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,* François-Xavier Francou,‡ 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 intra-cellular 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 bacteria living in a stable environmental niche, like the intra-cellular 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 (Tanas 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; Bocard 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
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 BamHI-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 Sau3A1-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× coverage 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 BstXI 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 tbl2seq (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 (Suze 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 syneny, protein sequence comparisons were extracted from SAMDB to be readable under ACT (Rutherford et al. 2000). Duplicated genes in the TIRs were named SAMTnnnn, whereas those specific only either to the left or right arms were annotated as SAMLnnnn and SAMRnnnn.

**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.webglm.scbiol.ups-nancy.fr/ambofaciens/. The sequences were deposited in European Molecular Biology Laboratory under the AM238665 (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
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. 1A). 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 (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
has occurred. The *S. ambofaciens/S. coelicolor* comparison using BAC alignments (fig. 1B) 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.

**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.
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 *ttaA*, 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 (*spaB*, *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 parcelled 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.
with \textit{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 \textit{S. ambofaciens} CDSs are orthologues with \textit{S. coelicolor} and \textit{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 \textit{S. ambofaciens}/\textit{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 \textit{S. ambofaciens}, a 34.5-kb cluster (28 genes) implied in secondary metabolism is species specific. In \textit{S. coelicolor}, this region is replaced by 30.8-kb–specific cluster (31 genes, SCO0850–SCO0880). In \textit{S. avermitilis}, the same region (SAV7356–SAV7422) corresponds to a secondary metabolite gene cluster (\textit{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 \textit{S. ambofaciens} left arm (fig. 2). The \textit{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 \textit{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 \textit{S. ambofaciens} when it is compared with \textit{S. avermitilis} (fig. 2). The slope of GOC variation is stronger when compared with \textit{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 \textit{S. coelicolor} and \textit{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 \textit{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 \textit{Streptomyces} (Pandza et al. 1998; Yamasaki and Kinashi 2004) as well as in \textit{Borrelia}.
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 adaptive 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 adaptive 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 syntenic 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 actinomyces 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’Éducation 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 PCRD T (“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


Martin Embley, Associate Editor

Accepted September 1, 2006