Complex prokaryotic genome structure: rapid evolution of chromosome II

Anish Bavishi, Ankur Abhishek, Lin Lin, and Madhusudan Choudhary

Abstract: Although many bacteria with two chromosomes have been sequenced, the roles of such complex genome structuring are still unclear. To uncover levels of chromosome I (CI) and chromosome II (CII) sequence divergence, Mauve 2.2.0 was used to align the CI- and CII-specific sequences of bacteria with complex genome structuring in two sets of comparisons: the first set was conducted among the CI and CII of bacterial strains of the same species, while the second set was conducted among the CI and CII of species in *Alphaproteobacteria* that possess two chromosomes. The analyses revealed a rapid evolution of CII-specific DNA sequences compared with CI-specific sequences in a majority of organisms. In addition, levels of protein divergence between CI-specific and CII-specific genes were determined using phylogenetic analyses and confirmed the DNA alignment findings. Analysis of synonymous and nonsynonymous substitutions revealed that the structural and functional constraints on CI and CII genes are not significantly different. Also, horizontal gene transfer estimates in selected organisms demonstrated that CII in many species has acquired higher levels of horizontally transferred segments than CI. In summary, rapid evolution of CII may perform particular roles for organisms such as aiding in adapting to specialized niches.

Key words: multiple chromosomes, genome analysis, horizontal gene transfer.

Résumé : Bien que les génomes de plusieurs bactéries dotées de deux chromosomes aient été séquencés, les rôles d'une telle structuration complexe du génome restent nébuleux. Afin de déterminer le degré de divergence entre séquences CI et CII, le logiciel Mauve 2.2.0 a été employé pour aligner les séquences spécifiques de CI et CII chez des bactéries aux génomes complexes dans le cadre de deux séries de comparaisons. La première comparaison a été réalisée entre les souches CI et CII de la même espèce, tandis que la seconde comparaison a été faite au sein des espèces CI et CII parmi les alphaprotéobactéries qui possèdent deux chromosomes. Ces analyses ont révélé une évolution rapide des séquences d'ADN spécifiques à CII par rapport à l'évolution observée chez les séquences spécifiques de CI chez une majorité d'espèces. De plus, les degrés de divergence au niveau des séquences peptidiques entre les gènes spécifiques de CI ou de CII ont été examinés par analyse phylogénétique et ont confirmé les observations faites au niveau de l'ADN. L'analyse des substitutions synonymes et non-synonymes a révélé que les contraintes structurales et fonctionnelles sur les gènes CI et CII ne sont pas significativement différentes. Aussi, une évaluation de la fréquence des transferts horizontaux chez des espèces choisies a démontré que, chez plusieurs espèces, CII aurait acquis davantage de segments obtenus suite à un transfert horizontal que CI. En résumé, l'évolution rapide de CII pourrait jouer certains rôles pour les organismes comme aider à s'adapter à des niches spécialisées.

Mots-clés : chromosomes multiples, analyse génomique, transfert horizontal de gènes.

[Traduit par la Rédaction]

Introduction

For years, a bacterium was thought to have a single circular chromosome that contained much, if not all, of its essential genetic material. However, since the discovery of two chromosomes in *Rhodobacter sphaeroides* (Suwanto and Kaplan 1989, 1992), this dogma of the prokaryotic genome structure has been fundamentally revised. Since this finding, many more prokaryotic organisms have been reported to have more than one chromosome, some of which include *Agrobacterium tumefaciens* (Allardet-Servent et al. 1993), *Brucella melitensis* (Michaux et al. 1993), *Burkholderia cepacia* (Rodley et al. 1995; Lessie et al. 1996), *Leptospira interrogans* (Zuerner et al. 1993), *Vibrio cholerae* (Trucksis et al. 1998), and *Vibrio parahaemolyticus* (Yamaichi et al. 1999). The complex prokaryotic genome structure is no longer an exception but rather an accepted paradigm of genome variation among prokaryotes.

Although what specifically constitutes an additional chromosome in bacteria is debatable, a chromosome must contain essential genes necessary for the functioning of the organism (Egan et al. 2005). For bacteria with two chromosomes, there is usually one larger primary chromosome (CI) that houses a great deal of the genetic information necessary for the fundamental functioning of the organism and one smaller chromosome (CII) that not only contains a number of essential genes but also encodes a greater percentage of hypothetical proteins compared with CI (Egan et al. 2005). In several organisms in the genus *Vibrio*, the larger chromo-

Received 7 February 2010. Accepted 21 May 2010. Published on the NRC Research Press Web site at genome.nrc.ca on .

Corresponding Editor: B. Golding.

A. Bavishi, A. Abhishek, L. Lin, and M. Choudhary.[1]
Department of Biological Sciences, Sam Houston State University, Huntsville, TX 77341, USA.

[1]Corresponding author (e-mail: mchoudhary@shsu.edu).

some retains many essential genes while the smaller chromosome remains more flexible, encompassing both essential and nonessential genes (Okada et al. 2005).

Studies indicate that CI and CII may perform different, specialized roles in certain organisms. For instance, when *V. cholerae* was grown in aerobic laboratory conditions and in rabbit intestine, many CI genes were expressed in both cases but significantly more CII genes were expressed in the rabbit intestine than in laboratory conditions (Xu et al. 2003). Similarly, even though it has been determined that both chromosomes in *R. sphaeroides* contain vital genetic information necessary for the functioning of the organism (Choudhary et al. 1994), a comparison among three strains of *R. sphaeroides* revealed that the DNA sequences of CII diverge significantly more than the DNA sequences of CI (Choudhary et al. 2007).

This study sought to uncover the divergence within and among genomes of bacterial species that possess multiple chromosomes and ascertain the possible advantages of such genome structures. An explosion of fully sequenced genomes of microbial species has occurred in recent years (Koonin and Wolf 2008) and has provided data that can be used to understand such complex bacterial genome organization. This study employed several independent approaches to provide measures of CI and CII divergence and evolution within and between different bacterial species: global DNA sequence alignments, phylogenetic analyses, measurement of synonymous and nonsynonymous substitution rates, and horizontal gene transfer estimates. The results obtained from these independent methods demonstrated that CII-specific sequences have diverged faster than CI-specific sequences in several bacterial species. Additional analysis revealed that CI and CII genes seem to have similar structural and functional constraints. Also, observations of current and previous studies are discussed and related to the study findings to elucidate the role of accessory chromosomes in bacteria.

Materials and methods

Global genome alignment

This study focused only on fully sequenced bacteria with two chromosomes to allow for a comparative genomics approach. All DNA and protein sequences used in this study were obtained from the NCBI database. DNA sequences were aligned using Mauve 2.2.0, which allows for rapid genome alignments without sacrificing accuracy (Darling et al. 2004). Then, data were acquired from each alignment to measure CI and CII genome divergence. Mauve classifies colinear subsets of the multiple maximal unique matches into regions called local colinear blocks (LCBs) that serve as anchors for the global alignment. As each LCB contains significantly preserved sections of nucleotide subsequences, the regions together reveal the common areas determined to have actual homology among the genomes analyzed. To maintain consistency as well as preserve accuracy, a minimum seed length of 15 was set for all comparisons conducted. An LCB weight whose value is 3 times the seed length eliminates many false arrangements and groupings in the final alignment (Darling et al. 2004). In accord, a minimum LCB weight of 45 was utilized.

Since nucleotide sequences are aligned only once from one genome to each of the other genomes in an alignment, Mauve recognizes only orthologous regions so paralogous regions are not input into the final alignment. Orthologous regions are DNA regions that originated from a common ancestor, whereas paralogous regions originated within a single specific lineage. More specifically, orthologous regions are homologous regions that are related across speciation events, whereas paralogous regions are homologous regions that are related across duplication events. Mauve also takes into account rearrangements such as translocations and inversions. Therefore, it can be used to align genomes that have significantly diverged.

For each alignment, the number of determined LCBs was obtained and the total length of the LCBs was determined by adding the total number of nucleotides within all of the anchored portions. The percent identity was computed by measuring the retained identity in the common LCBs. To test whether the levels of CII nucleotide retention were statistically different from the levels of CI nucleotide retention, expected values of conserved CII nucleotide length were calculated by multiplying the levels (%) of CI identity retention by the lengths of the common LCBs in CII. These values was compared with the observed values of CII nucleotide retention, the levels (%) of CII identity retention multiplied by the lengths of the common LCBs in CII, using a χ^2 test ($\alpha = 0.05$). Chi-square tests were performed for both the within-species comparison and the across-genera comparison.

There were two sets of comparisons: within-species comparisons and across-genera comparisons. Each of the within-species comparisons was conducted between sequenced strains of the same species. Data extraction for comparisons could be performed only between two strains or among three strains. In total, there were nine within-species comparisons conducted among strains of *Brucella abortus*, *Brucella melitensis*, *Brucella suis*, *Burkholderia mallei*, *Burkholderia pseudomallei*, *Ralstonia eutropha*, *Vibrio cholerae*, *Vibrio fischeri*, and *Vibrio vulnificus*.

In an extension of our analyses of sequence divergence within CI and CII, pairwise comparisons among species from different genera were performed. Since the *Alphaproteobacteria* contains several organisms with complex genomes, the analyses were limited to this group to sustain a level of similarity among the organisms. In accord, five model strains (*Agrobacterium tumefaciens* C58, *Brucella melitensis* 16M, *Ochrobactrum anthropi* ATCC 49188, *Paracoccus denitrificans* PD1222, and *Rhodobacter sphaeroides* 2.4.1) from *Alphaproteobacteria* were selected for the comparisons.

Phylogenetic analysis

Phylogenetic analyses were performed among four related organisms with multiple chromosomes: *Agrobacterium tumefaciens* C58, *Brucella melitensis* 16M, *Paracoccus denitrificans* PD1222, and *Rhodobacter sphaeroides* 2.4.1. OrthoMCL was used to cluster CI-specific and CII-specific protein sequences (Li et al. 2003). OrthoMCL performs an all-versus-all BLAST and determines protein families from the BLAST results. More specifically, the proteins from each organism are blasted to one another and those with ho-

mology are paired into protein clusters. From the determined clusters, only those present in a single copy in each of the compared genomes were extracted. These clusters thus represent genes present only in a single copy in each replicon (orthologs). Thirty of these clusters were randomly selected and the protein files for these clusters were concatenated for each organism. Then, each of these concatenated protein files was aligned using MUSCLE (Edgar 2004). PhyML (Guindon and Gascuel 2003), using the JTT model (Jones et al. 1992), was utilized to construct maximum likelihood trees from the corresponding MUSCLE alignments. In addition, for each tree, 100 replications were performed for the computation of a bootstrap value. For these analyses, both MUSCLE and PhyML were utilized through Geneious, a resourceful bioinformatics suite (Drummond et al. 2010).

Functional constraints analysis

For the functional constraints analysis, five organisms were utilized: *Agrobacterium tumefaciens*, *Brucella melitensis*, *Ochrobactrum anthropi*, *Paracoccus denitrificans*, and *Rhodobacter sphaeroides*. OrthoMCL clustering was performed on CI and CII independently among these organisms. Since 24 single-copy, homologous gene (ortholog) clusters were found among the CIIs, a similar clustering was performed on CI and 24 ortholog clusters were randomly selected. These genes were then utilized for functional constraints analysis. The synonymous and nonsynonymous substitution rates, as well as the nonsynonymous/synonymous rate ratio, were calculated by a modified Yang–Nielsen algorithm (Yang and Nielsen 2000; Zhang et al. 2006*a*). Amino acid sequences were aligned using MUSCLE (Edgar 2004). The aligned sequences were transformed into the original DNA sequences and then each pair of DNA sequences was analyzed with the KaKs_Calculator (Zhang et al. 2006*b*) to calculate the synonymous substitution rate (K_s), the nonsynonymous substitution rate (K_a), and the nonsynonymous/synonymous rate ratio ($\omega = K_a/K_s$). Under the modified Yang–Nielsen model, ω = 0.3, 1, and 3 were used for negative (purifying), neutral, and positive selection, respectively (Yang and Nielsen 2000; Zhang et al. 2006*a*). Since each gene was pairwise compared against an ortholog in another organism, there were 10 calculations for each gene cluster to give 240 data points for CI and the same for CII. C++ ROOT (Antcheva et al. 2009) was used to perform a two-sample Kolmogorov–Smirnov test to determine whether the distributions of the CI and CII constraint data were significantly different.

Horizontal gene transfer

Estimates of horizontal gene transfer (HGT) were calculated for CI and CII independently using Alien_hunter, a program that uses interpolated variable order motifs (IVOMs) to determine regions of probable HGT origin (Vernikos and Parkhill 2006). The IVOM method utilizes compositional biases to ascertain regions of potential HGT origin. A threshold value is calculated using the sequence configuration of the designated genome, among other factors, so that abnormal areas (those of putative HGT origin) are those that have values above the determined threshold. The Alien-hunter output was viewed using Artemis (Rutherford et al. 2000).

Although Alien_hunter has been found to have high recall, as it finds a significant number of foreign genes, it also has low precision, as it presents many false positives (Azad and Lawrence 2007; Langille et al. 2010). To provide a set of comparison points, IslandViewer, a publicly available resource that identifies genomic islands on completed genomes in the NCBI database (Langille and Brinkman 2009), was also used. Two methods that it utilizes are the SIGI-HMM method (Waack et al. 2006) and the IslandPath-DIMOB method (Hsiao et al. 2003, 2005). While SIGI-HMM and IslandPath-DIMOB are both highly accurate and precise compared with Alien_hunter, their recall is lower (Langille et al. 2008, 2010). SIGI-HMM utilizes a hidden Markov model architecture in relation to compositional biases to determine genomic islands (Waack et al. 2006). Similarly, IslandPath-DIMOB utilizes compositional features such as G+C content and dinucleotide bias to predict genomic islands (Hsiao et al. 2003, 2005). IslandViewer provides an integrated output of the two methods so that regions that one method does not identify may be picked up by the other.

Results and discussion

Genome characteristics of bacterial species sampled for this study are described in Table 1. Although CII is smaller than CI in all of the organisms, its size is variable in proportion to CI. Even among strains, the sizes of CI and CII vary substantially. The minimum difference in CI and CII sizes within all of the organisms used in this study occurs in *Brucella suis* ATCC 23445, with a CI–CII difference of ~523 kb and CI constituting ~57.8% of the total genome size. The maximum difference in CI and CII sizes occurs in *Rhodobacter sphaeroides*, with a CI–CII difference of ~2.24 Mb and CI constituting ~77.2% of the total genome size. The G+C compositions of CI and CII of the organisms examined were very similar, with the minimum difference in percent G+C content at 0% and the maximum difference at 2.0%.

Rapid evolution of CII

The within-species pairwise comparisons are detailed in Table 2. Also, typical Mauve global alignments of the CI and CII of three strains of *Burkholderia mallei* are displayed as examples in Fig. 1. Percent identity in CI varied widely, ranging from 66.75% to 99.89%. CII identity varied widely as well, ranging from 57.31% to 99.90%. The mean of the CI identities was 90.82%, while that of the CII identities was 87.23%. There was no explicit pattern among the comparisons regarding the number of LCBs in either chromosome. Of the nine organisms examined, seven displayed greater CII sequence divergence compared with that of CI. Of the remaining two, one organism, *Brucella abortus*, displayed fairly equal CI and CII identity retention values, 99.89% and 99.90%, respectively, while the other organism, *Burkholderia pseudomallei*, displayed marginally lower CI identity retention (86.59%) compared with that of CII (87.27%).

As depicted in Table 3, for most of the across-genera pairwise comparisons, the CI and CII percent identities were notably smaller than those for the within-species com-

Table 1. Genome characteristics of bacterial species.

Bacterial species	Group	CI size (nt)	CI G+C (%)	CI NCBI accession No.	CII size (nt)	CII G+C (%)	CII NCBI accession No.	CI/CII size ratio	Pathogenic
Agrobacterium tumefaciens C58	*Alphaproteobacteria*	2 841 580	59.4	NC_003062	2 075 577	59.3	NC_003063	1.37	Yes
Brucella abortus biovar 1 strain 9-941	*Alphaproteobacteria*	2 124 241	57.2	NC_006932	1 162 204	57.2	NC_006933	1.83	Yes
Brucella abortus S19	*Alphaproteobacteria*	2 122 487	57.2	NC_010742	1 161 449	57.3	NC_010740	1.83	No
Brucella melitensis 16M	*Alphaproteobacteria*	2 117 144	57.2	NC_003317	1 177 787	57.3	NC_003318	1.80	Yes
Brucella melitensis biovar Abortus 2308	*Alphaproteobacteria*	2 121 359	57.2	NC_007618	1 156 948	57.3	NC_007624	1.83	Yes
Brucella suis 1330	*Alphaproteobacteria*	2 107 794	57.2	NC_004310	1 207 381	57.3	NC_004311	1.75	Yes
Brucella suis ATCC 23445	*Alphaproteobacteria*	1 923 763	57.1	NC_010169	1 400 844	57.3	NC_010167	1.37	Yes
Burkholderia mallei ATCC 23344	*Betaproteobacteria*	3 510 148	68.2	NC_006348	2 325 379	69.0	NC_006349	1.51	Yes
Burkholderia mallei NCTC 10229	*Betaproteobacteria*	3 458 208	68.2	NC_008836	2 284 095	68.9	NC_008835	1.51	Yes
Burkholderia mallei NCTC 10247	*Betaproteobacteria*	3 495 687	68.2	NC_009080	2 352 693	69.0	NC_009079	1.49	Yes
Burkholderia pseudomallei 1710b	*Betaproteobacteria*	4 126 292	67.6	NC_007434	3 181 762	68.5	NC_007435	1.30	Yes
Burkholderia pseudomallei 668	*Betaproteobacteria*	3 912 947	68.0	NC_009074	3 127 456	68.6	NC_009075	1.25	Yes
Burkholderia pseudomallei K96243	*Betaproteobacteria*	4 074 542	67.7	NC_006350	3 173 005	68.5	NC_006351	1.28	Yes
Ochrobactrum anthropi ATCC 49188	*Alphaproteobacteria*	2 887 297	56.1	NC_009667	1 895 911	56.2	NC_009668	1.52	Yes
Paracoccus denitrificans PD1222	*Alphaproteobacteria*	2 852 282	66.7	NC_008686	1 730 097	66.8	NC_008687	1.65	No
Ralstonia eutropha H16	*Betaproteobacteria*	4 052 032	66.5	NC_008313	2 912 490	66.8	NC_008314	1.39	No
Ralstonia eutropha JMP134	*Betaproteobacteria*	3 806 533	64.7	NC_007347	2 726 152	65.0	NC_007348	1.40	No
Rhodobacter sphaeroides 2.4.1	*Alphaproteobacteria*	3 188 609	69.0	NC_007493	943 016	69.0	NC_007494	3.38	No
Vibrio cholerae O395	*Gammaproteobacteria*	3 024 069	46.9	NC_009456	1 108 250	47.8	NC_009457	2.73	Yes
Vibrio cholerae O1 biovar eltor strain N16961	*Gammaproteobacteria*	2 961 149	47.7	NC_002505	1 072 315	46.9	NC_002506	2.76	Yes
Vibrio fischeri ES114	*Gammaproteobacteria*	2 897 536	39.0	NC_006840	1 330 333	37.0	NC_006841	2.18	No
Vibrio fischeri MJ11	*Gammaproteobacteria*	2 905 029	38.9	NC_011184	1 418 848	37.2	NC_011186	2.05	No
Vibrio vulnificus CMCP6	*Gammaproteobacteria*	3 281 944	46.4	NC_004459	1 844 853	47.1	NC_004460	1.78	Yes
Vibrio vulnificus YJ016	*Gammaproteobacteria*	3 354 505	46.4	NC_005139	1 857 073	47.2	NC_005140	1.81	Yes

Table 2. Within-species comparisons.

Comparison	CI			CII		
	Common LCBs	Length of common LCBs (nt)	Nucleotide identity (%)	Common LCBs	Length of common LCBs (nt)	Nucleotide identity (%)
Alphaproteobacteria						
Brucella abortus biovar 1 strain 9-941 / *Brucella abortus* S19	1	2 124 438	99.89	1	1 162 354	99.90
Brucella melitensis 16M / *Brucella melitensis* biovar Abortus 2308	10	2 126 972	98.87	8	1 193 542	95.42
Brucella suis 1330 / *Brucella suis* ATCC 23445	5	2 024 549	93.61	1	1 416 117	84.05
Rhodobacter sphaeroides 2.4.1 / *Rhodobacter sphaeroides* ATCC 17025 / *Rhodobacter sphaeroides* ATCC 17029*	263	2 896 258	76.80	119	451 250	71.40
Betaproteobacteria						
Burkholderia mallei ATCC 23344 / *Burkholderia mallei* NCTC 10229 / *Burkholderia mallei* NCTC 10247	46	3 530 202	95.09	16	2 417 983	92.42
Burkholderia pseudomallei 1710b / *Burkholderia pseudomallei* 668 / *Burkholderia pseudomallei* K96243	18	4 280 362	86.59	28	3 320 503	87.27
Ralstonia eutropha H16 / *Ralstonia eutropha* JMP134	177	3 522 606	66.75	337	1 816 082	57.31
Gammaproteobacteria						
Vibrio cholerae O1 biovar eltor strain N16961 / *Vibrio cholerae* O395	4	3 076 500	93.71	31	1 104 530	91.91
Vibrio fischeri ES114 / *Vibrio fischeri* MJ11	35	3 044 396	85.45	66	1 504 706	71.10
Vibrio vulnificus CMCP6 / *Vibrio vulnificus* YJ016	170	3 341 213	85.97	16	1 919 284	84.82

*Results from Choudhary et al. 2007. Their study performed pairwise comparisons using the same alignment parameters as this study and their results are comparable to the data analyzed here.

parisons, with CI sequence identities ranging from 36.17% to 57.59% and CII sequence identities ranging from 14.26% to 46.27%; the mean of the CI nucleotide identities was 42.65%, while the mean of the CII nucleotide identities was 27.60%. The overall pattern of CI and CII sequence identities is consistent with the within-species comparisons. However, the genome comparison between *Agrobacterium tumefaciens* C58 and *Ochrobactrum anthropi* ATCC 49188 showed higher identity for CII than for CI owing to the presence of a more conserved region of relatively small DNA segments. The observed distributions of conserved nucleotides in CII significantly differed from the expected distributions in both the within-species and across-genera comparisons (χ^2 tests, $p < 0.0001$), indicating that most CII sequences diverged more rapidly than CI sequences.

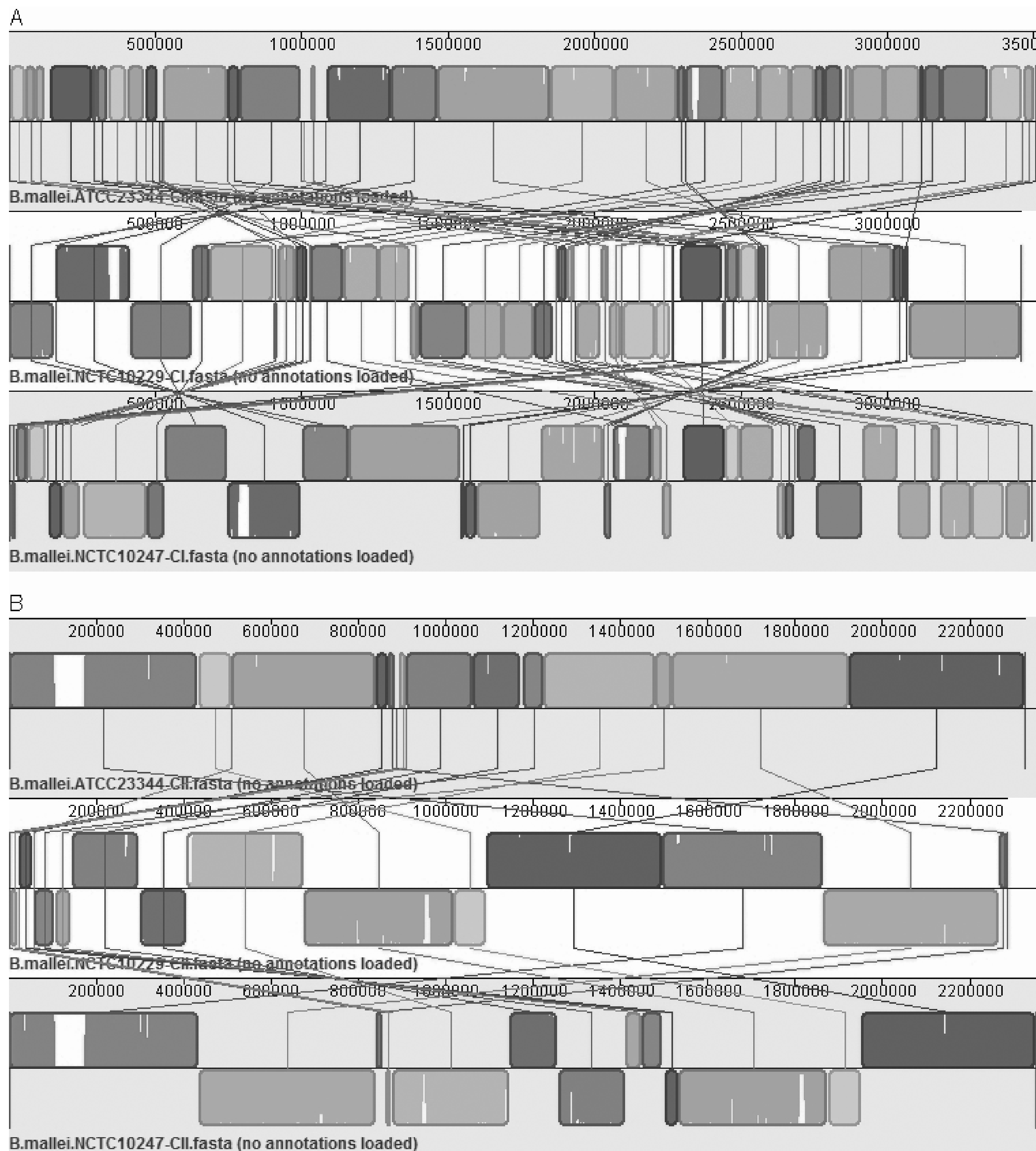
The results of the within-species and across-genera comparisons suggest a trend, namely that CII may evolve faster than CI in many organisms with such complex genome structures. In regions with demonstrated homology, CII generally diverged to a greater extent than CI. As would be expected, the across-genera comparisons demonstrated significantly lower magnitudes of identity retention than the within-species comparisons.

Phylogenetic analyses for CI-specific proteins and CII-specific proteins are displayed in Fig. 2. The aligned lengths of the concatenated CI and CII protein sequences were 12 291 and 13 303 amino acids, respectively. The bootstrap value for both trees was 100. The offshoot branch lengths confirm the global genome alignment findings. More specifically, the branch lengths represent amino acid substitutions per site and are similar to extents of divergence. All offshoot branch lengths representing organismal CII protein divergence rates are greater (1.3–2.4 times) than corresponding CI divergence rates.

Among all of the species, the maximum G+C content difference between CI and CII was 2.0%, lending significant credence to a long-term relationship between CI and CII in each of the organisms. Furthermore, in many of the organisms CII may have coexisted with CI in an ancient partnership, as has been demonstrated in *Rhodobacter sphaeroides* (Choudhary et al. 2004). Likewise, both CI and CII seemingly coexisted prior to the divergence and speciation of organisms in the genus *Vibrio* (Thompson et al. 2004). An extended evolutionary association between CI and CII would suggest that not only are both chromosomes essential for growth and expansion of these species, but also that CII may play a larger and more necessary role than has previously been considered.

Although the importance of rapid CII evolution and the CI–CII relationship still need to be explicitly discerned, some clues are available. For example, in a comparison between *Brucella melitensis* 16M and *Brucella melitensis* biovar Abortus 2308, CII regions diverged more than CI regions. Even though the two organisms have been shown to be highly related in evolutionary lineage and genome composition (Gándara et al. 2001; Chain et al. 2005), the two also have different host species specificities (Halling et al. 2005). Such differing host specificities and altered pathogenicity may be partly explained by CII divergence and adaptation; the second chromosome can serve as an adaptive

Fig. 1. Visual representations of two anchored Mauve 2.2.0 pairwise comparisons conducted among the CI (Fig. 1A) and CII (Fig. 1B) of three strains of *Burkholderia mallei*. Both chromosomes of the species were independently aligned among the strains to measure identity retention in common local colinear blocks (LCBs). An LCB weight of 45 was used. The linked blocks in the alignment represent the common LCBs among the compared genomes and inside the LCBs are similarity plots depicting the level of DNA sequence identity retained within those homologous regions. The lines at the end of each genome represent chromosome boundaries. The horizontal lines to which the LCBs are affixed depict the orientation of the LCBs such that LCBs anchored underneath the line represent reverse complements in relation to the first genome in the alignment.



Pagination not final/Pagination non finale

Table 3. Across-genera comparisons (within *Alphaproteobacteria*).

Organism 1 / organism 2 comparison	CI			CII		
	Common LCBs	Length of common LCBs (nt)	Nucleotide identity (%)	Common LCBs	Length of common LCBs (nt)	Nucleotide identity (%)
Rhizobiales/Rhizobiales						
Agrobacterium tumefaciens C58 / *Brucella melitensis* 16M	232	1 616 926	40.81	87	497 538	27.72
Agrobacterium tumefaciens C58 / *Ochrobactrum anthropi* ATCC 49188	278	1 915 476	36.91	122	397 268	45.93
Brucella melitensis 16M / *Ochrobactrum anthropi* ATCC 49188	41	2 591 685	57.59	82	1 113 797	46.27
Rhodobacteraceae/Rhodobacteraceae						
Paracoccus denitrificans PD1222 / *Rhodobacter sphaeroides* 2.4.1	356	1 349 469	51.02	85	291 560	27.17
Rhizobiales/Rhodobacteraceae						
Agrobacterium tumefaciens C58 / *Paracoccus denitrificans* PD1222	262	538 702	41.55	82	331 239	23.77
Agrobacterium tumefaciens C58 / *Rhodobacter sphaeroides* 2.4.1	319	719 063	43.05	67	362 051	17.89
Brucella melitensis 16M / *Paracoccus denitrificans* PD1222	209	455 765	40.36	72	271 562	24.23
Brucella melitensis 16M / *Rhodobacter sphaeroides* 2.4.1	259	625 629	41.69	36	222 914	14.26
Ochrobactrum anthropi ATCC 49188 / *Paracoccus denitrificans* PD1222	231	516 993	37.33	72	275 765	31.06
Ochrobactrum anthropi ATCC 49188 / *Rhodobacter sphaeroides* 2.4.1	254	748 661	36.17	58	270 193	17.71

ground in which ecological and nutritional processes can be modified to better suit the host environment.

As has been previously found, *Brucella suis* ATCC 23445 underwent an interchromosomal recombination between its CI and CII (Wattam et al. 2009). This translocation seems to be fairly recent, since the regions have not diverged significantly compared with the homologous regions in *Brucella suis* 1330 (greater than 99% identity was retained). As bacterial plasmids and chromosomes can undergo genetic recombination (Hopwood and Wright 1976; Broda 1979; Pandza et al. 1998) so that certain genes are shuffled and recombined to provide potential evolutionary advantages, bacteria with multiple chromosomes may similarly perform a comparable type of gene shuffling between chromosomes.

Origin and evolution of multiple chromosomes

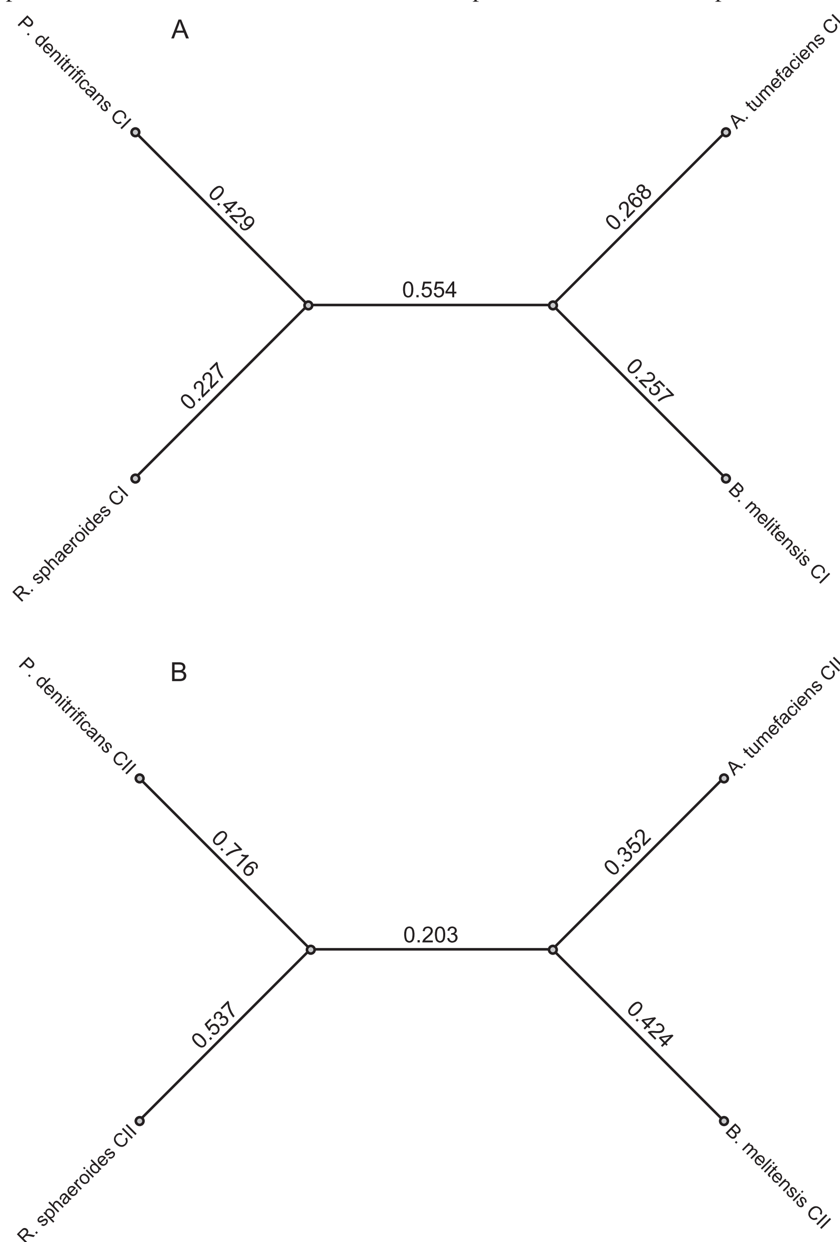
Although the occurrence of multiple chromosomes is widespread in many distantly related groups of bacteria, the advantage of multiple chromosomes is unclear; however, it has been speculated that the larger the chromosome, the more inefficient its ability to replicate compared with normal-sized chromosomes. However, there are some bacteria whose chromosomes are relatively large, such as *Myxococcus xanthus* (∼9 Mb) and *Sorangium cellulosum* So ce56 (∼13 Mb) (He et al. 1994; Schneiker et al. 2007). Therefore, size alone cannot explain the development of multiple chromosomes. In addition, a Mauve comparison between two strains of *Rhodopseudomonas palustris*, which possesses a single chromosome, reveals significant nucleotide divergence, with chromosome nucleotide identity at 55.87% (see Table 4). A Mauve comparison between two strains of *Burkholderia cenocepacia*, an organism with three chromosomes, shows a pattern similar to that found in the CI–CII comparisons, as shown in Table 4. More specifically, CIII retained less sequence identity (76.94%) than CI (89.40%) and CII (92.70%); this may also suggest that in organisms with three chromosomes, different chromosomes experience different levels of evolutionary pressures.

Comparative genome analysis of bacterial species that possess multiple chromosomes suggests that the second chromosome (CII) originated from plasmids (Mackenzie et al. 2004; Slater et al. 2009; Wattam et al. 2009). In this model, a resident plasmid in an organism would acquire a few clusters of essential genes by intragenomic gene transfer from the primary chromosome and by HGT from related species; over time, this mechanism would facilitate the establishment of an accessory chromosome.

There are several lines of evidence that support the above proposals. Plasmid replication among the *Rhizobiaceae* is generally under the control of the *repABC* system. Recent studies confirm that the intragenomic movement of essential genes including *repABC* genes may assist in the formation of a second chromosome (Cevallos et al. 2008; Slater et al. 2009). For instance, a recent comparative genome analysis of *Brucella* strains has revealed that a 210-kb segment of CI has been translocated to CII in one strain, indicating that a large cluster of essential genes have been resituated from CI to CII (Wattam et al. 2009). Also, the biovar I genome of *Agrobacterium tumefaciens* C58 harbors a linear CII derived from a plasmid to which large blocks of DNA, including rRNA and other essential genes, have been transferred

PROOF/ÉPREUVE

Fig. 2. Maximum likelihood trees for genes found exclusively on the CI (Fig. 2A) and CII (Fig. 2B) of four organisms: *Agrobacterium tumefaciens* C58, *Brucella melitensis* 16M, *Paracoccus denitrificans* PD1222, and *Rhodobacter sphaeroides* 2.4.1. Trees were constructed using PhyML. The bootstrap values for both trees are 100. The numbers represent the substitutions per site in relation to the divergent node.



from CI (Slater et al. 2009). Furthermore, CII appears to be more plastic than CI, with shared internal rearrangements among strains. A segment of approximately 700 kb in CII is a shared inversion among the three *Brucella abortus* genomes (Wattam et al. 2009). Taken all together, these data corroborate a generalized plasmid-mediated model of secondary chromosome formation among bacteria.

Mechanisms of CII evolution

Although a plasmid-mediated origin is the favored model explaining the origin of additional chromosomes in bacteria, accessory chromosomes may acquire or develop new genes through different mechanisms such as gene transfer (intergenomic or intragenomic), genetic recombination, and gene duplication. The identity retained within each chromosome would depend upon several factors such as the mutation rate, the recombination rate, the level of constraints, HGT, and ecological selection. Since these factors are by no means mutually exclusive, combinations of these factors would explain why the divergence of CII varies across bacteria.

Because fewer structural and functional constraints on CII genes would explain their increased divergence rate compared with CI genes, functional constraints analysis was performed on homologous CI and CII genes across five *Alphaproteobacteria* (*Agrobacterium tumefaciens* C58, *Brucella melitensis* 16M, *Ochrobactrum anthropi* ATCC 49188, *Paracoccus denitrificans* PD1222, and *Rhodobacter sphaeroides* 2.4.1). The correlations of nonsynonymous (K_a) and synonymous (K_s) substitution rates for CI genes and CII

Table 4. Single and triple chromosome comparisons.

Comparison	CI			CII			CIII		
	Common LCBs	Length of common LCBs (nt)	Nucleotide identity (%)	Common LCBs	Length of common LCBs (nt)	Nucleotide identity (%)	Common LCBs	Length of common LCBs (nt)	Nucleotide identity (%)
Rhodopseudomonas palustris BisB18 / *Rhodopseudomonas palustris* CGA009	565	4 585 536	55.87	—	—	—	—	—	—
Burkholderia cenocepacia AU 1054 / *Burkholderia cenocepacia* HI2424	17	3 187 567	89.40	5	3 003 016	92.70	21	823 329	76.94

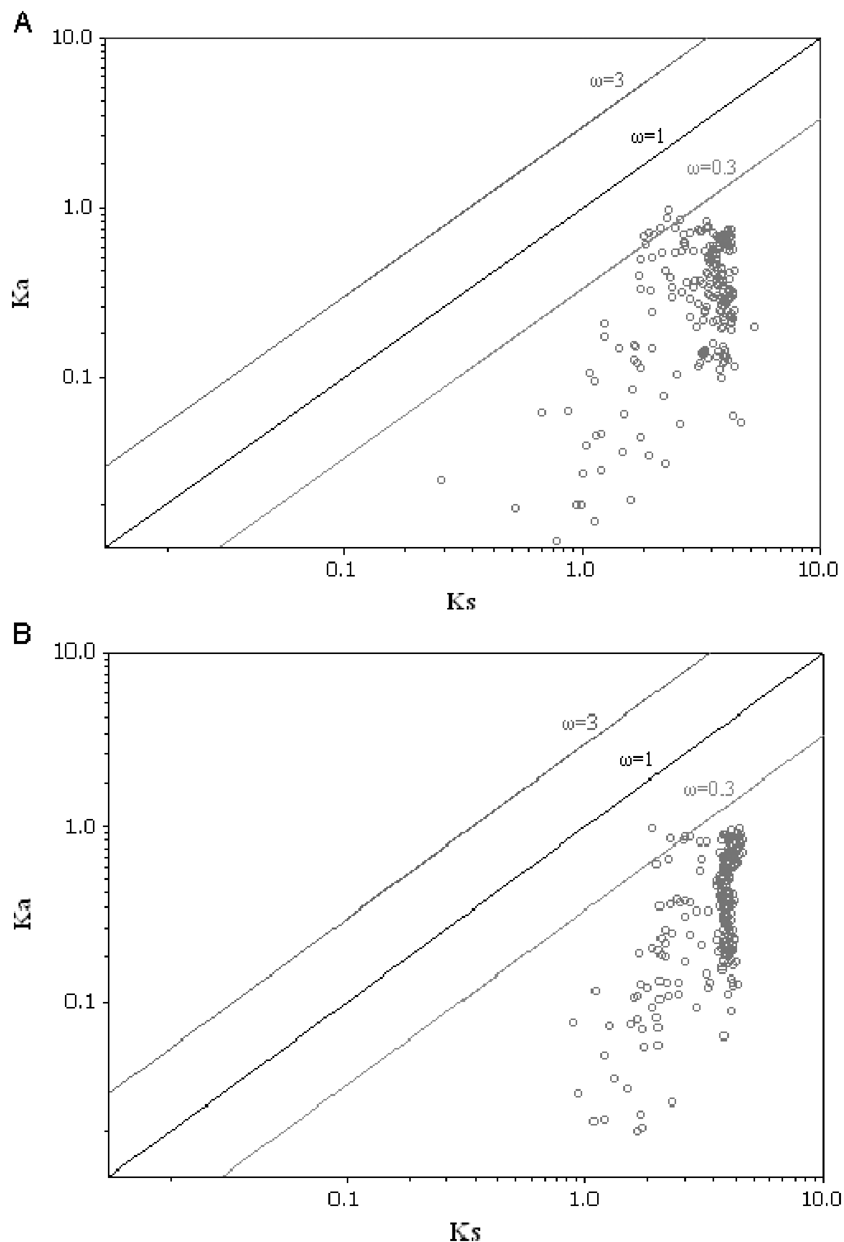
genes are shown in Fig. 3. Under the modified Yang–Nielsen algorithm, ω = 0.3, 1, and 3 were used for negative, neutral, and positive selection, respectively (Yang and Nielsen 2000; Zhang et al. 2006*a*). The correlation data reveal that most CI and CII data points cluster similarly, with the majority of ω values less than 0.3, indicative of negative or purifying selection operating on these orthologs. The mean ω values for CI and CII were 0.1176 and 0.1160, respectively, and the standard deviations of ω for CI and CII were 0.0790 and 0.074, respectively. The two-sample Kolmogorov–Smirnov test, which was performed to test the distribution of CI and CII data points, was insignificant (p = 0.8088), suggesting that the levels of constraints on CI and CII genes are not markedly different.

Since the constraints on CI and CII genes seem to be similar, differences in CI and CII replication could explain the more rapid CII divergence rates. In certain organisms such as *Vibrio cholerae*, CI and CII replication have been shown to possess both shared and unshared features (Egan and Waldor 2003; Egan et al. 2005). For instance, in *Vibrio cholerae*, CI and CII replication are initiated synchronously once during the cell cycle, and replication in both chromosomes is coordinated by common factors such as DnaA and Dam methylation (Egan and Waldor 2003; Egan et al. 2004). In the same organism, however, RctB, exclusive to CII, binds to the CII origin and acts like an autorepressor, regulating CII replication (Egan et al. 2006). Moreover, it has been shown that when *Vibrio cholerae* is placed in rabbit intestine, CII expression increases dramatically compared with that in aerobic laboratory conditions (Xu et al. 2003), and so gene regulation could be controlled independently on the two chromosomes (Heidelberg et al. 2000).

Horizontal gene transfers constitute another significant mechanism by which bacteria evolve and adapt (Jain et al. 1999; de la Cruz and Davies 2000; Ochman et al. 2000). In particular, HGTs have played an important role in the development of pathogenic bacteria (Fuchs 1998) and in the process of speciation (Lawrence and Ochman 1998; Garcia-Vallvé et al. 2000). As the CI of bacterial organisms contains essential housekeeping genes, CII may have evolved, through selection, to serve as an adaptive reservoir through which bacteria can suit ecological niches and nutritional constraints. Alien_hunter horizontal gene transfer estimates for selected organisms are detailed in Table 5 along with integrated IslandViewer estimates. Examining the Alien_hunter estimates, CI HGT levels range from 2.04% to 17.59%, while CII HGT levels range from 9.66% to 27.72%. Of the nine organisms examined, seven possess HGT levels higher in CII (percentage-wise) than in CI. For these seven organisms, CII HGT values are 1.03 to 5.72 times greater than corresponding CI values. The IslandViewer estimates present a slightly different picture, as the CI HGT levels range from 1.92% to 9.76% while the CII HGT levels range from 3.58% to 14.65%. With this method, five of the nine organisms possess higher levels of HGT in CII (percentage-wise) than in CI; for these five organisms, CII HGT values are 1.09 to 2.89 times greater than corresponding CI values.

The estimates reveal that the levels of HGT in CI and CII are not simply functions of chromosome size, even though in all organisms examined, CII is smaller than CI. Although the estimates also reveal that there are significant levels of

Fig. 3. K_a–K_s correlation of 24 chromosome-specific homologous genes in CI (Fig. 3A) and CII (Fig. 3B) compared pairwise across five *Alphaproteobacteria* (240 data points for each replicon). K_a and K_s values were estimated using MYN (modified Yang–Nielsen algorithm); ω = 0.3, 1, and 3 were used for negative, neutral, and positive selection, respectively.



HGT on CII, it is difficult to ascertain whether CII truly does have higher levels of HGT compared with CI. It must be noted that HGT events and the relative occurrence of these events are difficult to ascertain. It is likely that older HGT events would have homogenized to the structure and makeup of the overall genome and likely would not be readily distinguishable. Furthermore, genes or regions even recently transferred from an organism of similar genomic composition (i.e., G+C content, codon bias, etc.) may not be detectable in an examined genome. Moreover, if the second chromosome in many bacteria evolved from megaplasmids that possessed genes from various sources, then it could be that the regions identified as horizontally transferred were present prior to the formation of CII and simply have not taken on the sequence features of the overall genome in which they are present. In addition, in many organisms, CI and CII could also possess different functional classes of genes that would explain their different evolution rates.

Significant CII divergence may occur through only one of these factors or a combination of these factors. In analyses of several *R. sphaeroides* strains, the size of CI stays fairly constant while the size of CII varies (Nereng and Kaplan 1999). Seemingly, this is true also in several species in the genus *Vibrio*, where the CII substantially varies (Okada et al. 2005). Such CII variation points to a variety of scenarios involving possible CII dispositions for genetic recombination and HGT as well as differing replication mechanisms and thereby different mutation rates for CI and CII. As such, since genome differentiation and sequence divergence

Table 5. Horizontal gene transfer estimates in selected organisms.

	Alien_hunter estimates				Integrated IslandViewer estimates			
Organism	Total length of CI HGT regions	HGT regions as % of CI genome	Total length of CII HGT regions	HGT regions as % of CII genome	Total length of CI HGT regions	HGT regions as % of CI genome	Total length of CII HGT regions	HGT regions as % of CII genome
Agrobacterium tumefaciens C58	456 136	16.05	200 592	9.66	54 503	1.92	83 549	4.03
Brucella melitensis 16M*	227 518	10.75	202 519	17.19	108 481	5.12	42 140	3.58
Burkholderia mallei ATCC 23344	617 567	17.59	470 431	20.23	204 777	5.83	201 916	8.68
Burkholderia pseudomallei K96243	630 046	15.46	503 043	15.85	315 233	7.74	171 979	5.42
Ochrobactrum anthropi ATCC 49188	439 832	15.23	457 533	24.13	190 789	6.61	136 207	7.18
Paracoccus denitrificans PD1222	427 522	14.99	205 022	11.85	278 247	9.76	106 278	6.14
Ralstonia eutropha JMP134	307 527	8.08	755 703	27.72	218 327	5.74	97 609	3.58
Rhodobacter sphaeroides 2.4.1*	65 005	2.04	110 009	11.66	147 460	4.62	69 409	7.36
Vibrio cholerae O1 biovar eltor strain N16961	287 528	9.71	245 020	22.85	150 188	5.07	157 091	14.65

*Alien_hunter estimates from A. Bavishi, L. Lin, K. Schroeder, A. Peters, and M. Choudhary (unpublished data). Their regions analysis was conducted in a similar manner to those performed in the current study.

among strains may require a combination of these factors, the strains of species that display small but significant differences between CI and CII may not have been exposed to new elements and environments to spur CII adaptation.

Gene expression studies using microarray analysis conducted in differing growth conditions, as in *Vibrio cholerae* (Xu et al. 2003), would shed more light on the roles of CII. More precisely, if CII serves to adapt an organism to distinctive hosts for specific nutritional exploitation, then CII expression would be expected to be amplified in such in vivo conditions when compared with defined laboratory growth conditions. In such processes, it could be expected that CII would go through rapid degeneration and pseudogene formation, as such degeneration could specialize the function of CII in intracellular pathogenic organisms. For instance, *Brucella* genomes contained a high fraction (4.6%) of pseudogenes suggestive of genome degradation, and when broken down by replicon, the pseudogene fraction was 3.9% for CI and 6% for CII (Wattam et al. 2009). A more detailed examination of the variable CII sequences would also aid in confirming the advantages of the second chromosome. Furthermore, examination of such sequences would possibly unveil the mechanisms by which the bacteria target CII as a genetic reservoir for adaption. Moreover, there are bacteria with three chromosomes and exploring the roles of these chromosomes in different bacteria may shed more light on complex prokaryotic genome structuring.

Acknowledgements

This work was supported by the Enhancement Grant for Research (EGR) from Sam Houston State University to Madhusudan Choudhary. We thank the two anonymous reviewers and Dr. Brian Golding for their helpful comments and suggestions.

References

Allardet-Servent, A., Michaux-Charachon, S., Jumas-Bilak, E., Karayan, L., and Ramuz, M. 1993. Presence of one linear and one circular chromosome in the *Agrobacterium tumefaciens* C58 genome. J. Bacteriol. **175**(24): 7869–7874. PMID:8253676.

Antcheva, I., Ballintijn, M., Bellenot, B., Biskup, M., Brun, R., Buncic, N., et al. 2009. ROOT — a C++ framework for petabyte data storage, statistical analysis and visualization. Comput. Phys. Commun. **180**(12): 2499–2512. doi:10.1016/j.cpc.2009.08.005.

Azad, R.K., and Lawrence, J.G. 2007. Detecting laterally transferred genes: use of entropic clustering methods and genome position. Nucleic Acids Res. **35**(14): 4629–4639. doi:10.1093/nar/gkm204. PMID:17591616.

Broda, P. 1979. Plasmids. W.H. Freeman & Co., Oxford, England.

Cevallos, M.A., Cervantes-Rivera, R., and Gutiérrez-Ríos, R.M. 2008. The *repABC* plasmid family. Plasmid, **60**(1): 19–37. doi:10.1016/j.plasmid.2008.03.001. PMID:18433868.

Chain, P.S., Comerci, D.J., Tolmasky, M.E., Larimer, F.W., Malfatti, S.A., Vergez, L.M., et al. 2005. Whole-genome analyses of speciation events in pathogenic Brucellae. Infect. Immun. **73**(12): 8353–8361. doi:10.1128/IAI.73.12.8353-8361.2005. PMID:16299333.

Choudhary, M., Mackenzie, C., Nereng, K.S., Sodergren, E., Weinstock, G.M., and Kaplan, S. 1994. Multiple chromosomes in bacteria: structure and function of chromosome II of *Rhodobacter sphaeroides* 2.4.1^{T}. J. Bacteriol. **176**(24): 7694–7702. PMID: 8002595.

Choudhary, M., Fu, Y.X., Mackenzie, C., and Kaplan, S. 2004. DNA sequence duplication in *Rhodobacter sphaeroides* 2.4.1: evidence of an ancient partnership between chromosomes I and II. J. Bacteriol. **186**(7): 2019–2027. doi:10.1128/JB.186.7.2019-2027.2004. PMID:15028685.

Choudhary, M., Zanhua, X., Fu, Y.X., and Kaplan, S. 2007. Genome analyses of three strains of *Rhodobacter sphaeroides*: evidence of rapid evolution of chromosome II. J. Bacteriol. **189**(5): 1914–1921. doi:10.1128/JB.01498-06. PMID:17172323.

Darling, A.C., Mau, B., Blattner, F.R., and Perna, N.T. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. **14**(7): 1394–1403. doi:10.1101/gr.2289704. PMID:15231754.

de la Cruz, F., and Davies, J. 2000. Horizontal gene transfer and the origin of species: lessons from bacteria. Trends Microbiol. **8**(3): 128–133. doi:10.1016/S0966-842X(00)01703-0. PMID:10707066.

Drummond, A., Kearse, M., Heled, J., Moir, R., Thierer, T., Ashton, B., et al. 2010. Geneious [computer program]. Version 4.8. Available from http://www.geneious.com.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**(5): 1792–1797. doi:10.1093/nar/gkh340. PMID:15034147.

Egan, E.S., and Waldor, M.K. 2003. Distinct replication requirements for the two *Vibrio cholerae* chromosomes. Cell, **114**(4): 521–530. doi:10.1016/S0092-8674(03)00611-1. PMID:12941279.

Egan, E.S., Løbner-Olesen, A., and Waldor, M.K. 2004. Synchronous replication initiation of the two *Vibrio cholerae* chromosomes. Curr. Biol. **14**(13): R501–R502. doi:10.1016/j.cub.2004.06.036. PMID:15242627.

Egan, E.S., Fogel, M.A., and Waldor, M.K. 2005. Divided genomes: negotiating the cell cycle in prokaryotes with multiple chromosomes. Mol. Microbiol. **56**(5): 1129–1138. doi:10.1111/j.1365-2958.2005.04622.x. PMID:15882408.

Egan, E.S., Duigou, S., and Waldor, M.K. 2006. Autorepression of RctB, an initiator of *Vibrio cholerae* chromosome II replication. J. Bacteriol. **188**(2): 789–793. doi:10.1128/JB.188.2.789-793.2006. PMID:16385068.

Fuchs, T.M. 1998. Molecular mechanisms of bacterial pathogenicity. Naturwissenschaften, **85**(3): 99–108. doi:10.1007/s001140050463. PMID:9562989.

Gándara, B., Merino, A.L., Rogel, M.A., and Martínez-Romero, E. 2001. Limited genetic diversity of *Brucella* spp. J. Clin. Microbiol. **39**(1): 235–240. doi:10.1128/JCM.39.1.235-240.2001. PMID:11136777.

Garcia-Vallvé, S., Romeu, A., and Palau, J. 2000. Horizontal gene transfer in bacterial and archaeal complete genomes. Genome Res. **10**(11): 1719–1725. doi:10.1101/gr.130000. PMID:11076857.

Guindon, S., and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52**(5): 696–704. doi:10.1080/10635150390235520. PMID:14530136.

Halling, S.M., Peterson-Burch, B.D., Bricker, B.J., Zuerner, R.L., Qing, Z., Li, L.L., et al. 2005. Completion of the genome sequence of *Brucella abortus* and comparison to the highly similar genomes of *Brucella melitensis* and *Brucella suis.* J. Bacteriol. **187**(8): 2715–2726. doi:10.1128/JB.187.8.2715-2726.2005. PMID:15805518.

He, Q., Chen, H., Kuspa, A., Cheng, Y., Kaiser, D., and Shimkets, L.J. 1994. A physical map of the *Myxococcus xanthus* chromosome. Proc. Natl. Acad. Sci. U.S.A. **91**(20): 9584–9587. doi:10.1073/pnas.91.20.9584. PMID:7937810.

Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., et al. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature, **406**(6795): 477–483. doi:10.1038/35020000. PMID:10952301.

Hopwood, D.A., and Wright, H.M. 1976. Genetic studies on SCP1-prime strains of *Streptomyces coelicolor* A3(2). J. Gen. Microbiol. **95**(1): 107–120. PMID:956772.

Hsiao, W., Wan, I., Jones, S.J., and Brinkman, F.S. 2003. IslandPath: aiding detection of genomic islands in prokaryotes. Bioinformatics, **19**(3): 418–420. doi:10.1093/bioinformatics/btg004. PMID:12584130.

Hsiao, W.W., Ung, K., Aeschliman, D., Bryan, J., Finlay, B.B., and Brinkman, F.S. 2005. Evidence of a large novel gene pool associated with prokaryotic genomic islands. PLoS Genet. **1**(5): e62. doi:10.1371/journal.pgen.0010062. PMID:16299586.

Jain, R., Rivera, M.C., and Lake, J.A. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. Proc. Natl. Acad. Sci. U.S.A. **96**(7): 3801–3806. doi:10.1073/pnas.96.7.3801. PMID:10097118.

Jones, D.T., Taylor, W.R., and Thornton, J.M. 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. **8**(3): 275–282. PMID:1633570.

Koonin, E.V., and Wolf, Y.I. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res. **36**(21): 6688–6719. doi:10.1093/nar/gkn668. PMID:18948295.

Langille, M.G., and Brinkman, F.S. 2009. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. Bioinformatics, **25**(5): 664–665. doi:10.1093/bioinformatics/btp030. PMID:19151094.

Langille, M.G., Hsiao, W.W., and Brinkman, F.S. 2008. Evaluation of genomic island predictors using a comparative genomics approach. BMC Bioinformatics, **9**(1): 329. doi:10.1186/1471-2105-9-329. PMID:18680607.

Langille, M.G.I., Hsiao, W.W.L., and Brinkman, F.S.L. 2010. Detecting genomic islands using bioinformatics approaches. Nat. Rev. Microbiol. **8**(5): 373–382. doi:10.1038/nrmicro2350. PMID:20395967.

Lawrence, J.G., and Ochman, H. 1998. Molecular archaeology of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. U.S.A. **95**(16): 9413–9417. doi:10.1073/pnas.95.16.9413. PMID:9689094.

Lessie, T.G., Hendrickson, W., Manning, B.D., and Devereux, R. 1996. Genomic complexity and plasticity of *Burkholderia cepacia.* FEMS Microbiol. Lett. **144**(2-3): 117–128. doi:10.1111/j.1574-6968.1996.tb08517.x. PMID:8900054.

Li, L., Stoeckert, C.J., Jr., and Roos, D.S. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. **13**(9): 2178–2189. doi:10.1101/gr.1224503. PMID:12952885.

Mackenzie, C., Kaplan, S., and Choudhary, M. 2004. Multiple chromosomes. *In* Microbial evolution: gene establishment, survival, and exchange. *Edited by* R.V. Miller and M.J. Day. ASM Press, Washington, D.C. pp. 82–101.

Michaux, S., Paillisson, J., Carles-Nurit, M.J., Bourg, G., Allardet-Servent, A., and Ramuz, M. 1993. Presence of two independent chromosomes in the *Brucella melitensis* 16M genome. J. Bacteriol. **175**(3): 701–705. PMID:8423146.

Nereng, K.S., and Kaplan, S. 1999. Genomic complexity among strains of the facultative photoheterotrophic bacterium *Rhodobacter sphaeroides*. J. Bacteriol. **181**(5): 1684–1688. PMID:10049404.

Ochman, H., Lawrence, J.G., and Groisman, E.A. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature, **405**(6784): 299–304. doi:10.1038/35012500. PMID:10830951.

Okada, K., Iida, T., Kita-Tsukamoto, K., and Honda, T. 2005. Vibrios commonly possess two chromosomes. J. Bacteriol. **187**(2): 752–757. doi:10.1128/JB.187.2.752-757.2005. PMID:15629946.

Pandza, S., Biuković, G., Paravić, A., Dadbin, A., Cullum, J., and Hranueli, D. 1998. Recombination between the linear plasmid pPZG101 and the linear chromosome of *Streptomyces rimosus* can lead to exchange of ends. Mol. Microbiol. **28**(6): 1165–1176. doi:10.1046/j.1365-2958.1998.00877.x. PMID:9680206.

Rodley, P.D., Römling, U., and Tümmler, B. 1995. A physical genome map of the *Burkholderia cepacia* type strain. Mol. Microbiol. **17**(1): 57–67. doi:10.1111/j.1365-2958.1995.mmi_17010057.x. PMID:7476209.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. 2000. Artemis: sequence visualization and annotation. Bioinformatics, **16**(10): 944–945. doi:10.1093/bioinformatics/16.10.944. PMID:11120685.

Schneiker, S., Perlova, O., Kaiser, O., Gerth, K., Alici, A., Altmeyer, M.O., et al. 2007. Complete genome sequence of the myxobacterium *Sorangium cellulosum*. Nat. Biotechnol. **25**(11): 1281–1289. doi:10.1038/nbt1354. PMID:17965706.

Slater, S.C., Goldman, B.S., Goodner, B., Setubal, J.C., Farrand, S.K., Nester, E.W., et al. 2009. Genome sequences of three *Agrobacterium* biovars help elucidate the evolution of multichromosome genomes in bacteria. J. Bacteriol. **191**(8): 2501–2511. doi:10.1128/JB.01779-08. PMID:19251847.

Suwanto, A., and Kaplan, S. 1989. Physical and genetic mapping of the *Rhodobacter sphaeroides* 2.4.1 genome: presence of two unique circular chromosomes. J. Bacteriol. **171**(11): 5850–5859. PMID:2808300.

Suwanto, A., and Kaplan, S. 1992. Chromosome transfer in *Rhodobacter sphaeroides*: Hfr formation and genetic evidence for two unique circular chromosomes. J. Bacteriol. **174**(4): 1135–1145. PMID:1735708.

Thompson, F.L., Iida, T., and Swings, J. 2004. Biodiversity of vibrios. Microbiol. Mol. Biol. Rev. **68**(3): 403–431. doi:10.1128/MMBR.68.3.403-431.2004. PMID:15353563.

Trucksis, M., Michalski, J., Deng, Y.K., and Kaper, J.B. 1998. The *Vibrio cholerae* genome contains two unique circular chromosomes. Proc. Natl. Acad. Sci. U.S.A. **95**(24): 14464–14469. doi:10.1073/pnas.95.24.14464. PMID:9826723.

Vernikos, G.S., and Parkhill, J. 2006. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. Bioinformatics, **22**(18): 2196–2203. doi:10.1093/bioinformatics/btl369. PMID:16837528.

Waack, S., Keller, O., Asper, R., Brodag, T., Damm, C., Fricke, W.F., et al. 2006. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. BMC Bioinformatics, **7**(1): 142. doi:10.1186/1471-2105-7-142. PMID:16542435.

Wattam, A.R., Williams, K.P., Snyder, E.E., Almeida, N.F., Jr., Shukla, M., Dickerman, A.W., et al. 2009. Analysis of ten *Brucella* genomes reveals evidence for horizontal gene transfer despite a preferred intracellular lifestyle. J. Bacteriol. **191**(11): 3569–3579. doi:10.1128/JB.01767-08. PMID:19346311.

Xu, Q., Dziejman, M., and Mekalanos, J.J. 2003. Determination of the transcriptome of *Vibrio cholerae* during intraintestinal growth and midexponential phase *in vitro*. Proc. Natl. Acad. Sci. U.S.A. **100**(3): 1286–1291. doi:10.1073/pnas.0337479100. PMID:12552086.

Yamaichi, Y., Iida, T., Park, K.S., Yamamoto, K., and Honda, T. 1999. Physical and genetic map of the genome of *Vibrio parahaemolyticus*: presence of two chromosomes in *Vibrio* species. Mol. Microbiol. **31**(5): 1513–1521. doi:10.1046/j.1365-2958.1999.01296.x. PMID:10200969.

Yang, Z., and Nielsen, R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. **17**(1): 32–43. PMID:10666704.

Zhang, Z., Li, J., and Yu, J. 2006*a*. Computing Ka and Ks with a consideration of unequal transitional substitutions. BMC Evol. Biol. **6**(1): 44. doi:10.1186/1471-2148-6-44. PMID:16740169.

Zhang, Z., Li, J., Zhao, X.-Q., Wang, J., Wong, G.K.-S., and Yu, J. 2006*b*. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genomics Proteomics Bioinformatics, **4**(4): 259–263. doi:10.1016/S1672-0229(07)60007-2. PMID:17531802.

Zuerner, R.L., Herrmann, J.L., and Saint Girons, I. 1993. Comparison of genetic maps for two *Leptospira interrogans* serovars provides evidence for two chromosomes and intraspecies heterogeneity. J. Bacteriol. **175**(17): 5445–5451. PMID:7690025.