Nancy-Université
Université Henri Poincaré

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.

Toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

➢ Contact SCD Nancy 1 : theses.sciences@scd.uhp-nancy.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

Nancy-Université
Université Henri Poincaré

FACULTE DES SCIENCES ET TECHNIQUES

UFR Sciences & Techniques Biologiques
Ecole Doctorale Biologie Santé Environnement

Thèse

Présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy-I

En Génétique Moléculaire

par

Alexandre GALLOIS

DYNAMIQUE DES EXTREMITES DU CHROMOSOME LINEAIRE DE *STREPTOMYCES AMBOFACIENS*

Thèse soutenue le 12 Novembre 2007

Membres du jury :

Rapporteurs :	M. Dominique SCHNEIDER	Professeur, Université de Grenoble
	M. Jean-Nicolas VOLFF	Professeur, ENS Lyon
Examinateurs :	Mme Annie DARY,	Professeur, U.H.P.-I.U.T, Nancy I
	M. Bertrand AIGLE,	Maître de Conférence, U.H.P., Nancy I
	M. Bernard DECARIS,	Professeur, U.H.P., Nancy I (Président de jury)
	M. Pierre LEBLOND,	Professeur, U.H.P., Nancy I (Directeur de thèse)

Laboratoire de Génétique et Microbiologie – UMR UHP/INRA 1128 – IFR 110
Faculté des Sciences et Techniques – 54506 Vandoeuvre-les-Nancy

Nancy-Université
Université Henri Poincaré

FACULTE DES SCIENCES ET TECHNIQUES

UFR Sciences & Techniques Biologiques
Ecole Doctorale Biologie Santé Environnement

Thèse

Présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy-I

En Génétique Moléculaire

par

Alexandre GALLOIS

DYNAMIQUE DES EXTREMITES DU CHROMOSOME LINEAIRE DE *STREPTOMYCES AMBOFACIENS*

Thèse soutenue le 12 Novembre 2007

Membres du jury :

Rapporteurs :	M. Dominique SCHNEIDER	Professeur, Université de Grenoble
	M. Jean-Nicolas VOLFF	Professeur, ENS Lyon
Examinateurs :	Mme Annie DARY,	Professeur, U.H.P.-I.U.T, Nancy I
	M. Bertrand AIGLE,	Maître de Conférence, U.H.P., Nancy I
	M. Bernard DECARIS,	Professeur, U.H.P., Nancy I (Président du jury)
	M. Pierre LEBLOND,	Professeur, U.H.P., Nancy I (Directeur de thèse)

Laboratoire de Génétique et Microbiologie – UMR UHP/INRA 1128 – IFR 110
Faculté des Sciences et Techniques – 54506 Vandoeuvre-les-Nancy

« As-tu déjà fait de ces rêves qui ont l'air plus vrai que la réalité.
Si tu étais dans l'incapacité de sortir d'un de ces rêves.
Comment ferais-tu la différence entre le monde du rêve
Et le monde réel ? »

REMERCIEMENTS

Je souhaite remercier le Pr. Bernard DECARIS d'avoir accepté que je réalise mes travaux au sein du laboratoire de Génétique et Microbiologie – UMR INRA/UHP 1128 de l'Université Henri Poincaré Nancy 1.

Un très grand merci à mon directeur de thèse, Pr. Pierre LEBLOND, pour m'avoir proposé de travailler au sein de son équipe sur l'étude de l'évolution et des mécanismes évolutifs du chromosome de *Streptomyces ambofaciens,* et de m'avoir fait confiance. Merci pour sa disponibilité au quotidien, ses conseils et ses remarques sur mon travail de recherche. Merci également pour nos grandes discussions du déjeuner portant sur des sujets sortant du cadre de nos activités de recherche.

Un merci à Bertrand AIGLE pour m'avoir mis le pied à l'étrier durant mon stage de DEA et mes premières années de thèse sur *Streptomyces*, et les nombreux conseils techniques.

Merci à Messieurs Dominique SCHNEIDER, Professeur, et Jean-Nicolas VOLFF, Professeur, d'avoir accepté d'être les rapporteurs de ma thèse.
Merci à Mme Annie DARY, Professeur, d'avoir accepté de faire partie de ce jury de thèse.

Ce travail a fait l'objet de nombreux contacts et collaboration entre autres avec le Laboratoire de Génétique et Microbiologie de Nancy, l'Institut de Génétique et Microbiologie d'Orsay et le Génoscope. Ainsi je remercie ces personnes pour leur implication dans ce projet :

- Merci à Jean-Luc PERNODET, Michel GUERINEAU, Claude GERBAUD et François-Xavier FRANCOU de l'Institut de Génétique et Microbiologie.
- Merci à Valérie BARBE, Sophie MANGENOT, et Chantal TRUONG du Génoscope.
- Un remerciement spécial à l'ensemble de l'équipe de Béatrice SEGURENS (Nathalie, Manu et tous les autres) pour leur accueil, leur sympathie et leur aide durant mes deux séjours au sein du Génoscope.
- Merci à Frédéric BORGES et Frédéric CHOULET qui complètent l'équipe du laboratoire ayant participé à ce travail.
- Merci à Chloé AMBROSET d'avoir initié le travail sur la partition chromosomique et de m'avoir supporté durant son stage de DEA.

Je tiens à remercier beaucoup de personnes qui durant toutes ces années ont su être présent même juste un instant. Bien sûr, citer tout le monde est très difficile (et pas uniquement faute de place) donc si votre nom n'apparaît pas, n'en soyez pas vexé, mais sachez que je pense à vous au moment d'écrire ces quelques lignes.

Je tiens à remercier l'ensemble des personnes du laboratoire que j'ai côtoyé durant mes années passées au sein du Laboratoire de Génétique et Microbiologie, pour les discussions que l'on a pu avoir, l'aide mutuelle que l'on a pu s'apporter.
Plus particulièrement, je tiens également à remercier Paule, Laurence, Alexandra, pour son amitié et sa présence lors de moments particuliers, Sylvain et Manon, ainsi que Brigitte et Christian pour nos repas et soirées couscous chez Fatie, à qui je fais un clin d'œil. Merci également à Katy pour sa complicité pour organiser certains pots au labo et pour nos discussions.
Je ne veux pas oublier Guillaume PAVLOVIC qui m'a initié au travail de recherche durant mes deux stages de maîtrise. Merci pour tous les moments passés avec toi et l'autre Guillaume à user les fauteuils des cinémas de Nancy ou de ton appart pour les matchs de foot.
Un très grand merci à Céline FOURRIER, pas uniquement pour l'aide technique qu'elle m'a apporté durant ma thèse mais également pour sa sympathie et son amitié.

Je remercie également mon groupe laboratoire auquel se rattache d'autres éléments. Merci à Sibel, Kevin, Dorian, Axelle, Jérôme (et leurs deux futurs bouts de choux), Magali, Jean-Yves, Philippe, Elie, Céline, Romain, Nicolas E. (allez Nancy !!), Clémence, Jessy, Audrey, Ludo et Guillaume (il risque d'être cité plusieurs fois donc ne trouvez pas ça bizarre). Merci pour nos soirées, week-end à Grenoble, murder et leur présence quotidienne. Merci spécial à Magali qui me supporte depuis de nombreuses années déjà (8 ans). Egalement à Jean-Baptiste et Virginie pour nos soirées poker et à Nicolas G. Merci aux personnes du labo de médecine pour m'avoir initié au poker.

Un merci à ma troupe de théâtre estivale avec qui j'ai passé deux séjours fort agréables (Camille, Mathias, Nicolas, Claire, Laetitia, Caroline, Franck, Joëlle,…) et à Valprivas (Brigitte, Jean-Pierre, Anne, Jean….). Un grand merci à Franck de m'avoir accueilli et hébergé par deux fois à Paris lors de mes séjours au Génoscope. Une pensée émue à Nelly qui nous a quittés durant l'écriture de ce manuscrit.

Merci aux différentes personnes qui ont participé à l'aventure chorale CIES durant ces quatre années. Plus particulièrement à Marie-Amélie (tu sais pourquoi), Piotr, Lars, Perrine, Jean-François, Bertrand, Pierre E…. Et bien sûr pas uniquement pour la chorale mais aussi pour les randos et soirées. La chorale est morte, vive la chorale !

Un énorme merci à Emilie, Sebastien, Thomas, Muriel et Sandra pour leur soutien tout au long de cette thèse, leur amitié et pleins de choses encore…. Les choses sont ce qu'elles sont mais ce que nous avons vécu tous ensemble restera à jamais. Bienvenue p'tite Lucie…..

Merci à Guillaume, Sandra, Magali, Perrine et surtout à Muriel d'avoir corrigé ce manuscrit.

Une pensée à Nico M. qui me supporte depuis plus de deux décennies et même si nos situations géographiques se sont un peu éloignées, je sais que je peux compter sur lui.

Bien évidemment, je ne pouvais oublier l'ami Guillaume. Notre complicité, notre envie commune de réaliser et d'organiser différentes activités en dehors du laboratoire, nos discussions diverses et variées ont été très importantes durant toutes ces années et se poursuivront encore et encore….

Merci à certaines d'avoir pimenté ma vie au cours de la rédaction de ce manuscrit, dans mes souvenirs, elle n'en gardera que plus de saveurs.

Pour terminer, je remercie toute ma famille, plus particulièrement mes parents qui ont toujours été présents, qui m'ont soutenu dans mes choix et ont parfois fait preuve de patience.

Cette thèse est dédiée à mes parents, grands-parents, Jean-Baptiste, Loréne, Léa et Emilie.

ABREVIATIONS

aa :	Acide Aminé
Apra :	Apramycine
ATCC :	American Type Culture Collection
ATP :	Adénosine TriPhosphate
AUD :	Amplifiable Unit of DNA
BAC :	Bacterial Artificial Chromosomes
BES :	BAC End Sequence
BET :	Bromure d'Ethidium
BIR :	Break Induced Replication
BLAST :	Basic Local Alignment Search Tool
bom :	Basis of Mobility
CDS :	Coding Sequence
CFP :	Cassure-Fusion-Pont
ChCl :	Chloroforme
ClCs :	Chlorure de Césium
clt :	*cis*-Acting Locus
CML :	Chronic Myelogenous Leukemia
CNS :	Centre National de Séquençage
CTP :	Cytosine TriPhosphate
DO :	Densité Optique
DIG :	Digoxygénine
DIG dUTP :	Digoxygénine-11-déoxyUridine TriPhosphate
DMSO :	DiMéthyl SulfOxyde
DSM :	Deutsche Sammlung von Mikroorganismen
EDTA :	EthyleneDiamine Tetraacetic Acid
FRT :	FLP Recognition Target
GOC :	Gene Order Conservation
GTP :	Guanosine TriPhosphate
Hfr :	High Frequency of Recombinaison
HT :	Hickey Tresner
ICE :	Integrative and Conjugative Element
IFR :	Institut Fédératif de Recherche
IPTG :	Isopropyl-D-thio-Galacto-Pyranoside
IS :	Inserted Sequence
kb :	Kilobase
LB :	Luria Bertani
Mb :	Mégabase
MHB :	Mycorrhiza Helper Bacteria
µF :	MicroFarraday
migS :	Migration Site
NAHR :	Non Allelic Homologous Recombination
NHEJ :	Non-Homologous End Joining
NR :	Non Redundant
NSA :	Nancy *Streptomyces ambofaciens*
nt :	Nucléotides
ORF :	Open Readind Frame
pb :	Paire de Bases
PCR :	Polymerase Chain Reaction
PFGE :	Pulse Field Gel Electrophoresis

PPI :	Protein-Protein Interaction
RBS :	Ribosome Binding Site
rpm :	Round Per Minute
SamDB :	*Streptomyces ambofaciens* DataBase
Sav :	*Streptomyces avermitilis*
Sco :	*Streptomyces coelicolor*
SDS :	Sodium Dodecyl Sulfate
sp. :	Species
Spec :	Spectinomycine
SSC :	Standard Sodium Citrate
Ssc :	*Streptomyces scabies*
TAE :	Tris-Acetate-EDTA
Tap :	Terminal Associated Protein
TBE :	Tris-Borate-EDTA
TIGR :	The Institute of Genomic Research
TIR :	Terminal Inverted Repeat
Tn :	Transposon
Tpg :	Terminal Protein Gene
TTP :	Thymidine TriPhosphate
UE :	Unité Enzymatique
UV :	Ultra Violet
X-Gal :	5-bromo-4-chloro-3-indolyl-D-galactopyranoside
YEME:	Yeast Extract Malt Extract

SOMMAIRE

PREAMBULE **1**

INTRODUCTION **5**

I-GENOMIQUE BACTERIENNE COMPARATIVE **5**

A- TAILLE DES GENOMES 5
B- GEOMETRIE DES GENOMES 6
B-1 La linéarité chez les bactéries *6*
B-2 Structures des extrémités des chromosomes linéaires bactériens *7*
C- CARACTERE POLYGENOMIQUE 8
D- COMPARTIMENTATION CHROMOSOMIQUE 10
E- POLARISATION DU CHROMOSOME 11
E-1 Orientation des gènes *11*
E-2 Composition en bases *12*
E-3 G/C skew *13*
E-4 Taux en bases GC en troisième position de codons (G+C3) *14*

II- PLASTICITE GENOMIQUE **15**

A- REARRANGEMENTS CHROMOSOMIQUES 15
A-1 Localisation des réarrangements *15*
A-2 Impact du nombre de répétitions sur la stabilité génomique *15*
A-3 Influence des facteurs environnementaux *16*
A-4 Forces s'opposant aux réarrangements *17*
A-5 Réparation des cassures double brin et réarrangements chromosomiques *18*
A-5a Agents responsables de cassures double brin 18
A-5a1 Partition chromosomique 18
A-5A1.1 Centromère et ségrégation chromosomique chez les eucaryotes 18
A-5A1.2 Systèmes de partition *parABS* chez les bactéries 19
A-5a2 Défaut de septation 19
A-5B Mécanismes de réparation chez les bactéries 20
A-5b1.1 Mécanismes de recombinaison homologue 20
A-5b1.2 Mécanisme NHEJ 20
A-5B2 Mécanismes de réparation des cassures double brin chez la levure 21
A-5B3 Mécanismes de réparation chez l'Homme 22
A-5C Cycle CFP 23
B- TRANSFERT HORIZONTAL CHEZ LES BACTERIES 25
B-1 Mécanismes du transfert horizontal *25*
B-1a Transformation naturelle 26
B-1b Transduction 26
B-2 la Conjugaison bactérienne *26*
B-2a "Integrated Conjugative Element" (ICE) 27
B-2b Les plasmides conjugatifs 28
B-5 Maintien des séquences acquises horizontalement *31*
B-5a Contribution au fitness 31
B-5b Contribution de l'interaction avec d'autres protéines sur le maintien de gènes transférés 32

III- OBJECTIFS DE LA THESE **34**

A- LES STREPTOMYCES 34

A-1 Cycle de développement *34*
A-2 Le génome des Streptomyces *35*
A-3 Génomique comparée chez les Streptomyces *36*
A-4 Instabilité génétique chez les Streptomyces *37*
B- OBJECTIFS DE LA THESE 38

MATERIEL ET METHODES 41

I- SOUCHES BACTERIENNES, VECTEURS ET CONDITIONS DE CULTURE 41

A- SOUCHES BACTERIENNES 41
B- VECTEURS 42
C- MILIEUX DE CULTURE 52

II- MANIPULATION DES ACIDES NUCLEIQUES 52

III- TRANSFORMATION DES SOUCHES D'E. COLI 53

A- LIGATION 53
B- ELECTROPORATION DES CELLULES D'E. COLI 54
B-1 Préparation de cellules compétentes d'E. coli *54*
B-2 Introduction des cosmides ou BAC dans E. coli BW25113/pIJ790 et E. coli BW25113/pKD20 ..... *54*
B-3 Construction des cosmides modifiés *55*

IV- CONJUGAISON INTERGENERIQUE E. COLI-S. AMBOFACIENS 55

V- HYBRIDATION MOLECULAIRE 56

VI- REACTION DE SEQUENÇAGE 56

VII- STRATEGIE DE REMPLACEMENT OU D'INCORPORATION AU LOCUS CHROMOSOMIQUE CHEZ STREPTOMYCES 57

A- CHOIX DES AMORCES ET FABRICATION DU PRODUIT DE PCR 57
B- CONSTRUCTION DE L'ALLELE MODIFIE SUR LE COSMIDE RECOMBINANT 57
C- TRANSFERT CONJUGATIF DU COSMIDE OU DU BAC MODIFIE D'E. COLI VERS STREPTOMYCES ET SELECTION DES RECOMBINANTS. 58

VIII- REACTION DE POLYMERISATION EN CHAINE (PCR) 58

RESULTATS 59

I- SEQUENÇAGE DES REPETITIONS TERMINALES INVERSEES DE LA SOUCHE DE *S. AMBOFACIENS* DSM40697 59

A- CLONAGE DANS LE VECTEUR M13 60
B- CLONAGE DANS PCNS 62
C- FINITION DES SEQUENCES 63
C-1 Méthode de lissage des séquences *64*
C-2 Séquençage du fragment terminal *65*
C-2a Régions terminales de la souche de *S. ambofaciens* DSM40607 66
C-2b Régions terminales de la souche de *S. ambofaciens* ATCC23877 67
D- BILAN DU SEQUENÇAGE 69
E- ANALYSE BIO-INFORMATIQUE DES SEQUENCES 70
E-1 Module d'assemblage SAMASSEMBLER *70*
E-2 Module d'annotation SAMANNOT *71*
E-3 Module de génomique comparée : SAMCOMP *71*

II- VARIABILITE INTRASPECIFIQUE DES REPETITIONS TERMINALES INVERSEES DU CHROMOSOME DE *S. AMBOFACIENS* 73

A- FRONTIERES DES TIR 73
B- COMPARAISON DU CONTENU EN GENES DES TIR DES DEUX SOUCHES DE S. AMBOFACIENS 74
B-1 La région commune entre les TIR des deux souches 75
B-2 Régions souche-spécifiques 76
C- CONSERVATION DES TELOMERES 77

III- EVOLUTION DU GENOME DES STREPTOMYCES 79

A- CONSERVATION DE LA REGION CENTRALE 79
B- LES REGIONS SPECIFIQUES D'ESPECES PAR COMPARAISON AVEC LES AUTRES GENOMES DE STREPTOMYCES 81
C- PROPORTION DE CDS SPECIFIQUES DANS LES REGIONS TERMINALES 83
D- REGIONS DE SYNTENIE DEGENEREE ET FLUX DE GENES 84

IV- ETUDE DYNAMIQUE DU CHROMOSOME : CLONAGE MULTIPLE DE SITES PARS DANS LES REGIONS INSTABLES 87

A- PARTITION CHROMOSOMIQUE CHEZ LES STREPTOMYCES 87
B- INSERTION DE SITES PARS 88
B-1 La technique de PCR-Targeting 88
B-2 Choix de la localisation de l'insertion des sites parS 89
B-3 Clonage réitératif de sites parS 91
B-3a Amplicon *parS* 91
B-3b Dimère d'amplicons *parS* 92
B-3c Clonage réitératif 93
B-3c1 Clonage dans le plasmide pIJ2925 93
B-3c2 Clonage dans le plasmide pIJ773* 94
B-3c2.1 Construction du plasmide pIJ773* 94
B-3c2.2 Clonage réitératif dans le plasmide pIJ773*1 95
C- MODIFICATION DES COSMIDES. 98
C-1 Amplification des cassettes Apra-parS 98
C-2 Incorporation des sites parS dans les cosmides 99
D- OBTENTION DES SOUCHES S. AMBOFACIENS DSM40697 MODIFIEES 101
D-1 Analyse des souches de S. ambofaciens DSM40697 101

DISCUSSION 105

I- LE GENOME DES *STREPTOMYCES* EST COMPARTIMENTE 105

II- DYNAMIQUE DE L'EVOLUTION DES TIR 107

A- HOMOGENEISATION DU CONTENU DES TIR 107
B- EVOLUTION DE LA TAILLE DES TIR 109

III- MECANISMES DE LA VARIABILITE DES REGIONS TERMINALES 110

A- CONJUGAISON CHEZ LES STREPTOMYCES 110
A-1 Les plasmides conjugatifs 110
A-2 Le transfert conjugatif 110
A-2 Modèle "End First" 111
A-3 Gradient de transfert des régions terminales 112
B- MECANISMES DE REPARATION DES CASSURES DOUBLE BRIN 113
B-1 recombinaison homologue chez les Streptomyces 113
B-2 Système NHEJ chez les Streptomyces 113
B-3 Intervention des mécanismes de réparation dans la variabilité des régions terminales 114

IV- REGIONS TERMINALES ET ROLE DANS L'ADAPTATION 114

A- ANTIBIOTIQUES ET ROLE DANS L'ADAPTATION 114

B- INTERACTIONS AVEC D'AUTRES MICROORGANISMES 115

V- MODE D'ACQUISITION DES NOUVELLES FONCTIONS DANS LES TIR 116

A- INTERACTIONS ENTRE REPLICONS LINEAIRES 116
B- REMPLACEMENT DES EXTREMITES ENTRE REPLICONS LINEAIRES 117

VI- MAINTIEN DES SEQUENCES ACQUISES SUR LE LONG TERME 117

VII- ETUDE DYNAMIQUE DES REGIONS TERMINALES DES *STREPTOMYCES.* 118

A- STRUCTURE CHROMOSOMIQUE POTENTIELLE DES MUTANTS DE S. AMBOFACIENS DSM40697 118
B- INFLUENCE DE LA PRESENCE DES SITES PARS SUR LES DIFFERENTS PROFILS OBTENUS. 121
C- LIEN ENTRE PARTITION CHROMOSOMIQUE ET REARRANGEMENTS DES EXTREMITES 122

PERSPECTIVES 123

I- A COURTS TERMES 123

II- A MOYENS ET LONGS TERMES 125

REFERENCES BIBLIOGRAPHIQUES 127

INTRODUCTION

PREAMBULE

L'émergence de nouvelles techniques de biologie moléculaire a permis de préciser et d'appréhender l'étude des génomes et en particulier des génomes bactériens. Le séquençage de nombreux génomes bactériens a été réalisé (520 génomes à la date du 20 Août 2007) et d'autres sont actuellement en cours. L'ensemble de ces données a permis de montrer la forte hétérogénéité dans l'organisation et les structures des génomes, ainsi que les forces évolutives agissant sur ces derniers. Ainsi, une importante diversité en taille, configuration et composition entre les différents génomes peut être observée.

Parmi les bactéries présentant des caractéristiques génomiques remarquables, le genre bactérien *Streptomyces* a fait l'objet de ce travail. Les *Streptomyces* possédent un chromosome linéaire de grande taille (8-11 Mb) à fort taux en bases G-C (71-73%). De façon appliquée, ces bactéries sont utilisées pour leur capacité à synthétiser des molécules à intérêts en pharmacie ou encore dans l'agro-alimentaire. Ces bactéries sont également étudiées au sein du laboratoire de Génétique et Microbiologie (Nancy Universités) depuis de nombreuses décennies pour le phénomène d'instabilité génétique et chromosomique important qui affecte le chromosome. Ces phénomènes d'instabilité provoquent de grands réarrangements au sein du chromosome conduisant à des pertes importantes de l'information génétique au niveau des extrémités chromosomiques.

Au début de ce travail, deux génomes de *Streptomyces* étaient entièrement séquencés. L'analyse et la comparaison des génomes de *Streptomyces coelicolor* (par Bentley *et al.*, 2002), modèle d'étude des *Streptomyces*, et *Streptomyces avermitilis* (par Ikeda *et al.*, 2003) ont permis de corroborer des études préalables sur l'organisation de l'information génétique et d'obtenir une vision dynamique de l'évolution des génomes de ces bactéries. Ainsi, les gènes dits « essentiels » sont confinés dans la région centrale du chromosome alors que les gènes dits « de contingence » sont situés au niveau des régions terminales, encore appelées « bras chromosomiques ». La comparaison entre ces deux génomes a permis de préciser que les régions terminales sont spécifiques d'espèces alors que les régions centrales sont conservées. La comparaison entre le génome de *Mycobacterium tuberculosis* (agent de la tuberculose et actinomycète apparenté à *Streptomyces*) et *S. coelicolor* a permis d'émettre l'hypothèse d'une acquisition de la linéarité par recombinaison entre le chromosome circulaire ancestral et un réplicon linéaire.

Un projet de séquençage des régions terminales (environ 3 Mb) et le séquençage partiel de la région centrale de la souche de *Streptomyces ambofaciens* ATCC23877 avait été initié avant mon arrivée au laboratoire. Ce projet est le fruit d'une collaboration entre le laboratoire (équipe de Pierre Leblond, Nancy), l'Institut de Génétique et Microbiologie (équipe de Jean-Luc Pernodet, Orsay) et le Génoscope (équipe de Valérie Barbe, Evry). La comparaison des séquences obtenues avec les génomes de *S. coelicolor* et *S. avermitilis* a permis l'observation du mode d'évolution des régions terminales et des régions centrales conservées.

Au niveau des extrémités des bras chromosomiques, le chromosome des *Streptomyces* présente des régions répétées identiques aux deux extrémités. Ces régions sont appelés TIR pour

« Terminal Inverted Repeat » (ou Répétitions Terminales Inversées). Parallèlement au projet de séquençage des régions terminales de la souche de *S. ambofaciens* ATCC23877, j'ai eu pour objectif le séquençage des TIR de l'autre souche de *S. ambofaciens* étudiée au sein du laboratoire, la souche DSM40697, à leur analyse et comparaison avec les TIR de la souche ATCC23877. Ces résultats ont permis de montrer la localisation précise des régions spécifiques, cette fois-ci, entre souches de la même espèce.

L'étude et la caractérisation de mutants spontanés pour la pigmentation des colonies de *S. ambofaciens* ont permis d'émettre l'hypothèse d'un lien entre instabilité chromosomique et partition chromosomique. Pour permettre d'étudier cette hypothèse, des souches de *S. ambofaciens* contenant deux centres de partition chromosomique ont été construits. Ce type de mécanisme pourrait être à la base des grands réarrangements et de l'évolution des génomes des *Streptomyces*.

L'introduction de cette thèse présente la diversité des caractéristiques des génomes bactériens, ainsi que les mécanismes possibles engendrant les réarrangements et l'impact du transfert horizontal sur l'évolution des génomes.

Les résultats présentent dans un premier temps la stratégie et le séquençage des TIR de la souche de *S. ambofaciens* DSM40697 et leur comparaison avec la souche de *S. ambofaciens* ATCC23877, montrant l'évolution des ces extrémités par intégration et échanges de matériels génétiques (article dans *Journal of Bacteriology*). De plus, la comparaison des régions terminales de la souche de *S. ambofaciens* ATCC23877 avec les génomes de *S. coelicolor*, *S. avermitilis* et *Streptomyces scabies* est étudiée. Les forces évolutives agissant sur ces extrémités sont ainsi discutées (article dans *Molecular Biology and Evolution*). Enfin, dans une dernière partie, l'obtention de souches possédant un second centre de partition chromosomique est présentée. L'analyse de ces souches était en cours au moment de l'écriture de ce manuscrit.

La discussion propose des hypothèses sur l'obtention des extrémités chromosomiques, les mécanismes moléculaires intervenant dans la spécificité des régions terminales et du gradient de fréquence croissante de réarrangements vers ces extrémités. Parallèlement, des stratégies alternatives pour l'obtention de souches présentant un second centre de partition sont proposés.

Certains résultats sont résumés dans la partie « Résultats ». Les publications suivantes détaillent ces résultats et sont insérées dans les chapitres correspondants :

Publication n°1 :
Choulet F., Gallois A. (co-auteur), 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. Intraspecific variability of the terminal inverted repeat of the linear chromosome of *Streptomyces ambofaciens*. **Journal of Bacteriology**, Sept. 2006, p. 6599-6610, Vol. 188, No. 18.

Publication n°2:
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. Evolution of the terminal regions of the *Streptomyces* linear chromosome. **Molecular Biology and Evolution**, Déc. 2006, p. 2361-2369, Vol. 23, No 12.

Ces travaux ont été exposés à différents congrés nationaux et internationaux :

Gallois et al., Variabilité intraspécifique des répétitions terminales inversées chez *Streptomyces ambofaciens*. **Journées des Actinomycètes 2006**, *Université Claude Bernard Lyon I, Lyon, France, Juin 2006*

Choulet et al., Horizontal transfer and evolution of the terminal regions of the linear chromosome of *Streptomyces*. **Journées des Actinomycètes 2006***, Université Claude Bernard Lyon I, Lyon, France, Juin 2006*

Choulet et al. Horizontal transfer and evolution of the terminal regions of the linear chromosome of *Streptomyces*. **Rencontres des Microbiologistes de l'INRA***, Dourdan, France, Juin 2006*

Choulet et al., DNA rearrangements and integration of exogenous information in the terminal regions of the linear chromosome of *Streptomyces ambofaciens* revealed by comparative genomics, **Prokagen 2005**, *2nd European Conference on Prokaryotic Genomes, University of Göttingen, Germany, Sept. 2005*

Choulet et al., Sequence annotation of the terminal regions of the chromosome of *Streptomyces ambofaciens*. Comparative genomics and evidence for an important interspecific variability, **5èmes Journées Ouvertes de Biologie, Informatique et Mathématiques***, Montréal, Canada, Juin 2004, article JO72*

Choulet et al., Mise en évidence d'une importante variabilité entre les régions terminales des chromosomes de deux espèces phylogénétiquement proches : *Streptomyces ambofaciens* et *Streptomyces coelicolor*, **Rencontre du club des Actinomycètes***, Centre d'Ingénierie des protéines, Université de Liége, Belgique, Mai 2004*

Choulet et al., Evidence for a High Heterogeneity Between the Terminal Regions of the Chromosomes of Two Phylogenetically Close Species, *Streptomyces ambofaciens* and *Streptomyces coelicolor* by Genome Comparison, **Dissemination meeting***, Guilford, University of Surrey, UK, Juil 2003*

Choulet et al., Sequence analysis of the *Streptomyces ambofaciens* ATCC23877 larce Terminal Invertes Repeat (210 kb) and evidence of hugh heterogeneity with the Streptomyces coelicolor A3(2) genome, **Biology of streptomycetes and related actinomycetes***, University of Munster, Germany, Fév. 2003*

I NTRODUCTI ON

Le séquençage d'un nombre toujours croissant de génomes bactériens permet de comparer la structure de génomes bactériens d'espèces proches et d'espèces éloignées donnant une image dynamique de l'évolution génomique chez les bactéries.

I-Génomique bactérienne comparative

La comparaison des différents génomes séquencés a confirmé une forte hétérogénéité en termes de taille et de configuration notamment.

A- Taille des génomes

Une grande variabilité en taille est observée au sein des génomes bactériens, l'amplitude de cette variabilité pouvant être de plus de 20 fois. Cette variation des génomes est le résultat de différents événements génétiques, comme la duplication de gènes, l'acquisition horizontale ou encore la perte d'information génétique.

La taille des génomes au sein d'un même phylum peut être extrêmement variable. Ainsi *Tropheryma whipplei*, dont le génome mesure 0,93 Mb ainsi que *Streptomyces*, dont la taille varie entre 8 et 10 Mb, appartiennent au phylum des actinobactéries. De même, chez les protéobactéries, *Carsonella ruddii* (γ-protéobactérie) possède un génome de 0,16 Mb et *Bradyrhizobium japonicum*, en revanche, un génome de 9,1 Mb.

La niche écologique joue un rôle majeur dans cette variabilité. Les organismes vivant dans une niche complexe et changeante présentent un génome de grande taille, comme par exemple *B. japonicum* (9 105 828 pb) et *Streptomyces scabies* (10 148 695 pb), bactéries vivant dans le sol, qui font partie des plus grands génomes séquencés. En revanche, les bactéries, dont la niche est stable, souvent en association étroite avec un organisme hôte, présentent les génomes les plus petits et un degré de compaction de ce génome important (Klasson et Andersson 2004). Ceci est le cas par exemple des génomes de *Nanoarchaeum equitans* (490 885 pb), de *Mycoplasma genitalium* (580 076 pb) et de *C. ruddii* (159 662 pb), bactéries symbiotiques obligatoires et qui font partie des plus petits génomes séquencés (Fraser *et al.* 1995; Nakabachi *et al.* 2006).

Une relation directe entre taille du génome et nombre de phases codantes a été soulignée. Ainsi, un nombre important de gènes rencontrés dans les bactéries à cycle de vie autonome peut être associé à la capacité d'adaptation aux changements des conditions environnementales. En revanche, les bactéries vivant dans une niche stable possèdent un génome beaucoup plus réduit.

B- Géométrie des génomes

B-1 LA LINEARITE CHEZ LES BACTERIES

La plupart des bactéries étudiées possèdent un chromosome circulaire. La linéarité chromosomique est sporadique. Ainsi, dans trois phyla bactériens différents, un chromosome sous forme linéaire est retrouvé: chez les actinobactéries, avec les *Streptomyces* et *Rhodococcus* (Lin *et al.* 1993; Pisabarro *et al.* 1998); chez les spirochètes, avec *Borrelia burgdorferi* (Ferdows et Barbour 1989) et chez les α-protéobactéries avec *Agrobacterium tumefaciens* (Allardet-Servent *et al.* 1993). L'éloignement phylogénétique entre ces trois phyla et la distribution du caractère linéaire du chromosome au sein d'un arbre phylogénétique suggèrent fortement que la linéarité est apparue de façon indépendante dans ces lignées (Volff et Altenbuchner 2000).

Au sein même des actinomycètes, la distribution de la linéarité indique des événements de linéarisation indépendants. La comparaison de la séquence chromosomique de *Rhodococcus* sp. RHA1 et de son protéome avec les génomes d'autres actinomycètes indique que les membres du phylum phylogénétiquement les plus proches sont *Nocardia farcinica* et *Mycobacterium tuberculosis* H37Rv (Fig. 1). Ce résultat confirme l'étude phylogénétique préalable des ARN 16s (Gurtler *et al.* 2004). Or, ces deux bactéries possèdent un chromosome circulaire. En revanche, les 300 premières paires de bases étant similaires entre *Rhodoccoccus* et *Streptomyces,* l'origine des extrémités chromosomiques pourrait être commune entre ces deux espèces (McLeod *et al.* 2006). Et des actinomycètes dont le génome a été entièrement séquencé seuls les *Streptomyces* possèdent un chromosome linéaire. Cette observation montre que la linéarisation des chromosomes s'est probablement produite plus d'une fois durant l'évolution des actinomycètes. Une hypothèse expliquant l'apparition de la linéarité chez les bactéries, est l'ouverture d'un chromosome circulaire ancestral par intégration d'un réplicon linéaire.

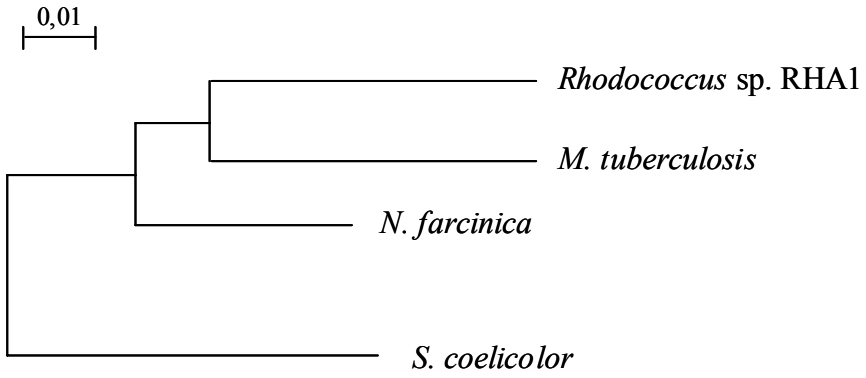


Figure 1 : Reconstruction des relations phylogénétiques de quelques actinomycètes.
Le chromosome de *Rhodococcus* et de *S. coelicolor* sont linéaires alors que celui de *M. tuberculosis* et *N. farcinica* sont circulaires.

Le changement de la géométrie chromosomique est observable en conditions de laboratoire. Ainsi, l'apparition de mutants présentant un chromosome circulaire comme chez *Borrelia* ou *Streptomyces* a été constatée (Lin *et al.* 1993; Ferdows *et al.* 1996; Fischer *et al.*

1997b), montrant que les réplicons linéaires peuvent se maintenir sous forme circulaire.

De même, la linéarisation du chromosome circulaire d'*Escherichia coli* par intégration des extrémités du phage N15, en conditions de laboratoire, a été mis en évidence (Cui *et al.* 2007). La structure linéaire du chromosome est stable au fil des générations. Ce changement de conformation ne modifie pas le niveau d'expression des gènes. En effet, l'expression de 4300 gènes a été comparée entre les souches à chromosome circulaire et linéaire. Seulement trois de ces gènes, *avtA* (Alanine-Valine transaminase), *sulA* (gène d'inhibiteur du cycle cellulaire) et *ygdB*_b2824 (codant une protéine hypothétique) présentent un niveau d'expression significativement différent. L'expression des gènes *avtA* et *sulA* augmente respectivement de 6,49 et 5,55 fois alors que celle du gène *ygdB*_b2824 diminue de 0,22 fois.

B-2 STRUCTURES DES EXTREMITES DES CHROMOSOMES LINEAIRES BACTERIENS

Chez les bactéries, deux types d'extrémité sont retrouvés (Volff et Altenbuchner 2000). Chez les *Streptomyces*, les extrémités sont constituées de sept séquences palindromiques très conservées. Une séquence d'environ 280 nt est laissée sous forme simple brin après réplication du chromosome (Chang et Cohen 1994). Une structure secondaire sous forme tige boucle a été prédite (Fig. 2). Un complexe formé de protéines terminales, nommées Tpg (Terminal Protein Gene), fixées à ces séquences télomériques, au niveau des palindromes II et III, par un domaine de fixation à l'ADN (HTH), et Tap (Terminal Associated Protein), associée à Tpg, est rencontré au niveau de ces extrémités. Ces protéines sont nécessaires au maintien des chromosomes sous forme linéaire, l'inactivation de leur gène conduisant à la perte du télomère et la circularisation du chromosome (Bao et Cohen 2003). Le phage linéaire Φ29 de *Bacillus subtilis* possède également une protéine terminale liée de façon covalente à ses extrémités. Cette protéine est nécessaire pour la réplication de ce phage à partir des extrémités (Salas 1991). En effet, à la différence des *Streptomyces*, où la réplication est initiée à partir du centre du chromosome, la réplication du phage Φ29 débute à partir des extrémités grâce à la protéine terminale et se poursuit de façon continue le long du génome.

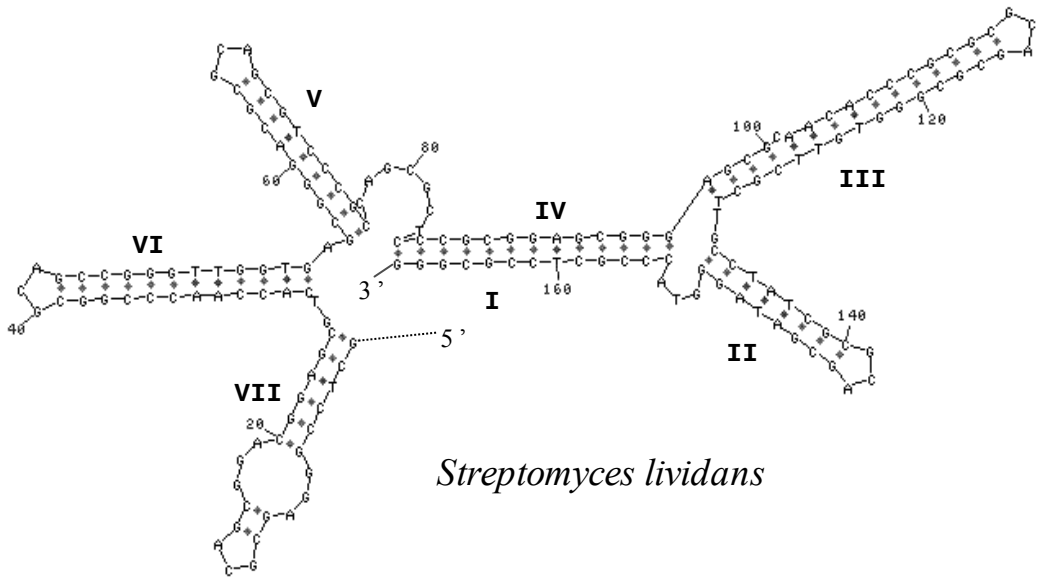


Figure 2 : Structures secondaires putatives des télomères de *S. lividans*
La structure putative présentée a été obtenue avec DNA fold 3.1 (NaCl 1M, 30°C).

Chez *Bo. burgdorferi* et *A. tumefaciens*, les extrémités présentent une structure tige boucle fermée (Barbour et Garon 1987; Allardet-Servent *et al.* 1993) similaires aux extrémités du prophage N15 d'*E. coli* (Ravin *et al.* 2000).

Les structures télomériques différentes entre chromosomes linéaires de même que la distribution phylogénétique des lignées présentant un chromosome linéaire appuient l'hypothèse d'une acquisition indépendante de la linéarité chez les bactéries.

C- Caractère polygénomique

Les génomes bactériens présentent un nombre de réplicons variable. Le tableau 1 présente la composition génomique chez quelques espèces bactériennes.

Bactérie	Taille du génome	Nature du réplicon	Géométrie du réplicon
Bo. burgdorferi	1,5 Mb	Chromosome (0,9 Mb)	Linéaire
		Plasmide (610 kb)	Linéaire (12)/Circulaire (9)
Burkholderia xenovorans	9,73 Mb	Chromosome (4,9 Mb)	Circulaire
		Chromosome (3,36 Mb)	Circulaire
		Mégaplasmide (1,47 Mb)	Circulaire
Vibrio cholerae	4,03 Mb	Chromosome (2,96 Mb)	Circulaire
		Chromosome (1,07 Mb)	Circulaire
Deinococcus radiodurans	3,28 Mb	Chromosome (2,65 Mb)	Circulaire
		Chromosome (0,41 Mb)	Circulaire
		Mégaplasmide (177 kb)	Circulaire
		Plasmide (45,7 kb)	Circulaire
Streptomyces coelicolor	9 Mb	Chromosome (8,6 Mb)	Linéaire
		Plasmide (356 kb)	Linéaire
		Plasmide (31 kb)	Circulaire
A. tumefaciens	5,67 Mb	Chromosome (2,84 Mb)	Circulaire
		Chromosome (2,07 Mb)	Linéaire
		Plasmide (543 kb)	Circulaire
		Plasmide (214 kb)	Circulaire

Tableau 1: Exemple de bactéries possédant plusieurs réplicons
La taille des différents réplicons est indiquée entre parenthèses. Pour *Bo. Burgdorferi,* la somme des tailles des 21 plasmides (12 linéaires, 9 circulaires) est indiquée. Mb : mégabases ; kb : kilobases.

Chez les *Streptomyces*, les souches isolées du sol possèdent un chromosome unique et linéaire. Des souches peuvent posséder un ou plusieurs plasmides circulaires et/ou linéaires. La distinction entre plasmide et chromosome est difficile à définir en général. Les chromosomes ont été définis comme les réplicons portant les gènes essentiels, comme chez *Ralstonia solanacearum* (Salanoubat *et al.* 2002) ou *Yersinia pestis* (Parkhill *et al.* 2001). Jusqu'à peu, les réplicons portant des gènes essentiels étaient considérés comme « chromosome ». La découverte de gènes de « ménage » plasmidique (réplication, partition) fait émerger l'hypothèse du transfert de gènes essentiels vers des plasmides ancestraux comme chez *Vibrio cholerae* ou *A. tumefaciens* (Bentley et Parkhill 2004).

D- Compartimentation chromosomique

La comparaison du contenu en gènes, entre génomes d'espèces proches ou entre souches d'une même espèce, a permis de mettre en évidence une localisation préférentielle des gènes le long du chromosome suivant leur fonction. Ainsi, les gènes du métabolisme central (réplication, traduction…) seraient essentiellement localisés autour de l'origine de réplication et les gènes accessoires plutôt autour du terminus de réplication.

La comparaison des génomes des souches d'*Escherichia coli* O157:H7 et K12 (Perna *et al.* 2001) illustre ce phénomène. En effet, leur génome est colinéaire, à l'exception d'une inversion de 422 kb chevauchant le terminus de réplication. Cette colinéarité est interrompue par la présence d'îlots génomiques. Ainsi, *E. coli* O157:H7 possède 177 îlots génomiques, contenant des gènes accessoires, recouvrant 1,34 Mb, principalement situés autour du terminus de réplication.

Le chromosome de *Saccharopolyspora erythraea* présente une région de 4,4 Mb, entourant l'origine de réplication, qui porte 85% des gènes prédits essentiels (Oliynyk *et al.* 2007). En dehors de cette région, le reste du chromosome subit la majorité des événements d'insertions. De plus, dans cette région, un très grand nombre d'*IS* est retrouvé (93 correspondant à 13 familles différentes) représentant 2,3% du chromosome.

La comparaison des trois génomes séquencés de *Frankia* (5,43 Mb, 7,5 Mb et 9,04 Mb) montre une synténie qui décline en approchant du terminus de réplication (Normand *et al.* 2007). Cette perte de synténie serait la conséquence de réarrangements de gènes, des duplications, le transfert horizontal de gènes ou des délétions à proximité du terminus de réplication. Ces événements peuvent en partie expliquer les différences de tailles observées entre ces génomes. Ces bactéries sont en symbiose avec les plantes. Les différences de taille reflèteraient les niches écologiques divergentes ou des partenaires symbiotiques distincts. La souche CcI3 est spécialiste de plantes isolées en Australie et dans les îles du pacifique, présentant un spectre symbiotique étroit, alors que la souche EAN, qui possède le plus grand génome, est retrouvée dans diverses niches écologiques et possède donc un spectre plus large. Cette dernière souche a subi le plus grand nombre de duplication de gènes parmi les trois et la grande majorité de ces duplications est localisée dans les régions terminales. Le génome de la souche ACN est plus stable au contraire des deux autres souches, avec le plus grand nombre et pourcentage d'ORF orphelines. La stabilité de cette souche reflète une pression de sélection plus faible en raison d'une niche relativement riche et stable, ne nécessitant pas l'apport ou la duplication d'informations génétiques.

Chez *Streptomyces coelicolor* et *Streptomyces avermitilis*, la région centrale (ou « core ») du chromosome contient la majorité des fonctions dites « essentielles » à la croissance végétative de la bactérie (Bentley *et al.* 2002). En revanche, les régions au-delà de cette région centrale, nommées bras chromosomiques, portent les gènes aux fonctions accessoires.

La localisation des gènes non essentiels préférentiellement dans les régions terminales est une caractéristique commune avec les eucaryotes. Chez *Saccharomyces cerevisiae*, l'étude par délétion de 22 gènes sur une région de 70 kb, localisés sur le chromosome XI, a montré que les mutants ne présentant pas de phénotypes particuliers présentaient des délétions dans les régions subtélomériques alors que les délétions létales touchaient les régions centrales et donc les gènes essentiels à la survie de la levure en conditions de laboratoire (Fairhead *et al.* 1998). Ces régions contiendraient donc des gènes non essentiels pour *Sa. cerevisiae*. Une étude, à plus grande échelle,

a permis l'analyse de 2026 ORF (Winzeler *et al.* 1999). Cette analyse a confirmé que les gènes non essentiels à la croissance de *Sa. cerevisiae* en conditions de laboratoire sont situés dans les régions terminales.

E- Polarisation du chromosome

E-1 ORIENTATION DES GENES

La majorité des gènes est retrouvée sur le brin direct de réplication. Une première hypothèse pour expliquer cette observation est la contre sélection d'une localisation sur le brin indirect de réplication pour prévenir des collisions possibles entre l'ADN polymérase et la machinerie de transcription (Nomura et Morgan 1977). En effet, la réplication et la transcription peuvent se produire simultanément sur la même molécule d'ADN. Ainsi, cette contrainte provoquerait un biais d'orientation des gènes.

La vitesse de croissance de la bactérie et la réplication seraient également impliquées dans l'orientation des gènes sur le brin direct. Ce biais dépend fortement de la composition de l'ADN polymérase. En effet, les bactéries qui possèdent une ADN polymérase à deux sous-unités α distinctes présentent le plus fort biais d'orientation avec en moyenne 78% des gènes situés dans le sens de la réplication (Rocha 2002). Chaque sous-unité serait dédiée à la réplication d'un brin d'ADN (Dervyn *et al.* 2001). De plus, parmi les bactéries contenant deux sous-unités d'ADN polymérase, les biais les plus importants sont retrouvés dans les bactéries à faible vitesse de croissance comme *My. genitalium* (80%) par rapport aux bactéries à croissance rapide comme *Ba. subtilis* (74%). Cette observation est également vérifiée pour les bactéries avec une seule sous-unité α où *M. tuberculosis* présente un biais supérieur par rapport à *E. coli* (59% contre 55%).

Ce biais d'orientation des gènes est encore plus marqué pour les gènes essentiels (Rocha et Danchin 2003b). Ainsi, chez *E. coli,* 55% de ses gènes et 76% de gènes essentiels sont situés sur le brin direct. Une situation analogue est observée chez *Ba. subtilis* où 75% de gènes et 94% des gènes essentiels sont localisés sur le brin direct. Chez *Ba. subtilis*, parmi les gènes fortement exprimés, une orientation sur le brin direct est plus fréquente pour les gènes essentiels. En revanche, l'orientation des gènes non essentiels hautement exprimés est moins biaisée que les gènes essentiels. L'orientation d'un gène dépend donc de son caractère « essentiel » au sein du génome.

Une hypothèse pour expliquer l'orientation des gènes essentiels sur le brin direct est, dans le cas de la rencontre entre la machinerie transcriptionnelle et de réplication, que cette rencontre va provoquer un arrêt de la transcription pouvant produire au final des protéines tronquées. Ces protéines tronquées sont généralement non fonctionnelles. Mais si une protéine fait partie d'un complexe multimérique, un phénotype de dominant négatif est souvent observé (Pakula et Sauer 1989). Ce phénotype est délétère, les complexes inactifs de protéines essentiels désorganisant les fonctions essentielles nécessaires à l'organisation cellulaire, comme par exemple les composants des ribosomes ou de l'ARN polymérase (Rocha et Danchin 2003a).

E-2 COMPOSITION EN BASES

La composition nucléotidique d'un génome varie d'une espèce bactérienne à l'autre. Ce pourcentage varie de 26,5% pour *Wigglesworthia glossinidia* (Akman *et al.* 2002) à 74,9% pour *Anaeromyxobacter dehalogenans* (Sanford *et al.* 2002). Le pourcentage en bases G-C varie peu entre souches d'une même espèce mais est extrêmement variable entre les espèces.

Une corrélation entre pourcentage en bases G-C et niche écologique peut être mise en évidence. Les endosymbiotes ont une tendance à un enrichissement en A-T. L'analyse de 52 génomes bactériens a montré que le pourcentage moyen en bases G-C est de 38% pour les génomes de bactéries dépendantes d'hôtes contre 49% pour ceux des bactéries à vie autonome (Fig. 3) (Rocha et Danchin 2002).

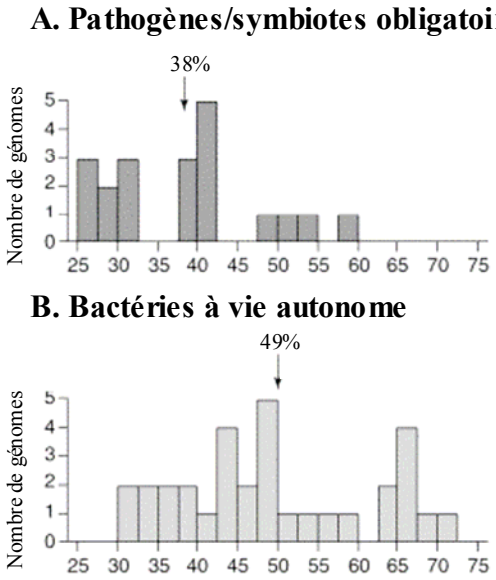


Figure 3 : Pourcentage en bases G+C des génomes bactériens (d'après Rocha et Danchin, 2002)
A. Génomes d'organismes pathogènes (ou symbiotiques) obligatoires
B. Organismes à vie autonome
Les flèches indiquent la valeur moyenne pour chaque catégorie.

Les éléments mobiles, comme les plasmides, les phages ou les séquences d'insertion (IS) présentent un fort taux en bases A-T par rapport au chromosome (enrichissement en A-T en moyenne de 4% et 2,7% respectivement pour les phages et les plasmides par rapport au chromosome de l'hôte). L'analyse de 245 gènes plasmidiques a révélé un taux en bases A-T supérieur de 2% en moyenne par rapport à leurs homologues chromosomiques (Rocha et Danchin 2002).

Plusieurs hypothèses peuvent expliquer ce biais. Les bactéries à petit génome ne présentent généralement pas les fonctions de réparation de l'ADN et de recombinaison (Moran 2002). Le mécanisme mutateur le plus rencontré est la déamination de la cytosine en uracile, qui sera répliqué en thymine. Ce type de mutation va engendrer un enrichissement en bases A-T. Une autre hypothèse est la disponibilité des différents nucléotides. En effet, en raison de son implication dans le métabolisme énergétique, la quantité en ATP est plus importante que celle en GTP ou CTP.

Rocha et Danchin privilégient plutôt une hypothèse sélectionniste. Les auteurs proposent que les génomes des bactéries pathogènes obligatoires ou symbiotiques tendent à devenir plus riches en A-T, comme les éléments génétiques mobiles, pour exploiter au mieux les ressources de la cellule. Ceci est en partie dû au coût énergétique élevé pour la synthèse des nucléotides GTP/CTP par rapport aux nucléotides ATP/TTP. Ainsi, dans un contexte où les ressources sont limitées, la sélection d'un biais mutationnel en faveur d'un contenu plus fort en A-T serait favorisée, cette richesse devenant un avantage en raison du coût moindre que représente la synthèse des nucléotides ATP/TTP. De même, les éléments extra-chromosomiques sont en compétition avec le reste du génome pour leur maintien, une utilisation préférentielle des bases A et T plus abondantes favoriserait la compétition en leur faveur (Rocha et Danchin 2002).

E-3 G/C SKEW

Un critère compositionnel remarquable est la proportion de bases G par rapport aux bases C (G-C/G+C) sur un brin appelé G/C skew, et réciproquement la proportion de bases A par rapport aux bases T. En effet la majorité des génomes bactériens séquencés présente un excès de bases G par rapport à C sur le brin continu et une inversion au niveau des sites d'initiation et de terminaison de la réplication, loci où le brin continu change de phase. Ce biais permet donc de faciliter la localisation de l'origine de réplication et le terminus. Ces transitions sont particulièrement bien marquées chez la plupart des bactéries, comme pour le chromosome d'*E. coli* K12, à de rares exceptions comme par exemple *Nostoc* sp. (Nikolaou et Almirantis 2005) ou encore chez *S. avermitilis* où aucun biais n'est observé (Fig. 4). Bien que le mécanisme responsable de ce biais ne soit pas connu, la distribution de motifs comme les séquences Chi et Rag, impliquées dans les événements de recombinaison homologue, riches en bases G, préférentiellement sur le brin direct, jouerait un rôle sur le GC skew (Lobry et Sueoka 2002).

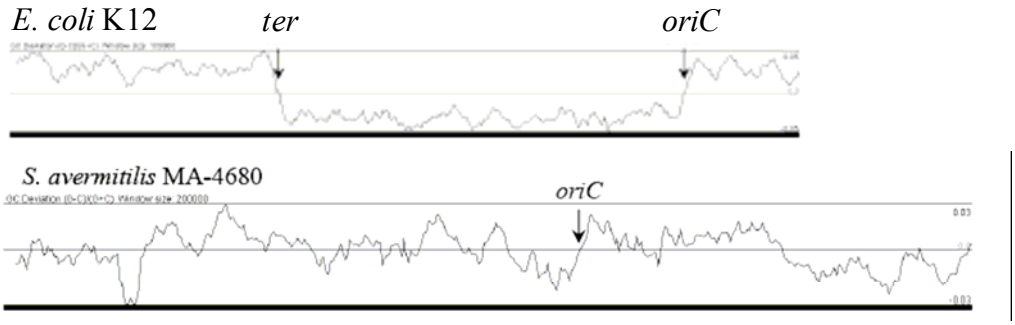


Figure 4 : Profils du GC skew du chromosome d'*E. coli* K12 et de *S. avermitilis*
Aucun biais du GC skew n'est identifiable pour *S. avermitilis*. Le terminus de réplication (*ter*) et l'origine (*ori*) sont indiqués.

Pour *Nostoc* sp., l'absence de transition pourrait être la conséquence de la présence de plusieurs origines de réplication. La recherche de la séquence des boîtes DnaA d'*E. coli* chez *Nostoc* sp. a montré l'existence possible de plusieurs origines de réplication fonctionnelles. En revanche, chez les archées, deux origines de réplication ont été découvertes dans le génome de *Aeropyrum pernix*, situées à chaque extrémité du chromosome circulaire. Ces deux origines

coïncident avec deux inversions du GC skew du génome (Zhang et Zhang 2005).

E-4 TAUX EN BASES GC EN TROISIEME POSITION DE CODONS (G+C3)

Une analyse de la composition en troisième position de codon a permis de distinguer deux groupes de bactéries en fonction de leur G+C3 (Daubin et Perriere 2003). Le premier groupe (comprenant *Ba. subtilis* et *E. coli*) est caractérisé par une baisse du G+C3 à proximité du terminus de réplication du chromosome.

En revanche, dans le second groupe, constitué entre autres de *Pseudomonas aeruginosa*, *Bacillus halodurans* ou encore *Mycoplasma pneumoniae*, la structure de leur génome est plus complexe. Des événements de transferts horizontaux ou de réarrangements génomiques auraient altéré la structure du G+C3 originale. Ainsi, chez *My. genitalium*, le génome présente une variation du contenu en G+C3 le long du chromosome sans relation avec la position vis à vis de l'origine et du terminus de la réplication (Kerr *et al.* 1997).

Plusieurs hypothèses peuvent expliquer le faible pourcentage du G+C3 à proximité du terminus de réplication. Chez *E. coli*, une étude a montré que les gènes acquis par transfert horizontal, étaient localisés de façon préférentielle au niveau de la région du terminus (Lawrence et Ochman 1998), incluant plusieurs prophages de *Ba. subtilis* (Kunst *et al.* 1997). Or, les séquences, issues de transferts horizontaux, notamment véhiculées par des éléments génétiques mobiles, présentent le plus souvent un pourcentage en G-C faible (voir paragraphe E2).

Une autre hypothèse est basée sur la superhélicité de l'ADN, de cette région terminale, impliquée dans la résolution des dimères de chromatides. La protéine Fis intervient dans la structuration de la chromatide et la régulation de la transcription par fixation en amont de plusieurs gènes (Finkel et Johnson 1992; Schneider *et al.* 2001; Travers *et al.* 2001). Cette protéine possède un grand nombre de sites de fixation, riches en A-T, dans une région de 1 Mb autour du terminus de réplication chez *E. coli* (Ussery *et al.* 2001).

Une autre hypothèse est l'utilisation de systèmes spécifiques de réparation. Chez *E. coli*, quand le complexe de réplication rencontre une lésion, celui-ci s'arrête et cette lésion est réparée (Cox *et al.* 2000). Le système SOS intervient et les polymérases (polIII et polV) sont connues pour suivre la « règle du A », c'est à dire l'incorporation préférentielle d'un nucléotide A au site abasique (Ide *et al.* 1995). L'intervention de ce système pourrait être à l'origine du biais en A-T au niveau du terminus de réplication. Cet enrichissement en A-T au niveau du terminus de réplication conduira à une chute du taux en G-C et donc du G+C3 au niveau du terminus de réplication.

II- Plasticité Génomique

A- Réarrangements chromosomiques

Différents mécanismes peuvent intervenir dans les réarrangements chromosomiques au sein d'un génome, comme par exemple les mécanismes de réparation des lésions causées à l'ADN. Ces phénomènes de recombinaison vont engendrer différents types de réarrangements comme des inversions, des translocations, des amplifications et des délétions, provoquant, par exemple, une modification de la localisation des gènes ou leur expression.

A-1 LOCALISATION DES REARRANGEMENTS

Plusieurs génomes présentent des zones de plasticité où délétions, inversions et insertions de matériel génétique sont beaucoup plus tolérées que dans le reste du chromosome (Alm *et al.* 1999; Glaser *et al.* 2002; Sibley et Raleigh 2004). Chez *Helicobacter pylori*, une inversion entre les génomes des souches J99 et 26695, nommée 3, est associée à une région dont le pourcentage en bases G-C est plus faible que le reste du génome (35% pour 39%) et nommée « zone de plasticité ». Cette région contient respectivement 46% et 48% des gènes qui sont uniques chez les souches 26695 et J99. La présence d'homologues aux gènes *vir*, d'IS et d'un pourcentage plus faible en bases G-C indique que cette région représente un îlot de pathogénicité (Alm *et al.* 1999). La comparaison de ces deux génomes indique la présence d'une dizaine de segments (de 1 à 83 kb) variables.

La comparaison de deux régions (région de l'antigène O et région RM) de 11 souches d'*E. coli* a montré le caractère hypervariable de ces deux régions (Milkman *et al.* 2003). Chez *Pirulella*, de grandes inversions ont été caractérisées autour du terminus de réplication (Glockner *et al.* 2003).

L'instabilité chromosomique chez les *Streptomyces* est corrélée à de grands réarrangements touchant les extrémités chromosomiques. Ces réarrangements conduisent à de grandes délétions, des remplacements de séquences ou encore la circularisation du chromosome (Fischer *et al.* 1997a; Chen *et al.* 2002; Wenner *et al.* 2003). La comparaison des génomes entièrement séquencés de *S. coelicolor* et de *S. avermitilis* a montré la présence de grands réarrangements au niveau des régions terminales spécifiques d'espèces ainsi qu'une inversion autour de l'origine de réplication (Ikeda *et al.* 2003).

A-2 IMPACT DU NOMBRE DE REPETITIONS SUR LA STABILITE GENOMIQUE

La présence d'un grand nombre de répétitions conduirait à une augmentation de la fréquence des réarrangements (Frank *et al.* 2002; Parkhill *et al.* 2003). Ainsi, chez les γ-protéobactéries, les taux de réarrangements les plus importants sont retrouvés dans les génomes présentant le plus grand nombre de répétitions (Rocha 2003).

Par exemple, les bactéries du genre *Neisseria* constituent le cas extrême dans l'abondance de répétitions, qui engendrent des phénomènes de variations antigéniques (Saunders *et al.* 2000). Ainsi, 66 860 paires de répétitions inversées sont présentes chez *Neisseria meningitidis*. En

revanche, aucune répétition n'est retrouvée chez *Chlamydia trachomatis*, parasite intracellulaire obligatoire, protégé du système immunitaire de l'hôte (Achaz *et al.* 2003).

Le nombre de répétitions retrouvées dans un génome va influencer la stabilité de ce dernier. Une étude a porté sur le génome de *Sinorhizobium meliloti* pour identifier de grandes répétitions, notamment les IS, dans les trois réplicons (un chromosome et deux mégaplasmides) (Guo *et al.* 2003). Sur les 133 répétitions identifiées, 91 ont été caractérisées comme points chauds de recombinaison. Les recombinaisons entre ces répétitions conduisent à de nombreux réarrangements provoquant une nouvelle architecture génomique, comme par exemple des souches ne possédant qu'un seul réplicon de 6,68 Mb ou des souches à deux réplicons.

Des bactéries phylogénétiquement proches avec un nombre très différent de répétitions peuvent présenter une perte de la synténie. Par exemple, les souches KIM et CO92 de *Y. pestis* (99,9% de similarité en protéines) montrent un grand nombre de réarrangements dans leur génome (Deng *et al.* 2002). *Salmonella enterica typhi* et *Salmonella enterica typhimurium* (98,6% de similarité en protéines) ne présentent que deux grands réarrangements. Chez *Y. pestis*, environ 5000 répétitions sont retrouvées pour moins de 1000 à *Sl. enterica typhimurium*, pour des génomes de taille similaire.

A-3 INFLUENCE DES FACTEURS ENVIRONNEMENTAUX

L'environnement va agir sur le taux de réarrangements en déréprimant certains mécanismes mutateurs, comme le système SOS, provoquant des variations de certains loci. Ainsi, les chromosomes de bactéries pathogènes intracellulaires obligatoires, comme *Chlamydia*, *Spirochetes* ou *Rickettsia*, présentent très peu d'éléments répétés. Ceci peut être mis en relation avec leur environnement intracellulaire relativement protecteur.

Suivant la phase du cycle cellulaire, l'efficacité des mécanismes de recombinaison sera variable. Par exemple, en phase stationnaire, les concentrations de MutS et de MutH, appartenant au système MMR (Methyl directed Mismatch Repair), sont plus faibles respectivement d'un facteur 10 et 3, la quantité de MutL restant constante (Feng *et al.* 1996; Tsui *et al.* 1997; Kivisaar 2003; Saint-Ruf et Matic 2006). Cette inhibition est dépendante de l'expression du facteur σ^S. En effet, chez un mutant *rpoS*-, le taux de MutS et MutH est similaire en phase exponentielle et en phase stationnaire, démontrant que la baisse des concentrations de ces deux protéines lors de l'entrée en phase stationnaire est *rpoS*-dépendante (Tsui *et al.* 1997). Le taux de σ^S augmente également lorsque la bactérie est confrontée à un stress nutritif, osmotique, acide ou thermique. Chez *S. pneumoniae*, le système Hex est homologue au système MMR. Comme le système MMR, ce système intervient comme barrière génétique en interférant avec la recombinaison. Son efficacité est plus forte que le système de *Ba. subtilis*, qui a un rôle marginal dans la prévention de la recombinaison homéologue. Ce rôle minime peut être expliqué par le fait que *Ba. subtilis* n'est compétent que lors d'une limitation en nutriments, durant laquelle les protéines du système MMR sont dégradées. En revanche, chez *S. pneumoniae*, la compétence ne nécessite pas de carence nutritive, les protéines HexA et HexB ne sont pas dégradées et restent aptes à prévenir la recombinaison homéologue (Majewski *et al.* 2000).

En 1988, Cairns et son équipe observent que des mutations ont lieu dans des cellules en conditions de non-croissance et proposent qu'elles soient produites en réponse à une induction de

l'environnement (Cairns *et al.* 1988). La capacité de mutants, incapables de métaboliser le lactose (frameshift +1 dans le gène *lacI*) à réverter a conforté cette hypothèse (Cairns et Foster 1991). Lors de cultures avec le lactose comme unique source de carbone, donc en phase de non-croissance, l'apparition de mutants de type réversion, restaurant le phénotype Lac+, est fréquente alors que ces mutations sont rares en phase exponentielle de croissance (Cairns et Foster 1991). Ces mutations « dirigées » vers le gène dont la réversion conférait un avantage adaptatif ont été appelées « mutations adaptatives ».

Les mécanismes de recombinaison interviennent également dans les processus de variation de phase ON↔OFF. Chez les bactéries pathogènes, la variation de phase affecte très souvent des gènes impliqués dans la synthèse de structures de surface comme les pili, les fibrilles, les flagelles, les protéines de la membrane externe ou encore les LPS (Dybvig 1993). Par exemple, la synthèse de pili de type IV chez *Neisseria gonorrhoeae* dépend de ce mécanisme de régulation (Mehr et Seifert 1998). Le gène mutable est le gène *pilE* (E pour « expressed ») qui code la sous-unité structurale du pilus. Suite à des événements de conversion génique, la région 5' codante peut être délétée. Les allèles résultants, nommés *pilS* (S pour « silent ») codent des protéines solubles excrétées hors de la cellule (Haas et Meyer 1987). Le passage d'un état ON (synthèse du pilus) à un état OFF (non-production) ou inversement se produit à haute fréquence. D'autres exemples sont rencontrés comme la régulation de la synthèse des pili de type I chez *E. coli* et *Sl. typhimurium* (Abraham *et al.* 1986).

A-4 FORCES S'OPPOSANT AUX REARRANGEMENTS

Les principaux facteurs qui s'opposent aux réarrangements sont l'effet dose, la symétrie chromosomique *oriC*/*ter* ou encore le biais d'orientation des gènes sur les brins d'ADN (Hill et Gray 1988; Campo *et al.* 2004).

Chez *Lactococcus lactis*, des inversions non symétriques autour de l'origine ou le terminus de réplication, conduisant à un rapprochement du terminus vers l'origine de réplication, sont associées à une diminution de la croissance, le chromosome ayant perdu sa symétrie origine de réplication à un pôle et terminus à l'autre (symétrie *oriC*/*ter*) (Campo *et al.* 2004). Cette diminution de la croissance peut être aussi la conséquence de la perte de l'effet dose de gènes.

Chez *L. lactis*, environ 84% des gènes et environ 90% des gènes essentiels sont situés sur le brin direct (Rocha et Danchin 2003b). Toutes les inversions, ayant provoqué le passage de gènes du brin direct au brin indirect, amènent à la létalité.

Les inversions les plus fréquemment observées sont donc symétriques autour de l'origine et du terminus de réplication chez les bactéries (Tillier et Collins 2000) et chez les archées (Zivanovic *et al.* 2002). En effet, ces réarrangements minimisent la désorganisation du chromosome au niveau de la réplication et de la ségrégation (Hill et Gray 1988; Mackiewicz *et al.* 2001). De plus, les inversions symétriques autour de l'origine ou du terminus de réplication permettent de conserver la distance entre les gènes et *oriC*/*ter* ou encore l'organisation des gènes sur les brins directs ou indirects.

A-5 REPARATION DES CASSURES DOUBLE BRIN ET REARRANGEMENTS CHROMOSOMIQUES

A-5a Agents responsables de cassures double brin

Divers agents exogènes et endogènes à la cellule peuvent engendrer des cassures double brin au niveau de l'ADN. Les cassures double brin sont considérées comme une forme sévère de dommage à l'ADN. En effet, elles peuvent conduire à l'arrêt du cycle cellulaire et à la mort cellulaire et si elles sont incorrectement réparées, elles peuvent mener à une instabilité génomique.

Le génome est soumis à de nombreux agents mutagènes physiques ou chimiques. Un exemple d'agents physiques qui provoque des cassures double brin est l'émission de radiations ionisantes comme les rayons X. Des éléments intégrés au génome peuvent également provoquer des cassures de l'ADN. Ainsi lors de l'excision d'IS ou d'éléments génétiques mobiles, le site d'intégration peut rester ouvert et donc provoquer ces cassures.

D'autres événements peuvent provoquer des cassures double brin. Ainsi la présence de courtes séquences répétées sur l'ADN matrice au moment de la réplication peut conduire à des glissements de la fourche de réplication (Viguera *et al.* 2001). Ces glissements peuvent conduire à des pauses et des arrêts de la machinerie de réplication produisant des cassures double brin.

Chez les eucaryotes, les extrémités chromosomiques sont un complexe nucléo-protéique contenant de courtes séquences répétées (Blackburn 2001). Ces télomères interviennent dans de nombreuses fonctions, dont la protection des extrémités et la prévention dans la fusion chromosomique. Dans les cellules somatiques, ces télomères vont raccourcir marquant le début de la sénescence cellulaire (Herbig *et al.* 2004; Itahana *et al.* 2004). La perte des télomères conduit donc à la formation de cassures au niveau des extrémités chromosomiques.

Autres mécanismes pouvant conduire à la formation de cassures au niveau de l'ADN sont les systèmes de partition et de ségrégation chromosomique et de septation.

A-5A1 PARTITION CHROMOSOMIQUE

Chez les bactéries, contrairement aux plasmides à grand nombre de copies dont la distribution dans les cellules filles est passive, les copies nouvellement répliquées des plasmides à faible nombre de copies sont, à l'instar du/des chromosome(s), réparties grâce à un mécanisme actif, appelé partition.

A-5A1.1 Centromère et ségrégation chromosomique chez les eucaryotes

Suite à la réplication du matériel génétique, la division cellulaire s'achève, accompagnée de la ségrégation d'un jeu chromosomique dans chacune des futures cellules filles (Fukagawa 2004). Le centromère joue un rôle clé dans ce mécanisme de ségrégation chromosomique. Il est le siège de formation du kinétochore. Le kinétochore est un complexe nucléo-protéique permettant l'ancrage des chromatides au fuseau achromatique. Ainsi, *via* cet ancrage, les chromosomes peuvent être tractés aux pôles du fuseau ce qui assure leur ségrégation équitable dans les cellules filles.

A-5A1.2 Systèmes de partition *parABS* chez les bactéries

Chez les bactéries, la partition chromosomique et plasmidique fait intervenir un système à trois composantes nommé *parABS*. La bactérie *E. coli* ne possède pas ce système mais un système qui lui est spécifique. Une séquence, nommée *migS* (pour « **mig**ration **s**ite »), peut être considérée comme le centromère du génome d'*E. coli* (Yamaichi et Niki 2004). En revanche, les acteurs intervenant avec la séquence *migS* restent encore à définir.

En revanche, chez les autres bactéries, le système actif de distribution des réplicons néosynthétisés dans les cellules filles est nommé système *par*. Tous les systèmes *par* connus sont constitués de trois éléments : un site agissant en *cis* et deux protéines agissant en *trans*. La protéine ParB se fixe au niveau de la séquence *parS*, qui est considérée comme l'analogue fonctionnel du centromère eucaryote. Une fois la protéine ParB fixée, la protéine ParA vient former un complexe et à partir de son activité ATPasique, permet la distribution des réplicons dans les différentes cellules filles.

Chez *Ba. subtilis*, bactérie modèle de l'étude des systèmes de partition, 10 sites *parS* sont présents et la séquence consensus est composée de 16 nucléotides (5'-TGTTNCACGTGAAACA-3') (Lin et Grossman 1998). Tous ces sites sont situés dans la région de l'*oriC*. Cette séquence consensus a été utilisée pour déterminer la présence de sites *parS* chez *S. coelicolor*. Ainsi, 24 sites potentiels de 16 pb ont été identifiés. La majorité de ces sites (21/24) est située dans une région de 400 kb entourant l'origine de réplication (Kim *et al.* 2000). A partir de ces séquences, une séquence consensus *parS* de *S. coelicolor* a été définie 5'-tGTTTCACGTGAAACa-3'.

A-5A2 DEFAUT DE SEPTATION

Sur milieu solide, les spores de *Streptomyces* germent et se développent sous forme d'un mycélium végétatif, à partir duquel des hyphes aériens se forment pour donner des chaînes de spores ne contenant qu'une seule copie du génome. En revanche les hyphes végétatifs contiennent plusieurs copies du génome (Flardh 2003).

Au moment de l'étape de la sporulation, une copie du chromosome est distribuée dans chaque spore (Schwedock *et al.* 1997). La protéine FtsK intervient chez *E. coli* au niveau du septum formé par les protéines FtsZ (encore appelé « anneau Z ») pour permettre la cytokinèse et la résolution des dimères de chromosomes (Begg *et al.* 1995; Yu *et al.* 1998). Cette protéine présente également une fonction d'ADN translocase pour distribuer les copies génomiques dans chaque cellule fille (Pease *et al.* 2005). Un homologue de la protéine FtsK est présent chez *S. coelicolor* (Bentley *et al.* 2002). La mutation de l'homologue de *ftsK* chez *S. coelicolor* confère une hétérogénéité de la morphologie des colonies bactériennes de la descendance (Wang *et al.* 2007). Ces descendants sont caractérisés par des délétions des extrémités pouvant atteindre jusqu'à 10% du génome. La protéine FtsK interagirait avec les extrémités du chromosome pour distribuer ce dernier dans chaque spore. Ainsi, chez le mutant, au moment de la formation du septum des spores, les extrémités chromosomiques seraient toujours au niveau du plan de division. La fermeture du septum sur le chromosome provoquerait alors une cassure de ce dernier, conduisant à des délétions de ses extrémités (phénoméne de guillotine).

A-5B Mécanismes de réparation chez les bactéries

Les systèmes de réparation par recombinaison homologue sont dépendants de la protéine RecA. Suivant le type de dommage, différentes voies vont intervenir (pour revue (Kowalczykowski 2000)):

- Le système RecBCD est le principal système de réparation des cassures double brin.
- Le système RecFOR intervient pour la réparation des cassures simple brin.

Un autre système de réparation par recombinaison illégitime, système NHEJ (pour « Non Homologous End Joining ») peut également intervenir pour la réparation des cassures double brin.

Cependant, ces mécanismes ne sont pas distribués chez toutes les bactéries. Une étude de la distribution des gènes impliqués dans la réparation de l'ADN dans 117 bactéries dont le génome est entièrement séquencé a été réalisée en 2005 (Rocha *et al.* 2005) permettant de déterminer la présence et la distribution de ces différents mécanismes chez les bactéries.

A-5B1.1 MECANISMES DE RECOMBINAISON HOMOLOGUE

Deux mécanismes principaux de recombinaison homologue interviennent dans les réparations des cassures double brin. Pour le système RecBCD, toutes les fonctions, hélicase, nucléase et chargement de RecA, sont assemblées dans une holoenzyme (Singleton *et al.* 2004). Cette holoenzyme se fixe aux extrémités double brin, déroule et dégrade l'ADN jusqu'à rencontrer un site χ. Au niveau de ce site, l'activité de RecBCD est modifiée et un ADN simple brin est produit permettant le chargement de RecA (Kowalczykowski 2000). Puis le système RuvABC intervient pour permettre la résolution de la croix de Holliday. Le système RecFOR intervient dans la réparation des cassures simple brin. La protéine RecF interviendrait dans la fixation de RecA et dans l'hydrolyse d'ATP (Lovett et Kolodner 1989). La protéine RecO présente des activités dans l'hybridation de l'ADN complémentaire et le chargement de RecA sur l'ADN simple brin. Enfin RecR s'associe à RecF pour limiter l'extension des filaments RecA au-delà de la cassure simple brin.

Les gènes de la recombinaison homologue ne sont pas tous présents dans les génomes bactériens. Les gènes *recA* et *ruvAB* sont absents uniquement des génomes des bactéries intracellulaires *Buchnera* et *Blochmania*. Certains acteurs, comme RuvC, présentent des analogues fonctionnels, comme RecU chez les firmicutes et les mollicutes. De même, lorsque le complexe RecBCD (exonucléase V impliquée dans la voie majeure de recombinaison homologue) est absent, un complexe analogue codé par les gènes *addAB* est trouvé.

Ce système RecFOR est présent dans de nombreux génomes contenant le système RecBCD. Ceci reflète la spécificité de ces deux systèmes dans les différents types de lésions, le système RecBCD pour la réparation des cassures double brin et le système RecFOR dans celui des cassures simple brin. Ainsi, chez des mutants de *Mycobacterium smegmatis* RecBCD-, le système RecFOR compense l'absence de ce système pour les étapes présynaptiques (Chedin et Kowalczykowski 2002).

A-5B1.2 MECANISME NHEJ

Quelques bactéries possèdent les gènes intervenant dans le mécanisme NHEJ bactérien. Les gènes de NHEJ, codant les principaux acteurs de la recombinaison illégitime NHEJ, Ku et ligase IV, ont été identifiés dans 24 génomes sur les 117 étudiés. Ce mécanisme nécessite uniquement des extrémités complémentaires et pas de séquences d'homologies. La protéine Ku se

fixe au niveau des extrémités de la cassure double brin et va jouer un rôle de pontage entre les deux extrémités. Puis la ligase IV intervient pour associer les extrémités des deux morceaux d'ADN et permettre ainsi la réparation (Weller *et al.* 2002). Chez *M. tuberculosis* et *M. smegmatis*, le système NHEJ intervient dans la recircularisation des plasmides (Gong *et al.* 2004; Gong *et al.* 2005). Ce système intervient de façon spécifique durant la phase stationnaire de croissance lors de dommages causés par des rayonnements infrarouges chez *M. smegmatis*, indiquant que ce système est important durant les phases non-réplicatives du cycle cellulaire chez cette bactérie (Stephanou *et al.* 2007).

La distribution de ces gènes est sporadique car ils sont aussi bien présents chez les firmicutes, les actinomycètes et plusieurs groupes de protéobactéries. Dans certains génomes, plusieurs copies des gènes de NHEJ sont retrouvées comme par exemple, chez *A. tumefaciens*, qui possède six copies du gène de la protéine Ku et huit copies du gène de la ligase. Chez les bactéries possédant un chromosome linéaire, seuls les *Streptomyces* et *A. tumefaciens* possèdent les gènes du système NHEJ. Mais de nombreux génomes bactériens ne possèdent pas ce système NHEJ, indiquant que la recombinaison homologue apparaît ainsi être le système principal de réparation des lésions chez les bactéries.

A-5B2 Mécanismes de réparation des cassures double brin chez la levure

Chez les eucaryotes, les cassures double brin sont réparées par des systèmes analogues à ceux décrits précédemment chez les bactéries. Comme les *Streptomyces*, les chromosomes de la levure et de l'homme sont linéaires. Une analogie entre les événements de recombinaison observés dans les régions subtélomériques chez les eucaryotes et les *Streptomyces* peut être abordée.

Les systèmes de réparation le long du chromosome XI de cellules haploïdes de *Sa. cerevisiae* ont été examinés (Ricchetti *et al.* 2003). Plus particulièrement, cette étude a porté sur les régions subtélomériques, qui ont la particularité d'être faibles en densité de gènes et de porter seulement des gènes non-essentiels (Louis 1995; Fairhead *et al.* 1998).

Chez *Sa. cerevisiae*, la réparation de 90% des cassures double brin est réalisée par recombinaison homologue et 10% par recombinaison illégitime (Resnick et Martin 1976; Kramer *et al.* 1994; Fairhead *et al.* 1996). Cette étude a montré que le système NHEJ est le système exclusif de réparation des cassures double brins sur la quasi-totalité du chromosome, dans des conditions où la recombinaison homologue ne peut pas intervenir, mais que sa fréquence d'intervention diminue dans les régions subtélomériques (Ricchetti *et al.* 2003). Ces réparations s'accompagnent parfois de délétions de petites tailles et d'insertions de fragments d'ADN mitochondrial (de 47 à 97 pb) (Ricchetti *et al.* 1999). Mais trois autres événements sont observés plus particulièrement dans les régions subtélomériques.

Le premier va conduire à la combinaison entre le plasmide utilisé dans cette étude et le chromosome grâce à la présence de microhomologies de séquences. Ainsi le matériel génétique plasmidique est incorporé au chromosome, accompagné de la perte de la région subtélomérique ayant subi la cassure double brin.

Le second mécanisme est appelé BIR (pour « Break Induced Replication »). Le chromosome cassé va initier un événement de recombinaison homologue (jusqu'à 20 nt) et copier la séquence « donneuse » jusqu'à l'extrémité (Fig. 5) (Malkova *et al.* 1996). La région, qui sera utilisée comme matrice, peut atteindre plusieurs centaines de kb. Ainsi, la région « matrice » du chromosome donneur sera dupliquée et remplacera l'extrémité perdue.

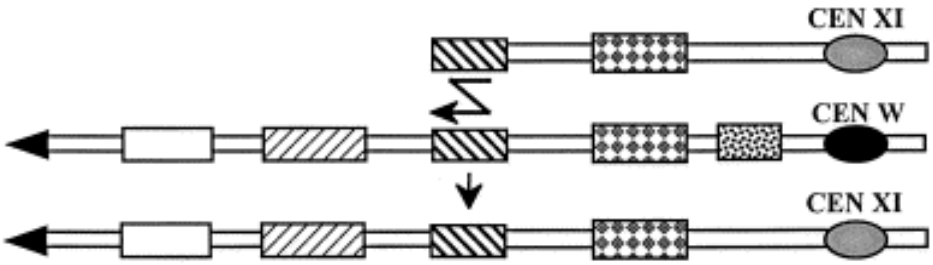


Figure 5 : Réparation des cassures double brin par le mécanisme de BIR (d'après Richetti *et al.*, 2003)
Après la cassure double brin, le chromosome XI, ayant perdu son télomère, va envahir un autre chromosome (W) grâce à la présence de microhomologies. La région terminale du chromosome W va servir de matrice pour la réplication et permettre au chromosome XI d'acquérir une nouvelle région terminale.

Enfin, le troisième mécanisme est la cicatrisation (« healing ») de la cassure double brin par l'addition d'un télomère. Ce mécanisme, rare, s'accompagne de la perte des séquences séparant le point de cassure du télomère. Ces mécanismes sont Rad52-dépendants car dans un contexte *rad52-* , le mécanisme NHEJ devient le mécanisme de réparation majoritaire dans les régions subtélomériques alors qu'aucun événement de réparation par le mécanisme de BIR par exemple n'est détecté. En revanche, en dehors du mécanisme de NHEJ, le mécanisme d'addition de télomères devient majoritaire (67% des cas).

A-5B3 Mécanismes de réparation chez l'Homme

Dans le génome des primates, les régions subtélomériques de chromosomes non homologues partagent des blocs de séquences homologues (Mefford et Trask 2002). En outre, de grands segments dupliqués sont courants dans ces régions (Bailey *et al.* 2001). L'étude dynamique des régions subtélomériques montre une grande plasticité à l'intérieur et entre espèces (Mefford et Trask 2002). En effet, les régions subtélomériques de diverses espèces telles que *Sa. cerevisiae*, *Plasmodium falciparum*, *Drosophila melanogaster* et l'Homme sont structuralement similaires, composées de divers éléments répétés mais leur taille et leur séquence varient suivant l'organisme. Ces régions contiennent diverses fonctions comme par exemple, chez l'homme, les gènes des récepteurs olfactifs (Trask *et al.* 1998).

L'observation de grands segments chromosomiques dupliqués suggère que les structures subtélomériques résultent d'événements de translocations plutôt que d'événements de transpositions (Linardopoulou *et al.* 2005). En effet, les régions subtélomériques ne présentent aucun transposons ou sites d'insertions connus (Smit et Riggs 1996). Ces translocations seraient le résultat de réparations aberrantes par le mécanisme de NHEJ ou par la recombinaison homologue, mécanismes principaux de réparation de cassures double brin de l'ADN chez les mammifères (Pfeiffer *et al.* 2000; Rothkamm *et al.* 2003). Pour déduire la contribution relative du mécanisme NHEJ ou de la recombinaison homologue dans la juxtaposition des blocs subtélomériques, tous les points de jonctions ont été analysés. La présence d'éléments répétés aux frontières suggère fortement un mécanisme de réparation homologue. En revanche, l'absence d'éléments répétés tronqués au niveau des frontières est indicative du mécanisme de NHEJ. Ainsi, dans une étude de la structure subtélomérique, les auteurs ont identifié un jeu de 56 alignements de jonctions de

séquences, issues d'un seul événement de translocation. Sur ces 56 jonctions, 53 seraient issus de mécanismes de réparation dont 49 (soit 92%) par le mécanisme de NHEJ. Les quatre dernières jonctions sont le fruit de la recombinaison homologue entre loci non alléliques (NAHR pour « Non Allelic Homologous Recombination »).

A-5C Cycle CFP

Le cycle CFP (pour Cassure-Fusion-Pont) peut être initié par la perte d'un télomère (suite à un dysfonctionnement télomérique lié à la sénescence par exemple) ou par cassure double brin (McClintock 1941). Il débute par la fusion de chromatides sœurs issues d'un chromosome cassé. Cet événement donne naissance à un chromosome dicentrique, portant deux centromères actifs. Le chromosome dicentrique casse sous l'effet de la traction physique lorsque les centromères sont entraînés vers les pôles opposés durant l'anaphase de mitose. Ce cycle peut se perpétuer au cours des cycles cellulaires suivants. Ces changements chromosomiques correspondent à des translocations, des amplifications de gènes, la perte ou le gain de grandes portions de chromosomes. Ces changements peuvent également conduire à la fusion promoteur-phase codante provoquant une dérépression de l'expression génique (Mathieu *et al.* 2004). Un exemple célèbre, chez l'homme, de la conséquence d'une translocation à la suite d'un cycle CFP, est le chromosome de Philadelphie dans la leucémie myélineuse chronique (CML) qui est le résultat de la translocation t(9;22) (translocation du chromosome 9 vers le chromosome 22) (Nowell et Croce 1986).

La figure 6 présente les différents réarrangements générés par le cycle CFP.

Les facteurs de la régulation du cycle cellulaire auraient un impact sur l'instabilité génomique. De nombreuses observations montrent que la mutation de la protéine p53, intervenant dans la régulation du cycle cellulaire, conduit à une augmentation importante du nombre de carcinomes, résultant de réarrangements chromosomiques dus à un cycle CFP (Artandi *et al.* 2000; Rudolph *et al.* 2001).

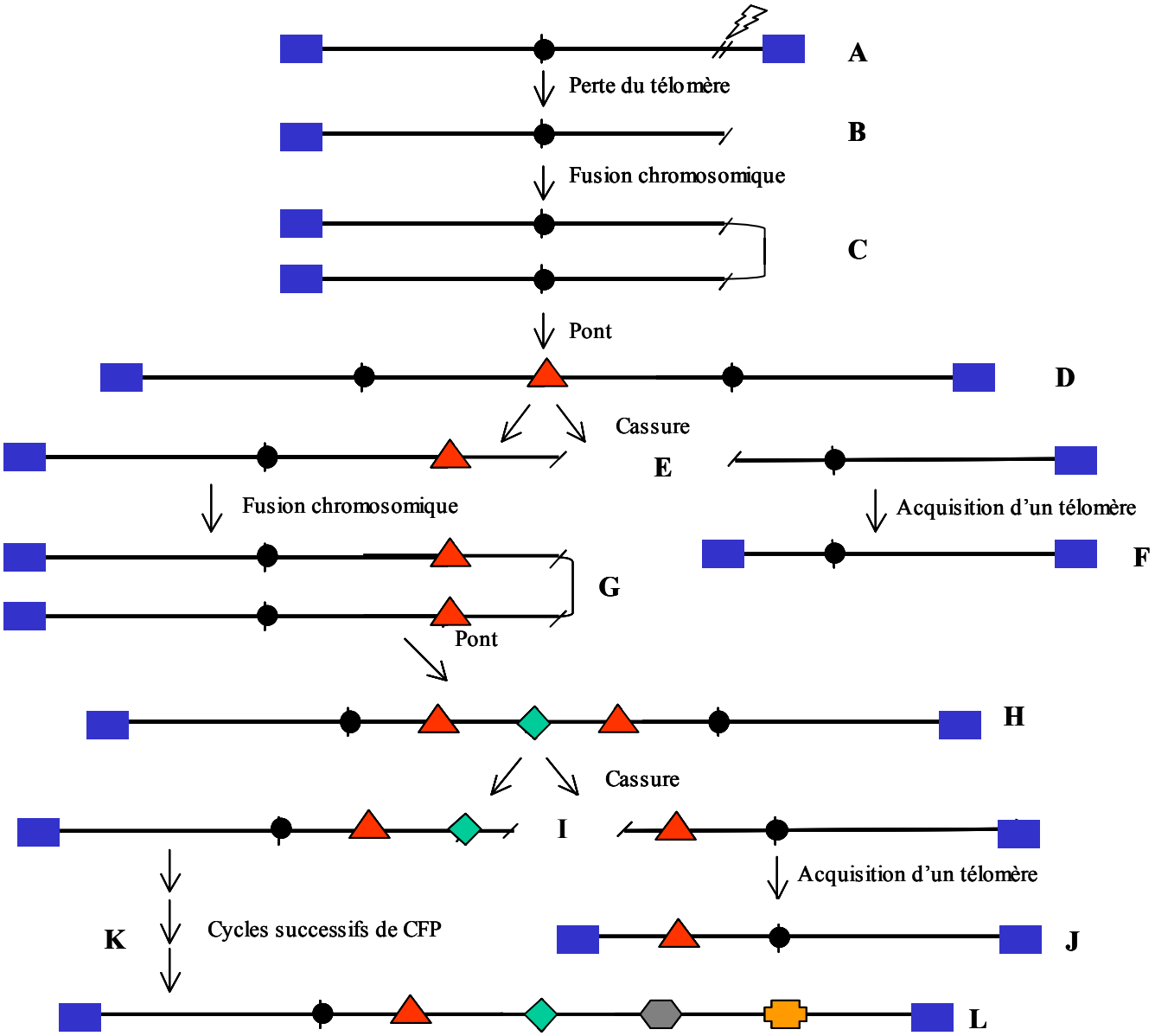


Figure 6: Cycle de CFP et réarrangements chromosomiques
A. Le chromosome perd un télomère (dysfonctionnement de la télomérase lié à la sénescence cellulaire ou par cassure double brin) et B. présente la structure chromosomique après la perte. C. Deux chromatides vont fusionner au niveau des extrémités cassées hautement recombinogènes et D. présenter la structure chromosomique suivante. E. Sous l'effet de la traction physique des centromères, le chromosome dicentrique va casser dans un locus situé entre les deux centromères. F. L'acquisition d'un nouveau télomère par l'un des deux fragments conduit à la sortie du cycle CFP et en résulte une grande délétion des régions télomériques. G. L'autre fragment va poursuivre son cycle CFP par la fusion avec une autre chromatide et H. présente une nouvelle structure chromosomique. I. De nouveau, une cassure va se produire entre les deux centromères. J. Un des fragments pourra sortir du cycle CFP en acquérant un télomère. Des réarrangements de la région terminale sont le résultat de deux cycles CFP. K. L'autre fragment chromosomique peut subir de nombreux cycles CFP successifs et L. l'acquisition d'un télomère sort le chromosome du cycle CFP. Ces cycles successifs ont conduit à la perte des régions terminales initiales, et à l'acquisition/transposition d'informations génétiques contenues dans les régions terminales d'autres chromosomes.

Les carrés bleus correspondent aux télomères, les ronds noirs aux centromères, les autres formes géométriques (triangle rouge, losange vert etc.) représentent les points de fusion entre deux chromatides.

Chez les *Streptomyces*, les réarrangements sont localisés au niveau des extrémités comme pour les chromosomes eucaryotes (Fischer *et al.* 1997b). Par analogie avec ce qui est observé chez les eucaryotes, les mécanismes de réparation de cassures de l'ADN ou encore un cycle CFP pourraient être les mécanismes à l'origine de ces réarrangements. Un mutant de *Streptomyces ambofaciens*, présentant un chromosome résultant de la fusion de deux chromosomes partiellement délétés sur le même bras, montre des analogies avec les chromosomes dicentriques eucaryotes (Wenner *et al.* 2003).

.

B- Transfert horizontal chez les bactéries

Les mutations ponctuelles ou géniques n'expliquent qu'une partie de la diversité génétique observée aux niveaux intra- et interspécifiques chez les bactéries. De plus, ce mécanisme d'évolution permet l'émergence de nouvelles fonctions par accumulation de divergences et correspond à un mécanisme lent. Le transfert horizontal, quant à lui, permet l'acquisition de nouvelles fonctions en un nombre d'événements réduits. Les trois processus, transformation naturelle, transduction et conjugaison, aboutissent à un échange génétique unidirectionnel d'ADN de la cellule donneuse vers une cellule réceptrice (Milkman et Bridges 1993).

Suivant le mécanisme considéré, la taille de l'ADN transféré peut varier de quelques kilobases pour la transformation naturelle, à plusieurs dizaines de kilobases pour le mécanisme de transduction phagique et potentiellement à des centaines de kilobases par conjugaison. Pour permettre l'intégration de ces fragments dans le chromosome de la cellule donneuse, les mécanismes de recombinaison vont intervenir. Les mécanismes de recombinaison homologue sont efficaces pour promouvoir la recombinaison entre isolats proches (de la même espèce ou d'espèces proches). La recombinaison illégitime permet de faciliter l'intégration de fragments entre organismes phylogénétiquement éloignés (Spratt *et al.* 2001).

L'expression de gènes participant à la résistance de drogues ou d'antibiotiques, au caractère d'invasivité/pathogénicité/virulence chez les pathogènes ou encore à la capacité de métabolisation de nouveaux substrats est comprise comme un apport substantiel au fitness. Suivant l'impact de ces séquences sur le « fitness » de la bactérie réceptrice, elles seront conservées ou pas à long terme. Le fitness ou valeur adaptative d'un génome correspond à la capacité de transmettre son génotype à la descendance et à être maintenu au sein de la population.

B-1 MECANISMES DU TRANSFERT HORIZONTAL

Pour permettre le transfert d'information génétique, la séquence acquise doit être incorporée dans un réplicon ou se maintenir de façon autonome (plasmide, chromosome..) dans la cellule receveuse. Selon les microorganismes, la part du génome résultant du transfert horizontal sera différente. Par exemple, le génome d'*E. coli* présente 10 à 16% de ses gènes acquis horizontalement (Medigue *et al.* 1991; Lawrence et Ochman 1997).

La niche écologique intervient dans l'impact des événements de transfert horizontal. Les bactéries intracellulaires, comme *Ricketssia prowazekii* ou *My. genitalium,* sont dans un environnement qui va limiter le transfert de l'ADN exogène. Des exemples de transfert entre

bactérie intracellulaire et le noyau de la cellule hôte, comme les mitochondries et les chloroplastes ont été décrits (Fox *et al.* 1980; Bergthorsson *et al.* 2003). Au contraire des bactéries dont l'habitat naturel est le sol par exemple, le transfert d'ADN exogène sera plus fréquent (Ochman *et al.* 2000).

B-1a Transformation naturelle

La transformation requiert la présence d'ADN nu dans l'environnement et permet théoriquement le transfert d'information génétique entre organismes phylogénétiquement très éloignés. Pour que l'ADN puisse pénétrer dans la cellule hôte, celle-ci doit développer un état appelé « état de compétence ». Certaines bactéries, comme *Ne. gonorrhoeae* et *Haemophilus influenzae* sont en permanence en état de compétence, alors que d'autres bactéries comme *Ba. subtilis* ou *Streptococcus pneumoniae* deviennent compétentes à certains stades physiologiques (Dubnau 1999; Thomas et Nielsen 2005).

De plus, certaines espèces bactériennes, comme par exemple *Ne. gonorrhoeae* et *Ha. influenzae*, sélectionnent l'ADN qui va traverser la membrane alors que d'autres espèces intègrent l'ADN quelle que soit sa séquence (Lorenz et Wackernagel 1994). Ainsi le transfert d'ADN venant d'espèces phylogénétiquement éloignées sera fortement restreint chez *Ne. gonorrhoeae* et *Ha. influenzae*.

B-1b Transduction

Le matériel génétique peut aussi être introduit dans une cellule donneuse par le biais d'un bactériophage. Lors de la formation de nouveaux virions, du matériel génétique de la cellule donneuse peut être accidentellement encapsidé. Il s'agit, soit de fragments d'ADN pris au hasard dans le génome (transduction généralisée), soit de séquences d'ADN adjacentes au site d'attachement du phage (transduction spécialisée) au moment de l'encapsidation du génome phagique.

La quantité d'ADN pouvant être encapsidée correspond à la taille du génome phagique. Après la lyse de la cellule donneuse, les particules phagiques contenant l'ADN de la cellule donneuse peuvent infecter une cellule receveuse. L'ADN de la cellule donneuse peut alors être incorporé par recombinaison dans le génome de la cellule receveuse. Comme la transformation, la transduction ne nécessite pas la présence simultanée des bactéries donneuses et receveuses.

Pour la transduction, la présence des protéines d'ancrage de ces phages ou la présence dans la même niche écologique que le phage sont des prérequis pour permettre la transduction. De même, les îlots de pathogénicité contiennent souvent des intégrases de bactériophages, suggérant leur acquisition par transduction comme chez *Staphylococcus aureus* (Lindsay *et al.* 1998). Chez *V. cholerae*, les gènes de toxine sont situés à l'intérieur du génome du phage CTXΦ (Waldor et Mekalanos 1996).

B-2 LA CONJUGAISON BACTERIENNE

Chez les *Streptomyces*, le principal mécanisme d'acquisition de matériel génétique exogène serait la conjugaison (Chater et Hopwood 1984). Ce mécanisme nécessite un contact physique entre les bactéries donneuse et réceptrice. Ce mécanisme est caractérisé par un spectre d'hôtes très large. En effet, il permet le transfert de matériel génétique entre bactéries et eucaryotes (par exemple, entre bactéries et plantes ou bactéries et champignons) (Buchanan-Wollaston *et al.* 1987;

Heinemann et Sprague 1989). La mobilisation de marqueurs chromosomiques peut se faire en *cis* (souche *E. coli* Hfr) ou en *trans* (ex.: pSAM2)

B-2a "Integrated Conjugative Element" (ICE)

Historiquement, les premiers éléments chromosomiques conjugatifs intégratifs ont été définis comme des transposons conjugatifs. Le premier transposon conjugatif décrit fut *Tn916* d'*Enterococcus faecalis* (Clewell et al. 1995). L'analyse de nombreux transposons conjugatifs et des éléments conjugatifs et intégratifs site spécifique a montré qu'une majorité présentait souvent une grande spécificité d'intégration chez un hôte et une plus faible chez un autre hôte. Par exemple, *Tn916* peut s'intégrer à différents sites chez *En. faecalis* alors qu'il ne s'intègre qu'à un seul site chez *Clostridium difficile* (Wang *et al.* 2000). Ainsi la distinction entre transposons conjugatifs et éléments intégratifs site spécifique et conjugatifs n'est pas pertinente. Ces éléments ont été regroupés sous le terme d'ICE (pour « Integrated Conjugative Element ») (Burrus *et al.* 2002). Les ICE sont des éléments qui s'excisent par recombinaison site spécifique sous forme circulaire, codent pour leur propre transfert conjugatif sous forme simple brin et s'intègrent dans le génome de l'hôte après réplication, quels que soient la spécificité et les mécanismes d'intégration et de conjugaison.

Ces éléments possèdent un gène *int,* qui code une intégrase catalysant la réaction de recombinaison entre le site *attP* et le site *attB* chromosomique. Le produit du gène *xis* est une excisionase, qui catalyse l'excision de l'élément par recombinaison entre les sites *attL* et *attR*.

Ces ICE contiennent des modules de régulation de l'intégration et de l'excision. De plus, chaque élément porte des séquences leur conférant leurs propriétés spécifiques. Ces éléments évoluent par échanges de modules entre ICE mais également entre ICE et plasmides conjugatifs ou encore par l'intégration d'éléments mobiles non conjugatifs. Par exemple, *Tn5267* de *L. lactis* porte les gènes de synthèse de la lantibiotic nisine (van der Meer *et al.* 1993). Une région de 23 kb de l'élément SXT de *V. cholerae* contient les gènes de résistance à divers antibiotiques, comme la streptomycine et le chloramphénicol. Cette région présente une structure similaire à un transposon composite de type I, flanquée par des IS (Hochhut *et al.* 2001).

La figure 7 résume le transfert conjugatif de ces éléments.

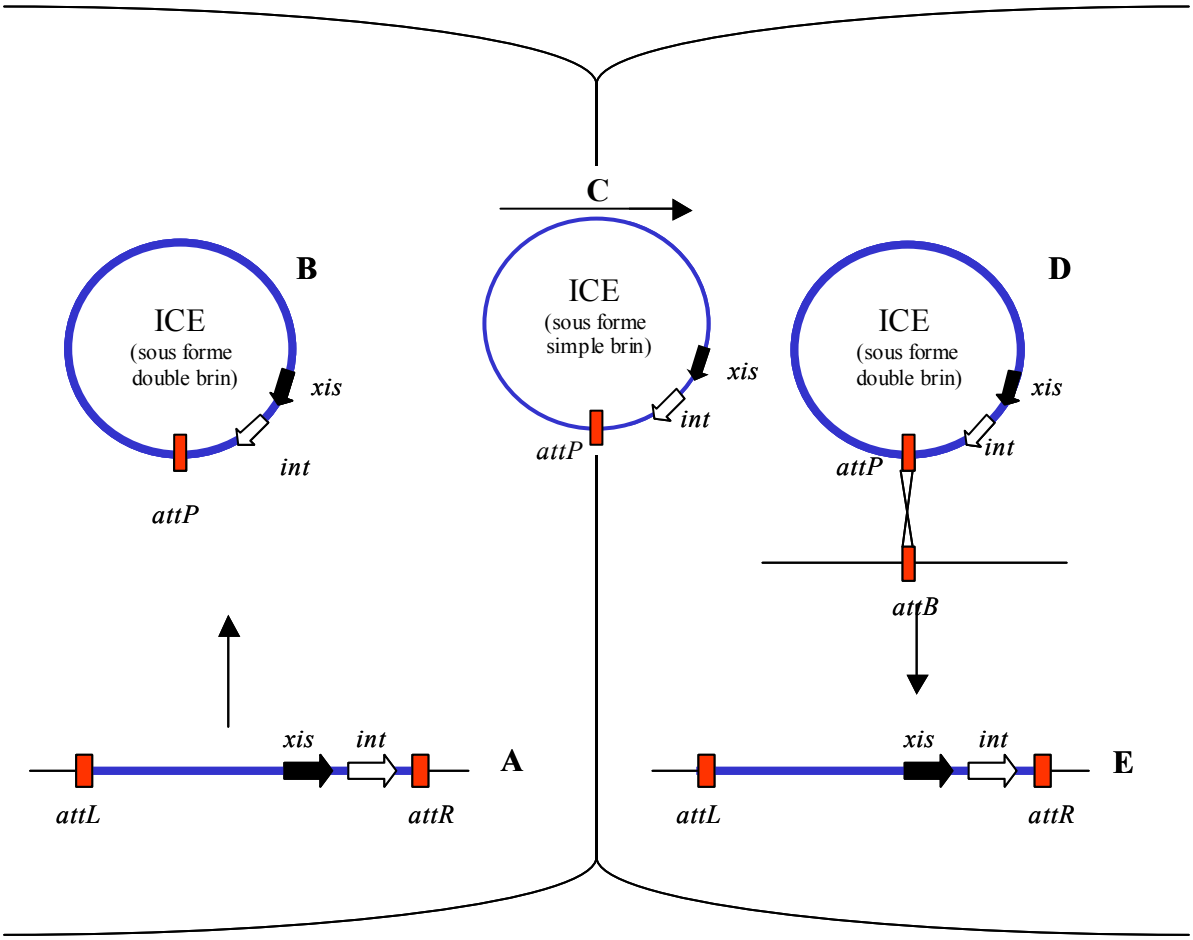


Figure 7: Transfert conjugatif des ICE
A. L'ICE est intégré au chromosome de la cellule hôte. Seuls les gènes *xis* et *int*, responsables du transfert de cet élément sont représentés ainsi que les sites *attL* et *attR*.
B. Après excision par recombinaison entre les sites *attL* et *attR*, l'ICE se trouve sous forme circulaire.
C. Réplication et transfert conjugatif de l'élément sous forme simple brin.
D. Dans la cellule réceptrice, après réplication, l'élément va s'intégrer au niveau de son site d'intégration par recombinaison entre ce site (*attB*) et le site *attP* de l'élément.
E. Après recombinaison, l'élément est intégré au chromosome de la cellule réceptrice.

B-2b Les plasmides conjugatifs

Historiquement, de nombreux exemples de transferts conjugatifs de gène de résistance portés par des plasmides ont été décrits entre bactéries de même coloration Gram comme entre bactéries Gram+ et bactéries Gram - (pour revue (Courvalin 1994)). Par exemple, le gène *aphA-3* de résistance à l'amikacine et à la kanamycine retrouvé chez *Campylobacter coli* est porté par un plasmide d'*Enterococcus* (Trieu-Cuot *et al.* 1985).

Le plasmide conjugatif modèle est le facteur F découvert chez *E. coli* K12 par Tatum et Lederberg (1946). Ces éléments sont retrouvés chez de nombreuses espèces bactériennes Gram +, comme chez *Streptococcus*, *Enterococcus* et *Staphylococcus* par exemple (Scott et Churchward 1995; Grohmann *et al.* 2003).

Ces plasmides contiennent l'ensemble des fonctions nécessaires à leur transfert conjugatif (pour revue (Frost *et al.* 1994)). Pour le transfert du facteur F, un filament extracellulaire codé par la région de transfert, va créer un contact avec la cellule réceptrice. Quand le contact est stable, l'enzyme TraI va couper le site *oriT* sur un seul brin (Fekete et Frost 2000). Le brin T simple brin

est ensuite transféré dans une direction 5'-3' vers la cellule réceptrice. Le transfert entier du facteur F (100 kb) prend approximativement 5 minutes. Dans la cellule réceptrice, le brin transféré est religué puis la synthèse du brin complémentaire est initiée.

Les différences majeures entre la conjugaison des bactéries Gram + et des bactéries Gram – tiennent sur les mécanismes qui permettent l'établissement du contact cellulaire. En effet, la membrane externe et la paroi de peptidoglycane des bactéries Gram – doivent être traversées pour permettre le passage de l'ADN d'une cellule à l'autre.

Les *Streptomyces* ne semblent pas être naturellement transformables (Chater et Hopwood 1984). Dans les génomes de *Streptomyces* séquencés, aucun homologue aux gènes de la compétence de *Ba. subtilis* n'a été retrouvé. De plus, un seul système de transduction généralisé a été décrit, avec le phage ΦSV1 de *Streptomyces venezuelae* (Chater et Hopwood 1984).

En revanche, de nombreux éléments conjugatifs ont été identifiés (Hopwood 2006). Un grand nombre de plasmides conjugatifs a été décrit : des plasmides de petites tailles, codant uniquement des fonctions de transfert et réplication, comme pIJ101 de *Streptomyces lividans* (Kieser *et al.* 1982). De grands plasmides intégratifs linéaires tels que SCP1 (360 kb) de *S. coelicolor* (Hopwood *et al.* 1983) contiennent d'autres fonctions, en plus de leurs fonctions de transfert et de réplication.

Un des plasmides les mieux connus est le plasmide pSAM2 de *S. ambofaciens*. A la différence du plasmide F, ce plasmide est intégré dans un site unique, au niveau de l'extrémité 3' du gène de $ARNt^{Pro}$ (Raynal *et al.* 1998). Les gènes de réplication (*repSA*), d'excision (*xis*) et d'intégration (*int*) sont regroupés en opéron sous le contrôle du produit du gène *pra* (Sezonov *et al.* 1995). La transcription de *pra* est réprimée en absence de cellule réceptrice et pSAM2 reste sous forme intégré. Après contact physique entre cellules donneuse et réceptrice, l'opéron est transcrit, pSAM2 s'excise et se réplique par cercle roulant (Fig. 8) (Possoz *et al.* 2001). Une des copies va se réintégrer dans le génome de la cellule donneuse. La protéine Tra permet le transfert de pSAM2 vers la cellule réceptrice. A la différence du plasmide F, pSAM2 se transfère sous forme double brin (Possoz *et al.* 2001). Chez les *Streptomyces*, le transfert conjugatif de ce type d'élément s'accompagne d'une inhibition transitoire de la croissance de la cellule réceptrice, formant des zones de faible croissance nommées « pock » (Hopwood et Kieser 1993), phénomène non observé lors du transfert du plasmide F. Après transfert et réplication, une copie de pSAM2 s'intègre de façon site spécifique dans le génome de la cellule donneuse et receveuse. Après acquisition dans le mycélium récepteur, un mécanisme de dispersion intramycélien se met en place, codé par les gènes *spd* (« spread ») (Kieser *et al.* 1982). Le plasmide pSAM2 n'est pas un véritable plasmide conjugatif mais pourrait être considéré comme un ICE (Burrus *et al.* 2002). A la différence du plasmide F, qui nécessite près de 30 gènes pour son transfert, le transfert conjugatif intermycélien chez les *Streptomyces* ne nécessite que le seul gène *tra* (Hopwood 2006).

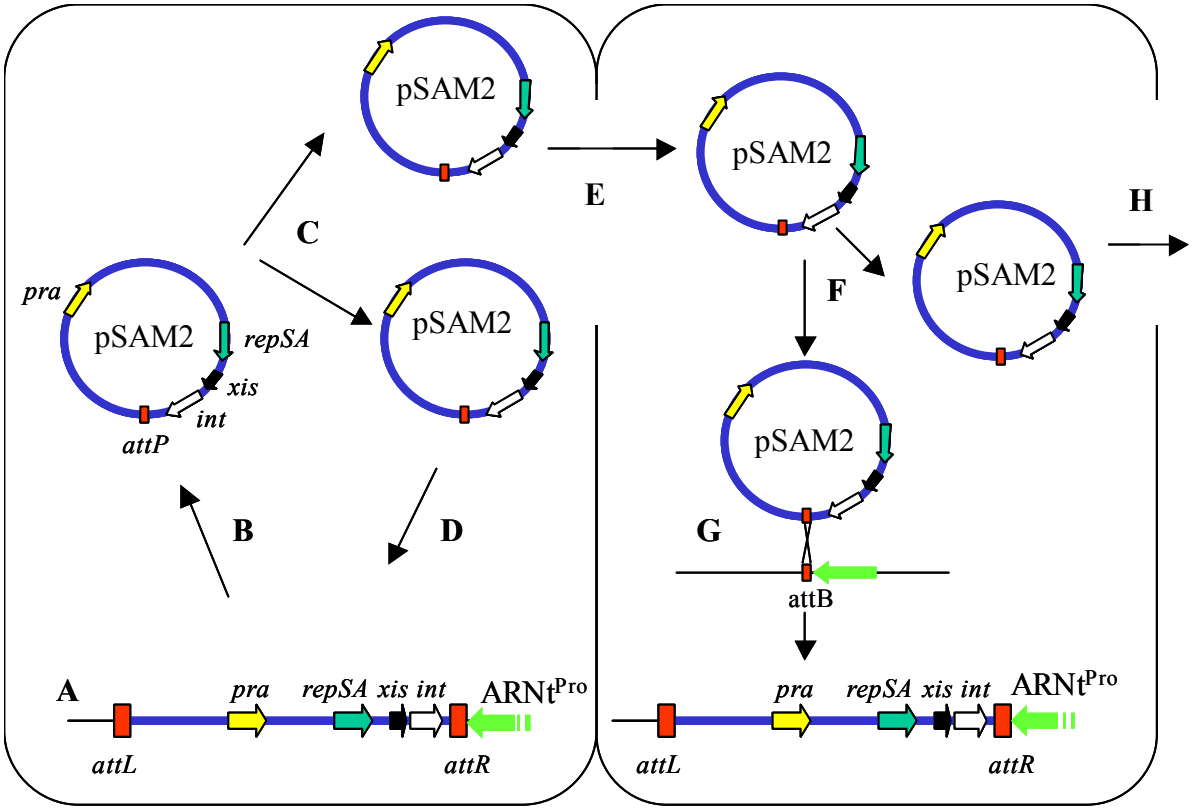


Figure 8 : Modèle de conjugaison du plasmide pSAM2 de *S. ambofaciens*.
A. Le plasmide pSAM2 intégré en 3' du gène de l'ARNt^{Pro}.
B. Excision par recombinaison entre les sites *attL* et *attR*.
C. Réplication du plasmide.
D. Une des deux copies répliquées va se réintégrer dans le génome de la cellule donneuse.
E. L'autre copie est transférée sous forme double brin vers la cellule réceptrice.
F. La copie transférée va se répliquer dans la cellule réceptrice.
G. Une copie du plasmide répliqué va s'intégrer au génome en 3' du gène de l'ARNt^{Pro}.
H. Les autres copies du plasmide vont être transférées dans les cellules voisines du mycélium.

Le transfert des plasmides conjugatifs chez *Streptomyces* est très efficace, pouvant atteindre 100% dans certains cas (Kieser *et al.* 1982). La mobilisation de marqueurs chromosomiques est observée à des fréquences entre 0,1 et 1% (Hopwood *et al.* 1985). Une hypothèse du transfert conjugatif du chromosome à partir d'une ou des extrémités a été proposée par C.W. Chen et nommée modèle « End-first » (Chen 1996). Les télomères constitueraient les origines de transfert dans ce modèle. Ce modèle décrit le transfert d'ADN sous forme simple brin, alors que le transfert de pSAM2 se fait sous forme double brin (Possoz *et al.* 2001). La présence du gène *ttrA* (pour « terminal transfer ») dans l'ensemble des régions subtélomériques des chromosomes et des plasmides de *Streptomyces* est un argument en faveur de ce modèle. Ce gène coderait une hélicase qui joue un rôle dans le transfert conjugatif du plasmide SLP2 et le chromosome de *S. lividans* (Huang *et al.* 2003). De plus, de nombreux exemples de recombinaison entre plasmide conjugatif et régions terminales ont été décrits. Ainsi, un échange, entre les extrémités du bras droit du plasmide SLP2 et le chromosome ancestral de *S. lividans*, a été décrit grâce à la présence d'une région de 15,4 kb, contenant le gène *ttrA*, identique entre ces deux réplicons (Huang *et al.* 2003).

Le mécanisme conjugatif chez les *Streptomyces* diffère de celui décrit chez les autres

bactéries. Chez *M. smegmatis*, la localisation des fonctions de transfert conjugatif serait chromosomique, cette bactérie ne possédant pas de plasmide. L'étude d'une banque cosmidique a permis de mettre en évidence la présence de nombreuses régions qui agissent en *cis* dans le transfert conjugatif, comme pour les souches Hfr d'*E. coli*, permettant potentiellement une probabilité égale pour chaque région du génome d'être transférée (Bhatt et Jacobs 2003; Wang *et al.* 2003). Cette région est plus complexe que les origines de transfert conjugatif traditionnelles et a été nommée *bom* pour « basis of mobility ». Cette séquence a été définie comme un homologue fonctionnel au site *oriT* car sa seule présence permet le transfert conjugatif de plasmides initialement non conjugatifs (Wang *et al.* 2003). Le transfert conjugatif à partir de cette séquence nécessite d'autres facteurs présents dans la cellule donneuse mais non encore identifiés.

B-5 MAINTIEN DES SEQUENCES ACQUISES HORIZONTALEMENT

Dans le contexte d'une contrainte de taille exercée sur les chromosomes, l'acquisition de gènes est contre-balancée par une perte de gènes. Cet équilibre conduit à une compétition entre gènes pour le maintien au sein du génome, établissant un flux de gènes.

B-5a Contribution au fitness

Lawrence et Roth ont proposé une hiérarchisation des gènes en fonction de leur apport au fitness (Fig. 9) (Lawrence et Roth, 1999). Ainsi les gènes seront plus ou moins maintenus au sein du génotype ou au contraire, susceptibles d'être éliminés, selon que la fonction apportée favorisera la capacité reproductrice de la bactérie. Une valeur, nommée « s », a été proposée. En dessous de cette valeur « s », le gène acquis ne sera pas maintenu. Tout matériel acquis et fixé entraîne la perte de gènes.

Plusieurs paramètres vont intervenir sur cette valeur « s ». Plus le taux de mutation sera élevé, plus la valeur adaptative du gène devra être élevée pour permettre son maintien. La taille de la population intervient également sur la valeur adaptative. Plus la population sera petite, plus la perte de gènes par dérive sera accrue et donc plus la valeur adaptative du gène devra être élevée. Enfin, le taux de recombinaison intraspécifique va jouer sur la sélection des allèles mutés. Plus cette recombinaison est efficace plus elle pourra éliminer ces allèles délétés. Réciproquement, plus ce taux de recombinaison diminue, plus un gène devra contribuer de façon significative au fitness pour son maintien.

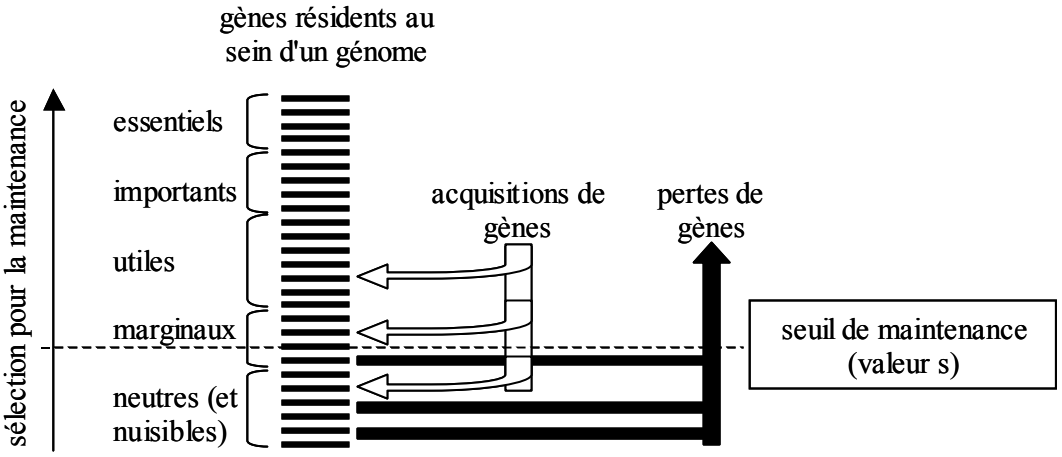


Figure 9 : Flux de gènes et évolution (d'après Lawrence et Roth, 1999)
Chaque trait représente un gène. Tous les gènes dont la valeur adaptative sera inférieure à la valeur seuil ne pourront pas être maintenus.

B-5b Contribution de l'interaction avec d'autres protéines sur le maintien de gènes transférés

Un autre paramètre pouvant contribuer au fitness d'une bactérie est l'interaction du produit d'un gène avec d'autres protéines. Le réseau d'interactions protéiques PPI (pour Protein-Protein Interaction), déterminé par Butland *et al.* pour *E. coli*, a été utilisé pour déterminer le réseau d'interactions protéiques d'autres bactéries (Butland *et al.* 2005). Pour *E. coli*, 1000 ORF ont été l'objet d'un remplacement ciblé par un gène de résistance à un antibiotique. Ainsi, par exemple, les auteurs ont déterminé que 118 des protéines testées ne présentaient aucun partenaire détectable. A partir de ces données, les auteurs ont compilé les données des génomes de 7 espèces bactériennes et construit un génome « ancestral » par comparaison entre ces génomes et les données d'interactions d'*E. coli* (Ochman *et al.* 2007). Ainsi, le génome « ancestral » contient 57% des 4288 gènes du génome d'*E. coli*, ainsi que 70% des 1339 protéines ayant des interactions, et 86% des 5868 interactions protéiques détectées chez *E. coli*.

Ainsi, un gène acquis par transfert horizontal sera plus favorablement conservé s'il appartient à un réseau de protéines (Ochman *et al.* 2007). Ce cas de figure est observé par le recrutement de gènes acquis codant des voies métaboliques chez *E. coli* (Light *et al.* 2005). Seulement 20% de protéines acquises interagissent avec un seul partenaire mais cette proportion augmente de façon surprenante avec le nombre de partenaires (28%, 32% et 50% pour deux, trois et quatre partenaires protéiques respectivement). Ainsi, les protéines, interagissant avec de nombreux partenaires d'une voie établie, sont plus conservées et donc leur proportion dans le génome augmente.

Chez *Buchnera*, une protéine interagissant avec un seul partenaire a cinq fois moins de chance d'être maintenue que les protéines ayant plus de 15 partenaires. Cette observation avait déjà été développée chez les eucaryotes (Jeong *et al.* 2001; Krylov *et al.* 2003; Campillos *et al.* 2006). Ainsi, si le PPI d'une protéine augmente, cette protéine aura plus de chance d'être maintenue et réciproquement.

Ce PPI va être diminué avec l'inactivation de gènes. En effet, le symbiote facultatif

Sodalis glossinidia est un cas extrême de la dégradation génomique, où un quart de son génome (soit un peu plus de 1 Mb) est constitué de pseudogènes (Toh *et al.* 2006). De façon significative, la grande majorité des protéines ancestrales, à un seul partenaire, a été inactivée. Ainsi, si le (ou les) partenaire(s) d'une protéine est (sont) inactivé(s), le PPI de cette protéine va diminuer. La diminution du PPI ménera à la perte ou à l'inactivation accrue de cette protéine.

III- Objectifs de la thèse

A- Les Streptomyces

Les *Streptomyces* sont des bactéries présentant une coloration Gram positive, à haut pourcentage en bases G+C (70-72%), appartenant à l'ordre des *Actinomycétales*. Ce sont des bactéries filamenteuses dont le biotope est le sol. Les *Streptomyces* présentent un cycle de différenciation complexe, caractérisé par une différenciation morphologique et biochimique (Chater 1993).

Ces bactéries présentent un grand intérêt industriel. Elles sont à l'origine de la grande majorité des antibiotiques utilisés en thérapie humaine et vétérinaire. Le métabolisme secondaire se met en place dans les étapes tardives du développement morphologique des *Streptomyces*.

A-1 CYCLE DE DEVELOPPEMENT

Le cycle de différenciation débute par la germination d'une spore. A partir de cette spore, un mycélium végétatif à croissance radiale va se développer (Fig. 10). Ce mycélium est constitué de filaments qui contiennent de nombreuses copies du génome. Quand le milieu devient limitant en ressources nutritives, le mycélium végétatif se différencie en mycélium aérien. Ce mycélium aérien est formé d'hyphes contenant plusieurs copies du génome de la bactérie. Ces hyphes forment des chaînes de spores par septation. Chaque spore contient une seule copie du génome de *Streptomyces*.

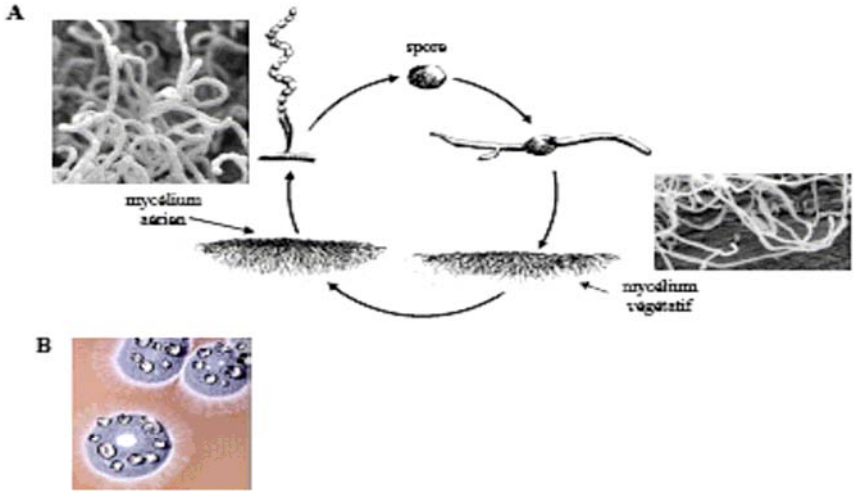


Figure 10: Cycles de développement de *Streptomyces*
A. Cycle de différenciation des *Streptomyces*.
B. Colonies de *S. ambofaciens* sur milieu solide.

A-2 LE GENOME DES *STREPTOMYCES*

Les *Streptomyces* présentent des caractéristiques génomiques originales au sein des bactéries. En effet, chaque génome caractérisé se compose d'un chromosome linéaire de grande taille (8,7 Mb pour *S. coelicolor*, 9,1 Mb pour *S. avermitilis*) avec un pourcentage très élevé en bases G+C (71% à 73%). De nombreuses espèces possèdent en plus des plasmides linéaires et/ou circulaires. Le chromosome des *Streptomyces* présente des Répétitions Terminales Inversées ou TIR de composition et de tailles variables. Les réplicons linéaires sont répliqués depuis une origine bidirectionnelle localisée plus ou moins à équidistance des extrémités (Chang et Cohen 1994; Musialowski *et al.* 1994).

Au niveau des extrémités, le brin 3' ne peut être répliqué de façon continue. Environ 280 nucléotides restent sous forme simple brin (Chang et Cohen 1994). Ces séquences télomériques contiennent 7 palindromes qui sont susceptibles de se replier en structures tiges-boucles (Huang *et al.* 1998). Cette structure, ainsi que les séquences télomériques, sont très conservées au sein des *Streptomyces* (Fig. 2). Cependant, des télomères atypiques ont été mis en évidence sur le chromosome de *Streptomyces griseus* (Goshi *et al.* 2002), le plasmide linéaire SCP1 de *S. coelicolor* (Kinashi *et al.* 1991) ou encore les plasmides pRL1 et pRL2 respectivement des souches de *Streptomyces* 44030 et 44414 (Zhang *et al.* 2006). Ceux-ci sont également constitués de palindromes mais leurs séquences sont différentes entre elles et avec celles classiquement retrouvées chez les *Streptomyces*. Ces différences de séquences conduisent à des structures secondaires distinctes de celles habituellement retrouvées. Pour le plasmide SCP1, par exemple, les boucles sont constituées de quatre nucléotides au lieu de trois (Bentley *et al.* 2004). Pour le plasmide pRL2, les tiges sont plus courtes, la structure secondaire ne compte que trois tiges-boucles, dont deux contiennent un triplet AAG au lieu de ACG chez les autres télomères de *Streptomyces* (Fig. 11) (Zhang *et al.* 2006).

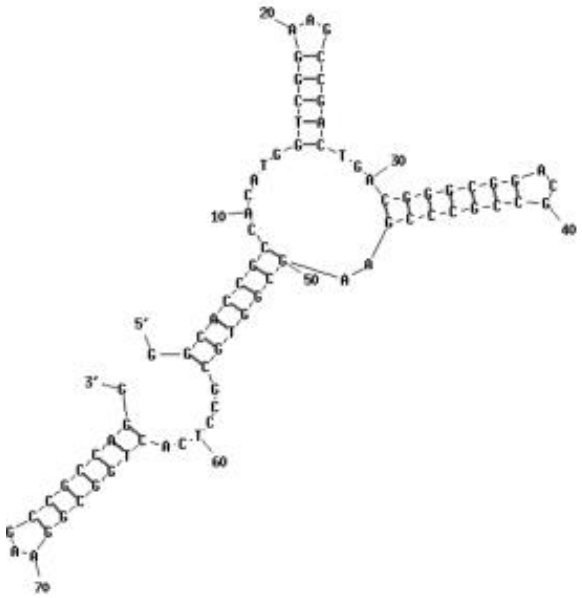


Figure 11 : Structure secondaire des télomères du plasmide pRL2 de la souche de *Streptomyces* 44414 (d'après Zhang et al., 2006)

Les extrémités des réplicons linéaires sont liées de façon covalente, par le brin 5', à une protéine terminale, nommée Tpg (pour Terminal Protein Gene, (Bao et Cohen 2001)). Dans le voisinage immédiat du gène *tpg* est localisé le gène *tap* (Terminal Associated Protein). Ces deux gènes présenteraient vraisemblablement une structure opéronique (Bao et Cohen 2003). Ces deux protéines sont associées au niveau des palindromes II et III du télomère, pour former un complexe protéique responsable du maintien de la forme linéaire du chromosome (Bao et Cohen 2003).

Ce complexe protéique servirait dans la réplication des extrémités chromosomiques. La protéine Tap se fixe au niveau du télomère. Puis elle recruterait la protéine Tpg, servant d'amorce à la réplication pour initier la réplication des télomères sur le brin discontinu. PolA (ADN Polymérase) et TopA (Topoisomérase I) font également partie de ce complexe télomérique (Bao et Cohen 2004).

A-3 GENOMIQUE COMPAREE CHEZ LES *STREPTOMYCES*

Le séquençage et l'annotation du génome complet de *S. coelicolor* a permis de définir différentes régions chromosomiques, basées sur les fonctions prédites des différentes CDS retrouvées (Bentley *et al.* 2002). La région centrale, nommée région « core » de 4,9 Mb, contiendrait la quasi-totalité des gènes dits « essentiels » au développement végétatif de *S. coelicolor* (intervenant dans la division cellulaire, la réplication etc.). Les régions, contenant les gènes probablement non essentiels à la croissance végétative, comme les gènes du métabolisme secondaire, ont été définies comme régions de contingence. Ces régions correspondent aux régions terminales et elles mesurent 2,3 Mb au niveau du bras droit et 1,5 Mb au niveau du bras gauche chez *S. coelicolor* (Fig. 12).

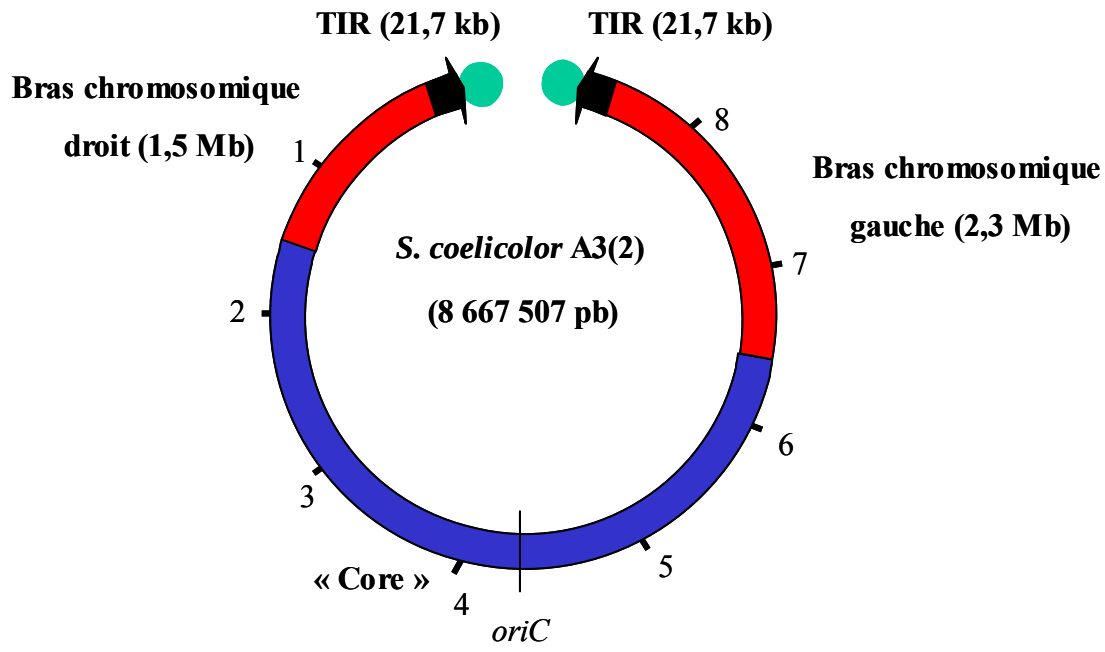


Figure 12 : Génome de *S. coelicolor* (d'après Bentley *et al.*, 2002)
La région bleue correspond à la région nommée « core » qui contient la quasi-totalité des gènes dits « essentiels ». Les bras chromosomiques, porteurs des gènes de contigence, sont représentés en rouge. Aux extrémités des bras chromosomiques se situent les Répétitions Terminales Inversées (ou TIR) en noir. La taille du génome, des bras chromosomiques et des TIR est indiquée.

La région « core » montre une synténie avec les chromosomes circulaires d'autres actinomycètes comme *M. tuberculosis*, *Frankia* ou encore *N. farcinica* (Cole *et al.* 1998; Ishikawa *et al.* 2004; Normand *et al.* 2007). Celle-ci constituerait donc la partie ancestrale du chromosome.

Le séquençage du génome complet de *S. avermitilis* a été réalisé en 2003 (Ikeda *et al.* 2003). Sa comparaison avec le génome de *S. coelicolor* a montré que les gènes définis comme « essentiels » sont retrouvés dans la région de conservée entre ces deux espèces (Fig. 13).

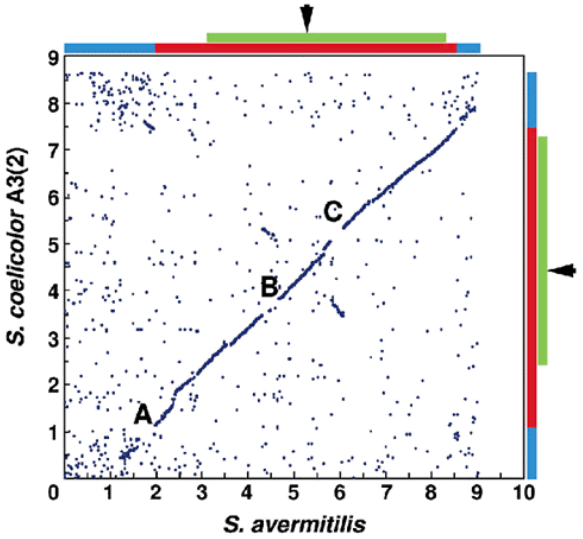


Figure 13 : Comparaison par dot-plot des génomes de *S. coelicolor* et *S. avermitilis* (d'après Ikeda et al., 2003)
Chaque point correspond à la localisation d'un gène homologue entre ces deux espèces. Les points sur la diagonale représentent la région de synténie entre *S. coelicolor* et *S. avermitilis*. A, B et C indiquent les trois événements de réarrangement. Les flèches représentent les origines de réplication *oriC* de chaque génome. La barre verte correspond aux régions conservées avec les chromosomes circulaires d'autres actinomycètes, les barres bleues les régions subtélomériques et les barres rouges les régions « core » de ces génomes.

Trois événements de réarrangement (inversions) (notés A, B et C sur la figure 13), dont deux autour de l'origine de réplication, sont observés. En revanche, les régions proches des extrémités chromosomiques sont moins conservées. Ces régions subtélomériques variables représentent 2 Mb et 0,5 Mb respectivement pour les bras chromosomiques gauche et droit de *S. avermitilis* et 1,1 Mb pour chaque bras chromosomique de *S. coelicolor*. Ces régions terminales ont été définies comme spécifiques d'espèces.

A-4 INSTABILITE GENETIQUE CHEZ LES *STREPTOMYCES*

Chez les *Streptomyces*, les taux de mutations spontanées sont très élevés (pour revues : (Leblond et Decaris 1994; Volff et Altenbuchner 2000; Chen *et al.* 2002)). Chez *S. ambofaciens*, les variants affectés dans la pigmentation des colonies apparaissent à la fréquence de 1% (colonies totalement dépigmentées) (Leblond *et al.* 1989; Martin *et al.* 1998). Ce phénomène est observé chez tous les *Streptomyces* étudiés.

Cette instabilité a été reliée à la formation de grands réarrangements chromosomiques affectant les régions terminales (Birch *et al.* 1989; Leblond *et al.* 1989). L'amplification de régions particulières, nommées AUD (Amplifiable Unit of DNA) ou des délétions sont à la base de cette instabilité. Ces régions instables mesurent 1 Mb chez *S. lividans* (Redenbach *et al.* 1993) et elles

représentent 2,3 Mb chez *S. ambofaciens* (Fischer *et al.* 1997a). Dans la descendance de certains mutants, l'instabilité génétique atteint un niveau extrêmement élevé (87%). Ce phénomène a été appelé « hypervariabilité » (Leblond *et al.* 1989). Les délétions peuvent se produire de façon interne au chromosome, n'affectant pas les télomères. Une conformation circulaire, qu'elle soit retrouvée dans une population de mutants ou provoquée artificiellement indique que la réplication et la ségrégation du chromosome ne sont pas affectées par la conformation circulaire du chromosome (Fischer *et al.* 1997a; Volff *et al.* 1997). Parmi les mutants isolés de *S. ambofaciens*, le chromosome des mutants NSA27 et NSA65 résulte de la fusion de deux chromosomes partiellement délétés sur le même bras chromosomique (Wenner *et al.* 2003). Cette structure est faiblement héritée dans la descendance et une très grande majorité de mutants présente de grands réarrangements chromosomiques. L'analyse de cette descendance et de la structure parentale a permis de faire un lien potentiel entre instabilité et partition chromosomique. En effet, par analogie avec ce qui est observé avec les chromosomes dicentriques eucaryotes, ce mutant présente une instabilité de sa structure chromosomique. Ce mutant rentrerait dans un cycle CFP (Cassure-Fusion-Pont) qui engendrerait des réarrangements génomiques.

B- Objectifs de la thèse

Le séquençage des régions instables de la souche ATCC23877 a été initié, conjointement entre le Laboratoire de Génétique et Microbiologie (Nancy), le Génoscope (Centre National de Séquençage, Evry) et l'Institut de Génétique et Microbiologie (J-L Pernodet, Orsay) pour poursuivre l'étude de l'instabilité observée chez *S. ambofaciens*. La comparaison des génomes de *S. coelicolor* et *S. avermitilis* a montré une forte conservation de la région centrale alors que les bras chromosomiques ont été définis comme spécifiques d'espèces (Ikeda *et al.* 2003). La comparaison de ces régions terminales avec deux autres espèces de *Streptomyces*, une phylogénétiquement proche (*S. coelicolor*) et l'autre phylogénétiquement plus éloignée (*S. avermitilis*) permettra d'aborder l'évolution de ces régions terminales et aborder les questions suivantes :

- Comment évolue la variabilité génomique en fonction de la distance phylogénétique entre ces différentes espèces ?
- Quelle est l'intensité de cette variabilité au niveau des régions terminales par rapport à la région centrale ?

Parallèlement, les séquences des TIR des deux souches de *S. ambofaciens* ATCC23877 et DSM40697 ont été comparées. Ces deux souches sont deux isolats indépendants du sol, appartenant à la même espèce suivant les critères morphologiques et physiologiques classiques comme par exemple la production des antibiotiques spiramycine, congocidine et alpomycine (Pinnert-Sindico 1954; Pang *et al.* 2004). L'appartenance à la même espèce a été confirmée par des méthodes moléculaires. Les profils de restriction *Ase*I du chromosome présentent de grandes similitudes (Leblond *et al.* 1996). Enfin la comparaison des ITS (Internal Transcribed Spacers) des six opérons ribosomiques, couramment utilisées pour préciser les relations phylogénétiques à courtes distances évolutives, a montré que les deux souches de *S. ambofaciens* partagent les mêmes séquences (100% d'identité) alors qu'un polymorphisme existe avec les autres espèces de

Streptomyces dont *S. coelicolor* (Wenner *et al.* 2002).

L'analyse des séquences des TIR de la souche de *S. ambofaciens* DSM40697 à partir d'une banque cosmidique couvrant la totalité de ces dernières est un des objectifs de la thèse. La comparaison des séquences obtenues avec celle de la souche de *S. ambofaciens* ATCC23877 permettra d'étudier la variabilité terminale au niveau intraspécifique et d'aborder le passé évolutif proche des extrémités chromosomiques.

Enfin, l'aspect dynamique des régions terminales du chromosome des *Streptomyces* a été étudié. Parmi les mécanismes potentiels dont découlent les réarrangements observés et l'évolution des régions terminales des chromosomes des *Streptomyces*, le lien entre défauts de partition et instabilité a été abordé. La construction de souches présentant un second centre de partition actif au niveau des régions terminales a été réalisée pour déterminer si la partition chromosomique jouerait un rôle analogue à ce qui est observé chez les eucaryotes dans les réarrangements chromosomiques.

MATERIELS ET METHODES

MATERIEL ET METHODES

I- Souches Bactériennes, vecteurs et conditions de culture

A- Souches bactériennes

Les souches de *S. ambofaciens* DSM40697 et ATCC23877 sont utilisées dans ce travail (Pinnert-Sindico *et al.* 1954-1955; Hutter 1967).

La souche d'*E. coli* DH5α (Stratagene®) est utilisée comme hôte de propagation des dérivés recombinants pGEM®-T easy (Promega), de pIJ2925, pIJ773 et pIJ778. Son génotype est (*supE*44, *ΔlacU*169 (ϕ80 *lacZ*ΔM15), *hsdR*17, *recA*1, *endA*1, *gyrA*96, *thi*-1, *relA*1).

La souche d'*E. coli* JM101 est utilisée pour la propagation des phages M13 recombinants. Son génotype est (F′ *traD*36 *proA*$^{+}$*B*$^{+}$ *lacI*^{q} Δ*(lacZ)M*15/ Δ*(lac-proAB) glnV thi*).

La souche *E. coli* JM109 (Promega) est utilisée comme hôte pour la préparation d'ADN simple brin des dérivés recombinants de pGEM®-T Easy et du phage M13. Son génotype est (*recA*1, *supE*44, *endA*1, *hsdR*17, *gyrA*96, *relA*1 *thi*Δ(*lac-proAB*) F'[*traD*36, *proAB*$^{+}$, *lacI*^{q}, *lacZ*ΔM15]).

La souche d'*E. coli* SURE (Stratagene®) est l'hôte des cosmides recombinants de SuperCos1 couvrant les TIR (« Terminal Inverted Repeat ») des souches de *S. ambofaciens* DSM40697 et ATCC23877. Son génotype est (*endA*1 *glnV*44 *thi*-1 *gyrA*96 *relA*1 *lac recB recJ sbcC umuC*::Tn5 uvrC e14- Δ(*mcrCB-hsdSMR-mrr*)171 F'[*proAB*$^{+}$ *lacI*^{q} *lacZ*ΔM15 *Tn*10]).

La souche d'*E. coli* DH10B (InVitroGene®) sert d'hôte aux dérivés recombinants du vecteur pCNS pour la création de la banque de plasmides couvrant les TIR des souches de *S. ambofaciens* ATCC23877 et DSM40697. Son génotype est (F^{-} *endA*1 *recA*1 *galU galK deoR nupG rpsL* Δ*lacX*74 Φ80*lacZ*ΔM15 *araD*139 Δ(*ara,leu*)7697 *mcrA* Δ(*mrr-hsdRMS-mcrBC*) λ^{-}).

La souche d'*E. coli* BW25113 est utilisée comme hôte pour les cosmides recombinants et les plasmides pIJ790 et pKD20 dans les expériences de recombinaison entre le cosmide et le produit de PCR pour l'introduction ou le remplacement allélique par des sites *parS* par la méthode de PCR-Targeting. Son génotype est (Δ(*araD-araB*)567 Δ*lacZ*4787(::*rrnB-4*) *lacIp*-4000(lacI^{Q}) λ^{-}, *rpoS*369(Am) *rph*-1 Δ(*rhaD-rhaB*)568 *hsdR*514).

La souche ET12567 contient le plasmide pUZ8002 permettant la mobilisation du cosmide recombinant lors de la conjugaison intergénérique entre *E. coli* et *S. ambofaciens*. Son génotype est (*dam dcm hsd*S *Cm*^{R}).

B- Vecteurs

Le vecteur pGEM®-T Easy (Promega, Figure 14) est utilisé pour un clonage direct des produits de PCR.

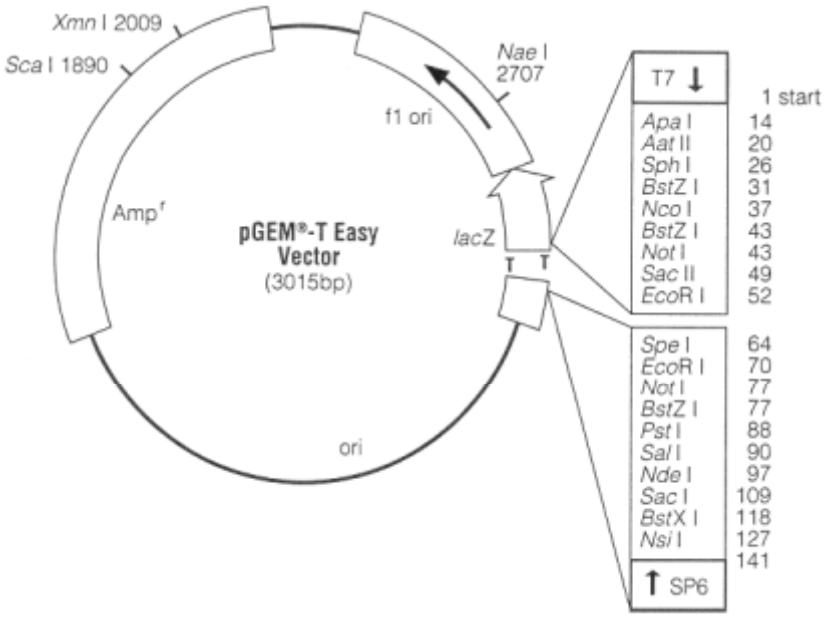


Figure 14 : Carte du plasmide pGEM®-T Easy
Amp^r : gène codant une β-lactamase confèrant la résistance à l'ampicilline.
ori : origine de réplication pour son maintien dans *E. coli*.
lacZ : gène codant la β-Galactosidase. Il contient la cassette multiple de clonage.
f1 ori : origine de réplication du phage filamanteux f1.
T7 et SP6 : localisation et direction des promoteurs des phages T7 et SP6.

En effet, il est fourni sous forme linéarisée avec un nucléotide T protubérant à chaque extrémité 3'. De nombreuses polymérases utilisées lors de réactions d'amplification rajoutent un A à l'extrémité 3' du brin néo-synthétisé. Cela permet un clonage direct des produits de PCR dans le vecteur (appariement entre les bases A protubérantes des produits et les bases T protubérantes du vecteur). Le vecteur porte un gène codant une β-lactamase conférant la résistance à l'ampicilline. Le vecteur a été linéarisé par coupure dans le gène *lacZ*, ce qui permet d'identifier les dérivés recombinants par le phénotype conféré à savoir la coloration des colonies sur un milieu contenant de l'X-gal (80 μg/mL) et d'IPTG (0,5mM). Ce plasmide possède une origine de réplication de type f1 permettant la préparation d'ADN simple brin. Cet ADN est ensuite utilisé comme matrice lors de réactions de séquençage.

Les plasmides pIJ773 et pIJ778 portent les cassettes utilisées pour les remplacements de gènes. Pour pIJ773 (Fig. 15), ces cassettes se composent, du gène *aac3(IV)* conférant la résistance à l'apramycine et de l'origine de transfert *oriT* du plasmide RK2 (Pansegrau *et al.* 1994).

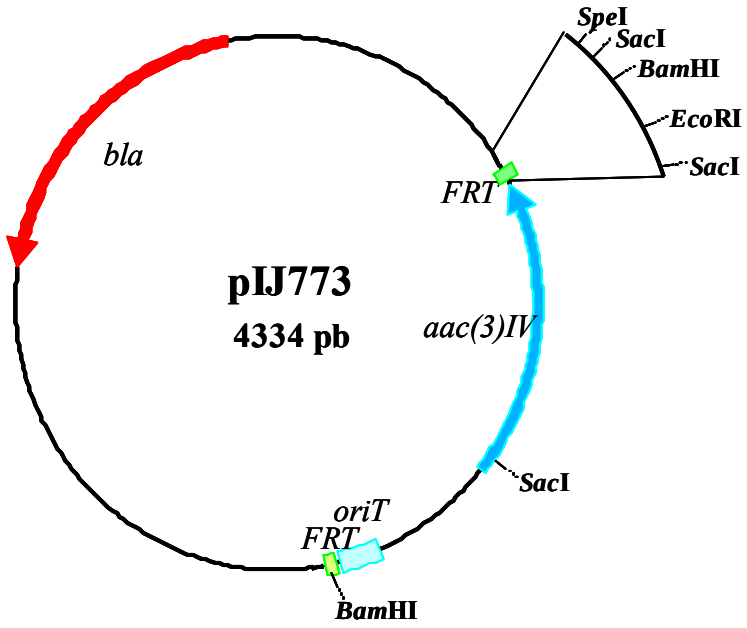


Figure 15 : Carte du plasmide pIJ773 (d'après Gust *et al.*, 2002)
Le plasmide pIJ773 porte la cassette de résistance à l'apramycine utilisée pour l'insertion de séquences par la technique de mutagénèse ciblée par PCR.
bla : gène codant une β-lactamase qui confère la résistance à l'ampicilline.
aac(3)IV : gène codant l'apramycine acétyltransférase qui confère la résistance à l'apramycine.
oriT : origine de transfert du plasmide RK2.
FRT : (pour « FLP Recognition Targets ») sites reconnus par la FLP recombinase.

Pour pIJ778, la cassette se compose du gène *aadA* conférant la résistance aux antibiotiques spectinomycine et streptomycine, ainsi que du locus *oriT* (Fig. 16). Chacune de ces cassettes est flanquée des sites FRT pour « FLP Recognition Target ».

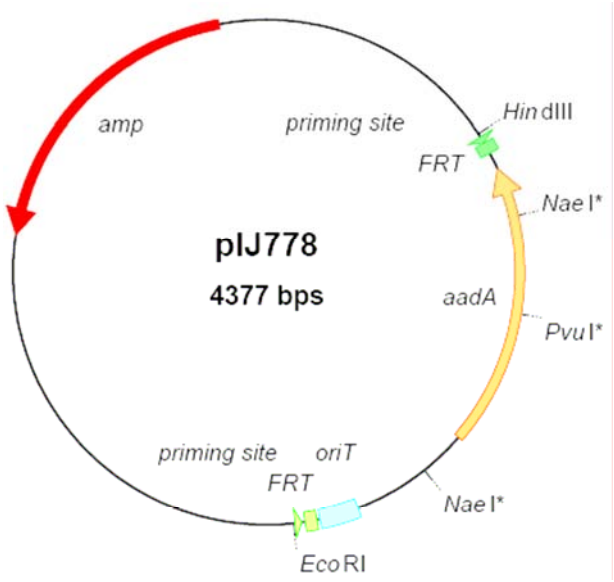


Figure 16 : Carte du plasmide pIJ778 (d'après Gust *et al.*, 2002)
Le plasmide pIJ778 porte la cassette de résistance à la spectinomycine/streptomycine utilisée pour le remplacement de gène par la technique de mutagénèse ciblée par PCR.
amp : gène codant une β-lactamase. Confère la résistance à l'ampicilline.
aadA : gène codant une spectinomycine adényltransférase ? Confère la résistance à la spectinomycine et à la strptomycine.
oriT : origine de transfert du plasmide RK2.
FRT : (pour « FLP Recognition Targets ») sites reconnus par la FLP recombinase.

Le plasmide pKD20 (Fig. 17) porte le système de recombinaison λRED ainsi que le gène *bla* de résistance à l'ampicilline (Datsenko et Wanner 2000). Ce plasmide est utilisé lors des expériences de remplacement du système de résistance des BAC recombinants dérivant de pBeloBAC11 par le système de conjugaison du plasmide pSET152. Ce plasmide a été utilisé car il possède un système de résistance différent de celui du vecteur pBeloBAC11.

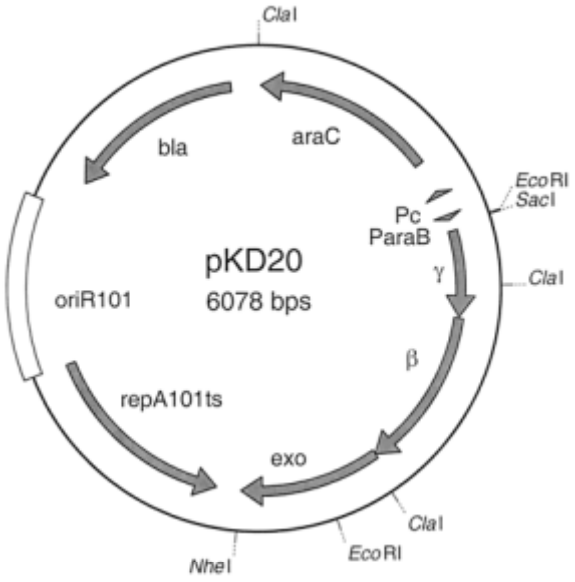


Figure 17 : Carte du plasmide pKD20 (d'après Datsenko et Wanner, 2000)
Le plasmide pKD20 porte le système de recombinaison λRED.
araC : gène codant l'activateur transcriptionnel AraC.
bet: gène codant la protéine Bet, une protéine se fixant sur l'ADN simple brin.
bla : gène codant une β-lactamase, conférant une résistance à l'ampicilline.
exo : gène codant l'exonucléase qui promeut la recombinaison avec Bet.
gam : gène codant la protéine Gam, un inhibiteur de l'exonucléase V de l'hôte.
oriR101 : origine de réplication.
*repA*101ts : origine de réplication thermosensible
P*araBAD* : promoteur de l'opéron arabinose inductible par le L-arabinose.

Le plasmide pUZ8002, dérivant du plasmide RP4, fournit en trans les fonctions de transfert nécessaires à la conjugaison mais n'est pas mobilisable (absence du locus *oriT*) (Paget *et al.* 1999). Ce plasmide porte le gène de résistance à la kanamycine.

Le vecteur pGEM®T easy exige la présence du phage « helper » R408 (Promega®) qui permet sa propagation, son empaquetage et sa sécrétion sous forme d'ADN simple-brin.

Le plasmide pIJ790 (Fig. 18) dérive du plasmide pKD20 portant le système de recombinaison λRED. Le gène *bla* de pKD20 a été remplacé dans pIJ790 par le gène *cat* conférant la résistance au chloramphénicol. Le système λRED de ce plasmide est utilisé pour le remplacement allélique ou insertion de sites *parS* dans les cosmides recombinants.

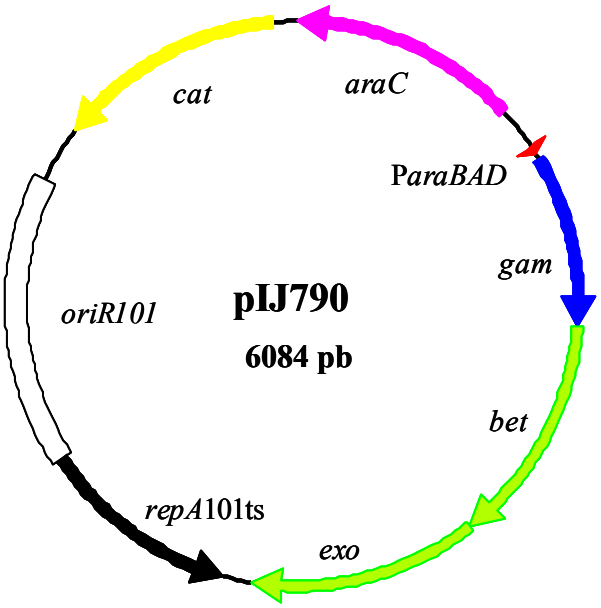


Figure 18 : Carte du plasmide pIJ790 (d'après Gust *et al.*, 2002)
Le plasmide pIJ790 dérive du plasmide pKD20 (Datsenko et Wanner, 2000) portant le système de recombinaison λ-RED, constitué des gènes gam, bet et exo. Le gène bla de pKD20 conférant la résistance à l'ampicilline a été remplacé dans pIJ790 par le gène cat conférant la résistance au chloramphénicol.
araC : gène codant l'activateur transcriptionnel AraC.
bet: gène codant la protéine Bet, une protéine se fixant sur l'ADN simple brin.
cat : gène codant une chloramphénicol acétyltransférase qui confère la résistance au chloramphénicol.
exo : gène codant l'exonucléase qui promeut la recombinaison avec Bet.
gam : gène codant la protéine Gam, un inhibiteur de l'exonucléase V de l'hôte.
oriR101 : origine de réplication.
*repA*101ts : origine de réplication thermosensible
P*araBAD* : promoteur de l'opéron arabinose inductible par le L-arabinose.

Le vecteur M13mp18 (Fig. 19), dérivant du phage M13, contient un site multiple de clonage dans le gène *lacZ* de la β-Galactosidase. En présence de X-Gal et d'IPTG, les vecteurs recombinants peuvent être distingués des vecteurs natifs par l'absence de coloration bleue des plages de lyse. La forme double-brin intracellulaire (RF-I) de M13mp18 permet de le manipuler comme un plasmide. Ce phage permet la synthèse d'ADN simple brin. Cet ADN simple brin est encapsidé dans la capsule phagique et libéré à travers la paroi bactérienne dans le milieu extracellulaire. Cet ADN simple brin est ensuite utilisé comme matrice pour le séquençage.

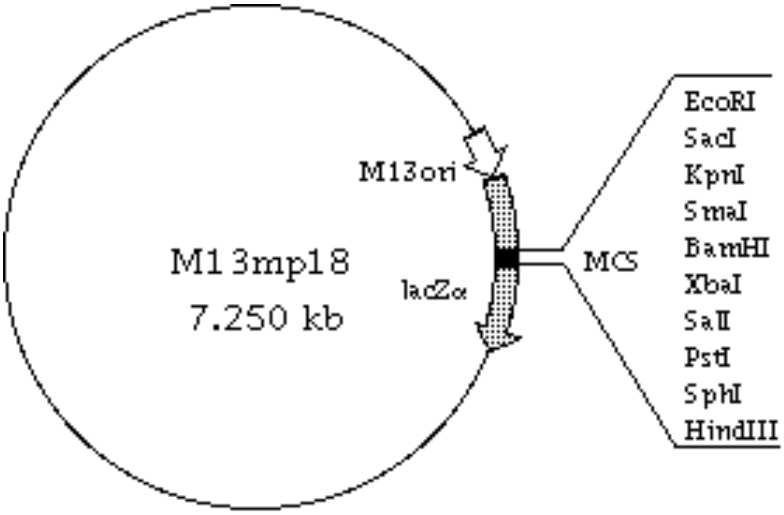


Figure 19 : Carte du phage M13mp18
Ce vecteur est utilisé pour la réalisation de la banque plasmidique et pour la synthèse d'ADN sous forme simple brin.
lacZ : gène de la β-Galactosidase.
MCS : Site Multiple de Clonage.
M13 ori : origine de réplication.

Le plasmide pCNS dérive du plasmide pSU18 (Bartolome *et al.* 1991), et utilisé comme vecteur de clonage lors de la réalisation des banques plasmidiques couvrant les TIR de *S. ambofaciens* ATCC23877 et DSM40697. Ce plasmide possède un site multiple de clonage dans le gène *lacZ*, et un gène de résistance au chloramphénicol.

Le plasmide pIJ2925 (Fig. 20) sert de vecteur de clonage pour la multimérisation des sites *parS*. Ce plasmide possède le gène *bla* de résistance à l'ampicilline ainsi qu'un site multiple de clonage dans le gène *lacZ*.

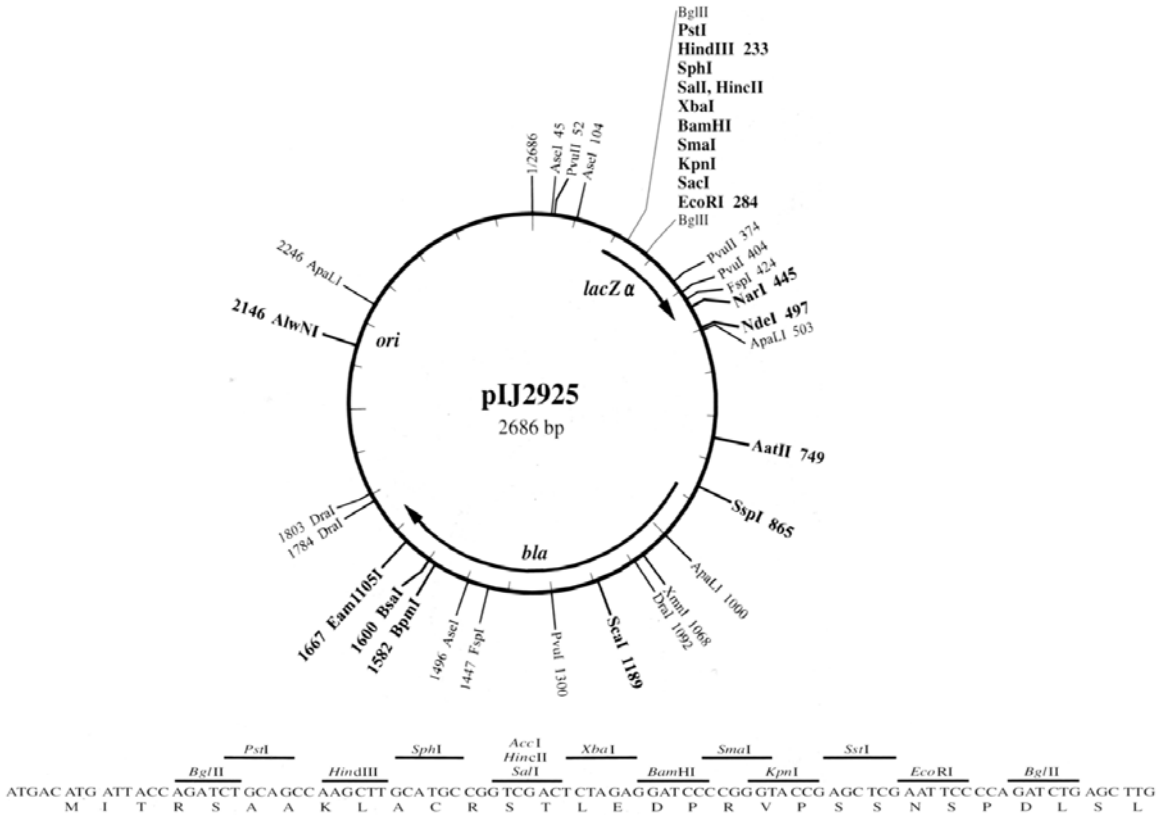


Figure 20 : Carte du plasmide pIJ2925 (Kieser *et al.*, 2000)
Le plasmide pIJ2925 porte le système blanc/bleu codé par le gène de la β-Galactosidase et possède un site de clonage multiple dans ce gène.
lacZα : gène codant le β-Galactosidase.
bla : gène codant la β-Lactamase qui confère la résistance à l'ampicilline.
ori : origine de réplication du plasmide.

La banque cosmidique, réalisée dans *E. coli* SURE comme hôte, recouvrant les TIR de *S. ambofaciens* ATCC23877 et DSM40697, a été réalisées dans le vecteur SuperCos1 (Evans *et al.* 1989) (Fig. 21). Ce vecteur possède le gène *néo* de résistance à la néomycine et *bla* de résistance à l'ampicilline. L'ADN cloné est inséré au site *Bam*HI et la taille est de 35 à 40 kb.

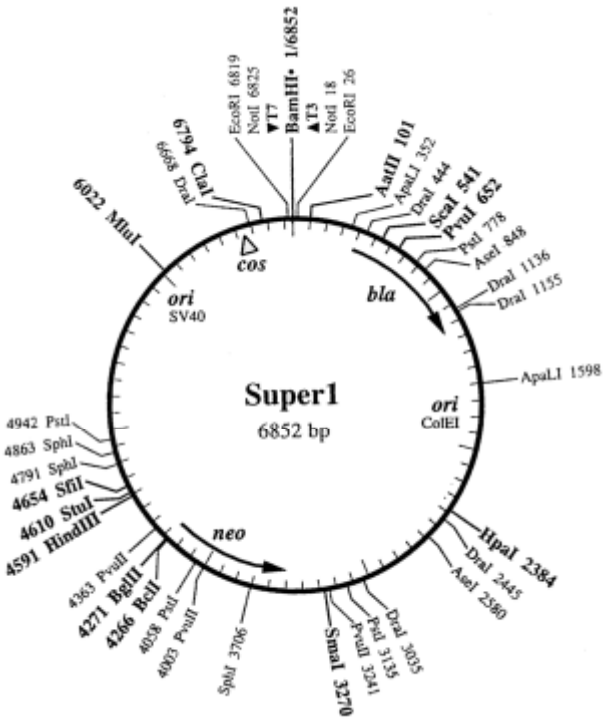


Figure 21 : Carte du vecteur Super1
Vecteur cosmidique (Super1) (Evans *et al.*, 1989) dérivé de Supercos1 aprés recombinaison de l'insert au site *Bam*HI. L'ADN cloné (35 à 40 kb) est inséré au site *Bam*HI (1/6852).
bla : code la β-Lactamase de résistance à l'ampicilline.
cos : extrémités cohésives.
neo : gène de résistance à la néomycine.
*ori*ColE1 : origine de réplication du plasmide ColE1 d'*E. coli.*

Le vecteur pBeloBAC11 (Fig. 22) est un vecteur de clonage utilisé pour la création de Bacterial Artificial Chromosomes (BACs). En raison de son origine de réplication de type F, ce vecteur est présent en copie unique par cellule. Ce vecteur a été utilisé lors de la réalisation de la banque de BAC, couvrant l'ensemble du chromosome de la souche de *S. ambofaciens* ATCC23877.

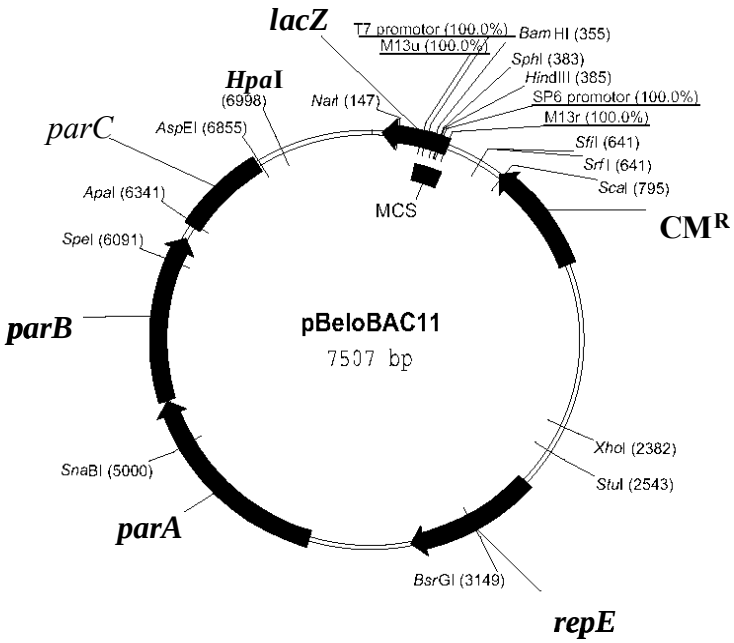


Figure 22 : Carte du vecteur pBELOBac11
Ce vecteur est utilisé pour le clonage de la région centrale du chromosome de *Streptomyces ambofaciens*.
parABC : système de partition du plasmide F.
repE : facteur d'initiation intervenant dans l'assemblage du complexe de réplication au niveau de l'origine de réplication.
CM^r : gène de résitance au chloramphénicol.
lacZ : gène de la β-Galactosidase.

Le plasmide pSET152 dérive du phage des *Streptomyces* ΦC31 et est utilisé comme matrice pour le clonage de son système de conjugaison et intégration (Fig. 23) lors des expèriences de modification des BAC.

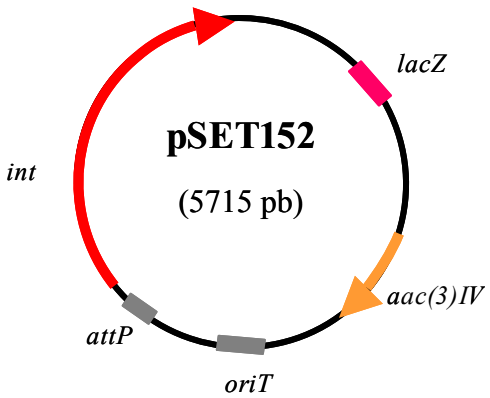


Figure 23 : Carte du plasmide pSET152
lacZ : gène codant la β-Galactosidase.
int : code l'intégrase du phage ΦC31.
attP : site d'attachement du phage ΦC31.
oriT : origine de transfert conjugatif du phage ΦC31.
aac(3)IV : code l'apramycine acétyltransférase qui confère la résistance à l'apramycine.

Les cosmides D50, D118, 25E1 et 14C4 ont été utilisés lors des expériences de clonage réitératif d'amplicons *parS*. La carte de restriction *Bam*HI de chaque cosmide est représentée dans la figure 24.

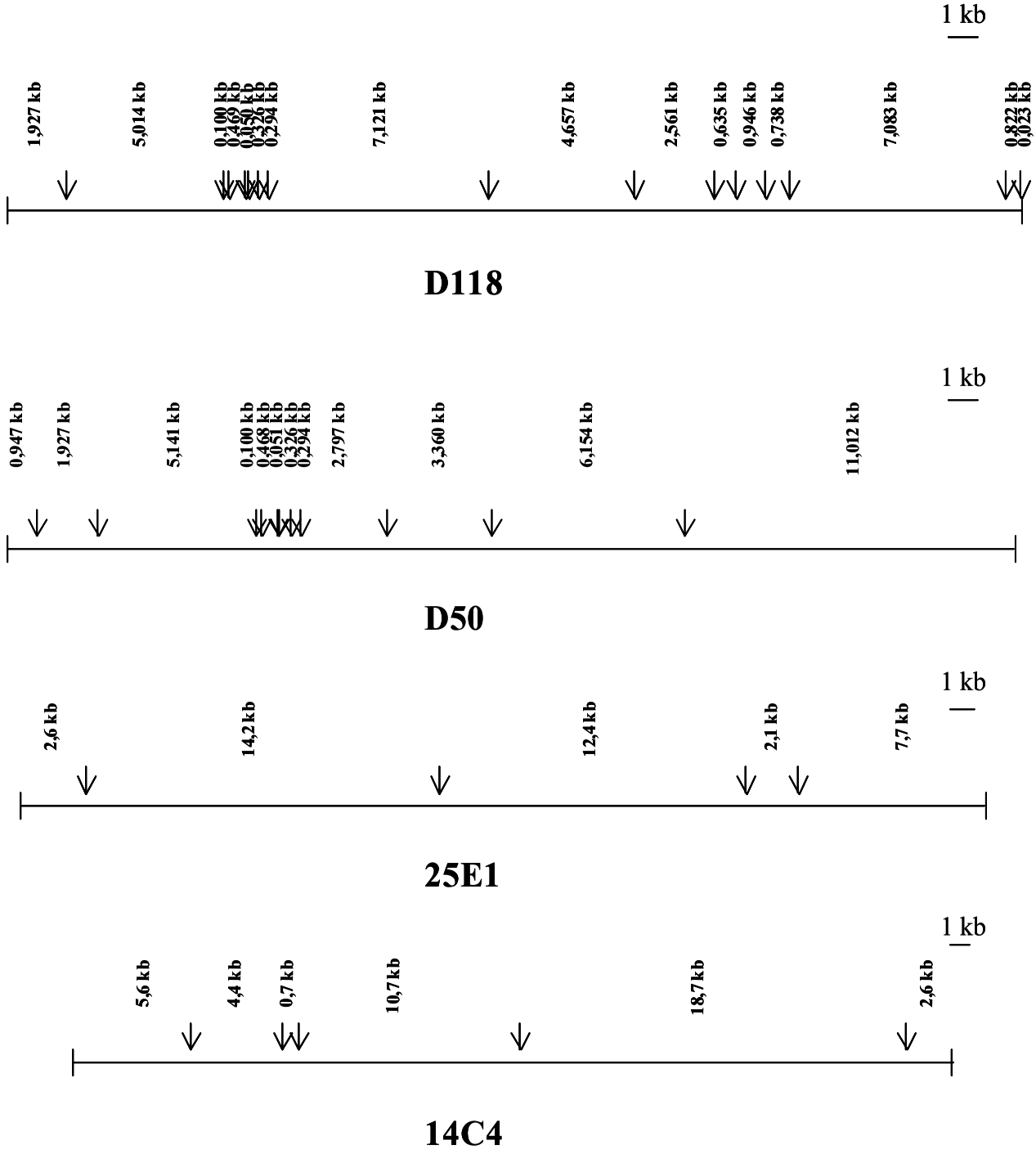


Figure 24 : Carte de restriction des cosmides recombinants D118, D50, 25E1 et 14C4
Carte *Bam*HI des inserts des cosmides recombinants D118, D50, 25E1 et 14C4.

C- Milieux de culture

Les cultures d'*E. coli* sont réalisées en milieu LB solide ou liquide à 30°C et 37°C (Sambrook *et al.* 1989). Les cultures d'*E. coli* réalisées dans le cadre de la conjugaison intergénérique *E. coli-S. ambofaciens* et les cultures permettant la purification d'ADN simple brin de M13, sont réalisées en milieu 2xYT. Dans le cas de la préparation de cellules compétentes d'*E. coli* BW25113, les cultures sont réalisées en milieu SOB. Après l'électroporation de cellules compétentes de souches d'*E. coli*, autre que la souche BW25113, les cellules transformées sont reprises en milieu SOC. Sur milieu solide, la souche *E. coli* JM109 est cultivée sur milieu minimum M9 additionné de thiamine (0,01% final) pour s'assurer du maintien du facteur F' (Sambrook *et al.*, 1989).

Les souches sauvages et mutantes de *S. ambofaciens* sont cultivées sur milieu HT (Pridham *et al.* 1957). Pour les expériences de conjugaison intergénérique *E. coli-S. ambofaciens*, le milieu solide SFM (Kieser *et al.* 2000) supplémenté en $MgCl_2$ (10 mM final) est utilisé. Ce milieu est également utilisé pour les cultures sur milieu gélosé pour la récolte de spores des souches sauvages ou mutantes de *S. ambofaciens*. Le milieu YEME 10% (Kieser *et al.* 2000) (complémenté en $MgCl_2$ et glycine, concentrations respectives finales de 5 mM et 0,5%) est utilisé pour les extractions d'ADN de *Streptomyces.*

Les spores de *S. ambofaciens* sont récoltées, après 7 jours de culture à 30°C sur milieu HT ou SFM solide, par ajout de 3 ml de glycérol 20% (v/v) et par grattage du mycélium à l'aide d'une anse stérile et congélation à -20°C. Le glycérol joue le rôle de cryoprotecteur. La filtration des suspensions de spores obtenues par culture sur milieu HT ou SFM (7 jours à 30°C), permet par passage dans un cône à filtre stérile, de concentrer les spores et d'éliminer les fragments mycéliens (Kieser *et al.* 2000).

II- Manipulation des acides nucléiques

L'ADN plasmidique ou cosmidique d'*E. coli* est extrait suivant la méthode de lyse alcaline dont le protocole est décrit par Sambrook *et al.* (1989). Une extraction par phénol-ChCl (1 :1) est pratiquée pour purifier les acides nucléiques. Pour les cosmides, une étape de purification par utilisation de ClCs a été entreprise suivant le protocole décrit par (Kieser *et al.* 2000) pour éliminer toutes traces d'ADN génomique.

L'extraction de l'ADN simple brin à partir du phage M13mp18 se déroule suivant le protocole décrit par (Bankier 1993). L'extraction d'ADN simple brin des dérivés recombinants de pGEM®-T Easy est réalisée selon Sambrook *et al.* (1989) à partir de surnageants de culture de transformants d'*E. coli* JM109 infectée par le phage « helper » R408 (Russel *et al.* 1986).

Les enzymes de restriction (New England Biolabs et Roche) sont utilisées pour les digestions partielles ou totales suivant les recommandations du fournisseur. Pour le PFGE, 20 UE sont nécessaires pour la digestion de l'ADN génomique contenu dans une lamelle. Pour les digestions partielles, le contrôle de l'action de l'endonucléase est réalisé par le temps d'incubation (1 UE par µg d'ADN).

Les électrophorèses en gel d'agarose sont réalisées en tampon TAE (Tris-acétate 0,04M ; EDTA 0,001M) ou TBE (Tris-borate 0,09M ; EDTA 0,002M). Le système de traitement d'images GelDoc 1000 (Biorad) permet de visualiser les gels après immersion dans une solution de BET à 0,5 µg/ml, puis de traiter et de stocker les données. L'ADN du bactériophage λ (Daniels *et al.* 1983) digéré par *Pst*I et *Hin*dIII ou le cosmide E8 de la souche de *S. ambofaciens* ATCC23877 digérée par *Sac*I sont utilisés comme marqueurs de taille.

La purification de l'ADN se fait par la méthode « Gene-clean » (Crosa *et al.* 1994). Les purifications des fragments « cassés » mécaniquement, effectués au Centre National de Séquençage, ont été réalisées par utilisation du kit NucleoSpin (Macherey Nagel) utilisé suivant les recommandations du fournisseur.

La purification d'une bande de taille déterminée sur gel d'agarose est réalisée grâce à l'insertion dans le gel d'électrophorèse d'une membrane de DEAE cellulose en amont de la bande choisie. Lorsque le champ électrique est de nouveau appliqué, la migration de la bande est arrêtée par la membrane. L'ADN est récupéré après lavage de la membrane (5 ml de tampon de lavage : Tris-HCl 50 mM pH 8, NaCl 0,15 M, EDTA 10 mM pH 8) et après deux étapes d'élution (30 min puis 15 min à 68°C) dans du tampon d'élution (Tris-HCl 50mM pH 8, NaCl 1M, EDTA 10 mM pH 8). L'ADN est ensuite précipité par ajout de 1/10 de volume d'acétate de sodium (3M, pH 5,2) et de deux volumes d'isopropanol, séché et dissous dans 20 µl d'eau ultra-pure.

L'activité polymérase 5'-3' de la Klenow est utilisée pour combler les extrémités cohésives. Plus spécifiquement, elle est employée ici pour muter un des sites *Bam*HI portés par le plasmide pIJ773. Cinquante ng de ce plasmide préalablement digéré partiellement par *Bam*HI sont incubés en présence de tampon 2 (Biolbas), 1 UE de Klenow (Roche) et 0,5 µL de dNTP (Biolabs, 10 nM) dans un volume final de 20 µL. Le mélange de réaction est incubé 30 min à température ambiante, puis l'enzyme est inactivée par chauffage 20 minutes à 75°C. Les 20 µL sont ensuite déposés pendant 20 min sur un filtre Millipore (diamètre : 13 mm ; pore : 0,025 µm) posé sur de l'eau UP pour éliminer les sels. Ces 20 µL sont ensuite récupérés et utilisés pour les réactions de ligation.

III- Transformation des souches d'E. coli

A- Ligation

Les ligations des produits de PCR dans pGEM®-T Easy sont réalisées avec le kit "2X Rapid Ligation" (Promega) selon les recommandations du fournisseur. Le ratio molaire insert:vecteur est de 3:1. Les mélanges réactionnels sont incubés à 4°C pendant 16 heures.

La DNA ligase (New England Biolabs) est utilisée pour les réactions de ligations dans les autres vecteurs de ce travail selon les recommandations du fournisseur. Le ratio molaire insert:vecteur est compris entre 3:1 et 10:1 pour chaque couple insert-vecteur. Les mélanges réactionnels sont incubés à 16°C pendant 16 heures.

*B- Electroporation des cellules d'*E. coli

Les électroporations sont réalisées avec l'appareil GenePulserII de BioRad calibré à 200 Ω, 25 µF et 2,5 kV. Les constantes de temps attendues varient entre 3,8 et 4,9 msec.

B-1 PREPARATION DE CELLULES COMPETENTES D'*E. COLI*

Une préculture d'*E. coli* est réalisée dans 3 ml de LB, additionnée ou non d'antibiotiques, sur la nuit à 37°C sous agitation à 250 rpm. Une culture est réalisée en inoculant 50 ml de LB avec la préculture (1% final). Les cellules sont mises en culture à 37°C sous agitation à 250 rpm jusqu'à l'obtention d'une DO_{600nm} comprise entre 0,4 et 0,6. Les cellules sont récupérées par centrifugation 10 minutes à 4 000 rpm, à 4°C, puis lavées deux fois avec un volume de glycérol 10%. Le culot est repris dans 100 µl de glycérol 10% et 50 µl de ce culot sont utilisés lors de l'électroporation. Les quantités de mélange vecteur-insert sont précisés dans les résultats. Après électroporation, les cellules sont reprises dans 1 ml de SOC et incubée à 37°C sous agitation à 250 rpm pendant une heure. Les cellules sont ensuite étalées sur milieu gélosé sélectif et incubées sur la nuit à 37°C. Les cellules obtenues sont enfin analysées par restriction enzymatique sur gel d'électrophorése d'agarose.

B-2 INTRODUCTION DES COSMIDES OU BAC DANS *E. COLI* BW25113/PIJ790 ET *E. COLI* BW25113/PKD20

Une préculture de 3 ml d'*E. coli* BW25113/pIJ790 et BW25113/pKD20 est réalisée en LB contenant 25 µg/ml de chloramphénicol et 50 µg/ml d'ampicilline respectivement, sur la nuit à 30°C sous agitation à 250 rpm. Une culture est réalisée en inoculant 50 ml de SOB (Hanahan 1983), supplémenté à 20 mM de $MgSO_4$ et 25 µg/ml de chloramphénicol pour le plasmide pIJ790 ou 50 µg/mL en présence du plasmide pKD20, avec la préculture (1% final). Les cellules sont placées à 30°C sous agitation à 250 rpm jusqu'à une DO_{600nm} comprise entre 0,4 et 0,6.

Les cellules sont récupérées par centrifugation à 4 000 rpm à 4°C pendant 10 minutes. Les cellules sont ensuite lavées deux fois avec un volume de glycérol 10% puis resuspendues dans un volume de 100 µL de glycérol à 10% froid. Un volume de 50 µl de cellules compétentes est électroporé avec 100 ng de cosmide (14C4, 25E1, D50 et D118) ou de BAC (pBeloBAC11, A8 et F5). Les cellules sont immédiatement reprises dans 1 ml de LB froid et incubées à 30°C sous agitation à 250 rpm.

Les cellules sont ensuite étalées sur LB solide contenant du chloramphénicol (25µg/mL) et de la kanamycine (50µg/mL) dans le cas des cosmides et d'ampicilline (50µg/mL) et de chloramphénicol (25µg/mL) dans le cas des BAC, et incubées à 30°C (température permissive pour les plasmides pIJ790 et pKD20) sur la nuit. Quelques clones résistants sont analysés pour la présence d'ADN extra-chromosomiques (plasmides et cosmide ou BAC) par profil de restriction enzymatique sur gel d'électrophorèse.

B-3 CONSTRUCTION DES COSMIDES MODIFIES

Les souches d'*E. coli* BW25113/pIJ790 contenant les cosmides et *E. coli* BW25113/pKD20 contenant les BAC sont transformées par électroporation (protocole analogue à celui décrit précédemment) avec 100 ng de produit de PCR correspondant aux cassettes de sélection, avec les sites *parS* avec la cassette Apra (50 µg/mL) ou le site *attB* du phage ΦC31 avec la cassette Spec (50 µg/mL) ou 100 ng de produit de PCR correspondant au gène *aac3(IV)* de résistance à l'apramycine (50 µg/mL), l'origine de transfert de RK2, ainsi que le site d'attachement *attP* et le gène codant l'intégrase du phage ΦC31 (Bierman *et al.* 1992) du plasmide pSET152 pour la modification des BAC. Pour induire le système λRED, 10 mM final de L-arabinose est ajouté à la culture. Pour augmenter le taux de réussite de l'électroporation, les produits de PCR sont au préalable débarrassés des sels par dépôts sur un filtre Millipore (diamètre : 13 mm ; pore : 0,025 µm). Suite à l'électroporation, les cellules sont incubées à 37°C pour provoquer la perte des plasmides pIJ790 et pKD20. L'ADN cosmidique ou les BAC sont contrôlés par PCR et par profil de restriction sur gel d'électrophorèse.

IV- Conjugaison intergénérique E. coli-S. ambofaciens

Les souches sauvages de *S. ambofaciens* sont modifiées par conjugaison intergénérique avec *E. coli* ET12567/pUZ8002 contenant les cosmides ou les BAC modifiés (voir III B-C). Le plasmide pUZ8002 possède les fonctions de transfert nécessaires à la conjugaison mais n'est pas mobilisable puisqu'il ne porte pas d'origine de transfert (Paget *et al.* 1999).

Pour chaque construction, la souche d'*E. coli*/pUZ8002/cosmide ou BAC modifié est amenée à une DO_{600nm} comprise entre 0,4 et 0,6 (culture LB avec 25 µg/mL de chloramphénicol, 50 µg/mL de kanamycine et 50 µg/mL d'apramycine ou 50 µg/mL de spectinomycine à 37°C à 250 rpm). Tandis que ces cellules sont lavées deux fois avec un volume de LB pour éliminer les antibiotiques qui inhiberaient la croissance des *Streptomyces* puis reprises dans 1/10 de volume initial de LB, environ 10^8 spores de *S. ambofaciens* pour chaque conjugaison sont diluées dans 0,5 ml de milieu de prégermination (2xYT (Kieser *et al.* 2000)) et placées à 50°C pendant 10 minutes (choc thermique permettant la prégermination des spores).

Un mélange de 0,5 mL de cellules compétentes d'*E. coli* et 0,5 mL de spores est réalisé et des suspensions/dilutions de ce mélange sont étalées sur milieu SFM ($MgCl_2$ 10 mM) sans pression de sélection. Après incubation à 30°C de 16h à 20h, les boîtes sont recouvertes par 1 ml d'eau stérile, contenant 0,5 mg d'acide nalidixique (contre-sélection d'*E. coli*) et 1,25 mg d'apramycine ou de spectinomycine (sélection de l'intégration de la cassette dans le chromosome de *S. ambofaciens*), puis de nouveau incubées à 30°C.

Lorsque des colonies de *Streptomyces* sont visibles et sporulantes (après environ une semaine d'incubation), des répliques sont effectuées sur SFM avec acide nalidixique (25 µg/ml), apramycine (50 µg/ml) ou spectinomycine (200µg/ml) avec ou sans kanamycine (50 µg/ml). Les ex-conjugants apparus par double *crossing-over* sont sensibles à la kanamycine [Kan^S] et résistants à l'apramycine [$Apra^R$] ou à la spectinomycine [$Spec^R$]. Les clones positifs, à savoir [$Apra^R$][Kan^S] ou [$Spec^R$][Kan^S] sont sous-clonés de façon à obtenir une population homogène et

analysé par PCR.

V- Hybridation moléculaire

L'ADN constituant la sonde est marqué par la méthode des amorces aléatoires par incorporation de DIG-dUTP (déoxyUridine TriPhosphate couplée à la digoxygénine) à l'aide du kit "DIG DNA Labeling and Detection " (Roche).

Après migration sur gel d'agarose, le transfert de l'ADN est réalisé sur membrane de Nylon (type N, Amersham) à l'aide du système Vacugene (Pharmacia-LKB). Une étape de dépurination (HCl 0,25 N, 20 min) est suivie d'une étape de dénaturation (NaOH 0,5 N ; NaCl 1,5 M, 20 minutes) puis d'une étape de transfert (solution SSC 20X : NaCl 3 M, citrate de Na 0,3 M, 1 heure). La membrane est rincée dans une solution de SSC 2X, puis séchée. L'ADN est fixé à la membrane par exposition à un rayonnement UV (λ=254 nm) pendant 25 secondes.

Les étapes de pré-hybridation, hybridation, lavages des membranes ainsi que l'étape de révélation sont réalisées d'après les instructions du kit "DIG DNA Labeling and Detection" (Roche). Les signaux sont détectés grâce à une caméra CCD (Fluor'S max, Biorad) couplée à une carte d'acquisition de données gérée par le logiciel Quantity One^{TM} (Biorad).

VI- Réaction de séquençage

Le séquençage est réalisé sur la plate-forme génomique de l'IFR 110 selon la méthode de Sanger (1977). Les produits d'élongation sont séparés par électrophorèse capillaire en gel de poly N,N-diméthylacrylamide et analysés automatiquement à l'aide du système Ceq2000XL (Beckman Coulter).

Le séquençage à haut débit des banques plasmidiques, issues des cosmides recouvrant les TIR de la souche de *S. ambofaciens* DSM40697, a été réalisé au sein du Centre National de Séquençage (Génoscope, Ivry). L'obtention et l'analyse des séquences obtenues sont réalisées par le système ABI Prism 300 (Applied Biosystems). La détection du signal est effectuée grâce à l'incorporation de didéoxynucléotides couplés à une molécule fluorescente.

Les séquences obtenues sont analysées par le logiciel Sequencher (http://genecodes.com/sequencher/). Les programmes BLAST (Basic Local Alignment Search Tool) ont été utilisés pour rechercher les pourcentages d'identité avec les séquences nucléotidiques ou protéique des banques de données via le National Center for Biotechnology Information (http://www.ncbi.nih.gov/). Le logiciel CodonCode Aligner a été utilisé pour réaliser les contigs de séquences. L'alignement de séquence est réalisé via l'utilisation du logiciel BioEdit. Le programme ClustalW a permis de détecter les homologies en nucléotides ou acides aminés entre les séquences ou produits de gènes.

L'annotation des séquences des TIR s'est faite au sein de la plate-forme génomique du laboratoire, mise en place par Frédéric CHOULET.

VII- Stratégie de remplacement ou d'incorporation au locus chromosomique chez Streptomyces

La stratégie de « PCR-Targeting » se divise en trois étapes majeures. La différence est qu'une étape préalable permet le clonage de plusieurs site *parS* ou du site *attB* du phage ΦC31 à proximité direct de la cassette des plasmides pIJ773 et pIJ778. Toutes les autres étapes sont parfaitement identiques à celles décrites.

A- Choix des amorces et fabrication du produit de PCR

Les sites choisis, pour recevoir les sites *parS* ou *attB* de ΦC31, sont situés sur des cosmides recombinants (dérive de Supercos1, (Evans *et al.* 1989)) ou pour recevoir le système conjugaison-intégration ainsi que le gène de résistance à l'apramycine du plasmide pSET152, sur les BAC recombinants dérivant de pBeloBAC11 (New England Biolabs). Deux amorces de grande taille (58 nucléotides) sont nécessaires. Chacune de ces amorces se caractérise par deux régions distinctes : une séquence de 39 nucléotides, situés en 5' de l'amorce, est complémentaire des régions adjacentes du locus d'intégration ou du gène à remplacer, et en 3', une séquence de 19 ou 20 nucléotides est complémentaire des extrémités 5' et 3' de la cassette, englobant les éléments génétiques clonés. Ainsi, la cassette peut être orientée en combinant au sein des amorces les séquences homologues à la cassette et au locus considéré.

Outre le marqueur de résistance et les éléments génétiques clonés, ces cassettes comprennent l'origine de transfert *oriT* du plasmide RK2 (Pansegrau *et al.* 1994) et deux séquences FRT (FLP Recognition Target), sites de reconnaissance pour la recombinase site-spécifique FLP. La présence des sites FRT permet, après remplacement de l'ORF ou après l'incorporation des éléments génétiques étudiés, l'excision par la protéine FLP de la partie centrale de la cassette, les éléments génétiques étant clonés en dehors des sites FRT.

B- Construction de l'allèle modifié sur le cosmide recombinant

Le cosmide ou le BAC recombinant recouvrant le locus d'intérêt est introduit par électroporation dans respectivement *E. coli* BW25113/pIJ790 ou *E. coli* BW25113/pKD20. Les plasmides pIJ790 et pKD20 possèdent le système λRED et un gène conférant la résistance respectivement au chloramphénicol et à l'ampicilline. Ce système comporte trois gènes de recombinaison du bactériophage λ, *exo*, *bet* et *gam* (Murphy 1998). La protéine Gam inhibe l'exonucléase V du système RecBCD de la souche hôte, alors que les protéines Bet et Exo permettent l'accès aux extrémités des fragments d'ADN pour promouvoir la recombinaison. Ce système présente un taux de recombinaison 15 à 130 fois supérieurs que des souches *recBC* ou *recD* (Murphy 1998).

Ce système est placé sous le contrôle d'un promoteur inductible par le L-arabinose. Après induction, la souche d'*E. coli* BW25113/pIJ790 ou pKD20/cosmide ou BAC d'intérêt est transformée le produit de PCR. Les transformants sont ensuite sélectionnés pour les résistances portées par le cosmide ou le BAC (kanamycine et chloramphénicol respectivement) et par la

cassette (apramycine ou spectinomycine). A cette étape, la température de culture d'*E. coli* est amenée à 37°C provoquant la perte des plasmides pIJ790 et pKD20, leur réplication étant initiée à partir d'une origine de réplication thermosensible. Après sélection des recombinants, le cosmide ou le BAC modifié est électroporé dans la souche non méthylante *E. coli* ET12567/pUZ8002. Les transformants sont sélectionnés de façon analogue à la sélection des transformants de la souche d'*E. coli* BW25113 modifiée. Cette étape permet de contourner les éventuels systèmes de restriction-modification présents chez certaines espèces de *Streptomyces*.

*C- Transfert conjugatif du cosmide ou du BAC modifié d'*E. coli *vers* Streptomyces *et sélection des recombinants.*

La technique de conjugaison intergénérique *E. coli- Streptomyces* est plus efficace que la transformation de protoplastes. Cette conjugaison est décrite dans le chapitre IV de cette section. Le cosmide ou le BAC ne possédant pas d'origine de réplication lui permettant de se maintenir de manière autonome chez *Streptomyces*, la résistance conférée par ce vecteur ne peut être active par intégration dans le chromosome. La recombinaison entre le cosmide ou le BAC modifié et le chromosome est favorisée par la longueur d'homologie entre les régions flanquant la cassette et le locus d'intérêt, longueur variant de quelques kilobases à une vingtaine de kilobases.

L'intégration de la cassette de résistance au sein du chromosome étant stable, la pression de sélection peut être levée. Ceci permet la comparaison des souches sauvages et mutantes dans les mêmes conditions expérimentales.

VIII- Réaction de Polymérisation en Chaîne (PCR)

Deux polymérases sont utilisées : la *Taq* polymérase (Sigma ou Biolabs) et la LA *Taq* polymérase (Takara). L'enzyme *Taq* polymérase de Sigma ou de Biolabs a été utilisée pour les PCR de routine, cette enzyme ne présentant pas d'activité 3'→ 5' exonucléasique (activité lecture d'épreuve). L'enzyme LA *Taq* polymérase est pourvue d'une activité lecture d'épreuve, ce qui augmente considérablement la fidélité de la réaction d'amplification. Cette enzyme est donc choisie pour amplifier les régions d'ADN à séquencer. Les amorces utilisées sont fournies par Invitrogen ou Eurogentec. Elles sont choisies pour assurer une composition en bases G-C équivalente et garantir des températures de fusion équivalentes. L'utilisation du logiciel DNA Amplify 2.5 (Université du Wisconsin) ou de Primer3 Input (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) permet de valider le choix des amorces : absence de dimérisation des amorces et de formation de structures secondaires.

Le contrôle négatif consiste en une réaction dans laquelle la matrice ADN est remplacée par un volume équivalent d'eau. Les produits de PCR sont purifiés après migration sur gel d'agarose à l'aide du kit "High pure PCR product purification" (Roche).

RESULTATS

RESULTATS

Les régions terminales du chromosome des *Streptomyces* sont le siège d'une instabilité importante. La corrélation entre instabilité phénotypique des *Streptomyces* et instabilité génomique a été étudiée depuis plusieurs décennies, et a révélé le caractère non essentiel des régions affectées. L'avènement des méthodologies de séquençage à haut débit permet actuellement de réaliser la génomique comparée aux niveaux interspécifique et intraspécifique. Ainsi la comparaison des séquences génomiques de *S. coelicolor* (Bentley *et al.* 2002) et *S. avermitilis* (Ikeda *et al.* 2003) a démontré la compartimentation génétique du génome des *Streptomyces*, avec une région centrale conservée contenant les gènes nécessaires à la croissance végétative et des régions terminales regroupant des gènes « accessoires » ou de « contingence ».

Au laboratoire, un programme de séquençage des régions terminales chez la souche de *S. ambofaciens* ATCC23877 a été initié avec le Génoscope (Centre National de Séquençage, Evry) et l'Institut de Génétique et Microbiologie (J.-L. Pernodet, Evry). Le séquençage des TIR (Terminal Inverted Repeats) d'une autre souche de *S. ambofaciens*, DSM40697, a été mené pour évaluer la variabilité terminale au niveau intraspécifique.

La comparaison des génomes entre espèces bactériennes a permis de montrer la présence de nombreux réarrangements (inversion, délétions…) qui ont été fixés au cours de l'évolution. L'étude de la formation des réarrangements permet d'aborder les mécanismes responsables de la plasticité chromosomique. Ainsi, l'étude de la structure et de la stabilité chromosomique chez *S. ambofaciens* a permis de suggérer un lien entre partition chromosomique et stabilité. Ainsi un phénomène analogue au cycle CFP affectant les chromosomes dicentriques des cellules eucaryotes a été préalablement caractérisé au laboratoire (Wenner *et al.* 2003). Ainsi, parallèlement à l'analyse des événements chromosomiques fixés au cours de l'évolution des souches et des espèces, mon travail a consisté à tester l'hypothèse soulevée en construisant des souches dont le chromosome mime la présence de deux centromères bactériens.

I- Séquençage des Répétitions Terminales Inversées de la souche de S. ambofaciens DSM40697

Préalablement à cette étude, une première approche a été réalisée à partir de sondes couvrant certaines régions des Répétitions Terminales Inversées (TIR) et a montré que la souche DSM40697 présentait une région spécifique d'environ 51 kb aux extrémités par rapport à la souche de *S. ambofaciens* ATCC23877. De même, la souche ATCC23877 possède environ 33 kb à l'extrémité des TIR qui lui sont propres, le reste étant commun entre les deux souches. Pour déterminer la variabilité entre ces deux régions au niveau nucléotidique, les TIR ont été séquencées. Les séquences des TIR de la souche ATCC23877 étant comprises dans le projet de séquençage des régions terminales, leurs séquences étaient déjà disponibles, à l'exception du fragment terminal, portant les télomères.

Une banque cosmidique, recouvrant la totalité des TIR de la souche de *S. ambofaciens* DSM40697 était déjà disponible au laboratoire. La figure 25 représente l'alignement des différents cosmides utilisés lors de cette étude le long des TIR de la souche de *S. ambofaciens* DSM40697. Deux approches ont été utilisées pour obtenir une banque plasmidique de chacun des cosmides en vue de leur séquençage puis assemblage.

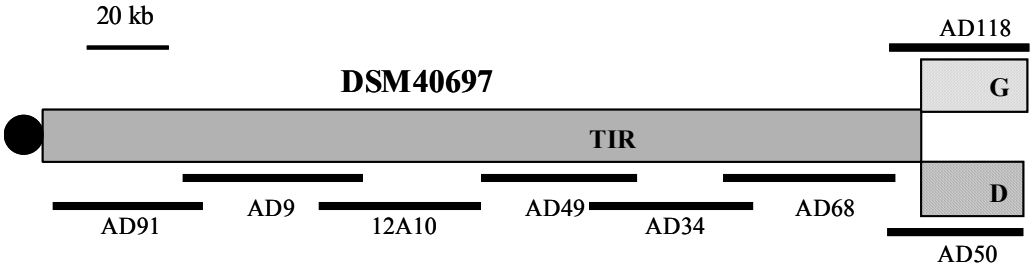


Figure 25 : Alignement des cosmides recombinants le long des Répétitions Terminales Inversées (TIR) de la souche de *S. ambofaciens* DSM40697
Les régions hachurées correspondent aux régions adjacentes aux TIR. G : Bras chromosomique gauche ; D : Bras chromosomique droit.

A- Clonage dans le vecteur M13

La première approche est la digestion des ADN par une enzyme de restriction aux sites de reconnaissance fréquents. Le choix s'est porté sur l'enzyme *Sau*3A (5'-GATC-3') dont la fréquence de coupure est de 1/256.

Les réactions de digestion sont partielles (temps d'incubation ou quantité d'enzyme utilisée contrôlés). Ainsi, de nombreux fragments de restriction de tailles différentes contiendront plusieurs sites de restriction. L'alignement des séquences de ces fragments permettra un assemblage facilité pour former un seul contig.

La taille des fragments de restriction retenue pour les clonages est de 1 à 1,4 kb. Ces fragments sont extraits d'un gel d'électrophorèse à l'aide d'une membrane de DEAE cellulose (Fig. 26).

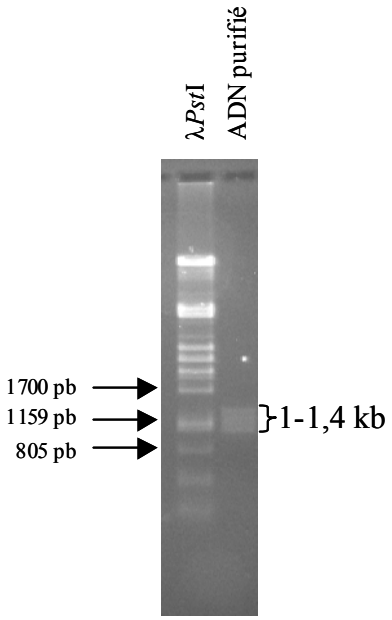


Figure 26 : Migration des fragments de digestion partielle *Sau*3A de l'ADN du cosmide AD91 purifiés par DEAE cellulose (Gel agarose 0,8% TAE 1X à 5 V/cm)

Le vecteur de clonage choisi dans cette approche est le phage M13, qui permet la synthèse d'un ADN sous forme simple brin, présentant moins de formation de structure secondaire ce qui permet à la polymérase une processivité plus importante sans arrêt de progression le long de l'ADN et donc d'obtenir un meilleur rendement. Le vecteur est digéré par *Bam*HI puis déphosphorylé.

Pour valider une banque, 24 clones sont analysés. L'analyse de l'ADN plasmidique de clones par digestion enzymatique *Eco*RI-*Hin*dIII et migration électrophorétique a montré que la majorité (19/24) présente un insert de taille moyenne estimée conforme à la taille souhaitée.

Le surnageant de culture de 20 clones est traité suivant la méthode décrite par Bankier (1993) puis l'ADN simple brin purifié a été séquencé. L'analyse des séquences obtenues a confirmé la présence de plusieurs sites de restriction *Sau*3A au sein de chaque séquence.

Enfin, le nombre de réactions de séquençage pour chaque cosmide a été déterminé afin d'atteindre une probabilité de couverture de 99% pour chaque position nucléotidique. Ainsi, pour le cosmide AD91 dont l'insert mesure 33,7 kb, 289 réactions de séquençage seront nécessaires pour un taux de recouvrement de 4,5 fois. Après l'obtention et l'assemblage d'environ 200 séquences, 7 contigs ont été obtenus. Un contig regroupant plus de 60% des séquences obtenues indique un fort biais de représentativité. Ce biais découle probablement de l'utilisation du vecteur M13. Une approche distincte a été développée en collaboration avec le Génoscope.

B- Clonage dans pCNS

La seconde approche est la cassure de l'ADN de façon mécanique par Hydro-Shear (GeneMachines).

Cette banque a été réalisée au sein de l'équipe de Béatrice SEGURENS au Génoscope (CNS, Evry) responsable également de la réalisation des banques plasmidiques des régions variables de la souche ATCC23877. En raison de la similarité des séquences entre les deux souches (% en G-C moyen…), une approche analogue à celle utilisée pour la souche ATCC23887 a été choisie.

Ainsi, un plasmide (pCNS) présentant un faible nombre de copies par cellule (ici 20 copies/cellule) et une réplication de type θ a été utilisé. De plus, ce plasmide possède un site multiple de clonage dans le gène *lacZ* de l'enzyme β-galactosidase, permettant de distinguer les plasmides natifs des plasmides recombinants.

Après cassure, des fragments de taille aléatoire, dont les extrémités sont soit franches, soit cohésives en 5' ou en 3', sont obtenus. Pour pouvoir cloner l'ensemble de ces fragments, leurs extrémités sont réparées par l'action de la T4 Polymérase en présence de dNTP. Cette dernière va combler les extrémités en synthétisant l'ADN complémentaire au brin cohésif 3'. Dans le cas des extrémités cohésives en 5', l'activité exonucléasique de l'enzyme permettra la dégradation de cette extrémité. Ainsi, la grande majorité des fragments présenteront des extrémités franches.

Puis, des adaptateurs (ici *Bst*XI) sont ligués à ces fragments avec la T4 DNA ligase (New England Biolabs). L'ensemble des fragments, ayant un adaptateur à chacune de ces extrémités, présentera des extrémités similaires et sera cloné dans le même site au sein du plasmide utilisé. Des fragments de 3 kb sont ensuite extraits d'un gel par « Gene-Clean ». Le gel est immergé dans un bain contenant du Sybr Green, permettant la visualisation de faible quantité (à partir de 5 ng) et ne provoquant aucune mutation.

Trois extractions pour des tailles englobant les 3 kb sont réalisées. L'exemple présenté ici est le dosage des fragments des cosmides AD91 et AD9 (Fig. 27). Un mélange de ligation contenant le vecteur pCNS, et les fragments purifiés est réalisé avec la T4 DNA ligase (New England Biolabs). Les transformants issus de ce mélange de ligation sont sélectionnés sur LB gélosé additionné de chloramphénicol, de X-Gal et d'IPTG. A titre d'exemple, pour la cosmide AD9, 1,9% des colonies étaient bleues pour 98,1% de colonies blanches.

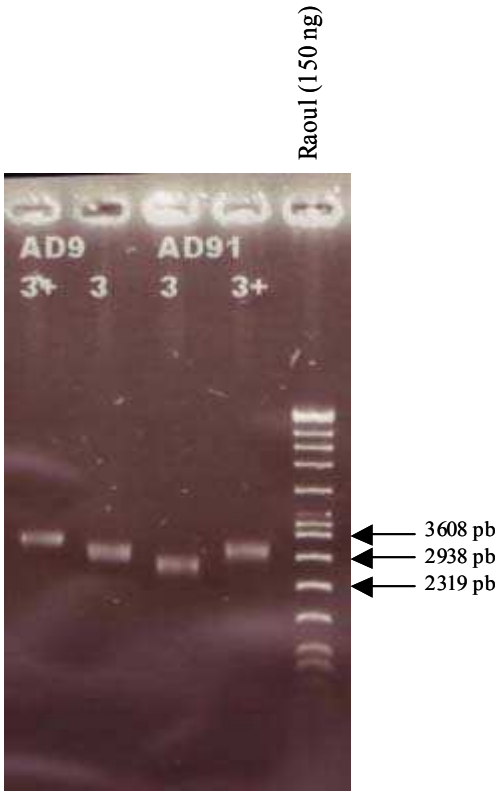


Figure 27 : Gel de vérification des extractions sur gel pour les cosmides AD91 et AD9.
Dépôt de 1/25 du volume de solution de chaque fragment purifié. Gel agarose 0,7% TBE 1X à 10 V/cm.

Les réactions de séquence nécessitent une quantité de matrice définie qui est dépendante de la taille de cette dernière. Pour pouvoir séquencer un grand nombre d'échantillons sans quantifier chacun de ces échantillons, la banque plasmidique doit être homogène en taille. Les purifications et les mélanges réactionnels seront standardisés. Pour cela, 16 clones, pour chaque banque plasmidique réalisée, seront vérifiés par analyse électrophorétique. L'ADN plasmidique natif de ces clones est déposé. Ceci permet de vérifier l'homogénéité de la taille des plasmides recombinants.

A partir des banques plasmidiques obtenues pour chacun des cosmides, 384 séquences pour les banques 12A10, AD49, AD34, AD68, D50 et D118 et 768 séquences pour les banques des cosmides AD91 et AD9 ont été réalisées. Ces derniers constituent les bancs d'essais de notre programme de séquençage.

C- *Finition des séquences*

La combinaison des séquences réalisées à partir des deux banques plasmidiques a permis d'obtenir 18 « contigs » dont les tailles s'échelonnaient de 544 pb dans le cosmide AD68 à 22085 pb dans le cosmide D50. La taille des contigs obtenus est en moyenne de 6 kb. Ainsi 17 lacunes ou « gaps » de séquences ainsi que 5 incertitudes sur certaines positions nucléotidiques restaient à résoudre. Le séquençage des fragments terminaux des souches ATCC23877 et DSM40697, contenant les télomères, non représentés dans les banques cosmidiques, a été abordé. La principale difficulté est l'obtention des derniers nucléotides formant le télomère, le séquençage direct sur le

chromosome étant difficilement réalisable.

C-1 METHODE DE LISSAGE DES SEQUENCES

Deux incertitudes sont situées dans les cosmides D50 et D118, trois dans les cosmides AD91 et AD49, quatre dans le cosmide AD68 et une dans les cosmides 12A10 et AD34 ainsi que la jonction entre les cosmides 12A10-AD34.

L'amplification par PCR des régions, contenant les incertitudes et les lacunes, et le séquençage de ces produits d'amplification ont été réalisés.

Le choix des amorces, pour la synthèse de ces produits de PCR, a été guidé par l'estimation de la taille des lacunes. Pour estimer cette taille, deux approches ont été utilisées :

- Dans les régions fortement homologues de la séquence de la souche ATCC23877, la taille des lacunes a été estimée sur la base de la conservation de la séquence.

- Dans les régions spécifiques de la souche DSM40697, l'estimation est réalisée sur la base de la taille du fragment de restriction *Bam*HI portant la lacune. L'analyse des séquences pour chaque cosmide permet de positionner les sites de restriction le long de l'insert et de déterminer ainsi la taille du fragment où se situe la lacune. Connaissant le nombre de nucléotides séquencés, la taille de la lacune est déduite.

Des produits d'amplification sont séquencés dans différentes conditions et avec différentes « chimies » de séquençage : kit de séquençage CEQ Dye Terminator (Beckman Coulter), kit de séquençage BigDye Terminator (Applied Biosystems) avec des matrices double brin, correspondant aux produits d'amplifications.

Le chromatogramme obtenu lors du séquençage de certains produits de PCR présente un arrêt brutal de la réaction de séquençage, le plus souvent associé à la présence d'une structure secondaire potentielle sur l'ADN matrice. Ces structures correspondent le plus souvent à la présence d'un terminateur de transcription (Fig. 28).

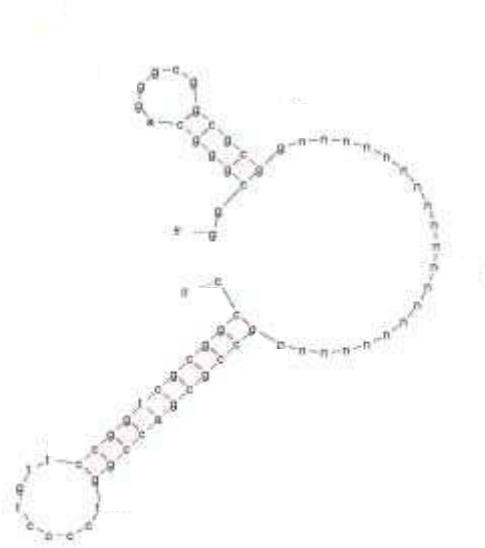


Figure 28 : Structure secondaire potentielle rencontrée dans le cosmide AD91
La lettre n représente les nucléotides non déterminés. Cette structure putative a été obtenue avec DNA *mfold* 3.1 (NaCl 1M, 30°C).

Pallier ce problème :

- L'ADN simple brin, utilisé comme matrice de séquençage, facilite le passage de l'ADN polymérase dans les régions contenant des structures secondaires. Les produits de PCR sont clonés dans un plasmide (ici pGEM-T easy, Promega) permettant la synthèse de cet ADN simple brin grâce à l'utilisation d'un phage « helper ».

- Une autre solution est de déstabiliser ces structures secondaires. Pour cela, des agents chimiques chaotropiques, tels que le DMSO (Diméthyl Sulfoxide), sont additionnés aux mélanges réactionnels lors des réactions de séquençage. Le DMSO est utilisé pour empêcher la formation de structures secondaires par interaction avec l'ADN.

C-2 SEQUENÇAGE DU FRAGMENT TERMINAL

Pour séquencer les fragments terminaux des deux souches de *S. ambofaciens*, une approche par amplification a été retenue. Pour cela, une amorce est choisie dans les séquences télomériques conservées et la seconde dans les séquences déterminées des cosmides AD91 et C7 pour respectivement les souches de *S. ambofaciens* DSM40607 et ATCC23877.

La comparaison des séquences télomériques montre une forte homologie entre les réplicons linéaires des *Streptomyces*. Par exemple, les séquences télomériques de *S. coelicolor* et *S. lividans* présentent 92% d'identité sur les 166 premiers nucléotides (153/166).

Ainsi, le choix de l'amorce, nommée T1, repose sur l'hypothèse de la forte conservation des séquences télomériques entre souches et espèces proches (Fig. 29). *S. coelicolor*, *S. lividans* et *S. ambofaciens* sont phylogénétiquement étroitement apparentées (reconstruction phylogénétique basée sur les ARN 16S et disponible à http://avermitilis.ls.kitasato-u.ac.jp/tree.html). Cette amorce a été dessinée à partir de la séquence télomérique de *S. coelicolor* et chevauche partiellement les séquences palindromiques I et II, évitant théoriquement la formation d'une structure tige boucle (Fig. 30).

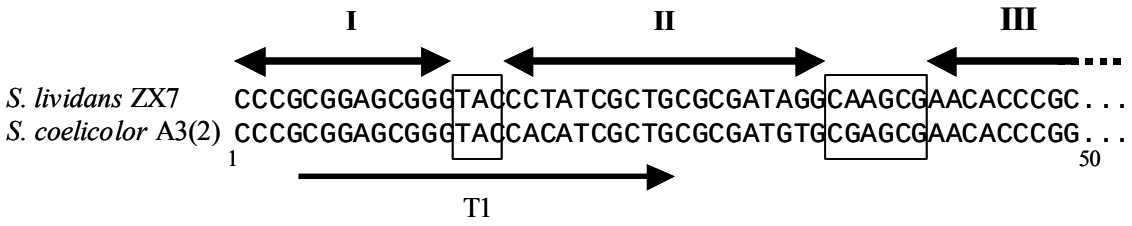


Figure 29 : Alignements des 50 derniers nucléotides des télomères de *S. lividans* ZX7 et de *S. coelicolor* A3(2) et la position de l'amorce T1

Les chiffres romains et les doubles flèches représentent les séquences palindromiques. Les cadres délimitent les séquences séparant les différents palindromes.

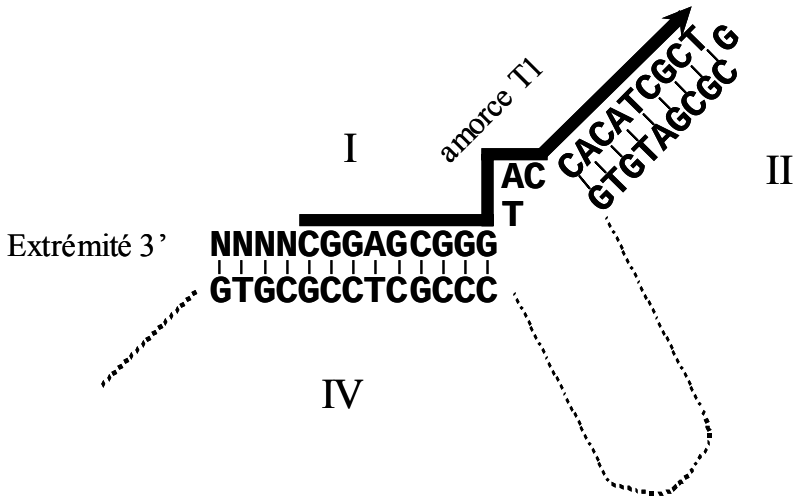


Figure 30 : Localisation de l'amorce T1 dans la structure secondaire des télomères de *S. ambofaciens*

Les chiffres romains correspondent aux séquences palindromiques. Seuls les nucléotides appareillés avec l'amorce sont représentés.

C-2a Régions terminales de la souche de *S. ambofaciens* DSM40607

Une région, dont la taille est estimée à 3 kb, n'est pas représentée dans la banque cosmidique. Une amplification de cette région par PCR a été réalisée (Fig. 31). La séquence de l'amorce T1 a été choisie à partir de la séquence télomérique de *S. coelicolor* et la seconde amorce A1 à partir des premières séquences du cosmide AD91 et est située à 162 pb à l'intérieur du cosmide.

Les réactions de séquençage sont réalisées avec les amorces A1 et T1 et de nouvelles amorces sont choisies à partir des séquences obtenues.

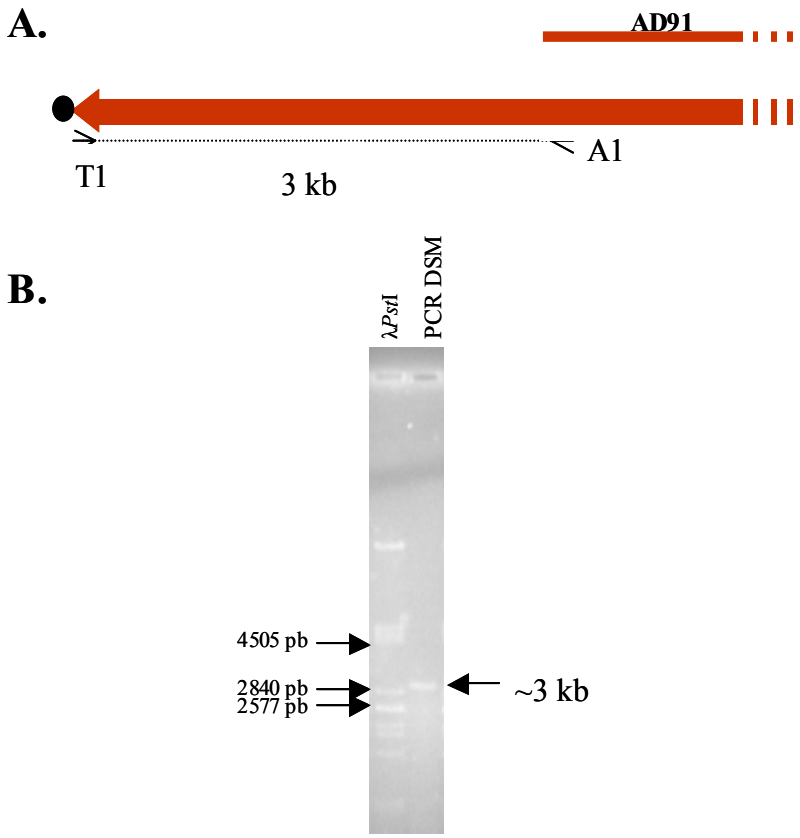


Figure 31 : Amplification du fragment terminal du chromosome de *S. ambofaciens* DSM40697 par le couple d'amorce T1-A1
A. Extrémité du chromosome de *S. ambofaciens* DSM40697.
B. Migration électrophorétique du produit de PCR. Gel agarose 0,8% TAE 1X à 3V/cm.

C-2b Régions terminales de la souche de *S. ambofaciens* ATCC23877

Une région, dont la taille est estimée à 2,7 kb, est absente des banques cosmidiques pour les mêmes raisons que le fragment terminal de la souche DSM40697. Cette région recouvre la région subtélomérique et les télomères de la souche ATCC23877. L'amorce T1 est celle utilisée lors de la réaction d'amplification de la région subtélomérique de la souche DSM40697. La seconde amorce, C7, a été choisie à partir des séquences du cosmide C7 de la banque.

Des échantillons d'ADN génomique "-" (non-traité à la protéinase K) ont été préparés. L'absence de traitement à la protéinase K permet de préserver la liaison covalente entre la protéine terminale et l'extrémité du brin 5'. Cette protéine empêche la migration électrophorétique du fragment d'ADN auquel elle est liée. La sonde constituée par le produit de PCR de 2,7 kb a été hybridée sur des profils *Bam*HI et *Bcl*I d'ADN génomique "+" (traité à la protéinase K) et "–" (Fig. 32 A). Des signaux sont observés uniquement dans les pistes "+", et correspondent respectivement à des fragments *Bam*HI et *Bcl*I de 2,7 kb et 9 kb, fragments préalablement identifiés comme les fragments terminaux libérés par ces deux endonucléases (Borges 2001). Le fragment *Bam*HI (2,7 kb) révélé dans l'ADN "+" comigre avec la bande obtenue par digestion *Bam*HI du produit de PCR de 2,8 kb (Fig. 32 B). Ces résultats montrent que les séquences amplifiées correspondent aux régions terminales non représentées dans la banque de cosmides

recombinants.

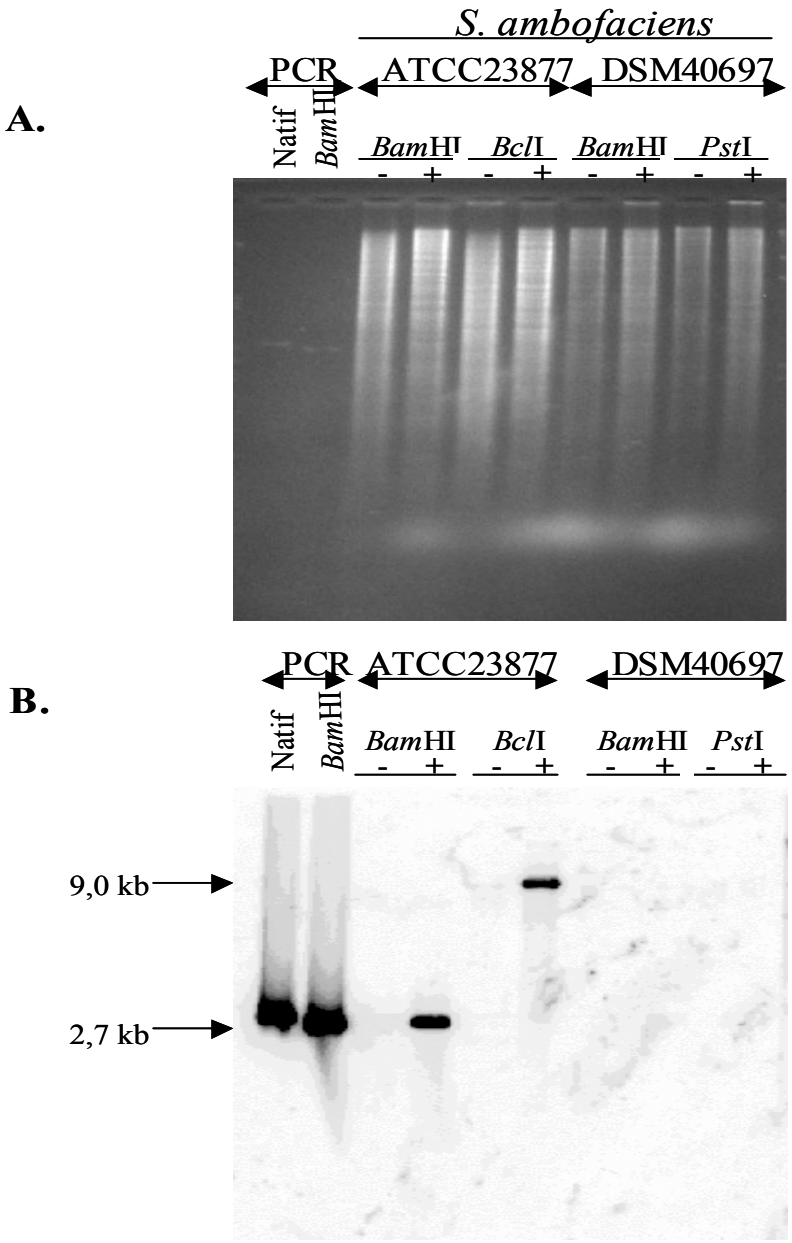


Figure 32 : Hybridation du produit de PCR (couple d'amorces T1-C7) sur le fragment terminal du chromosome de *S. ambofaciens* ATCC23877

A. Migration électrophorétique des profils de restriction *Bam*HI et *Bcl*I de l'ADN total de *S. ambofaciens* ATCC23877 et *Bam*HI et *Pst*I de l'ADN total de *S. ambofaciens* DSM40697.
Gel agarose 0,8% TAE 1X à 2 V/cm.
B. Hybridation du produit de PCR marqué sur les profils de restriction de l'ADN génomique de *S. ambofaciens* ATCC23877 et DSM40697.

De façon analogue au fragment de la souche DSM40697, une séquence de 1,56 kb a ainsi pu être déterminée à partir du produit de PCR. En revanche, le séquençage à partir de l'amorce T1 montre un arrêt à la position nucléotidique 193 (position définie par rapport au premier nucléotide de T1 ; Figure 33).

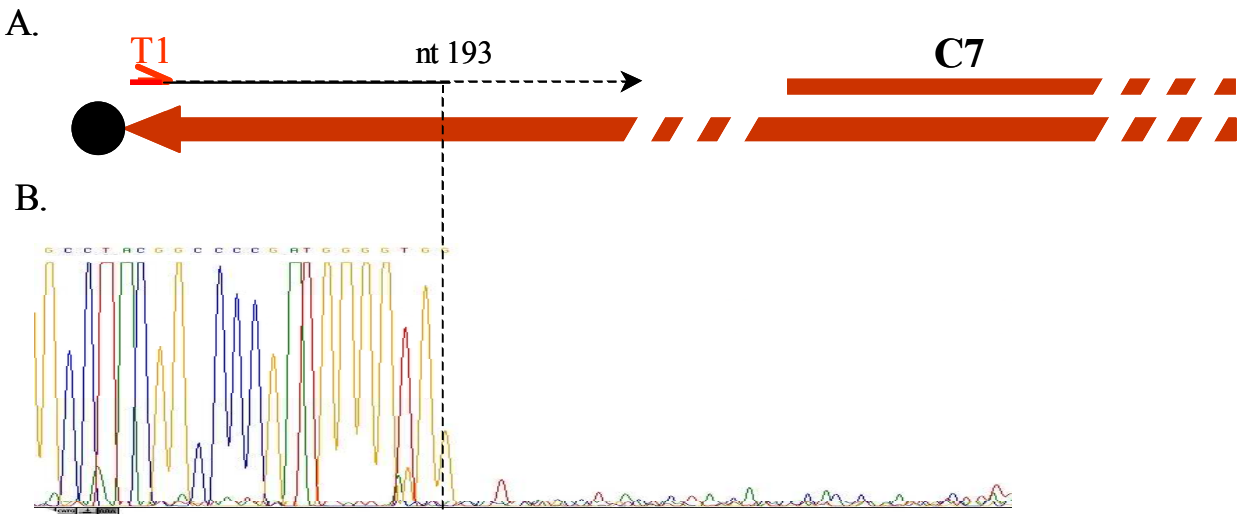


Figure 33 : Electrophorégramme obtenu à partir de l'amorce T1
A. Schéma de la région terminale du chromosome de *S. ambofaciens* ATCC23877. Seule la partie d'intérêt est présentée.
B. Electrophorégramme de la réaction de séquençage du produit de PCR de 2,8 kb à partir de l'amorce T1.

Cet arrêt de la synthèse d'ADN a été observé de manière reproductible quelle que soit la matrice utilisée (issue de différentes réactions de PCR), et quelles que soient les amorces utilisées, bien que ces dernières permettent l'obtention de produit de PCR de tailles attendues à partir de l'ADN génomique de la souche ATCC23877.

Pour confirmer certaines positions nucléotidiques et tenter de résoudre ce problème de structure secondaire, les produits de PCR ont été clonés dans le plasmide pGEM-T® easy (Promega). A partir d'un clone contenant le plasmide recombinant, un ADN simple brin est extrait et utilisé comme matrice lors des réactions de séquençage. La synthèse et l'utilisation de cette matrice simple brin a permis l'obtention des séquences de l'ensemble de la région subtélomérique des deux souches de *S. ambofaciens*. Dans le cas de la souche de *S. ambofaciens* ATCC23877, les séquences ont permis de prédire la structure secondaire responsable de l'arrêt de réaction de séquençage.

D- Bilan du séquençage

La compilation des séquences obtenues par ces deux approches a permis d'obtenir la quasi-totalité des séquences des TIR avec un taux de recouvrement de 10 fois. Ainsi deux contigs de 238 517 pb pour le bras gauche et de 237 382 pb pour le bras droit ont été obtenus.

Sur 17 incertitudes, seules quatre n'ont pas été résolues. Bien que les séquences télomériques présentent une forte similarité avec celles de *S. coelicolor*, il n'est pas possible de déterminer la séquence correspondante à l'amorce T1, cette dernière ayant été dessinée à partir de la séquence de *S. coelicolor*. De même, il n'est pas possible de déterminer le nombre et la nature des nucléotides localisés en 5' de la séquence de l'amorce. En se basant sur la longueur des télomères de *S. coelicolor*, seuls 4 nucléotides resteraient indéterminer en 5'.

La figure 34 résume l'ensemble de la localisation de ces différents cosmides, des lacunes de séquences et des produits de PCR pour le fragment terminal et la jonction des cosmides 12A10-

AD49.

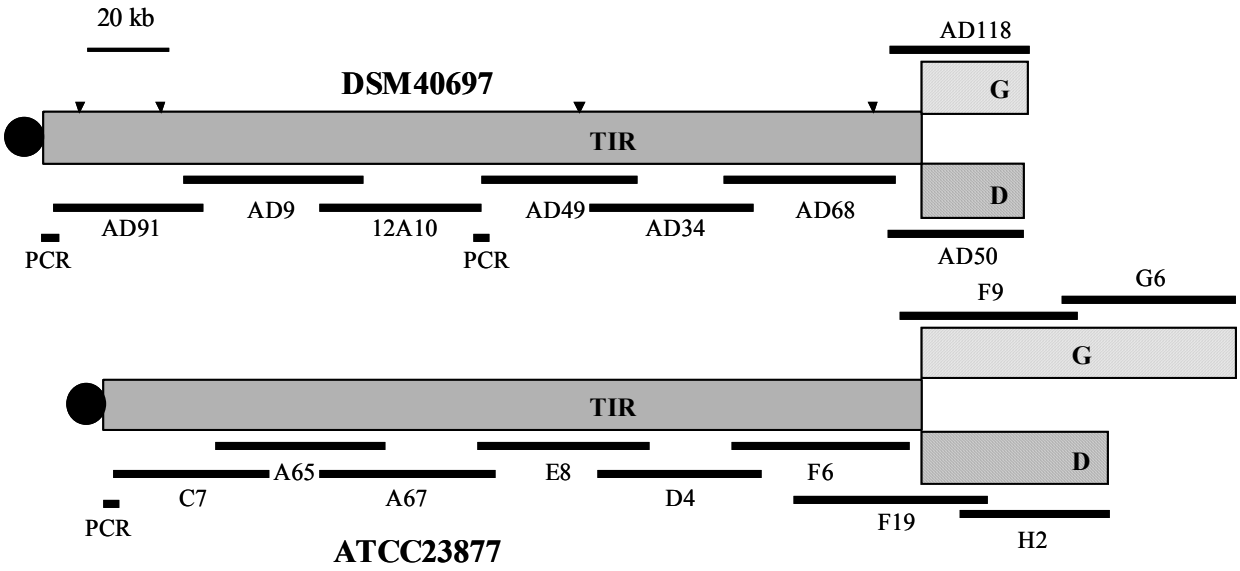


Figure 34 : Représentation graphique des TIR et des régions adjacentes des deux souches de *S. ambofaciens*

Les régions adjacentes sont hachurées (G : Bras Gauche ; D : Bras Droit). La localisation des cosmides et des produits de PCR est indiquée. La position des lacunes est représentée par un triangle.

E- Analyse bio-informatique des séquences

Les séquences obtenues ont été traitées sur la plate-forme génomique mise en place au sein du laboratoire par F. Choulet (Thèse 2006 UHP Nancy I). La base de données SAMDB contient l'intégralité des données issues de l'annotation des régions terminales de *S. ambofaciens* et celles issues de la génomique comparée des *Streptomyces*. Trois modules ont été mis en place pour l'assemblage, l'annotation et la génomique comparée des séquences : SAMASSEMBLER, SAMANNOT et SAMCOMP. L'articulation des applications au sein des différents modules et l'architecture partielle de SAMDB sont schématisées dans la figure 35.

E-1 MODULE D'ASSEMBLAGE SAMASSEMBLER

La suite logicielle phred/phrap/consed a été utilisée pour le traitement et l'assemblage des séquences issues des électrophorégrammes (Ewing et Green 1998; Ewing *et al.* 1998; Gordon *et al.* 1998). Des programmes permettent d'éliminer les traces de vecteur de clonage.

E-2 MODULE D'ANNOTATION SAMANNOT

La première étape est l'identification des CDS. Pour cela, le programme Glimmer2 a été utilisé. Ce programme prédit la position des gènes grâce à des modèles de Markov (Delcher *et al.* 1999). L'apprentissage de Glimmer2 a été réalisé avec 3000 ORF de *S. coelicolor*. Pour corriger les positions des codons d'initiation de la traduction, le programme RBSFinder a permis la recherche des RBS de traduction en amont des CDS.

A partir des séquences nucléotidiques de chaque CDS, le programme BlastFromEmbl.pl traduit chaque CDS en séquence protéique. Le programme BLASTP est utilisé pour rechercher des similarités de séquences avec les banques de données généralistes NR (Non Redundant) du National Center for Biotechnology Information (NCBI), regroupant toutes les séquences disponibles. Une approche manuelle est indispensable pour éviter au maximum des erreurs de prédiction.

La validation des CDS s'est faite sous Artemis (Rutherford *et al.* 2000). Deux paramètres doivent être contrôlés à cette étape, à savoir la validité des CDS prédites comme gène potentiel et la position du codon d'initiation. Les phases ouvertes de lecture présentent peu de codons stop caractérisés par une richesse en bases A et T. Ainsi un autre paramètre a été utilisé. Un biais très prononcé en G+C en troisième position de codon est observé dans les génomes à haut G+C (Wright et Bibb 1992). Ce biais atteint 92% chez *S. coelicolor* et 91% chez *S. avermitilis*. Pour les deux autres positions de codons, ce pourcentage est de respectivement 73% et 52% pour la première et la seconde position chez *S. coelicolor*. Cette particularité permet souvent d'éliminer les erreurs de prédiction de CDS. Un autre critère est l'existence d'homologues dans les banques de données. A partir de la détection du RBS et de la prédiction des CDS, les différentes phases de lecture ont été validées.

Pour la suite, différents programmes, comme PRODOM (Corpet *et al.* 1999) ou PROSITE (Hofmann *et al.* 1999) ont été utilisés pour la recherche de domaines protéiques ou encore la classification en catégories fonctionnelles.

E-3 MODULE DE GENOMIQUE COMPAREE : SAMCOMP

Ce dernier module est employé pour la comparaison de paire de génomes. Ce module intègre les fonctionnalités permettant la réalisation de comparaison ou encore la visualisation sous forme graphique de ces comparaisons. Pour cela, différents programmes sont utilisés. Le « dot-plot » permet de représenter les positions relatives des gènes homologues entre deux réplicons sur un graphique en deux dimensions. La mesure du GOC (Gene Order Conservation) est calculée pour étudier la stabilité des génomes bactériens (Rocha 2006). Cet indice définit la fréquence de conservation des paires de gènes contigus et de leurs orthologues entre deux génomes. Pour visualiser les régions homologues entre deux séquences nucléotidiques, le logiciel ACT (Artemis Comparison Tool) est utilisé. Ce logiciel permet la représentation graphique des résultats issus d'une comparaison réalisée par BLASTN.

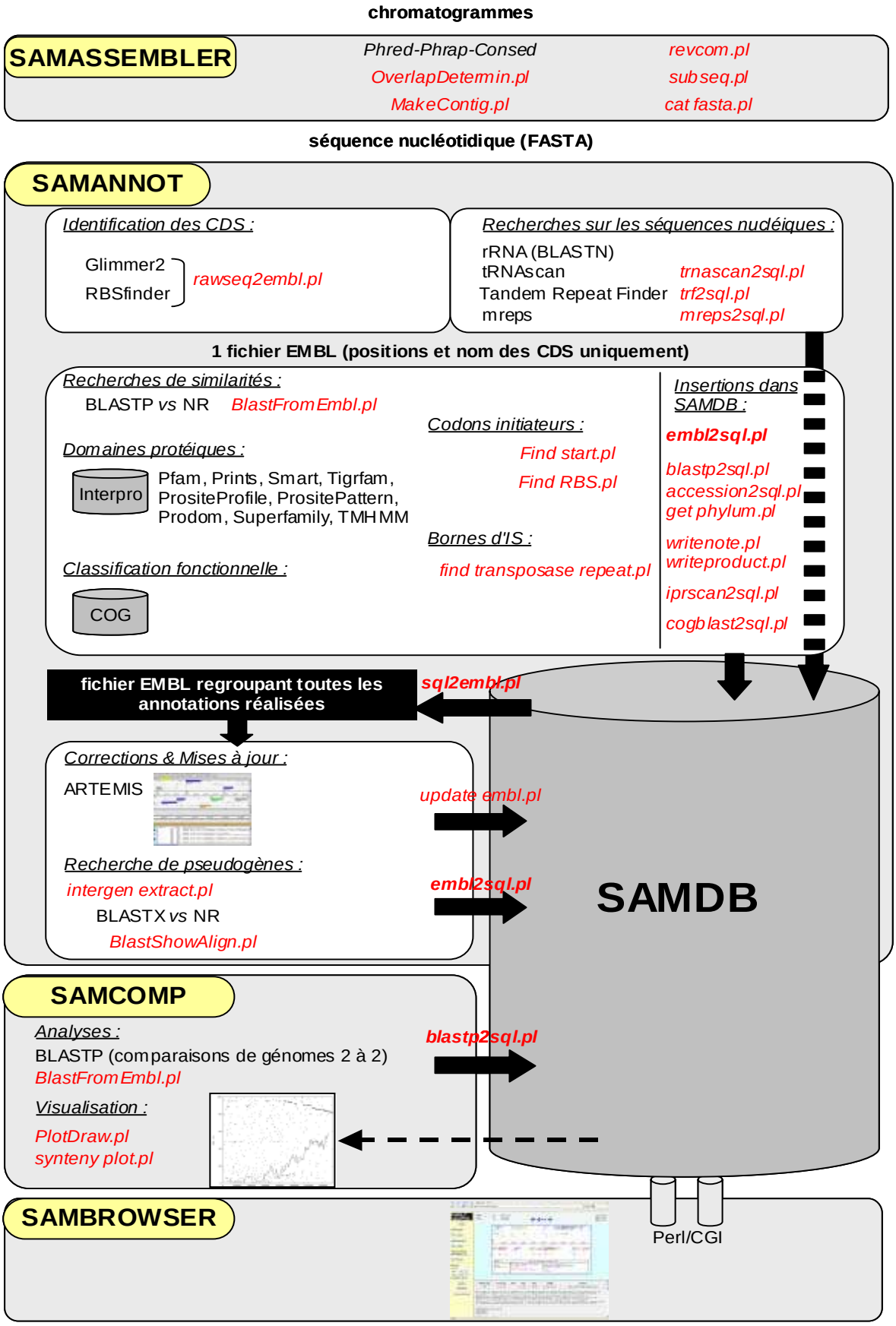


Figure 35 : Schéma de la plateforme dédiée à l'assemblage (SAMASSEMBLER), l'annotation (SAMANNOT) et la génomique comparée (SAMCOMP) du chromosome de *S. ambofaciens*. SAMBROWSER permet de parcourir l'annotation à travers une interface graphique et permet l'interrogation de

II- Variabilité intraspécifique des répétitions terminales inversées du chromosome de S. ambofaciens

Ces résultats sont rapportées dans :

Publication N°1

Choulet F., Gallois A. (co-auteur), 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.

Intraspecific variability of the terminal inverted repeat of the linear chromosome of *Streptomyces ambofaciens*.

Journal of Bacteriology, Sept. 2006, p. 6599-6610, Vol. 188, No. 18.

Cet article décrit la structure, composition et contenu des répétitions terminales inversées des deux souches de *S. ambofaciens* DSM40697 et ATCC23877, et permet de proposer des mécanismes d'évolution de ces structures terminales.

Pour les deux souches de *S. ambofaciens*, les séquences des Répétitions Terminales Inversées (TIR) ont été obtenues à partir d'une encyclopédie de cosmides. A partir des banques plasmidiques réalisées pour chaque cosmide, deux contigs de 238 517 pb (bras gauche) et 237 382 pb (bras droit) ont été obtenus pour la souche DSM40697. Les séquences des TIR de la souche de *S. ambofaciens* ATCC23877 ont été obtenues à partir du programme de séquençage des régions terminales de *S. ambofaciens*. L'assemblage de deux contigs de respectivement 1 544 032 et 1 367 119 pb pour les bras chromosomiques gauche et droit a été obtenu. Ces contigs contiennent les séquences des TIR de cette souche.

Les TIR de la souche de *S. ambofaciens* DSM40697 mesurent 212 655 pb alors que les TIR de la souche de *S. ambofaciens* ATCC23877 mesurent 197 936 pb.

A- Frontières des TIR

Les TIR sont définies comme des séquences répétées de façon identique aux extrémités chromosomiques. Ainsi, la frontière des TIR est définie par le premier nucléotide divergent entre les deux bras chromosomiques d'un même génome. Ainsi, les duplications terminales ne se terminent pas à la même position nucléotidique (Fig. 36). Cette divergence est le résultat de mutations ponctuelles et d'événements insertions/délétions, situés dans la même région. Cette région contient deux CDS, une cytochrome P450 putative (DSMT0215, SAML/R0195) et un homologue de *afsA* (DSML/R0216, SAML/R0196) séparées par une région riche en répétitions C/A. Une autre série de motifs riches en C/A est localisée dans la partie 3' d'*afsA*, et la variation du nombre de ces motifs définit trois allèles de ce gène parmi quatre copies (deux allèles par

souche).

Ainsi, la frontière des TIR pour la souche ATCC23877 est située après une ORF codant une transposase tronquée (SAMT0194) alors que, pour la souche DSM40697, elle se trouve dans la région intergénique entre la cytochrome P450 (DSMT0215) et l'homologue de *afsA* (DSML/R0216).

Les deux bras divergent totalement à partir de la même position nucléotidique pour les deux souches, montrant une localisation commune de l'ancienne frontière ancestrale.

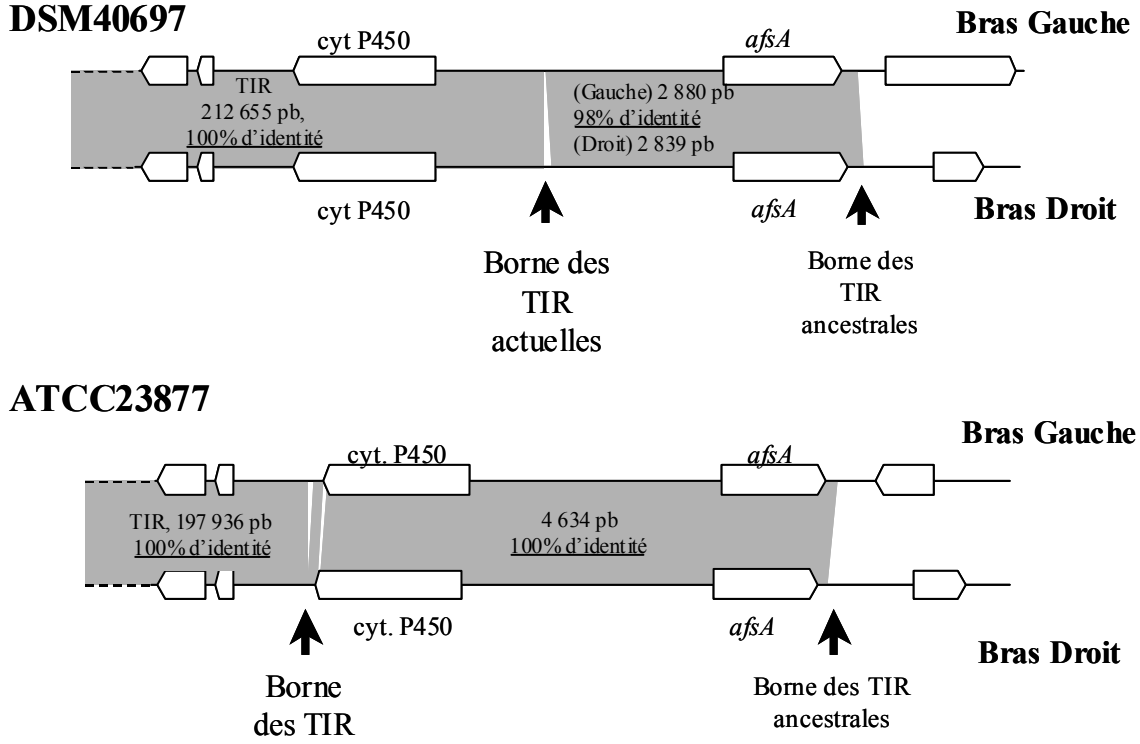


Figure 36 : Comparaison intrachromosomique des régions des bras droit et gauche environnant les frontières des TIR chez les souches DSM40697 et ATCC23877 de *S. ambofaciens*
Les régions dupliquées pour chaque bras sont représentées en gris.

B- Comparaison du contenu en gènes des TIR des deux souches de S. ambofaciens

L'analyse des séquences a permis de dresser une carte de comparaison entre les TIR des deux souches. Une grande région commune de 150 kb présente près de 99% d'identité en nucléotide. En revanche, cinq régions spécifiques, pour un total de 62 kb, pour la souche DSM40697 et de 49 kb divisée en quatre régions pour la souche ATCC23877, sont spécifiques de chacune de ces régions (Fig. 37). Les TIR de la souche ATCC23877 portent 194 CDS pour 215 CDS dans les TIR de la souche DSM40697, dont 45 et 65 sont spécifiques respectivement de chacune.

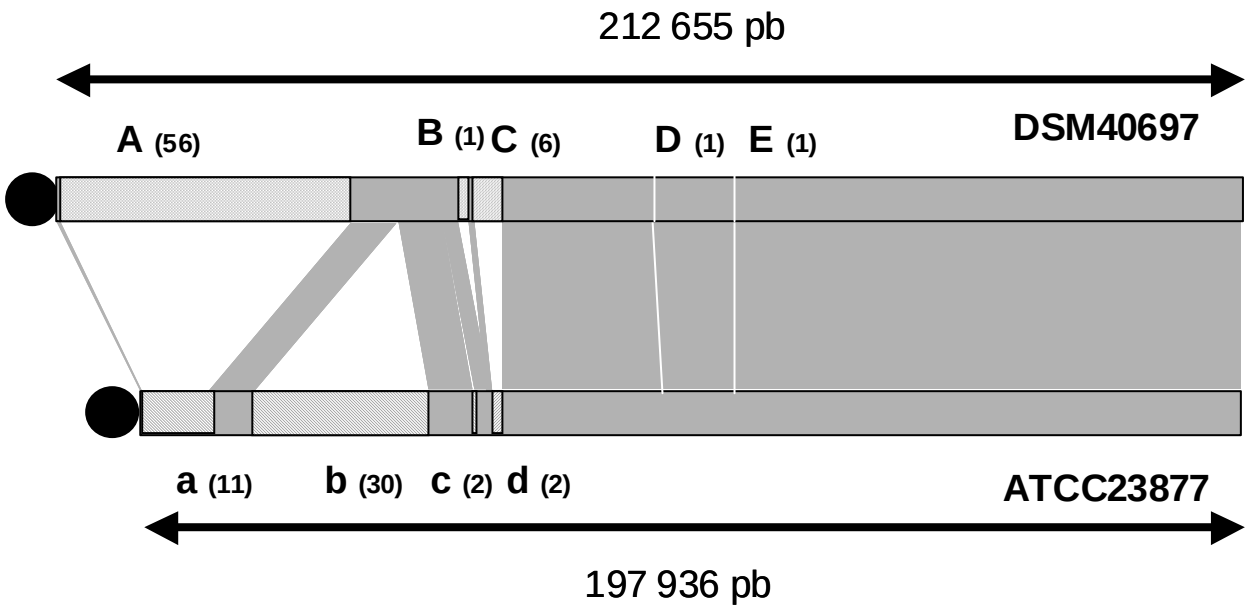


Figure 37 : Comparaison des séquences des TIR de *S. ambofaciens* ATCC23877 et DSM40697
Les régions communes sont en gris. Les régions souches-spécifiques sont nommées de la façon suivante : de A à E pour la souche DSM40697 et de a à d pour la souche ATC23877. Le nombre de gènes (y compris les pseudo-gènes) pour chaque région est indiqué entre parenthèses.

Les TIR des deux souches de *S. ambofaciens* sont caractérisées par une faible densité en séquences codantes. En considérant les pseudogènes comme de l'ADN non codant, cette densité est de respectivement 77,1% et 75,7% pour la souche DSM40697 et la souche ATCC23877. Même si les pseudogènes sont considérés comme des phases codantes, cette densité passe à 80,1% pour la souche DSM40697 et à 77,1% pour la souche ATCC23877, alors que la densité en ADN codant chez *S. coelicolor* et *S. avermitilis* est de respectivement 88,9% et 86,2% (Bentley *et al.* 2002; Ikeda *et al.* 2003).

B-1 LA REGION COMMUNE ENTRE LES TIR DES DEUX SOUCHES

Dans la région commune aux deux souches, 149 CDS sont retrouvées. La figure 38 représente la répartition de ces CDS en fonction de leur homologie dans les banques de données.

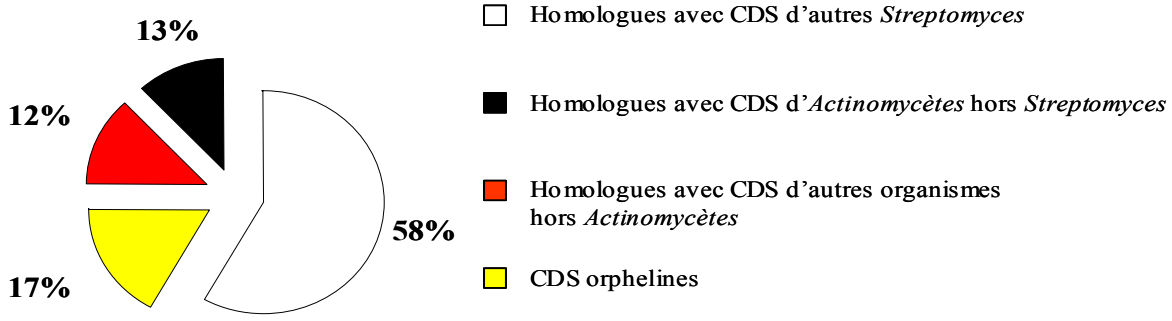


Figure 38

Concernant les CDS homologues avec des CDS d'autres *Streptomyces* (camembert blanc), 49% des CDS (soit 73/149 CDS) présentent la meilleure identité avec des séquences de *S. coelicolor* et *S. avermitilis*. Seules 3 présentent plus de 80% d'identité en acides aminés (DSMT0109-SAMT0091, DSMT0110-SAMT0092 et DSMT0203-SAMT0183).

Trois transposases sont retrouvées, situées à la frontière des régions spécifiques de chaque souche et la dernière à proximité des bornes des TIR. La présence de mutations « frameshift », de délétions ou encore l'absence de répétitions inversées montre qu'elles ne seraient pas fonctionnelles.

Enfin, le pourcentage en bases G-C de cette région est décrit en deux régions. Une première zone de 61 kb présente un pourcentage moyen de 74,4% alors que la seconde zone contient un pourcentage en bases G-C plus faible (66,5%).

B-2 REGIONS SOUCHE-SPECIFIQUES

La région spécifique des TIR de la souche DSM40697 mesure 62 kb et contient 65 CDS. Le pourcentage en G-C de cette région souche spécifique est faible (69,2%) comparé au pourcentage moyen de l'ensemble du chromosome de *S. coelicolor* (72,1%).

Le tableau 1 résume les caractéristiques des régions spécifiques de souche.

	Souche ATCC23877	Souche DSM40697
Taille (en kb)	49	62
Nombre de CDS	45	65
% G-C	68,8	69,2
Nombre de CDS d'origine plasmidique	5 (11%)	16 (24,6%) (8 CDS de SCP1 de *S. coelicolor* 5 CDS de SAP1 de *S. avermitilis*)
Homologue des CDS liés à des plasmides	- *ttrA* (84% d'identité en aa avec *ttrA* de SLP2) - *kilB* (48% d'identité en aa avec pRL2.23 de pRL2) - Transposase (71% d'identité en aa avec *orf118* de pSLA2-L)	- Facteur σ d'ARN polymérase (37% d'identité en aa avec SAP1_87 de SAP1) - Esterase (93% d'identité en aa avec SCP1.199c de SCP1) - *tpg* (pseudogène) (75% d'identité en aa avec *tpgSLP2* de SLP2) - Oxydoréductase FAD dép. (58% d'identité en aa avec pREL1_0108 de pREL1) - ADN ligase ATP dép. (64% et 59% d'identité en aa avec SAP1_90 de SAP1)
Exemples de gènes accessoires	- Gène *ard2* de résistance à l'antibiotique A201A (54% d'identité en aa avec *Saccharothrix mutabilis subsp. Capreolus*)	- Phosphinothricine N-acétyltransférase (77% d'identité en aa avec *S. coelicolor*) - Nourseothricine acétyltransférase (69% d'identité en aa avec *Streptomyces noursei*) - Cluster de résistance à l'arsenate (5 CDS de 78 à 87% d'identité en aa avec *S. coelicolor*)

Tableau 2 : Caractéristiques des séquences des régions spécifiques et de leur contenu
dép. : dépendante ; aa : acides aminés

Les CDS présentent souvent une organisation conservée, ce qui pourrait révéler un événement de transfert unique. Par exemple, chez la souche DSM40697, quatre CDS (DSMT0042 à DSMT0045) présentent leur meilleure homologie au locus comprenant quatre CDS du plasmide SCP1 (SCP1.202 à SCP1.199c) mais dans l'ordre inversé.

C- Conservation des télomères

La comparaison des séquences télomériques entre les deux souches a montré une identité s'élevant à 96,1% sur 180 pb. Les sept palindromes, permettant la structure secondaire typique des télomères de *Streptomyces*, sont conservés (Bey *et al.* 2000; Huang *et al.* 2003). La comparaison avec les séquences télomériques d'autres *Streptomyces* montre des divergences localisées dans les motifs III et V, et correspondent à des variations compensatoires au sein de la structure secondaire prédite (Fig. 39).

Chez les *Streptomyces*, cette structuration de télomères est prédite pour la grande majorité des réplicons. Néanmoins, quelques exceptions ont été décrites comme pour le plasmide SCP1 de *S. coelicolor* et le chromosome de *S. griseus* (Kinashi *et al.* 1991; Goshi *et al.* 2002). La structure secondaire prédite pour le plasmide SCP1 présente des boucles contenant quatre nucléotides au lieu de trois. Pour le chromosome de *S. griseus*, de nombreux palindromes sont prédits mais leur séquence diverge de celles des télomères typiques des *Streptomyces.*

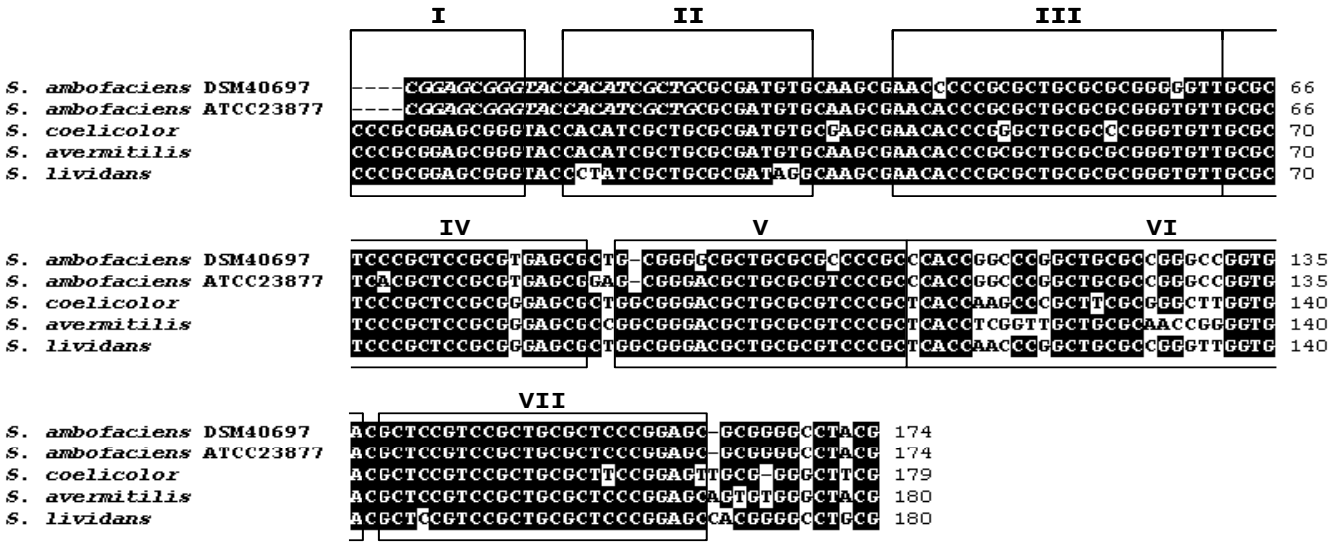


Figure 39 : Alignement des séquences télomériques des deux souches de *S. ambofaciens* avec les télomères de *S. coelicolor* A3(2), *S. lividans* ZX7 et *S. avermitilis* MA-4680
Le numéro des palindromes est indiqué en chiffre romain.

Tous ces résultats montrent que la variabilité terminale serait issue d'évènements récents avec l'existence d'un flux de gènes façonnant ces extrémités chromosomiques, par échanges d'extrémités entre réplicons linéaires en raison des similarités retrouvées avec des CDS associées à des plasmides de *Streptomyces*. Ces échanges pourraient être à l'origine de la diversification de ces extrémités au niveau intraspécifique.

JOURNAL OF BACTERIOLOGY, Sept. 2006, p. 6599–6610 Vol. 188, No. 18
0021-9193/06/$08.00+0 doi:10.1128/JB.00734-06
Copyright © 2006, American Society for Microbiology. All Rights Reserved.

Intraspecific Variability of the Terminal Inverted Repeats of the Linear Chromosome of *Streptomyces ambofaciens*

Frédéric Choulet,[1]# Alexandre Gallois,[1]# Bertrand Aigle,[1] Sophie Mangenot,[2] Claude Gerbaud,[3] Chantal Truong,[2] François-Xavier Francou,[3] Frédéric Borges,[1] Céline Fourrier,[1] Michel Guérineau,[3] Bernard Decaris,[1] Valérie Barbe,[2] Jean-Luc Pernodet,[3] and Pierre Leblond[1]*

Laboratoire de Génétique et Microbiologie, UMR INRA 1128, IFR 110, Université Henri Poincaré Nancy 1, Faculté des Sciences et Techniques, BP239, 54506 Vandœuvre-lès-Nancy, France[1]; Génoscope, Centre National de Séquençage, 2 rue Gaston Crémieux CP5706 91057 Evry cedex, France[2]; and Institut de Génétique et Microbiologie, UMR CNRS 8621, Université Paris-Sud 11, Bâtiment 400, 91405 Orsay cedex, France[3]

Received 22 May 2006/Accepted 7 July 2006

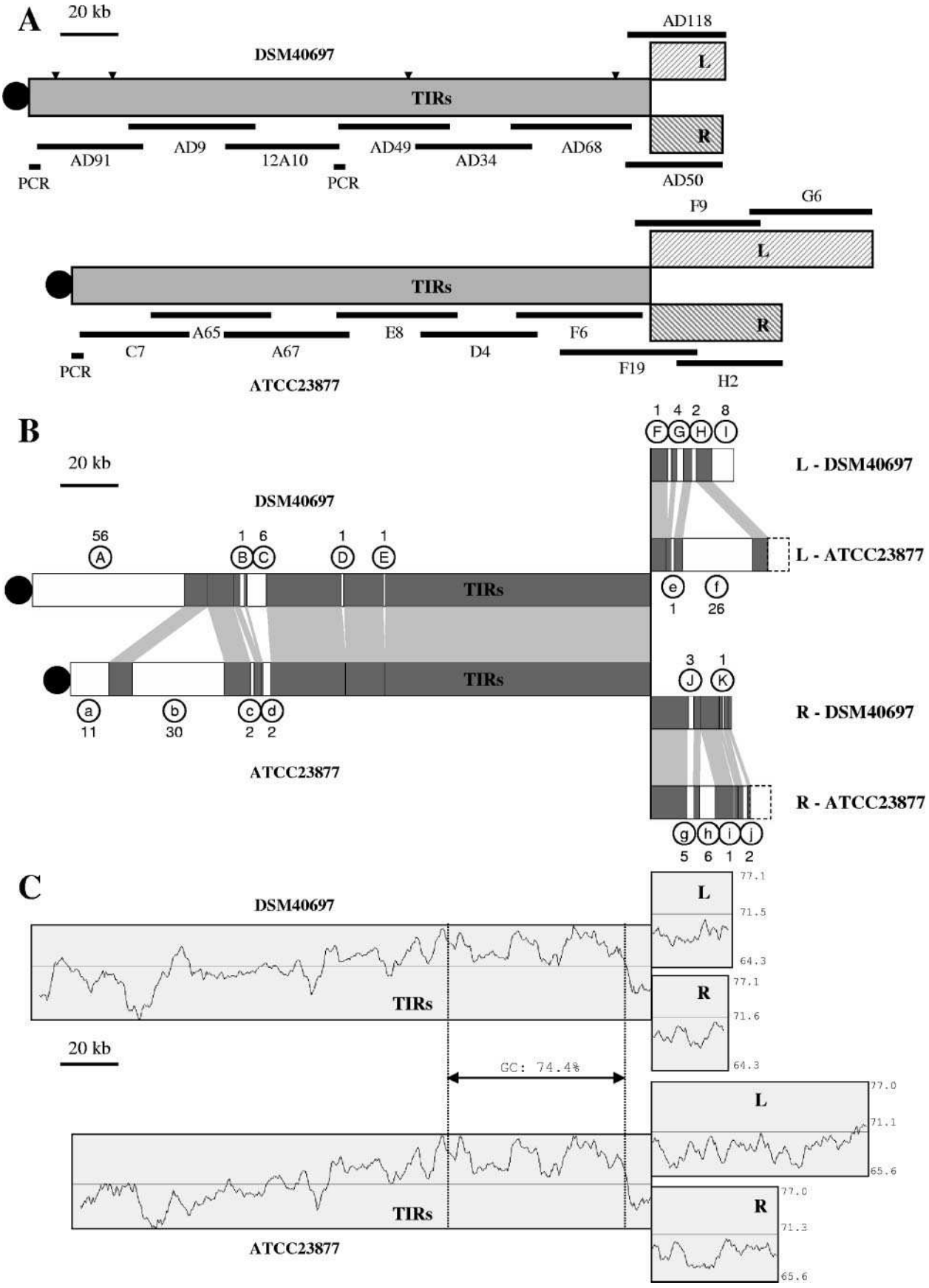
The sequences of the terminal inverted repeats (TIRs) ending the linear chromosomal DNA of two *Streptomyces ambofaciens* strains, ATCC23877 and DSM40697 (198 kb and 213 kb, respectively), were determined from two sets of recombinant cosmids. Among the 215 coding DNA sequences (CDSs) predicted in the TIRs of strain DSM40697, 65 are absent in the TIRs of strain ATCC23877. Reciprocally, 45 of the 194 predicted CDSs are specific to the ATCC23877 strain. The strain-specific CDSs are located mainly at the terminal end of the TIRs. Indeed, although TIRs appear almost identical over 150 kb (99% nucleotide identity), large regions of DNA of 60 kb (DSM40697) and 48 kb (ATCC23877), mostly spanning the ends of the chromosome, are strain specific. These regions are rich in plasmid-associated genes, including genes encoding putative conjugal transfer functions. The strain-specific regions also share a G+C content (68%) lower than that of the rest of the genome (from 71% to 73%), a percentage that is more typical of *Streptomyces* plasmids and mobile elements. These data suggest that exchanges of replicon extremities have occurred, thereby contributing to the terminal variability observed at the intraspecific level. In addition, the terminal regions include many mobile genetic element-related genes, pseudogenes, and genes related to adaptation. The results give insight into the mechanisms of evolution of the TIRs: integration of new information and/or loss of DNA fragments and subsequent homogenization of the two chromosomal extremities.

Streptomyces chromosomal DNA is linear and is among the largest described for bacteria, typically 8 to 10 Mb (6, 19). *Streptomyces* linear replicons (chromosomes and plasmids) share an invertronic structure including the presence of terminal inverted repeat sequences (TIRs) ended by bacterial telomeres covalently linked to terminal proteins (35). The lengths and sequences of the TIRs are extremely variable, and their sizes are not correlated to that of the replicon. The terminal duplications can be as large as several hundreds of kilobases, e.g., in the chromosome of *Streptomyces coelicolor* M600 (~1 Mb) (43), or can be restricted to the telomeric palindromes, e.g., in the chromosome of *Streptomyces avermitilis* (167 bp) (19). Two *Streptomyces* annotated genomes have been released so far: that for *S. coelicolor* A3(2), with 7,825 predicted coding DNA sequences (CDSs) (6), and that for *Streptomyces avermitilis* MA-4680, with 7,577 putative CDSs (19). Comparison of the two genomes revealed a common general organization with a conserved central region of about 5 Mb and terminal regions (or "arms") carrying mainly nonessential variable genes (6, 19). The terminal regions appear poorly conserved at the level of gene content and organization, which contrasts with the strong synteny observed for the central region. The fact that the terminal regions are dispensable for vegetative growth in laboratory conditions was revealed earlier by the characterization of instability phenomena (25). In *S. ambofaciens*, up to 2.5 Mb located at the ends of the chromosome can be lost and is not essential for vegetative growth in laboratory growth conditions. In addition, large DNA rearrangements, such as duplications, deletions, and amplifications, frequently occur in the subtelomeric regions. One of the most spectacular rearrangements affects the size of the TIRs, which can vary from 5 kb to 1.4 Mb in spontaneous mutant strains (46), while the wild-type strain DSM40697 harbors 210-kb TIRs (26). This variation implies nonreciprocal translocations of chromosomal extremities that in some cases result from homologous recombination between duplicated genes (14). Another phenomenon triggering TIR variation implies exchanges of replicon extremities between plasmids and chromosomes. This was demonstrated for a strain of *S. coelicolor* in which chimeric chromosomes can be generated by crossover of the wild-type chromosome and the linear plasmid SCP1 (47). This phenomenon was also shown for *Streptomyces rimosus* by interaction between plasmid pZG101 and the chromosome (30) and strongly suggested by the analysis of the terminal structure of the *Streptomyces lividans* chromosome, which could result from partial integration of the linear plasmid SLP2 (17).

In order to gain further insight into the mechanisms of chromosomal end diversification in *Streptomyces*, we analyzed

* Corresponding author. Mailing address: Laboratoire de Génétique et Microbiologie, UMR INRA 1128, IFR 110, Faculté des Sciences et Techniques, Université Henri Poincaré - Nancy 1, Boulevard des Aiguillettes, BP239, 54506 Vandœuvre-lès-Nancy, France. Phone: 33(0)3 83 68 42 07. Fax: 33(0)3 83 68 44 99. E-mail: leblond@nancy.inra.fr.

# F.C. and A.G. contributed equally to this report.


A
20 kb
DSM40697
AD118
L
TIRs
R
AD91
AD9
12A10
AD49
AD34
AD68
PCR
PCR
AD50
F9
G6
L
TIRs
R
A65
E8
F6
C7
A67
D4
F19
H2
PCR
ATCC23877
B
20 kb
DSM40697
1 4 2 8
F G H I
L - DSM40697
56
A
1 6
B C
1
D
1
E
L - ATCC23877
TIRs
e
1
f
26
TIRs
3 1
J K
a
11
b
30
c d
2 2
R - DSM40697
ATCC23877
R - ATCC23877
g h i j
5 6 1 2
C
DSM40697
L
77.1
71.5
64.3
TIRs
R
77.1
71.6
64.3
20 kb
GC: 74.4%
L
77.0
71.1
65.6
TIRs
R
77.0
71.3
65.6
ATCC23877

the variability of the TIRs of two isolates belonging to the *S. ambofaciens* species. It was previously shown by cross-hybridization experiments that they could be distinguished by the terminal regions (13). These two independent soil isolates, ATCC23877 (32) and DSM40697 (18), were assigned to the same species according to classical morphological and physiological traits, e.g., antibiotic production. These two isolates indeed share the same antibiotic synthesis profile, both producing the three known antibiotic compounds spiramycin, congocidin, and alpomycin (31, 32). In contrast, the closely related species *S. coelicolor* (1.1% divergence of 16S rDNA sequence from that of *S. ambofaciens* ATCC23877) shows a completely different secondary metabolite profile.

This phenetic classification was supported by more-recent molecular analyses. First, at the whole-genome scale, pulsed-field gel electrophoresis analysis, which is a highly sensitive approach to distinguishing and identifying bacterial isolates, showed that the two strains differ only slightly, whereas the *S. coelicolor* chromosome diverges much more (26). Further, mapping experiments with large chromosomal DNA fragments and linking clones showed that most of them were common to both strains, leading to the conclusion that the two chromosomes show a colinear organization (26). An additional piece of evidence is the presence of a recent gene duplication affecting a sigma factor-encoding gene (*has* gene). This duplication is common to the two *S. ambofaciens* isolates, while a single copy of the *has* gene is present in the two complete genome sequences of *S. coelicolor* and *S. avermitilis* (34).

Furthermore, the most convincing evidence for their close relationships comes from the analysis of their 16S-23S internal transcribed spacer sequences, whose evolution is the most effective (among the *rrn* operon) to infer close phylogenetic relationships. The six internal transcribed spacer regions were isolated from the two strains, and their sequences were compared to those of *S. coelicolor* (45). While the two isolates share identical sequences, these sequences differ from that of *S. coelicolor*, thereby showing the divergence of the strains. Although gene conversion between the *rrn* loci could be identified, no crossover has occurred between them, maintaining the colinearity of the two chromosomes (45).

Altogether, these data show that the two isolates are closely related strains and that their assignment to the same species is supported by a multicriterion analysis. In this work, we undertook a detailed sequence analysis of the TIRs of these two *S. ambofaciens* strains to determine the strain-specific gene content and to gain a deeper insight into the mechanisms important for the evolution of chromosomal extremities.

MATERIALS AND METHODS

Sequencing. For each strain, ATCC23877 and DSM40697, a cosmid library was constructed from partially BamHI-digested *S. ambofaciens* genomic DNA cloned into the Supercos1 (Stratagene) vector. As the size of the *S. ambofaciens* TIRs greatly exceeds that of a fragment readily clonable into a cosmid vector, each copy of the TIRs cannot be isolated as a single recombinant molecule. Consequently, the sequences were obtained from a set of ordered recombinant cosmids (Fig. 1A). Note that for cosmids with insert sequences entirely corresponding to the TIRs, e.g., from cosmid C7 to F6 for strain ATCC23877, the chromosomal origin (i.e., from the right or left arm) of the recombinant cosmids cannot be deduced. Therefore, the TIR sequence is likely to consist of a chimera between the left and right repeats.

For sequencing, cosmids were mechanically fragmented and cloned with a BstXI adaptor into either pcDNA2.1 vector (Invitrogen) or pCNS (a derivate of pSU18 [4]). Ligation products were then introduced into *Escherichia coli* DH10B. It was not possible to obtain a cosmid clone containing the terminal fragment that includes the telomeres, and a PCR strategy was therefore used to amplify and sequence this region in both strains. Primers were designed from the cosmid C7 sequence ($^{5'}$CACCCAGCGAGCCCAGCA$^{3'}$) for strain ATCC23877 and from the cosmid AD91 sequence ($^{5'}$AGCTGCAACGGTGCGTTCTATTG GG$^{3'}$) for strain DSM40697, and a third primer was designed according to a consensus of the telomere sequences derived from several *Streptomyces* species ($^{5'}$CGGAGCGGGTACCACATCGCTG$^{3'}$). PCR was performed using 50 ng of DNA, with 800 μM deoxynucleoside triphosphates, 2 units of LA *Taq* polymerase (Takara), 2.5% dimethyl sulfoxide, and 20 pmol of each primer in a 50-μl final volume. After a denaturation step (95°C, 5 min), 30 cycles of denaturation (95°C, 30 s), annealing (58°C, 30 s), and polymerization (68°C, 3 min) were used to amplify the terminal fragment.

The 10 cosmids of strain ATCC23877 were completely sequenced (mean coverage, 10×). For strain DSM40697 (mean coverage, 10×), 17 gaps were remaining after the shotgun sequencing of the eight cosmids. PCR products were obtained for each gap and were cloned into pGEM-T Easy vector (Promega) for production of single-stranded DNA. For each gap, sequencing reactions were performed on all available templates, i.e., PCR product, double-stranded recombinant pGEM-T Easy vector, and the single-stranded DNA. In addition, sequencing reactions were carried out using two different reagents, CEQ Dye terminator (Beckman) and BigDye Terminator (Applied Biosystems). Four gaps remained, all localized in intergenic regions, suggesting they probably resulted from the formation of intrastrand secondary structures (e.g., terminators). Gap 1 (between DSMT0010 and DSMT0011) and gap 2 (between DSMT0032 and DSMT0033) are included in a strain-specific region (cosmid AD91) (Fig. 1A), and they are estimated to be less than 50 bp in length. Gap 3 (between DSMT0146 and DSMT00147) and gap 4 (between DSMT0204 and DSMT0205) belong to regions highly conserved between the two strains, and their sizes can be estimated as less than 10 bp and 200 bp, respectively, by comparison to the sequence of strain ATCC23877.

Annotation. The gene finder Glimmer2.10 (11) was used for CDS prediction, with a minimum size of 40 codons arbitrarily chosen as the threshold. Results were then refined by RBSfinder (39). The Basic Local Alignment Search Tool (BLAST 2.2.6) was used to find similarities (1), and the Interpro package was used to describe protein domains (48). CDSs were assigned a functional category where their best cluster of orthologous groups (COG) homologue is classified (40). Then, BLASTX translations were realized for each intergenic region in order to detect initially unpredicted CDSs and pseudogenes. While comparing the predicted protein sequences with BLASTP, proteins sharing more than 30% identity over at least 80% of the length of the query sequence were considered to be homologous.

FIG. 1. (A) Schematic representation of the *S. ambofaciens* ATCC23877 and DSM40697 TIRs and adjacent regions (L, left arm; R, right arm) and of the cosmids and PCR products used for the sequencing. Gap positions are represented by triangles (see Materials and Methods). (B) Comparison of the TIR sequences and flanking regions of *S. ambofaciens* ATCC23877 and DSM40697. Strain-specific regions are represented by white rectangles, whereas conserved regions are represented in gray. The strain-specific regions are named as follows: from A to K for strain DSM40697 and from a to j for strain ATCC23877, and the number of genes (including pseudogenes) carried by each specific region is indicated. Since the regions sequenced outside of the TIRs are larger for strain ATCC23877, sequences which cannot be compared because they were not sequenced in the second strain are represented by dashed rectangles. The terminal protein, covalently bound to the telomere, is represented by a black circle. (C) G+C content of TIRs and adjacent regions for the two strains (displayed with a 5,000-bp window size). Minimum, maximum, and mean (represented by a straight line) values of G+C percentages are indicated at the right side. These values are calculated for the whole contigs including the TIRs and the left or right adjacent region.

TABLE 1. General features of the TIR sequences of the *S. ambofaciens* strains ATCC23877 and DSM40697

Strain	Size of TIR (bp)	G+C content (%)	No. of predicted CDSs (no. of pseudogenes)	No. of proteins with assigned function (%)	No. conserved with proteins of unknown function (%)	No. of orphans (%)
ATCC23877	197,936	71.8	194 (3)	125 (65)	49 (25)	20 (10)
DSM40697	212,655	71.9	215 (10)	135 (63)	54 (25)	26 (12)

For strain ATCC23877, duplicated genes in the TIRs were named SAMT*nnnn*, whereas those specific only to either the left or right arm were annotated as SAML*nnnn* or SAMR*nnnn*, respectively (underlining indicates the type of specificity). A similar principle was adopted for strain DSM40697, with the prefix "DSM" used instead of "SAM." Sequences of left and right contigs for the two strains and their corresponding annotations are available through the SAMDB web server at http://www.weblgm.scbiol.ambofaciens.uhp-nancy.fr/.

Nucleotide sequence accession numbers. The sequences were deposited in EMBL under the following accession numbers: AJ937740 (ATCC23877 left TIR), AJ937741 (ATCC23877 right TIR), AM279694 (DSM40697 left TIR), and AM279695 (DSM40697 right TIR).

RESULTS

Intraspecific variability at the chromosomal ends. The large TIR sequences of two *S. ambofaciens* strains were determined using sets of recombinant cosmids spanning the terminal regions of the chromosome (Fig. 1A). Chimeric sequences of 197,936 bp and 212,655 bp (see Materials and Methods) were produced for strains ATCC23877 and DSM40697, respectively (see Table 1 for general features).

Considering a single genome, the 100% nucleotide identity of the two TIR copies was supported by two lines of evidence. First, hybridization of DNA probes corresponding to the TIRs onto genomic DNA did not reveal any polymorphism (26). Second, no mismatch was found within the overlaps of sequenced cosmids, despite the fact that they can originate from either of the chromosomal arms.

The ends of the TIRs will be defined as the first nucleotide of divergence (noted as "internal boundary of the TIRs" in Fig. 2A). Consequently, the terminal duplication does not end at exactly the same nucleotide position in the two strains. However, this situation results from point mutations and small insertions/deletions causing minor differences between the arms in both strains in the flanking small region (over about 5 kb). These regions show near identity and contain two CDSs, a probable cytochrome P450 (DSMT0215, SAML/R0195) and a homologue of *afsA* (DSML/R0216, SAML/R0196), separated by a large intergenic region that includes stretches of short repeated C/A-rich motifs (Fig. 2A). Another stretch of C/A-rich motifs is located in the 3′ part of *afsA*, and the variation in the number of motifs results in differences between the four *afsA* copies (two copies in each strain). Consequently, the *afsA* coding sequences have different sizes (three different alleles for four copies). In addition, these open reading frames (ORFs) might be pseudogenes, since only the 5′ end (about 200 codons) shows similarity (between 38% and 46%) with the *afsA*-like gene of *Streptomyces rochei* (28).

The two arm sequences diverge totally at the same nucleotide proximal to *afsA* in the two strains. Thus, the two strains share the same ancestral boundaries of TIRs (Fig. 2A).

Thus, 194 CDSs (from SAMT0001 to SAMT0194) belong to the TIRs of strain ATCC23877. This last ORF is similar to that encoding a truncated transposase. In strain DSM40697, the end of the sequences assumed to be identical occurs within the long intergenic region separating DSMT0215 (putative cytochrome P450) and DSML/R0216 (*afsA* homologues) (Fig. 2A).

When the TIRs of the two strains are compared, two regions can be distinguished: a terminal strain-specific region and a conserved region with a more internal location. In total, the syntenic regions extend over about 150 kb and include 149 genes sharing an average 99% nucleotide identity (Fig. 1B).

Within the whole TIRs, seven strain-specific segments resulting from DNA rearrangements such as insertions, deletions, and/or gene replacements involving from 1 to 56 CDSs can be delimited (Fig. 1B). For strain DSM40697, five specific regions (A to E) represent 62 kb and contain 65 CDSs (including probable pseudogenes). Region A on its own includes 56 of the 65 strain-specific CDSs and represents a quarter of the TIR size. For strain ATCC23877, the specific loci are scattered into four regions (a to d) spanning 49 kb and including 45 strain-specific CDSs. Again, 41 of these 45 specific genes are clustered into two loci, one of which constitutes the chromosomal end (region a, 14 kb).

Although the chromosomal ends are strain specific, highly similar telomere sequences were found in the two strains (96.1% nucleotide identity over 180 bp) (Fig. 3). The seven typical palindromes predicted to form the secondary structures found in other known *Streptomyces* telomeres are present in both strains (7, 17) (Fig. 3).

Outside the TIRs, intraspecific variability can also be noticed. Within the areas sequenced, at least four and five DNA rearrangements have occurred in the left and right arms, respectively (Fig. 1B; also see below).

Altogether, these results suggest that, despite their variability in gene content, the current TIRs of the two *S. ambofaciens* strains appear to derive from a single ancestral event of terminal duplication, since they share the same ancestral internal boundary.

Low G+C content. Despite the differences in gene content observed between the two strains, profiles of G+C percentages are quite similar (Fig. 1C). For each strain, a decrease in G+C content characterized the strain-specific extremities as well as the sequences outside of the TIRs. Thus, the terminal 50 kb of the ATCC23877 and DSM40697 chromosomes show G+C contents of 68.8% and 69.2%, respectively. Similar values are observed for the regions sequenced outside of the TIRs (69.1% in strain ATCC23877 and 68.9% in strain DSM40697). In contrast, G+C contents of *Streptomyces* chromosomes are 72.1% for *S. coelicolor* and 70.7% for *S. avermitilis*. In fact, the TIRs of *S. ambofaciens* seem to be present in regions of even lower G+C content, but the presence of a 61-kb cluster (which includes the alpomycin cluster [31]) showing a G+C content of

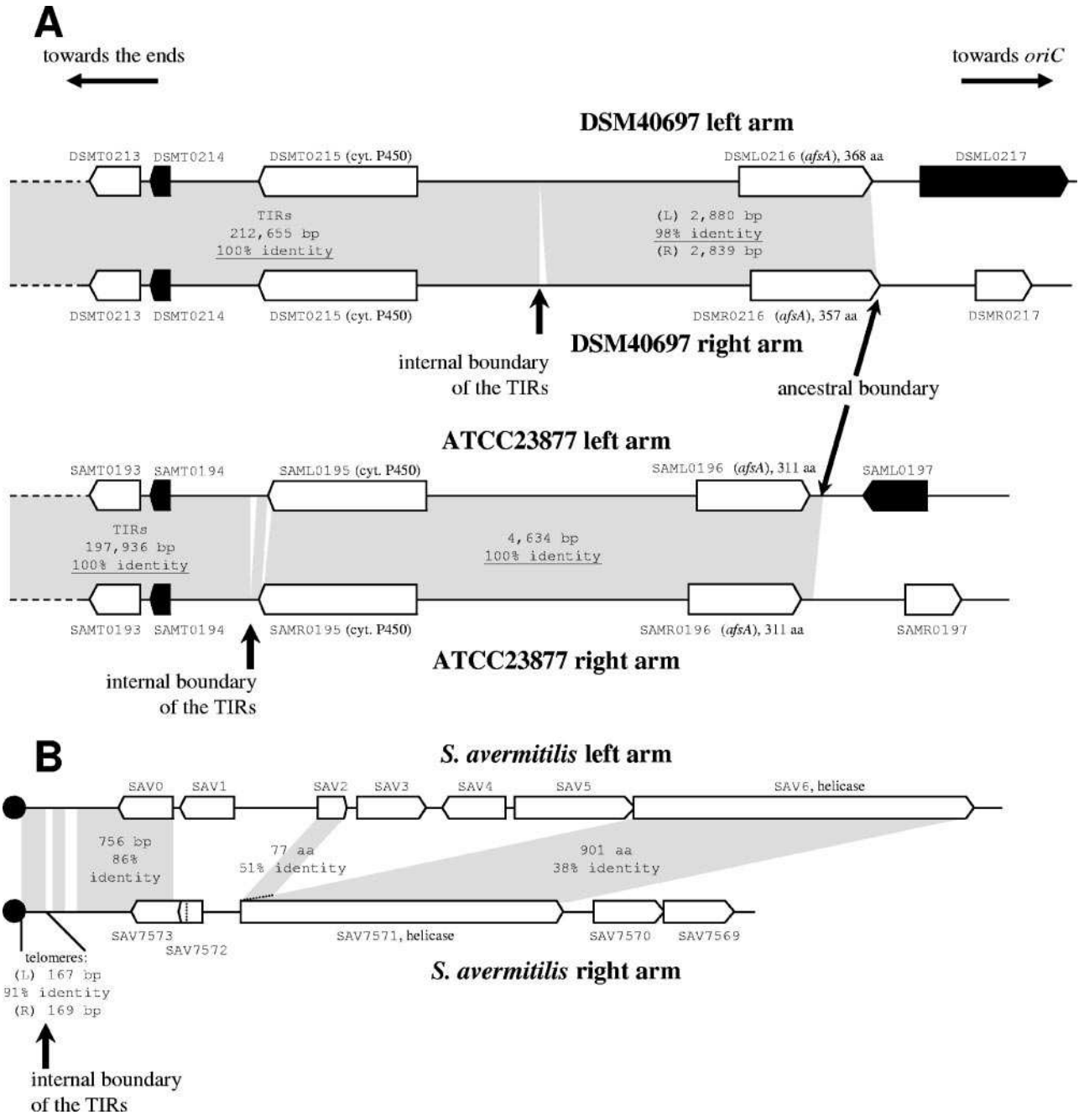


FIG. 2. Intrachromosomal comparisons of the left and right arm regions surrounding the ends of the TIRs of *S. ambofaciens* strains ATCC23877 and DSM40697 (A) and *S. avermitilis* (B). Duplicated regions between both arms are shaded in gray, and their sizes and percentages of identity are indicated. Black arrows represent ORFs similar to genes encoding transposases (or truncated transposases), and the terminal proteins are represented by the black circles. The SAV0 putative CDS was not predicted in reference 19 and was added in this work. cyt., cytochrome. aa, amino acid.

74.4% skews the data by causing a local increase in G+C content (Fig. 1C).

The low G+C content of the chromosomal extremities is reminiscent of that of *Streptomyces* plasmids (SLP2, 68.4% [17]; pSV2, 69.7% [38]; SCP1, 69.0% [5]; and SAP1, 69.2% [19] [see "Plasmid-associated genes" below]). The lower G+C content observed in the regions outside of the TIRs is related to a remarkable abundance of insertion sequences (ISs) and related genes in the strain-specific regions (found in regions F, G, H, and K in strain DSM40697 and in regions f, h, and j in strain ATCC23877) (Fig. 1B; also see "Mobile genetic element-related genes" below). The lower G+C content of mobile genetic elements in bacteria has been discussed previously (33), and their presence, together with an observed low G+C content, suggests acquisition by horizontal gene transfer.

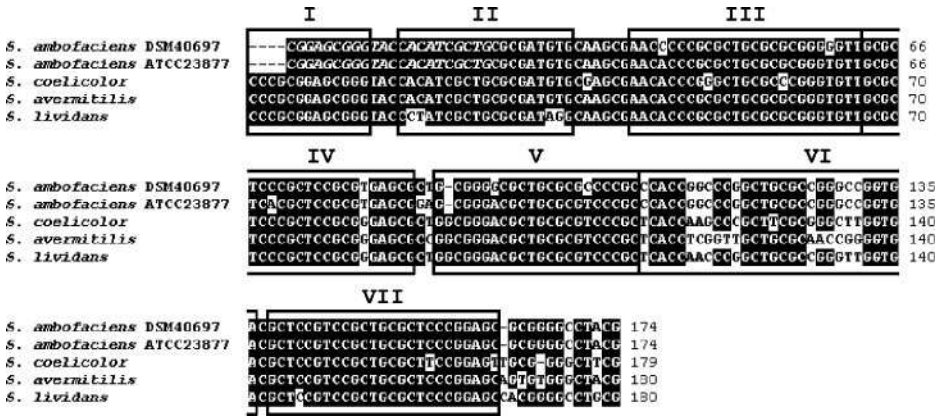


FIG. 3. Sequence alignment of the telomeres of the two *S. ambofaciens* strains with the chromosomal telomeres of *S. coelicolor* A3(2), *S. avermitilis* MA-4680, and *S. lividans* ZX7. The palindromes are labeled, and the primer sequence used for the amplification of the telomeres (see Materials and Methods) is represented in italic. This PCR strategy explains the missing nucleotides at the end of the *S. ambofaciens* telomeres. The boxes labeled I to VII indicate the positions of palindromes starting from the end of the chromosomal DNA. Numbers to the right of the sequences indicate the cumulative length for each of the aligned sequences.

Plasmid-associated genes. A large proportion of the strain-specific CDSs included in the TIRs may have a plasmid origin (see Table 2). This situation is particularly striking for strain DSM40697, for which 16 of the 56 CDSs located in the strain-specific region A show best similarity with plasmid-associated genes. For half of them, the level of identity is particularly high, i.e., more than 80% amino acid identity. Five gene products share best similarity with proteins encoded by linear plasmid SAP1 from *S. avermitilis* (19) and eight with linear plasmid SCP1 from *S. coelicolor* (5). Interestingly, one gene fragment (DSMT0048) is similar to *tpg* from *S. lividans* linear plasmid SLP2, which encodes the terminal protein involved in the replication of the telomeres (17). Some of the best hits were also with plasmids from other *Actinomycetales* spp., i.e., pNF1 (circular) from *Nocardia farcinica* (20) and pREL1 (linear) from *Rhodococcus erythropolis* (37). Furthermore, syntenic clusters between these strain-specific regions and linear plasmids (Table 2) were found, which strongly supports the hypothesis of integration of plasmid DNA into the terminal regions (9, 42). For example, the DSMT0042-45 cluster is conserved with the cluster SCP1.199-202 from linear plasmid SCP1.

The same conclusion can be inferred from the analysis of region a from strain ATCC23877. First, the probable DNA helicase TtrA (SAMT0002) shows the best BLASTP hit (84% amino acid identity) with that of plasmid SLP2 of *S. lividans*. This helicase is implicated in the conjugal transfer of the SLP2 plasmid (8). The *ttrA* gene is present at the extremities of the chromosome of each *S. ambofaciens* strain, although it is located in the strain-specific regions. It should be pointed out that a *ttrA* homologue is present close to the telomeres in almost all *Streptomyces* replicons. In fact, TtrA proteins of the two *S. ambofaciens* strains, which share 45% identity, may have a different origin: TtrA from strain ATCC23877 shares 84% identity with TtrA from plasmid SLP2, whereas that of strain DSM40697, which is truncated, shares 83% identity with TtrA from the chromosome of *S. avermitilis*. In addition, the SAMT0010 protein is similar to the KilB protein from different *Streptomyces* plasmids, pRL2 (linear) and pIJ101 (circular), in which it is implicated in conjugal transfer and intramycelial spread (36).

Mobile genetic element-related genes. Genomic islands often carry genes implicated in mobility, such as those encoding integrases, recombinases, and transposases (15). Thus, the presence of such genes in variable regions supports the idea of acquisition by horizontal gene transfer. In the conserved part of the TIRs, 149 pairs of orthologues between the two *S. ambofaciens* strains were predicted. Among them, three gene products showing similarity with transposases (DSMT0057/SAMT0012, DSMT0058/SAMT0013, and DSMT0214/SAMT0194) and one with an integrase/recombinase (DSMT0060/SAMT0015) were annotated. For three of them, the best similarity is found with transposase (or integrase) from *Frankia* species (*Actinomycetales*). Two transposase ORFs are located in a conserved region of the TIRs, just at the borders of the strain-specific regions A and a. The third one is located at the internal boundary of the TIRs as described above. However, none of the transposase-encoding genes seem to constitute a functional IS, the transposase being either inactivated by frameshift mutations or truncated, with no detectable flanking inverted repeats. The terminal strain-specific region A (DSM40697) contains a truncated IS (DSMT0046) similar to an IS from *Frankia* sp. strain Cci3 and a phage integrase (DSMT0003), close to the telomeres, sharing no homology to *Streptomyces* but with homology to *Nocardioides* species (see Table 2 for more details). Reciprocally, in the strain ATCC23877, one truncated IS (SAMT0032) is present in strain-specific region b.

Interestingly, many transposase-encoding genes are found close to the ancestral boundary of the TIRs (four in strain DSM40697 and three in strain ATCC23877), and this is a common feature of *Streptomyces* replicons. In *S. coelicolor* A3(2), an IS constitutes the ends of the chromosomal TIRs (6), while in plasmid SCP1, Tn*5714* is located 3 kb outside the left TIR and IS*466* is located at the end of the right one (5).

Given the close relationship between *S. ambofaciens* strains, it is even possible to spot recent IS- or transposon-mediated rearrangements. Indeed, outside of the TIRs, a putative com-

TABLE 2. Strain-specific genes predicted in the TIRs of *S. ambofaciens* strains DSM40697 and ATCC23877

Strain	Specific region	Gene	Product[a]	Identity (%)[b]	Overlap (%)[c]	Gene name	Plasmid	Organism
DSM40697	A	DSMT0001	Unknown					
		DSMT0002	Putative transcriptional regulator					
		DSMT0003	Putative phage integrase	35	93	NocaDRAFT_4522		*Nocardioides* sp. strain JS614
		DSMT0004	Unknown					
		DSMT0005	Putative serine/threonine protein kinase					
		DSMT0006	Unknown					
		DSMT0007	Unknown					
		DSMT0008	Putative helicase	83	88	SAV7571		*Streptomyces avermitilis*
		DSMT0009	Putative ATP-dependent DNA ligase	64	98	SAP1_90 (*lig*)	SAP1	*Streptomyces avermitilis*
		DSMT0010	Putative integral membrane transport protein	72	100	SCO6809		*Streptomyces coelicolor*
		DSMT0011	Putative secreted protein	75	100	SCO6811		*Streptomyces coelicolor*
		DSMT0012	Putative FAD-dependent oxidoreductase	58	98	pREL1_0108	pREL1	*Rhodococcus erythropolis*
		DSMT0013	Putative transcriptional regulator	95	96	pFQ25.11		*Streptomyces* sp. strain F2
		DSMT0014	Putative phosphinothricin *N*-acetyltransferase	77	99	pFQ25.10		*Streptomyces* sp. strain F2
		DSMT0015	Unknown					
		DSMT0016	Conserved hypothetical protein	85	99	pnf1840	pNF1	*Nocardia farcinica*
		DSMT0017	Putative membrane protein	76	100	SCO3280		*Streptomyces coelicolor*
		DSMT0018	Putative glycosyl hydrolase, BNR repeat	36	98	ArthDRAFT_2101		*Arthrobacter* sp. strain FB24
		DSMT0019	Putative lipoprotein	72	100	SCO4458		*Streptomyces coelicolor*
		DSMT0020	Putative membrane protein	60	96	SCO4459		*Streptomyces coelicolor*
		DSMT0021	Putative monooxygenase	78	94	SCO6838		*Streptomyces coelicolor*
		DSMT0022	Putative arsenic resistance membrane transport protein	84	99	SCO6837		*Streptomyces coelicolor*
		DSMT0023	Putative transcriptional regulator	82	100	SCO3699		*Streptomyces coelicolor*
		DSMT0024	Putative arsenate reductase	87	95	SCO6835		*Streptomyces coelicolor*
		DSMT0025	Putative thioredoxin reductase	79	99	SCO6834		*Streptomyces coelicolor*
		DSMT0026	Conserved hypothetical protein	25	75	Tfu_2935		*Thermobifida fusca*
		DSMT0027	Unknown					
		DSMT0028	Conserved hypothetical protein	87	100	SCP1.257	SCP1	*Streptomyces coelicolor*
		DSMT0029	Conserved hypothetical protein	36	83	SAP1_88	SAP1	*Streptomyces avermitilis*
		DSMT0030	Putative RNA polymerase sigma factor	37	97	SAP1_87 (*sig*)	SAP1	*Streptomyces avermitilis*
		DSMT0031	Conserved hypothetical protein	35	93	SAP1_86	SAP1	*Streptomyces avermitilis*
		DSMT0032	Putative secreted protein	84	100	SCP1.323c	SCP1	*Streptomyces coelicolor*
		DSMT0033	Putative secreted protein	89	100	SCP1.261c	SCP1	*Streptomyces coelicolor*
		DSMT0034	Conserved hypothetical protein	84	94	SCP1.262	SCP1	*Streptomyces coelicolor*
		DSMT0035	Ttra helicase fragment (pseudogene)	80	95	SAV7571		*Streptomyces avermitilis*
		DSMT0036	Conserved hypothetical protein DUF1099	63	95	ShewDRAFT_1466		*Shewanella* sp. strain PV-4
		DSMT0037	Unknown					
		DSMT0038	Conserved hypothetical protein (pseudogene)	70	99	SCO0085		*Streptomyces coelicolor*
		DSMT0039	Putative ATP-dependent DNA ligase	59	100	SAP1_90 (*lig*)	SAP1	*Streptomyces avermitilis*
		DSMT0040	Conserved hypothetical protein	71	100	SCO0048		*Streptomyces coelicolor*
		DSMT0041	Unknown					
		DSMT0042	Conserved hypothetical protein	29	95	SCP1.202	SCP1	*Streptomyces coelicolor*
		DSMT0043	Conserved hypothetical protein	84	100	SCP1.201	SCP1	*Streptomyces coelicolor*
		DSMT0044	Putative secreted protein	94	100	SCP1.200c	SCP1	*Streptomyces coelicolor*
		DSMT0045	Putative secreted esterase	93	100	SCP1.199c	SCP1	*Streptomyces coelicolor*
		DSMT0046	Putative transposase	55	71	Francci3_3385		*Frankia* sp. strain CcI3
		DSMT0047	Putative transporter	55	93	nfa29650		*Nocardia farcinica*
		DSMT0048	Tpg protein fragment (pseudogene)	75	97	*tpgSLP2*	SLP2	*Streptomyces lividans*
		DSMT0049	Putative AraC-family transcriptional regulator (pseudogene)	82	97	SCO3804		*Streptomyces coelicolor*
		DSMT0050	Conserved hypothetical protein	73	48	SCO3803		*Streptomyces coelicolor*
		DSMT0051	Putative nourseothricin acetyltransferase	69	100	*nat1*		*Streptomyces noursei*
		DSMT0052	Conserved hypothetical protein	30	93	Francci3_1866		*Frankia* sp. strain CcI3
		DSMT0053	Conserved hypothetical protein	47	76	Francci3_1863		*Frankia* sp. strain CcI3
		DSMT0054	Unknown					
		DSMT0055	Putative peptidase	38	98	SCO3610		*Streptomyces coelicolor*
		DSMT0056	Putative major facilitator superfamily	29	93	NocaDRAFT_1725		*Nocardioides* sp. strain JS614
	B	DSMT0083	Putative 3-oxoacyl-ACP synthase III	40	100	ArthDRAFT_2448		*Arthrobacter* sp. strain FB24
	C	DSMT0085	Putative ABC transport system ATP-binding protein	64	93	SCO0121		*Streptomyces coelicolor*
		DSMT0086	Putative integral membrane protein	44	93	SCO0120		*Streptomyces coelicolor*
		DSMT0087	Conserved hypothetical protein	46	76	Tfu_1509		*Thermobifida fusca*
		DSMT0088	Conserved hypothetical protein	31	98	SCO0123		*Streptomyces coelicolor*
		DSMT0089	Putative polyprenyl synthetase	61	98	SCO0568		*Streptomyces coelicolor*

Continued on following page

TABLE 2—*Continued*

Strain	Specific region	Gene	Product[a]	Identity (%)[b]	Overlap (%)[c]	Gene name	Plasmid	Organism
		DSMT0090	Putative geranylgeranyl diphosphate synthase	48	100	*ggdps*		*Streptomyces* sp. strain KO-3988
	D	DSMT0118	Conserved hypothetical protein	84	79	SCP1.218c	SCP1	*Streptomyces coelicolor*
	E	DSMT0134	Unknown					
ATCC23877	a	SAMT0001	Conserved hypothetical protein	62	92	SAV7573		*Streptomyces avermitilis*
		SAMT0002	Putative helicase	84	100	*ttrA*	SLP2	*Streptomyces lividans*
		SAMT0003	Unknown					
		SAMT0004	Conserved hypothetical protein	42	100	pFRL1.57	pFRL1	*Streptomyces* sp. strain FR1
		SAMT0005	Unknown	64	59	pFRL1.57	pFRL1	*Streptomyces* sp. strain FR1
		SAMT0006	Putative NTP pyrophosphohydrolase	39	67	PFL_4894		*Pseudomonas fluorescens*
		SAMT0007	Conserved hypothetical protein	52	95	nfa38470		*Nocardia farcinica*
		SAMT0008	Putative glyoxalase	64	96	nfa38460		*Nocardia farcinica*
		SAMT0009	Putative ferredoxin NADPH reductase	55	97	nfa38450		*Nocardia farcinica*
		SAMT0010	KilB-like protein, role in intramycelial spread	48	95	pRL2.23	pRL2	*Streptomyces* sp. strain 44414
		SAMT0011	Putative acetyltransferase	50	100	SAV2967		*Streptomyces avermitilis*
	b	SAMT0023	Unknown					
		SAMT0024	Putative hydrolase	53	90	Adeh_0548		*Anaeromyxobacter dehalogenans*
		SAMT0025	Conserved hypothetical protein	38	100	SCO7248		*Streptomyces coelicolor*
		SAMT0026	Putative phosphotransferase	54	91	*ard*		*Streptomyces avermitilis*
		SAMT0027	Unknown					
		SAMT0028	Putative secreted protein	47	80	SCO0072		*Streptomyces coelicolor*
		SAMT0029	Putative secreted protein	61	89	SCO0072		*Streptomyces coelicolor*
		SAMT0030	Unknown	30	67	Francci3_0808		*Frankia* sp. strain CcI3
		SAMT0031	Unknown					
		SAMT0032	Putative truncated transposase	71	80	*orf118*	pSLA2-L	*Streptomyces rochei*
		SAMT0033	Putative secreted protein	58	97	SCO0072		*Streptomyces coelicolor*
		SAMT0034	Putative secreted protein	86	100	SCO0072		*Streptomyces coelicolor*
		SAMT0035	Conserved hypothetical protein	60	100	SCO0073		*Streptomyces coelicolor*
		SAMT0036	Putative SAM-dependent methyltransferase	67	98	SCO2653		*Streptomyces coelicolor*
		SAMT0037	Conserved hypothetical protein	82	100	SCO0031		*Streptomyces coelicolor*
		SAMT0038	Putative transmembrane restriction endonuclease	87	100	SCO7763		*Streptomyces coelicolor*
		SAMT0039	Unknown					
		SAMT0040	Putative stress response protein	77	100	SCO3763		*Streptomyces coelicolor*
		SAMT0041	Unknown	54	48	Francci3_1126		*Frankia* sp. strain CcI3
		SAMT0042	Putative membrane protein	37	100	SAV3190		*Streptomyces avermitilis*
		SAMT0043	Unknown					
		SAMT0044	Putative regulator	39	81	SAV1103		*Streptomyces avermitilis*
		SAMT0045	Putative anti-sigma factor antagonist (pseudogene)	42	81	SCO3692		*Streptomyces coelicolor*
		SAMT0046	Putative regulator	71	96	*prpC3*		*Streptomyces avermitilis*
		SAMT0047	Putative hydrolase	79	100	SAV923		*Streptomyces avermitilis*
		SAMT0048	Putative urease beta/gamma subunit	50	94	DR_A0319		*Deinococcus radiodurans*
		SAMT0049	Putative urease alpha subunit	65	98	SCO1234		*Streptomyces coelicolor*
		SAMT0050	Putative *ureF*-like urease accessory protein	76	100	*ureF*		*Streptomyces avermitilis*
		SAMT0051	Putative *ureG*-like urease accessory protein	75	98	*ureG*		*Streptomyces avermitilis*
		SAMT0052	Putative *ureD*-like urease accessory protein	54	81	SCO1231		*Streptomyces coelicolor*
	c	SAMT0065	Putative truncated transposase	41	100	SAV18		*Streptomyces avermitilis*
		SAMT0066	Putative haloacid dehalogenase	89	99	SAV737		*Streptomyces avermitilis*
	d	SAMT0071	Putative transcriptional regulator	41	96			*Saccharopolyspora erythraea*
		SAMT0072	Putative esterase	38	100	SCO4392		*Streptomyces coelicolor*

[a] FAD, flavin adenine dinucleotide; BNR, bacterial neuraminidase repeat; ACP, acyl carrier protein; NTP, nucleoside triphosphate.
[b] Results for the best BLASTP hit are summarized.
[c] "Overlap" corresponds to the ratio between the length of the BLASTP alignment and the length of the query protein (for cases in which this ratio was >100%, 100% overlap was indicated).

posite transposon that is absent from the DSM40697 strain was detected in strain ATCC23877 (SAMR0213 to SAMR0218 [complete region h in Fig. 1B]). This transposon (5.1 kb) consists of two almost identical IS elements (99% nucleotide identity) flanking four CDSs, the products of two of which are also related to IS transposases, plus two orphans. Interestingly, the flanking ISs do not share any homology with *Streptomyces* sequences but do share homology with a transposase from *Frankia* sp. strain EAN1pec (67% amino acid identity).

Coding density and pseudogenes. Acquisition of DNA by horizontal gene transfer and loss of useless genes are the main causes of intraspecific variability, and an equilibrium between these two phenomena leads to genomic flux that shapes bacterial genomes (24). The first step of gene loss is the creation of pseudogenes either by point mutations or by truncations.

In *S. ambofaciens*, the TIRs are characterized by a low coding density. When considering pseudogenes as noncoding DNA, the coding densities of the sequenced regions are

77.1% and 75.7% for strains DSM40697 and ATCC23877, respectively. These values are not biased significantly by the presence of pseudogenes, since the coding densities are 80.1% (DSM40697) and 77.1% (ATCC23877) when pseudogenes are included as CDSs. These values contrast with values of 88.9% and 86.2% predicted for the complete chromosomes of *S. coelicolor* (6) and *S. avermitilis* (19), respectively. This is again a common trait of *Streptomyces* plasmids, such as SAP1 from *S. avermitilis*, for which the density falls to 79% (19).

Another striking feature of the terminal regions is the strong presence of pseudogenes, notably in the TIRs of strain DSM40697 (10, including two truncated transposases, representing 5% of the gene content). In contrast, pseudogenes represent less than 1% of the CDSs predicted in the whole genome of *S. coelicolor* (6). Eight out of the 10 pseudogenes are carried by the strain-specific region A. In addition to the truncation of the *ttrA* gene (DSMT0008; 219 amino acid residues out of 835 in *S. avermitilis*), two additional TtrA fragments are encoded in strain-specific regions A (DSMT0035; 44 residues [Table 2]) and G (DSML0224; 31 residues), outside of the TIRs. The first two fragments are homologous to different parts of the same gene (*ttrA* from *S. avermitilis*). However, the latter pseudogene is not related to them, suggesting acquisitions of extra copies by different integrations of parts of linear replicons. This hypothesis is further supported by the finding, in the same region A, of a gene fragment (DSMT0048; 99 bp) showing best identity with *tpgC* encoding the terminal protein of the linear plasmid SLP2 and of three gene fragments (DSMT0009, DSMT0038, and DSMT0039) similar to the *lig* gene encoding a ligase from the *S. avermitilis* linear plasmid SAP1 (SAP1_90).

In the regions sequenced outside of the TIRs, 10 additional probable pseudogenes were predicted for strain DSM40697, of which 6 belong to the strain-specific regions.

The number of pseudogenes is probably underestimated, and the low coding density could be a consequence of a high mutation rate. Altogether, these data support the hypothesis that the TIRs and the terminal regions of the genome constitute a hot spot for horizontal gene transfer events mediated by linear plasmids.

Horizontal transfer of accessory genes. Horizontal transfer mostly involves accessory genes that are able to confer a selective advantage to the recipient cell. Housekeeping genes are more recalcitrant to transfer (23). In *S. ambofaciens*, functions of many genes predicted for the strain-specific regions, such as resistance to toxic compounds, are related to adaptation to the environment. They are more particularly abundant in the terminal specific region A of strain DSM40697 (Fig. 1B). A good example is the five-gene cluster DSMT0021-25, which is highly similar (from 78% to 87% amino acid identity) to the *S. coelicolor* cluster SCO6838-34, which is implicated in the transport of and resistance to arsenate.

Antibiotic resistance genes are also present. For example, the DSMT0051 gene product shows best similarity (69%) with the *Streptomyces noursei* nourseothricin acetyltransferase, Nat1 (22). In the same way, a homologue of the Ard2-encoding gene (SAMT0026; 54% identity) from *Saccharothrix mutabilis* subsp. *capreolus* (*Actinomycetales*), which confers resistance to A201A antibiotic (3), is present in the ATCC23877 strain-specific chromosomal end.

The DSM40697 strain-specific regions B and C also carry functions related to adaptation. Indeed, the seven specific genes identified here (DSMT0083, DSMT0085-90) all have associations with secondary metabolism: DSMT0083 is a probable 3-oxoacyl-ACP synthase III-encoding gene having homologues in different *Actinomycetales* spp.; DSMT0085/86/88 are conserved, with three membrane and transport protein-encoding genes of *S. coelicolor* adjacent to the eicosapentaenoic acid cluster (SCO0124-0129); DSMT0087 shares similarity with a CDS of unknown function located in the fredericamycin biosynthesis gene cluster from *Streptomyces griseus* (44); and DSMT0089/90 are similar to genes encoding a putative polyprenyl synthase and a geranylgeranyl diphosphate synthase, respectively, from *Streptomyces* sp. strain KO-3988 (21). Interestingly, although these genes all seem to be implicated in secondary metabolism, they do not correspond to a whole conserved cluster but rather show similarity to genes from many different clusters among *Actinomycetales*. This could therefore be an example of a locus created by the association of genes derived from different pathways. Examples of transfer of secondary metabolism clusters have already been discussed in reference 27. The *alp* polyketide synthase cluster present in the TIRs of the two *S. ambofaciens* strains could also constitute an example of a chimeric cluster (31). While its left part is mostly similar to the kinamycin-biosynthetic cluster of *S. murayamaensis* (accession number AY228175), most of the right part shows high similarity to and the same genetic organization as a locus identified in the *S. rochei* linear plasmid pSLA2-L (28).

In addition, the variable region C is replaced by region d in strain ATCC23877 (Fig. 1B), in which one of the two specific CDSs identified encodes a probable transcriptional regulator sharing best identity with the OrfD regulator belonging to a cluster involved in the biosynthesis of a pigment in *Saccharopolyspora erythraea* (10).

However, a significant part of the strain-specific genes have no known function and are, in some cases, orphans.

Species specificity of the *S. ambofaciens* TIR conserved genes. The part of the TIRs conserved at the intraspecific level is highly variable at the interspecific level. No synteny can be observed the other *Streptomyces* with genomes. The regions conserved with *S. coelicolor* are limited to eight small syntenic clusters comprising two to seven CDSs (e.g., a urea degradation cluster). In contrast to what is found for *S. avermitilis*, only one cluster of four CDSs is syntenic.

Seventeen percent of the 149 pairs of orthologues between the two strains do not have any homologue in the NR database. In addition, among the 37 proteins showing best similarity with those from organisms other than *Streptomyces*, the majority (19/37) show highest similarity with those from other *Actinomycetales* genera (especially *Frankia*, *Arthrobacter*, *Nocardia*, *Nocardioides*, and *Kineococcus*). Many of these organisms are fellow soil-dwelling bacteria, which plausibly suggests lateral gene transfer events.

DISCUSSION

Acquisition of new functions by exchange of extremities between linear replicons. The *S. ambofaciens* TIRs contain a large proportion of strain- and species-specific genes as well as

sequences potentially involved in genome plasticity. Genetic organization of the TIRs is consistent with the idea that DNA rearrangements of endogenous sets of genes (duplication and translocation), integration of exogenous information, and/or deletions have been fixed during evolution from their common ancestor.

Several lines of evidence corroborate the acquisition of the terminal strain-specific regions by exchange of replicon extremities with linear plasmids: the presence of gene clusters homologous to plasmid-associated ones; the presence of genes implicated in conjugal transfer (*kilB* and *ttrA*); and their low G+C content (68.8% and 69.2%), which is characteristic of many *Streptomyces* plasmids. In addition, many similarities are found with genes carried by the extremities of the linear plasmids (*ttrA*, *tpgC*, *lig*, and *kilB*). Since the last telomere-proximal ORF conserved in the two strains is a truncated IS, it is plausible that exchange of extremities could have happened by homologous recombination involving two IS copies.

This comparison made by use of *S. ambofaciens* describes for the first time such exchanges at the origin of intraspecific variability in natural isolates. Insertion of DNA extremities of linear replicons may be a favored mechanism for gene acquisition in *Streptomyces*. Indeed, exchange of the terminal parts requires a single crossover event, the success of which would be guaranteed by the presence of the telomeres in the two replicons. In addition, the presence of a helicase-like gene (*ttrA*) closely associated with the telomere is a trait common to almost all *Streptomyces* linear replicons. However, it is truncated in strain DSM40697, and no deleterious effect could be assigned in laboratory growth conditions to the mutation of both copies of *ttrA* in *S. lividans* (17). The strong conservation of a gene in a variable region appears paradoxical, but it suggests a role for conferring some long-term advantage, e.g., conjugal transfer, as suggested by C.W. Chen in his "end-first" model (8). This model predicts that this probable helicase would be involved in conjugal transfer by acting on the DNA terminus which would correspond to an origin of transfer (7, 8).

Many of the functions predicted for the strain-specific regions are potentially related to adaptation, which consolidates the idea that horizontal transfer is at the origin of the *S. ambofaciens* terminal variability. In general, genes successfully transferred are responsible for adaptation to the environment. The presence in the DSM40697 strain chromosomal extremities of multiple resistance genes may have conferred advantages responsible for their maintenance in the bacterial population.

In addition to exchange of extremities, multiple events of insertions/deletions have occurred recently both inside and outside of the TIRs. For example, specific region b of the strain ATCC23877 chromosome could be the result of an integration event, but the absence of a detectable target site suggests either integration by illegitimate recombination or deletion of this locus in strain DSM40697. The specific regions C (DSM40697) and d (ATCC23877), which are two different variable regions located at the same locus, could be the consequence of a DNA exchange by double crossover or by different deletions at the same locus in the two strains.

All these multiple rearrangements are posterior to the formations of the TIRs that are themselves totally different from those of the very close *S. coelicolor* genome. Thus, genome plasticity is extremely strong in the terminal regions, which could be privileged targets for environmental adaptation by loss, acquisition, and creation of new functions. They might be considered as a vector for genetic transfer and also as the "Swiss army knife" of the "boy scout" *Streptomyces* (6, 16).

Origin and evolution of the size of the TIRs. Although the presence of two telomeres is known to be essential for the maintenance of chromosome linearity (2), no role has so far been attributed to the TIRs. Indeed, some replicons, such as the *S. avermitilis* chromosome, have TIRs restricted to a part of the telomeres, while others carry very large TIRs (up to 1.4 Mb [46]). In addition, the duplication of the genes located in the TIRs does not appear to be a mechanism for gene regulation. Thus, there is no obvious difference between the transcriptional levels of the genes duplicated in the *S. coelicolor* M600 TIRs (1 Mb) and those for the same set of genes present in single copy in *S. coelicolor* M145 (22-kb TIRs [43]).

The formation of TIRs could be the consequence of terminal recombination induced either by dysfunction of a telomere or by formation of double-strand breaks (DSBs). Chromosomal rescue could then be achieved by recombination with a DNA fragment including a telomere similar to the endogenous one. This fragment may be a broken daughter chromatid (46) or a plasmidic or chromosomal extremity either endogenously present or resulting from horizontal gene transfer (e.g., conjugal transfer). DSB repair might also result from the break-induced replication as described previously for *Saccharomyces cerevisiae* (29). Finally, circularization can also trigger chromosomal rescue (reviewed in reference 25).

Analysis of the *S. ambofaciens* TIR boundary reveals a preliminary step in TIR shortening (Fig. 2A). Thus, while the strict border of the TIRs is located at different positions in the two strains, the imperfect duplication includes the same genes. Point mutations and variability in tandem repeated motifs contribute to the decrease in length of the TIRs. In *S. avermitilis*, the terminal repetitions are restricted to the telomeres (19), which suggests the loss of ancestral TIRs in this species. Indeed, detailed analysis of the subtelomeric sequences reveals putative traces of ancestral TIRs (Fig. 2B). The last predicted ORF, SAV7573, is in fact imperfectly duplicated at the other end of the chromosome (Fig. 2B), and a putative CDS, which we tentatively called SAV0, can be detected between the telomere and SAV1. SAV0 (372 bp) and SAV7573 (411 bp) share 87% nucleotide identity over most of the CDS length (271/311 nucleotides). Furthermore, although only weakly similar (38% amino acid identity), a duplicated helicase gene is also present at both ends of the chromosome (SAV6/SAV7571). SAV2 also shows similarity with the helicase-like gene (Fig. 2B) and could correspond to the duplication of the 5′ end of SAV6 (69% nucleotide identity). These data suggest that once formed, the TIRs may accumulate point or insertional mutations that result in divergence which might progressively reduce and even prevent the frequent homogenization between TIRs and consequently accelerate the rate of divergence. The TIRs would then tend to degenerate by progressive shortening. The divergence of the duplicated genes present in the TIRs may result in pseudogene formation or leave functional coding sequences evolving toward new functions (34). Alternatively, the loss of TIRs in *S. avermitilis* could result from a single recombination

event between the ancestral chromosome and an exogenous DNA molecule.

The reduction in TIR size may be balanced by recombination events leading to expansion. Enlarged TIRs have been reported for mutants of *S. griseus* and *S. ambofaciens* (12, 41). In both cases, recombination events occurring between duplicated genes (each copy located on a different arm and in a divergent orientation) led to a dramatic increase of the TIR size (from 210 kb to 480 kb and 850 kb in *S. ambofaciens* and from 24 kb to 450 kb in *S. griseus*). For *S. coelicolor*, TIR length variation has been shown to be associated with strain lineage (43). The formation of large TIRs could be ascribed to homologous recombination between transposed IS*110* copies. TIR shortening from 1.06 Mb to 22 kb was observed (43). It therefore seems possible that TIR formation corresponds to a by-product of chromosomal rescue mechanisms. However, once formed, terminal duplications would provide an appropriate substrate for homologous recombination and thus for DNA repair of DSBs occurring in the terminal parts of the *Streptomyces* linear chromosome. The fact that the *S. ambofaciens* strain-specific regions are duplicated on both arms in a given chromosome strongly suggests a homogenization mechanism. In other words, these data show that DNA rearrangements occurring within a copy of TIRs can be followed by homogenization of both arms, probably by intrachromosomal recombination between duplicated sequences.

ACKNOWLEDGMENTS

F.C. and A.G. were recipients of a grant from the "Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche" (M.E.N.E.S.R.). This research was supported by the PAI (ALLIANCE), the "ACI Microbiologie 2003" programs funded by the M.E.N.E.S.R., and the VIth PCRDT ("ActinoGen").

Many thanks are due to K. Chater, G. Chandra, and T. Kieser (John Innes Centre, Norwich, United Kingdom) for their warm welcome and their help in the development of the computational methods. Many thanks to B. Segurens (Génoscope, CNS) for her help. We are grateful to A. Hesketh (John Innes Centre, Norwich, United Kingdom) and Eriko Takano (University of Groningen, Groningen, The Netherlands) for critical reading of the manuscript.

REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
2. **Bao, K., and S. N. Cohen.** 2001. Terminal proteins essential for the replication of linear plasmids and chromosomes in *Streptomyces*. Genes Dev. **15:** 1518–1527.
3. **Barrasa, M. I., J. A. Tercero, and A. Jimenez.** 1997. The aminonucleoside antibiotic A201A is inactivated by a phosphotransferase activity from *Streptomyces capreolus* NRRL 3817, the producing organism. Isolation and molecular characterization of the relevant encoding gene and its DNA flanking regions. Eur. J. Biochem. **245:**54–63.
4. **Bartolome, B., Y. Jubete, E. Martinez, and F. de la Cruz.** 1991. Construction and properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives. Gene **102:**75–78.
5. **Bentley, S. D., S. Brown, L. D. Murphy, D. E. Harris, M. A. Quail, J. Parkhill, B. G. Barrell, J. R. McCormick, R. I. Santamaria, R. Losick, M. Yamasaki, H. Kinashi, C. W. Chen, G. Chandra, D. Jakimowicz, H. M. Kieser, T. Kieser, and K. F. Chater.** 2004. SCP1, a 356,023 bp linear plasmid adapted to the ecology and developmental biology of its host, *Streptomyces coelicolor* A3(2). Mol. Microbiol. **51:**1615–1628.
6. **Bentley, S. D., K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, and D. A. Hopwood.** 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature **417:**141–147.
7. **Bey, S. J., M. F. Tsou, C. H. Huang, C. C. Yang, and C. W. Chen.** 2000. The homologous terminal sequence of the *Streptomyces lividans* chromosome and SLP2 plasmid. Microbiology **146:**911–922.
8. **Chen, C. W.** 1996. Complications and implications of linear bacterial chromosomes. Trends Genet. **12:**192–196.
9. **Chen, C. W., C. H. Huang, H. H. Lee, H. H. Tsai, and R. Kirby.** 2002. Once the circle has been broken: dynamics and evolution of *Streptomyces* chromosomes. Trends Genet. **18:**522–529.
10. **Cortes, J., J. Velasco, G. Foster, A. P. Blackaby, B. A. Rudd, and B. Wilkinson.** 2002. Identification and cloning of a type III polyketide synthase required for diffusible pigment biosynthesis in *Saccharopolyspora erythraea*. Mol. Microbiol. **44:**1213–1224.
11. **Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27:**4636–4641.
12. **Fischer, G., B. Decaris, and P. Leblond.** 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.
13. **Fischer, G., A. Kyriacou, B. Decaris, and P. Leblond.** 1997. Genetic instability and its possible evolutionary implications on the chromosomal structure of *Streptomyces*. Biochimie **79:**555–558.
14. **Fischer, G., T. Wenner, B. Decaris, and P. Leblond.** 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. USA **95:**14296–14301.
15. **Hacker, J., and E. Carniel.** 2001. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. EMBO Rep. **2:**376–381.
16. **Hopwood, D. A.** 2003. The *Streptomyces* genome—be prepared! Nat. Biotechnol. **21:**505–506.
17. **Huang, C. H., C. Y. Chen, H. H. Tsai, C. Chen, Y. S. Lin, and C. W. Chen.** 2003. Linear plasmid SLP2 of *Streptomyces lividans* is a composite replicon. Mol. Microbiol. **47:**1563–1576.
18. **Hütter, R.** 1967. Systematik der *Streptomycete*. Karger Verlag, Basel, Switzerland.
19. **Ikeda, H., J. Ishikawa, A. Hanamoto, M. Shinose, H. Kikuchi, T. Shiba, Y. Sakaki, M. Hattori, and S. Omura.** 2003. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat. Biotechnol. **21:**526–531.
20. **Ishikawa, J., A. Yamashita, Y. Mikami, Y. Hoshino, H. Kurita, K. Hotta, T. Shiba, and M. Hattori.** 2004. The complete genomic sequence of *Nocardia farcinica* IFM 10152. Proc. Natl. Acad. Sci. USA **101:**14925–14930.
21. **Kawasaki, T., T. Kuzuyama, Y. Kuwamori, N. Matsuura, N. Itoh, K. Furihata, H. Seto, and T. Dairi.** 2004. Presence of copalyl diphosphate synthase gene in an actinomycete possessing the mevalonate pathway. J. Antibiot. (Tokyo) **57:**739–747.
22. **Krugel, H., G. Fiedler, C. Smith, and S. Baumberg.** 1993. Sequence and transcriptional analysis of the nourseothricin acetyltransferase-encoding gene *nat1* from *Streptomyces noursei*. Gene **127:**127–131.
23. **Lawrence, J. G., and H. Hendrickson.** 2005. Genome evolution in bacteria: order beneath chaos. Curr. Opin. Microbiol. **8:**572–578.
24. **Lawrence, J. G., and J. R. Roth.** 1999. Genomic flux: genome evolution by gene loss and acquisition in bacterial genomes, p. 263–289. *In* R. L. Charlebois (ed.), Organization of the prokaryotic genome. American Society for Microbiology, Washington D.C.
25. **Leblond, P., and B. Decaris.** 1999. "Unstable" linear chromosomes: the case of *Streptomyces*, p. 235–261. *In* R. L. Charlebois (ed.), Organization of the prokaryotic genome. American Society for Microbiology, Washington, D.C.
26. **Leblond, P., G. Fischer, F. X. Francou, F. Berger, M. Guérineau, and B. Decaris.** 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.
27. **Metsa-Ketela, M., L. Halo, E. Munukka, J. Hakala, P. Mantsala, and K. Ylihonko.** 2002. Molecular evolution of aromatic polyketides and comparative sequence analysis of polyketide ketosynthase and 16S ribosomal DNA genes from various *Streptomyces* species. Appl. Environ. Microbiol. **68:**4472–4479.
28. **Mochizuki, S., K. Hiratsu, M. Suwa, T. Ishii, F. Sugino, K. Yamada, and H. Kinashi.** 2003. The large linear plasmid pSLA2-L of *Streptomyces rochei* has an unusually condensed gene organization for secondary metabolism. Mol. Microbiol. **48:**1501–1510.
29. **Morrow, D. M., C. Connelly, and P. Hieter.** 1997. "Break copy" duplication: a model for chromosome fragment formation in *Saccharomyces cerevisiae*. Genetics **147:**371–382.
30. **Pandza, S., G. Biukovic, A. Paravic, A. Dadbin, J. Cullum, and D. Hranueli.** 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.
31. **Pang, X., B. Aigle, J. M. Girardet, S. Mangenot, J. L. Pernodet, B. Decaris,**

and P. Leblond. 2004. Functional angucycline-like antibiotic gene cluster in the terminal inverted repeats of the *Streptomyces ambofaciens* linear chromosome. Antimicrob. Agents Chemother. **48:**575–588.
32. **Pinnert-Sindico, S.** 1954. Une nouvelle espèce de *Streptomyces* productrice d'antibiotiques: *Streptomyces ambofaciens* n. sp. caractères culturaux. Ann. Inst. Pasteur (Paris) **87:**702–707.
33. **Rocha, E. P., and A. Danchin.** 2002. Base composition bias might result from competition for metabolic resources. Trends Genet. **18:**291–294.
34. **Roth, V., B. Aigle, R. Bunet, T. Wenner, C. Fourrier, B. Decaris, and P. Leblond.** 2004. Differential and cross-transcriptional control of duplicated genes encoding alternative sigma factors in *Streptomyces ambofaciens*. J. Bacteriol. **186:**5355–5365.
35. **Sakaguchi, K.** 1990. Invertrons, a class of structurally and functionally related genetic elements that includes linear DNA plasmids, transposable elements, and genomes of adeno-type viruses. Microbiol. Rev. **54:**66–74.
36. **Schully, K. L., and G. S. Pettis.** 2003. Separate and coordinate transcriptional control mechanisms link expression of the potentially lethal KilB spread locus to the upstream transmission operon on *Streptomyces* plasmid pIJ101. J. Mol. Biol. **334:**875–884.
37. **Sekine, M., S. Tanikawa, S. Omata, M. Saito, T. Fujisawa, N. Tsukatani, T. Tajima, T. Sekigawa, H. Kosugi, Y. Matsuo, R. Nishiko, K. Imamura, M. Ito, H. Narita, S. Tago, N. Fujita, and S. Harayama.** 2006. Sequence analysis of three plasmids harboured in *Rhodococcus erythropolis* strain PR4. Environ. Microbiol. **8:**334–346.
38. **Spatz, K., H. Kohn, and M. Redenbach.** 2002. Characterization of the *Streptomyces violaceoruber* SANK95570 plasmids pSV1 and pSV2. FEMS Microbiol. Lett. **213:**87–92.
39. **Suzek, B. E., M. D. Ermolaeva, M. Schreiber, and S. L. Salzberg.** 2001. A probabilistic method for identifying start codons in bacterial genomes. Bioinformatics **17:**1123–1130.
40. **Tatusov, R. L., D. A. Natale, I. V. Garkavtsev, T. A. Tatusova, U. T. Shankavaram, B. S. Rao, B. Kiryutin, M. Y. Galperin, N. D. Fedorova, and E. V. Koonin.** 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. **29:**22–28.
41. **Uchida, T., M. Miyawaki, and H. Kinashi.** 2003. Chromosomal arm replacement in *Streptomyces griseus*. J. Bacteriol. **185:**1120–1124.
42. **Volff, J. N., and J. Altenbuchner.** 2000. A new beginning with new ends: linearisation of circular chromosomes during bacterial evolution. FEMS Microbiol. Lett. **186:**143–150.
43. **Weaver, D., N. Karoonuthaisiri, H. H. Tsai, C. H. Huang, M. L. Ho, S. Gai, K. G. Patel, J. Huang, S. N. Cohen, D. A. Hopwood, C. W. Chen, and C. M. Kao.** 2004. Genome plasticity in *Streptomyces*: identification of 1 Mb TIRs in the *S. coelicolor* A3(2) chromosome. Mol. Microbiol. **51:**1535–1550.
44. **Wendt-Pienkowski, E., Y. Huang, J. Zhang, B. Li, H. Jiang, H. Kwon, C. R. Hutchinson, and B. Shen.** 2005. Cloning, sequencing, analysis, and heterologous expression of the fredericamycin biosynthetic gene cluster from *Streptomyces griseus*. J. Am. Chem. Soc. **127:**16442–16452.
45. **Wenner, T., V. Roth, B. Decaris, and P. Leblond.** 2002. Intragenomic and intraspecific polymorphism of the 16S–23S rDNA internally transcribed sequences of *Streptomyces ambofaciens*. Microbiology **148:**633–642.
46. **Wenner, T., V. Roth, G. Fischer, C. Fourrier, B. Aigle, B. Decaris, and P. Leblond.** 2003. End-to-end fusion of linear deleted chromosomes initiates a cycle of genome instability in *Streptomyces ambofaciens*. Mol. Microbiol. **50:**411–425.
47. **Yamasaki, M., and H. Kinashi.** 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.
48. **Zdobnov, E. M., and R. Apweiler.** 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics **17:** 847–848.

III- Evolution du génome des Streptomyces

Ces résultats sont rapportés dans :

Publication N°2

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.

Evolution of the terminal regions of the *Streptomyces* linear chromosome

Molecular Biology and Evolution, Déc. 2006, p. 2361-2369, Vol. 23, No 12

Cet article présente la comparaison des régions terminales de la souche de *S. ambofaciens* ATCC23877 avec le génome de *S. coelicolor* et *S. avermitilis*. Ces résultats ont permis d'émettre des hypothèses sur l'évolution des génomes des *Streptomyces* par accumulation d'événements d'insertions/délétions dans certaines régions.

Le programme de séquençage des régions terminales de la souche de *S. ambofaciens* ATCC23877 a été initié avec le Génoscope (Centre National de Séquençage, Evry) et l'Institut de Génétique et Microbiologie (J.-L. Pernodet, Evry). L'objectif de ce programme était le séquençage complet des régions délétées chez les mutants spontanés. La plus grande délétion terminale observée est de 2335 kb chez le mutant NSA857 dérivant de la souche de *S. ambofaciens* DSM40697 (Fischer *et al.* 1997a).

Le programme de séquençage a permis l'obtention d'une séquence de 1,54 Mb et 1,37 Mb respectivement pour le bras gauche et le bras droit. L'ensemble de ces séquences contient 2 532 CDS, incluant 43 pseudogènes. Ces séquences ont été annotées et analysées au sein de la plate-forme génomique mise en place au laboratoire par F. Choulet.

A- Conservation de la région centrale

La région centrale a été séquencée à un taux de recouvrement de 0,4 fois. Pour comparer les régions centrales de *S. ambofaciens* et *S. coelicolor*, les extrémités des BAC (BES pour BAC End Sequence) ont été alignées par BlastN avec la séquence du chromosome de *S. coelicolor* et a ainsi confirmé une forte conservation de cette région entre les deux espèces. Une région a été considérée comme conservée quand la distance entre deux BES est compatible avec la longueur de l'insert d'un BAC (entre 25 et 80 kb).

Le traitement de ces données par Dot-plot permet la visualisation de l'alignement des CDS entre deux génomes. Chaque point de Dot-plot représente la position d'une CDS d'un génome et de son homologue (meilleur score de BLASTP, au moins 30% d'identité sur plus de 80% de la longueur) chez l'autre génome. La présence d'une forme en X sur le graphique de Dot-plot est caractéristique d'inversions péricentriques (Fig. 40 A). Ainsi, depuis la divergence entre *S. ambofaciens* et *S. coelicolor*, deux inversions péricentriques ont été fixées. Ces inversions

correspondent à deux des quatre inversions observées entre *S. coelicolor* et *S. avermitilis* (Ikeda *et al.* 2003).

La comparaison avec les génomes de *S. avermitilis* et *S. scabies* a permis de placer ces inversions sur l'arbre phylogénétique (Fig. 40 B). Les chromosomes de *S. coelicolor* et *S. avermitilis* présentent quatre événements d'inversions tandis que les chromosomes de *S. ambofaciens* et *S. scabies* sont colinéaires (Fig. 40 C).

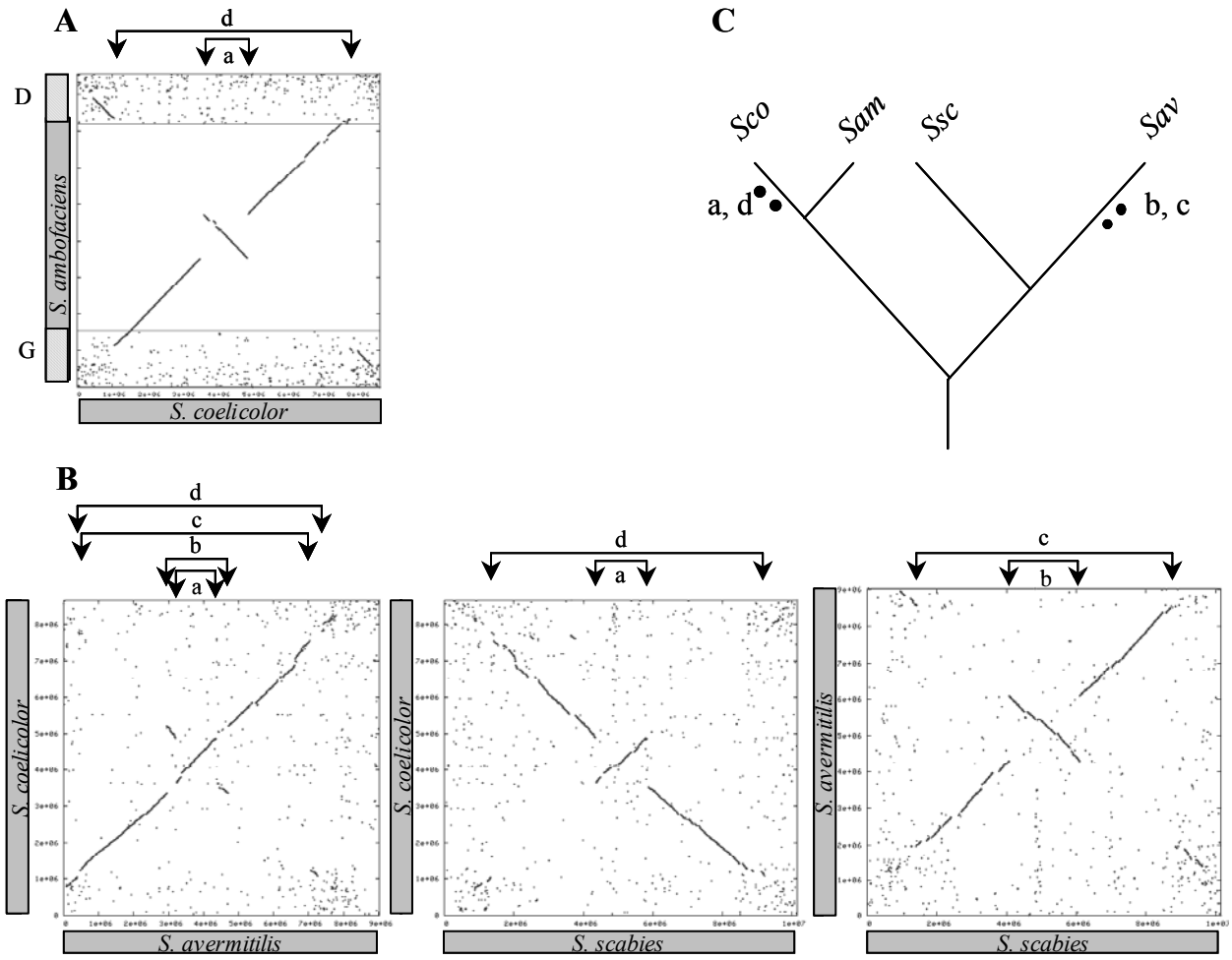


Figure 40: Comparaison par paires des chromosomes de *Streptomyces*.

A. Comparaison par dot plot du chromosome de *S. ambofaciens* et *S. coelicolor*. Pour *S. ambofaciens*, les rectangles hachurés représentent les bras chromosomiques totalement séquencés (G: bras gauche, D: bras droit). Chaque point représente la position d'une CDS de *S. ambofaciens* et de son homologue (meilleur score de BLASTP) chez *S. coelicolor*. L'alignement de la partie centrale du chromosome a été dessiné à partir de la comparaison des BES par BLASTN. Pour représenter l'alignement en deux dimensions, les critères suivants ont été utilisés: les régions identifiées comme conservées à partir des analyses de BES ont été considérées de même taille chez les deux espèces; comme la taille des régions spécifiques du chromosome de *S. ambofaciens* ne peut pas être déterminée par cette approche, une taille fixe de 80 kb (longueur maximale d'un insert) leur a été attribuée pour l'alignement.

B. Comparaison par dot plot des chromosomes de *S. coelicolor*, *S. avermitilis* et *S. scabies*. Les points représentent les CDS orthologues. Les inversions centrées sur l'origine de réplication sont indiquées par des flèches, numérotées.

C. positionnées sur l'arbre phylogénétique (en respectant le principe d'un maximum de parcimonie).

Le séquençage complet de trois BAC indépendants (117 kb pour 91 CDS prédites), pris au hasard, a permis de montrer la conservation de l'organisation de ces gènes et un haut degré d'identité avec les génomes d'autres *Streptomyces*. En moyenne, ces gènes prédits présentent respectivement 89%, 82% et 80% d'identité en acides aminés avec *S. coelicolor*, *S. avermitilis* et *S. scabies*.

Cependant, 12 régions spécifiques, situées dans la région centrale du chromosome, dont les tailles varient de 26 à 149 kb, sont présentes chez *S. coelicolor* et absentes chez *S. ambofaciens*. Parmi ces clusters, certains interviennent dans la synthèse d'antibiotiques comme par exemple l'actinorhodine (*act*), l'undécylprodigiosine (*red*) et l'antibiotique dépendant du calcium (*cda)* (calcium dependent antibiotic). Chez *S. ambofaciens*, quatre régions, au moins, sont absentes chez *S. coelicolor*, dont le cluster de biosynthèse de la spiramycine. En raison de la méthode utilisée pour l'alignement des BAC de la région centrale par rapport au génome de *S. coelicolor* (alignement à partir des résultats de BLASTN avec les BES de *S. ambofaciens* ATCC23877), des régions peuvent être absentes de *S. coelicolor* sans toutefois être détectées. Ainsi, les limites des régions spécifiques seront, pour les mêmes raisons, difficilement définissables.

Sur la base du taux en G+C et du contenu en gènes, Bentley *et al*. ont défini 14 régions potentiellement acquises récemment par transfert horizontal dans le chromosome de *S. coelicolor* (Bentley *et al.* 2002). La comparaison avec les données de *S. ambofaciens* confirme l'origine récente pour la moitié de ces régions (7/14), car elles sont absentes chez *S. ambofaciens*, mais rien ne peut exclure leur perte chez *S. ambofaciens* et leur maintien chez *S. coelicolor*. Ces régions contiennent les gènes de synthèse de l'actinorhodine (Malpartida et Hopwood 1986) ou encore de résistance aux métaux lourds.

B- Les régions spécifiques d'espèces par comparaison avec les autres génomes de Streptomyces

La comparaison des génomes de *S. coelicolor* et *S. avermitilis* a permis de montrer la localisation des régions spécifiques d'espèces (Ikeda *et al.* 2003). Ces régions sont confinées aux régions terminales. Cette caractéristique se confirme entre espèces proches comme *S. ambofaciens* et *S. coelicolor*. Cependant, la frontière de ces régions spécifiques d'espèces varie suivant le couple de génomes étudié. En effet, une frontière nette, au gène près, ne peut être déterminée. Le niveau de GOC (Gene Order Conservation) a été utilisé pour mesurer le niveau de synténie le long du chromosome. La valeur de GOC correspond au pourcentage de paires d'orthologues contigus communs aux deux génomes comparés. Quand la valeur de GOC est inférieure à 20%, les régions comparées sont définies comme non synténiques et donc spécifiques de chacune de ces deux espèces. Au-delà de 60%, ces régions comparées sont synténiques. Ainsi, si la valeur de GOC est comprise entre 20 et 60%, nous avons appelé ces régions, régions de synténie dégénérées.

Le calcul de la valeur de COG entre les régions terminales de *S. ambofaciens* et *S. coelicolor* indique que les régions terminales spécifiques de *S. ambofaciens* mesurent 1279 kb (619 et 660 kb pour respectivement les régions terminales droite et gauche) par comparaison avec *S. coelicolor* (Fig. 41). Par comparaison avec *S. avermitilis*, les régions terminales spécifiques de *S. ambofaciens* s'étendent à 1878 kb (889 kb et 989 kb) et 1883 kb (889 kb et 994 kb) par rapport à *S. scabies*.

Ainsi la taille des régions spécifiques est plus grande entre espèces phylogénétiquement éloignées. Réciproquement, la taille de la région centrale conservée, héritée de l'ancêtre commun, diminue avec la distance phylogénétique qui sépare les espèces comparées. Ceci suggère donc que l'acquisition de matériels génétiques spécifiques se réalise de façon préférentielle dans les extrémités du chromosome des *Streptomyces* et que ce flux de gènes, qui fait disparaître la synténie, serait relativement constant depuis l'ancêtre des *Streptomyces*.

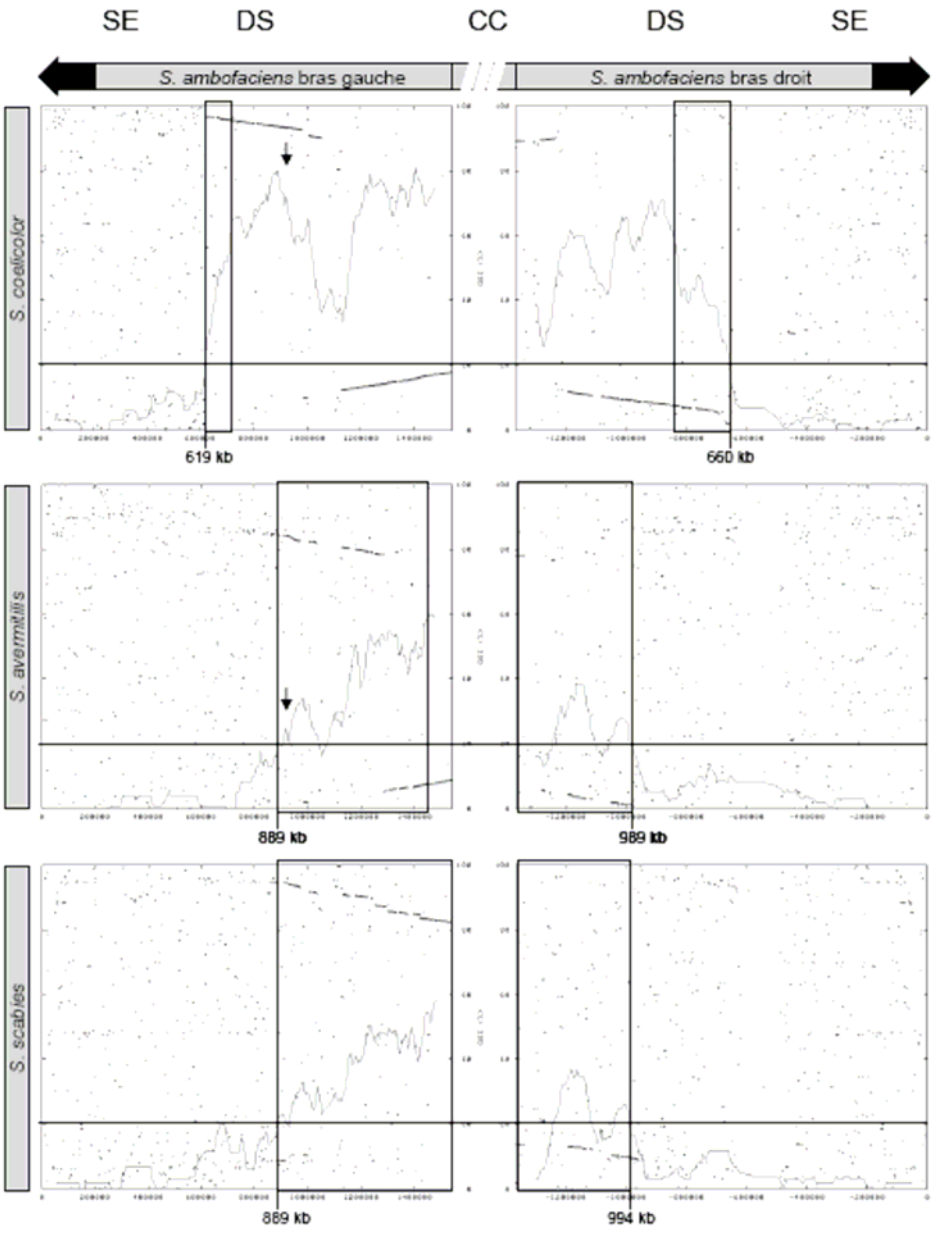


Figure 41 : Profil de GOC le long des bras chromosomiques gauche et droit de *S. ambofaciens* comparés aux génomes de *S. coelicolor*, *S. avermitilis* et *S. scabies*.
Sur chaque graphique, la courbe du GOC est superposée au dot-plot. Un seuil minimal de GOC de 20% a été choisi pour délimiter les régions terminales spécifiques d'espèces. SE : extrémités spécifiques ; DS : régions de synténie dégénérée ; CC : région centrale conservée. Les régions de synténie dégénérée (entre 20 et 60% de GOC) sont encadrées. Calculs réalisés avec une fenêtre glissante de 100 gènes et un pas de 5 gènes.

C- Proportion de CDS spécifiques dans les régions terminales

Dans les régions spécifiques de la souche de *S. ambofaciens* ATCC23877, où le niveau de GOC est faible, 1082 CDS sont prédites comme spécifiques des extrémités du chromosome de *S. ambofaciens* pour 1279 kb. Le pourcentage de spécificité et d'identité des CDS de *S. ambofaciens* avec *S. coelicolor* et *S. avermitilis* est représenté dans la figure 42.

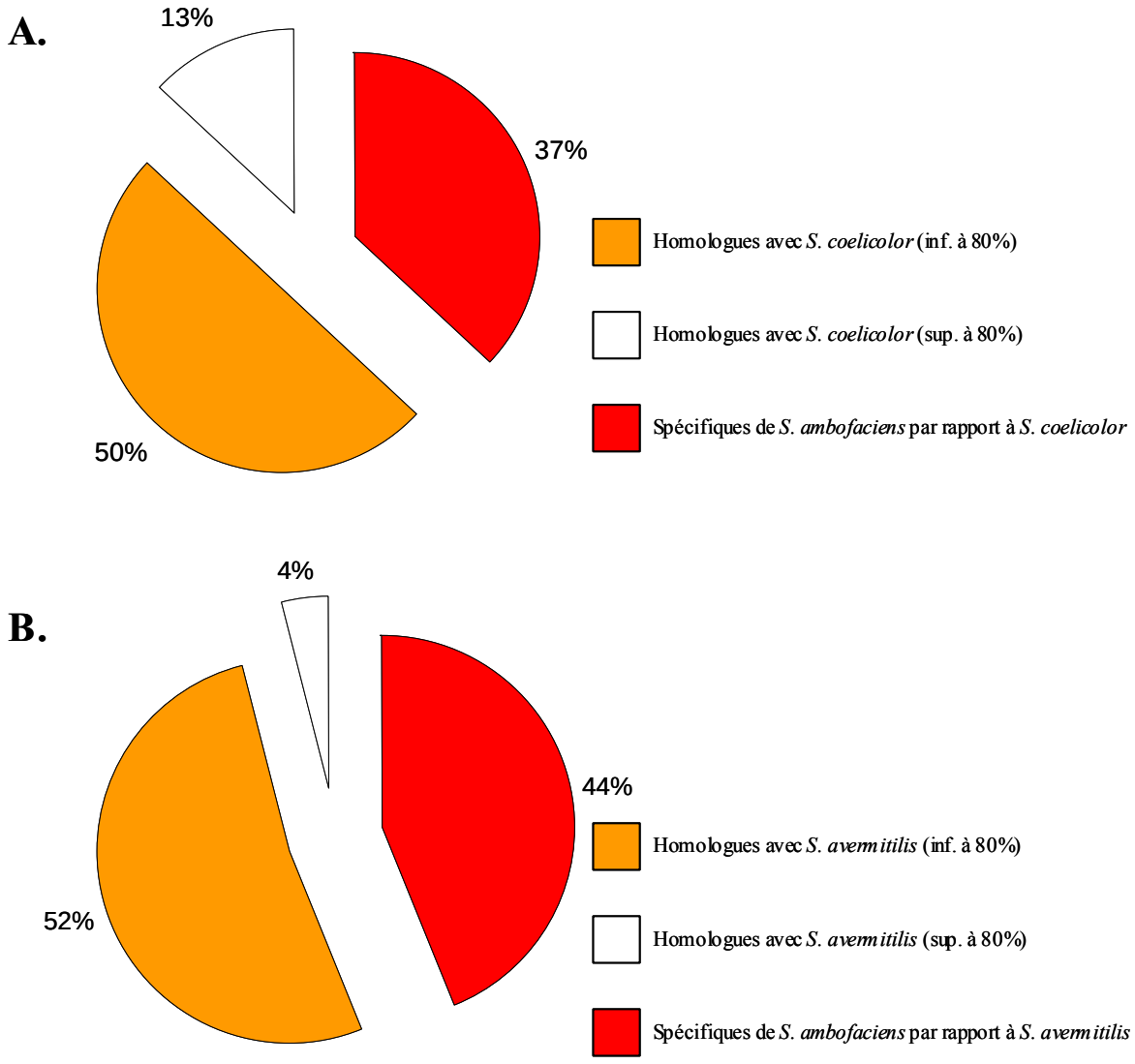


Figure 42 : Proportion de CDS spécifiques et homologues de *S. ambofaciens* par rapport à *S. coelicolor* et *S. avermitilis*

Les zones orange correspondent aux CDS présentant moins de 80% d'homologie avec *S. coelicolor* et *S. avermitilis*, les zones blanches plus de 80% d'homologies et en rouge la proportion de CDS spécifiques de *S. ambofaciens*.
A. Comparaison avec *S. coelicolor*
B. Comparaison avec *S. avermitilis*

Ainsi, la proportion de CDS spécifiques augmente avec l'éloignement phylogénétique. Ces valeurs contrastent avec celles retrouvées pour la région de 117 kb de la région centrale où 100% des CDS prédites sont des homologues forts.

D- Régions de synténie dégénérée et flux de gènes

Entre la région centrale conservée et les régions terminales spécifiques, des régions de synténie dégénérée ont été caractérisées. Ces régions de synténie dégénérée ont été définies quand la valeur de GOC est comprise entre 20% et 60%.

La synténie de la région centrale diminue progressivement sur plusieurs centaines de kilobases jusqu'à la perte totale de cette synténie (régions spécifiques). De multiples insertions/délétions, fixées au cours de la divergence, sont à l'origine de cette dégénérescence. Cette chute de la synténie ne révèle pas seulement une augmentation du niveau de réarrangements de l'information endogène mais aussi d'une évolution rapide du contenu génétique par accumulation de nouveaux gènes et perte de l'information ancestrale.

Le nombre de ruptures de synténie a été évalué à 10 depuis leur divergence sur une région de 100 kb (pour 99 CDS) sur le bras gauche de *S. ambofaciens* (SAML0798-SAML0896) vis-à-vis de celui de *S. coelicolor* (SC07238-SCO7327). La figure 43 illustre ce phénomène sur une région de 30 kb. En comparaison avec *S. avermitilis* et *S. scabies*, 29 et 25 ruptures respectivement sont observées. Le nombre de réarrangements est probablement sous-estimé pour des distances phylogénétiques importantes à cause d'un phénomène de saturations.

Le nombre de gènes orthologues dans cette région de 100 kb varie suivant la distance phylogénétique et est de 83, 52 et 49, respectivement, avec *S. coelicolor*, *S. avermitilis* et *S. scabies*, ceci soulignant l'importance du flux de gènes dans ces régions. Le nombre de paralogues augmente donc avec la distance phylogénétique.

Certains événements de délétions pourraient intéresser plus de 10 CDS, comme par exemple des clusters de gènes impliqués dans le métabolisme secondaire. La perte d'un seul gène de ce cluster favoriserait la perte de l'ensemble, celui-ci devenant probablement moins actif. Cette situation est observée dans les régions de synténie dégénérée. Par exemple, chez *S. ambofaciens*, un cluster de 34,5 kb, comprenant 28 gènes, du métabolisme secondaire est spécifique d'espèce. Chez *S. coelicolor*, cette région est remplacée par un cluster spécifique de 30,8 kb (31 gènes). Chez *S. avermitilis*, cette même région correspond à un cluster de gène de métabolisme secondaire *pks1* (Ikeda *et al.* 2003) différent des deux premiers.

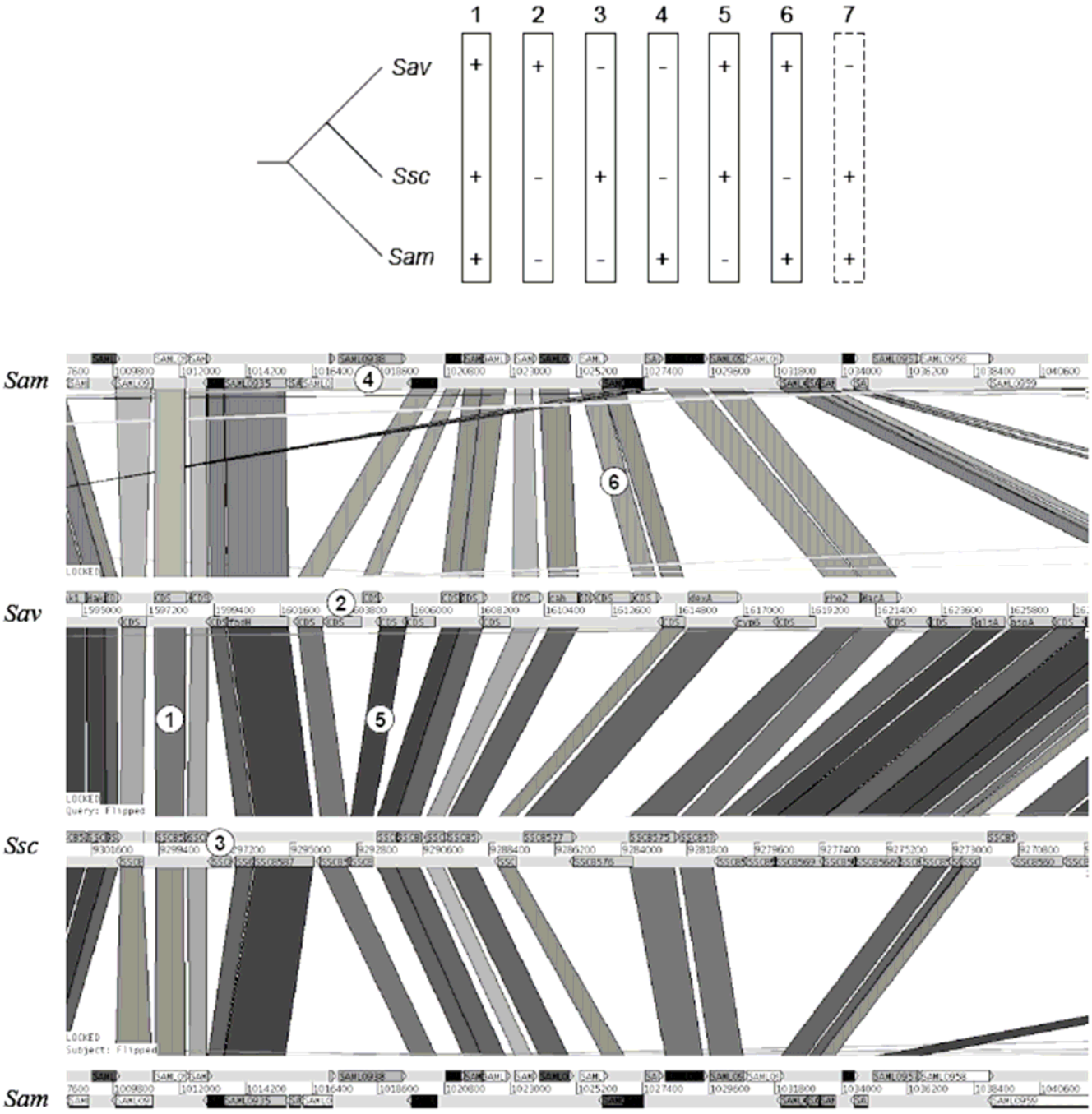


Figure 43 : Illustration de la synténie dégénérée sur une région d'environ 30 kb entre *S. scabies* (*Ssc*), *S. ambofaciens* (*Sam*) et *S. avermitilis* (*Sav*).
La région de *S. ambofaciens* a été représentée en bas et en haut de cette figure pour pouvoir visualiser la synténie avec les deux autres chromosomes. Toutes les situations possibles de partage de gènes entre les 3 espèces sont représentées sur l'arbre (+ : gène présent ; - : gène absent) et numérotées : 1. gène partagé entre les trois génomes, 2. gène spécifique de *Sav*, 3. gène spécifique de *Ssc*, 4. gène spécifique de *Sam*, 5. gène partagé entre *Sav* et *Ssc* et absent de *Sam*, 6. gène partagé entre *Sav* et *Sam* 7. gène partagé entre *Ssc* et *Sam*. Ces situations sont presque toutes retrouvées dans ces 30 kb. Seule la dernière n'est pas observable ici mais est présente dans d'autres régions du chromosome.

La taille des régions affectées par cette dégénérescence augmente avec l'éloignement phylogénétique. Ainsi, la région commune centrale entre deux espèces sera d'autant plus petite que ces deux espèces sont éloignées phylogénétiquement. Réciproquement, la taille des régions spécifiques de chacune de ces espèces augmente avec l'éloignement. Un autre indice GOC a été

évalué de façon à diminuer l'impact des insertions/délétions des îlots génomiques sur l'évolution des courbes GOC.

Le GOC_2 calcule le nombre de paires de gènes orthologues contigus dans deux génomes rapporté au nombre de gènes orthologues dans une fenêtre glissante de 300 CDS. Les résultats obtenus montrent que le niveau de conservation diminue graduellement vers les extrémités chromosomiques (inf. à 60%) alors que les régions centrales sont synténiques (Fig. 44).

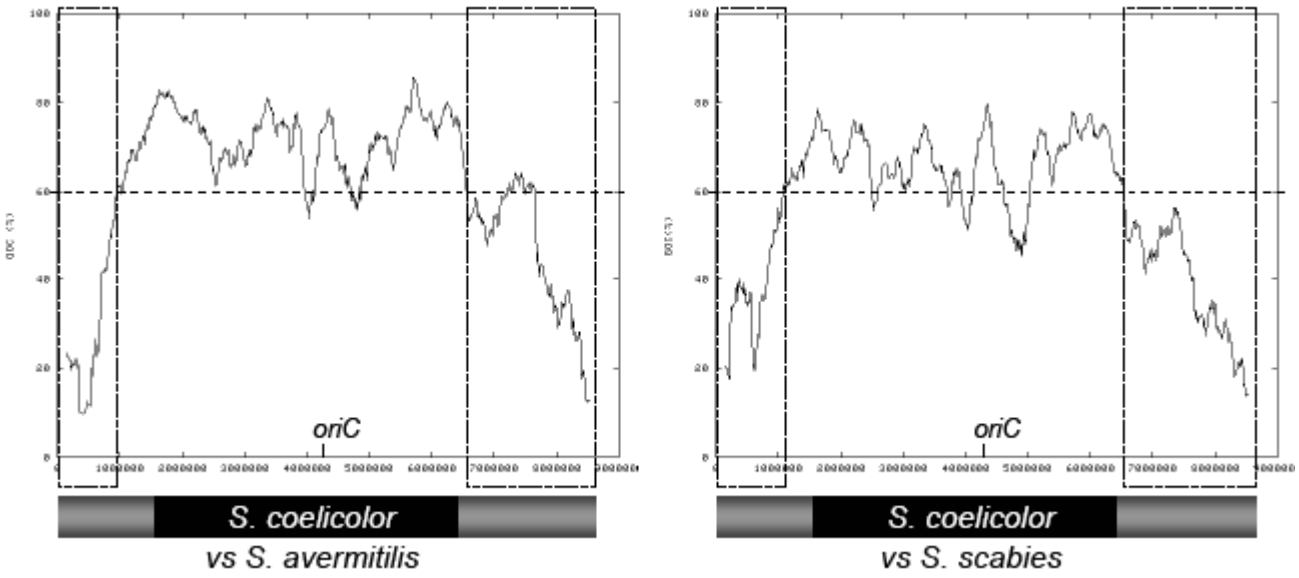


Figure 44 : Mesure du GOC_2 entre *S. coelicolor*-*S. avermitilis* et *S. coelicolor*-*S. scabies*
Les régions grisées représentent les régions du chromosome où la synténie diminue de façon graduelle jusqu'aux extrémités. Une valeur maximale de 60% GOC_2 a été choisie comme seuil minimal de conservation des régions de synténie. La région « core » et les extrémités (régions de contingence) définies par Bentley *et al.* 2002 sont représentées respectivement en noir et gris. Calculs réalisés avec une fenêtre glissante de 300 gènes et un pas de 10 gènes.

Evolution of the Terminal Regions of the *Streptomyces* Linear Chromosome

Frédéric Choulet, Bertrand Aigle,* Alexandre Gallois,* Sophie Mangenot,† Claude Gerbaud,‡ Chantal Truong,† François-Xavier Francou,‡ Céline Fourrier,* Michel Guérineau,‡ Bernard Decaris,* Valérie Barbe,† Jean-Luc Pernodet,‡ and Pierre Leblond**

*Laboratoire de Génétique et Microbiologie, UMR INRA 1128, IFR 110, Université Henri Poincaré Nancy 1, Faculté des Sciences et Techniques, Vandoeuvre-lès-Nancy, France; †Génoscope, Centre National de Séquençage, Evry, France; and ‡Institut de Génétique et Microbiologie, UMR CNRS 8621, Université Paris-Sud 11, Orsay, France

Comparative analysis of the *Streptomyces* chromosome sequences, between *Streptomyces coelicolor*, *Streptomyces avermitilis*, and *Streptomyces ambofaciens* ATCC23877 (whose partial sequence is released in this study), revealed a highly compartmentalized genetic organization of their genome. Indeed, despite the presence of specific genomic islands, the central part of the chromosome appears highly syntenic. In contrast, the chromosome of each species exhibits large species-specific terminal regions (from 753 to 1,393 kb), even when considering closely related species (*S. ambofaciens* and *S. coelicolor*). Interestingly, the size of the central conserved region between species decreases as the phylogenetic distance between them increases, whereas the specific terminal fraction reciprocally increases in size. Between highly syntenic central regions and species-specific chromosomal parts, there is a notable degeneration of synteny due to frequent insertions/deletions. This reveals a massive and constant genomic flux (from lateral gene transfer and DNA rearrangements) affecting the terminal contingency regions. We speculate that a gradient of recombination rate (i.e., insertion/deletion events) toward the extremities is the force driving the exclusion of essential genes from the terminal regions (i.e., chromosome compartmentalization) and generating a fast gene turnover for strong adaptation capabilities.

Introduction

Comparisons of complete genome sequences have revealed that the level of variability in bacteria is variable. This can be related to the life style of the organism. For bacteria living in a stable environmental niche, like the intracellular pathogens, the rate of genomic variation is low compared with free-living bacteria (Mira et al. 2002). For example, no rearrangement or gene acquisition has occurred in the genome of the obligate host-associated *Buchnera aphidicola* in the past 50–70 Myr (Tamas et al. 2002). In contrast, only 39.2% of the set of proteins from 3 *Escherichia coli* strains are common to them all (Welch et al. 2002). Genome reduction by gene loss is frequent for adaptation to a stable environment (Gil et al. 2002), whereas acquisition of useful functions by lateral gene transfer (LGT) plays an important role in the evolution of free-living bacteria (Ochman et al. 2000).

Although gene content can be highly different between related organisms, the structure of the chromosome is under strong selection and is highly organized. Beneath an apparent disorder, selective pressures to maintain information sets and valuable aspects of chromosome structure restrict the rate at which diversity can be added to a genome (Lawrence and Hendrickson 2005). Many organizational features, such as gene distribution and nucleotide composition (Lobry and Louarn 2003), are related to the replication process (Rocha 2004; Boccard et al. 2005). For example, gene dosage effects would constrain the position of genes along the genome (Couturier and Rocha 2006). The terminus of replication appears as a privileged target for DNA rearrangements (Suyama and Bork 2001). Thus, the control of the level of variability is dependent on location within the genome. In addition, genome size is constrained and implies a competition between genes for their maintenance. This competition is at the origin of genomic flux (Lawrence and Roth 1999).

Streptomyces are soil bacteria belonging to the Actinomycetales order. They present a complex cell cycle characterized by both morphological and biochemical differentiation processes (Chater 1993). They exhibit a remarkable phenotypic diversity typified by the diversity of the secondary metabolites produced and are consequently of great economic interest for applications in medicine, agriculture, and biotechnology. Their chromosome is linear with a central replication origin and among the largest in bacteria, ranging from 8.7 Mb in *Streptomyces coelicolor* (Bentley et al. 2002) to 10.1 Mb in *Streptomyces scabies* (http://www.sanger.ac.uk/Projects/S_scabies/). All *Streptomyces* species studied so far have been found to be subject to a high degree of genetic instability, correlated with the formation of large rearrangements (large-scale deletions and amplifications) occurring in the terminal chromosomal regions (Leblond and Decaris 1999). The frequent loss of the terminal regions (up to 2.3 Mb in *Streptomyces ambofaciens* [Fischer et al. 1997]) in laboratory growth conditions indicates that they do not contain genes essential for vegetative growth. This organization was corroborated by the analysis of the *S. coelicolor* chromosome in which all known essential genes are located in a central "core" region (4.9 Mb), whereas the chromosomal "arms" were defined as contingency (i.e., nonessential) regions (Bentley et al. 2002). The core corresponds to the region common to both genomes of *S. coelicolor* and of the actinomycete *Mycobacterium tuberculosis* (Bentley et al. 2002).

In this article, we report the analysis of the partial genome sequence of *S. ambofaciens* focusing on comparative genomic analysis with the other available *Streptomyces* genomes: *S. coelicolor* A3(2) (Bentley et al. 2002) and *Streptomyces avermitilis* (Ikeda et al. 2003). *Streptomyces ambofaciens* and *S. coelicolor* are extremely close phylogenetically (16S rRNA divergence: 1.1%), whereas *S. avermitilis* is a more distantly related species (16S rRNA divergence from *S. ambofaciens*: 2.9%). The pairwise

Key words: comparative genomics, linear bacterial chromosome, lateral gene transfer, contingency regions, *Streptomyces*.

E-mail: leblond@nancy.inra.fr.

Mol. Biol. Evol. 23(12):2361–2369. 2006
doi:10.1093/molbev/msl108
Advance Access publication September 6, 2006

© The Author 2006. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org

comparison of species phylogenetically closely or distantly related provides insights into the evolutionary mechanisms that shape the chromosome of *Streptomyces*.

Materials and Methods

Sequencing

Terminal inverted repeats (TIRs) were sequenced using 10 ordered cosmids constructed from partially *Bam*HI-digested *S. ambofaciens* ATCC23877 genomic DNA cloned into the Supercos1 (Stratagene, La Jolla, CA) vector. As the size of the *S. ambofaciens* TIRs (~198 kb) greatly exceeds that of a fragment readily clonable into a cosmid vector, each copy of the TIRs cannot be isolated as a single recombinant molecule. DNA libraries were constructed from partially *Sau*3A1-digested *S. ambofaciens* ATCC23877 genomic DNA cloned into the pBeloBAC11 (a derivate of pBAC108L [Shizuya et al. 1992]). A total of 4,809 recombinant bacterial artificial chromosomes (BACs) were isolated (insert average size: 37.9 ± 9.8 kb) representing a 21× covering rate. A systematic sequencing of the extremities of each of the BAC inserts was performed leading to 0.4× coverage of the complete chromosome. Finally, 88 ordered BACs and 10 cosmids were sequenced to cover both terminal regions. For sequencing, BACs and cosmids were mechanically fragmented and cloned with a *Bst*XI adaptor into either pcDNA2.1 vector (Invitrogen, Carlsbad, CA) or pCNS (Bartolome et al. 1991) and ligation products were then introduced into *E. coli* DH10B.

Annotation

Each BAC sequence was assembled using the Phred–Phrap–Consed suite (Ewing and Green 1998; Ewing et al. 1998; Gordon et al. 1998), and bl2seq (Altschul et al. 1997) was finally used for the overall assembly. The gene finder Glimmer2.10 (Delcher et al. 1999), trained with 3,000 coding DNA sequences (CDSs) of *S. coelicolor*, were used for CDS prediction. Results were then refined using the RBSfinder tool (Suzek et al. 2001). A potential ribosome binding site (RBS) was considered when 4 of the consensus bases 5′-GGAGG-3′ were detected upstream from the start codon (Strohl 1992). Blast 2.2.6 was used to find similarities (Altschul et al. 1997). The Interpro package was also used to describe protein domains (Zdobnov and Apweiler 2001). CDSs were assigned a functional category where their best homolog in the clusters of orthologous groups of proteins is classified (Tatusov et al. 2001). The choice of the start codon was guided by the presence of a RBS, Blast results, and the G + C frameplot pattern (Ishikawa and Hotta 1999). The most upstream start codon, minimizing overlap with the previous gene, was chosen when the situation was not clear. An annotation platform was developed using Perl scripts and the Bioperl library (Stajich et al. 2002) to deal with program outputs and sequence manipulation. A manual validation of each CDS was performed using Artemis (Rutherford et al. 2000). BlastP alignments were performed against the Non Redundant database and also against each individual *Streptomyces* proteome for comparative genomics. While comparing the predicted protein sequences with BlastP, those sharing more than 30% identity over at least 80% of the protein length were considered homologues; 2 proteins were regarded as orthologues if they are reciprocal best hits according to these criteria. Pseudogenes were identified by comparisons with their functional counterparts. A sequence was considered as a pseudogene when a coding DNA sequence has been inactivated by nonsense mutations, frameshifts, truncations, or a combination of these mechanisms. A relational database, SAMDB, is integrated into the platform to organize annotation and comparative genomics data. To visualize the degenerated synteny, protein sequence comparisons were extracted from SAMDB to be readable under ACT (Rutherford et al. 2000). Duplicated genes in the TIRs were named SAM**T**nnnn, whereas those specific only either to the left or right arms were annotated as SAM**L**nnnn and SAM**R**nnnn.

Genome Comparison using BAC End Sequences

A total of 8,457 BESs (average size: 417 ± 122 nt) were obtained from systematic sequencing of each BAC insert extremity. It resulted in 0.4× covering rate of the chromosome (~8.5 Mb), which represents approximately 1 BAC end sequence (BES) every kilobase. BlastN of each BES against the *S. coelicolor* chromosome was performed when both ends of a BAC were available and a relational database was developed in order to store the resulting data and to be able to align the *S. ambofaciens* BACs on the *S. coelicolor* central region (see fig. 1). This led us to localize the species-specific regions within the chromosome as explained in figure 1 legend.

Gene Order Conservation

For each pairwise comparison, the level of gene order conservation (GOC) was estimated along a chromosome, using a sliding window (100 CDSs with 5 CDS steps), by calculating the number of pairs of orthologues that are contiguous in the 2 compared chromosomes divided by the number of genes in the window (fig. 2). The GOC profile of the whole chromosome of *S. coelicolor* and *S. avermitilis* compared with each other (Supplementary Material online) was estimated by calculating the number of pairs of orthologues that are contiguous in the 2 chromosomes divided by the total number of orthologues in the window as defined by Rocha (2006).

The *S. ambofaciens* chromosomal arm annotation is available through the SAMDB web server on http://www.weblgm.scbiol.uhp-nancy.fr/ambofaciens/. The sequences were deposited in European Molecular Biology Laboratory under the AM238663 (left arm) and AM238664 (right arm) accession numbers.

Results

Conservation of the Central Chromosomal Region among the *Streptomyces* Genus

The terminal regions of the *S. ambofaciens* chromosome were completely sequenced over 1,544 kb and

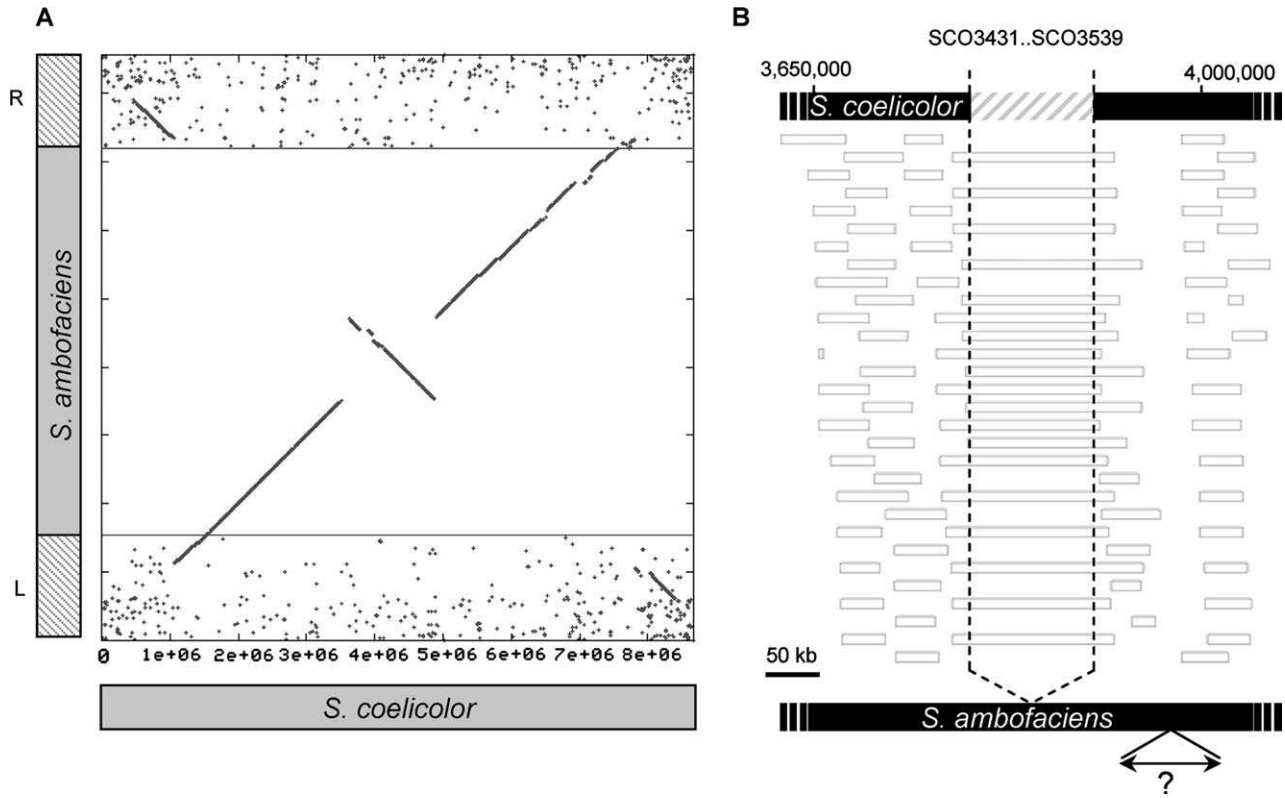


FIG. 1.—Pairwise comparison of *Streptomyces ambofaciens* and *Streptomyces coelicolor* chromosome. (*A*) Dot plot comparison of *S. ambofaciens* and *S. coelicolor*. For *S. ambofaciens*, hatched rectangles represent the completely sequenced extremities (labeled L and R). Each dot included in these extremities represents the best BlastP hit of each *S. ambofaciens* protein compared with *S. coelicolor*. The central synteny was designed using BlastN comparison of BES with *S. coelicolor* resulting in an alignment of the central chromosomal part as illustrated in *B*. For a 2-dimensional representation of this alignment, the following criteria were used: conserved regions detected with BAC alignments are considered to have the same size in both species; because the exact size of the *S. ambofaciens* central–specific regions cannot be assessed, gaps of 80 kb (i.e., the maximum insert size) were created in the alignment. (*B*) Alignment of *S. ambofaciens* BACs on the *S. coelicolor* chromosome. Each rectangle represents 1 BAC insert according to BlastN results of each BES against the *S. coelicolor* chromosome. When the distance between 2 BES matches is compatible with the length of a BAC insert (between 25 and 80 kb), the corresponding region is considered to be conserved between both species. In contrast, the set of BACs represented with a larger size reveals the presence of a *S. coelicolor*–specific region (in that case: SCO3431–SCO3539) absent in *S. ambofaciens*. Finally, the *S. ambofaciens*–specific regions can be localized by gaps in the alignment (e.g., region marked with a question mark); however, its size cannot be assessed precisely. Such gaps can also correspond to rearrangement loci, for example, inversion.

1,367 kb for the left and right extremities, respectively, and contain 2,532 CDSs, including 43 pseudogenes.

The central region was sequenced at a 0.4× covering rate using a BAC end sequencing approach (see Materials and Methods). For the comparison of the central region of the *S. ambofaciens* chromosome with *S. coelicolor*, the BESs were aligned on the *S. coelicolor* chromosome using BlastN analyses (fig. 1). This analysis reveals a high level of synteny between the central parts of the chromosome of these 2 species. Chromosome structure comparisons reveal the occurrence of 2 inversion events centered on the origin of replication, detectable as broken X patterns in the data (fig. 1*A*). They correspond to 2 out of the 4 inversions previously reported in the *S. coelicolor*/*S. avermitilis* genome comparison (Ikeda et al. 2003). Most (6/8) of these inversion points cannot be precisely localized because they are located in species-specific regions. For example, a break point falls near the pSAM2-like element (Sezonov et al. 1998) adjacent to the calcium-dependent antibiotic (*cda*) cluster in *S. coelicolor*. It thus seems that the inversion events are more likely to be fixed within regions where rearrangements are counterselected less efficiently. Further, in order to assess the level of identity between central homologous genes, 3 independent BACs from the *S. ambofaciens* central chromosomal region were arbitrarily chosen and sequenced in their entirety (91 genes; 117 kb). All predicted genes share a conserved organization and a high level of identity with the other *Streptomyces* genomes following the phylogenetic relationships (89% and 82% of amino acid identity with *S. coelicolor* and *S. avermitilis*, respectively; 90% and 84%, respectively, at the nucleotide level; data not shown) inferred by rDNA sequence analysis (http://avermitilis.ls.kitasato-u.ac.jp/tree2.html). A single synteny break is observed among the 117 kb between *S. ambofaciens* and *S. coelicolor*. Dot plots comparing homologous proteins of the 2 other *Streptomyces* reveal that the central region is highly conserved even at long evolutionary distances as shown in Ikeda et al. (2003). The synteny includes the central part, whereas the replicon extremities appear species specific.

Although the level of synteny is high in the central regions, the insertion of several specific genomic islands

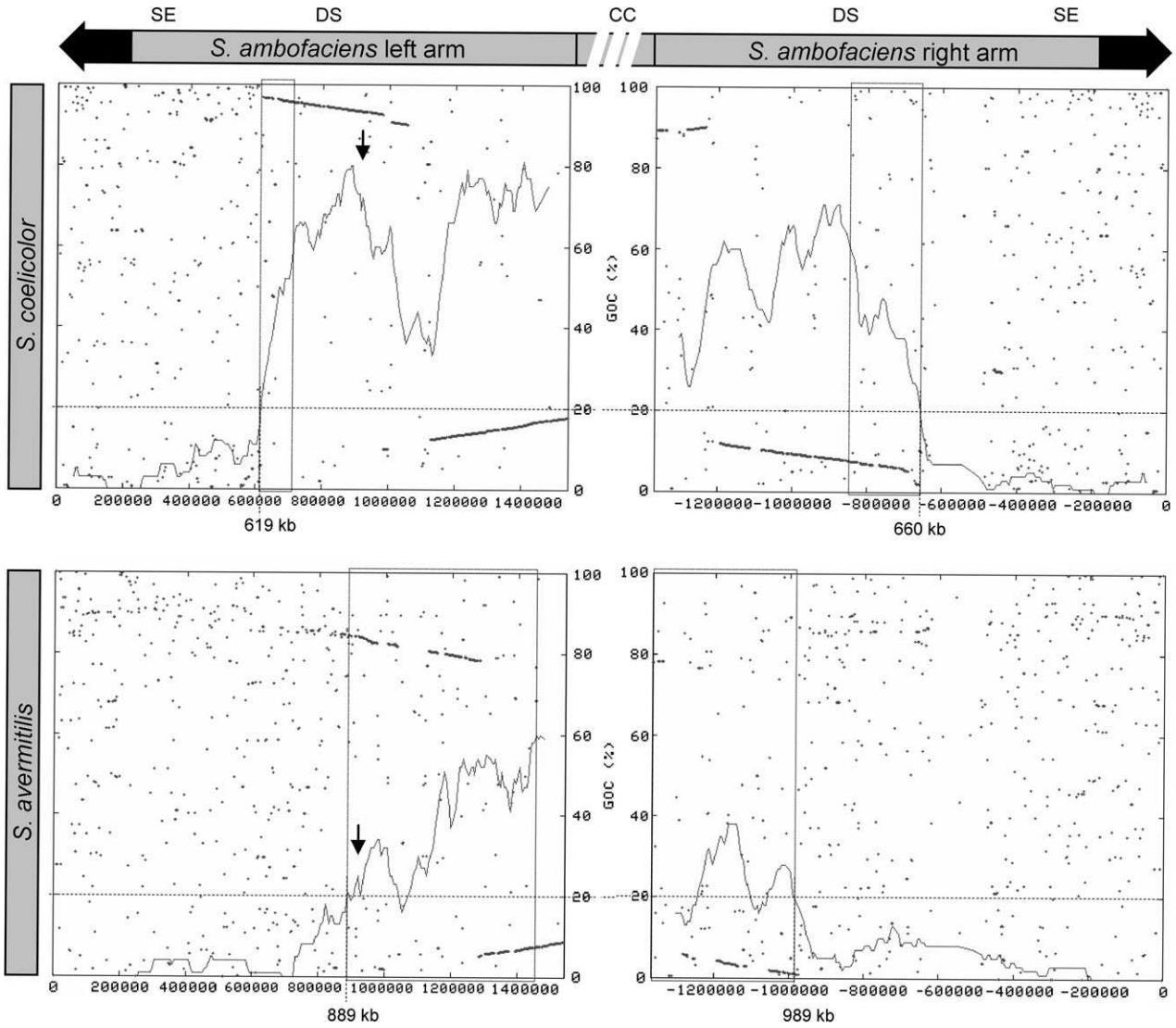


FIG. 2.—Pairwise comparison of the *Streptomyces ambofaciens* chromosomal arms with the 2 other chromosomes. Profile of GOC and protein dot plot are superimposed for a given pairwise comparison. GOC values correspond to the percent of pairs of orthologues that are contiguous in the *S. ambofaciens* genome and in the compared genome (see Materials and Methods). GOC value (*y* coordinate) was calculated using a sliding window (100 CDSs with 5 CDS steps); *x* coordinate corresponds to the position of the 50th CDS in the window. A threshold of 20% GOC (dotted horizontal lines) was chosen to delimit the end of specific extremities for each pairwise comparison, the size of which is indicated under the *x* axis. Framed areas represent the degenerated syntenic regions of *S. ambofaciens*, the location of which varies according to species compared (a threshold of 60% GOC was chosen for their internal limit). Small arrows indicate the region of *S. ambofaciens* zoomed in figure 3. Black arrows at the end of the *S. ambofaciens* chromosome represent the TIRs. SE: specific extremity; DS: degenerated syntenic part; CC: conserved central part.

has occurred. The *S. ambofaciens*/*S. coelicolor* comparison using BAC alignments (fig. 1*B*) revealed 4 clusters (larger than the maximum length of a BAC insert, i.e., circa 80 kb) specific to *S. ambofaciens*. One of them corresponds to the spiramycin antibiotic gene cluster (Richardson et al. 1990). Reciprocally, 12 *S. coelicolor*–specific regions whose sizes vary from 26 to 149 kb are located in the essential core region (Supplementary Material online). Bentley et al. (2002) defined 14 regions potentially recently laterally acquired in the *S. coelicolor* chromosome according to the G/C and gene contents. The present analysis confirms that half of them (7/14) could have a recent origin in *S. coelicolor* because they are absent in the *S. ambofaciens* chromosome, although one cannot exclude the possibility that they were once present in the *S. ambofaciens* lineage but have since been lost. They include the actinorhodin (Malpartida and Hopwood 1986), undecylprodigiosin (Rudd and Hopwood 1980), and *cda* (Wright and Hopwood 1976) gene clusters and a 41-kb region (from SCO3677 to SCO3725 including genes relevant to resistance to heavy metals) adjacent to a tRNA gene. Although tRNA genes are known targets for integrative elements, no other specific signature could be identified within this genomic island. The 7 other regions putatively recently laterally acquired have a small size (about 10 kb), and the BES analysis reveals that at least 2 of them are present in the *S. ambofaciens* chromosome.

The Size of the Terminal Species-Specific Regions Increases as the Phylogenetic Distance between Compared Species Increases

The confinement of variability at the chromosomal extremities seems a general trait of the *Streptomyces* genome. Indeed, when a given species is compared with the other 2, the minimal size of the specific information can be deduced. Thus, the specific extremities cover 1,279 kb (619 + 660 kb) in *S. ambofaciens* as illustrated in figure 2. Considering the absence of a drastic drop in synteny, the limits of the specific terminal regions were defined using a threshold of 20% of GOC for each pairwise comparison (see Materials and Methods for GOC calculation). Interestingly, the size of the species-specific regions increases with the phylogenetic distance. This is shown in figure 2, where the size of the *S. ambofaciens*–specific extremities increases from 1,279 kb (619 + 660 kb) compared with *S. coelicolor* to 1,878 kb (889 + 989 kb) compared with *S. avermitilis*. Reciprocally, the size of the central conserved part decreases with the distance. These data suggest that the specific information preferentially accumulates in the terminal regions along the evolutionary time.

When all pairwise comparisons of the terminal regions of the 2 other *Streptomyces* were carried out using the same approach (GOC calculation with the same parameters), the minimal size of the species-specific extremities was estimated to be 753 kb in *S. coelicolor* and 1,393 kb in *S. avermitilis*. Enrichment of the terminal ends in specific information is highlighted by the fact that, for example, the 753 kb specific to *S. coelicolor* represents only 9% of the whole chromosome but corresponds to 41% of the total of the regions estimated to be absent in *S. ambofaciens*.

Level of Variability of the Terminal Species-Specific Regions

Figure 2 shows the location of the ends of the synteny between *S. ambofaciens* chromosomal arms and the other genomes. The *S. ambofaciens*–specific chromosomal extremities (619 + 660 kb) are characterized by a very low level of GOC. Although conserved clusters could be observed, they are limited to very small regions (generally less than 8 genes), whatever the phylogenetic relationship considered. These terminal species-specific regions include 1,082 CDSs in *S. ambofaciens* of which only 9% share more than 60% end-to-end identity in the 2 other *Streptomyces*. When compared with its close relative *S. coelicolor*, 37% of the 1,082 CDSs are absent. Only 13% are highly conserved (>80% of identity). This contrasts with the conservation estimated in the central part, where 100% of the CDSs identified (over 117 kb) share more than 80% of identity (average: 89%). The level of variability is higher with *S. avermitilis* with 44% of the 1,082 CDSs that do not share any similarity and only 4% that are highly conserved.

The low levels of conservation (synteny and identity) between homologues strongly suggest that the majority of them are not orthologues but rather result from the massive introduction of foreign alleles (xenologues) by LGT. Thus, the specificity of the extremities probably originates from LGT. Our data reveal that the level of terminal variability is extremely high even between closely related species extending the preliminary conclusions resulting from the comparison of the *S. coelicolor* and *S. avermitilis* genome sequences (Ikeda et al. 2003).

The origins of the terminal specificity could also result from the presence of mobile genetic elements. Indeed, the terminal regions are enriched in mobile elements (insertion sequence–like elements) as previously reported (Chen et al. 2002) and proposed to mediate DNA rearrangements and integration of horizontally transferred DNA sequences. In the *S. ambofaciens* chromosomal arms, 53 transposase-encoding genes (including pseudogenes) were predicted with a strong bias in the terminal species-specific regions (50 of 53). Furthermore, some terminal CDSs show highest similarity to plasmid-associated genes (Choulet et al. 2006). In *S. ambofaciens*, 4 homologues to plasmid-associated genes lie in the terminal 50 kb, for example, the helicase-encoding gene *ttrA*, which is conserved in the 2 other species but which shares best similarity with its homologue of the plasmid SLP2 of *Streptomyces lividans* (Huang et al. 2003). Three CDSs similar to genes encoding plasmid transfer functions (*spdB*, *traB2*, and *traA2*) are also detected in the *S. avermitilis* chromosomal extremities. In *S. ambofaciens*, this observation is correlated with an average GC content (68.8% in the 50 terminal kb) that is slightly lower than that of the rest of the genome (72.3%). A lower GC content is typical from *Streptomyces* linear plasmids (Spatz et al. 2002; Huang et al. 2003; Ikeda et al. 2003; Bentley et al. 2004) and more generally from mobile elements.

Degenerated Syntenic Regions and Massive Gene Flux

Synteny analysis of the 117 kb sequenced in the center of the *S. ambofaciens* chromosome revealed a high GOC with *S. coelicolor*, and a similar level of synteny is described when compared with *S. avermitilis*. Interestingly, the synteny observed between central regions degenerates progressively over several hundreds of kilobases before reaching the terminal species-specific regions (framed areas in fig. 2). This is true whatever the pairs of species considered. In these regions, the synteny appears as gradually parceled out by multiple insertions/deletions (indels) of genes. Degeneration of synteny not only reflects an increase in the level of rearrangement of endogenous information but also reflects a rapid evolution of the gene content by gene flux, that is, by accumulation of new genes and loss of ancestral information. When the GOC level falls under 20%, the synteny becomes undetectable and the regions were consequently considered as specific. Figure 3 illustrates this degeneration phenomenon using a protein-to-protein comparison between *S. ambofaciens*, *S. coelicolor*, and *S. avermitilis* and shows that degeneration follows the phylogenetic distances.

More significantly, the number of synteny breaks (expressing the level of degeneration) was estimated over 100 kb (99 CDSs) including the locus detailed in figure 3, of the *S. ambofaciens* left arm (from SAML0798 to SAML0896). Compared with *S. coelicolor*, 10 rearrangements (from SCO7238 to SCO7327) are observed, whereas at least 29 events have led to the current genome divergence

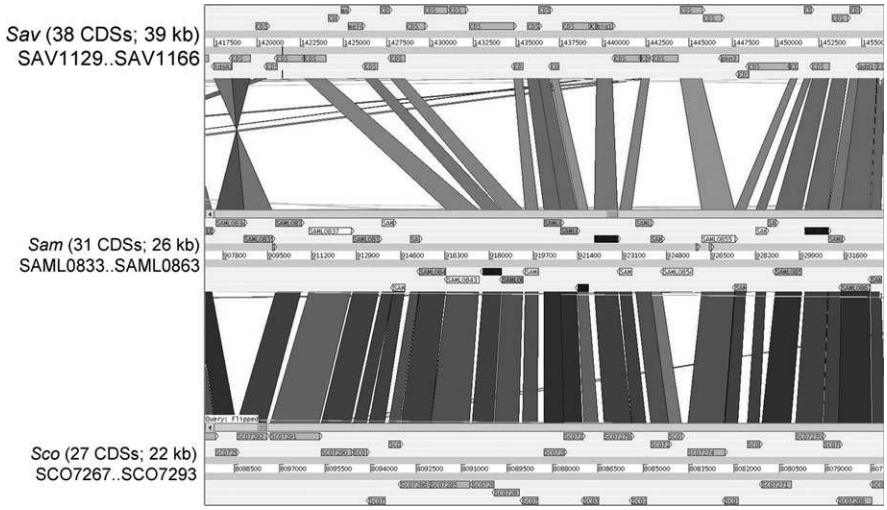


FIG. 3.—Protein-to-protein comparisons displayed with Artemis Comparison Tool (Rutherford et al. 2000). CDSs are represented by arrows in the 6 frames, and pairs of homologues are linked by a gray area. Example of a 26-kb region of the *Streptomyces ambofaciens* (Sam) left arm (from position 907,000–933,000) included into the degenerated syntenic regions compared with *Streptomyces coelicolor* (Sco) and *Streptomyces avermitilis* (Sav) (see arrows in fig. 2). As the sizes of the compared regions are not identical in the 3 species, the scales are different to fit in window.

with *S. avermitilis* (from SAV1085 to SAV1268) (in contrast, 1 and 4 synteny breaks were observed along the 117 kb of central sequences, respectively). Correlated to the decreasing GOC, the proportion of remaining orthologous genes also varies according to the phylogenetic distance (83 and 52 of 99 *S. ambofaciens* CDSs are orthologues with *S. coelicolor* and *S. avermitilis*, respectively), highlighting the importance of gene flux in these regions. However, at long phylogenetic distances, the number of indels probably becomes underestimated. Indeed, assuming an indel size of 1–10 genes (estimated with the *S. ambofaciens*/*S. coelicolor* comparison), a synteny break probably masks multiple indel events. Obviously, some indel events can involve more than 10 genes (e.g., operonic structure or gene clusters such as those involved in secondary metabolite biosynthesis), the insertion of the whole set of genes being necessary for the achievement of a selectable function. Conversely, the loss of one of these genes would favor the loss of the complete gene set. This situation was indeed observed in these degenerated synteny regions. For example, in *S. ambofaciens*, a 34.5-kb cluster (28 genes) implied in secondary metabolism is species specific. In *S. coelicolor*, this region is replaced by 30.8-kb–specific cluster (31 genes, SCO0850–SCO0880). In *S. avermitilis,* the same region (SAV7356–SAV7422) corresponds to a secondary metabolite gene cluster (*pks1*, [Ikeda et al. 2003]) different from the 2 other species. Such large genomic islands result in a drop of the GOC profile, for example, from position 1,050,000–1,125,000 of the *S. ambofaciens* left arm (fig. 2). The *S. ambofaciens* right arm exhibits higher variability than the left arm. Large specific islands inserted in the terminal regions result in falls of the GOC profile.

As shown by GOC variation (fig. 2), the regions affected by the degeneration are more internal when distant *Streptomyces* are compared. Altogether, these data suggest that the terminal regions are prone to a massive and constant gene flux. Strikingly, the profile of GOC increases in a gradual way from the terminal specific regions to the internal conserved part. The gradient of degeneration toward the extremities extends over 600 kb of the left arm of *S. ambofaciens* when it is compared with *S. avermitilis* (fig. 2). The slope of GOC variation is stronger when compared with *S. coelicolor*, but the gradient of degeneration remains detectable using a smaller sliding window as well as by dot plot comparison (data not shown).

When the whole chromosome of *S. coelicolor* and *S. avermitilis* are compared with each other, GOC profile reveals that the degenerated syntenic regions extend over the entire contingency regions as defined by Bentley et al. (2002) (Supplementary Material online). Although the GOC is high and constant in the central region, it gradually decreases toward the extremities.

Discussion

Comparative genomics within the *Streptomyces* genus has revealed a highly compartmentalized structure where variability is mostly confined to the extremities of the linear chromosome. Indeed, the core region defined by the presence of essential genes (Bentley et al. 2002) appears to be highly syntenic throughout the genus, whereas the chromosomal arms, which contain contingency genes, contain highly variable and species-specific genes. In addition, a high level of chromosome instability has long been known through genomic characterization of mutants showing terminal rearrangements (Fischer et al. 1998; Wenner et al. 2003).

Therefore, the high level of genomic diversity would occur in a highly organized structure where gene maintenance might be dependent on chromosomal location. Thus, a part of variability probably results from exchange of extremities between linear replicons (plasmids or chromosomes). Such events have already been observed in *Streptomyces* (Pandza et al. 1998; Yamasaki and Kinashi 2004) as well as in *Borrelia*

(Casjens et al. 1997, 2000). This hypothesis is supported by the identification of genes similar to plasmid-associated genes proximal to the chromosomal ends. Numerous linear conjugative plasmids are known in *Streptomyces* (Hopwood 1999), which could self-mobilize and/or mobilize chromosomal regions. The "end first" model of mobilization proposed by Chen (1996), that is, the transfer of linear replicon from one end, could explain why terminal regions might be favored during conjugational events. The transfer of terminal information could lead to replacement of the whole chromosomal end as suggested by the analysis of strain-specific regions at the end of the *S. ambofaciens* TIRs (Choulet et al. 2006). This type of fast evolution mechanism could be an advantage conferred by chromosomal linearity.

Comparative genomics revealed that the frontier between the specific and the conserved regions corresponds to a region of degenerated synteny (figs. 2 and 3). This phenomenon is observable for each pairwise comparison, and the level of degeneration is also correlated to the phylogenetic distance. This type of genome divergence cannot result from exchange of replicon extremities. These data rather reveal that *Streptomyces* terminal regions are subject to a massive and constant gene flux occurring as a result of insertions, deletions, and replacements. These events accumulate with time and gradually erase the GOC. Because the frontier between the specific and the conserved regions is not the same when considering different pairs of species, this degeneration does not seem to be locus specific but would rather affect the whole contingency regions. The earlier 2 *Streptomyces* species diverge, the shorter the conserved region. Reciprocally, the size of the species-specific regions increases with the phylogenetic distance.

Lawrence and Roth suggested that the acquisition of foreign genes contributing to cell fitness results in the loss of resident functions of lower selective value, that is, contributing weakly or not at all to the fitness (Lawrence and Roth 1999). This model applies because multiple constraints limit genome-size expansion. A minimal selective advantage, the *s* value corresponding to the maintenance threshold, is required for maintenance of a gene in a genome. Indeed, genes with an adaptative value under the threshold *s* can be either fixed or lost in a context of low or high mutation rate, respectively. In the context of a high rate of rearrangements (e.g., deletion), a stronger selective coefficient is required for a gene to be maintained. Conversely, genes located in regions of high recombination rate must have a higher adaptative value to be maintained.

Interestingly, analysis of the *Streptomyces* genomes supports a region-dependent selection pressure. In other words, we proposed that the minimal *s* value required for maintenance would be dependent on the chromosomal location. As the rearrangement rate (i.e., deletion rate) gradually increases while approaching the chromosomal ends, the threshold for gene maintenance and, consequently, the frequency of gene loss also gradually increases.

An alternative hypothesis would be that rearrangements preferentially occur in the nonessential regions because they are tolerated better (Chen et al. 2002). In other words, because terminal regions lack essential genes, the terminal variability would reflect a higher frequency of fixation of mutations instead of a higher frequency of rearrangements. The tendency of essential genes to be located in the central region would be the result of gene dosage effect, that is, highly expressed genes would tend to be close to the origin of replication in fast-growing bacteria as shown by Couturier and Rocha (2006). However, *Streptomyces* are slow-growing bacteria, and highly expressed genes are also present in the terminal regions, although enrichment can be noticed in the central part (Wu et al. 2005). According to this hypothesis, the gradual degeneration of the synteny would reveal that the genes are distributed according to their fitness contribution.

The increasing level of rearrangements toward the extremities may be the force driving the compartmentalization that excludes essential genes from the extremities and generates a high rate of gene flux for adaptation capabilities. Thus, the fact that the terminal regions are more tolerant to rearrangements would be true, but it would only be a consequence of the particular organization driven by a variable level of instability along the chromosome. These higher recombination frequencies could result from formation of double strand breaks (DSBs) generated by arrests of the replication fork. DSBs initiated during termination of replication were demonstrated to stimulate genetic instability in *E. coli* (Michel et al. 1997). It can also be speculated that conjugational mechanisms favor the introduction of replicon extremities into the recipient and/or the formation of DSBs in the chromosome of the donor and may thus stimulate DNA recombination in the terminal regions. The presence of multiple mobile genetic elements could also account for a high frequency of DNA breaks in the terminal regions (Gunes et al. 1999; Chen et al. 2002). These phenomena might constitute an advantage conferred by chromosomal linearity to acquired and rearranged DNA in these regions.

The core region of the *Streptomyces* genome is syntenic with the whole chromosome of *M. tuberculosis*. *Mycobacterium tuberculosis* belongs to the actinomycetes and possesses a circular chromosome. A parsimonious hypothesis about the evolution of the *Streptomyces* chromosome would be that a single event resulted in the acquisition of the contingency regions and chromosomal extremities by the ancestral chromosome (integration of a linear replicon within the ancestral chromosome) (Volff and Altenbuchner 2000). This hypothesis does not explain how the current terminal regions diverged from their original ancestral version. Our analysis has shed light on these mechanisms, proposing a gradient of chromosome instability (i.e., indel frequency) toward the chromosomal extremities that generate a gradient of selection pressure that eliminates the genes contributing weakly to the cell fitness. Hence, the closer a gene is to the end of the linear chromosome, the lower is its probability of being maintained.

Supplementary Material

1) A table outlining the locations and features of *S. coelicolor* regions absent in the *S. ambofaciens* genome and 2) GOC profiles of the *S. coelicolor* and *S. avermitilis* chromosomes compared with each other are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

F.C. and A.G. were recipients of a grant from the "Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche" (M.E.N.E.S.R.). This research was supported by a Programme d'Actions Intégrées (ALLIANCE), the "ACI Microbiologie 2003" program (funded by M.E.N.E.S.R.), and the VIth PCRDT ("ActinoGen"). Many thanks are due to K. Chater, G. Chandra, and T. Kieser (John Innes Centre, Norwich, United Kingdom) for their warm welcome and their help in the development of the computational methods. We are grateful to A. Hesketh (John Innes Centre, Norwich, United Kingdom) for critical reading of the manuscript.

Literature Cited

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Bartolome B, Jubete Y, Martinez E, de la Cruz F. 1991. Construction and properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives. Gene. 102:75–78.

Bentley SD, Brown S, Murphy LD, et al. (18 co-authors). 2004. SCP1, a 356,023 bp linear plasmid adapted to the ecology and developmental biology of its host, *Streptomyces coelicolor* A3(2). Mol Microbiol. 51:1615–1628.

Bentley SD, Chater KF, Cerdeno-Tarraga AM, et al. (43 co-authors). 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature. 417:141–147.

Boccard F, Esnault E, Valens M. 2005. Spatial arrangement and macrodomain organization of bacterial chromosomes. Mol Microbiol. 57:9–16.

Casjens S, Murphy M, DeLange M, Sampson L, van Vugt R, Huang WM. 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, Palmer N, van Vugt R, et al. (15 co-authors). 2000. 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.

Chater KF. 1993. Genetics of differentiation in *Streptomyces*. Annu Rev Microbiol. 47:685–713.

Chen CW. 1996. Complications and implications of linear bacterial chromosomes. Trends Genet. 12:192–196.

Chen CW, Huang CH, Lee HH, Tsai HH, Kirby R. 2002. Once the circle has been broken: dynamics and evolution of *Streptomyces* chromosomes. Trends Genet. 18:522–529.

Choulet F, Gallois A, Aigle B, et al. (14 co-authors). 2006. Intraspecific variability of the terminal inverted repeats of the linear chromosome of *Streptomyces ambofaciens*. J Bacteriol. 188:6599–6610.

Couturier E, Rocha EP. 2006. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. Mol Microbiol. 59: 1506–1518.

Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 27:4636–4641.

Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8: 186–194.

Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8:175–185.

Fischer G, Decaris B, 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, 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 USA. 95:14296–14301.

Gil R, Sabater-Munoz B, Latorre A, Silva FJ, Moya A. 2002. Extreme genome reduction in *Buchnera* spp.: toward the minimal genome needed for symbiotic life. Proc Natl Acad Sci USA. 99:4454–4458.

Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. Genome Res. 8:195–202.

Gunes G, Smith B, Dyson P. 1999. Genetic instability associated with insertion of IS6100 into one end of the *Streptomyces lividans* chromosome. Microbiology. 145:2203–2208.

Hopwood DA. 1999. Forty years of genetics with *Streptomyces*: from in vivo through in vitro to in silico. Microbiology. 145: 2183–2202.

Huang CH, Chen CY, Tsai HH, Chen C, Lin YS, Chen CW. 2003. Linear plasmid SLP2 of *Streptomyces lividans* is a composite replicon. Mol Microbiol. 47:1563–1576.

Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, Sakaki Y, Hattori M, Omura S. 2003. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat Biotechnol. 21: 526–531.

Ishikawa J, Hotta K. 1999. FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G + C content. FEMS Microbiol Lett. 174:251–253.

Lawrence JG, Hendrickson H. 2005. Genome evolution in bacteria: order beneath chaos. Curr Opin Microbiol. 8:572–578.

Lawrence JG, Roth JR. 1999. Genomic flux: genome evolution by gene loss and acquisition in bacterial genomes. In: Charlebois RL, editor. Organization of the prokaryotic genome. Washington (DC): American Society for Microbiology. p. 263–289.

Leblond P, Decaris B. 1999. 'Unstable' linear chromosomes: the case of *Streptomyces*. In: Charlebois RL, editor. Organization of the prokaryotic genome. Washington (DC): American Society for Microbiology. p. 235–261.

Lobry JR, Louarn JM. 2003. Polarisation of prokaryotic chromosomes. Curr Opin Microbiol. 6:101–108.

Malpartida F, Hopwood DA. 1986. Physical and genetic characterisation of the gene cluster for the antibiotic actinorhodin in *Streptomyces coelicolor* A3(2). Mol Gen Genet. 205:66–73.

Michel B, Ehrlich SD, Uzest M. 1997. DNA double-strand breaks caused by replication arrest. Embo J. 16:430–438.

Mira A, Klasson L, Andersson SG. 2002. Microbial genome evolution: sources of variability. Curr Opin Microbiol. 5:506–512.

Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature. 405: 299–304.

Pandza S, Biukovic G, Paravic A, Dadbin A, Cullum J, 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.

Richardson MA, Kuhstoss S, Huber ML, Ford L, Godfrey O, Turner JR, Rao RN. 1990. Cloning of spiramycin biosynthetic

genes and their use in constructing *Streptomyces ambofaciens* mutants defective in spiramycin biosynthesis. J Bacteriol. 172: 3790–3798.
Rocha EP. 2004. The replication-related organization of bacterial genomes. Microbiology. 150:1609–1627.
Rocha EP. 2006. Inference and analysis of the relative stability of bacterial chromosomes. Mol Biol Evol. 23:513–522.
Rudd BA, Hopwood DA. 1980. A pigmented mycelial antibiotic in *Streptomyces coelicolor*: control by a chromosomal gene cluster. J Gen Microbiol. 119:333–340.
Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. Bioinformatics. 16:944–945.
Sezonov G, Duchene AM, Friedmann A, Guerineau M, Pernodet JL. 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.
Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc Natl Acad Sci USA. 89: 8794–8797.
Spatz K, Kohn H, Redenbach M. 2002. Characterization of the *Streptomyces violaceoruber* SANK95570 plasmids pSV1 and pSV2. FEMS Microbiol Lett. 213:87–92.
Stajich JE, Block D, Boulez K, et al. (21 co-authors). 2002. The Bioperl toolkit: Perl modules for the life sciences. Genome Res. 12:1611–1618.
Strohl WR. 1992. Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. Nucleic Acids Res. 20:961–974.
Suyama M, Bork P. 2001. Evolution of prokaryotic gene order: genome rearrangements in closely related species. Trends Genet. 17:10–13.
Suzek BE, Ermolaeva MD, Schreiber M, Salzberg SL. 2001. A probabilistic method for identifying start codons in bacterial genomes. Bioinformatics. 17:1123–1130.
Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS, Wernegreen JJ, Sandstrom JP, Moran NA, Andersson SG. 2002. 50 million years of genomic stasis in endosymbiotic bacteria. Science 296:2376–2379.
Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV. 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. 29:22–28.
Volff JN, Altenbuchner J. 2000. A new beginning with new ends: linearisation of circular chromosomes during bacterial evolution. FEMS Microbiol Lett. 186:143–150.
Welch RA, Burland V, Plunkett G 3rd, et al. (19 co-authors). 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc Natl Acad Sci USA. 99:17020–17024.
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.
Wright LF, Hopwood DA. 1976. Actinorhodin is a chromosomally-determined antibiotic in *Streptomyces coelicolor* A3(2). J Gen Microbiol. 96:289–297.
Wu G, Culley DE, Zhang W. 2005. Predicted highly expressed genes in the genomes of *Streptomyces coelicolor* and *Streptomyces avermitilis* and the implications for their metabolism. Microbiology. 151:2175–2187.
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.
Zdobnov EM, Apweiler R. 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics. 17:847–848.

Martin Embley, Associate Editor

Accepted September 1, 2006

IV- Etude dynamique du chromosome : Clonage multiple de sites parS dans les régions instables

Les études portant sur l'instabilité ont permis de caractériser la structure du chromosome de différents mutants isolés de *S. ambofaciens* (Fischer *et al.* 1997b). Un de ces mutants, nommé NSA27, résulte de la fusion de deux chromosomes partiellement délétés (Wenner *et al.* 2003). Cette structure dupliquée est strictement associée à un cycle d'instabilité phénotypique et chromosomique. La descendance de ce mutant présente 70% de réarrangements chromosomiques, tels que des translocations ou des circularisations du chromosome. Cette instabilité est directement associée à l'héritabilité de la structure parentale fusionnée. En effet, les descendants ayant perdu la fusion retrouvent les caractéristiques d'instabilité de type sauvage (Wenner *et al.* 2003).

Des analogies avec ce qui est observé chez les eucaryotes présentant un chromosome dicentrique sont retrouvées. En raison de la présence de deux centromères actifs au sein de la même molécule, le chromosome dicentrique subit une tension physique quand les deux centromères sont tractés aux pôles opposés de la cellule (Sabatier *et al.* 2005). Cette tension conduit à une cassure chromosomique générant des extrémités d'ADN non télomériques. Ces extrémités sont fortement recombinogènes et d'autres réarrangements génomiques seront induits durant les cycles cellulaires suivants. Ces chromosomes dicentriques rentrent dans un cycle d'instabilité génomique appelé cycle cassure-fusion-pont (CFP) découvert chez le maïs par Barbara McClintock (1951).

Par analogie, les chromosomes fusionnés des mutants de *S. ambofaciens* présenteraient des défauts de partition en raison de la présence d'une région chromosomique surnuméraire impliquée dans le positionnement des chromatides néorépliquées dans les cellules en cours de division cellulaire (vraisemblablement au cours de la sporulation). Pour tester l'hypothèse d'un lien entre la présence de deux centres de partition ou centromère bactérien et instabilité chromosomique, nous avons envisagé la construction d'un second centromère dans les régions instables du chromosome sauvage et observer le niveau d'instabilité des souches obtenues.

A- Partition chromosomique chez les Streptomyces

La partition chromosomique est un mécanisme actif permettant la bonne distribution des chromosomes néosynthétisés dans chacune des cellules filles. Pour se faire, la protéine ParB se fixe sous forme de dimère sur un élément *cis*, séquence *parS* (5'- TGTTNCACGTGAAACA-3') (Lin et Grossman 1998).

A partir de la séquence génomique de *S. coelicolor*, 24 motifs *parS*, homologues de la séquence définie chez *Ba. subtilis,* ont été identifiés (Kim *et al.* 2000). La majorité de ces séquences (21/24) est située dans une région de 400 kb entourant l'origine de réplication. Ces sites présentent une séquence consensus palindromique (5' tGTTTCACGTGAAACa 3'), présentant une symétrie typique des sites de fixation de protéines dimériques (Jakimowicz *et al.* 2002). De plus, 16 de ces sites sont situés dans des régions intergéniques, 2 sont situés dans la région

promotrice de l'opéron *parAB* et sont impliqués dans la régulation de l'expression de l'opéron.

Des expériences d'immuno-précipitation, utilisant la protéine ParB purifiée, ont été réalisées sur 14 sites *parS* potentiels (Jakimowicz *et al.* 2002). Ainsi 8 sites sont clairement des cibles de la protéine ParB. Les six autres candidats sont soit faiblement reconnus par ParB (4/6) soit pas du tout (2/6). L'analyse de ces sites a montré que cinq de ces six sites présentent une variation par rapport au consensus dans leur séquence interne. De plus, deux de ces six sites sont localisés à 500 et 100 kb de l'*oriC* et les quatre autres sont situés à l'intérieur d'une phase codante. Ainsi, les sites *parS* « forts » sont préférentiellement situés dans des régions intergéniques, présentent la séquence consensus de 16 nucléotides et sont situés à proximité de l'origine de réplication, *oriC* (Jakimowicz *et al.* 2002).

A partir de ces données, un locus contenant deux sites *parS*, séparés par 32 nucléotides, situé dans une région intergénique (SCO3896 et SCO3897) a été choisi. Les deux gènes flanquants sont convergents, SCO3896 coderait une RNA nucléotidyltransférase, alors que SCO3897 coderait une protéine hypothétique. Ces deux sites, lors des expériences d'immuno-précipitation (Jakimowicz *et al.* 2002), présentaient une très forte fixation de la protéine ParB, suggérant leur caractère fonctionnel probable.

B- Insertion de sites parS

B-1 LA TECHNIQUE DE PCR-TARGETING

La technique de PCR-Targeting a été mise au point pour déléter un locus donné et le remplacer par une cassette de résistance à un antibiotique (Fig. 45) (Gust *et al.* 2003). Cette technique a été utilisée ici pour permettre l'insertion des sites *parS* dans un locus choisi au préalable, sans délétion associée (insertion propre) ceci pour éviter d'éventuelles modifications de l'expression des gènes avoisinants.

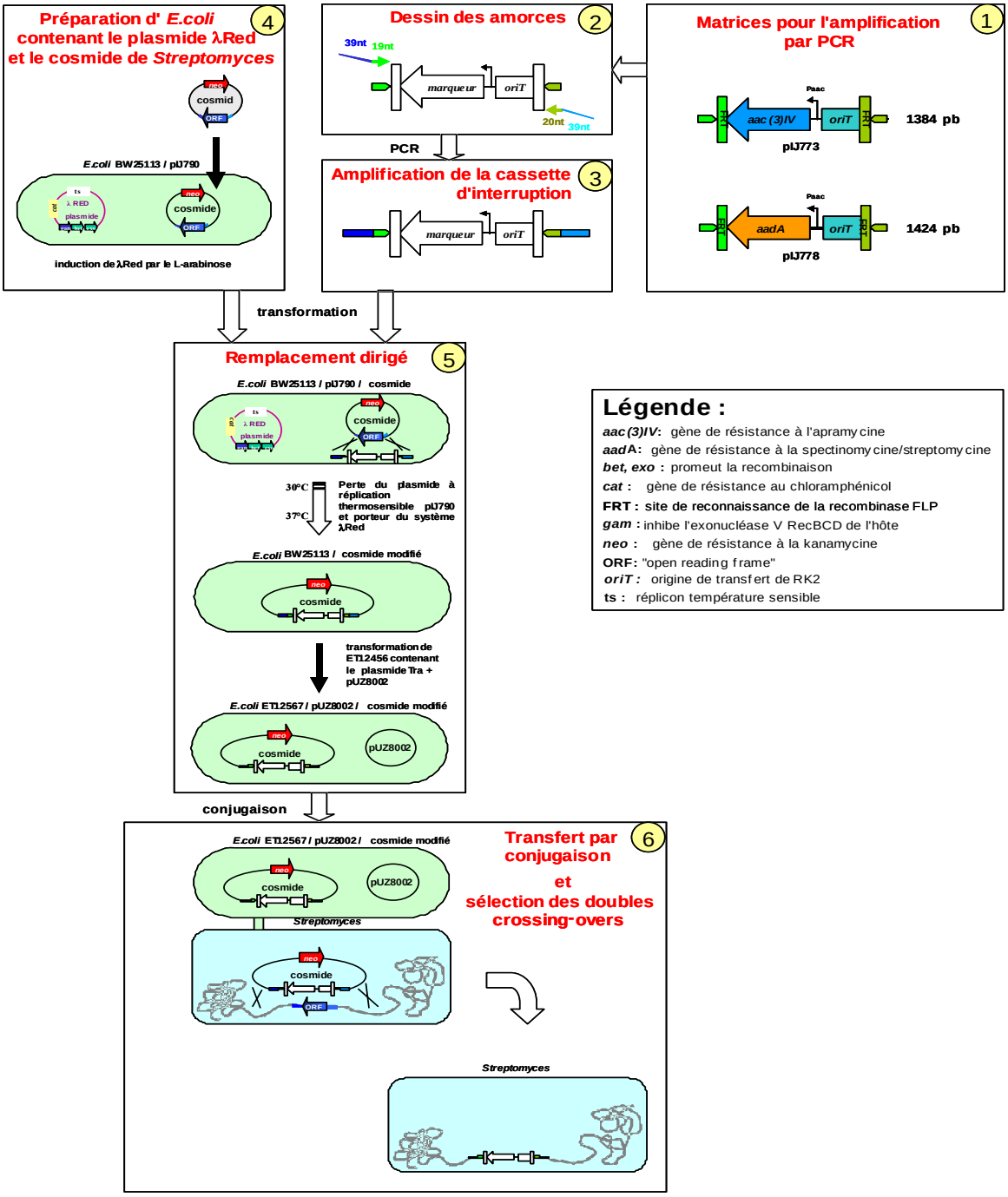


Figure 45 : Stratégie de remplacement de gène par « PCR Targeting » (d'après Gust et al., 2002)

B-2 CHOIX DE LA LOCALISATION DE L'INSERTION DES SITES *PARS*

Les régions choisies pour l'insertion des sites *parS* doivent répondre à différents critères :

- Localisation dans les régions terminales du chromosome de *S. ambofaciens*, ces régions pouvant être perdues sans affecter la croissance de la bactérie en conditions de laboratoire.

- Le point d'insertion devait se situer entre deux gènes convergents pour conserver la même organisation génétique que celle de la majorité des sites *parS*.

Ainsi, quatre sites d'insertion ont été choisis, deux dans le bras chromosomique droit et

deux dans le bras chromosomique gauche. Pour chacun des bras, un site est situé à proximité de la frontière (mais en dehors) des TIR et l'autre site est situé à l'intérieur du chromosome (Fig. 46).

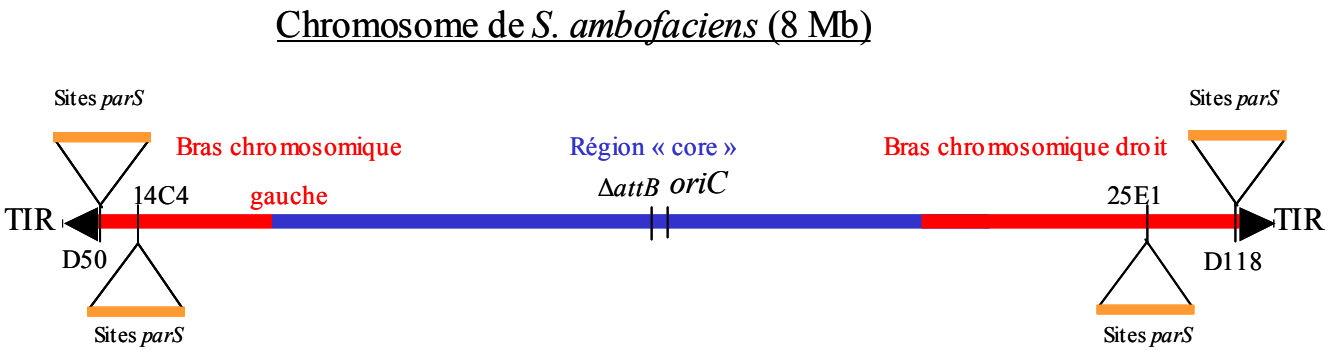


Figure 46 : Localisation des sites d'insertion des sites *parS* sur le chromosome de *S. ambofaciens* DSM40697
Les têtes de flèche noire correspondent aux TIR. Les régions en rouge sont les régions terminales du chromosome et la région bleue la région centrale. Le nom des cosmides où se situe l'insertion est indiqué sous les flèches.

Dans le bras droit, le premier site d'insertion est localisé dans le cosmide D50, à 21 965 nt après la frontière de la TIR (Fig. 47).

Dans le bras gauche, le premier point d'insertion est situé dans le cosmide D118, à 22 863 nt après la frontière de la TIR (Fig. 48).

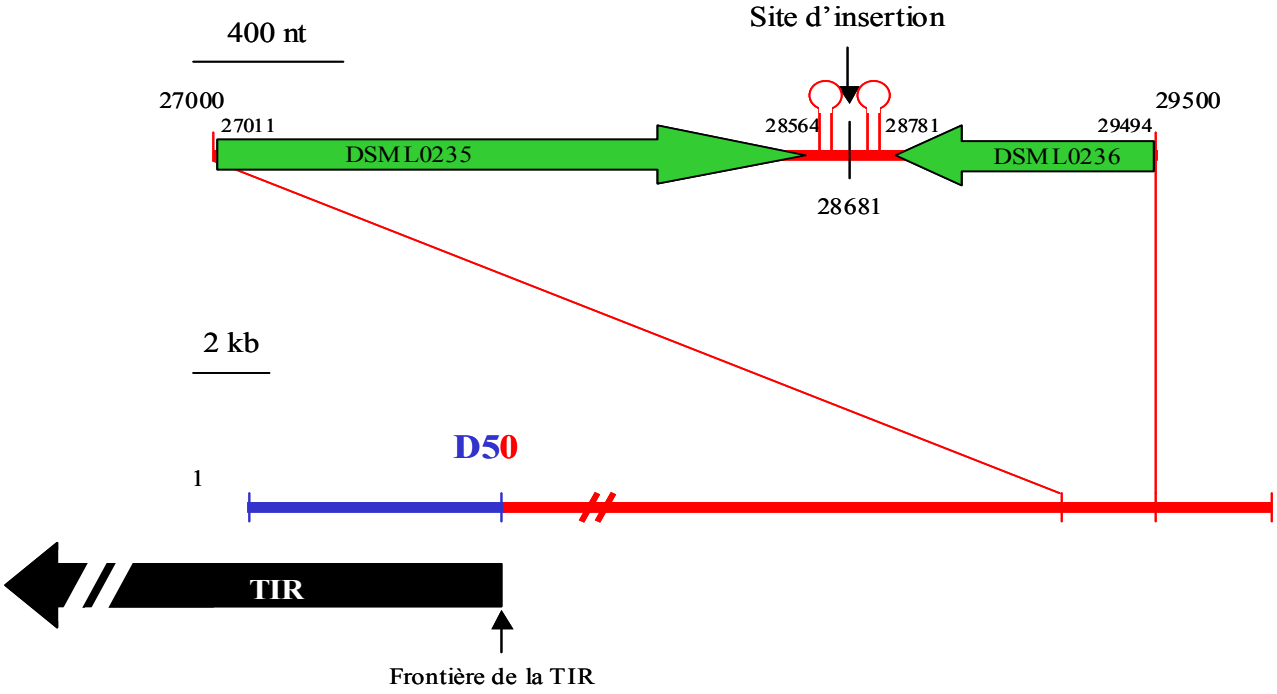


Figure 47 : Localisation du site d'insertion dans le bras droit à la frontière des TIR
Le trait bleu indique la région de l'insert du cosmide D50 comprise dans les TIR, le trait rouge, la région terminale du bras droit en dehors des TIR. Le symbole ♀ représente les terminateurs de transcription.

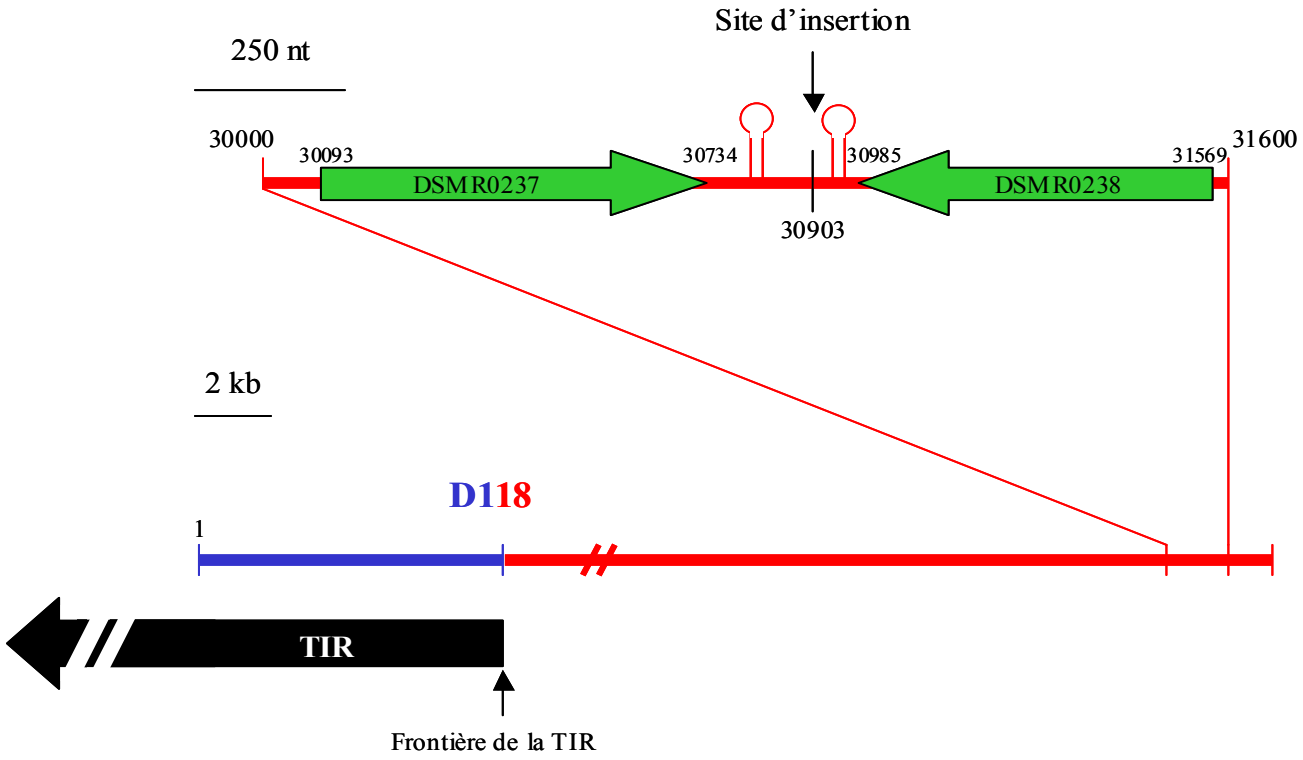


Figure 48 : Localisation du site d'insertion dans le bras gauche à la frontière des TIR
Le trait bleu indique la région de l'insert du cosmide D118 comprise dans les TIR, le trait rouge, la région terminale du bras gauche en dehors des TIR. Le symbole ♈ représente les terminateurs de transcription.

Les autres sites d'insertion correspondent aux gènes *hasL* et *hasR* respectivement pour le bras gauche et le bras droit. Ces gènes codent des facteurs sigma alternatifs. Des études portant sur la délétion de ces gènes ont montré que leur absence n'avait aucune conséquence sur l'instabilité chromosomique chez *S. ambofaciens* (Roth *et al.* 2004). Ainsi, si une instabilité est observée lors du remplacement de ces gènes par les sites *parS*, cette instabilité ne sera due qu'à la présence des sites *parS* et non pas à la délétion de ces gènes. De plus, le remplacement de ces gènes par des cassettes de résistance a déjà été réalisé au sein du laboratoire.

Ces gènes *hasL* et *hasR* sont situés respectivement à 450 kb et à 850 kb des extrémités du chromosome.

B-3 CLONAGE REITERATIF DE SITES *PARS*

Le centre de partition chromosomique chez les *Streptomyces* semble requérir environ une dizaine de sites *parS* pour une fonctionnalité maximale (Jakimowicz *et al.* 2002). Ainsi, pour mimer la présence d'un second centromère, l'insertion de multiples sites *parS* en tandem, grâce à une stratégie de clonage, appelé « réitératif », a été développée.

B-3a Amplicon *parS*

Ce clonage réitératif a été initié par C. Ambroset (DEA MEN Nancy, 2005). Les deux sites *parS* sont amplifiés par PCR (Fig. 49) à partir des séquences du génome de *S. coelicolor*. Ce produit d'amplification de 142 pb est bordé par deux sites *Eco*RI, sites dont la séquence est ajoutée aux extrémités 5' des amorces. Ce produit est appelé « amplicon *parS* ».

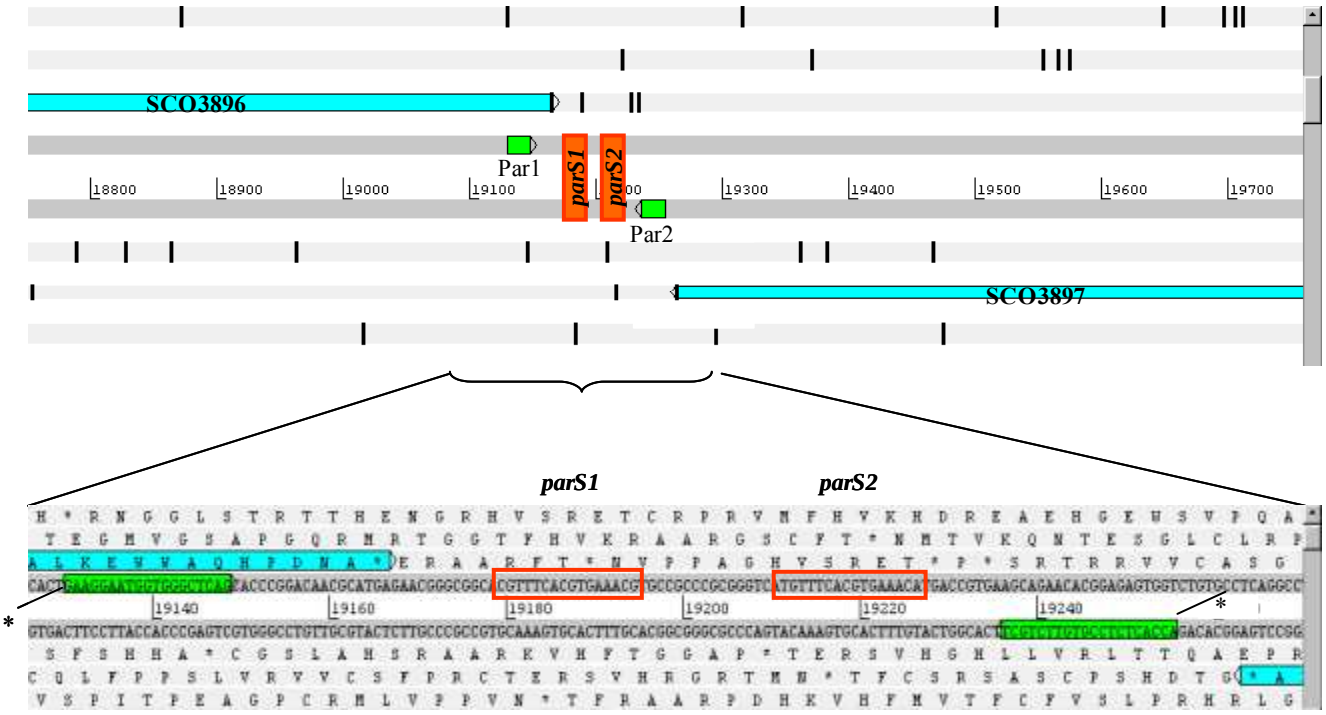


Figure 49 : Localisation de la séquence de *S. coelicolor* des motifs *parS*
Séquence nucléotidique de l'amplicon *parS* visualisée dans une fenêtre ARTEMIS. Les amorces par1 et par2 avec leur extension 5' et le site *Eco*RI (matérialisé par un astérisque) sont schématisées par des boîtes vertes. Les séquences *parS1* et *parS2* sont matérialisées par des cadres rouges.

B-3b Dimère d'amplicons *parS*

Ce produit est ensuite digéré par *Eco*RI et ligué pour permettre la concaténation de ces produits. Des concatémères sont observés jusqu'à 7 copies d'amplicons *parS* (Fig. 50). La purification de ces concatémères a été réalisée sur gel (DEAE cellulose).

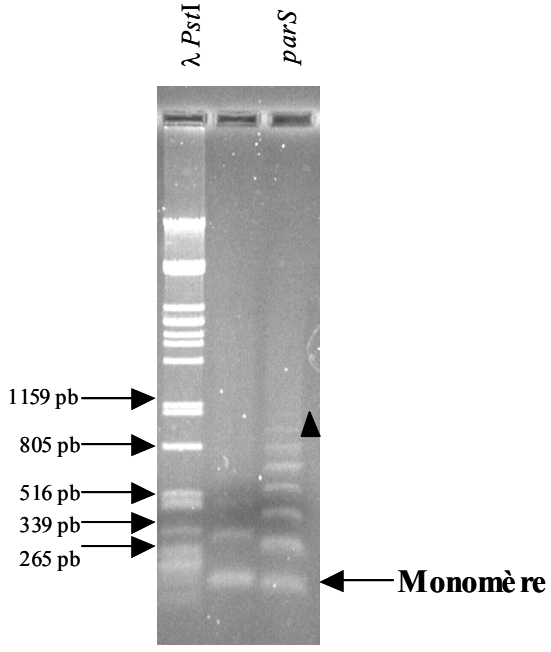


Figure 50 : Ligation des monomères d'amplicons *parS*
Gel agarose 0,8% TBE 1X à 6V/cm. Le triangle noir indique la présence d'un produit de ligation contenant 14 sites *parS*.

B-3c Clonage réitératif

B-3c1 Clonage dans le plasmide pIJ2925

Le plasmide pIJ2925 a été utilisé comme vecteur de clonage (voir Matériels). L'analyse de 50 clones recombinants a permis l'isolation d'un dimère d'amplicon *parS* (Fig. 51). Un dimère d'amplicons *parS* correspond à la ligation de deux produits de PCR de 142 nt. Ce dimère contient alors quatre sites *parS*, chaque produit de PCR contenant deux sites.

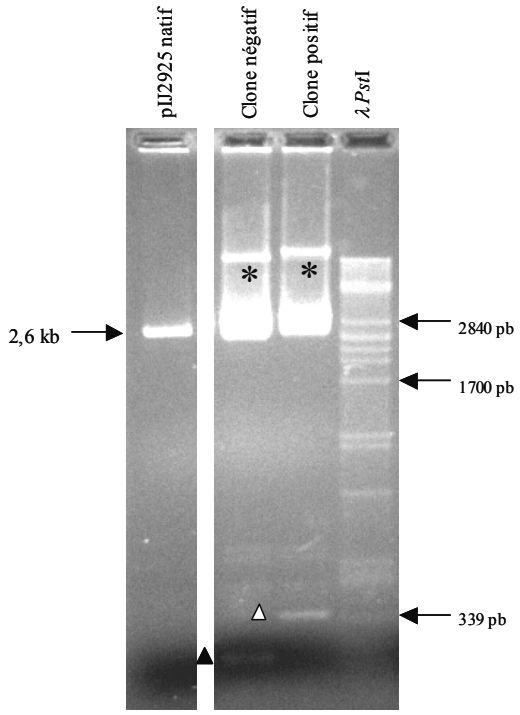
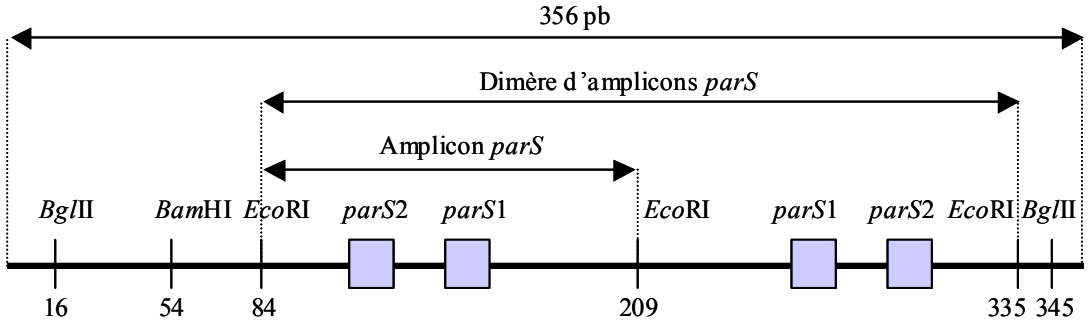


Figure 51 : Clonage du dimère d'amplicon *parS* dans le plasmide pIJ2925.
Vérification de l'insertion du dimère d'amplicon *parS* dans le plasmide pIJ2925 par digestion *Bgl*II. Le triangle noir indique le fragment *Bgl*II sans insert, le triangle blanc indique le fragment *Bgl*II portant l'insert. Gel d'agarose 1,5 % TAE 1X à 6 V/cm.

A partir de ce dimère d'amplicon *parS* cloné dans le plasmide pIJ2925, un produit d'amplification par PCR de 356 pb, comprenant quatre sites *parS*, a été synthétisé (Fig. 52). Ce produit sera digéré par *Bgl*II en raison de la présence à ses deux extrémités des sites de restriction.



*parS*1: CGTTTCACGTGAAACG
*parS*2: TGTTTCACGTGAAACA

Figure 52 : Carte du produit d'amplification du dimère d'amplicons *parS*
La position des différents sites de restriction est indiquée sur le schéma. Les blocs gris correspondent aux séquences *parS*. La séquence des différents *parS* est indiquée sous le schéma.

B-3c2 Clonage dans le plasmide pIJ773*

Le plasmide pIJ773 est utilisé comme vecteur de clonage dans la méthode de clonage réitératif (voir Matériels).

B-3c2.1 Construction du plasmide pIJ773*

La première étape est d'éliminer un des deux sites de restriction *Bam*HI contenu sur le plasmide. En effet, le produit de PCR à cloner contient un site *Bam*HI et est digéré *Bgl*II, compatible *Bam*HI. Après clonage de ce produit, ni le site *Bam*HI ni le site *Bgl*II ne sont régénérés. Pour cela, une digestion partielle de l'ADN de pIJ773 par *Bam*HI a été réalisée (C. Ambroset, DEA MEN Nancy 2005). Les extrémités cohésives sont ensuite comblées par l'activité polymérase 5'-3' de la Klenow. Ainsi, deux plasmides, dérivant du plasmide pIJ773 et présentant un site unique *Bam*HI, ont été obtenus et nommés pIJ773*1 et pIJ773*2 (Fig. 53). L'élimination dans le plasmide pIJ773 de l'un des deux sites *Bam*HI permettra, avec le clonage de l'insert, d'introduire un site *Bam*HI unique qui pourra à nouveau servir de site d'insertion d'un fragment *Bgl*II. Ce cycle de clonage peut être répété. Pour la suite de la présentation, seuls les résultats avec le plasmide pIJ773*1 seront présentés, les expériences décrites ayant été réalisées à partir de plasmides recombinants dérivant de pIJ773*1.

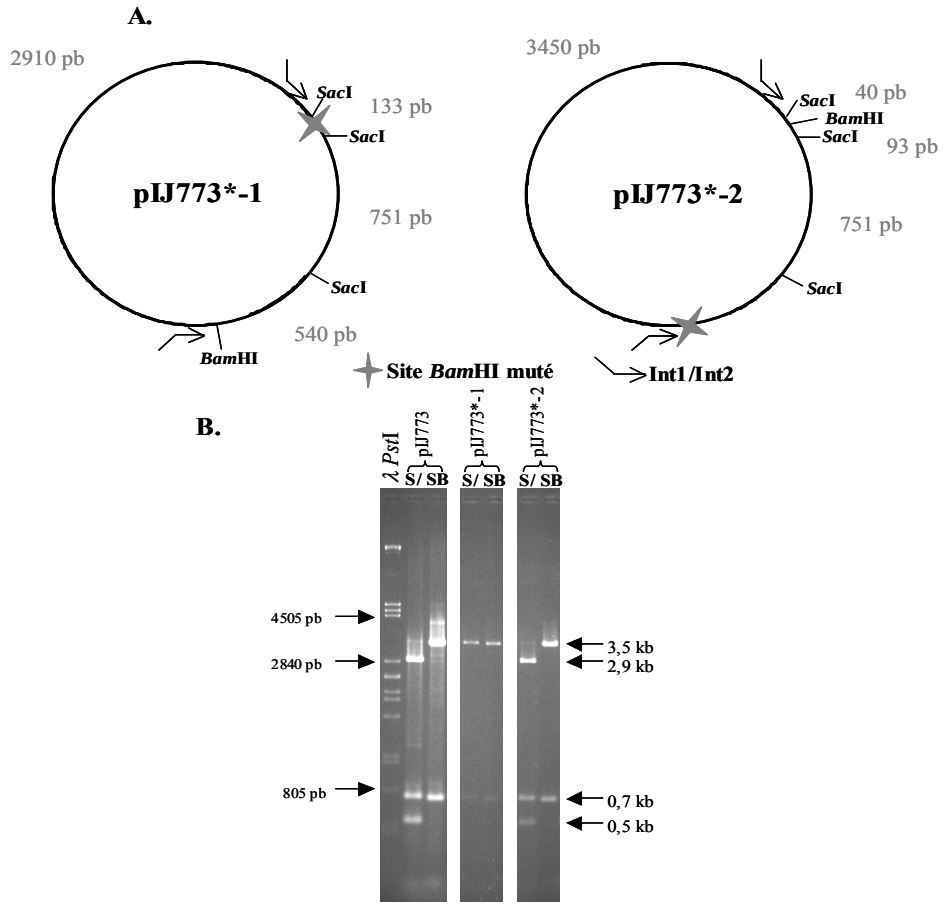


Figure 53 : Construction des plasmides pIJ773*1 et pIJ773*2 par mutation d'un site *Bam*HI
A. Schématisation des sites *Bam*HI mutés. Les tailles des fragments obtenus à l'issue de la double digestion *Sac*I/*Bam*HI sont indiquées. Les amorces Int1 et Int2 sont les amorces utilisées lors des expériences d'amplification pour la modification des différents cosmides.
B. Migration électrophorétique des digestions *Sac*I (S) et *Sac*I/*Bam*HI (S/B) de l'ADN plasmidique de pIJ773 dans le cas d'une inactivation des sites *Bam*HI. Gel d'agarose 1,2% TAE 1X à 1,5 V/cm.

B-3c2.2 Clonage réitératif dans le plasmide pIJ773*1

Chaque plasmide pIJ773* est linéarisé par digestion *Bam*HI. Le produit de PCR *parS*, digéré *Bgl*II, est cloné dans ce plasmide. Ces étapes sont réitérées jusqu'à l'obtention de plasmides recombinants contenant le maximum de sites *parS* (Fig. 54). Pour vérifier le nombre de sites *parS* clonés, un couple d'amorces est utilisé pour amplifier l'insert des plasmides recombinants. La taille du produit d'amplification obtenu permettra de préciser le nombre de sites *parS* clonés.

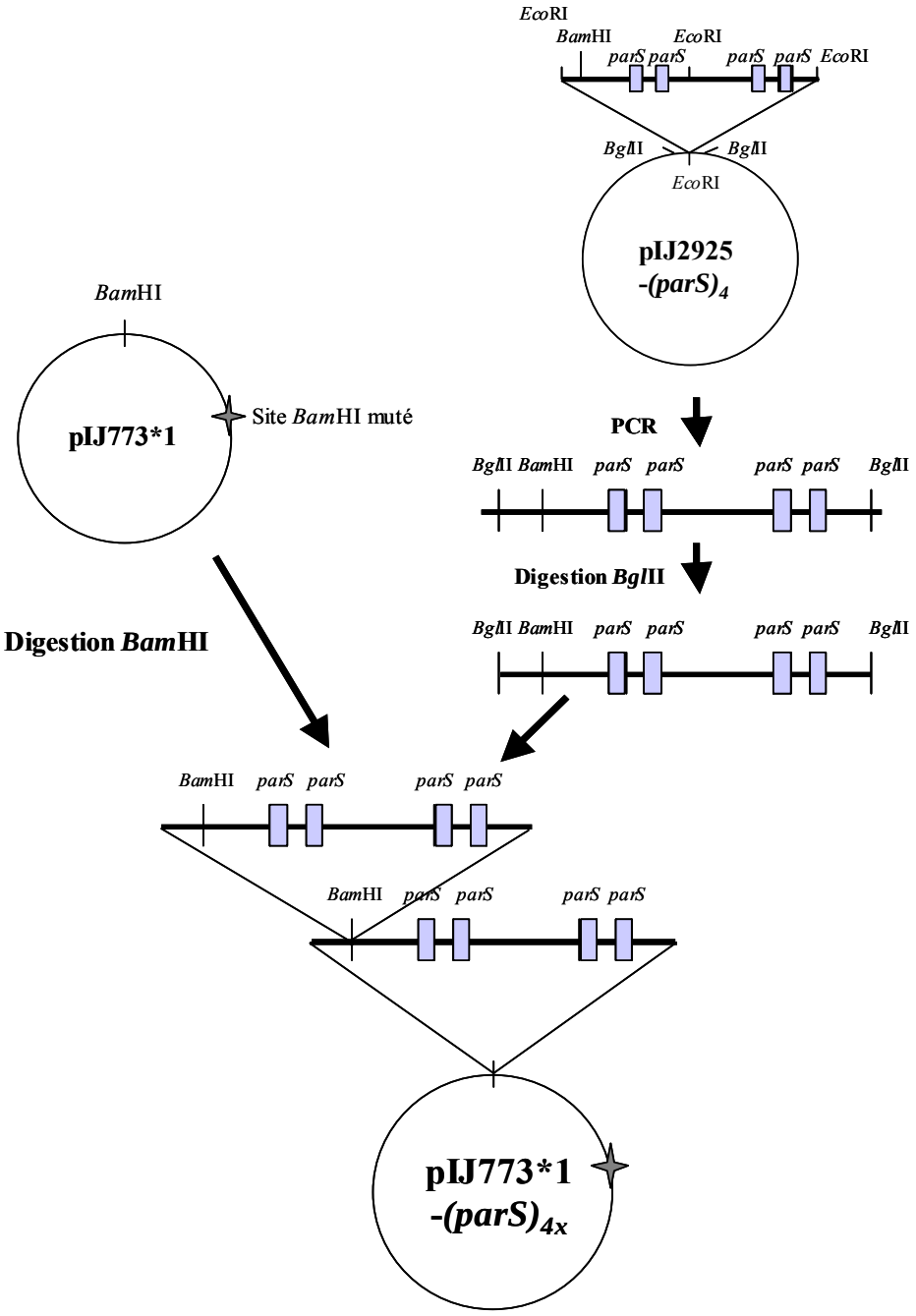


Figure 54 : Schématisation du cycle de clonage réitératif du plasmide pIJ773*1 avec les dimères d'amplicons *parS*
X désigne le nombre de dimères d'amplicons insérés dans le plasmide.

La suite de la démarche est équivalente à celle décrite lors d'un remplacement allélique, les sites *parS* étant introduits aux sites choisis.

Le dimère d'amplicons a été amplifié par PCR, digéré par *Bgl*II et cloné dans le site unique *Bam*HI du plasmide pIJ773*-1. Des amorces, qui encadrent le site de clonage, ainsi que l'ensemble de la cassette Apra, sont utilisées pour vérifier la présence d'amplicons *parS*. Un produit d'amplification de 1993 pb est obtenu lors du clonage du dimère d'amplicons *parS* pour 1660 pb à la cassette Apra seule. La figure 55 présente le résultat de la migration en gel d'électrophorèse de deux produits d'amplification, l'un correspondant au plasmide pIJ773*1 avec

un dimère d'amplicons *parS* (piste 1) et l'autre à la cassette Apra seule. Ces plasmides sont nommés pIJ773*D2 par la suite. Ces plasmides contiennent donc quatre sites *parS*.

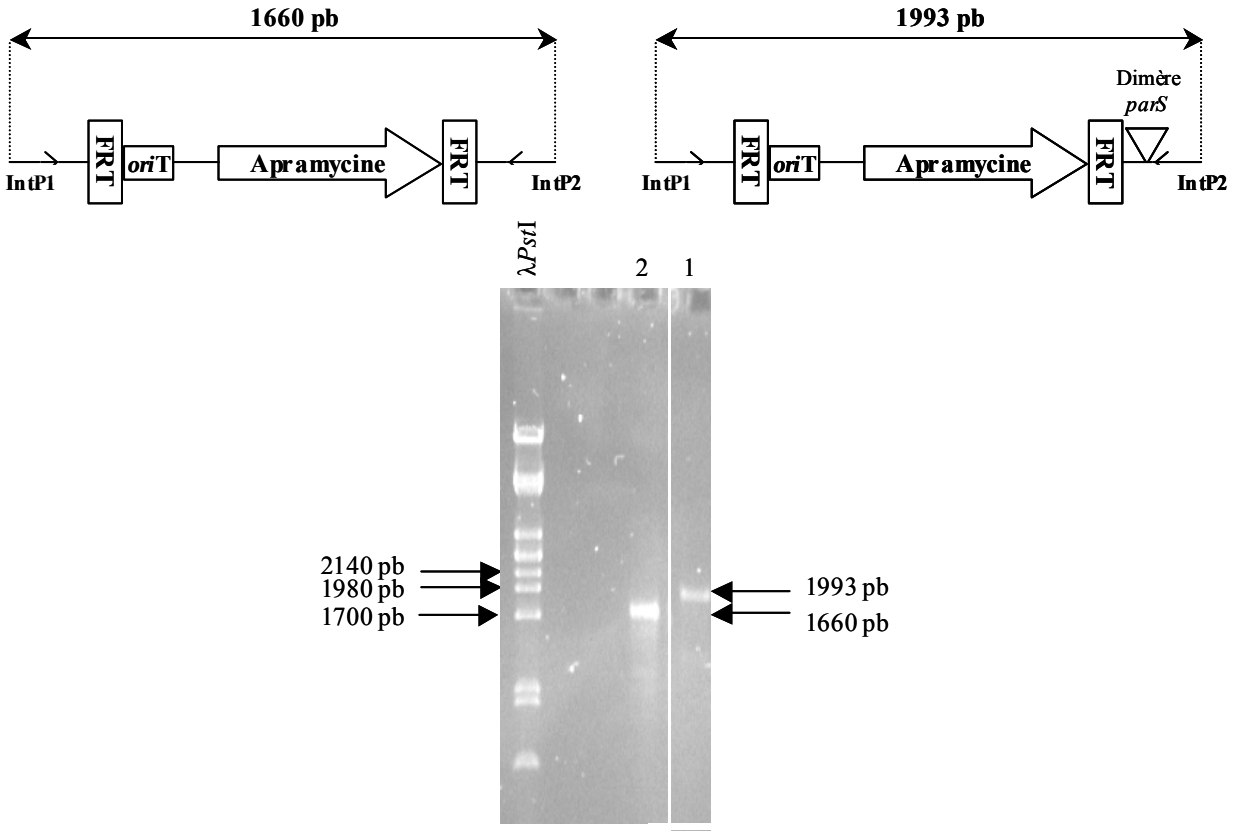


Figure 55 : Vérification du clonage du dimère d'amplicons *parS* dans le plasmide pIJ773*-1 par PCR (couple d'amorce IntP1/IntP2)
La piste 1 présente un produit d'amplification dont la taille correspond au plasmide recombinant alors que le produit d'amplification de la piste 2 correspond au plasmide natif. Gel agarose 0,8% TBE 1X à 5V/cm.

Le clonage réitératif a permis de construire des plasmides de D4 à D10 (Fig. 56). La lettre correspond à dimère et le chiffre, au nombre d'amplicons *parS* initiaux clonés. Ainsi, un plasmide nommé D4 contient 4 amplicons initiaux soit 8 sites *parS*, les amplicons initiaux contenant deux sites *parS*.

Ces plasmides vont servir de matrice lors des réactions d'amplification.

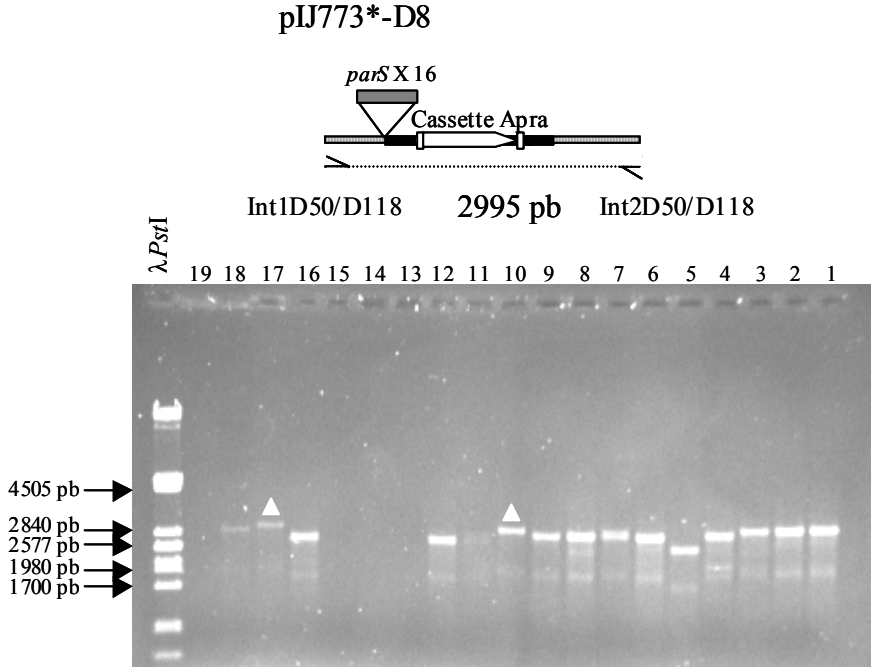


Figure 56 : Migration électrophorétique de produits d'amplification pIJ773*-D8
Les triangles blancs indiquent les clones retenus comme positifs. Gel agarose 0,6 % TBE 1X à 4V/CM.

Après l'obtention d'un plasmide recombinant contenant 20 sites *parS*, la construction de la version D12 a été entreprise. Un clone, présentant le plasmide recombinant pIJ773*1-D12, a été obtenu mais le très faible rendement des réactions d'amplification au vu de l'expérience de PCR-Targeting, ne nous a pas permis de l'utiliser pour la suite des expériences.

C- Modification des cosmides.

C-1 AMPLIFICATION DES CASSETTES APRA-*PARS*

L'amplification des cassettes est réalisée à partir de l'ADN des plasmides pIJ773*1, pIJ773*D8-10 et pIJ773*D10-40 et grâce aux couples d'amorces Int1D50/Int2D50, Int1D118/Int2D118, Int125E1/Int225E1 et Int114C4/Int214C4. Ces amorces contiennent 18-19 nucléotides qui vont s'hybrider d'un côté de la cassette Apra (Int1) et du côté des sites *parS* (Int2). Le reste de la séquence de ces amorces est spécifique de chaque cosmide et est homologue avec les régions encadrant le site d'insertion.

Les plasmides sont, au préalable, digérés par *Ase*I, enzyme qui possède deux sites de restriction encadrant la cassette et les sites *parS*. Cette digestion permet d'éviter une contamination, durant les transformations des souches d'*E. coli* BW25113/pIJ790 contenant le cosmide d'intérêt (14C4, D118…) par le plasmide utilisé comme matrice durant les réactions d'amplification (Fig. 57).

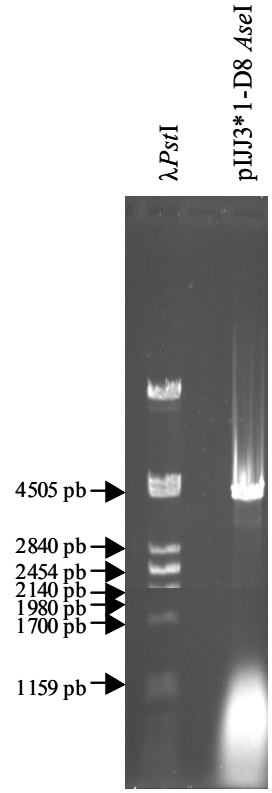


Figure 57 : Profil de restriction *Ase*I du plasmide pIJ773*1-D8
Gel agarose 0,8% TBE 1X à 5V/cm.

Ainsi, 12 produits d'amplification ont été obtenus, trois pour chaque couple d'amorces. Quel que soit le couple d'amorces, les régions des amorces s'hybridant avec le plasmide recombinant étant communes à chaque amorces, la taille des produits d'amplification avec le plasmide pIJ773*, pIJ773*D8-10 et pIJ773*D10-40 est respectivement de 1,66 kb, 2,98 kb et 3,3 kb. Des produits d'amplification correspondant à ces tailles ont été obtenus et utilisés lors des expériences de modifications des différents cosmides.

C-2 INCORPORATION DES SITES *PARS* DANS LES COSMIDES

Chaque produit d'amplification est utilisé pour transformer des cellules compétentes d'*E. coli* BW25113/pIJ790/cosmide. Les cellules compétentes ont été préparées à partir de cultures supplémentées en L-arabinose pour induire le système λRED. Plusieurs clones pour chaque cosmide ont été analysés par profil de restriction *Bam*HI sur gel d'électrophorèse et/ou par PCR. Les résultats présentés seront ceux obtenus pour le cosmide D118. Pour les autres cosmides, le même schéma a été utilisé pour permettre la vérification des cosmides modifiés.

Pour le cosmide D118, un fragment *Bam*HI de 7 kb disparaît au profil de deux fragments *Bam*HI. L'intégration de la cassette Apra seule provoque l'apparition de fragments *Bam*HI de 6,2 et 2,5 kb. En présence de 16 sites *parS*, ces deux fragments présentent une taille de 7,5 et 2,7 kb.

Parallèlement, l'amplification avec le couple d'amorces vD118P1-vD118P2 (1,95 kb pour la cassette vide et 3,27 kb pour 16 sites *parS*) (Fig. 58 et 59) est utilisée pour vérifier l'intégration

des sites *parS* au locus choisi sur le cosmide modifié. Ainsi, les cosmides D118 contenant la cassette seule et D8 ont été obtenus.

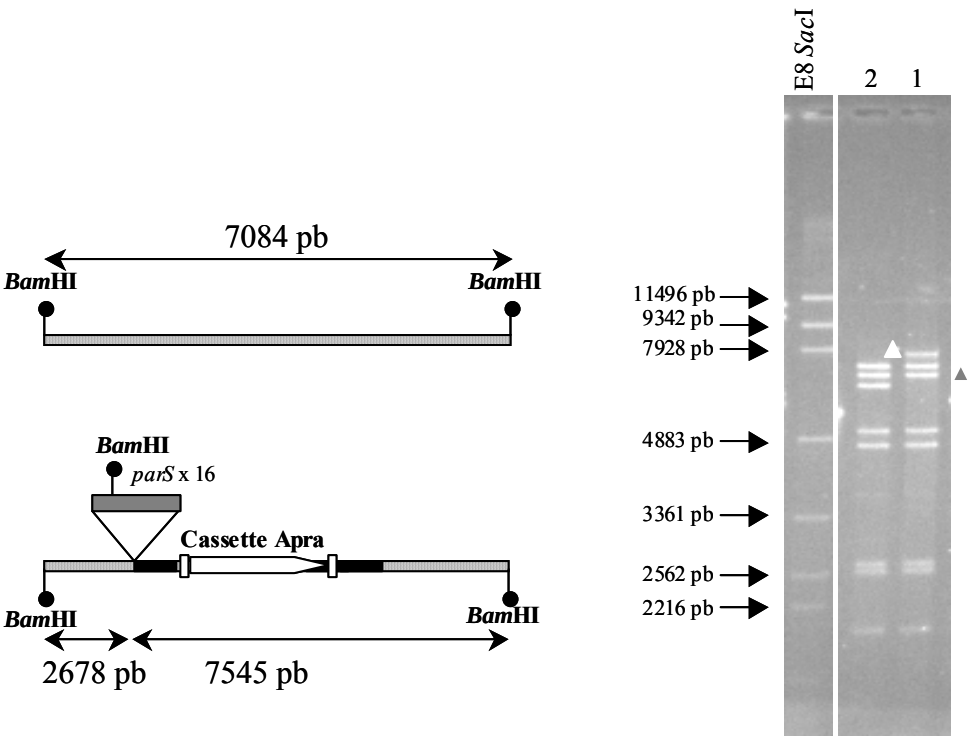


Figure 58 : Profil de restriction *Bam*HI de cosmides D118-D8 recombinants
La piste 2 présente le profil du cosmide D118 natif. La piste 1 présente un profil de restriction de cosmide D118-D8. Le triangle gris indique la bande qui disparaît après insertion au profil de la bande marquée par le triangle gris.
Gel agarose 0,8% TAE 1Xà 2 V/cm.

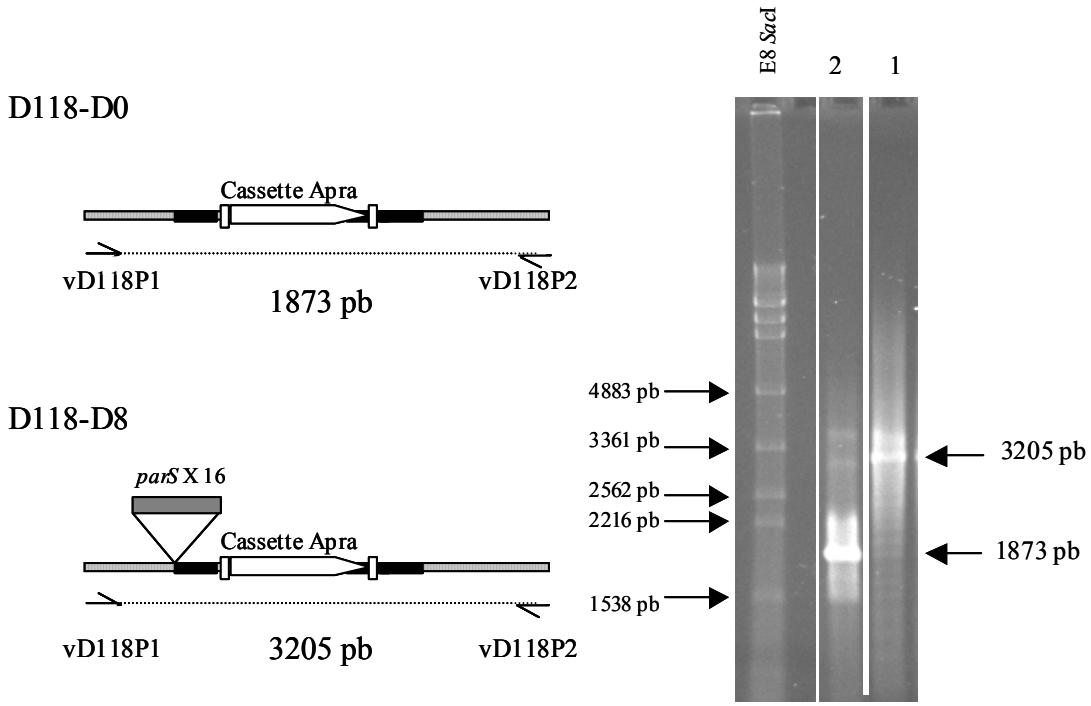


Figure 59 : Migration électrophorétique des produits d'amplification de vérification de cosmides recombinants D118
La piste 1 correspond au cosmide recombinant D118-D8 dont le produit d'amplification mesure 3205 pb.
La piste 2 correspond au cosmide D118-D0, présentant un produit d'amplification de 1873 pb.
Gel d'agarose 0,8% TAE 1X à 2 V/cm.

Pour le cosmide D50, des vecteurs contenant la cassette Apra seule et le cosmide D50-D8 ont été obtenus. Pour le cosmide 14C4, des vecteurs avec la cassette apra seule et cette cassette avec 8 sites *parS* ont été obtenus alors que pour le cosmide 25E1, la cassette apra seule a été introduite. Ces différentes constructions ont été électroporées dans la souche d'*E. coli* ET12567/pUZ8002 et vérifiées par profil de restriction. Ces différentes souches d'*E. coli* ET12567/pUZ8002 sont ensuite utilisées pour les expériences de conjugaison entre *E. coli* et *S. ambofaciens*.

D- Obtention des souches S. ambofaciens *DSM40697 modifiées*

D-1 ANALYSE DES SOUCHES DE *S. AMBOFACIENS* DSM40697

Les surnageants de culture des transconjugants obtenus, sont utilisés comme matrice lors de réaction d'amplification par PCR. Ces réactions d'amplification permettent de vérifier l'intégrité du chromosome. Pour cela, cinq couples d'amorces sont employés (tableau 4):

Couple d'amorce	Taille des produits d'amplification (en pb)
CL1-CL2	1161
CR1-CR2	1182
T1-18.1.2	369
OriC1-OriC2	1121
12A10-49-2	500

Tableau 4 : Couples d'amorces utilisés et taille des produits d'amplification.

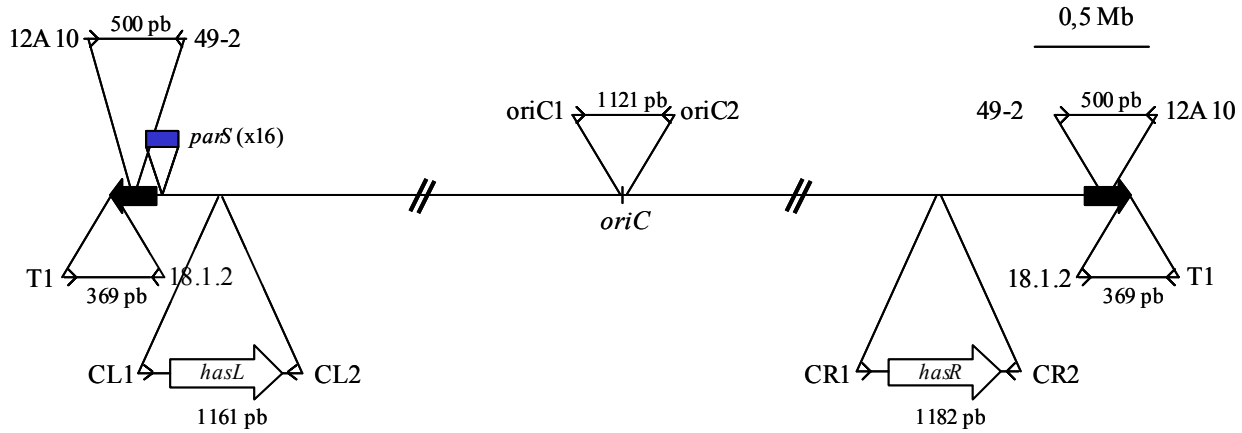
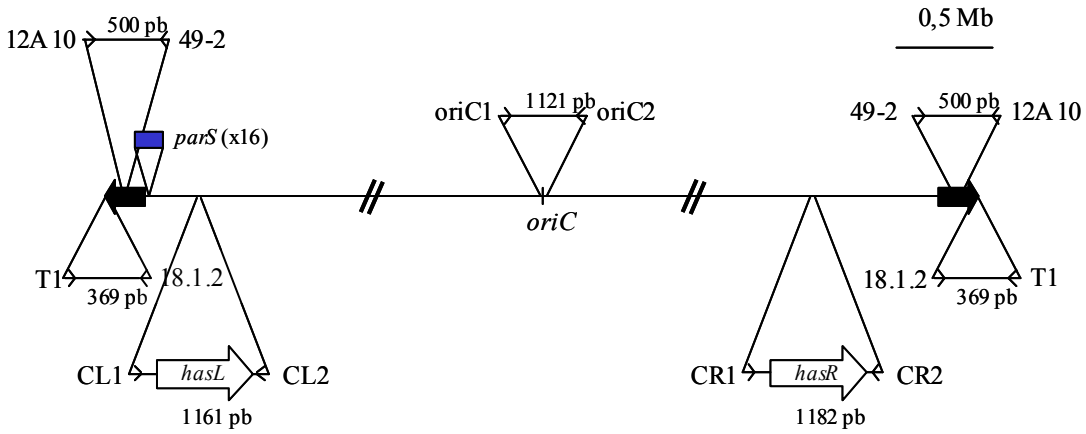


Figure 60 : Localisation des différents couples d'amorces
Les têtes de flèches représentent les amorces. Leur nom est indiqué à côté de chaque tête de flèche. La taille des différents produits d'amplification sont indiqués pour chaque couple d'amorces. Les flèches noires correspondent aux TIR. Le carré bleu représente les 16 sites *parS*.

Ainsi, deux souches modifiées (une avec la cassette Apra seule, l'autre avec les 16 sites *parS* situés dans le bras gauche) pour les lignées DSM40697-1, 2 et 3 ont été analysées. Pour chacune de ces six souches, 12 clones indépendants ont été cultivés en milieu HT liquide pendant 72h à 30°C. Le surnageant de culture a été utilisé comme matrice pour les réactions d'amplification avec les différents couples d'amorces.

Le couple d'amorces OriC1-OriC2 sert de témoin positif, la présence de l'origine de réplication *oriC* étant indispensable à la survie des cellules. Ainsi, 12 clones ne présentent aucun signaux d'amplification pour ce couple d'amorces révélant un probléme technique lié à la matrice d'ADN. Ces clones n'ont pas été retenus pour la suite de cette analyse.

Les autres clones peuvent être regroupés en huit « profils » suivant la présence ou l'absence de produits d'amplification. Des produits d'amplification sont détectés avec les cinq couples d'amorces pour 17 clones (« profil 1 »). La figure 61 résume le profil et l'effectif de ces différentes catégories.



	T1/18.1.2 (Télomères)	12A10/49-2 (TIR)	CL1/CL2 (*hasL*)	OriC1/OriC2 (origine de réplication)	CR1/CR2 (*hasR*)	12A10/49-2 (TIR)	T1/18.1.2 (Télomères)	Effectif DO	D8	Total
Profil 1:	—	—	—	—	—	—	—	8/32	9/28	17
Profil A:	—	—	—	—	.....	—	—	0/32	5/28	5
Profil B:	—	—	.....	—	—	—	—	7/32	4/28	11
Profil C:	.....	—	.....	—	.....	—	.....	8/32	1/28	9
Profil D:	.....	—	—	—	—	—	.....	2/32	3/28	5
Profil E:	.....	—	—	—	.....	—	.....	6/32	3/28	9
Profil F:	.....	—	.....	—	—	—	.....	0/32	3/28	1
Profil G:	—	—	.....	—	.....	—	—	1/32	0/28	3

Figure 61: Schématisation des profils d'amplification obtenus.
Les traits pleins indiquent l'obtention d'un produit de PCR et les pointillés l'absence d'amplification. Les effectifs observés pour la cassette Apra seule et 16 sites *parS* pour chaque catégorie sont indiqués par rapport à l'effectif total sur le nombre de clones contenant la cassette Apra et les clones avec 16 sites *parS* étudiés.

A partir de ces données, la structure potentielle des génomes de ces différents clones peut être envisagée.

Les effectifs des différents types de profils dans les populations de souches portant la cassette Apra seule ou la cassette avec les sites parS ont été comparés grâce au test du $\chi 2$. Seuls les profils A et C pourraient être influencés par la présence/absence des motifs *parS* au seuil de 5%. En revanche, au seuil de 1%, aucune différence significative n'est observée pour l'ensemble des profils obtenus.

DISCUSSION-PERSPECTIVES

DISCUSSION

Les *Streptomyces* ont été étudiés depuis des décennies entre autres raisons pour l'instabilité génétique observée (Leblond et Decaris 1999; Volff et Altenbuchner 2000; Chen *et al.* 2002). Cette instabilité est associée à des réarrangements d'ADN, tels que la circularisation du chromosome, des délétions ou des fusions de chromosomes (Fischer *et al.* 1997a; Fischer *et al.* 1998; Wenner *et al.* 2003). Ces grands réarrangements affectent les régions terminales du chromosome des *Streptomyces*. L'analyse fonctionnelle des génomes séquencés de *S. coelicolor* et *S. avermitilis* a montré que ces régions terminales étaient enrichies en gènes spécifiques d'espèces, contrairement à la région centrale qui contient les gènes conservés et essentiels à la croissance végétative de la bactérie (Ikeda *et al.* 2003).

La comparaison des régions terminales de la souche de *S. ambofaciens* ATCC23877 avec les autres espèces de *Streptomyces*, dont le génome est entièrement séquencé, a permis l'identification d'un phénomène nommé synténie dégénérée, où le niveau de synténie entre deux génomes diminue graduellement vers les extrémités alors que les régions centrales sont fortement conservées. Cette perte graduelle de la synténie serait la conséquence d'événements d'insertions/délétions.

Au niveau intraspécifique, la comparaison des séquences des répétitions terminales inversées (ou TIR) des souches de *S. ambofaciens* ATCC23877 et DSM40697 a montré que les régions variables spécifiques de souches sont situées à l'extrémité. Cette variabilité résulterait d'événements d'échanges d'extrémités entre réplicons linéaires. Les fonctions et les caractéristiques génomiques de ces séquences spécifiques permettent de faire des hypothèses sur l'histoire évolutive du génome chez les *Streptomyces*.

I- Le génome des Streptomyces est compartimenté

L'étude des régions terminales de la souche de *S. ambofaciens* ATCC23877 a permis de montrer que la longueur des régions de synténie varie en fonction de la paire de génomes comparés. Le niveau de GOC (Gene Order Conservation) a été utilisé comme indice de mesure du niveau de synténie le long du chromosome. Ainsi, lors de la comparaison entre deux espèces phylogénétiquement proches, *S. ambofaciens* et *S. coelicolor*, la taille des régions spécifiques pour *S. ambofaciens* est de 1279 kb et de 753 kb pour *S. coelicolor*. En revanche, par comparaison avec *S. avermitilis*, espèce phylogénétiquement plus éloignée, la taille des régions spécifiques de *S. ambofaciens* est de 1878 kb et celle de *S. avermitilis* de 1393 kb (Fig. 62).

Ainsi, pour deux espèces de *Streptomyces* phylogénétiquement éloignées, la taille de leurs régions spécifiques sera plus grande qu'entre deux espèces étroitement apparentées. Réciproquement, la taille de la région centrale héritée de l'ancêtre commun diminue avec la distance phylogénétique qui sépare les espèces comparées.

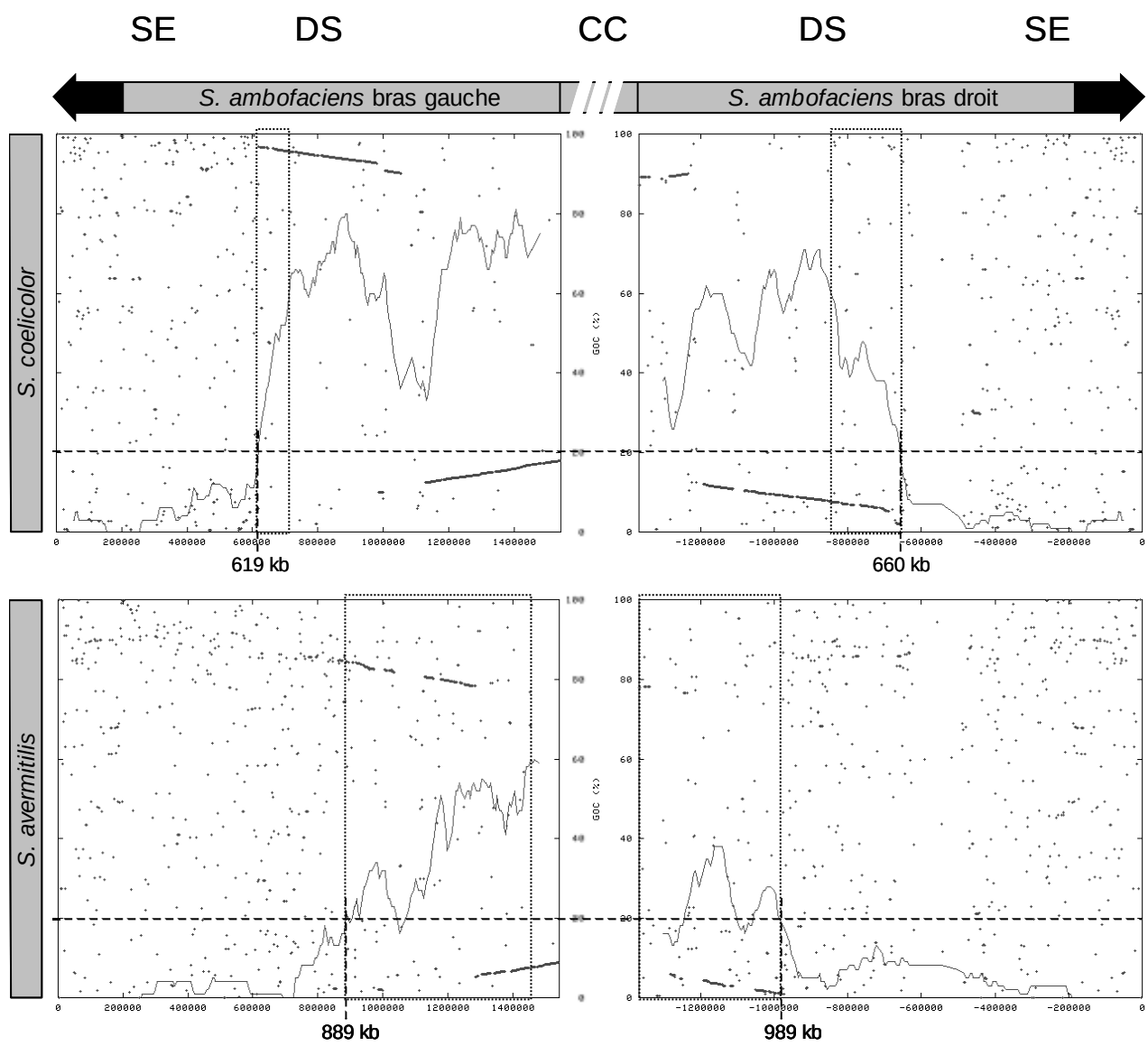


Figure 62 : Mesure du GOC le long des bras chromosomiques de *S. ambofaciens* comparés aux génomes de *S. coelicolor* et *S. avermitilis*

Sur chaque graphique, le niveau de GOC est superposé au dot-plot. Un seuil minimal de GOC de 20% a été choisi pour délimiter les régions terminales spécifiques d'espèces. SE : extrémités spécifiques ; DS : régions de synténie dégénérée ; CC : région centrale conservée. Les régions de synténie dégénérée (entre 20 et 60% de GOC) sont encadrées. La taille de ces régions de synténie est plus grande entre *S. coelicolor* et *S. ambofaciens*, phylogénétiquement plus proches qu'entre *S. ambofaciens* et *S. avermitilis*, espèces plus éloignées.

Calculs réalisés avec une fenêtre glissante de 100 gènes et un pas de 5 gènes.

La présence de cette synténie dégénérée montre que les gènes présents dans ces régions ne sont pas issus d'un échange entre extrémités de réplicons linéaires mais vraisemblablement d'événements de recombinaison multiples. En effet le remplacement d'extrémités ferait varier l'ensemble du contenu en gènes depuis le point de recombinaison jusqu'aux télomères.

Les régions terminales évolueraient principalement par fixation d'événements d'insertions/délétions (indels) au cours des temps évolutifs. La dégénérescence graduelle de la synténie montre que la fréquence de ces événements est variable suivant la localisation chromosomique et augmente à proximité des extrémités. Par saturation en indels, la synténie ancestrale est progressivement effacée.

La comparaison des TIR entre les deux souches de *S. ambofaciens* a montré que les régions spécifiques de souches étaient situées aux extrémités de ces TIR. Ces régions mesurent 60 kb et 48 kb respectivement pour la souche DSM40697 et ATCC23877. En revanche, les séquences télomériques sont très conservées.

Ces résultats mettent en évidence deux mécanismes permettant l'évolution de ces régions terminales :

- insertion/remplacement d'extrémités plasmidiques ou chromosomiques. Ce type d'événement serait « sporadique » et permettrait de générer un remplacement complet des fonctions terminales et l'acquisition potentielle de nombreux gènes en un seul événement.

- flux d'indels « constant » au cours des temps évolutifs. Le mécanisme permettant l'afflux d'information ainsi que les mécanismes de recombinaison permettant sa fixation à court terme restent à déterminer.

II- Dynamique de l'évolution des TIR

A- Homogénéisation du contenu des TIR

Les approches d'hybridation moléculaire comme le séquençage des régions frontières des TIR n'ont pas pu mettre en évidence de divergence entre les répétitions d'un même chromosome.

En revanche, la recombinaison fréquente entre les TIR pourrait homogénéiser ces dernières. Ainsi, l'étude du cluster *alp*, situé dans la région des TIR commune aux deux souches de *S. ambofaciens* est réalisée au laboratoire (B. Aigle, communication personnelle). Le remplacement allélique de certains gènes de ce cluster dans l'un des deux bras a pour conséquence le remplacement de l'homologue par l'allèle mutant.

La formation d'un crossing-over entre deux chromosomes fils ou la réparation d'une cassure double brin dans les TIR peuvent entraîner l'homogénéisation des séquences.

Ces deux mécanismes ont été proposés quant à la formation de translocations chromosomiques chez *S. ambofaciens* (Fischer *et al.* 1998). Ainsi, deux gènes dupliqués, *hasR* et *hasL*, localisés dans les régions spécifiques des régions terminales constituent le substrat d'événements de recombinaison homologue menant à la perte de l'extrémité subterminale d'un bras et son remplacement par une copie du bras intact (Fig. 63).

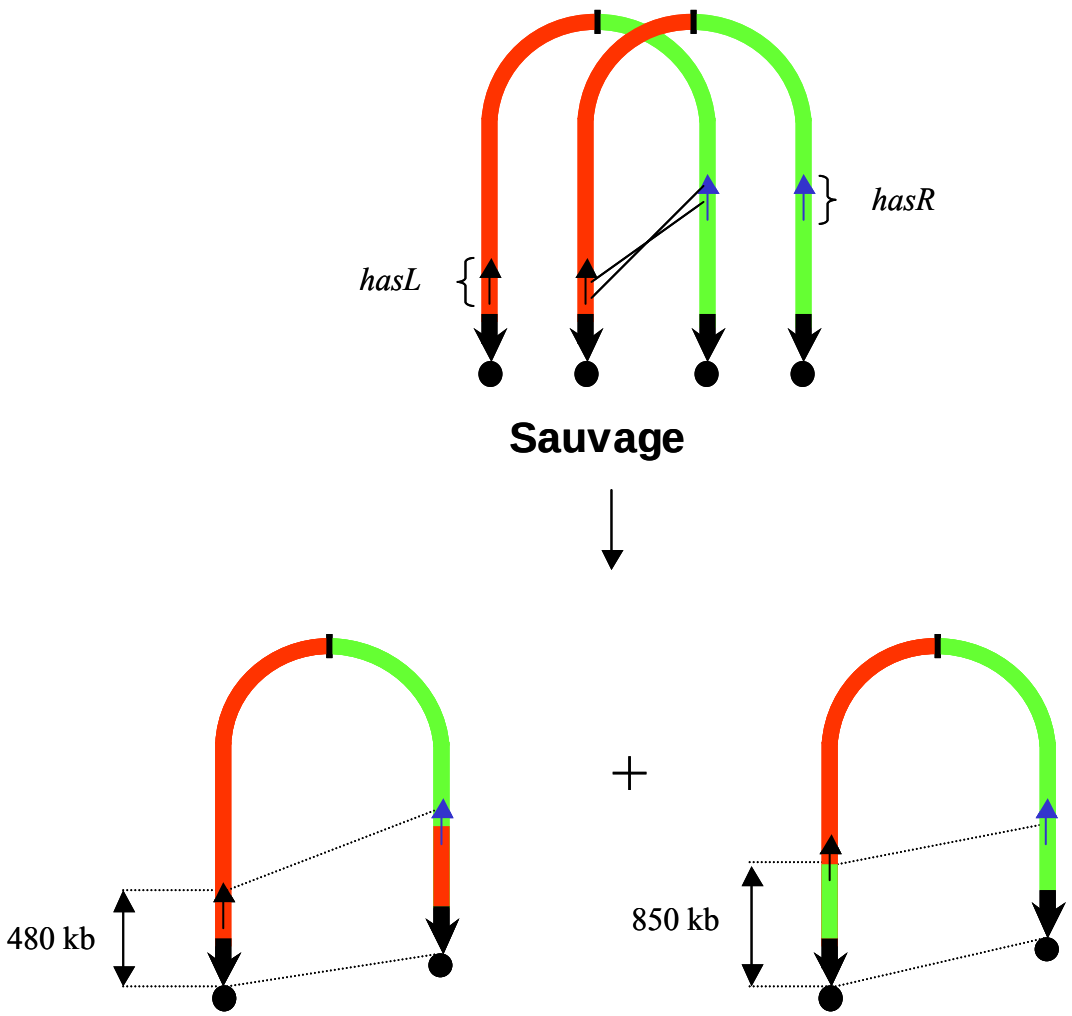


Figure 63 : Remplacement d'un bras chromosomique par recombinaison homologue entre les gènes *hasL* et *hasR* (d'après Fischer *et al*., 1998)
La recombinaison homologue entre les gènes *hasR* et *hasL* pourrait être à l'origine de l'apparition de souches de *S. ambofaciens* présentant des TIR de 480 et 850 kb.

La seconde hypothèse implique la formation de cassure double brin et leur réparation. Celles ci peuvent résulter de :

- l'arrêt de la fourche de réplication en raison de la présence de courts motifs répétés (Viguera *et al.* 2001). Le motif ACCCC, situé aux frontières des TIR, est surreprésenté (respectivement 59 et 47 fois dans les souches DSM40697 et ATCC23877).

- la cassure chromosomique par effet « guillotine » au cours de la division cellulaire. Ce phénomène a été mis récemment en évidence chez des mutants du gène *ftsK* chez *S. coelicolor* (Wang *et al.* 2007). La protéine FtsK intervient, chez *E. coli*, dans la ségrégation chromosomique et la formation du septum (Begg *et al.* 1995; Yu *et al.* 1998; Aussel *et al.* 2002). Cette protéine présente un rôle d'ADN translocase qui permet de distribuer les chromosomes néorépliqués dans les cellules filles. Un homologue du gène *ftsK* est présent chez *S. coelicolor* (SCO5750) dont la délétion provoque une hétérogénéité dans le phénotype des colonies, associée à des délétions

terminales (jusqu'à 10% de la taille du chromosome). La déficience de FtsK induirait un retard de cheminement du chromosome au cours de la partition dans les spores. La fermeture du septum sur le chromosome générerait alors une cassure de ce dernier. Dans le contexte sauvage, ce phénomène pourrait intervenir à basses fréquences et participer à l'instabilité terminale. Ceci pourrait induire des cassures soit préférentiellement dans les régions terminales, les plus éloignées du centromère bactérien, soit de façon équivalente tout au long du chromosome. Les cassures dans la région centrale seraient fortement contre-sélectionnées. Jakimowicz *et al*. (2007) ont montré que la délétion du gène *parA* provoque une septation irrégulière et une ségrégation anormale lors de la formation des spores. Ainsi la protéine ParA interagirait avec FtsZ pour la formation de l'anneau Z et pourrait être un acteur dans la formation de cassures double brin.

B- Evolution de la taille des TIR

La frontière des TIR a été définie comme le dernier nucléotide commun aux deux bras chromosomiques. Ainsi, entre les TIR des deux souches de *S. ambofaciens*, la frontière n'est pas située au même locus. La présence de motifs C/A à la borne des TIR peut provoquer des glissements de l'ADN polymérase et ainsi entraîner une divergence du nombre de répétitions entre les deux bras modifiant l'homologie entre les TIR des deux bras.

Chez *S. avermitilis*, ces TIR se limitent aux seules séquences télomériques (169 pb) mais l'analyse des extrémités chromosomiques révèle des traces de TIR ancestrales plus longues. En effet, à une extrémité, une CDS initialement non décrite, nommée SAV0, située entre les télomères et SAV1 sur le bras gauche, présente 87% d'identité en nucléotides avec SAV7573, avant dernière CDS décrite sur le bras droit (Fig. 64). De plus, les CDS SAV6 et SAV7571, présentes dans chacune des extrémités, sont également homologues (38% d'identité).

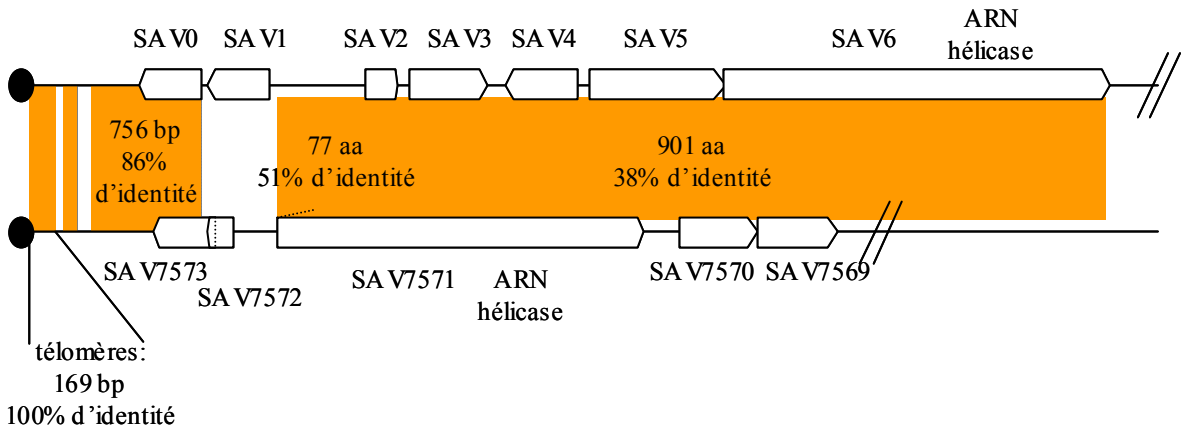


Figure 64 : Les extrémités chromosomiques de *S. avermitilis*
Les TIR de *S. avermitilis* se limitent aux 169 nucléotides télomériques. L'analyse des séquences montre la présence de SAV0 présentant des homologies avec SAV7573 et de SAV2 et SAV6 homologues de SAV7571.

Ces données permettent de spéculer que les séquences des TIR divergent progressivement par accumulation de mutations qui vont réduire l'efficacité de la recombinaison homologue limitant la capacité d'homogénéisation. Alternativement, chez *S. avermitilis*, la perte de ces TIR

pourrait s'expliquer par un événement unique de remplacement d'une extrémité chromosomique.

L'accumulation de mutations ponctuelles pourrait donc conduire à la réduction en taille des TIR. Cette réduction pourrait être contrebalancée par des événements de recombinaison provoquant leur expansion. Chez *S. coelicolor*, les TIR mesurent 22 kb mais des souches présentant des TIR de 1,06 Mb ont été isolées en laboratoire. Ces souches correspondraient à la situation ancestrale dans la généalogie des souches de *S. coelicolor* lors de son isolement du sol (Weaver *et al.* 2004). La réduction de la taille des TIR serait la conséquence d'un événement de recombinaison intrachromosomique entre deux séquences identiques de 1,7 kb incluant SCO0020 (Weaver *et al.* 2004). L'instabilité de la longueur des TIR peut être stimulée par l'exposition aux UV, vraisemblablement par l'induction de la recombinaison homologue au cours de la réponse SOS (Weaver *et al.* 2004). Expérimentalement, l'exposition aux UV réduit la longueur des TIR à une fréquence allant jusqu'à 30%.

III- Mécanismes de la variabilité des régions terminales

A- Conjugaison chez les Streptomyces

Contrastant avec les mécanismes conjugatifs chez les bactéries Gram -, et avec le paradigme du facteur F, les mécanismes conjugatifs chez les *Streptomyces* dépendent de la présence d'une seule protéine membranaire similaire aux ADN translocases, la protéine Tra.

A-1 LES PLASMIDES CONJUGATIFS

Les *Streptomyces* présentent un grand nombre de plasmides conjugatifs de configuration et de tailles variables : du plasmide circulaire pIJ101 isolé de *S. lividans* (Kieser *et al.* 1982) à de grands plasmides linéaires comme SCP1 (360 kb) de *S. coelicolor* (Hopwood *et al.* 1983). Certains de ces éléments sont des ICE (pour Integrative and Conjugative Element) comme le plasmide pSAM2 de *S. ambofaciens* (Pernodet *et al.* 1984). En effet, après son transfert, pSAM2 va s'intégrer au chromosome par recombinaison site spécifique au niveau de l'extrémité 3' d'un gène ARNt.

La plupart de ces plasmides est transférée à haute fréquence pouvant atteindre 100% dans certains cas (Kieser *et al.* 1982). La mobilisation de marqueurs chromosomiques a été observée à des fréquences allant de 0,1 à 1% (Hopwood *et al.* 1985).

A-2 LE TRANSFERT CONJUGATIF

Le transfert conjugatif, par rapport au transfert par exemple du plasmide F, est tout à fait remarquable en raison de l'intervention d'une seule protéine, nommée Tra, pour près de 30 protéines pour le transfert du plasmide F. Chez ces bactéries, l'initiation commence par le clivage de l'origine de transfert au niveau de l'ADN transféré, suivi du transfert sous forme simple brin et enfin de la ligation dans le génome de la cellule fille.

La conjugaison bactérienne chez *Streptomyces* peut être divisée en deux étapes :

- Le transfert intermycélien contrôlé par le gène *tra.*
- La diffusion de l'élément génétique à travers le mycélium receveur grâce aux gènes *spdA*-

D (Kataoka *et al.* 1991).

La protéine Tra présente des homologies avec les protéines de la famille SpoIIIE/FtsK, qui permettent la translocation d'ADN double brin à travers les septa bactériens (Errington *et al.* 2001). Cette protéine reconnaît le locus *clt* (*cis*-acting locus), nécessaire au transfert des plasmides mais pas des marqueurs chromosomiques (Pettis et Cohen 1994). Dans le plasmide pIJ101, le site *clt* mesure 54 pb et est situé dans l'extrémité 3' du gène *korB* et s'étend dans la région intergénique *korA-korB* (Pettis et Cohen 1994). Les produits de ces deux gènes sont des régulateurs transcriptionnels, KorA par exemple régulant l'expression du promoteur du gène *tra* (Stein et Cohen 1990).

Pour le plasmide pJV1 de *Streptomyces phaechromogenes*, le site *clt* est constitué de répétitions directes de 9 pb (Franco *et al.* 2003). La délétion de ces répétitions directes provoque une forte chute du taux de transfert plasmidique. En revanche, les acteurs permettant la fusion des cellules n'ont pas été identifiés (Hopwood 2006). Contrairement aux autres sites d'initiation de la conjugaison, le site *clt* n'est pas clivé. Un modèle du transfert conjugatif des plasmides implique le pompage intercellulaire des plasmides circulaires sous forme CCC par la protéine Tra localisée au niveau de la membrane (Grohmann *et al.* 2003). Enfin la protéine possède une activité de liaison non spécifique à l'ADN pouvant expliquer la mobilisation de marqueurs chromosomiques.

Lorsque le plasmide conjugatif est transféré dans le mycélium receveur, les produits des gènes *spd* interviennent pour permettre sa dissémination dans les autres cellules du mycélium. L'analyse des génomes totalement séquencés de *S. coelicolor* et *S. avermitilis* révèle la présence de plusieurs copies de gènes *spdA* au sein d'éléments génétiques intégrés au chromosome (Bentley *et al.* 2002; Ikeda *et al.* 2003).

A-2 MODÈLE "END FIRST"

Un transfert conjugatif du chromosome par l'une ou l'autre (ou les deux) extrémités a été proposé par C. W. Chen dans son modèle « End First » (Chen 1996) (Fig. 65).

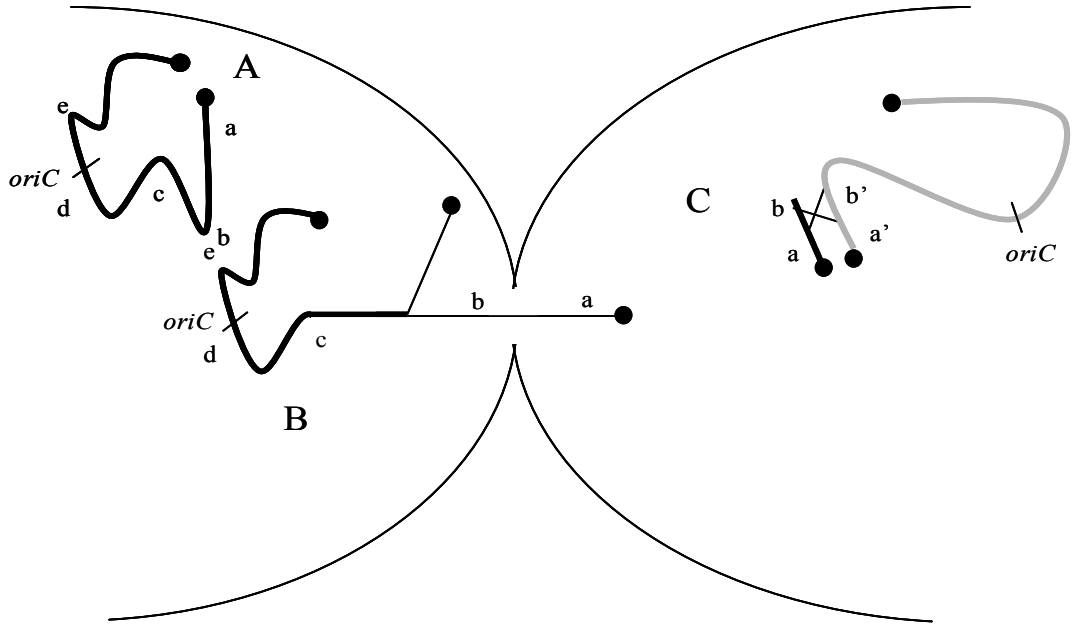


Figure 65 : Modèle de conjugaison « End First » chez les *Streptomyces*
A. La conjugaison est initiée à partir des télomères du chromosome de *Streptomyces*.
B. Les marqueurs chromosomiques des régions terminales (a-b) seront transférés en premier.
C. Par recombinaison, le fragment transféré sera incorporé au niveau des régions terminales du chromosome de la cellule réceptrice.

Dans ce modèle, les télomères servent d´origine de transfert par la fixation des protéines terminales, servant de protéines initiatrices du transfert conjugatif. En revanche, ce modèle, basé sur les mécanismes de conjugaison connus, propose le transfert de l'ADN sous forme simple brin. En revanche, des expériences, notamment avec le plasmide conjugatif pSAM2 de *S. ambofaciens*, ont montré que ce dernier était transféré à l'état double brin (Possoz *et al.* 2001). Un argument en faveur de ce modèle est la présence du gène *ttrA* (pour « terminal transfer ») présentant une localisation terminale. Chez *S. avermitilis*, ce gène est localisé à 5 kb du télomère gauche. En revanche, chez *S. coelicolor* et *S. lividans*, ce gène est placé à 1,5 kb des télomères (Bentley *et al.* 2002; Ikeda *et al.* 2003). Chez la souche de *S. ambofaciens* ATCC23877, le gène *ttrA* (SAMT0002) est situé à 2,6 kb des télomères. Pour la souche de *S. ambofaciens* DSM40697, le gène *ttrA* (DSMT0008) est tronqué (environ 200 codons sur 886) et donc coderait un produit potentiellement non fonctionnel.

Le produit de ce gène présente une activité hélicase et est responsable du transfert conjugatif du plasmide SLP2 et du chromosome de *S. lividans* (Huang *et al.* 2003). Le rôle proposé pour l'hélicase codée par *ttrA* est l'initiation du transfert conjugatif par déroulement de l'ADN.

A-3 GRADIENT DE TRANSFERT DES REGIONS TERMINALES

La variabilité des régions terminales serait en lien avec le transfert préférentiel à partir de ces extrémités, qui seraient donc les régions les plus fréquemment transférées, suivant le modèle « End First ». Chez *E. coli*, les souches Hfr voient leur chromosome mobilisé à partir de l'origine de transfert *oriT*, induisant un gradient de transfert des marqueurs chromosomiques suivant leur distance par rapport à l'origine de transfert (Reimmann et Haas 1993).

Chez *M. smegmatis*, la région *bom* est un homologue fonctionnel des origines de transfert. Le transfert conjugatif se fait de façon classique avec clivage des sites *bom* et transfert dans la cellule receveuse (Wang *et al.* 2003).

Par analogie, un gradient de transfert des régions terminales, à partir des télomères, chez les *Streptomyces*, provoquerait l'insertion préférentielle des marqueurs chromosomiques terminaux. Par recombinaison homologue ou illégitime, ces marqueurs transférés seraient intégrés préférentiellement au niveau des régions terminales. Ainsi, une corrélation existerait entre la mobilisation plus favorable des extrémités chromosomiques et la localisation des gènes fréquemment transférés.

B- Mécanismes de réparation des cassures double brin

Les cassures double brin peuvent être prises en charge par la recombinaison homologue ou illégitime. Chez *Sa. cerevisiae,* suivant la localisation de la cassure double brin, cette dernière ne sera pas prise en charge par les mêmes systèmes de réparation. Ainsi, le mécanisme de NHEJ (Non-Homologous End Joining) intervient préférentiellement au niveau de la région centrale. En revanche, aux extrémités, quatre événements différents sont observés (Ricchetti *et al.* 2003):

- le mécanisme de BIR (Break Induced Replication) qui va conduire à la duplication de l'extrémité d'un autre chromosome
- le remplacement de l'extrémité par l'extrémité d'un autre réplicon
- la conversion génique
- la cicatrisation par addition d'un télomère.

Chez les bactéries, les cassures double brin sont prises en charge majoritairement par la recombinaison homologue (Kowalczykowski 2000). En effet, le système de réparation par NHEJ (Non-Homologous End Joining) est absent de 80% des génomes bactériens étudiés, alors que les systèmes de recombinaison homologue RecBCD/AddAB sont retrouvés dans de nombreux génomes. En absence du système RecBCD, le système AddAB permet la réparation des cassures double brin.

B-1 RECOMBINAISON HOMOLOGUE CHEZ LES *STREPTOMYCES*

Chez les *Streptomyces*, la recombinaison homologue est très efficace (Wohlleben *et al.* 1994). La réduction de l'activité de RecA conduit à une forte augmentation du taux d'instabilité (jusqu'à un facteur 70) (Volff et Altenbuchner 1997). Un mutant RecA- présente un retard de croissance, des spores vides et une importante baisse de l'efficacité de recombinaison lors de la conjugaison (Huang et Chen 2006).

Les mécanismes de réparation des cassures double brin ne sont pas connus chez les *Streptomyces*. Les gènes homologues des systèmes RecFOR, de réparation des cassures simple brin ou RuvABC sont présents chez *S. coelicolor* et *S. avermitilis* (Rocha *et al.* 2005). Le système RecFOR fait partie du complexe présynaptique de résolution des cassures simple brin en se fixant au niveau de l'extrémité simple brin, en déplaçant les protéines SSB pour permettre la fixation de RecA (Kowalczykowski 2000). Le système RuvABC intervient dans la migration et la résolution des croix de Holliday.

B-2 SYSTEME NHEJ CHEZ LES *STREPTOMYCES*

Le système NHEJ serait présent chez *Streptomyces* (SCO5308/09 et SAV2945/46 respectivement pour *S. coelicolor* et *S. avermitilis*) et chez *Mycobacterium* (Aravind et Koonin 2001; Weller *et al.* 2002). Chez *Mycobacterium*, les protéines Ku forment un dimère qui reconnaît les extrémités des cassures pour recruter la ligase IV et permettre la réparation. Ces deux protéines sont actives et suffisantes pour la réparation de ce type de cassure chez *M. tuberculosis* (Pitcher *et*

al. 2005). Ce système est identifié dans environ 20% des espèces bactériennes étudiées (le gène Ku est présent dans 24 génomes des 117 analysés).

A partir de l'analyse des extrémités des BAC (BES pour BAC End Sequence, cf. Résultats II-A1) chez *S. ambofaciens*, le génome présente une situation analogue à celle de *S. coelicolor*. La comparaison des protéines putatives montre que la séquence protéique Ku est similaire aux protéines Ku des autres bactéries. En revanche, la ligase IV diffère entre les espèces, où cette protéine ne possède qu'un seul domaine catalytique (activité polymérase) au lieu de trois domaines rencontrés habituellement chez les autres ligases IV.

B-3 INTERVENTION DES MECANISMES DE REPARATION DANS LA VARIABILITE DES REGIONS TERMINALES

Par analogie avec ce qui est observé chez *Sa. cerevisiae*, il peut être envisagé que le système de réparation diverge suivant la localisation de la cassure double brin le long du chromosome. Ainsi, le système de recombinaison homologue interviendrait pour des cassures produites dans la région centrale du chromosome. Ces événements de recombinaison ne rompraient pas la synténie, mais aboutirait à la variation allélique. En revanche, le système NHEJ interviendrait préférentiellement pour les cassures au niveau des régions terminales, provoquant une variabilité de ces régions. La variabilité terminale, c'est à dire l'absence d'identités fortes entre les séquences terminales, serait elle-même un frein à l'implication de la recombinaison homologue dans les régions spécifiques. La présence de gènes essentiels serait contre-sélectionnée au cours de l'évolution de part l'instabilité forte agissant dans ces régions.

Une autre hypothèse est que les régions terminales sont plus fréquemment la cible de cassures double brin. Leur réparation pourrait conduire, suivant le mécanisme de réparation, à modifier l'information contenue dans ces régions. En revanche, si la probabilité de provoquer une cassure double brin est équivalente le long du chromosome, les conséquences sur les régions terminales pourraient être observées. En revanche, celles touchant la région centrale seraient vraisemblablement contre-sélectionnées.

IV- Régions terminales et rôle dans l'adaptation

La génomique comparée de *S. ambofaciens* avec *S. coelicolor*, *S. avermitilis* et *S. scabies*, et entre les deux souches de *S. ambofaciens* a permis de montrer que les régions terminales sont hautement variables au niveau interspécifique comme au niveau intraspécifique.

Ces régions sont riches en gènes d'adaptation comme par exemple ceux codant les voies de biosynthèse de métabolites secondaires ou des enzymes extracellulaires responsables de la dégradation de polymères de l'environnement (par exemple la cellulose ou la chitine).

A- Antibiotiques et rôle dans l'adaptation

Les fonctions codées dans les régions terminales jouent un rôle dans l'adaptation face aux

changements de l'environnement de ces bactéries. En effet, le biotope des *Streptomyces* est le sol. Ce milieu présente une grande diversité biotique et abiotique (paramètres physico-chimiques). La multitude de fonctions prédites dans ces régions sont autant de réponses pour permettre la survie et l'adaptation de la bactérie dans son milieu. La diversité des métabolites secondaires produits (antibiotique, antifongique…) de même que les résistances aux métaux lourds ou aux drogues (ex : Système de résistance à l'arsenate dans les TIR de la souche de *S. ambofaciens* DSM40697) sont des exemples des fonctions permettant une meilleure adaptation au biotope.

Chez les *Streptomyces*, un lien entre cycle de développement et synthèse des antibiotiques peut être observé. Cette synthèse se déroule au moment de la formation des hyphes aériens, où une partie des substances nutritives pour ce développement provient de la lyse du mycélium végétatif (Mendez *et al.* 1985). La production et la sécrétion d'antibiotiques élimineraient les autres microorganismes, laissant les nutriments obtenus aux seuls *Streptomyces* (Chater 2006).

Les antibiotiques pourraient jouer d'autres rôles que le seul rôle antagoniste (Yim *et al.* 2006). En effet, pour pouvoir inhiber la croissance ou tuer un autre microorganisme, ces molécules doivent être synthétisées en quantités suffisantes. Or, ces conditions sont rarement rencontrées dans les conditions environnementales, la fonction antibiotique ne pourrait être réalisée (Davies 2006). En revanche, ils pourraient jouer un rôle de signal. Chez *Pseudomonas*, par exemple, la pyocyanine agit comme signal intercellulaire pour les cellules en phase stationnaire de croissance, permettant l'expression de gènes contrôlés par le quorum-sensing (Price-Whelan *et al.* 2006).

Les molécules antibiotiques peuvent également induire des changements globaux au niveau de l'expression des gènes pour des concentrations 100 fois inférieures aux concentrations inhibitrices. Par exemple, chez *E. coli* et *Sl. typhimurium,* jusqu'à 5% des promoteurs seraient affectés par l'érythromycine et la rifampycine à de telles concentrations (Goh *et al.* 2002). La virulence ou encore le métabolisme carboné sont également modulés (Goh *et al.* 2002; Linares *et al.* 2006; Yim *et al.* 2006). Ainsi les antibiotiques pourraient jouer un rôle hormétique, à savoir un effet stimulant à faibles concentrations et inhibiteur à fortes concentrations.

B- Interactions avec d'autres microorganismes

Les *Streptomyces* interagissent également avec d'autres organismes de leur biotope. Certains microorganismes pourraient être considérés comme des « proies » (source de nutriments) pour les *Streptomyces* (Chater 2006). Par exemple, les *Streptomyces* produisent des chitinases pouvant dégrader la paroi cellulaire des champignons qui serviraient alors de source nutritive. Shi et Zusman (1993) ont décrit un mécanisme chez *Myxococcus xanthus* où, en conditions de carence nutritionnelle, cette dernière sécrète des molécules qui attirent les cellules d'*E. coli*, proie pour les myxobactéries (Shi et Zusman 1993).

La majorité des plantes forme des associations symbiotiques, nommées mycorhizes, avec des champignons filamenteux. Dans ces structures sont retrouvées des bactéries comme *Burkholderia* ou *Rhodococcus* et qui sont appelées MHB (pour « Mycorrhiza Helper Bacteria »). Deux espèces de *Streptomyces*, *Streptomyces* sp. nov. 505 (AcH 505) et *Streptomyces annulatus* 1003 (AcH 1003) ont été caractérisées suivant leur effet sur la croissance de champignons et la formation des mycorhizes (Maier *et al.* 2004). L'espèce AcH 505 stimule la croissance du

mycélium et du mycorhize de *Amanite muscaria* mais va inhiber la croissance de champignons pathogènes. L'ajout de *Streptomyces* sp. nov. 505 en co-culture avec *Am. muscaria* va modifier le niveau de la transcription de 113 gènes étudiés (Schrey *et al.* 2005). La grande majorité est liée à la voie de transduction de signaux (protéines kinases ou facteurs de transcription), du métabolisme (carbone, sulfate et de l'azote), la croissance cellulaire et la structure cellulaire. En revanche, le/les composé(s), responsable(s) de cette stimulation de la transcription, n'a(ont) pas été identifié(s). Ainsi, certains métabolites synthétisés par les *Streptomyces* permettraient le développement des champignons symbiotiques et inhiberait la croissance des champignons pathogènes au sein d'un mycorhize, participant à l'équilibre des populations de microorganismes des sols.

V- Mode d'acquisition des nouvelles fonctions dans les TIR

A- Interactions entre réplicons linéaires

De nombreux exemples d'interactions entre chromosome et plasmide ont été décrits. Ainsi, le chromosome de *S. coelicolor* et le plasmide SCP1 recombinent pour donner deux réplicons chimériques de 7,2 Mb et 1,8 Mb (Yamasaki et Kinashi 2004). Ces deux réplicons sont la conséquence d'un événement de recombinaison illégitime entre les séquences codantes SCP1.136 (hélicase putative) et SCO6388 (fonction inconnue). Cette recombinaison a eu pour effet de transférer 1,6 Mb terminaux du bras chromosomique droit sur le plasmide. Ces deux réplicons sont considérés comme des chromosomes car chacun d'eux porte des fonctions essentielles à la survie de la bactérie.

L'analyse des extrémités du plasmide linéaire SLP2 de *S. lividans* suggère également un échange d'extrémités entre le bras droit de ce plasmide et le chromosome ancestral de *S. lividans* (Huang *et al.* 2003). En effet, les 15,4 kb terminaux de l'extrémité droite sont identiques à ceux du chromosome (Fig. 66). Cette région d'homologie s'étend depuis les télomères jusqu'au transposon *Tn*4811. Dans cette région d'homologie, le gène *ttrA*, responsable du transfert conjugatif, est présent.

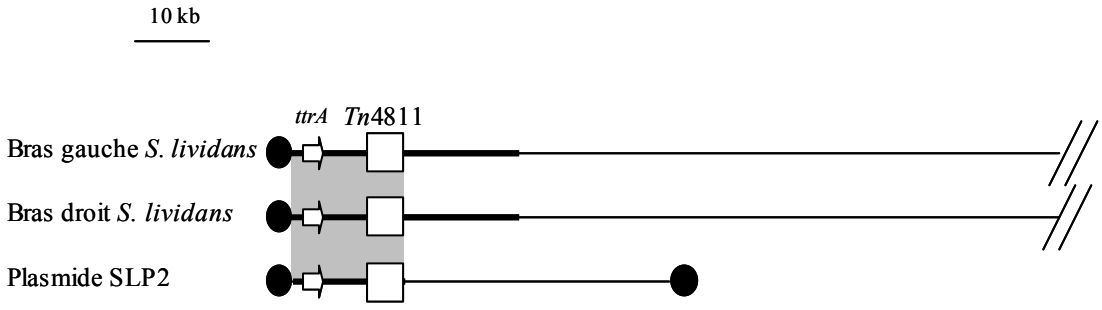


Figure 66 : Comparaison des extrémités du chromosome de *S. lividans* et du plasmide SLP2
La région terminale de 15,4 kb est homologue au bras droit du plasmide SLP2 (région grisée).
Les répétitions terminales inversées sont représentées en trait épais. Les ronds noirs correspondent aux protéines terminales. Seuls le transposon *Tn*4811 et l'ORF *ttrA*, responsable du transfert conjugatif du plasmide SLP2, sont représentés.

La présence de Tn*4811*, au niveau de la jonction des régions potentiellement échangées,

suggère un événement de transposition réplicative ou encore de recombinaison homologue entre copies plasmidique et chromosomique de l'élément. Cet événement pourrait être à l'origine de la structure du bras droit de SLP2.

B- Remplacement des extrémités entre réplicons linéaires

A coté de ces régions spécifiques hautement variables, les séquences télomériques sont très conservées entre les différents *Streptomyces* (Huang *et al.* 1998). Dans l'hypothèse d'un remplacement d'extrémités entre réplicons linéaires, un seul événement de recombinaison est nécessaire pour permettre ce remplacement. Ce remplacement pourrait ne pas être contre-sélectionné car il maintiendrait la structure linéaire fonctionnelle. Ainsi, la linéarité des réplicons apporterait un avantage en favorisant le flux de gènes dans ces régions. A la frontière des régions conservées des TIR dans les deux souches de *S. ambofaciens*, une *IS* tronquée est présente. Cette *IS* pourrait avoir servi de substrat à un événement de recombinaison homologue.

La comparaison des TIR des deux souches de *S. ambofaciens* suggère que l'une de ces souches a acquis des extrémités divergentes par interaction avec un réplicon linéaire. La seconde a soit conservé la séquence ancestrale (aux événements ponctuels près) soit a acquis également de nouvelles extrémités. Au vu des séquences associées à des plasmides linéaires, la souche DSM40697 aurait acquis des extrémités par échanges avec des plasmides linéaires.

VI- Maintien des séquences acquises sur le long terme

Le maintien à long terme d'une séquence acquise par transfert horizontal dépend de sa contribution à la valeur adaptative globale, le « fitness ». Lorsque la fonction acquise contribue positivement au fitness, la séquence nucléotidique subit le phénomène d'amélioration, c'est-à-dire une modification, à l'échelle évolutive, de la composition nucléotidique pour converger vers les paramètres génomiques de l'hôte.

Chez les *Streptomyces*, le biais de distribution et d'orientation des gènes par exemple est de faible intensité avec respectivement 55,5% et 56,2% des CDS de *S. coelicolor* et *S. avermitilis* situées sur le brin direct de réplication. Le biais d'orientation est variable suivant les régions étudiées. En effet, la région « core », qui contient l'ensemble des gènes essentiels, présente un biais de 58,7% et 58,4% respectivement chez *S. coelicolor* et *S. avermitilis* contre 51,4% et 52,9% pour les régions de contingence. La région centrale contient, entre autres, les gènes dits « essentiels » au développement végétatif, pouvant expliquer la divergence dans l'orientation des gènes entre ces deux régions. Une autre hypothèse pour expliquer ce biais serait l'acquisition préférentielle des gènes par transfert horizontal dans ces régions de contingence.

Une des caractéristiques génomiques des *Streptomyces* est leur pourcentage élevé en bases G-C (72,1% chez *S. coelicolor*). Chez *S. ambofaciens*, les gènes sans homologues chez *S. coelicolor* et *S. avermitilis*, possèdent les caractéristiques des génomes des *Streptomyces*. Ainsi, le pourcentage en G-C minimum identifié pour une CDS de *S. ambofaciens* est de 60,4%. De plus, toutes les CDS prédites présentent un biais marqué en bases G-C en troisième position de codons, pour une moyenne de 88,2% pour les CDS spécifiques (pour 92,1% chez *S. coelicolor*). Ceci

indique que les caractéristiques des génomes créent une barrière pour le maintien des séquences nouvellement acquises. Seules les CDS avec une composition nucléotidique proche de celles des *Streptomyces* sont observées et donc ont été maintenues.

Les donneurs potentiels d'informations génétiques doivent posséder un pourcentage en G-C élevé. Parmi les génomes bactériens séquencés, les *Streptomyces* et plus largement les actinomycètes présentent des caractéristiques compositionnelles proches (par exemple *Frankia* sp. Cci3 70%, *Rhodococcus* 67%). Ces espèces occupent le même biotope, à savoir le sol, et les échanges pourraient être favorisés.

Rhodococcus est un exemple de partenaire potentiel (McLeod *et al.* 2006). Le transfert conjugatif de certains plasmides de *Rhodococcus* a été décrit (Dabrock *et al.* 1994). Ces plasmides sont porteurs de gènes d'adaptation comme des gènes de résistance aux métaux lourds, fonctions retrouvées au niveau des extrémités du chromosome chez *S. ambofaciens*. De plus, le génome de *Rhodococcus* sp. RHA1 est constitué de quatre réplicons linéaires (McLeod *et al.* 2006). De façon analogue aux *Streptomyces*, ces réplicons présentent une structure invertronique. Ainsi, le chromosome possède des répétitions inversées de 12 kb pour 500 pb au plasmide pRHL1. Enfin, les répétitions inversées télomériques sont similaires à celles des *Streptomyces*.

VII- Etude dynamique des régions terminales des Streptomyces.

A- Structure chromosomique potentielle des mutants de S. ambofaciens *DSM40697*

Des souches de *S. ambofaciens* DSM40697,à partir des lignées de S. ambofaciens DSM40697-1,-2 et –3, contenant 16 sites *parS* situés dans le même locus dans le bras chromosomique gauche ont été obtenues. Pour chacune de ces souches, le témoin avec la cassette Apra seule a également été produit. Ainsi six souches ont été analysées pour déterminer la structure chromosomique de leur descendance.

L'analyse a porté sur 12 clones pour chacune des six souches, par amplification de différentes régions du chromosome (Fig. 60).

A partir de ces résultats, huit « profils » d'amplification ont pu être distingués (Fig. 67).

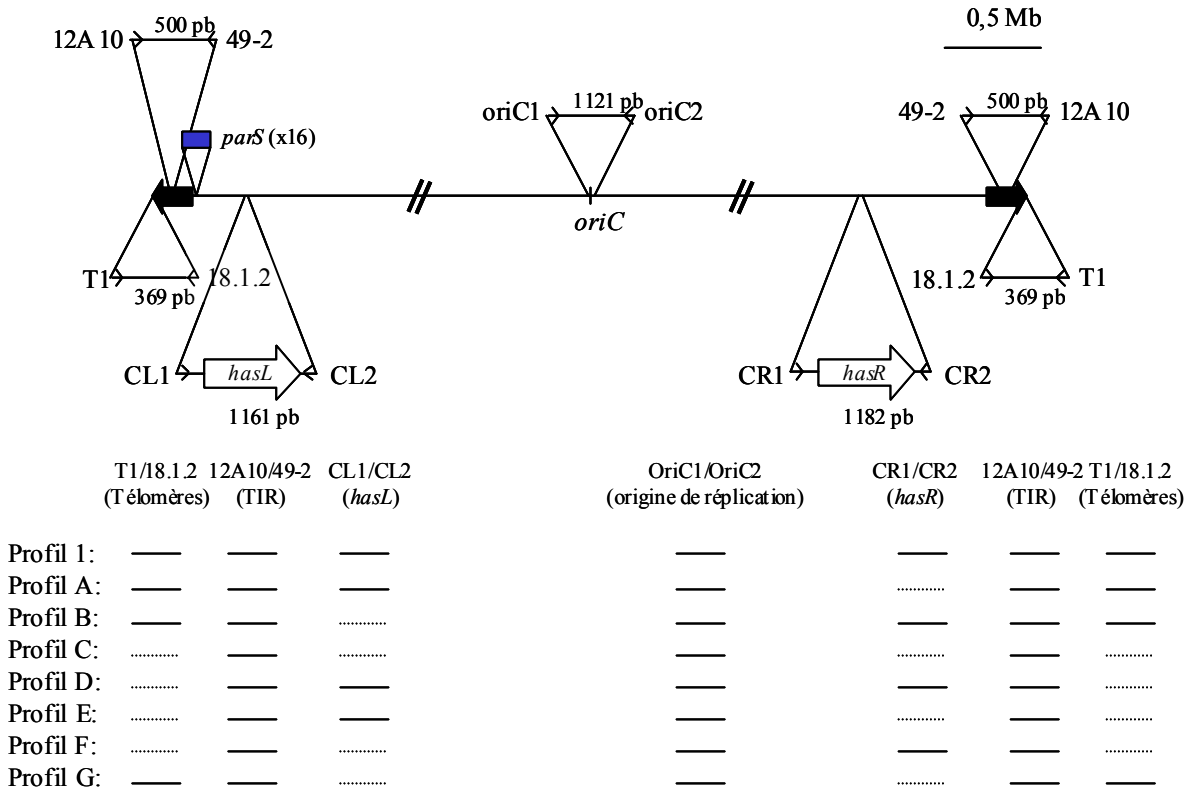


Figure 67 : Profils obtenus après analyse par réactions d'amplification.
Les traits pleins indiquent l'obtention de produits de PCR et les pointillés une absence de produits d'amplification. La localisation des cibles pour les différents couples d'amorces est indiquée entre parenthèses.

Pour les clones du profil 1, un produit d'amplification a été obtenu pour chaque couple d'amorces. Ces clones présenteraient donc un chromosome linéaire de type sauvage.

La présence en deux copies à chacune des extrémités des cibles des couples d'amorces T1/18.1.2 et 12A10/49-2 conduira à l'amplication de cette cible même dans le cas de la perte d'une seule extrémité. Ainsi à partir des résultats obtenus, une structure chromosomique potentielle peut être décrite pour ces différents « profils ».

Pour les clones présentant un « profil A», le gène *hasR* est absent. En revanche, tous les loci étudiés sont amplifiés. Ce profil pourrait s'expliquer par la perte d'une partie ou de la totalité du bras chromosomique droit (contenant le gène *hasR*) sans circularisation du chromosome, car un produit d'amplification correspondant aux télomères est observé avec le couple T1/18.1.2. Le « profil B » correspond au réciproque du « profil A » avec la perte du gène *hasL*. Pour ces deux profils, l'hypothèse expliquant la structure chromosomique est l'acquisition d'une nouvelle extrémité et délétion des gènes *has* (Fig. 68). De même ces deux profils peuvent correspondre à une structure chromosomique fusionnée entre les bras chromosomiques droits de deux chromatides sœurs pour le « profil A » ou entre deux chromosomiques gauches pour les clones du « profil B » (Fig. 69) à l'instar des souches obtenues spontanément NSA27 et NSA65 (Wenner *et al.* 2003).

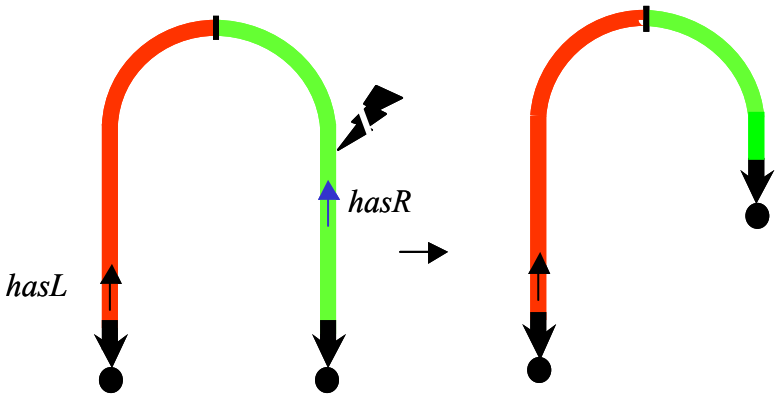


Figure 68: Structure hypothètique des clones du profil A.
Pour les clones du « profil B », la même hypothèse peut être formulée mais avec la perte du gène *hasL*.

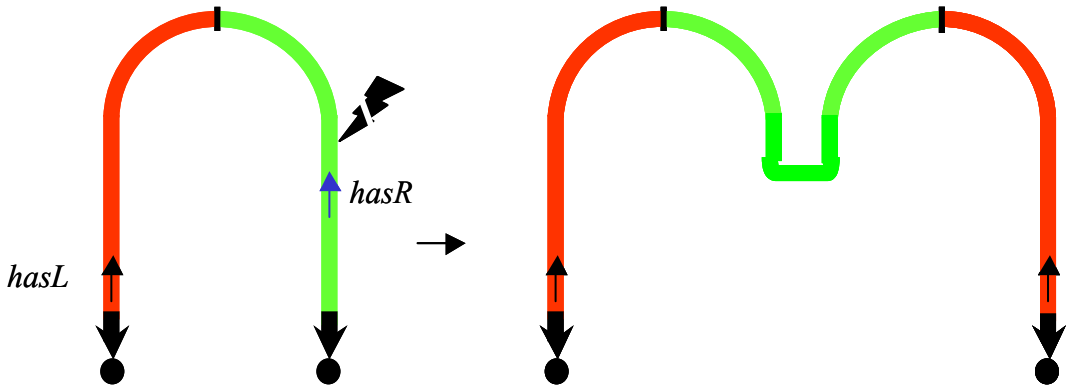


Figure 69 : Hypothèse de la fusion chromosomique pour les clones du profil A.

Le « profil C » montre la délétion des télomères ainsi que des gènes *has* mais conservation d'une partie des TIR. Ceci pourrait s'expliquer par la translocation de la cible du couple 12A10/49-2 et circularisation du chromosome.

Le « profil D » présente une absence de produit d'amplification des séquences télomériques. Ainsi, la structure hypothétique pour ces clones serait la circularisation du chromosome par recombinaison entre les deux loci situés dans les TIR entre les télomères et le locus cible du couple d'amorces 12A10/49.2.

Le « profil E » indique une absence de signaux pour le gène *hasR* et des télomères. Une structure chromosomique potentielle pour ce profil est la circularisation du chromosome entre un locus situé entre le télomères et la cible du couple 12A10/49.2, comme pour le « profil D », et un locus situé au-delà du gène *hasR* conduisant à la perte du bras droit, où sont situés les 16 sites *parS* (Fig. 70). Comme pour les « profils A et B », le « profil F » est la réciproque du « profil E » pour le bras gauche. Ainsi, une structure découlant de la recombinaison entre un locus situé dans les TIR et un locus au-delà du gène *hasL* conduit à la délétion de ce gène et à une structure chromosomique circulaire.

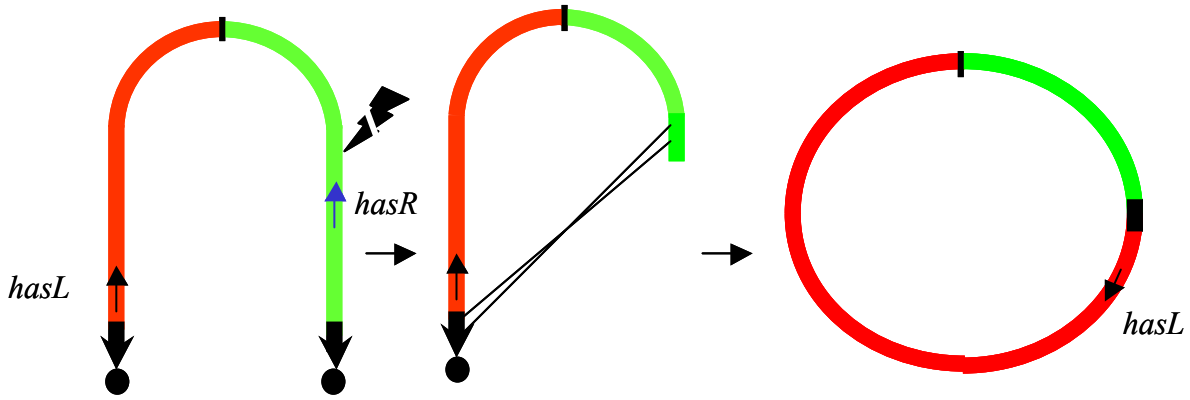


Figure 70: Structure hypothètique des clones du profil E.
Pour les clones du « profil F », le même cheminement peut être observé mais avec la perte du gène *hasL*

Les produits d'amplification des gènes *hasL* et *hasR* sont absents dans le « profil G». En revanche, les produits d'amplification des régions terminales sont observés. Ces résultats indiquent la délétion des deux gènes *has* sans circularisation chromosomique.

Ces différentes structures restent hypothétiques et seront donc à confirmer grâce aux techniques d'analyse génomique telles que l'électrophorèse en champ pulsé.

B- Influence de la présence des sites parS sur les différents profils obtenus.

L'influence de la présence d'un second site de partition potentiel sur la stabilité chromosomique a aussi pu être abordée.

Les fréquences des différents types de profils dans l'échantillonage de souches isogéniques présentant la cassette Apra seule ou Apra-sites parS ne sont pas significativement différents pour les profils 1, B, D et E.

En revanche, au risque de 5%, la fréquence du profil A et C sont influencés par le type de cassette. Dans le cas A, la fréquence apparaît plus forte grâce à la présence de sites *parS*. Dans le cas du profil C, la fréquence de réarrangement est significativement plus faible avec un second site de partition potentiel.

Alors que l'influence positive de la présence d'un second site de partition sur l'instabilité génomique est cohérente avec les connaissances sur les mécanismes de partition chromosomique, et tendrait à valider notre hypothése de travail. En revanche, les sitres *parS* sont situés sur le bras gauche et non sur le bras chromosomique droit. Ainsi, l'influence du second site de partition devra être vérifiée. Le résultat concernant le profil C semblerait indiquer une stabilisation de la structure chromosomique par la présence de séquences *parS* intégrées dans les régions terminales. De plus, au seuil de risque 1%, les différences observées ne sont plus significatives. Il convient donc

visiblement de renforcer les effectifs étudiés afin de confirmer ces tendances.

C- Lien entre partition chromosomique et réarrangements des extrémités

Chez les eucaryotes, la perte d'un télomère ou une cassure double brin peut conduire le chromosome dans un cycle d'instabilité nommé Cassure-Fusion-Pont (CFP). Ce cycle est caractérisé par la fusion de deux chromatides formant un chromosome dicentrique (Fig. 6). Les régions terminales peuvent alors subir des réarrangements chromosomiques par délétions des régions terminales ou encore des translocations non réciproques.

Les *Streptomyces* possèdent également un chromosome linéaire. Des mutants spontanés présentant un chromosome circulaire ont été obtenus. La perte d'un télomère et une structure fusionnée ont été observées chez un mutant spontané de *S. ambofaciens* (Wenner *et al.* 2003). Ce mutant voit sa structure parentale faiblement héritée dans la descendance. Par analogie avec ce qui est observé chez les chromosomes dicentriques eucaryotes, la présence d'un second centre de partition chromosomique serait à l'origine de cette instabilité qui pourrait conduire à de grands réarrangements des régions terminales.

Les systèmes de partition chromosomique interviennent principalement au moment de la sporulation chez les *Streptomyces* (Kim *et al.* 2000; Jakimowicz *et al.* 2006), le mycélium végétatif étant constitué de syncitia contenant de nombreuses copies du génome.

Après le développement du mycélium aérien, les hyphes se différencient en spores. Chaque spore contient une seule copie du génome. Le mécanisme de partition va permettre la distribution d'un génome par spore. Dans le cas où les deux centres de partition tractent le chromosome vers la même cellule fille, ce dernier sera incorporé sans réarrangement. En revanche, si la traction se fait vers les pôles opposés, des contraintes physiques vont s'exercer entre les deux centres de partition conduisant à des cassures chromosomiques. Dans la descendance, un grand nombre de spores non viables sera présent en raison de la perte de fonctions essentielles à la survie de la cellule. Les spores ayant subi des cassures au niveau des régions terminales pourraient entrer dans un cycle analogue à celui observé chez les eucaryotes. L'entrée dans ce cycle d'instabilité conduira, sur la base de ce qui est observé chez les eucaryotes, à la formation de grands réarrangements chromosomiques dans les régions terminales, à savoir des translocations, des circularisations ou des duplications de grandes régions chromosomiques.

PERSPECTIVES

I- A courts termes

Cette étude a permis l'obtention d'une souche de *S. ambofaciens* présentant un second centre de partition, par clonage de 16 sites *parS* seuls. La présence de nombreux sites *parS*, proches les uns des autres, peut provoquer un encombrement stérique lors de la fixation des protéines ParB sur leurs cibles. Cet encombrement pourrait conduire à une efficacité moindre de l'action de ce second centre de partition.

Outre la validation fonctionnelle du nouveau site de partition inséré dans le chromosome par des méthodes d'immunoprécipitation (Jakimowicz *et al.* 2002), nous envisageons la construction de souches où l'insertion sera distribuée le long des TIR et des régions terminales.

Cependant, le caractère fonctionnel de ce site de partition pourrait être remis en cause. En effet, les séquences *parS* multiples organisées en tandem pourraient ne pas permettre la fixation coopérative du complexe ParB-ParA. Ainsi, nous envisageons l'insertion des séquences *parS* dans leur disposition « spontanée » dans les régions terminales. Pour cela, des BAC recombinants recouvrant les régions riches en séquences *parS* seront insérés dans les régions terminales grâce au module d'insertion du phage ΦC31 (*attB-int*) dérivé du vecteur pSET152 (Fig. 71).

Ainsi, une souche délétée pour le site *attB* (localisée dans l'homologue du gène SCO3798) est en cours de construction, et par la même approche de PCR Targeting des sites *attB* sont réintroduits en des sites intergéniques dans les régions terminales.

Les BAC choisis comprennent respectivement 10 et 12 sites *parS*, l'un contenant l'origine de réplication *oriC*, l'autre en étant dépourvu. La partie vecteur des BAC recombinants sera modifiée pour y insérer les modules d'insertion et de mobilisation (*oriT*).

Ces souches seront ensuite analysées pour la stabilité de leur structure chromosomique, au travers des analyses génomiques classiques (PFGE, hybridations moléculaires) mais également grâce à l'utilisation de protéines Par couplées à l'EFGP qui permettra d'étudier la chorégraphie chromosomique *in situ*.

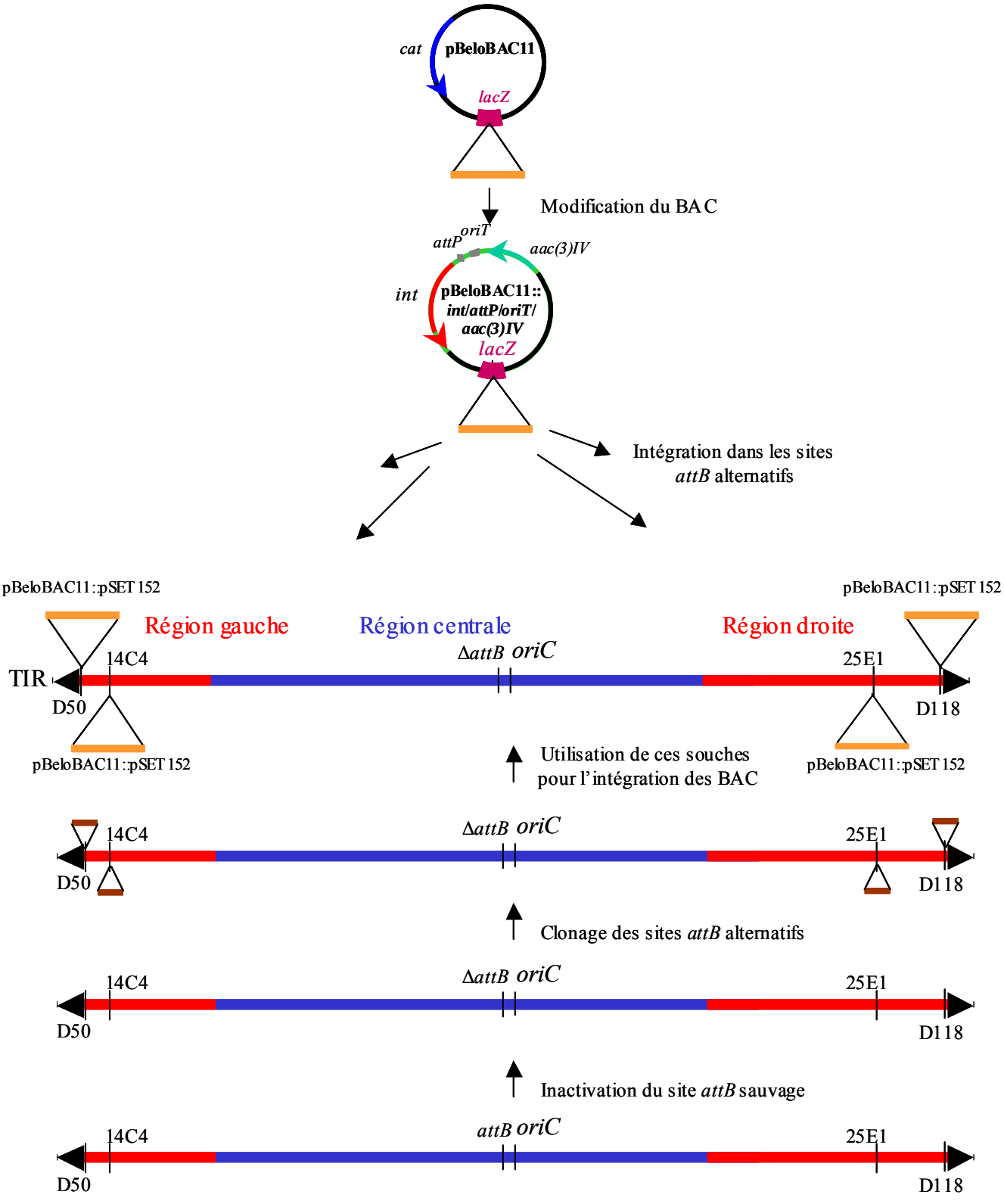


Figure 71 : Stratégie d'insertion des BAC recombinants au niveau de sites *attB* alternatifs.
Le site *attB* sauvage est inactivé et les sites *attB* alternatifs sont insérés dans les régions terminales. Parallèlement, les BAC recombinants sont modifiés pour devenir conjugatifs et intégratifs par l'incorporation du système d'intégration-conjugaison du plasmide pSET152. Enfin, les BAC sont incorporés dans les différents sites *attB*, après conjugaison, pour permettre l'obtention de souches de *S. ambofaciens* contenant un second système de partition chromosomique.

II- A moyens et longs termes

La recherche des mécanismes permettant la mobilisation de marqueurs chromosomiques sera abordée. L'absence de transformation naturelle démontrée et de bactériophages tempérés permettant la transduction généralisée suggère que la conjugaison bactérienne soit à l'origine de ce flux d'information génétique. A l'aide du plasmide conjugatif intégré (ICE) pSAM2, nous étudierons le transfert de fragments d'ADN chromosomique à partir de souches marquées. Nous pouvons aussi tester l'hypothèse de transfert « End First », par l'observation de la mobilisation de marqueurs placés à divers loci sur le chromosome, et/ou au contraire caractériser des séquences *cis* de mobilisation de type *clt* (Franco *et al.* 2003).

Le mode de tranfert serait'il responsable de la perte progressive de synténie ?

Alternativement, l'étude des mécanismes de recombinaison le long du génome sera abordée en caractérisant les modes de réparation des cassures double brin le long du chromosome. L'insertion de site de coupure par une endonucléase (I-*Sce*I) et l'expression contrôlée de cette dernière permettra d'induire la formation de cassures à des sites choisis et d'en étudier la réparation dans différents contextes génétiques : déficients pour la recombinaison homologue et /ou illégitime (Ku/ligIV dépendants).

Enfin, la génomique comparée sera poursuivie tant au niveau interspécifique, plusieurs génomes d'Actinomycètes et notamment des *Streptomyces* étant sur le point d'être accessibles (*S. peucetius, S. venezuelae, S. griseus*).

En ce qui concerne *S. ambofaciens*, la séquence du chromosome de la souche ATCC23877 est en cours de finition, et celle du génome de la souche DSM40697 est envisagée. Le niveau de variabilité intraspécifique permettra de déterminer le niveau de variabilité allélique d'une part, mais également de préciser les événements de réarrangements terminaux récents qui se sont accumulés, ainsi que les mécanismes probables.

REFERENCES BIBLIOGRAPHIQUES

REFERENCES BIBLIOGRAPHIQUES

Abraham, J. M., C. S. Freitag, R. M. Gander, J. R. Clements, V. L. Thomas et B. I. Eisenstein (1986). "Fimbrial phase variation and DNA rearrangements in uropathogenic isolates of *Escherichia coli*." Mol Biol Med **3**(6): 495-508.

Achaz, G., E. Coissac, P. Netter et E. P. Rocha (2003). "Associations between inverted repeats and the structural evolution of bacterial genomes." Genetics **164**(4): 1279-89.

Akman, L., A. Yamashita, H. Watanabe, K. Oshima, T. Shiba, M. Hattori et S. Aksoy (2002). "Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*." Nat Genet **32**(3): 402-7.

Allardet-Servent, A., S. Michaux-Charachon, E. Jumas-Bilak, L. Karayan et M. Ramuz (1993). "Presence of one linear and one circular chromosome in the *Agrobacterium tumefaciens* C58 genome." J Bacteriol **175**(24): 7869-74.

Alm, R. A., L. S. Ling, D. T. Moir, B. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. deJonge, G. Carmel, P. J. Tummino, A. Caruso, M. Uria-Nickelsen, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis et T. J. Trust (1999). "Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*." Nature **397**(6715): 176-80.

Aravind, L. et E. V. Koonin (2001). "Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system." Genome Res **11**(8): 1365-74.

Artandi, S. E., S. Chang, S. L. Lee, S. Alson, G. J. Gottlieb, L. Chin et R. A. DePinho (2000). "Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice." Nature **406**(6796): 641-5.

Aussel, L., F. X. Barre, M. Aroyo, A. Stasiak, A. Z. Stasiak et D. Sherratt (2002). "FtsK Is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases." Cell **108**(2): 195-205.

Bailey, J. A., A. M. Yavor, H. F. Massa, B. J. Trask et E. E. Eichler (2001). "Segmental duplications: organization and impact within the current human genome project assembly." Genome Res **11**(6): 1005-17.

Bankier, A. T. (1993). "M13 phage growth and DNA purification using 96 well microtiter trays." Methods Mol Biol **23**: 41-5.

Bao, K. et S. N. Cohen (2001). "Terminal proteins essential for the replication of linear plasmids and chromosomes in *Streptomyces*." Genes Dev **15**(12): 1518-27.

Bao, K. et S. N. Cohen (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**(6): 774-85.

Bao, K. et S. N. Cohen (2004). "Reverse transcriptase activity innate to DNA polymerase I and DNA topoisomerase I proteins of *Streptomyces* telomere complex." Proc Natl Acad Sci U S A **101**(40): 14361-6.

Barbour, A. G. et C. F. Garon (1987). "Linear plasmids of the bacterium *Borrelia burgdorferi* have covalently closed ends." Science **237**(4813): 409-11.

Bartolome, B., Y. Jubete, E. Martinez et F. de la Cruz (1991). "Construction and properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives." Gene **102**(1): 75-8.

Begg, K. J., S. J. Dewar et W. D. Donachie (1995). "A new *Escherichia coli* cell division gene, ftsK." J Bacteriol **177**(21): 6211-22.

Bentley, S. D., S. Brown, L. D. Murphy, D. E. Harris, M. A. Quail, J. Parkhill, B. G. Barrell, J. R. McCormick, R. I. Santamaria, R. Losick, M. Yamasaki, H. Kinashi, C. W. Chen, G. Chandra, D. Jakimowicz, H. M. Kieser, T. Kieser et K. F. Chater (2004). "SCP1, a 356,023 bp linear plasmid adapted to the ecology and developmental biology of its host, *Streptomyces coelicolor* A3(2)." Mol Microbiol **51**(6): 1615-28.

Bentley, S. D., K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill et D. A. Hopwood (2002). "Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2)." Nature **417**(6885): 141-7.

Bentley, S. D. et J. Parkhill (2004). "Comparative genomic structure of prokaryotes." Annu Rev Genet **38**: 771-92.

Bergthorsson, U., K. L. Adams, B. Thomason et J. D. Palmer (2003). "Widespread horizontal transfer of mitochondrial genes in flowering plants." Nature **424**(6945): 197-201.

Bey, S. J., M. F. Tsou, C. H. Huang, C. C. Yang et C. W. Chen (2000). "The homologous terminal sequence of the *Streptomyces lividans* chromosome and SLP2 plasmid." Microbiology **146 (Pt 4)**: 911-22.

Bhatt, A. et W. R. Jacobs, Jr. (2003). "Conjugal rites of mycobacteria." Nat Genet **34**(1): 3-4.

Bierman, M., R. Logan, K. O'Brien, E. T. Seno, R. N. Rao et B. E. Schoner (1992). "Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp." Gene **116**(1): 43-9.

Birch, A., A. Hausler, M. Vogtli, W. Krek et R. Hutter (1989). "Extremely large chromosomal deletions are intimately involved in genetic instability and genomic rearrangements in *Streptomyces glaucescens*." Mol Gen Genet **217**(2-3): 447-58.

Blackburn, E. H. (2001). "Switching and signaling at the telomere." Cell **106**(6): 661-73.

Borges, F. (2001). Variabilité intraspécifique des extrémités du chromosome linéaire de *Streptomyces ambofaciens*: transfert horizontal et spécifité de séquences terminales. DEA Microbiologie Enzymologie et Nutrition. Nancy, Université Henri Poincaré Nancy1.

Buchanan-Wollaston, V., J. E. Passiatore et F. Canon (1987). "The *mob* and *oriT* mobilization functions of a bacterial plasmid promote its transfer to plants." Nature **328**: 170-175.

Burrus, V., G. Pavlovic, B. Decaris et G. Guedon (2002). "Conjugative transposons: the tip of the iceberg." Mol Microbiol **46**(3): 601-10.

Butland, G., J. M. Peregrin-Alvarez, J. Li, W. Yang, X. Yang, V. Canadien, A. Starostine, D. Richards, B. Beattie, N. Krogan, M. Davey, J. Parkinson, J. Greenblatt et A. Emili (2005). "Interaction network containing conserved and essential protein complexes in *Escherichia coli*." Nature **433**(7025): 531-7.

Cairns, J. et P. L. Foster (1991). "Adaptive reversion of a frameshift mutation in *Escherichia coli*." Genetics **128**(4): 695-701.

Cairns, J., J. Overbaugh et S. Miller (1988). "The origin of mutants." Nature **335**(6186): 142-5.

Campillos, M., C. von Mering, L. J. Jensen et P. Bork (2006). "Identification and analysis of evolutionarily cohesive functional modules in protein networks." Genome Res **16**(3): 374-82.

Campo, N., M. J. Dias, M. L. Daveran-Mingot, P. Ritzenthaler et P. Le Bourgeois (2004). "Chromosomal constraints in Gram-positive bacteria revealed by artificial inversions." Mol Microbiol **51**(2): 511-22.

Chang, P. C. et S. N. Cohen (1994). "Bidirectional replication from an internal origin in a linear streptomyces plasmid." Science **265**(5174): 952-4.

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." Philos Trans R Soc Lond B Biol Sci **361**(1469): 761-8.

Chater, K. F. et D. Hopwood (1984). *Streptomyces* genetics. London, Academic Press.

Chedin, F. et S. C. Kowalczykowski (2002). "A novel family of regulated helicases/nucleases from Gram-positive bacteria: insights into the initiation of DNA recombination." Mol Microbiol **43**(4): 823-34.

Chen, C. W. (1996). "Complications and implications of linear bacterial chromosomes." Trends Genet **12**(5): 192-6.

Chen, C. W., C. H. Huang, H. H. Lee, H. H. Tsai et R. Kirby (2002). "Once the circle has been broken: dynamics and evolution of *Streptomyces* chromosomes." Trends Genet **18**(10): 522-9.

Clewell, D. B., D. D. Jaworski, S. E. Flannagan, L. A. Zitzow et Y. A. Su (1995). "The conjugative transposon *Tn916* of *Enterococcus faecalis:* structural analysis and some key factors involved in movement." Dev Biol Stand **85**: 11-7.

Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead et B. G. Barrell (1998). "Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence." Nature **393**(6685): 537-44.

Corpet, F., J. Gouzy et D. Kahn (1999). "Recent improvements of the ProDom database of protein domain families." Nucleic Acids Res **27**(1): 263-7.

Courvalin, P. (1994). "Transfer of antibiotic resistance genes between gram-positive and gram-negative bacteria." Antimicrob Agents Chemother **38**(7): 1447-51.

Cox, M. M., M. F. Goodman, K. N. Kreuzer, D. J. Sherratt, S. J. Sandler et K. J. Marians (2000). "The importance of repairing stalled replication forks." Nature **404**(6773): 37-41.

Crosa, J. H., M. E. Tolmasky, L. A. Actis et S. Falkow (1994). Plasmids. Washington D.C.

Cui, T., N. Moro-oka, K. Ohsumi, K. Kodama, T. Ohshima, N. Ogasawara, H. Mori, B. Wanner, H. Niki et T. Horiuchi (2007). "*Escherichia coli* with a linear genome." EMBO Rep **8**(2): 181-7.

Dabrock, B., M. Kesseler, B. Averhoff et G. Gottschalk (1994). "Identification and characterization of a transmissible linear plasmid from *Rhodococcus erythropolis* BD2 that encodes isopropylbenzene and trichloroethene catabolism." Appl Environ Microbiol **60**(3): 853-60.

Daniels, D. L., F. Sanger et A. R. Coulson (1983). "Features of bacteriophage lambda: analysis of the complete nucleotide sequence." Cold Spring Harb Symp Quant Biol **47 Pt 2**: 1009-24.

Datsenko, K. A. et B. L. Wanner (2000). "One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products." Proc Natl Acad Sci U S A **97**(12): 6640-5.

Daubin, V. et G. Perriere (2003). "G+C3 structuring along the genome: a common feature in prokaryotes." Mol Biol Evol **20**(4): 471-83.

Davies, J. (2006). "Are antibiotics naturally antibiotics?" J Ind Microbiol Biotechnol **33**(7): 496-9.

Delcher, A. L., D. Harmon, S. Kasif, O. White et S. L. Salzberg (1999). "Improved microbial gene identification with GLIMMER." Nucleic Acids Res **27**(23): 4636-41.

Deng, W., V. Burland, G. Plunkett, 3rd, A. Boutin, G. F. Mayhew, P. Liss, N. T. Perna, D. J. Rose, B. Mau, S. Zhou, D. C. Schwartz, J. D. Fetherston, L. E. Lindler, R. R. Brubaker, G. V. Plano, S. C. Straley, K. A. McDonough, M. L. Nilles, J. S. Matson, F. R. Blattner et R. D. Perry (2002). "Genome sequence of *Yersinia pestis* KIM." J Bacteriol **184**(16): 4601-11.

Dervyn, E., C. Suski, R. Daniel, C. Bruand, J. Chapuis, J. Errington, L. Janniere et S. D. Ehrlich (2001). "Two essential DNA polymerases at the bacterial replication fork."

Science **294**(5547): 1716-9.

Dubnau, D. (1999). "DNA uptake in bacteria." Annu Rev Microbiol **53**: 217-44.

Dybvig, K. (1993). "DNA rearrangements and phenotypic switching in prokaryotes." Mol Microbiol **10**(3): 465-71.

Errington, J., J. Bath et L. J. Wu (2001). "DNA transport in bacteria." Nat Rev Mol Cell Biol **2**(7): 538-45.

Evans, G. A., K. Lewis et B. E. Rothenberg (1989). "High efficiency vectors for cosmid microcloning and genomic analysis." Gene **79**(1): 9-20.

Ewing, B. et P. Green (1998). "Base-calling of automated sequencer traces using phred. II. Error probabilities." Genome Res **8**(3): 186-94.

Ewing, B., L. Hillier, M. C. Wendl et P. Green (1998). "Base-calling of automated sequencer traces using phred. I. Accuracy assessment." Genome Res **8**(3): 175-85.

Fairhead, C., B. Llorente, F. Denis, M. Soler et B. Dujon (1996). "New vectors for combinatorial deletions in yeast chromosomes and for gap-repair cloning using 'split-marker' recombination." Yeast **12**(14): 1439-57.

Fairhead, C., A. Thierry, F. Denis, M. Eck et B. Dujon (1998). "'Mass-murder' of ORFs from three regions of chromosome XI from *Saccharomyces cerevisiae*." Gene **223**(1-2): 33-46.

Fekete, R. A. et L. S. Frost (2000). "Mobilization of chimeric *oriT* plasmids by F and R100-1: role of relaxosome formation in defining plasmid specificity." J Bacteriol **182**(14): 4022-7.

Feng, G., H. C. Tsui et M. E. Winkler (1996). "Depletion of the cellular amounts of the MutS and MutH methyl-directed mismatch repair proteins in stationary-phase *Escherichia coli* K-12 cells." J Bacteriol **178**(8): 2388-96.

Ferdows, M. S. et A. G. Barbour (1989). "Megabase-sized linear DNA in the bacterium *Borrelia burgdorferi*, the Lyme disease agent." Proc Natl Acad Sci U S A **86**(15): 5969-73.

Ferdows, M. S., P. Serwer, G. A. Griess, S. J. Norris et A. G. Barbour (1996). "Conversion of a linear to a circular plasmid in the relapsing fever agent *Borrelia hermsii*." J Bacteriol **178**(3): 793-800.

Finkel, S. E. et R. C. Johnson (1992). "The Fis protein: it's not just for DNA inversion anymore." Mol Microbiol **6**(22): 3257-65.

Fischer, G., B. Decaris et P. Leblond (1997a). "Occurrence of deletions, associated with genetic instability in *Streptomyces ambofaciens*, is independent of the linearity of the chromosomal DNA." J Bacteriol **179**(14): 4553-8.

Fischer, G., A. Kyriacou, B. Decaris et P. Leblond (1997b). "Genetic instability and its possible evolutionary implications on the chromosomal structure of *Streptomyces*." Biochimie **79**(9-10): 555-8.

Fischer, G., T. Wenner, B. Decaris et P. Leblond (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**(24): 14296-301.

Flardh, K. (2003). "Growth polarity and cell division in *Streptomyces*." Curr Opin Microbiol **6**(6): 564-71.

Fox, L., J. Erion, J. Tarnowski, L. Spremulli, N. Brot et H. Weissbach (1980). "*Euglena gracilis* chloroplast EF-Ts. Evidence that it is a nuclear-coded gene product." J Biol Chem **255**(13): 6018-9.

Franco, B., G. Gonzalez-Ceron et L. Servin-Gonzalez (2003). "Direct repeat sequences are essential for function of the cis-acting locus of transfer (clt) of Streptomyces phaeochromogenes plasmid pJV1." Plasmid **50**(3): 242-7.

Frank, A. C., H. Amiri et S. G. Andersson (2002). "Genome deterioration: loss of repeated sequences and accumulation of junk DNA." Genetica **115**(1): 1-12.

Fraser, C. M., J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley, R. D. Fritchman, J. F. Weidman, K. V. Small, M. Sandusky, J. Fuhrmann, D. Nguyen, T. R. Utterback, D. M. Saudek, C. A. Phillips, J. M. Merrick, J. F. Tomb, B. A. Dougherty, K. F. Bott, P. C. Hu, T. S. Lucier, S. N. Peterson, H. O. Smith, C. A. Hutchison, 3rd et J. C. Venter (1995). "The minimal gene complement of *Mycoplasma genitalium*." Science **270**(5235): 397-403.

Frost, L. S., K. Ippen-Ihler et R. A. Skurray (1994). "Analysis of the sequence and gene products of the transfer region of the F sex factor." Microbiol Rev **58**(2): 162-210.

Fukagawa, T. (2004). "Assembly of kinetochores in vertebrate cells." Exp Cell Res **296**(1): 21-7.

Glaser, P., C. Rusniok, C. Buchrieser, F. Chevalier, L. Frangeul, T. Msadek, M. Zouine, E. Couve, L. Lalioui, C. Poyart, P. Trieu-Cuot et F. Kunst (2002). "Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease." Mol Microbiol **45**(6): 1499-513.

Glockner, F. O., M. Kube, M. Bauer, H. Teeling, T. Lombardot, W. Ludwig, D. Gade, A. Beck, K. Borzym, K. Heitmann, R. Rabus, H. Schlesner, R. Amann et R. Reinhardt (2003). "Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1." Proc Natl Acad Sci U S A **100**(14): 8298-303.

Goh, E. B., G. Yim, W. Tsui, J. McClure, M. G. Surette et J. Davies (2002). "Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics." Proc Natl Acad Sci U S A **99**(26): 17025-30.

Gong, C., P. Bongiorno, A. Martins, N. C. Stephanou, H. Zhu, S. Shuman et M. S. Glickman (2005). "Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C." Nat Struct Mol Biol **12**(4): 304-12.

Gong, C., A. Martins, P. Bongiorno, M. Glickman et S. Shuman (2004). "Biochemical and genetic analysis of the four DNA ligases of mycobacteria." J Biol Chem **279**(20): 20594-606.

Gordon, D., C. Abajian et P. Green (1998). "Consed: a graphical tool for sequence finishing." Genome Res **8**(3): 195-202.

Goshi, K., T. Uchida, A. Lezhava, M. Yamasaki, K. Hiratsu, H. Shinkawa et H. Kinashi (2002). "Cloning and analysis of the telomere and terminal inverted repeat of the linear chromosome of *Streptomyces griseus*." J Bacteriol **184**(12): 3411-5.

Grohmann, E., G. Muth et M. Espinosa (2003). "Conjugative plasmid transfer in gram-positive bacteria." Microbiol Mol Biol Rev **67**(2): 277-301, table of contents.

Guo, X., M. Flores, P. Mavingui, S. I. Fuentes, G. Hernandez, G. Davila et R. Palacios (2003). "Natural genomic design in *Sinorhizobium meliloti*: novel genomic architectures." Genome Res **13**(8): 1810-7.

Gurtler, V., B. C. Mayall et R. Seviour (2004). "Can whole genome analysis refine the taxonomy of the genus *Rhodococcus*?" FEMS Microbiol Rev **28**(3): 377-403.

Gust, B., G. L. Challis, K. Fowler, T. Kieser et K. F. Chater (2003). "PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin." Proc Natl Acad Sci U S A **100**(4): 1541-6.

Haas, R. et T. F. Meyer (1987). "Molecular principles of antigenic variation in *Neisseria gonorrhoeae*." Antonie Van Leeuwenhoek **53**(6): 431-4.

Hanahan, D. (1983). "Studies on transformation of *Escherichia coli* with plasmids." J Mol Biol **166**(4): 557-80.

Heinemann, J. A. et G. F. Sprague, Jr. (1989). "Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast." Nature **340**(6230): 205-9.

Herbig, U., W. A. Jobling, B. P. Chen, D. J. Chen et J. M. Sedivy (2004). "Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21(CIP1), but not p16(INK4a)." Mol Cell **14**(4): 501-13.

Hill, C. W. et J. A. Gray (1988). "Effects of chromosomal inversion on cell fitness in *Escherichia coli* K-12." Genetics **119**(4): 771-8.

Hochhut, B., Y. Lotfi, D. Mazel, S. M. Faruque, R. Woodgate et M. K. Waldor (2001). "Molecular analysis of antibiotic resistance gene clusters in vibrio cholerae O139 and O1 SXT constins." Antimicrob Agents Chemother **45**(11): 2991-3000.

Hofmann, K., P. Bucher, L. Falquet et A. Bairoch (1999). "The PROSITE database, its status in 1999." Nucleic Acids Res **27**(1): 215-9.

Hopwood, D. A. (2006). "Soil to genomics: the *Streptomyces* chromosome." Annu Rev Genet **40**: 1-23.

Hopwood, D. A. et T. Kieser (1993). Conjugative plasmids of *Streptomyces*. New York, Plenum Press.

Hopwood, D. A., T. Kieser, H. M. Wright et M. J. Bibb (1983). "Plasmids, recombination and chromosome mapping in *Streptomyces lividans* 66." J Gen Microbiol **129**(7): 2257-69.

Hopwood, D. A., F. Malpartida, H. M. Kieser, H. Ikeda, J. Duncan, I. Fujii, B. A. Rudd, H. G. Floss et S. Omura (1985). "Production of 'hybrid' antibiotics by genetic engineering."

Nature **314**(6012): 642-4.

Huang, C. H., C. Y. Chen, H. H. Tsai, C. Chen, Y. S. Lin et C. W. Chen (2003). "Linear plasmid SLP2 of *Streptomyces lividans* is a composite replicon." Mol Microbiol **47**(6): 1563-76.

Huang, C. H., Y. S. Lin, Y. L. Yang, S. W. Huang et C. W. Chen (1998). "The telomeres of *Streptomyces* chromosomes contain conserved palindromic sequences with potential to form complex secondary structures." Mol Microbiol **28**(5): 905-16.

Huang, T. W. et C. W. Chen (2006). "A *recA* null mutation may be generated in *Streptomyces coelicolor*." J Bacteriol **188**(19): 6771-9.

Hutter, R. (1967). "[Classification of the streptomycetes with special regard to the antibiotics formed from them]." Bibl Microbiol **6**: 1-382.

Ide, H., H. Murayama, S. Sakamoto, K. Makino, K. Honda, H. Nakamuta, M. Sasaki et N. Sugimoto (1995). "On the mechanism of preferential incorporation of dAMP at abasic sites in translesional DNA synthesis. Role of proofreading activity of DNA polymerase and thermodynamic characterization of model template-primers containing an abasic site." Nucleic Acids Res **23**(1): 123-9.

Ikeda, H., J. Ishikawa, A. Hanamoto, M. Shinose, H. Kikuchi, T. Shiba, Y. Sakaki, M. Hattori et S. Omura (2003). "Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*." Nat Biotechnol **21**(5): 526-31.

Ishikawa, J., A. Yamashita, Y. Mikami, Y. Hoshino, H. Kurita, K. Hotta, T. Shiba et M. Hattori (2004). "The complete genomic sequence of *Nocardia farcinica* IFM 10152." Proc Natl Acad Sci U S A **101**(41): 14925-30.

Itahana, K., J. Campisi et G. P. Dimri (2004). "Mechanisms of cellular senescence in human and mouse cells." Biogerontology **5**(1): 1-10.

Jakimowicz, D., K. Chater et J. Zakrzewska-Czerwinska (2002). "The ParB protein of *Streptomyces coelicolor* A3(2) recognizes a cluster of *parS* sequences within the origin-proximal region of the linear chromosome." Mol Microbiol **45**(5): 1365-77.

Jakimowicz, D., S. Mouz, J. Zakrzewska-Czerwinska et K. F. Chater (2006). "Developmental control of a parAB promoter leads to formation of sporulation-associated ParB complexes in Streptomyces coelicolor." J Bacteriol **188**(5): 1710-20.

Jeong, H., S. P. Mason, A. L. Barabasi et Z. N. Oltvai (2001). "Lethality and centrality in protein networks." Nature **411**(6833): 41-2.

Kataoka, M., T. Seki et T. Yoshida (1991). "Five genes involved in self-transmission of pSN22, a *Streptomyces* plasmid." J Bacteriol **173**(13): 4220-8.

Kerr, A. R., J. F. Peden et P. M. Sharp (1997). "Systematic base composition variation around the genome of *Mycoplasma genitalium*, but not *Mycoplasma pneumoniae*." Mol Microbiol **25**(6): 1177-9.

Kieser, T., M. J. Bibb, M. J. Buttner, K. F. Chater et D. A. Hopwood (2000). Pratical *Streptomyces* genetics. Norwich.

Kieser, T., D. A. Hopwood, H. M. Wright et C. J. Thompson (1982). "pIJ101, a multi-copy broad host-range *Streptomyces* plasmid: functional analysis and development of DNA cloning vectors." Mol Gen Genet **185**(2): 223-8.

Kim, H. J., M. J. Calcutt, F. J. Schmidt et K. F. Chater (2000). "Partitioning of the linear chromosome during sporulation of *Streptomyces coelicolor* A3(2) involves an *oriC*-linked *parAB* locus." J Bacteriol **182**(5): 1313-20.

Kinashi, H., M. Shimaji-Murayama et T. Hanafusa (1991). "Nucleotide sequence analysis of the unusually long terminal inverted repeats of a giant linear plasmid, SCP1." Plasmid **26**(2): 123-30.

Kivisaar, M. (2003). "Stationary phase mutagenesis: mechanisms that accelerate adaptation of microbial populations under environmental stress." Environ Microbiol **5**(10): 814-27.

Klasson, L. et S. G. Andersson (2004). "Evolution of minimal-gene-sets in host-dependent bacteria." Trends Microbiol **12**(1): 37-43.

Kowalczykowski, S. C. (2000). "Initiation of genetic recombination and recombination-dependent replication." Trends Biochem Sci **25**(4): 156-65.

Kramer, K. M., J. A. Brock, K. Bloom, J. K. Moore et J. E. Haber (1994). "Two different types of double-strand breaks in *Saccharomyces cerevisiae* are repaired by similar RAD52-independent, nonhomologous recombination events." Mol Cell Biol **14**(2): 1293-301.

Krylov, D. M., Y. I. Wolf, I. B. Rogozin et E. V. Koonin (2003). "Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution." Genome Res **13**(10): 2229-35.

Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S. K. Choi, J. J. Codani, I. F. Connerton, A. Danchin et et al. (1997). "The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*." Nature **390**(6657): 249-56.

Lawrence, J. G. et H. Ochman (1997). "Amelioration of bacterial genomes: rates of change and exchange." J Mol Evol **44**(4): 383-97.

Lawrence, J. G. et H. Ochman (1998). "Molecular archaeology of the *Escherichia coli* genome." Proc Natl Acad Sci U S A **95**(16): 9413-7.

Leblond, P. et B. Decaris (1994). "New insights into the genetic instability of S*treptomyces*." FEMS Microbiol Lett **123**(3): 225-32.

Leblond, P. et B. Decaris (1999). ""Unstable" linear chromosome: the case of *Streptomyces*." Organization of the Prokaryotic Genome. Charlebois, R.L. (ed)
Washington, D.C.: American Society for Microbiology: 235-261.

Leblond, P., P. Demuyter, L. Moutier, M. Laakel, B. Decaris et J. M. Simonet (1989). "Hypervariability, a new phenomenon of genetic instability, related to DNA amplification in *Streptomyces ambofaciens*." J Bacteriol **171**(1): 419-23.

Leblond, P., G. Fischer, F. X. Francou, F. Berger, M. Guerineau et B. Decaris (1996). "The unstable region of *Streptomyces ambofaciens* includes 210 kb terminal inverted repeats flanking the extremities of the linear chromosomal DNA." Mol Microbiol **19**(2): 261-71.

Light, S., P. Kraulis et A. Elofsson (2005). "Preferential attachment in the evolution of metabolic networks." BMC Genomics **6**: 159.

Lin, D. C. et A. D. Grossman (1998). "Identification and characterization of a bacterial chromosome partitioning site." Cell **92**(5): 675-85.

Lin, Y. S., H. M. Kieser, D. A. Hopwood et C. W. Chen (1993). "The chromosomal DNA of *Streptomyces lividans* 66 is linear." Mol Microbiol **10**(5): 923-33.

Linardopoulou, E. V., E. M. Williams, Y. Fan, C. Friedman, J. M. Young et B. J. Trask (2005). "Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication." Nature **437**(7055): 94-100.

Linares, J. F., I. Gustafsson, F. Baquero et J. L. Martinez (2006). "Antibiotics as intermicrobial signaling agents instead of weapons." Proc Natl Acad Sci U S A **103**(51): 19484-9.

Lindsay, J. A., A. Ruzin, H. F. Ross, N. Kurepina et R. P. Novick (1998). "The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*." Mol Microbiol **29**(2): 527-43.

Lobry, J. R. et N. Sueoka (2002). "Asymmetric directional mutation pressures in bacteria." Genome Biol **3**(10): RESEARCH0058.

Lorenz, M. G. et W. Wackernagel (1994). "Bacterial gene transfer by natural genetic transformation in the environment." Microbiol Rev **58**(3): 563-602.

Louis, E. J. (1995). "The chromosome ends of *Saccharomyces cerevisiae*." Yeast **11**(16): 1553-73.

Lovett, S. T. et R. D. Kolodner (1989). "Identification and purification of a single-stranded-DNA-specific exonuclease encoded by the *recJ* gene of *Escherichia coli*." Proc Natl Acad Sci U S A **86**(8): 2627-31.

Mackiewicz, P., D. Mackiewicz, M. Kowalczuk et S. Cebrat (2001). "Flip-flop around the origin and terminus of replication in prokaryotic genomes." Genome Biol **2**(12): INTERACTIONS1004.

Maier, A., J. Riedlinger, H. P. Fiedler et R. Hampp (2004). "Actinomycetales bacteria from a spruce stand: characterization and effects on growth of root symbiotic, and plant parasitic soil fungi in dual culture." Mycological Progress **3**: 129-139.

Majewski, J., P. Zawadzki, P. Pickerill, F. M. Cohan et C. G. Dowson (2000). "Barriers to genetic exchange between bacterial species: *Streptococcus pneumoniae* transformation." J Bacteriol **182**(4): 1016-23.

Malkova, A., E. L. Ivanov et J. E. Haber (1996). "Double-strand break repair in the absence of RAD51 in yeast: a possible role for break-induced DNA replication." Proc Natl Acad Sci U S A **93**(14): 7131-6.

Malpartida, F. et D. A. Hopwood (1986). "Physical and genetic characterisation of the gene cluster for the antibiotic actinorhodin in *Streptomyces coelicolor* A3(2)." Mol Gen Genet **205**(1): 66-73.

Martin, P., A. Dary et B. Decaris (1998). "Generation of a genetic polymorphism in clonal populations of the bacterium *Streptomyces ambofaciens*: characterization of different mutator states." Mutat Res **421**(1): 73-82.

Mathieu, N., L. Pirzio, M. A. Freulet-Marriere, C. Desmaze et L. Sabatier (2004). "Telomeres and chromosomal instability." Cell Mol Life Sci **61**(6): 641-56.

McClintock, B. (1941). "The Stability of Broken Ends of Chromosomes in *Zea Mays*." Genetics **26**(2): 234-82.

McLeod, M. P., R. L. Warren, W. W. Hsiao, N. Araki, M. Myhre, C. Fernandes, D. Miyazawa, W. Wong, A. L. Lillquist, D. Wang, M. Dosanjh, H. Hara, A. Petrescu, R. D. Morin, G. Yang, J. M. Stott, J. E. Schein, H. Shin, D. Smailus, A. S. Siddiqui, M. A. Marra, S. J. Jones, R. Holt, F. S. Brinkman, K. Miyauchi, M. Fukuda, J. E. Davies, W. W. Mohn et L. D. Eltis (2006). "The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse." Proc Natl Acad Sci U S A **103**(42): 15582-7.

Medigue, C., T. Rouxel, P. Vigier, A. Henaut et A. Danchin (1991). "Evidence for horizontal gene transfer in *Escherichia coli* speciation." J Mol Biol **222**(4): 851-6.

Mefford, H. C. et B. J. Trask (2002). "The complex structure and dynamic evolution of human subtelomeres." Nat Rev Genet **3**(2): 91-102.

Mehr, I. J. et H. S. Seifert (1998). "Differential roles of homologous recombination pathways in *Neisseria gonorrhoeae* pilin antigenic variation, DNA transformation and DNA repair." Mol Microbiol **30**(4): 697-710.

Mendez, C., A. F. Brana, M. B. Manzanal et C. Hardisson (1985). "Role of substrate mycelium in colony development in *Streptomyces*." Can J Microbiol **31**(5): 446-50.

Milkman, R. et M. M. Bridges (1993). "Molecular evolution of the *Escherichia coli* chromosome. IV. Sequence comparisons." Genetics **133**(3): 455-68.

Milkman, R., E. Jaeger et R. D. McBride (2003). "Molecular evolution of the *Escherichia coli* chromosome. VI. Two regions of high effective recombination." Genetics **163**(2): 475-83.

Moran, N. A. (2002). "Microbial minimalism: genome reduction in bacterial pathogens." Cell **108**(5): 583-6.

Murphy, K. C. (1998). "Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*." J Bacteriol **180**(8): 2063-71.

Musialowski, M. S., F. Flett, G. B. Scott, G. Hobbs, C. P. Smith et S. G. Oliver (1994). "Functional evidence that the principal DNA replication origin of the *Streptomyces coelicolor* chromosome is close to the *dnaA-gyrB* region." J Bacteriol **176**(16): 5123-5.

Nakabachi, A., A. Yamashita, H. Toh, H. Ishikawa, H. E. Dunbar, N. A. Moran et M. Hattori

(2006). "The 160-kilobase genome of the bacterial endosymbiont *Carsonella*." Science **314**(5797): 267.

Nikolaou, C. et Y. Almirantis (2005). "A study on the correlation of nucleotide skews and the positioning of the origin of replication: different modes of replication in bacterial species." Nucleic Acids Res **33**(21): 6816-22.

Nomura, M. et E. A. Morgan (1977). "Genetics of bacterial ribosomes." Annu Rev Genet **11**: 297-347.

Normand, P., P. Lapierre, L. S. Tisa, J. P. Gogarten, N. Alloisio, E. Bagnarol, C. A. Bassi, A. M. Berry, D. M. Bickhart, N. Choisne, A. Couloux, B. Cournoyer, S. Cruveiller, V. Daubin, N. Demange, M. P. Francino, E. Goltsman, Y. Huang, O. R. Kopp, L. Labarre, A. Lapidus, C. Lavire, J. Marechal, M. Martinez, J. E. Mastronunzio, B. C. Mullin, J. Niemann, P. Pujic, T. Rawnsley, Z. Rouy, C. Schenowitz, A. Sellstedt, F. Tavares, J. P. Tomkins, D. Vallenet, C. Valverde, L. G. Wall, Y. Wang, C. Medigue et D. R. Benson (2007). "Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography." Genome Res **17**(1): 7-15.

Nowell, P. C. et C. M. Croce (1986). "Chromosomes, genes, and cancer." Am J Pathol **125**(1): 7-15.

Ochman, H., J. G. Lawrence et E. A. Groisman (2000). "Lateral gene transfer and the nature of bacterial innovation." Nature **405**(6784): 299-304.

Ochman, H., R. Liu et E. P. Rocha (2007). "Erosion of interaction networks in reduced and degraded genomes." J Exp Zoolog B Mol Dev Evol **308**(1): 97-103.

Oliynyk, M., M. Samborskyy, J. B. Lester, T. Mironenko, N. Scott, S. Dickens, S. F. Haydock et P. F. Leadlay (2007). "Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338." Nat Biotechnol **25**(4): 447-453.

Paget, M. S., L. Chamberlin, A. Atrih, S. J. Foster et M. J. Buttner (1999). "Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2)." J Bacteriol **181**(1): 204-11.

Pakula, A. A. et R. T. Sauer (1989). "Genetic analysis of protein stability and function." Annu Rev Genet **23**: 289-310.

Pang, X., B. Aigle, J. M. Girardet, S. Mangenot, J. L. Pernodet, B. Decaris et P. Leblond (2004). "Functional angucycline-like antibiotic gene cluster in the terminal inverted repeats of the *Streptomyces ambofaciens* linear chromosome." Antimicrob Agents Chemother **48**(2): 575-88.

Pansegrau, W., E. Lanka, P. T. Barth, D. H. Figurski, D. G. Guiney, D. Haas, D. R. Helinski, H. Schwab, V. A. Stanisich et C. M. Thomas (1994). "Complete nucleotide sequence of Birmingham IncP alpha plasmids. Compilation and comparative analysis." J Mol Biol **239**(5): 623-63.

Parkhill, J., M. Sebaihia, A. Preston, L. D. Murphy, N. Thomson, D. E. Harris, M. T. Holden, C. M. Churcher, S. D. Bentley, K. L. Mungall, A. M. Cerdeno-Tarraga, L. Temple, K. James, B. Harris, M. A. Quail, M. Achtman, R. Atkin, S. Baker, D. Basham, N. Bason, I. Cherevach, T. Chillingworth, M. Collins, A. Cronin, P. Davis, J.

Doggett, T. Feltwell, A. Goble, N. Hamlin, H. Hauser, S. Holroyd, K. Jagels, S. Leather, S. Moule, H. Norberczak, S. O'Neil, D. Ormond, C. Price, E. Rabbinowitsch, S. Rutter, M. Sanders, D. Saunders, K. Seeger, S. Sharp, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, L. Unwin, S. Whitehead, B. G. Barrell et D. J. Maskell (2003). "Comparative analysis of the genome sequences of *Bordetella pertussis, Bordetella parapertussis* and *Bordetella bronchiseptica*." Nat Genet **35**(1): 32-40.

Parkhill, J., B. W. Wren, N. R. Thomson, R. W. Titball, M. T. Holden, M. B. Prentice, M. Sebaihia, K. D. James, C. Churcher, K. L. Mungall, S. Baker, D. Basham, S. D. Bentley, K. Brooks, A. M. Cerdeno-Tarraga, T. Chillingworth, A. Cronin, R. M. Davies, P. Davis, G. Dougan, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Leather, S. Moule, P. C. Oyston, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead et B. G. Barrell (2001). "Genome sequence of *Yersinia pestis*, the causative agent of plague." Nature **413**(6855): 523-7.

Pease, P. J., O. Levy, G. J. Cost, J. Gore, J. L. Ptacin, D. Sherratt, C. Bustamante et N. R. Cozzarelli (2005). "Sequence-directed DNA translocation by purified FtsK." Science **307**(5709): 586-90.

Perna, N. T., G. Plunkett, 3rd, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch et F. R. Blattner (2001). "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7." Nature **409**(6819): 529-33.

Pernodet, J. L., J. M. Simonet et M. Guerineau (1984). "Plasmids in different strains of *Streptomyces ambofaciens*: free and integrated form of plasmid pSAM2." Mol Gen Genet **198**(1): 35-41.

Pettis, G. S. et S. N. Cohen (1994). "Transfer of the plJ101 plasmid in *Streptomyces lividans* requires a cis-acting function dispensable for chromosomal gene transfer." Mol Microbiol **13**(6): 955-64.

Pfeiffer, P., W. Goedecke et G. Obe (2000). "Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations." Mutagenesis **15**(4): 289-302.

Pinnert-Sindico, S. (1954). "Not Available." Ann Inst Pasteur (Paris) **87**(6): 702-7.

Pinnert-Sindico, S., L. Ninet, J. Preud'homme et C. Cosar (1954-1955). "A new antibiotic spiramycin." Antibiot. Ann.: 724-727.

Pisabarro, A., A. Correia et J. F. Martin (1998). "Pulsed-field gel electrophoresis analysis of the genome of *Rhodococcus fascians*: genome size and linear and circular replicon composition in virulent and avirulent strains." Curr Microbiol **36**(5): 302-8.

Pitcher, R. S., L. M. Tonkin, A. J. Green et A. J. Doherty (2005). "Domain structure of a NHEJ DNA repair ligase from *Mycobacterium tuberculosis*." J Mol Biol **351**(3): 531-44.

Possoz, C., C. Ribard, J. Gagnat, J. L. Pernodet et M. Guerineau (2001). "The integrative element pSAM2 from *Streptomyces*: kinetics and mode of conjugal transfer." Mol Microbiol **42**(1): 159-66.

Price-Whelan, A., L. E. Dietrich et D. K. Newman (2006). "Rethinking 'secondary' metabolism: physiological roles for phenazine antibiotics." Nat Chem Biol **2**(2): 71-8.

Pridham, T. G., P. Anderson, C. Foley, L. A. Lidenfelser et R. G. Benedict (1957). "A selection of medium for maintenance and taxonomic study of *Streptomyces*." Antibiot Annu **1956/57**: 947-953.

Ravin, V., N. Ravin, S. Casjens, M. E. Ford, G. F. Hatfull et R. W. Hendrix (2000). "Genomic sequence and analysis of the atypical temperate bacteriophage N15." J Mol Biol **299**(1): 53-73.

Raynal, A., K. Tuphile, C. Gerbaud, T. Luther, M. Guerineau et J. L. Pernodet (1998). "Structure of the chromosomal insertion site for pSAM2: functional analysis in *Escherichia coli*." Mol Microbiol **28**(2): 333-42.

Redenbach, M., F. Flett, W. Piendl, I. Glocker, U. Rauland, O. Wafzig, R. Kliem, P. Leblond et J. Cullum (1993). "The *Streptomyces lividans* 66 chromosome contains a 1 MB deletogenic region flanked by two amplifiable regions." Mol Gen Genet **241**(3-4): 255-62.

Reimmann, C. et D. Haas (1993). Mobilization of Chromosomes and Nonconjugative Plasmids by Cointegrative Mechanisms. New York, Plenum Press.

Resnick, M. A. et P. Martin (1976). "The repair of double-strand breaks in the nuclear DNA of *Saccharomyces cerevisiae* and its genetic control." Mol Gen Genet **143**(2): 119-29.

Ricchetti, M., B. Dujon et C. Fairhead (2003). "Distance from the chromosome end determines the efficiency of double strand break repair in subtelomeres of haploid yeast." J Mol Biol **328**(4): 847-62.

Ricchetti, M., C. Fairhead et B. Dujon (1999). "Mitochondrial DNA repairs double-strand breaks in yeast chromosomes." Nature **402**(6757): 96-100.

Rocha, E. (2002). "Is there a role for replication fork asymmetry in the distribution of genes in bacterial genomes?" Trends Microbiol **10**(9): 393-5.

Rocha, E. P. (2003). "DNA repeats lead to the accelerated loss of gene order in bacteria." Trends Genet **19**(11): 600-3.

Rocha, E. P. (2006). "Inference and analysis of the relative stability of bacterial chromosomes." Mol Biol Evol **23**(3): 513-22.

Rocha, E. P., E. Cornet et B. Michel (2005). "Comparative and evolutionary analysis of the bacterial homologous recombination systems." PLoS Genet **1**(2): e15.

Rocha, E. P. et A. Danchin (2002). "Base composition bias might result from competition for metabolic resources." Trends Genet **18**(6): 291-4.

Rocha, E. P. et A. Danchin (2003a). "Essentiality, not expressiveness, drives gene-strand bias in bacteria." Nat Genet **34**(4): 377-8.

Rocha, E. P. et A. Danchin (2003b). "Gene essentiality determines chromosome organisation in bacteria." Nucleic Acids Res **31**(22): 6570-7.

Roth, V., B. Aigle, R. Bunet, T. Wenner, C. Fourrier, B. Decaris et P. Leblond (2004). "Differential and cross-transcriptional control of duplicated genes encoding alternative sigma factors in *Streptomyces ambofaciens*." J Bacteriol **186**(16): 5355-65.

Rothkamm, K., I. Kruger, L. H. Thompson et M. Lobrich (2003). "Pathways of DNA double-strand break repair during the mammalian cell cycle." Mol Cell Biol **23**(16): 5706-15.

Rudolph, K. L., M. Millard, M. W. Bosenberg et R. A. DePinho (2001). "Telomere dysfunction and evolution of intestinal carcinoma in mice and humans." Nat Genet **28**(2): 155-9.

Russel, M., S. Kidd et M. R. Kelley (1986). "An improved filamentous helper phage for generating single-stranded plasmid DNA." Gene **45**(3): 333-8.

Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream et B. Barrell (2000). "Artemis: sequence visualization and annotation." Bioinformatics **16**(10): 944-5.

Sabatier, L., M. Ricoul, G. Pottier et J. P. Murnane (2005). "The loss of a single telomere can result in instability of multiple chromosomes in a human tumor cell line." Mol Cancer Res **3**(3): 139-50.

Saint-Ruf, C. et I. Matic (2006). "Environmental tuning of mutation rates." Environ Microbiol **8**(2): 193-9.

Salanoubat, M., S. Genin, F. Artiguenave, J. Gouzy, S. Mangenot, M. Arlat, A. Billault, P. Brottier, J. C. Camus, L. Cattolico, M. Chandler, N. Choisne, C. Claudel-Renard, S. Cunnac, N. Demange, C. Gaspin, M. Lavie, A. Moisan, C. Robert, W. Saurin, T. Schiex, P. Siguier, P. Thebault, M. Whalen, P. Wincker, M. Levy, J. Weissenbach et C. A. Boucher (2002). "Genome sequence of the plant pathogen *Ralstonia solanacearum*." Nature **415**(6871): 497-502.

Salas, M. (1991). "Protein-priming of DNA replication." Annu Rev Biochem **60**: 39-71.

Sambrook, J., E. F. Fritsch et T. Maniatis (1989). Molecular Cloning: A Laboratory Manual. New York.

Sanford, R. A., J. R. Cole et J. M. Tiedje (2002). "Characterization and description of *Anaeromyxobacter dehalogenans* gen. nov., sp. nov., an aryl-halorespiring facultative anaerobic myxobacterium." Appl Environ Microbiol **68**(2): 893-900.

Saunders, N. J., A. C. Jeffries, J. F. Peden, D. W. Hood, H. Tettelin, R. Rappuoli et E. R. Moxon (2000). "Repeat-associated phase variable genes in the complete genome sequence of *Neisseria meningitidis* strain MC58." Mol Microbiol **37**(1): 207-15.

Schneider, R., R. Lurz, G. Luder, C. Tolksdorf, A. Travers et G. Muskhelishvili (2001). "An architectural role of the *Escherichia coli* chromatin protein FIS in organising DNA." Nucleic Acids Res **29**(24): 5107-14.

Schrey, S. D., M. Schellhammer, M. Ecke, R. Hampp et M. T. Tarkka (2005). "Mycorrhiza helper bacterium *Streptomyces* AcH 505 induces differential gene expression in the ectomycorrhizal fungus *Amanita muscaria*." New Phytol **168**(1): 205-16.

Schwedock, J., J. R. McCormick, E. R. Angert, J. R. Nodwell et R. Losick (1997). "Assembly of the cell division protein FtsZ into ladder-like structures in the aerial hyphae of

Streptomyces coelicolor." Mol Microbiol **25**(5): 847-58.

Scott, J. R. et G. G. Churchward (1995). "Conjugative transposition." Annu Rev Microbiol **49**: 367-97.

Sezonov, G., J. Hagege, J. L. Pernodet, A. Friedmann et M. Guerineau (1995). "Characterization of *pra*, a gene for replication control in pSAM2, the integrating element of *Streptomyces ambofaciens*." Mol Microbiol **17**(3): 533-44.

Shi, W. et D. R. Zusman (1993). "Fatal attraction." Nature **366**(6454): 414-5.

Sibley, M. H. et E. A. Raleigh (2004). "Cassette-like variation of restriction enzyme genes in *Escherichia coli* C and relatives." Nucleic Acids Res **32**(2): 522-34.

Singleton, M. R., M. S. Dillingham, M. Gaudier, S. C. Kowalczykowski et D. B. Wigley (2004). "Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks." Nature **432**(7014): 187-93.

Smit, A. F. et A. D. Riggs (1996). "Tiggers and DNA transposon fossils in the human genome." Proc Natl Acad Sci U S A **93**(4): 1443-8.

Spratt, B. G., W. P. Hanage et E. J. Feil (2001). "The relative contributions of recombination and point mutation to the diversification of bacterial clones." Curr Opin Microbiol **4**(5): 602-6.

Stein, D. S. et S. N. Cohen (1990). "Mutational and functional analysis of the korA and korB gene products of Streptomyces plasmid pIJ101." Mol Gen Genet **222**(2-3): 337-44.

Stephanou, N. C., F. Gao, P. Bongiorno, S. Ehrt, D. Schnappinger, S. Shuman et M. S. Glickman (2007). "Mycobacterial nonhomologous end joining mediates mutagenic repair of chromosomal double-strand DNA breaks." J Bacteriol **189**(14): 5237-46.

Thomas, C. M. et K. M. Nielsen (2005). "Mechanisms of, and barriers to, horizontal gene transfer between bacteria." Nat Rev Microbiol **3**(9): 711-21.

Tillier, E. R. et R. A. Collins (2000). "Genome rearrangement by replication-directed translocation." Nat Genet **26**(2): 195-7.

Toh, H., B. L. Weiss, S. A. Perkin, A. Yamashita, K. Oshima, M. Hattori et S. Aksoy (2006). "Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host." Genome Res **16**(2): 149-56.

Trask, B. J., C. Friedman, A. Martin-Gallardo, L. Rowen, C. Akinbami, J. Blankenship, C. Collins, D. Giorgi, S. Iadonato, F. Johnson, W. L. Kuo, H. Massa, T. Morrish, S. Naylor, O. T. Nguyen, S. Rouquier, T. Smith, D. J. Wong, J. Youngblom et G. van den Engh (1998). "Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes." Hum Mol Genet **7**(1): 13-26.

Travers, A., R. Schneider et G. Muskhelishvili (2001). "DNA supercoiling and transcription in *Escherichia coli:* The FIS connection." Biochimie **83**(2): 213-7.

Trieu-Cuot, P., G. Gerbaud, T. Lambert et P. Courvalin (1985). "In vivo transfer of genetic

information between gram-positive and gram-negative bacteria." Embo J **4**(13A): 3583-7.

Tsui, H. C., G. Feng et M. E. Winkler (1997). "Negative regulation of *mutS* and *mutH* repair gene expression by the Hfq and RpoS global regulators of *Escherichia coli* K-12." J Bacteriol **179**(23): 7476-87.

Ussery, D., T. S. Larsen, K. T. Wilkes, C. Friis, P. Worning, A. Krogh et S. Brunak (2001). "Genome organisation and chromatin structure in *Escherichia coli*." Biochimie **83**(2): 201-12.

van der Meer, J. R., J. Polman, M. M. Beerthuyzen, R. J. Siezen, O. P. Kuipers et W. M. De Vos (1993). "Characterization of the *Lactococcus lactis* nisin A operon genes *nisP*, encoding a subtilisin-like serine protease involved in precursor processing, and *nisR*, encoding a regulatory protein involved in nisin biosynthesis." J Bacteriol **175**(9): 2578-88.

Viguera, E., D. Canceill et S. D. Ehrlich (2001). "Replication slippage involves DNA polymerase pausing and dissociation." Embo J **20**(10): 2587-95.

Volff, J. N. et J. Altenbuchner (1997). "Influence of disruption of the recA gene on genetic instability and genome rearrangement in *Streptomyces lividans*." J Bacteriol **179**(7): 2440-5.

Volff, J. N. et J. Altenbuchner (2000). "A new beginning with new ends: linearisation of circular chromosomes during bacterial evolution." FEMS Microbiol Lett **186**(2): 143-50.

Volff, J. N., P. Viell et J. Altenbuchner (1997). "Artificial circularization of the chromosome with concomitant deletion of its terminal inverted repeats enhances genetic instability and genome rearrangement in *Streptomyces lividans*." Mol Gen Genet **253**(6): 753-60.

Waldor, M. K. et J. J. Mekalanos (1996). "Lysogenic conversion by a filamentous phage encoding cholera toxin." Science **272**(5270): 1910-4.

Wang, H., A. P. Roberts et P. Mullany (2000). "DNA sequence of the insertional hot spot of *Tn916* in the *Clostridium difficile* genome and discovery of a Tn916-like element in an environmental isolate integrated in the same hot spot." FEMS Microbiol Lett **192**(1): 15-20.

Wang, J., L. M. Parsons et K. M. Derbyshire (2003). "Unconventional conjugal DNA transfer in mycobacteria." Nat Genet **34**(1): 80-4.

Wang, L., Y. Yu, X. He, X. Zhou, Z. Deng, K. F. Chater et M. Tao (2007). "Role of an FtsK-like protein in genetic stability in *Streptomyces coelicolor* A3(2)." J Bacteriol **189**(6): 2310-8.

Weaver, D., N. Karoonuthaisiri, H. H. Tsai, C. H. Huang, M. L. Ho, S. Gai, K. G. Patel, J. Huang, S. N. Cohen, D. A. Hopwood, C. W. Chen et C. M. Kao (2004). "Genome plasticity in *Streptomyces:* identification of 1 Mb TIRs in the *S. coelicolor* A3(2) chromosome." Mol Microbiol **51**(6): 1535-50.

Weller, G. R., B. Kysela, R. Roy, L. M. Tonkin, E. Scanlan, M. Della, S. K. Devine, J. P. Day, A. Wilkinson, F. d'Adda di Fagagna, K. M. Devine, R. P. Bowater, P. A. Jeggo, S. P. Jackson et A. J. Doherty (2002). "Identification of a DNA nonhomologous end-joining complex in bacteria." Science **297**(5587): 1686-9.

Wenner, T., V. Roth, B. Decaris et P. Leblond (2002). "Intragenomic and intraspecific polymorphism of the 16S-23S rDNA internally transcribed sequences of *Streptomyces ambofaciens*." Microbiology **148**(Pt 3): 633-42.

Wenner, T., V. Roth, G. Fischer, C. Fourrier, B. Aigle, B. Decaris et P. Leblond (2003). "End-to-end fusion of linear deleted chromosomes initiates a cycle of genome instability in Streptomyces ambofaciens." Mol Microbiol **50**(2): 411-25.

Winzeler, E. A., D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J. D. Boeke, H. Bussey, A. M. Chu, C. Connelly, K. Davis, F. Dietrich, S. W. Dow, M. El Bakkoury, F. Foury, S. H. Friend, E. Gentalen, G. Giaever, J. H. Hegemann, T. Jones, M. Laub, H. Liao, N. Liebundguth, D. J. Lockhart, A. Lucau-Danila, M. Lussier, N. M'Rabet, P. Menard, M. Mittmann, C. Pai, C. Rebischung, J. L. Revuelta, L. Riles, C. J. Roberts, P. Ross-MacDonald, B. Scherens, M. Snyder, S. Sookhai-Mahadeo, R. K. Storms, S. Veronneau, M. Voet, G. Volckaert, T. R. Ward, R. Wysocki, G. S. Yen, K. Yu, K. Zimmermann, P. Philippsen, M. Johnston et R. W. Davis (1999). "Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis." Science **285**(5429): 901-6.

Wohlleben, W., V. Hartmann, D. Hillemann, K. Krey, G. Muth, B. Nussbaumer et S. Pelzer (1994). "Transfer and establishment of DNA in *Streptomyces* (a brief review)." Acta Microbiol Immunol Hung **41**(4): 381-9.

Wright, F. et M. J. Bibb (1992). "Codon usage in the G+C-rich *Streptomyces* genome." Gene **113**(1): 55-65.

Yamaichi, Y. et H. Niki (2004). "*migS*, a cis-acting site that affects bipolar positioning of *oriC* on the *Escherichia coli* chromosome." Embo J **23**(1): 221-33.

Yamasaki, M. et H. Kinashi (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**(19): 6553-9.

Yim, G., H. H. Wang et J. Davies (2006). "The truth about antibiotics." Int J Med Microbiol **296**(2-3): 163-70.

Yu, X. C., A. H. Tran, Q. Sun et W. Margolin (1998). "Localization of cell division protein FtsK to the *Escherichia coli* septum and identification of a potential N-terminal targeting domain." J Bacteriol **180**(5): 1296-304.

Zhang, R., Y. Yang, P. Fang, C. Jiang, L. Xu, Y. Zhu, M. Shen, H. Xia, J. Zhao, T. Chen et Z. Qin (2006). "Diversity of telomere palindromic sequences and replication genes among *Streptomyces* linear plasmids." Appl Environ Microbiol **72**(9): 5728-33.

Zhang, R. et C. T. Zhang (2005). "Identification of replication origins in archaeal genomes based on the Z-curve method." Archaea **1**(5): 335-46.

Zivanovic, Y., P. Lopez, H. Philippe et P. Forterre (2002). "*Pyrococcus* genome comparison evidences chromosome shuffling-driven evolution." Nucleic Acids Res **30**(9): 1902-10.

RESUME

L'avènement du séquençage systématique des génomes bactériens permet d'appréhender les mécanismes de l'évolution de ces génomes. Les *Streptomyces* sont des bactéries du sol utilisées dans l'industrie pour la synthèse de métabolites d'intérêt. Ces bactéries possèdent des caractéristiques génomiques remarquables avec un chromosome linéaire de grande taille (8-11 Mb) présentant un fort taux en bases G-C (72,1% chez *Streptomyces coelicolor*). Une autre caractéristique de *Streptomyces* est l'apparition à haute fréquence de mutants (jusqu'à 1% chez *Streptomyces ambofaciens*). L'apparition de ces mutants est accompagnée de grands réarrangements touchant les régions terminales du chromosome.

La comparaison génomique de différentes espèces de *Streptomyces* permet de visualiser les réarrangements fixés au cours de l'évolution. Une variabilité des régions terminales a été observée. L'analyse des séquences des régions terminales et des séquences partielles de la région centrale du chromosome de la souche ATCC23877 de *S. ambofaciens* et leur comparaison avec les génomes d'autres *Streptomyces* a permis de montrer que la taille de la région spécifique d'espèce augmente avec l'éloignement phylogénétique et que réciproquement, la taille de la région centrale conservée entre deux espèces de *Streptomyces* diminue. Cette perte progressive de synténie est le résultat d'évènements d'insertions/délétions (indels) de petits fragments d'ADN dans ces régions.

La comparaison des séquences des TIR (ou Répétition Terminale Inversée) de deux souches de *S. ambofaciens* a montré la présence de frontières ancestrales communes. Pour chaque souche, les régions spécifiques se situent aux extrémités de ces TIR. Cette variabilité serait la conséquence de remplacements d'extrémités de réplicons linéaires. Ces extrémités pourraient avoir une origine plasmidique au vue des fonctions prédites dans ces régions. Les séquences télomériques sont particulièrement conservées dans cette région très variable.

Plusieurs mécanismes pourraient avoir façonné les régions terminales. Ainsi l'intervention différentielle de mécanismes de réparation de cassures double brin ou une fréquence plus importante des cassures le long du chromosome pourraient être à la base de cette variabilité. Une conséquence de la formation de cassures double brin est l'entrée dans le cycle de cassure-fusion-pont (CFP) dont un intermédiaire est la formation de chromosomes portant deux centromères (ici bactériens), à l'instar de mutants spontanés isolés au laboratoire. L'entrée dans ce cycle serait la conséquence de la fusion de deux chromosomes délétés d'un bras chromosomique, conduisant à la présence de deux centres de partition potentiels au sein de la même molécule. La construction de souches de *S. ambofaciens* DSM40697 contenant un second centre de partition a été entreprise et l'analyse de la descendance de ces souches permettra de préciser l'implication des systèmes de partition dans l'instabilité et l'évolution du chromosome des *Streptomyces*.

Mots-clés : Instabilité chromosomique, évolution, génomique, partition chromosomique, *Streptomyces ambofaciens*