Original Article

Widespread homologous recombination within and between *Streptomyces* species

James R Doroghazi[1] and Daniel H Buckley[2]

[1]*Department of Microbiology, Cornell University, Ithaca, NY, USA* and [2]*Department of Crop and Soil Sciences, Cornell University, Ithaca, NY, USA*

Horizontal gene transfer (HGT) is widespread in the microbial world, but its impact on the origin and persistence of microbial species remains poorly defined. HGT can result in either acquisition of new genetic material or homologous replacement of existing genes. The evolutionary significance of homologous recombination in a population can be quantified by examining the relative rates at which polymorphisms are introduced from recombination (ρ) and mutation (θ_w). We used multilocus sequence analysis (MLSA) to quantify both intraspecies and interspecies homologous recombination among streptomycetes, multicellular Gram-positive bacteria ubiquitous in soil, which are an important source of antibiotics and bioactive compounds. Intraspecies recombination was examined using strains of *Streptomyces flavogriseus* isolated from soils at five locations spanning 1000 km. The strains had >99.8% nucleotide identity across the loci examined. We found remarkable levels of gene exchange within *S. flavogriseus* ($\rho/\theta_w = 27.9$), and found that the population was in linkage equilibrium (standardized index of association = 0.0018), providing evidence for a freely recombining sexual population structure. We also examined interspecies homologous recombination among different *Streptomyces* species in an MLSA data set and found that 40% of the species had housekeeping genes acquired through HGT. The recombination rate between these named species ($\rho/\theta_w = 0.21$) exceeds that observed within many species of bacteria. Despite widespread gene exchange in the genus, the intraspecies recombination rate exceeded the interspecies rate by two orders of magnitude suggesting that patterns of gene exchange and recombination may shape the evolution of streptomycetes.

Introduction

Evidence for extensive genomic heterogeneity and widespread gene exchange in microbial genomes has highlighted our limited understanding of microbial species, and caused the very idea of species to be questioned (Doolittle and Bapteste, 2007). Several models have been proposed to explain the origin and persistence of genetic clusters that correspond to microbial species (as reviewed in Fraser *et al.*, 2009). Ecotype models, which predict that adaptive mutations define a species as an occupant of a distinct ecological niche, are theoretically appealing and work well for some taxa (Koeppel *et al.*, 2008). Even in environments where the ability of periodic selection to purge diversity is limited by the patchy distribution of habitable sites in space and time, associations between genetic diversity and

environmental specialization have been observed (Hunt *et al.*, 2008). Ecotype-based models may be less appealing, however, for microbial populations characterized by high rates of gene exchange in which recombination limits the ability of periodic selection to purge genetic diversity (Fraser *et al.*, 2009). In such highly recombining populations, the origin and persistence of species may be defined primarily by biological, ecological or geographical barriers to gene exchange (Fraser *et al.*, 2007, 2009). For example, the recent elimination of ecological barriers to gene exchange has been suggested as a mechanism to explain patterns of admixture resulting from hybridization between *Campylobacter* species (Sheppard *et al.*, 2008). The general significance of homologous gene exchange in defining microbial species remains unclear, however, as rates of homologous exchange can vary widely, and available data mainly focus on pathogenic populations, which represent a narrow spectrum of the ecological and phylogenetic diversity in the microbial world (Vos and Didelot, 2009).

Streptomycetes represent an interesting model system for exploring the impact of gene exchange on the genetic diversity of microbial taxa. Originally

classified as fungi, streptomycetes are Gram-positive bacteria, which have a complex life cycle characterized by hyphal growth and mycelium formation followed by development of aerial hyphae and asexual production of spores. *Streptomyces* species have a large ($\geqslant$7–9 Mb) linear chromosome, which contains a central region that is highly conserved throughout the genus, and terminal regions that are variable in composition and organization (Hopwood, 2006). Conjugation in streptomycetes involves transfer of double-stranded DNA (dsDNA) and requires the product of only a single gene, *traB*, which encodes a septal DNA translocator protein (Kataoka *et al.*, 1991). Gene exchange takes place at the tips of elongating filaments and may be facilitated by hyphal fusion (Hopwood, 2006). Chromosomal markers are mobilized through this mechanism at a frequency of 0.1–1% (Hopwood *et al.*, 1985), with plasmid integration causing mobilization of chromosomal genes at an efficiency that approaches 100% (Chater *et al.*, 1982). This mechanism is in stark contrast to that observed in Gram-negative bacteria, which involves transfer of single-stranded DNA and requires the coordinated action of multiple proteins and an origin of transfer. The phylogenetic distribution of these two mechanisms of gene exchange among bacteria remains poorly characterized, but the dsDNA exchange system has been most clearly documented in multicellular Gram-positive bacteria. Surveys of homologous recombination rates do not currently include any populations that possess this dsDNA conjugation system.

Given the nature of gene exchange and recombination mechanisms in streptomycetes, we hypothesized that homologous recombination may have a significant impact on their evolution. We tested this hypothesis at the intraspecies level using strains from a single *Streptomyces* species. A collection of 37 isolates of *Streptomyces flavogriseus* was obtained from five sites spanning four counties in New York and one in Michigan. Multilocus sequence analysis (MLSA) of these isolates was used to estimate homologous recombination rates for the *S. flavogriseus* population. We also analyzed an existing MLSA data set of *Streptomyces* spp. to quantify rates of interspecies gene exchange and to evaluate the impact of horizontal gene transfer on the evolution of *Streptomyces*.

Materials and methods

Soil sampling and isolation of S. flavogriseus
S. flavogriseus isolates were obtained from soil samples (0–5 cm depth) taken from grassy fields identified in five locations in New York and Michigan. The sites sampled in New York were the Miner Institute, Chazy, Clinton County (N 44.884672, W −73.474429); Willsboro Farm, Willsboro, Washington County (N 44.385817, W −73.384850); Mitchell Street Field, Ithaca, Tompkins County (N 42.434654, W −76.471442); Caldwell Field, Ithaca, Tompkins County (N 42.450061, W−76.458782); and Harford Farm, Harford, Courtland County. The Michigan site was in Kentwood, Kent County (N 42.856419, W−85.622737). Fresh soil was diluted 1:100 in phosphate-buffered saline solution and 50 µl of this suspension was spread onto glycerol–arginine agar plates with a pH of 8.5–8.7 containing 300 mg l^{-1} cycloheximide (Elnakeeb and Lecheval, 1963). The most common streptomycete colony type on this media was white on the edges and become dark gray in the center after 5–7 days, at which point they are 1–2 mm diameter. The 16S rRNA and *rpoB* gene sequences of each isolate were screened (see MLSA below) and the 37 isolates obtained in this manner were found to be highly similar (>99.8% sequence similarity, Table 1).

The gene sequences from these 37 isolates were found to be highly similar (99.8% average sequence similarity across the five protein-coding genes examined, see below) to gene sequences found in the genome of *S. flavogriseus* IAF-45-CD (ATCC 33331, GenBank accession number: ACZH00000000). *S. flavogriseus* IAF-45-CD was originally isolated in Laval, Canada, at a site <100 km from Chazy, NY, USA (Ishaque and Kluepfel, 1980). In terms of morphology and carbon-source utilization profile (D-glucose (+), L-arabinose (+), D-xylose (+), raffinose (−), D-fructose (+), I-inositol (−), D-mannitol (+) and rhamnose (+)), our isolates match *S. flavogriseus* as described in the International *Streptomyces* Project (Shirling and Gottlieb, 1970) and in the Wink Compendium (http://www.gbif-prokarya.de/microorganisms/files/Wink-Compendium.xls). It should be noted that *S. flavogriseus* IAF-45-CD (ATCC33331) and the type strain *S. flavogriseus* Heim (ATCC25452) share only 93.5% similarity

Table 1 Estimated population parameters

	Length	π	Sites	Φ_w	I_a^s	ρ	θ_w	ρ/θ_w
Interspecies	3923	0.0568	808	2.6×10^{-15}	—	0.0097	0.0454	0.21
Intraspecies	2528	0.0018	21	0.0038	0.0018	0.0552	0.0020	27.9

Values for sequence length (length) and polymorphic sites (sites) are in nucleotides. The value for nucleotide diversity (π) is the average number of nucleotide differences per site. The value presented from the Φ_w test is the *P*-value calculated under the null hypothesis of no recombination. The value for the standardized index of association (I_a^s) is 0 for a population in linkage equilibrium. Values for ρ and θ_w were calculated using LDhat (McVean *et al.*, 2002) and are expressed per site for the concatenated gene sequences with ρ expressed as the rate of gene conversion, $\rho = 2N_e r/2$. Data for individual loci are provided in Supplementary Table S3.

across the five protein-coding genes examined and 97.1% 16S rRNA similarity, suggesting that our 37 isolates and *S. flavogriseus* IAF-45-CD represent a population that is genetically distinct from *S. flavogriseus* Heim. Collectively, we refer to our 37 strains and strain IAF-45-CD as *S. flavogriseus* phylogroup *pratensis* (from the Latin for 'growing in the meadow'). It should also be noted that the gene sequences provided for a strain described as '*Streptomyces griseoplanus*' in Guo *et al.* (2008) exactly match to those from strain IAF-45-CD, and given that the 16S rRNA gene provided for '*S. griseoplanus*' by Guo *et al.* (2008) does not match the 16S rRNA gene sequence deposited for the type strain of *S. griseoplanus* (ACCN: AB184138.1), it seems likely that strain IAF-45-CD was misidentified as '*S. griseoplanus*' in Guo *et al.* (2008).

MLSA of Streptomyces sp.

DNA was extracted and the MLSA scheme of Guo *et al.* (2008) was used to characterize the isolates. Owing to reported problems in Guo *et al.* (2008) with existing *gyrB* primers, we designed new primers using recently released genome data from additional *Streptomyces* species (gyrBF: 5′-CTG GACGCGGTCCGCAAGCG-3′; gyrBR: 5′-GTCTGGC CCTCGAACTGCGGGCT-3′). All reactions were carried out with the following 25 μl reaction using AmpliTaq Gold reagents (Applied Biosystems, Foster City, CA, USA): 11.75 μl H$_2$O, 2.5 μl 10× buffer, 3 μl 25 mM MgCl$_2$, 2 μl dNTP mixture (2.5 mM each dNTP, 10 mM total dNTPs, Promega, Madison, WI, USA), 1 μl forward primer from 10 μM stock, 1 μl reverse primer from 10 μM stock, 2.5 μl dimethyl sulfoxide, 0.25 μl AmpliTaq Gold (5 U μl^{-1}) and 1 μl template. For all primer sets, the following reaction conditions were used: 95 °C, 10 min for initial denaturation; 35 cycles of 95 °C for 20 s, 65 °C for 30 s, 72 °C for 45 s; 72 °C for 10 min as a final extension; and short-term storage at 4 °C. Sequences were assembled manually and trace files were inspected for all sequences at all polymorphic sites. To confirm results and verify the absence of cross-contamination, isolates Cald 193, Chazy 277, Harf 495, MS7 19, W25 20, W25 23, W25 25, W25 26 and W300 21 were removed from storage and the sequence of the *rpoB* and *traB* genes were determined for one to four different individual colonies from each isolate. In every case, the expected sequence type was recovered. Gene sequences are available from GenBank with accession numbers GU979234–GU979418.

Assessment of homologous recombination and population structure

The properties of individual loci used for interspecies and intraspecies comparisons are provided in Supplementary Table S1. Sequences of *Streptomyces* species were acquired from NCBI using

accession numbers provided in Guo *et al.* (2008). The 16S rRNA gene sequences were included in the concatenated alignment used for interspecies analyses, but were not used in intraspecies analysis because of a lack of polymorphism. The standardized index of association was calculated with LIAN v3.5 (Haubold and Hudson, 2000) for allelic data from the *S. flavogriseus* isolates. LDhat (McVean *et al.*, 2002) was used to estimate ρ both for concatenated sequences and single loci. Clonal-Frame was run with 100 000 burn-in updates followed by 100 000 more updates on data without genome positions included (Didelot and Falush, 2007). Maximum likelihood trees were created for each locus from the 53 *Streptomyces* species described in Guo *et al.* (2008) using PAUP v4.0Beta. Maximum likelihood trees were made using the tree bisection reconnection algorithm and incongruence between trees was then evaluated using the Shimodaira–Hasegawa test (Shimodaira and Hasegawa, 1999). Individual recombination events within the interspecies data set were found using the Recombination Detection Program v3b34 (Martin *et al.*, 2005). Reported *P*-values are calculated using the Bonferroni correction within the program. Neighbornet phylogenetic networks were created with Splitstree v4.10 (Huson and Bryant, 2006).

Given the extent of recombination observed between *Streptomyces* species, the program Structure was also used to examine population structure among these taxa. Structure was run on data with 20 000 burn-in and 100 000 updates using the linkage model with other parameters set to default (Falush *et al.*, 2003). Structure assumes that populations are in both linkage equilibrium and Hardy–Weinberg equilibrium. Although there is ample evidence for homologous gene exchange between species in *Streptomyces*, this evidence is insufficient in itself to confirm that these *Streptomyces* satisfy the assumptions of the Structure analysis. Although the linkage model is designed to relax these assumptions and permit clustering for a wide range of population structures (Falush *et al.*, 2003), it is important to consider the impact of violating these assumptions. The admixed model in Structure tries to find the largest populations that are in equilibrium with admixture introduced to cope with linkage disequilibrium (Falush *et al.*, 2003). True admixture is generally asymmetrically distributed across individuals. Thus, in the absence of actual structure, the default assumption would be the equal distribution of ancestral populations across individuals (Falush *et al.*, 2003).

Results and discussion

Intraspecies homologous recombination

A total of 31 sequence types were detected in the collection of 38 *S. flavogriseus* isolates (Figure 1). Identical alleles were observed in a variety of
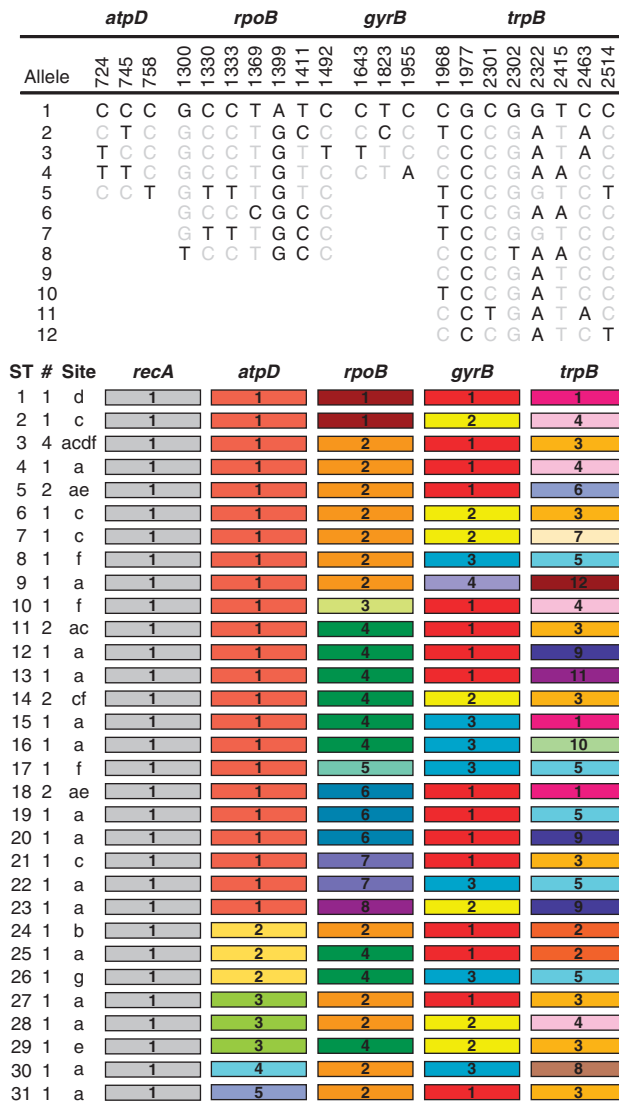
Allele	atpD			rpoB							gyrB			trpB							
	724	745	758	1300	1330	1333	1369	1399	1411	1492	1643	1823	1955	1968	1977	2301	2302	2322	2415	2463	2514
1	C	C	C	G	C	C	T	A	T	C	C	T	C	C	G	C	G	G	T	C	C
2	C	T	C	G	C	C	T	G	C	C	C	C	C	T	C	C	G	A	T	A	C
3	T	C	C	G	C	C	T	G	T	T	T	T	C	C	C	C	G	A	T	A	C
4	T	T	C	G	C	C	T	G	T	C	C	T	A	C	C	C	G	A	A	C	C
5	C	C	T	G	T	T	T	G	T	C				T	C	C	G	G	T	C	T
6				G	C	C	C	G	C	C				T	C	C	G	A	A	C	C
7				G	T	T	T	G	C	C				T	C	C	G	G	T	C	C
8				T	C	C	T	G	C	C				C	C	C	T	A	A	C	C
9														C	C	C	G	A	T	C	C
10														T	C	C	G	A	T	C	C
11														C	C	T	G	A	T	A	C
12														C	C	C	G	A	T	C	T

ST	#	Site	recA	atpD	rpoB	gyrB	trpB
1	1	d	1	1	1	1	1
2	1	c	1	1	1	2	4
3	4	acdf	1	1	2	1	3
4	1	a	1	1	2	1	4
5	2	ae	1	1	2	1	6
6	1	c	1	1	2	2	3
7	1	c	1	1	2	2	7
8	1	f	1	1	2	3	5
9	1	a	1	1	2	4	12
10	1	f	1	1	3	1	4
11	2	ac	1	1	4	1	3
12	1	a	1	1	4	1	9
13	1	a	1	1	4	1	11
14	2	cf	1	1	4	2	3
15	1	a	1	1	4	3	1
16	1	a	1	1	4	3	10
17	1	f	1	1	5	3	5
18	2	ae	1	1	6	1	1
19	1	a	1	1	6	1	5
20	1	a	1	1	6	1	9
21	1	c	1	1	7	1	3
22	1	a	1	1	7	3	5
23	1	a	1	1	8	2	9
24	1	b	1	2	2	1	2
25	1	a	1	2	3	1	2
26	1	g	1	2	4	3	5
27	1	a	1	3	2	1	3
28	1	a	1	3	2	2	4
29	1	e	1	3	4	2	3
30	1	a	1	4	2	3	8
31	1	a	1	5	2	1	3

Figure 1 Allele information for the *S. flavogriseus* phylogroup *pratensis* isolates. Each allele for each locus is represented by a unique number and color. The upper part of the figure provides information on the polymorphic sites for each allele at each locus, with nucleotide positions corresponding to the concatenated sequence alignment. ST indicates the sequence type; # indicates the number of times each ST was recovered; and site indicates the isolation source, with a and b representing sites in northern New York (in Willsboro and Chazy, respectively); c, d and e representing sites in central New York (Mitchell Street Field and Caldwell Field in Ithaca, and Harford Farm in Harford, respectively); f representing the site in Michigan and g representing *S. flavogriseus* IAF-45-CD isolated from Laval, Quebec, Canada.

sampling sites and in a wide variety of combinations that could only result from recombination. In total, 5 sequence types and 15 of the 30 alleles observed were present in two or more sites separated by >300 km (Figure 1). The standardized index of association for the population (Table 1) is one of the lowest ever calculated for a bacterial or archaeal population, indicating that the population is in almost perfect linkage equilibrium. This result could only be obtained if *S. flavogriseus* phylogroup

pratensis has a panmictic and freely recombining population structure, or if the population was recently in linkage equilibrium and has undergone an evolutionarily modern population expansion. As might be expected, given the observation of linkage equilibrium, the ratio of recombination to mutation rate, ρ/θ_w, for *S. flavogriseus* phylogroup *pratensis*, is among the highest observed for any bacterial or archaeal population (Table 1). It is important to note that the assumption of constant population size used to calculate ρ in LDhat has yet to be validated for microbial species. In addition, the lack of linkage disequilibrium also makes the calculation of an exact value for ρ impossible. Despite these limitations and pitfalls, this method has been used widely for microorganisms and is thought to allow meaningful comparison between microbial populations (for review see Perez-Losada *et al.* (2006)). In general, departure from assumptions should result in underestimation of the recombination rate and thus the value of ρ/θ_w that we provide should be treated as a lower bound. It was not possible to calculate the number of nucleotide substitutions because of recombination or mutation (r/m) with ClonalFrame (Didelot and Falush, 2007) due to the low levels of polymorphism and the high rates of recombination observed for *S. flavogriseus* phylogroup *pratensis*. As a result, an estimate of r/m was made using the single-locus variant approach ($r/m = 23.5$ for sites), but this method is known to underestimate recombination rates and thus this estimate should also be treated as a lower bound (Feil *et al.*, 1999).

Interspecies homologous recombination among *Streptomyces* species

In addition to measuring intraspecies gene exchange, we also examined gene exchange between *Streptomyces* species using an existing MLSA data set comprised of 53 different species affiliated with the *Streptomyces griseus* clade (Guo *et al.*, 2008). Although overclassification of species has been a problem for *Streptomyces* taxonomy (Anderson and Wellington, 2001), the housekeeping genes in this data set had an average nucleotide identity of 90.5% (Guo *et al.*, 2008), well below the 95% threshold that is thought to correspond with the conventional microbial species definition (Konstantinidis and Tiedje, 2005). The pairwise homeoplasy index test (Φ_w test) rejected the hypothesis of no recombination (Table 1), and the phylogenies for the *recA*, *atpD*, *rpoB*, *gyrB*, *trpB* and 16S rRNA genes (Supplementary Figure S1) were incongruent as determined by the Shimodaira–Hasegawa test (Shimodaira and Hasegawa, 1999) (Supplementary Table S2). The ratio of recombination to mutation rate (Table 1) and the ratio of nucleotide substitutions due to recombination or mutation ($r/m = 19.5$, as determined for sites by ClonalFrame) for the concatenated data set exceed the values reported for

1140

recombination within most bacterial species (Vos and Didelot, 2009), and exceed by several orders of magnitude interspecies recombination rates determined for other groups of bacteria and archaea (Eppley *et al.*, 2007; Papke *et al.*, 2007). Vos and Didelot (2009) describe r/m values calculated with ClonalFrame to be very high for a species when they exceed 10 and report an r/m of 13.6 for *Helicobacter pylori*. Owing to the number of polymorphisms in the sequences and the resulting lack of single-locus variants, it was not possible to estimate r/m using the single-locus variant method

Following these observations, we used the Recombination Detection Program (RDP 2.0) (Martin *et al.*, 2005) to document specific instances of interspecies gene exchange. A total of 13 interspecies gene transfer events were detected, affecting 40% of the lineages examined (Figure 2, Supplementary Table S3, Supplementary Figure S2). The criteria used to detect these events are insensitive to exchange of short or similar sequences, and should be viewed as a conservative lower bound on the actual number of recombination events among these taxa. ClonalFrame (Didelot and Falush, 2007) was used to explore the vertical pattern of inheritance in these sequences and the recombination events were mapped onto the resulting tree to contrast the vertical and horizontal patterns of inheritance (Figure 2). The widespread occurrence of horizontal gene exchange suggests that it is difficult to accurately depict the evolutionary history of *Streptomyces* using vertical models of inheritance. NeighborNet analysis (Bryant and Moulton, 2004) is able to depict phylogenetic signals resulting from reticulate evolutionary processes and thus this approach was also used to evaluate relationships among these taxa (Figure 3).

Given the extent of recombination, we chose to use Structure (Falush *et al.*, 2003) to further evaluate the genetic structure among these named species (Figure 3). Structure should be sensitive to admixture among the taxa that would not be detected in the previous analysis. We present a Structure model with three ancestral populations, although models with three to five populations had similar likelihood and probability of the data. It should be noted that it is difficult for Structure to infer the true number of ancestral populations for a given data set, as has been discussed by its creators (Falush *et al.*, 2003). It is likely that more than three ancestral populations are contributing to the ancestry of these taxa and that further genetic structure will be revealed by analysis of a greater diversity of *Streptomyces* and a greater diversity of loci. The purpose of this analysis was to examine patterns of admixture resulting from horizontal gene exchange, and $k = 3$ was selected because it represents the smallest value for the number of ancestral populations that captured meaningful structure in the data and was supported by the likelihood and probability of the data (as discussed in Falush *et al.*, 2003). The pattern of
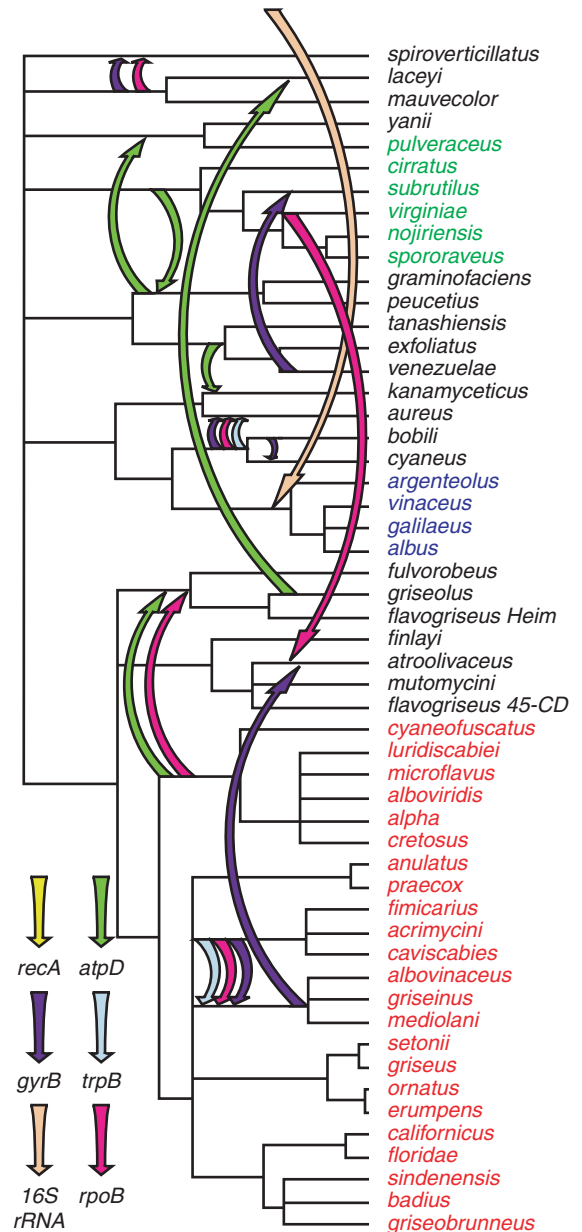

Figure 2 Evidence of widespread interspecies homologous recombination among *Streptomyces* species. Recombination events supported by multiple statistical tests (Supplementary Table S3) are mapped onto a 95% consensus ClonalFrame tree. Arrows represent recombination events, with different colors used to identify the gene or genes that were exchanged according to the legend. The direction and placement of arrows can be inferred from the events detected by RDP (Supplementary Table S3). The 16S rRNA gene transfer event originating outside the tree is inferred to be from an unknown donor (Supplementary Table S3). Species names are colored to reflect hypothetical ancestral populations assigned by Structure output in Figure 3.

ancestry generated by the Structure model was generally consistent with the recombination events depicted in Figure 2, while providing additional evidence for admixture in a range of taxa (Table 1). Two of the ancestral populations in the Structure model correspond to species clusters that have been proposed previously based on phenotypic
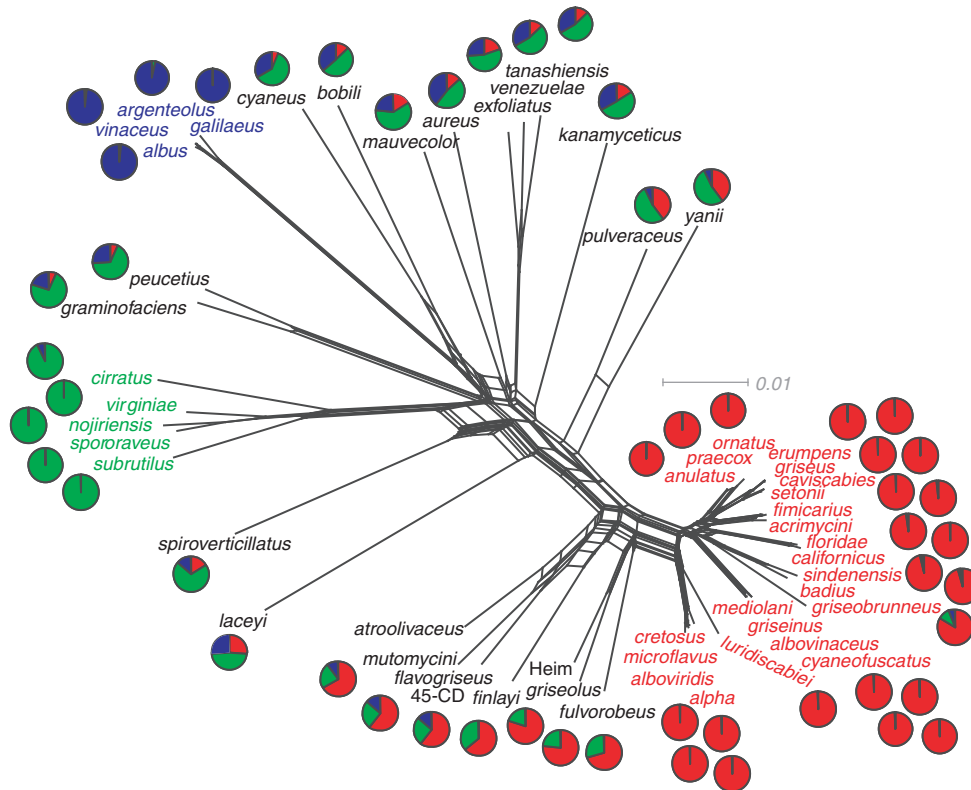
Figure 3 NeighborNet and Structure analyses of Guo *et al.* (2008) MLSA data. Boxes in the network represent uncertainty in the phylogeny and are expected if horizontal gene exchange has occurred. The scale bar is equal to 1% sequence divergence. The output from Structure analysis is indicated in the pie charts with colors indicating the proportion of ancestry estimated from each of three hypothetical ancestral populations. Two of the ancestral populations map onto species clusters that have been described previously (Kampfer *et al.*, 1991): the *S. griseus* clade (red) and the *S. lavendulae* clade (green). Heim is used for *S. flavogriseus* Heim.

properties (Kampfer *et al.*, 1991) (see Figure 3 legend for details). In several cases, lineages inferred to be admixed correspond to species for which various genetic and phenotypic markers have provided incongruent phylogenetic and taxonomic information (Kampfer *et al.*, 1991; Anderson and Wellington, 2001). For example, the species *S. atroolivaceus*, *S. finlayi*, *S. flavogriseus* and *S. griseolus* have been placed in as many as four separate species clusters on the basis of phenotypic characteristics (Kampfer *et al.*, 1991). Our results suggest that these four taxa contain housekeeping genes whose ancestry originates in the *S. griseus* and *Streptomyces lavendulae* clades, suggesting that reticulate processes have a major role in the evolution of these species.

The three ancestral populations found by Structure were also recapitulated in the NeighborNet analysis (Figure 3), with species inferred to be admixed by Structure found to occupy positions intermediate to the three ancestral populations in the NeighborNet network. Reticulation in the phylogenetic network generally supports the findings from Structure but cannot provide confirmation as reticulation in such networks may result either from genuine gene exchange or may be due to uncertainty in the phylogenetic signal captured in the data. Thus, it is possible that some of the lineages in

Figure 3 are not genuinely admixed but are inferred to be admixed due to a failure to find true structure in the data. We evaluated Structure models with 3–10 ancestral populations (data not shown) and found that the degree of admixture tended to increase asymmetrically across lineages with the number of ancestral populations included in the model, suggesting that evidence for admixture obtained with $k = 3$ is not an artifact of having too few populations in the model. Admixture remained largely absent from species in the *S. griseus* cluster regardless of the Structure model. In addition, the species *atroolivaceus*, *mutomycini*, *finlayi*, *flavogriseus*, *griseolus* and *fulvorobeus* were always composed of a mixture of two ancestral populations with a dominant contribution from the *S. griseus* ancestral population. These species clearly show asymmetrical distribution of ancestral populations (regardless of the Structure model used) with the greatest contribution from the most closely related lineage found in both the ClonalFrame tree (Figure 2) and the NeighborNet analysis (Figure 3).

Implications
The influence of horizontal gene exchange on the evolution of the *Streptomyces* seems to be profound. These data suggest that reticulation is widespread in

the *Streptomyces* phylogeny, with new lineages potentially arising as a result of hybridization between species clusters. There are reasons to suggest that the dsDNA conjugation system of streptomycetes may be associated with high levels of gene exchange among microbial populations. This conjugation system permits interspecies recombination between isolates in the laboratory (Alacevic, 1963), and can generate hybrid strains with genomes having nearly equal genetic contributions from each parent (Wang *et al.*, 1999). Such laboratory-generated hybrids are described to display new combinations of parent phenotypes, including changes in phage sensitivity (Lomovskaya *et al.*, 1977) and antibiotic production (Stoycheva *et al.*, 1994). These observations may go a long way toward explaining the reason why the taxonomy of the streptomycetes has been difficult to resolve historically. Variation in morphological, physiological and biochemical characteristics both within and between named species of the *Streptomyces* cause incongruence between phenotypic and genotypic groupings (Anderson and Wellington, 2001). Given the present data, it seems fair to hypothesize that this incongruence is due to horizontal exchange of phylogenetic markers and genes that encode particular phenotypic traits. Reticulate evolutionary processes are likely to blur the boundaries between *Streptomyces* species with gene exchange producing metaspecies that are difficult to classify. Fuzzy species boundaries and metaspecies have been indicated within other microbial groups (Hanage *et al.*, 2005; Papke *et al.*, 2007) and are well known in plants and animals (Mallet, 2008). The hypothesis of widespread gene exchange among streptomycetes would explain existing taxonomic inconsistencies among these organisms and would provide an evolutionary framework that could facilitate an understanding of their taxomony and phylogeny.

It is interesting to note that the intraspecies recombination rate within *S. flavogriseus* phylogroup *pratensis* exceeded the interspecies recombination rate by more than two orders of magnitude (Table 1). These data are consistent with the idea that recombination is acting as a cohesive force, which declines in strength with increasing sequence divergence (Fraser *et al.*, 2007). Although that may be the case, it is important to consider that the different *Streptomyces* species examined in this study were isolated from a wide range of localities and that information on their biogeography is not available. Thus, the contrast between the intraspecies and interspecies recombination rates could be a consequence of geographical or ecological barriers to gene exchange and not simply a function of sequence divergence. The stark discontinuity that we observed between rates of interspecies and intraspecies gene exchange would be expected to promote the persistence of cohesive genetic clusters. Given the high rate of interspecies gene exchange observed, we might expect the frequent introduction of foreign homologous genes into a population, with these genes frequently eliminated either by genetic drift or the cohesive effects of recombination. Introgression of foreign genes into the population or establishment of hybrid populations would be expected to result if the foreign genes have adaptive significance or as a consequence of demographic phenomena. To determine whether streptomycetes form cohesive genetic clusters that might be properly described as species, or rather represent points along a continuum of genetic exchange within the genus, will require investigation of multiple sympatric populations.

It is clear that further investigation of the population structure and biogeography of streptomycetes is needed to understand the ecological and evolutionary causes and consequences of these high rates of gene exchange. Such data should help to resolve longstanding inconsistencies in our understanding of streptomycete phylogeny and taxonomy. These issues are not merely of academic concern as streptomycetes are a preeminent source of antibiotics and bioactive compounds and progress in understanding the ecology and evolution of antibiotic production has been limited by our lack of a coherent phylogenetic framework for these organisms. In addition, an understanding of the biogeography of these organisms should lead to the development of rational sampling strategies for discovering novel genetic diversity and bioactive compounds within natural populations of streptomycetes.

Acknowledgements

References

Alacevic M. (1963). Interspecific recombination in *Streptomyces*. *Nature* **197**: 1323.

Anderson AS, Wellington EMH. (2001). The taxonomy of *Streptomyces* and related genera. *Int J Syst Evol Microbiol* **51**: 797–814.

Bryant D, Moulton V. (2004). Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* **21**: 255–265.

Chater KF, Hopwood DA, Kieser T, Thompson CJ. (1982). Gene cloning in *Streptomyces*. *Curr Top Microbiol Immunol* **96**: 69–95.

Didelot X, Falush D. (2007). Inference of bacterial microevolution using multilocus sequence data. *Genetics* **175**: 1251–1266.

Doolittle WF, Bapteste E. (2007). Pattern pluralism and the Tree of Life hypothesis. *Proc Natl Acad Sci USA* **104**: 2043–2049.

Elnakeeb MA, Lecheval HA. (1963). Selective isolation of aerobic actinomycetes. *Appl Microbiol* **11**: 75–85.

Eppley JM, Tyson GW, Getz WM, Banfield JF. (2007). Genetic exchange across a species boundary in the archaeal genus *Ferroplasma*. *Genetics* **177**: 407–416.

Falush D, Stephens M, Pritchard JK. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.

Feil EJ, Maiden MCJ, Achtman M, Spratt BG. (1999). The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. *Mol Biol Evol* **16**: 1496–1502.

Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP. (2009). The bacterial species challenge: making sense of genetic and ecological diversity. *Science* **323**: 741–746.

Fraser C, Hanage WP, Spratt BG. (2007). Recombination and the nature of bacterial speciation. *Science* **315**: 476–480.

Guo YP, Zheng W, Rong XY, Huang Y. (2008). A multilocus phylogeny of the *Streptomyces griseus* 16S rRNA gene clade: use of multilocus sequence analysis for streptomycete systematics. *Int J Syst Evol Microbiol* **58**: 149–159.

Hanage WP, Fraser C, Spratt BG. (2005). Fuzzy species among recombinogenic bacteria. *BMC Biol* **3**: 6.

Haubold B, Hudson RR. (2000). LIAN 3.0: detecting linkage disequilibrium in multilocus data. *Bioinformatics* **16**: 847–848.

Hopwood DA. (2006). Soil to genomics: the *Streptomyces* chromosome. *Annu Rev Genet* **40**: 1–23.

Hopwood DA, Lydiate DJ, Malpartida F, Wright HM. (1985). Conjugative sex plasmids of *Streptomyces*. *Basic Life Sci* **30**: 615–634.

Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF. (2008). Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320**: 1081–1085.

Huson DH, Bryant D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**: 254–267.

Ishaque M, Kluepfel D. (1980). Cellulase complex of a mesophilic streptomyces strain. *Can J Microbiol* **26**: 183–189.

Kampfer P, Kroppenstedt RM, Dott W. (1991). A numerical classification of the genera *Streptomyces* and *Streptoverticillium* using miniaturized physiological tests. *J Gen Microbiol* **137**: 1831–1891.

Kataoka M, Seki T, Yoshida T. (1991). Regulation and function of the *Streptomyces* plasmid psn22 genes involved in pock formation and inviability. *J Bacteriol* **173**: 7975–7981.

Koeppel A, Perry EB, Sikorski J, Krizanc D, Warner A, Ward DM *et al.* (2008). Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proc Natl Acad Sci USA* **105**: 2504–2509.

Konstantinidis KT, Tiedje JM. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* **102**: 2567–2572.

Lomovskaya ND, Voeykova TA, Mkrtumian NM. (1977). Construction and properties of hybrids obtained in interspecific crosses between *Streptomyces coelicolor* A3(2) and *Streptomyces griseus* Kr15. *J Gen Microbiol* **98**: 187–198.

Mallet J. (2008). Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philos Trans R Soc Lond B Biol Sci* **363**: 2971–2986.

Martin DP, Williamson C, Posada D. (2005). RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**: 260–262.

McVean G, Awadalla P, Fearnhead P. (2002). A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**: 1231–1241.

Papke RT, Zhaxybayeva O, Feil EJ, Sommerfeld K, Muise D, Doolittle WF. (2007). Searching for species in haloarchaea. *Proc Natl Acad Sci USA* **104**: 14092–14097.

Perez-Losada M, Browne EB, Madsen A, Wirth T, Viscidi RP, Crandall KA. (2006). Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. *Infect Genet Evol* **6**: 97–112.

Sheppard SK, McCarthy ND, Falush D, Maiden MCJ. (2008). Convergence of *Campylobacter* species: implications for bacterial evolution. *Science* **320**: 237–239.

Shimodaira H, Hasegawa M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* **16**: 1114–1116.

Shirling EB, Gottlieb D. (1970). Report of the International Streptomyces Project 5 Years of Collaborative Research. Prauser H (Ed). *The Actinomycetales. The Jena International Symposium on Taxonomy*, 439pIllus Veb Gustav Fischer Verlag: Jena E, Germany, 79–89.

Stoycheva Z, Todorov T, Peltekova V. (1994). Intergeneric crosses between *Streptomyces ambofaciens* and *Saccharopolyspora erythraea*. *Folia Microbiologica* **39**: 13–18.

Vos M, Didelot X. (2009). A comparison of homologous recombination rates in bacteria and archaea. *ISME J* **3**: 199–208.

Wang SJ, Chang HM, Lin YS, Huang CH, Chen CW. (1999). Streptomyces genomes: circular genetic maps from the linear chromosomes. *Microbiology* **145**: 2209–2220.