyces* ambofaciens element pSAM2. *Plasmid* **25**: 40–52.
- Sniegowski, P.D., Gerrish, P.J., and Lenski, R.E. (1997) Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* **387**: 703–705.
- Sokal, R.R., and Crovello, T.J. (1970) The Biological Species Concept: A Critical Evaluation. *The American Naturalist* **104**: 127–153.
- Sorek, R., Kunin, V., and Hugenholtz, P. (2008) CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* **6**: 181–186.

- Sosio, M., Kloosterman, H., Bianchi, A., Vreugd, P. de, Dijkhuizen, L., and Donadio, S. (2004) Organization of the teicoplanin gene cluster in *Actinoplanes teichomyceticus*. *Microbiology (Reading, Engl)* **150**: 95–102.
- Spatz, K., Köhn, H., and Redenbach, M. (2002) Characterization of the *Streptomyces violaceoruber* SANK95570 plasmids pSV1 and pSV2. *FEMS Microbiology Letters* **213**: 87–92.
- Spence, J.M., Mills, W., Mann, K., Huxley, C., and Farr, C.J. (2006) Increased missegregation and chromosome loss with decreasing chromosome size in vertebrate cells. *Chromosoma* **115**: 60–74.
- Stackebrandt, E., Frederiksen, W., Garrity, G.M., Grimont, P.A.D., Kämpfer, P., Maiden, M.C.J., *et al.* (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* **52**: 1043–1047.
- Stefanic, P., and Mandic-Mulec, I. (2009) Social Interactions and Distribution of *Bacillus subtilis* Pherotypes at Microscale. *Journal of Bacteriology* **191**: 1756–1764.
- Stegen, J.C., Lin, X., Fredrickson, J.K., Chen, X., Kennedy, D.W., Murray, C.J., *et al.* (2013) Quantifying community assembly processes and identifying features that impose them. *ISME J* **7**: 2069–2079.
- Stokes, H.W., and Hall, R.M. (1989) A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Mol Microbiol* **3**: 1669–1683.
- Suzuki, H., Marushima, K., Ohnishi, Y., and Horinouchi, S. (2008) A novel pair of terminal protein and telomere-associated protein for replication of the linear chromosome of *Streptomyces griseus* IFO13350. *Biosci Biotechnol Biochem* **72**: 2973–2980.
- Sylvia, D.M. (2005) *Principles and applications of soil microbiology*. Pearson Prentice Hall, Upper Saddle River, N.J.
- Syvanen, M. (2012) Evolutionary implications of horizontal gene transfer. *Annu Rev Genet* **46**: 341–358.
- Szczepanowski, R., Braun, S., Riedel, V., Schneiker, S., Krahn, I., Pühler, A., and Schlüter, A. (2005) The 120 592 bp IncF plasmid pRSB107 isolated from a sewage-treatment plant encodes nine different antibiotic-resistance determinants, two iron-acquisition systems and other putative virulence-associated functions. *Microbiology (Reading, Engl)* **151**: 1095–1111.
- Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P.H., and Godelle, B. (1997) Role of mutator alleles in adaptive evolution. *Nature* **387**: 700–702.
- Tamas, I., Klasson, L., Canbäck, B., Näslund, A.K., Eriksson, A.-S., Wernegreen, J.J., *et al.* (2002) 50 million years of genomic stasis in endosymbiotic bacteria. *Science* **296**: 2376–2379.
- Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C., and Julien, B. (2000) Cloning and Heterologous Expression of the Epothilone Gene Cluster. *Science* **287**: 640–642.
- Thibessard, A., and Leblond, P. (2016) Complete Genome Sequence of *Streptomyces ambofaciens* DSM 40697, a Paradigm for Genome Plasticity Studies. *Genome Announc* **4**.
- Thoma, L., and Muth, G. (2012) Conjugative DNA transfer in *Streptomyces* by TraB: is one protein enough? *FEMS Microbiol Lett* **337**: 81–88.
- Thoma, L., and Muth, G. (2015) The conjugative DNA-transfer apparatus of *Streptomyces*. *International Journal of Medical Microbiology* **305**: 224–229.

- Thomas, C.M., and Nielsen, K.M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* **3**: 711–721.
- Thomas, F., Lefèvre, T., and Raymond, M. (eds) (2016) *Biologie évolutive*. De Boeck Supérieur, Louvain-la-Neuve. <http://www.documentation.ird.fr/hor/fdi:010068078>. Accessed October 5, 2019.
- Tidjani, A.-R., Lorenzi, J.-N., Toussaint, M., Dijk, E. van, Naquin, D., Lespinet, O., *et al.* (2019a) Massive Gene Flux Drives Genome Diversity between Sympatric *Streptomyces* Conspecifics. *MBio* **10**.
- Tidjani, A.-R., Lorenzi, J.-N., Toussaint, M., Dijk, E. van, Naquin, D., Lespinet, O., *et al.* (2019b) Genome Sequences of 11 Conspecific *Streptomyces* sp. Strains. *Microbiol Resour Announc* **8**.
- Uchida, T., Miyawaki, M., and Kinashi, H. (2003) Chromosomal Arm Replacement in *Streptomyces griseus*. *J Bacteriol* **185**: 1120–1124.
- Udwary, D.W., Zeigler, L., Asolkar, R.N., Singan, V., Lapidus, A., Fenical, W., *et al.* (2007) Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc Natl Acad Sci USA* **104**: 10376–10381.
- Uroz, S., Oger, P., Morin, E., and Frey-Klett, P. (2012) Distinct Ectomycorrhizospheres Share Similar Bacterial Communities as Revealed by Pyrosequencing-Based Analysis of 16S rRNA Genes. *Appl Environ Microbiol* **78**: 3020–3024.
- Vandamme, P., Pot, B., Gillis, M., Vos, P. de, Kersters, K., and Swings, J. (1996) Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev* **60**: 407–438.
- Vernikos, G., Medini, D., Riley, D.R., and Tettelin, H. (2015) Ten years of pan-genome analyses. *Curr Opin Microbiol* **23**: 148–154.
- Vetsigian, K., Jajoo, R., and Kishony, R. (2011) Structure and Evolution of *Streptomyces* Interaction Networks in Soil and In Silico. *PLOS Biology* **9**: e1001184.
- Vicente, C.M., Thibessard, A., Lorenzi, J.-N., Benhadj, M., Hôtel, L., Gacemi-Kirane, D., *et al.* (2018) Comparative Genomics among Closely Related *Streptomyces* Strains Revealed Specialized Metabolite Biosynthetic Gene Cluster Diversity. *Antibiotics (Basel)* **7** <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6315706/>. Accessed October 6, 2019.
- Vogel, J., Normand, P., Thioulouse, J., Nesme, X., and Grundmann, G.L. (2003) Relationship between spatial and genetic distance in *Agrobacterium* spp. in 1 cubic centimeter of soil. *Appl Environ Microbiol* **69**: 1482–1487.
- Vogelmann, J., Ammelburg, M., Finger, C., Guezguez, J., Linke, D., Flötenmeyer, M., *et al.* (2011) Conjugal plasmid transfer in *Streptomyces* resembles bacterial chromosome segregation by FtsK/SpoIIIE. *EMBO J* **30**: 2246–2254.
- Volff, J.N., and Altenbuchner, J. (2000) A new beginning with new ends: linearisation of circular chromosomes during bacterial evolution. *FEMS Microbiol Lett* **186**: 143–150.
- Vos, M., Quince, C., Pijl, A.S., Hollander, M. de, and Kowalchuk, G.A. (2012) A Comparison of rpoB and 16S rRNA as Markers in Pyrosequencing Studies of Bacterial Diversity. *PLOS ONE* **7**: e30600.
- Vos, M., and Velicer, G.J. (2008) Isolation by Distance in the Spore-Forming Soil Bacterium *Myxococcus xanthus*. *Current Biology* **18**: 386–391.

- Vos, M., and Velicer, G.J. (2009) Social Conflict in Centimeter- and Global-Scale Populations of the Bacterium *Myxococcus xanthus*. *Current Biology* **19**: 1763–1767.
- Vos, M., Wolf, A.B., Jennings, S.J., and Kowalchuk, G.A. (2013) Micro-scale determinants of bacterial diversity in soil. *FEMS Microbiology Reviews* **37**: 936–954.
- Vries, J. de, and Wackernagel, W. (2002) Integration of foreign DNA during natural transformation of *Acinetobacter* sp. by homology-facilitated illegitimate recombination. *Proc Natl Acad Sci USA* **99**: 2094–2099.
- Waksman, S.A., Reilly, H.C., and Johnstone, D.B. (1946) Isolation of Streptomycin-producing Strains of *Streptomyces griseus*. *J Bacteriol* **52**: 393–397.
- Wallace, A.R. (1870) *Contributions to the Theory of Natural Selection: A Series of Essays*. Macmillan and Company, .
- Walsh, C.T. (2002) Combinatorial biosynthesis of antibiotics: challenges and opportunities. *Chembiochem* **3**: 125–134.
- Wang, J.J., and Bishop, P. I. (2005) Development of a Phosphate Ion-selective Microelectrode and Its Use in Studies of the Enhanced Biological Phosphorus Removal (ebpr) Process. *Environmental Technology* **26**: 381–388.
- Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, M.I., *et al.* (1987) Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *International Journal of Systematic and Evolutionary Microbiology*, **37**: 463–464.
- Weidlich, G., Wirth, R., and Galli, D. (1992) Sex pheromone plasmid pAD1-encoded surface exclusion protein of *Enterococcus faecalis*. *Mol Gen Genet* **233**: 161–168.
- Wenner, T., Roth, V., Fischer, G., Fourrier, C., Aigle, B., Decaris, B., and Leblond, P. (2003) End-to-end fusion of linear deleted chromosomes initiates a cycle of genome instability in *Streptomyces ambifaciens*. *Molecular Microbiology* **50**: 411–425.
- Wetterlind, J., Stenberg, B., and Söderström, M. (2010) Increased sample point density in farm soil mapping by local calibration of visible and near infrared prediction models. *Geoderma* **156**: 152–160.
- Whitaker, R.J., Grogan, D.W., and Taylor, J.W. (2003) Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301**: 976–978.
- Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998) Prokaryotes: The unseen majority. *PNAS* **95**: 6578–6583.
- Widenbrant, E.M., Tsai, H.-H., Chen, C.W., and Kao, C.M. (2007) *Streptomyces coelicolor* Undergoes Spontaneous Chromosomal End Replacement. *J Bacteriol* **189**: 9117–9121.
- Widenbrant, E.M., Tsai, H.-H., Chen, C.W., and Kao, C.M. (2008) Spontaneous Amplification of the Actinorhodin Gene Cluster in *Streptomyces coelicolor* Involving Native Insertion Sequence IS466. *J Bacteriol* **190**: 4754–4758.
- Wiedenbeck, J., and Cohan, F.M. (2011) Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiology Reviews* **35**: 957–976.

- Wielgoss, S., Barrick, J.E., Tenaillon, O., Cruveiller, S., Chane-Woon-Ming, B., Médigue, C., *et al.* (2011) Mutation Rate Inferred From Synonymous Substitutions in a Long-Term Evolution Experiment With *Escherichia coli*. *G3 (Bethesda)* **1**: 183–186.
- Wielgoss, S., Barrick, J.E., Tenaillon, O., Wisner, M.J., Dittmar, W.J., Cruveiller, S., *et al.* (2013) Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *PNAS* **110**: 222–227.
- Wielgoss, S., Fiegna, F., Rendueles, O., Yu, Y.-T.N., and Velicer, G.J. (2018) Kin discrimination and outer membrane exchange in *Myxococcus xanthus*: A comparative analysis among natural isolates. *Molecular Ecology* **27**: 3146–3158.
- Wietz, M., Duncan, K., Patin, N.V., and Jensen, P.R. (2013) Antagonistic interactions mediated by marine bacteria: the role of small molecules. *J Chem Ecol* **39**: 879–891.
- Woese, C.R. (1987) Bacterial evolution. *Microbiol Rev* **51**: 221–271.
- Wozniak, R.A.F., and Waldor, M.K. (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol* **8**: 552–563.
- Wright, D.A., Killham, K., Glover, L.A., and Prosser, J.I. (1995) Role of Pore Size Location in Determining Bacterial Activity during Predation by Protozoa in Soil. *Appl Environ Microbiol* **61**: 3537–3543.
- WRIGHT, S. (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the sixth international congress of Genetics* **1**: 356–366.
- Wu, K., Chung, L., Revill, W.P., Katz, L., and Reeves, C.D. (2000) The FK520 gene cluster of *Streptomyces hygroscopicus* var. *ascomyceticus* (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. *Gene* **251**: 81–90.
- Yamasaki, M., and Kinashi, H. (2004) Two Chimeric Chromosomes of *Streptomyces coelicolor* A3(2) Generated by Single Crossover of the Wild-Type Chromosome and Linear Plasmid SCP1. *J Bacteriol* **186**: 6553–6559.
- Yanai, K., Murakami, T., and Bibb, M. (2006) Amplification of the entire kanamycin biosynthetic gene cluster during empirical strain improvement of *Streptomyces kanamyceticus*. *Proc Natl Acad Sci USA* **103**: 9661–9666.
- Yang, C.-C., Huang, C.-H., Li, C.-Y., Tsay, Y.-G., Lee, S.-C., and Chen, C.W. (2002) The terminal proteins of linear *Streptomyces* chromosomes and plasmids: a novel class of replication priming proteins: Terminal proteins of *Streptomyces*. *Molecular Microbiology* **43**: 297–305.
- Young, M., and Cullum, J. (1987) A plausible mechanism for large-scale chromosomal DNA amplification in streptomycetes. *FEBS Lett* **212**: 10–14.
- Zhao, L., Zhong, L., and Qin, Z. (2014) Two distinct conjugal transfer systems on *Streptomyces* plasmid pZL1. *Acta Biochim Biophys Sin (Shanghai)* **46**: 1084–1086.
- Zhaxybayeva, O., Swithers, K.S., Lapierre, P., Fournier, G.P., Bickhart, D.M., DeBoy, R.T., *et al.* (2009) On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. *Proc Natl Acad Sci USA* **106**: 5865–5870.
- Zhou, H.-Q., Ning, L.-W., Zhang, H.-X., and Guo, F.-B. (2014) Analysis of the Relationship between Genomic GC Content and Patterns of Base Usage, Codon Usage and Amino Acid Usage in Prokaryotes:

Similar GC Content Adopts Similar Compositional Frequencies Regardless of the Phylogenetic Lineages. *PLoS One* **9** <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4177787/>. Accessed May 14, 2019.

Ziemert, N., Lechner, A., Wietz, M., Millán-Aguiñaga, N., Chavarria, K.L., and Jensen, P.R. (2014) Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *PNAS* **111**: E1130–E1139.

Zotchev, S.B. (2014) Genomics-Based Insights into the Evolution of Secondary Metabolite Biosynthesis in Actinomycete Bacteria.

Annexes : Figures supplémentaires de l'article
du chapitre II

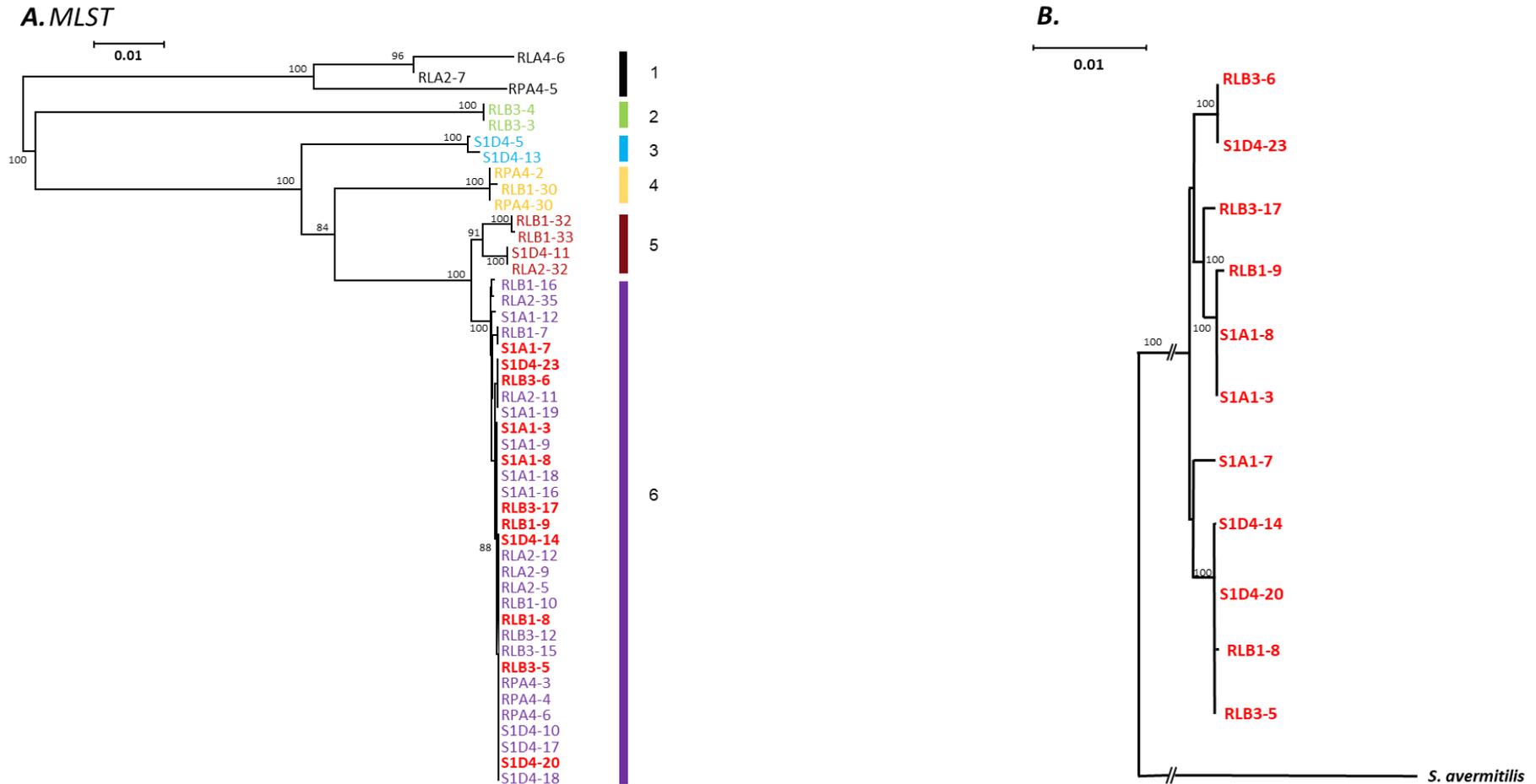


Fig. S1. Phylogenetic trees. A. Multi Locus sequence typing (MLST) tree of isolated strains. The tree is constructed by neighbor-joining with a K2P distance correction and 100 bootstrap replicates. It is based on the concatenated sequence alignment of five loci (*atpD*, *gyrB*, *recA*, *rpoB*, *trpB*). The final dataset has 3,178 positions in total. The scale bar indicates 1% estimated sequence divergence. Clades are represented by coloured boxes. Strains whose genomes were sequenced are represented in red inside clade 6. **B.** Phylogenomic analysis was performed with 5,213 genes (5,149,602 nucleotide positions) of the core genome of the population shared with *Streptomyces avermitilis*, which was used as a root. The tree was built in Maximum Likelihood with a GTR model and 100 bootstrap replicates.

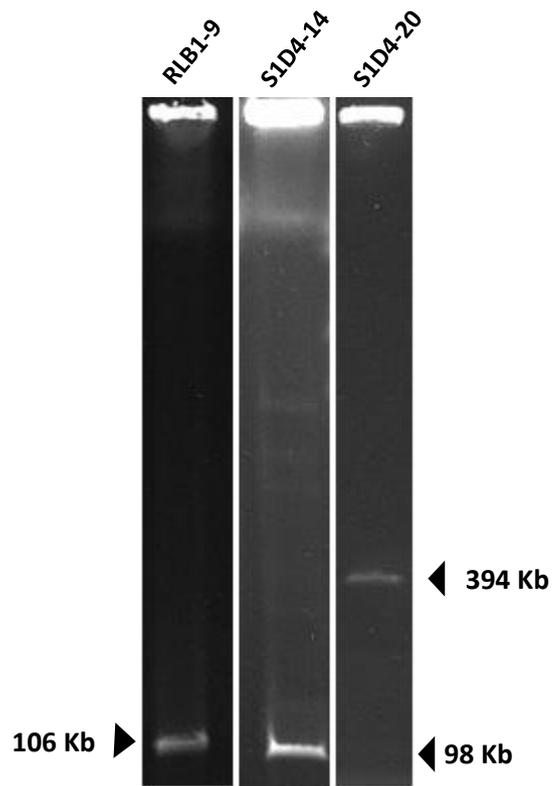


Fig. S2. Extrachromosomal DNA content. PFGE analysis of undigested DNA of strains RLB1-9, S1D4-14 and S1D4-20 after growth at 30° for 36 h. Running conditions were 6 V.cm⁻¹ for 20 h with ramped pulse times from 40 s to 160 s.

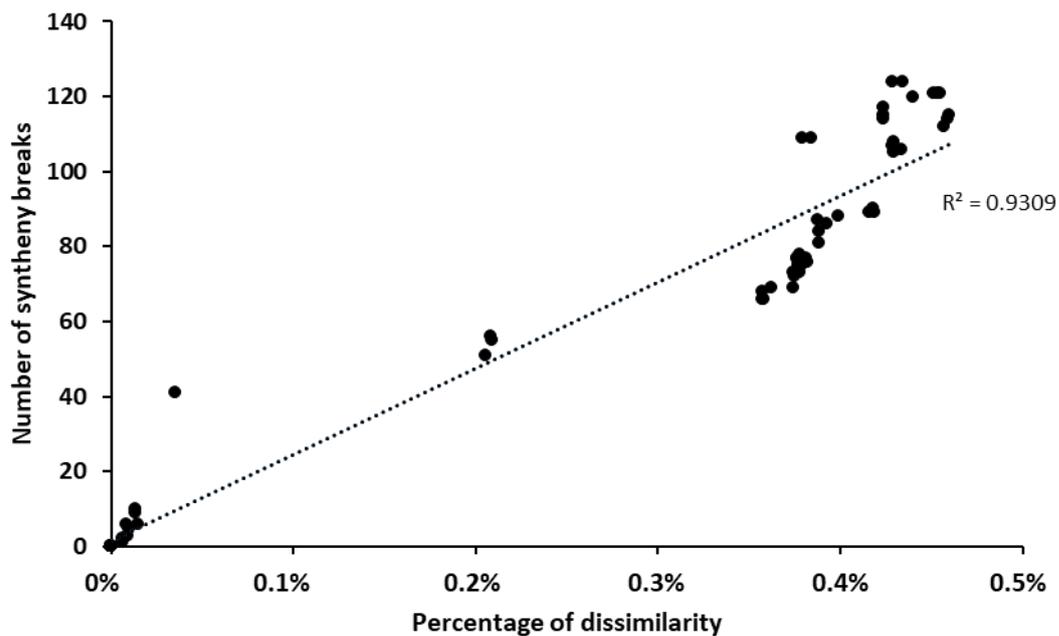


Fig. S3. Correlation between phylogenetic distance and synteny breaks. The phylogenetic distance was determined on the core genome and is expressed as a percentage of dissimilarity. Each dot represents the number of synteny breaks encountered in a pairwise comparison. The dotted line represents the correlation curve ($R^2=93.09$).

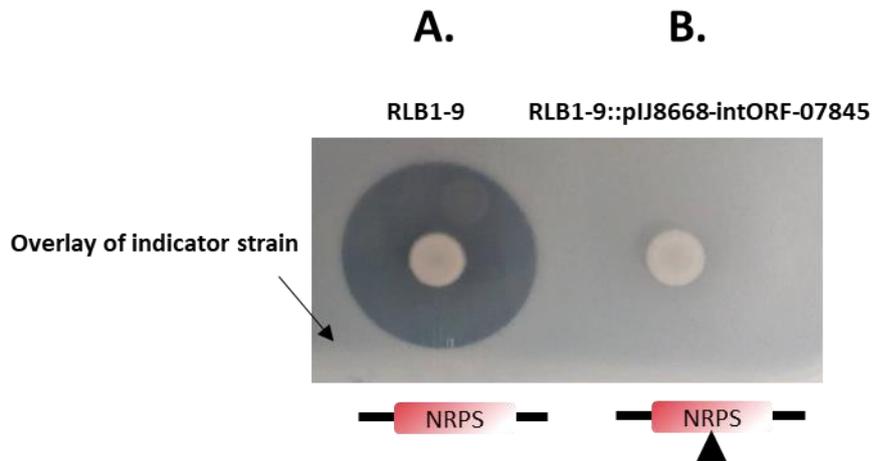


Fig. S4. Abolition of inhibitory activity of the strain RLB1-9 by NRPS disruption. Spore suspensions of RLB1-9 and of the RLB1-9 NRPS mutant were incubated 5 days in **A** and **B** respectively to enable the diffusion of specialized metabolites. They were then overlaid with a *Bacillus* indicator strain. In **A** the growth of *Bacillus* is inhibited by RLB1-9 as scored by the clearance zone when the NRPS mutant in **B** showed no inhibitory activity.

Table S1 Summary table of genome sequencing data. For Illumina sequencing, all strains were sequenced using 250 nt paired-ends except for RLB1-9 (*) and RLB1-8 (**) which were sequenced using 75 nt and 300 nt paired ends respectively.

Strain	Illumina sequencing		Nanopore sequencing				
	Number of reads	Coverage	Number of reads	Number of bases	Coverage	Number of contigs	Estimated genome size
RLB1-9	9,166,845*	107	88,694	488,157,099	41	3	12,146,841
RLB1-8	4,492,734**	150	655,492	1,794,236,654	150	2	11,652,134
RLB3-5	1,598,054	53	144,521	673,009,805	56	10	11,697,055
RLB3-6	1,637,136	55	299,155	618,705,318	52	5	12,454,816
RLB3-17	1,988,311	66	202,455	603,258,039	50	1	12,023,175
S1A1-3	1,621,592	54	197,777	612,096,268	51	7	11,988,482
S1A1-7	1,752,105	58	533,299	876,735,911	73	18	11,820,843
S1A1-8	1,595,659	53	780,962	632,637,759	53	16	12,040,123
S1D4-14	1,397,227	47	2,065,804	4,195,872,938	350	5	11,934,153
S1D4-20	1,663,586	55	1,425,070	3,098,746,284	258	5	12,255,226
S1D4-23	1,771,880	59	400,348	735,306,850	61	8	12,060,819

Table S2 Table of genomic data. L indicates an extrachromosomal replicon detected by PFGE. C indicates a circular extrachromosomal replicon identified in silico and consistently not visible in PFGE.

Strain	Replicon size	Genome size	Number of Coding Sequence	G+C %	Terminal Inverted Repeat size (kb)
RLB1-9	11,940,408	12,200,709	10,729	70.2	311
pRLB1-9.1	154,158 ^C		181	69.0	
pRLB1-9.2	106,143 ^L		120	68.7	24
RLB1-8	11,765,340	11,765,340	10,538	70.2	357
RLB3-5	11,898,970	11,898,970	10,632	70.2	365
RLB3-6	12,338,263	12,448,281	11,088	70.1	587
pRLB3-6.1	110,314 ^C		100	70.6	
RLB3-17	12,023,175	12,023,175	10,809	70.2	451
S1A1-3	12,042,091	12,042,091	10,801	70.2	393
S1A1-7	11,713,151	12,005,504	10,44	70.3	513
pS1A1-7.1	292,353 ^C		251	69.7	
S1A1-8	12,036,971	12,036,971	10,806	70.2	394
S1D4-14	11,723,487	11,934,498	10,472	70.2	369
pS1D4-14.1	112,196 ^L		117	68.7	0
pS1D4-14.2	98,815 ^C		150	69.1	
S1D4-20	11,851,257	12,245,276	10,642	70.2	373
pS1D4-20.1	394,019 ^L		372	69.1	68
S1D4-23	12,057,712	12,057,712	11	70.2	421

Table S3 Distributions of indels events along the chromosome. The chromosome of strain RLB1-8 was chosen as a reference for pairwise comparisons. The position of an indel is indicated according to the position of its first CDS in RLB1-8. This position corresponds to the CDS number in RLB1-8. Cells with identical colours indicate that the strains share the same indel at this position. The numbers in the cells indicate the number of CDS in the indel. * represents the presence of an AICE and # the presence of an ICE in the indel.

CDS position in RLB1-8 (Reference)	Strains										
	RLB1-8	S1D4-23	RLB1-9	S1A1-3	S1A1-8	S1A1-7	RLB3-6	RLB3-17	RLB3-5	S1D4-14	S1D4-20
61	3	3	3	3	3	3	49	3	3	3	3
82	0	0	3	3	3	0	0	0	0	0	0
97	4	4	4	4	4	4	2	4	4	4	4
110	3	3	3	3	3	3	3	3	3	3	3
135	64	64	64	64	64	64	86	64	64	64	64
136	4	4	4	2	2	4	4	2	4	4	4
147	52	52	52	52	52	12	52	52	52	52	52
183	32	19	32	32	32	32	32	32	32	32	32
190	3	0	0	0	0	3	3	0	3	3	3
196	3	3	3	8	3	3	3	3	3	3	3
199	0	0	6	0	6	0	0	0	0	0	0
206	9	9	9	9	9	41	41	9	9	9	9
220	2	2	2	2	2		5	2	2	2	2
231	8	8	8	8	8	4	8	8	8	8	8
235	9	9	12	12	12	9	9	12	9	9	9
247	0	0	0	0	0	9	0	0	0	0	0
249	3	3	0	0	0	3	3	3	3	3	3
259	3	5	3	3	3	5	5	3	3	3	3
265	11	1	11	11	11	4	4	11	11	11	11
282	11	4	5	5	5	4	4	11	11	11	11
321	1	7	1	1	1	1	7	1	1	1	1
339	0	0	0	0	0	0	0	0	0	6	0
341	35	64	35	35	35	35	35	5	35	35	35
343	4	4	4	4	4	2	4	4	4	4	4
355	13	13	1	1	1	13	13	13	13	13	13
365	5	5	5	5	5	5	5	5	5	5	5
376	4	4	4	4	4	4	4	5	4	4	4
378	2	2	3	3	3	2	2	2	2	2	2
384	1	1	3	3	3	1	1	1	1	1	1
422	1	57	1	1	1	1	57	1	1	1	1
448	0	3	0	0	0	0	0	0	0	0	0

455	5	5	5	5	5	5	5	6	5	5	5
465	0	0	4	4	4	0	0	0	0	0	0
467	3	3	6	6	6	3	3	1	3	3	3
469	2	2	2	2	2	3	2	2	2	2	2
474	5	1	5	5	5	5	1	5	5	5	5
476	0	0	2	2	2	3/0	0	3/0	0	0	0
510	6	6	2	2	2	6	6	2	6	6	6
513	3	0	3	3	3	3	0	3	3	3	3
522	5	2	5	5	5	5	2	5	5	5	5
525	2	2	2	2	2	4	2	2	2	2	2
530	0	0	5	5	5	0	0	0	0	0	0
531	3	4	3	3	3	3	4	3	3	3	3
537	20	10	20	20	20	20	10	20	20	20	20
543	14	14	12	12	12	5	14	12	14	14	14
565	0	0	0	0	0	16	0	4	0	0	0
578	2	2	2	2	2	2	5	2	2	2	2
582	0	8	0	0	0	0	4	0	0	0	0
584	3	3	3	3	3	3	4	3	3	3	3
600	2	2	3	3	3	2	2	3	2	2	2
610	14	14	37	37	37	33	14	33	14	14	14
616	1	1	1	1	1	1	3	1	1	1	1
626	3	3	2	2	2	2	3	3	3	3	3
630	3	3	3	3	3	3	3	3	3	3	3
639	0	0	3	3	3	0	0	0	0	0	0
643	35	2	35	35	35	2	2	35	35	35	35
674	5	5	5	5	5	5	5	3	5	5	5
693	3	3	3	3	3	3	4	3	3	3	3
701	1	1	1	4	4	1	1	1	1	1	1
705	4	4	4	2	2	4	4	2	4	4	4
712	2	4	2	2	2	2	4	2	2	2	2
731	1	17	1	1	1	1	17	1	1	1	1
755	2	7	3	3	3	2	3	2	2	2	2
760	3	1	1	1	1	3	1	3	3	3	3
766	7	7	4	4	4	4	7	4	7	7	7
769	4	7	4	4	4	4	7	4	4	4	4
779	0	3	0	0	0	0	3	0	0	0	0
810	6	1	6	6	6	6	1	6	6	6	6
852	1	6	1	1	1	1	6	1	1	1	1
855	4	5	4	4	4	4	5	4	4	4	4
930	6	4	4	4	4	6	4	6	6	6	6
938	12	0	0	0	0	12	0	12	12	12	12
1,011	0	59	59	59	59	59	59	0	0	0	0
1,013	3	5	3	3	3	5	3	3	3	3	3
1,037	1	7	1	1	1	1	7	1	1	1	1
1,054	1	8	1	1	1	1	8	1	1	1	1

1,077	0	7	0	0	0	0	6	0	0	0	0
1,104	1	1	1	4	4	1	1	1	1	1	1
1,216	3	0	0	0	0	0	0	0	3	3	3
1,252	5	5	1	1	1	1	5	5	5	5	5
1,411	3	7	12	13	13	3	3	3	3	3	3
1,425	103	95	172	175	175	95	95	18	103	103	103
1,433	58	58	58	58	58	20	20	58	58	58	58
1,523	20	20	20	20	20	20	20	8	20	20	20
1,530	10	2	10	10	10	3	1	10	10	10	10
1,532	3	3	0	0	0	3	3	3	3	3	3
1,562	2	2	2	2	2	2	5	2	2	2	2
1,597	7	7	7	7	7	7	7	0	7	7	7
1,619	1	3	1	1	1	3	3	5	1	1	1
1,630	4	6	4	4	4	4	4	4	4	4	4
1,638	28	6	28	28	28	2	3	87	28	28	28
1,642	2	2	5	5	5	2	2	2	2	2	2
1,658	3	3	4	4	3	3	3	3	3	3	3
1,672	0	0	6	5	5	0	0	0	0	0	0
1,677	17	4	17	17	17	5	17	17	17	17	17
1,687	1	1	4	4	4	1	1	1	1	1	1
1,700	4	20	4	4	4	4	4	4	4	4	4
1,713	4	4	4	4	4	11	12	11	4	4	4
1,727	105	0	105	105	105	0	0	105	105	105	105
1,737	12	12	12	12	12	12	12	12	12	12	12
1,741	5	5	4	5	5	5	5	5	5	5	5
1,745	5	1	5	5	5	1	2	5	5	5	5
1,758	5	0	5	5	5	1	0	5	5	5	5
1,829	2	2	2	2	2	2	2	11	2	2	2
1,845	4	0	4	4	4	0	4	4	4	4	4
1,876	3	3	2	2	2	27	38	5	3	3	3
1,878	13	10	13	13	13	13	13	13	13	13	13
1,903	0	4	0	0	0	0	4	0	0	0	0
2,017	2	12	2	2	2	2	37	2	2	2	2
2,067	4	3	11	11	11	11	7	4	4	1	1
2,080	3	8	3	3	3	3	8	3	3	3	3
2,153	12	14	12	12	12	12	14	12	12	12	12
2,296	0	7	0	0	0	0	7	0	0	0	0
2,359	2	2	2	2	2	4	2	2	2	2	2
2,509	2	2	2	2	2	2	2	3	2	2	2
2,614	0	3	0	0	0	0	3	0	0	0	0
2,940	0	0	0	0	0	0	0	3	0	0	0
2,977	41*	0	0	0	0	0	0	0	41*	41*	41*
3,222	0	4	0	0	0	0	0	0	0	0	0
3,485	0	0	20	20	20	20	0	0	0	0	0
3,778	1	6	1	1	1	1	6	1	1	1	1

3,785	0	4	0	0	0	0	4	0	0	0	0
3,881	0	0	0	0	0	0	24*	0	0	0	0
3,954	2	15*	2	2	2	17*	2	2	2	2	2
3,956	0	0	0	0	0	0	0	0	0	0	84
4,202	0	0	7	10	10	0	0	0	0	0	0
4,252	3	3	3	3	3	5	3	3	3	3	3
4,254	1	21*	1	1	1	1	21*	1	1	1	1
4,351	0	7	0	0	0	0	7	0	0	0	0
4,356	2	18*	2	2	2	2	4	2	2	2	2
4,371	4	2	4	4	4	4	2	4	4	4	4
4,380	8	6	8	8	8	8	6	8	8	8	8
4,382	6	6	1	1	1	1	6	6	6	6	6
4,384	4	4	4	4	4	4	4	0	4	4	4
4,566	19	6	4	4	4	13	4	19	19	19	19
4,916	3	22*	3	3	3	3	3	3	3	3	3
4,927	0	4	0	0	0	0	4	20*	0	0	0
4,971	1	6	1	1	1	1	6	1	1	1	1
4,985	23	6	6	6	6	5	6	4	23	23	23
4,992	11							2	11	11	11
5,074	3	0	0	0	0	0	0	0	3	3	3
5,250	0	0	0	0	0	0	18*	0	0	0	0
5,270	0	11	0	0	0	11	0	0	0	0	0
5,439	0	0	0	0	0	0	0	15*	0	0	0
5,464	0	0	0	0	0	0	13*	0	0	0	31*
5,476	3	16*	3	14*	16*	3	14*	22*	3	3	3
5,797	5	4	5	5	5	5	4	5	5	5	5
5,809	1	4	1	1	1	1	1	1	1	1	1
5,946	0	0	0	0	0	27*	0	0	0	0	0
5,991	85#	0	16*	16*	16*	0	0	0	85#	0	0
6,080	7	0	0	2	2	0	0	0	7	7	7
6,331	0	0	4	4	4	4	0	4	0	0	0
6,375	0	0	18*	18*	18*	0	3	18*	0	18*	18*
6,414	3	7	7	7	7	10*	7	7	3	3	3
6,745	0	0	0	0	0	6	0	0	0	0	0
6,807	4	2	2	2	2	4	2	2	4	4	4
7,111	8	13*	8	8	8	14	13*	46*	8	8	8
7,117	18*	18*	26*	29*	29*	18*	18*	18*	18*	17*	17*
7,127	8	8	8	8	8	9	8	9	8	8	8
7,135	0	0	0	0	0	0	0	0	0	0	27
7,232	3	3	3	3	3	3	3	3	3	3	5
7,408	0	0	0	0	0	3	0	0	0	0	0

7,441	1	3	1	1	1	1	3	1	1	1	1
7,473	0	0	0	0	0	0	0	36*	0	0	0
7,483	0	0	19*	19*	19*	0	15*	0	0	0	0
7,663	3	1	1	1	1	1	1	1	3	3	3
7,811	0	0	0	0	0	3	0	0	0	0	0
7,945	1	1	1	1	1	1	4	1	1	1	1
7,949	2	4	2	2	2	2	4	2	2	2	2
7,977	24	11	2	2	2	12	11	5	24	24	24
8,028	4	0	0	0	0	0	0	0	4	4	4
8,059	0	0	0	0	0	17	0	0	0	0	0
8,063	1	1	1	1	1	4	1	1	1	1	1
8,114	1	1	1	1	1	5	1	1	1	1	1
8,117	0	95	0	0	0	0	95	0	0	0	0
8,183	0	241	139	138	138	53	241	144	0	0	0
8,302	0	0	0	0	0	17	0	17	0	0	0
8,390	0	0	4	4	4	0	0	4	0	0	0
8,445	0	0	3	3	3	0	0	3	0	0	0
8,452	3	3	5	5	5	3	3	5	3	3	3
8,595	52	0	0	0	0	0	0	52	52	52	52
8,687	0	0	27	27	27	0	0	31	0	0	0
8,704	0	0	0	0	0	25*	0		0	0	0
9,051	1	1	3	1	1	1	1	1	1	1	1
9,063	1	3	1	1	1	1	3	1	1	1	1
9,082	0	0	0	0	0	3	0	3	0	0	0
9,139	4	4	4	4	4	4	4	5	4	4	4
9,163	4	4	4	4	4	4	5	4	4	4	4
9,362	2	6	2	2	2		6	2	2	2	2
9,374	7	8	7	7	7	8	12	7	7	7	7
9,400	6	0	6	6	6	0	0	6	6	6	6
9,403	3	3	0	0	0	3	3	0	3	3	3
9,409	1	1	1	1	1	1	3	1	1	1	1
9,428	4	4	3	3	3	3	4	4	4	4	4
9,435	7	7	7	7	7	7	10	7	7	7	7
9,444	0	3	0	0	0	0	0	0	0	0	0
9,610	4	1	4	4	4	4	1	4	4	4	4
9,735	1		1	1	1	1	1	1	87	1	1
9,831	1	34	1	1	1	1	32	1		1	1
9,851	0	0	3	3	3	0	0	0	0	0	0
9,962	2	5	7	7	7	5	5	7	2	2	2
9,986	9	6	9	9	9	9	6	9	9	9	9
10,169	2	4	2	2	2	7	4	2	2	2	2
10,189	4	8	4	4	4	4	4	4	4	4	4
10,192	1	1	1	1	1	1	6	1	1	1	1
10,197	4	4	5	5	5	4	4	5	4	4	4

10,211	3	0	0	0	0	0	0	0	3	3	3
10,239	1	1	4	4	4	1	1	4	1	1	1
10,254	10	10	2	2	2	10	10	2	10	10	10
10,261	3	3	3	3	3	4	3	3	3	3	3
10,283	14	2	14	14	14	2	2	14	14	14	14
10,299	8	1	8	8	8	1	1	8	8	8	8
10,325	1	1	1	1	1	1	1	5	1	1	1
10,328	2	9	2	2	2	9	9	2	2	2	2
10,337	5	5	5	5	5	10		5	5	5	5
10,347	0	0	0	0	0	0	0	3	0	0	0
10,351	3	3	3	3	3	12	3	3	3	3	3
10,376	7	22	7	7	7	7	22	7	7	7	7
10,400	2	2	2	2	2	5	2	2	2	2	2
10,430	5	5	5	5	5	5	10	5	5	5	5
10,441	1	1	1	1	1	1	7	1	1	1	1
10,448	29	7	29	29	29	29	11	29	29	29	29
10,470	2	2	2	4	2	2	2	2	2	2	2
10,472	0	0	0	0	0	0	0	3	0	0	0
10,487	4	4	4	4	4	2	7	4	4	4	4
10,497	4	4	2	5	5	4	4	2	4	4	4

Evolution génomique au sein d'une population naturelle de *Streptomyces*

Résumé :

Les *Streptomyces* sont des bactéries de la rhizosphère qui contribuent à la fertilité des sols (recyclage de la matière organique), et à la croissance et la santé des plantes. Elles possèdent parmi les plus grands génomes bactériens (12 Mb) et présentent une variabilité génétique importante. Cette variabilité connue au niveau interspécifique n'a jamais été abordée à l'échelle de la population, c'est-à-dire entre individus sympatriques appartenant à la même espèce (souches sœurs) au sein de la même niche écologique. L'objectif de ce travail est de rechercher cette diversité dans les populations de l'écosystème sol forestier, d'approcher sa dynamique et son rôle fonctionnel.

Après séquençage et comparaison des génomes complets, nous avons observé une grande diversité génomique en termes de taille, de présence/absence d'éléments extrachromosomiques, mais également en terme de présence/absence de gènes le long du chromosome. Un grand nombre d'événements d'insertions et délétions (indels) comprenant de 1 à 241 gènes différencient les individus de la population. Au vu des liens phylogénétiques étroits entre les individus, l'ancêtre commun de la population est récent, aussi la diversité génomique résulterait d'un flux massif et rapide de gènes. La forte prévalence d'éléments conjugatifs intégrés dans la population suggère que la conjugaison est le moteur prépondérant de cette diversité génomique.

La production différentielle de métabolites spécialisés (antibiotiques) a également été utilisée pour estimer l'impact de la diversité génétique sur le fonctionnement de la population. Nous avons pu montrer que cette production était liée à des gènes spécifiques de souches et qu'elle pouvait constituer un bien commun pour la population. Nous proposons que l'évolution rapide du génome participe au maintien des mécanismes de cohésion sociale chez ces bactéries du sol.

Mots clés : Genomique bactérienne, Evolution, Population, Transfert horizontal, Ecologie du sol forestier, *Streptomyces*.

Genomic evolution within a natural population of *Streptomyces*

Abstract:

Streptomyces are rhizospheric bacteria that contribute to soil fertility (recycling of organic matter), plant growth and health. They have among the largest bacterial genomes (12 Mb) with a high genetic variability. The genome variability, observed at the interspecific level has never been addressed within a population, i.e. between sympatric individuals belonging to the same species (Conspecific strains) within the same ecological niche. The objective of this work was to investigate this diversity in the forest soil ecosystem, to estimate its dynamics and its potential functional roles.

After sequencing and comparison of the complete genomes, we observed a wide genomic diversity in terms of size, presence/absence of extrachromosomal elements, but also in terms of presence/absence of genes along the chromosome. A large number of insertion and deletion events (indels) from 1 to 241 genes differentiate individuals in the population. Given the close phylogenetic relationship of these strains, the common ancestor of the population is recent, hence the genomic diversity would result from a massive and rapid gene flux. The high prevalence of integrative and conjugative elements in the population suggests that conjugation could act as a driving force of this diversity.

Differential production of specialized metabolites (antibiotics) was also used to estimate the impact of genetic diversity on population's ecology. We were able to show that this production was linked to strain specific genes and that it may constitute a « public good » for the population. We propose that the rapid evolution of the genome contributes to the maintenance of social cohesion mechanisms within these soil bacteria.

Keywords: Bacterial genomics, Evolution, Population, Horizontal gene transfer, Forest soil's ecology, *Streptomyces*.