Copyright © 2008 by the Genetics Society of America
DOI: 10.1534/genetics.107.085621

The Effect of Chromosome Geometry on Genetic Diversity

Pradeep Reddy Marri,*[,1] Leigh K. Harris,* Kathryn Houmiel,[†,‡] Steven C. Slater[†] and Howard Ochman*

**Department of Biochemistry and Molecular Biophysics, University of Arizona, Tucson, Arizona 85721, [†]Department of Applied Biological Sciences and The Biodesign Institute, Arizona State University, Mesa, Arizona 85212 and [‡]Department of Biology, Seattle Pacific University, Seattle, Washington 98119*

Manuscript received December 10, 2007
Accepted for publication February 19, 2008

ABSTRACT

Although organisms with linear chromosomes must solve the problem of fully replicating their chromosome ends, this chromosome configuration has emerged repeatedly during bacterial evolution and is evident in three divergent bacterial phyla. The benefit usually ascribed to this topology is the ability to boost genetic variation through increased recombination. But because numerous processes can impact linkage disequilibrium, such an effect is difficult to assess by comparing across bacterial taxa that possess different chromosome topologies. To test directly the contribution of chromosome architecture to genetic diversity and recombination, we examined sequence variation in strains of Agrobacterium Biovar 1, which are unique among sequenced bacteria in having both a circular and a linear chromosome. Whereas the allelic diversity among strains is generated principally by mutations, intragenic recombination is higher within genes situated on the circular chromosome. In contrast, recombination between genes is, on average, higher on the linear chromosome, but it occurs at the same rate as that observed between genes mapping to the distal portion of the circular chromosome. Collectively, our findings indicate that chromosome topology does not contribute significantly to either allelic or genotypic diversity and that the evolution of linear chromosomes is not based on a facility to recombine.

THE vast majority of bacterial genomes contain a single circular chromosome, often accompanied by one or more small circular plasmids (Casjens 1998; Tettelin *et al.* 2005; Canchaya *et al.* 2006; Makarova *et al.* 2006). That this chromosome structure is prevalent among bacteria suggests that it is both ancestral to cellular life forms and the most expedient to maintain in that it precludes the need for mechanisms that complete the replication of chromosome ends. With the advent of electrophoretic methods that could resolve large DNA fragments, the chromosome of the spirochete *Borrelia burgdorferi* was found to be linear (Ferdows and Barbour 1989), and it was subsequently shown that the genomes of bacteria from other phyla, notably some members of the Actinomycetes (Lin *et al.* 1993) and Alphaproteobacteria (Goodner *et al.* 1999), also harbor linear chromosomes.

The fact that linear chromosomes have originated multiple times in bacteria, and are ubiquitous among eukaryotes, suggests that this configuration, once established, might confer some advantage. However, when the normally circular chromosome of *Escherichia coli* was linearized synthetically, there were no obvious changes in growth rate, gene expression, or cell morphology (Cui *et al.* 2007). Similarly, when the linear chromosome of *Streptomyces lividans* is made circular (either artificially or spontaneously), cells of either chromosome configuration display the same phenotype and the same high levels of genetic instability (Lin and Chen 1997; Volff *et al.* 1997).

Although chromosome geometry has, so far, been shown to have negligible effects on cell fitness, it has been suggested that the advantage of linear chromosomes resides in an enhanced ability to recombine both because linearity can facilitate chromosome pairing and because gene exchange requires only a single recombination event (Volff and Altenbuchner 2000). But determining if linear chromosomes realize higher levels of gene exchange is not simply a matter of comparing the amount of recombination in species with linear *vs.* those with circular chromosomes. Numerous factors contribute to recombination frequencies, including genetic mechanisms that promote or suppress recombination (Matic *et al.* 1996; Redfield 2001) as well as external forces, such as natural selection, population substructure, and genetic drift (Smith *et al.* 1993). Even among bacterial species that have a single circular chromosome, some are strictly clonal (*i.e.*, not recombining) whereas others are virtually panmictic (*i.e.*, freely recombining) (Selander and Musser 1990; Feil *et al.*

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. EU334170–EU334349.

[1]*Corresponding author:* BIO5 Institute, Thomas W. Keating Bldg., 1657 E. Helen St., University of Arizona, Tucson, AZ 85721.
E-mail: pradeepm@email.arizona.edu

2001; Narra and Ochman 2006; Pérez-Losada *et al.* 2007).

In this study, we analyzed the genetic variation and linkage disequilibrium among loci in *Agrobacterium tumefaciens*, a bacterial species whose genomic organization is unique in having both a linear and a circular chromosome (Allardet-Servent *et al.* 1993). Because both chromosomes coexist in the same cellular lineage and experience the same molecular and population-level processes that can affect recombination rates, it is possible to directly evaluate the contribution of chromosome geometry to genetic diversity and recombination. In this study, we apply a multilocus sequencing approach to assess the degree of recombination within and among loci situated on each of the two chromosomes to determine whether linear chromosomes manifest higher rates of genetic diversity and exchange.

MATERIALS AND METHODS

Strain selection and DNA isolation: A total of 18 strains of *A. tumefaciens* and 1 each of *Agrobacterium rhizogenes* and *Agrobacterium radiobacter* were subjected to multilocus sequence analysis (supplemental Table 1). These 20 strains are closely related on the basis of 16S rRNA sequences and are assigned to Agrobacterium Biovar I (Sawada *et al.* 1993; Farrand *et al.* 2003; Young *et al.* 2003). Within the Agrobacteria, linear chromosomes are confined to Biovar I, and the presence of a circular and a linear chromosome in these strains has been ascertained by the presence of protelomerase and/or by pulsed-field gel electrophoresis (S. C. Slater, K. Houmiel and B. Goodner, unpublished results). Genomic DNAs were purified from 0.5 ml of overnight cultures of Agrobacterium using the QIAGEN (Valencia, CA) DNeasy kit per manufacturer's instructions.

Amplification and sequencing of loci: Nine genes, six situated on the circular chromosome and three on the linear chromosome, were selected on the basis of their (i) positions and spacing along the chromosomes, (ii) presence in related species, and/or (iii) use in other MLST analyses (Figure 1). The genes assayed also include two pairs of paralogs (*lysA*-I and *lysA*-II and *acsA*-I and *acsA*-II), each with one copy on the circular chromosome and the other on the linear chromosome, respectively.

Oligonucleotide primers were designed for the conserved regions of each gene on the basis of alignments of the homologous sequences from *A. tumefaciens* C58, *Agrobacterium vitis* S4, *A. radiobacter* K84, *Rhizobium leguminosarum*, and *Rhizobium elti* (supplemental Table 2). PCR amplifications were performed using 5 ng of genomic DNA as template under the following conditions: initial denaturation at 94° for 2 min followed by 35 cycles of 94° for 1 min, annealing at an empirically determined temperature for each primer pair for 1 min, and extension at 72° for 1.5 min, followed by a 6-min final extension at 72°. Touchdown and gradient PCR were used to recover the appropriately sized gene fragment for templates that did not yield a single amplification product under the initial PCR conditions. DNA sequencing was performed on both strands of the amplified products using an ABI Prism DNA sequencer, and the chromatograms were manually checked and curated with Sequencher (Gene Codes).

Phylogenetic analyses: Sequences were aligned in Clustal W (Thompson *et al.* 1994), and alignments edited with BioEdit v7.0.9.0 (Hall 1999). The evolutionary model that best fits the alignment for each gene was selected using Modeltest v3.06 (Posada and Crandall 1998), and a maximum-likelihood tree was generated in PAUP (Swofford 1993). The following models were selected for each of the genes: *pgi* (HKY + γ), *rpoD* (TrN + γ), *thrB* (HKY + γ), *glyA* (TrN + γ), *lysA-II* (HKY + γ), *acsA*-II (TrN + γ), *lysA*-I (TrN + γ), *acsA*-I (TrN + γ), and *ftsZ* (TrN + γ). Bootstrap support was obtained using 100 replicates, each based on 10 random replicates with a heuristic search. Congruence in the branching orders of each gene tree was tested against one another by applying the Shimodaira–Hasegawa (SH) test (Shimodaira and Hasegawa 1999).

Gene and population-level analyses: Analyses of the levels of nucleotide diversity and the relative contributions of mutation, selection, and recombination to sequence diversity were performed with DnaSp (Rozas and Rozas 1999), RDP (Martin and Rybicki 2000), START (Jolley *et al.* 2001), and software available on the MLST website (http://www.mlst.net). Haplotypes were assigned using the nonredundant databases. The pairwise nucleotide diversity (π), Watterson's θ, Tajima's D, and K_a/K_s values were calculated using DnaSp. The homoplasy ratio test (Smith and Smith 1998) for detection of recombination was done using START.

Detecting recombination: Additional analyses were carried out to quantify the extent of intra- and intergenic recombination within and among loci. Measures of linkage disequilibrium (or recombination) were obtained by using RDP, MaxChi, Chimera, and Geneconv (Posada and Crandall 2001), all of which implemented in RDP (Martin and Rybicki 2000) with a *P*-value cutoff set at 0.01 and Bonferroni corrections applied for multiple tests. To further investigate the linkage relationships among loci, we calculated the intergenic values of rho (ρ) on concatenated alignments of each gene pair and of all genes using LDhat implemented in RDP3 (http://darwin.uvigo.es/rdp/rdp.html).

RESULTS

Chromosome topology can potentially affect genetic diversity in a variety of ways. Recombination could act on a very local scale by influencing allelic diversity at a locus through intragenic recombination and/or, more broadly, by increasing haplotype diversity by reassorting alleles over loci. Moreover, assessments of recombination can be confounded by differences in nucleotide substitution rates among loci, making it necessary to partition several components of the observed variation.

Levels of polymorphism and nucleotide diversity: We sequenced an average of 525 bp from nine genes (six on the circular chromosome, three on the linear chromosome) in each of 20 Biovar 1 strains of Agrobacterium. The numbers of alleles ranged from 10 to 14 per locus, and there is no difference in the average number of alleles per locus for genes on the linear or circular chromosomes (Table 1). Furthermore, there were no significant differences in estimates of pairwise nucleotide diversity (π) of genes located on the linear and circular chromosomes; however, the values on the linear chromosome are slightly higher overall (0.07 *vs.* 0.05). Similarly, there were minor differences in estimates of the population mutation rate (θ) between loci on the linear and circular chromosomes, indicating that genes on the linear chromosome are evolving slightly faster.

TABLE 1

Genic diversity in Biovar 1 strains of Agrobacterium

Gene	Length amplified	Alleles	Π			θ per site	*r/m* per site	K_a/K_s	Tajima's *D*
			Per site	Synonymous sites	Nonsynonymous sites				
				Linear					
acsA-II	551	13	0.09	0.23	0.040	0.077	0.06	0.122	−0.26
ftsZ	568	13	0.07	0.25	0.006	0.060	0.12	0.016	−0.12
lysA-II	539	10	0.05	0.18	0.008	0.045	0.09	0.034	0.13
Average			0.07	0.22	0.018	0.061	0.09	0.058	−0.08
				Circular					
acsA-I	583	11	0.06	0.21	0.006	0.050	0.14	0.024	0.08
glyA	536	13	0.03	0.12	0.001	0.030	0.07	0.009	−0.70
lysA-I	470	11	0.07	0.20	0.030	0.070	0.14	0.110	−0.61
pgi	501	11	0.06	0.20	0.010	0.055	0.07	0.050	−0.22
rpoD	523	13	0.04	0.17	0.003	0.035	0.11	0.015	0.06
thrB	494	14	0.06	0.23	0.009	0.050	0.20	0.030	0.15
Average			0.05	0.19	0.012	0.048	0.12	0.040	−0.21

Π, pairwise nucleotide diversity, θ, Watterson's θ; *r/m*, ratio of recombination rate to mutation rate; K_a, no. of nonsynonymous substitutions per nonsynonymous sites; K_s, no. of synonymous substitutions per synonymous sites.

Because proteins evolve at highly different rates, it is not possible to determine if chromosome topology *per se* contributes directly to these differences in nucleotide diversity or whether less constrained loci happen to be situated on a particular chromosome (as is the case with accessory or supernumerary elements). In an attempt to control for this factor, we compared genes with paralogous counterparts on each chromosome with the thought that both copies might be under similar functional constraints and exhibit similar levels of nonsynonymous site diversity. Opposite trends were observed for each of the two pairs of paralogs that we examined: for *acsA*, the paralog situated on the linear chromosome displayed much higher nucleotide diversity at nonsynonymous sites, whereas for *lysA*, the copy on the linear chromosome had much lower nonsynonymous site diversity (Table 1).

We performed two tests of selection on these genes by computing both Tajima's *D*, which looks at disparities between π and θ and K_a/K_s ratios, which measure relative divergence at nonsynonymous and synonymous sites. No values of Tajima's *D* differed significantly from zero, indicating that the observed variation is selectively neutral, and the K_a/K_s tests recapitulated the findings on π-nonsynonymous values, as would be expected if silent substitution rates are similar across loci.

Role of recombination in producing allelic diversity: Allelic diversity can be generated both by mutation and by intragenic recombination, and several methods have been developed to estimate the relative contributions of each source to bacterial sequence polymorphisms (Guttman and Dykhuizen 1994; Feil *et al.* 1999, 2000). At all loci examined, the vast majority of polymorphisms arise through mutations, and averaging across all genes, new polymorphic nucleotides are introduced ~10 times more frequently by mutations than by recombination. On average, the contribution of recombination to nucleotide diversity is ~30% higher for genes on the circular chromosome (Table 1), although there is some degree of overlap of the ratio of recombination rate to mutation rate (*r/m*) values of genes on each of the two chromosomes.

We also applied the homoplasy ratio test, which compares the number of homoplasies in a phylogenetic tree to that expected by repeated mutations in a strictly clonal population. When applied to the nine loci, the homoplasy ratios were close to zero for most loci, suggesting a clonal population structure with very limited or no recombination. The recombination detection program RDP detected only a single event of recombination involving a 200-bp recombinant region in the *lysA*-I gene between the strains *A. radiobacter* K299 and *A. tumefaciens* KU12.

Cumulatively, these analyses suggest that recombination has played a minor role in producing allelic diversity in Biovar 1 strains of Agrobacterium. To quantify the extent of intragenic recombination in each gene, we calculated the association coefficient ρ, which could then be compared within and across loci as well as to genes in other species. Intragenic values of ρ are low, ranging from 1.28 for *glyA* to 7.46 for *lysA*-I, and the average ρ of ~3.5 is among the lowest observed for a bacterial species (Figure 1A). Despite this low level of recombination, but in accord with the situation observed for the *r/m* ratios, intragenic recombination was slightly higher, on average, in genes on the circular chromosome (3.76 in circular *vs.* 3.19 in linear); among paralogous pairs (*lysA*-I and *lysA*-II and *acsA*-I and *acsA*-II),

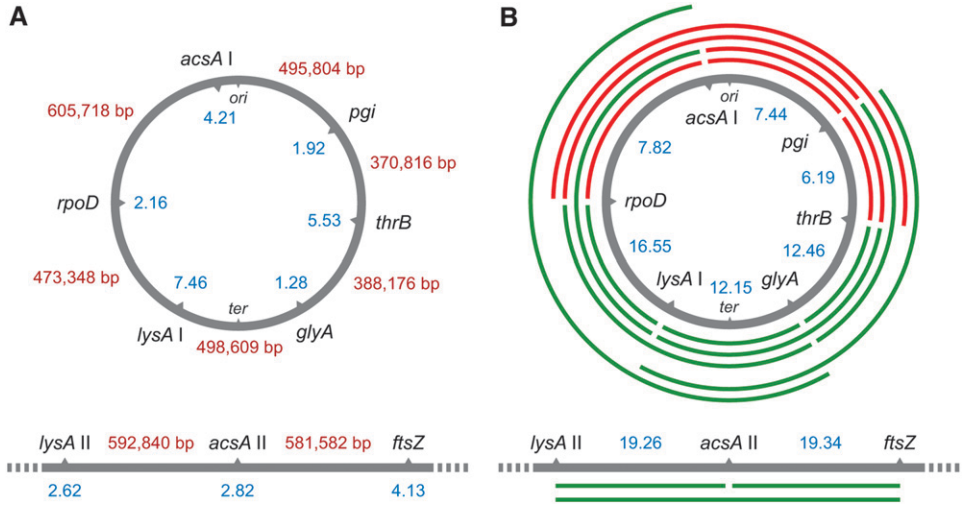

FIGURE 1.—(A) Chromosome assignment and map positions of the loci examined in Agrobacterium Biovar 1. Values in red are the physical distances in nucleotides between adjacent loci, and numbers in blue are the intragenic ρ-values for the corresponding genes. (B) Measuring intergenic recombination among loci. Numbers in blue are the intergenic ρ-values for pairs of adjacent loci, and arcs denote whether intergenic recombination was (green) or was not (red) detected between the corresponding pair of loci connected by the arc.

gene copies on the circular chromosome each had higher ρ-values than their counterparts on the linear chromosome.

Linkage relationships among loci: To detect possible cases of intergenic recombination, we first searched for inconsistencies among the maximum-likelihood phylogenies generated for each of the nine genes (supplemental Figure 1). The phylogenies of the three assayed genes on the linear chromosome, and of *acsA*-I, *rpoD*, and *thrB* on the circular chromosome, all yielded two major clades, each containing the same constituent strains but showing variation in the relationships among strains within each clade. The remaining three genes on the circular chromosome, *pgi*, *lysA*-I, and *glyA*, do not separate the strains into these same distinct groups. We applied the SH test to ascertain if any genes yielded significantly different tree topologies. For the 36 SH tests used to compare all pairwise combinations of genes, only seven pairs yielded trees that did not differ significantly (supplemental Table 1); among these, only two cases involved adjacent loci, suggesting that, in addition to the low background levels of intragenic recombination detected by other methods, some larger-scale events have occurred.

The extent of recombination among all pairs of genes was further evaluated by concatenating the two corresponding sequences and calculating the intergenic ρ-values. In addition, we performed a test using RDP3 that searched for breakpoints in the linkage relationships of the polymorphic sites in these concatenated alignments. The results from these analyses showed that, on average, recombination is more frequent among loci on the linear chromosome but that it occurs at about the frequency observed for loci situated closer to the replication terminus of the circular chromosome. Intergenic ρ-values for pairs of adjacent loci average 19.30 ($n = 2$) on the linear chromosome and 10.44 ($n = 6$) on the circular chromosome, but after scaling these values to a uniform interval length, the effect of chromosome position becomes apparent (ρ circular-proximate/100 kb = 1.49; ρ circular-distal/100 kb = 3.05; ρ linear/100 kb = 3.28) (Figure 1B). The intergenic ρ-values are similar for genes on the linear chromosome and those positioned in the distal position of the circular chromosome, and the values for genes in each of these regions differ significantly from those obtained for genes more proximate to the replication origin of the circular chromosome.

These results are also evident in the breakpoint analysis (red and green lines in Figure 1B). Those genes closest to the replication origin (*acsA*-I and *pgi*) display no evidence of recombination with adjacent loci; more distal genes (*rpoD* and *thrB*) found to have recombined with genes that are closer to the terminus but not those closer to the origin, and the genes closest to the replication terminus (*lysA*-I and *glyA*), recombined with all adjacent loci. All of the relationships among loci tested in this analysis were internally consistent, such that if genes 1 and 2 showed no evidence of intergenic recombination but genes 2 and 3 did, recombination was detected between genes 1 and 3.

DISCUSSION

Although organisms with linear chromosomes must invent mechanisms to overcome the end-replication problem, this chromosome structure has arisen independently at least three times over the course of bacterial evolution. The discovery that diverse bacterial species maintain linear chromosomes has prompted several speculations about the potential advantages conferred by this chromosome configuration. Such explanations have typically focused on the observed (or presumed) properties of the linear chromosomes, such as the instability of their chromosome ends or the facility with which they can recombine and/or integrate alien DNA.

The extent to which chromosome structure actually contributes to the observed levels of variation within

a species has been difficult to assess because numerous intrinsic and extrinsic factors will also impact genetic diversity. But because genomic and population-level processes will affect all chromosomes within the cell in a similar manner, it is possible to remove most of the potentially confounding factors, and thus directly evaluate the effects of chromosome geometry on genetic variation, by analyzing strains of Agrobacterium Biovar 1, whose genomes harbor both a linear and a circular chromosome.

Applying a multilocus sequence typing approach, the levels of genetic variation were not higher on the linear chromosome, and contrary to expectations, recombination has made a larger contribution to the allele diversity observed in loci situated on the circular chromosome. Despite this difference, the overall effect of intragenic recombination is slight: the *r/m* values for these nine genes in *A. tumefaciens* Biovar 1 are among the lowest detected in any bacterial species, but it remains to be seen whether the other Agrobacterium biovars (2 and 3), which do not have this chromosome organization, display similar levels and patterns of allelic diversity. Moreover, the average intragenic ρ in Agrobacterium Biovar 1 is among the lowest for any bacteria species, on the order of that observed in *Burkholderia pseudomallei*, *Enterococcus faecium*, and *Campylobacter jejuni* (Pérez-Losada *et al.* 2006).

With respect to other components of genic diversity, there are no systematic differences between loci on the linear and circular chromosomes. The ranges of values of per-site diversity and substitution rates overlap substantially for genes on each chromosome. Nonsynonymous substitution rates and K_a/K_s ratios are slightly, but not significantly, higher for genes on the linear chromosome, but there is no trend toward moving less essential loci to the linear chromosome. In fact, when examining paralogous genes, which are usually assumed to encode proteins of similar function, in one pair (*acsA*-I and *acsA*-II), the copy on the linear chromosome has a higher K_a/K_s ratio whereas in the other pair (*lysA*-I and *lysA*-II) the opposite is observed, indicating that the differences in evolutionary rates are not related to the particular chromosome on which a gene is situated.

Linear bacterial chromosomes have been previously shown to facilitate recombination leading to the high turnover of genes in the terminal regions of the linear chromosome (Choulet *et al.* 2006a,b). Whereas high rates of gene gain and loss might serve some evolutionary advantage, this type of variation is not limited to bacteria with linear chromosomes. Genome size and gene content can be highly variable among closely related strains of bacteria possessing a single circular chromosome (*e.g.*, Normand *et al.* 2007), and gene acquisition events are also known to be higher near the replication terminus of circular chromosomes (Lawrence and Ochman 1998).

Of the diversity indices that we measured, only the average rate of intergenic recombination differed between the circular and linear chromosomes of Agrobacterium Biovar 1 strains. However, while there is, on average, more recombination between genes on the linear chromosome, the level is no higher than that observed for loci situated closer to the replication terminus of the circular chromosome (Figure 1). Thus, the *raison d'être* for linear bacterial chromosomes is not in an enhanced ability for homologous exchange; in fact, any benefits to increased diversity would be incurred by genes nearer the replication terminus of circular bacterial chromosomes, which experience higher rates of both recombination (Louarn *et al.* 1994) and substitutions (Sharp *et al.* 1989; Mira and Ochman 2002).

We thank Becky Nankivell for preparing the figures. This work was supported by National Institutes of Health grant GM56120 to H.O. and National Science Foundation grant 0333297 to S.C.S. L.K.H. was supported, in part, by a National Science Foundation predoctoral fellowship.

LITERATURE CITED

Allardet-Servent, A., S. Michaux-Charachon, E. Jumas-Bilak, L. Karayan and M. Ramuz, 1993 Presence of one linear and one circular chromosome in the *Agrobacterium tumefaciens* C58 genome. J. Bacteriol. **175:** 7869–7874.

Canchaya, C., M. J. Claesson, G. F. Fitzgerald, D. Van Sinderen and P. W. O'Toole, 2006 Diversity of the genus *Lactobacillus* revealed by comparative genomics of five species. Microbiology **152:** 3185–3196.

Casjens, S., 1998 The diverse and dynamic structure of bacterial genomes. Annu. Rev. Genet. **32:** 339–377.

Choulet, F., A. Gallois, B. Aigle, S. Mangenot, C. Gerbaud *et al.*, 2006a Intraspecific variability of the terminal inverted repeats of the linear chromosome of *Streptomyces ambofaceins*. J. Bacteriol. **188:** 6599–6610.

Choulet, F., B. Aigle, A. Gallois, S. Mangenot, C. Gerbaud *et al.*, 2006b Evolution of the terminal regions of *Streptomyces* linear chromosome. Mol. Biol. Evol. **23:** 2361–2369.

Cui, T., N. Moro-Oka, K. Ohsumi, K. Kodama, T. Ohshima *et al.*, 2007 *Escherichia coli* with a linear genome. EMBO Rep. **8:** 181–187.

Farrand, S. K., P. B. Van Berkum and P. Oger, 2003 *Agrobacterium* is a definable genus of the family Rhizobiaceae. Int. J. Syst. Evol. Microbiol. **53:** 1681–1687.

Feil, E. J., M. C. Maiden, M. Achtman and B. G. Spratt, 1999 The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. Mol. Biol. Evol. **16:** 1496–1502.

Feil, E. J., J. M. Smith, M. C. Enright and B. G. Spratt, 2000 Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. Genetics **154:** 1439–1450.

Feil, E. J., E. C. Holmes, D. E. Bessen, M. S. Chan, N. P. Day *et al.*, 2001 Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. Proc. Natl. Acad. Sci. USA **98:** 182–187.

Ferdows, M. S., and A. G. Barbour, 1989 Megabase-sized linear DNA in the bacterium *Borrelia burgdorferi*, the Lyme disease agent. Proc. Natl. Acad. Sci. USA **86:** 5969–5973.

Goodner, B. W., B. P. Markelz, M. C. Flanagan, C. B. Crowell, Jr., J. L. Racette *et al.*, 1999 Combined genetic and physical map of the complex genome of *Agrobacterium tumefaciens*. J. Bacteriol. **181:** 5160–5166.

Guttman, D. S., and D. E. Dykhuizen, 1994 Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. Science **266:** 1380–1383.

Hall, T. A., 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. **41:** 95–98.

Jolley, K. A., E. J. Feil, M. S. Chan and M. C. J. Maiden, 2001 Sequence type analysis and recombinational tests (START). Bioinformatics **17:** 1230–1231.

Lawrence, J. G., and H. Ochman, 1998 Molecular archaeology of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. USA **95:** 9413–9417.

Lin, Y. S., and C. W. Chen, 1997 Instability of artificially circularized chromosomes of *Streptomyces lividans*. Mol. Microbiol. **26:** 709–719.

Lin, Y. S., H. M. Kieser, D. A. Hopwood and C. W. Chen, 1993 The chromosomal DNA of *Streptomyces lividans* 66 is linear. Mol. Microbiol. **10:** 923–933.

Louarn, J., F. Cornet, V. François, J. Patte and J. M. Louarn, 1994 Hyperrecombination in the terminus region of the *Escherichia coli* chromosome: possible relation to nucleoid organization. J. Bacteriol. **176:** 7524–7531.

Makarova, K., A. Slesarev, Y. Wolf, A. Sorokin, E. Koonin *et al.*, 2006 Comparative genomics of the lactic acid bacteria. Proc. Natl. Acad. Sci. USA **103:** 15611–15616.

Martin, D., and E. Rybicki, 2000 RDP: detection of recombination amongst aligned sequences. Bioinformatics **16:** 562–563.

Matic, I., F. Taddei and M. Radman, 1996 Genetic barriers among bacteria. Trends Microbiol. **4:** 69–72.

Mira, A., and H. Ochman, 2002 Gene location and bacterial sequence divergence. Mol. Biol. Evol. **19:** 1350–1358.

Narra, H. P., and H. Ochman, 2006 Of what use is sex to bacteria? Curr. Biol. **16:** R705–R710.

Normand, P., P. Lapierre, L. S. Tisa, J. P. Gogarten, N. Alloisio *et al.*, 2007 Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. Genome Res. **17:** 7–15.

Pérez-Losada, M., M. L. Porter, L. Tazi and K. A. Crandall, 2007 New methods for inferring population dynamics from microbial sequences. Infect. Genet. Evol. **7:** 24–43.

Posada, D., and K. A. Crandall, 1998 MODELTEST: testing the model of DNA substitution. Bioinformatics **14:** 817–818.

Posada, D., and K. A. Crandall, 2001 Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. USA **98:** 13757–13762.

Redfield, R. J., 2001 Do bacteria have sex? Nat. Rev. Genet **2:** 634–639.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Sawada, H., H. Ieki, H. Oyaizu and S. Matsumoto, 1993 Proposal for rejection of *Agrobacterium tumefaciens* and revised descriptions for the genus *Agrobacterium* and for *Agrobacterium radiobacter* and *Agrobacterium rhizogenes*. Int. J. Syst. Bacteriol. **43:** 4694–4702.

Selander, R. K., and J. M. Musser, 1990 Population genetics of bacterial pathogenesis, pp. 11–36 in *The Evolution of Bacterial Pathogens*, Vol. 11. Academic Press, New York.

Sharp, P. M., D. C. Shields, K. H. Wolfe and W. H. Li, 1989 Chromosomal location and evolutionary rate variation in enterobacterial genes. Science **246:** 808–810.

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

Smith, J. M., and N. N. Smith, 1998 Detecting recombination from gene trees. Mol. Biol. Evol. **15:** 590–599.

Smith, J. M., N. H. Smith, M. O'Rourke and B. G. Spratt, 1993 How clonal are bacteria? Proc. Natl. Acad. Sci. USA **90:** 4384–4388.

Swofford, D. L., 1993 *PAUP: Phylogenetic Analysis Using Parsimony*, Version 3.1. Illinois Natural History Survey, University of Illinois, Champaign, IL.

Tettelin, H., V. Masignani, M. J. Cieslewicz, C. Donati, D. Medini *et al.*, 2005 Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for microbial "pan-genome." Proc. Natl. Acad. Sci. USA **102:** 13950–13955.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

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.

Volff, J. N., P. Viell and 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:** 753–760.

Young, J. M., L. D. Kuykendall, E. Martínez-Romero, A. Kerr and H. Sawada, 2003 Classification and nomenclature of *Agrobacterium* and *Rhizobium*. Int. J. Syst. Evol. Microbiol. **53:** 1689–1695.

Communicating editor: J. Lawrence