Scientific Research Publishing

Advances in Microbiology, 2019, 9, 656-677
http://www.scirp.org/journal/aim
ISSN Online: 2165-3410
ISSN Print: 2165-3402

Multiple Chromosomes in Bacteria: Low Level of Evolutionary Constraint Drives the Rapid Genetic Divergence of Chromosome II

Cheramie Trahan[1], Ravi S. Pandey[2], Utkarsh Singh[2], Anushka Choudhary[1], Hyuk Cho[3], Rajeev K. Azad[2,4], Madhusudan Choudhary[1*]

[1]Department of Biological Sciences, Sam Houston State University, Huntsville, TX, USA
[2]Department of Biological Sciences, University of North Texas, Denton, TX, USA
[3]Department of Computer Science, Sam Houston State University, Huntsville, TX, USA
[4]Department of Mathematics, University of North Texas, Denton, TX, USA
Email: mchoudhary@shsu.edu

How to cite this paper: Trahan, C., Pandey, R.S., Singh, U., Choudhary, A., Cho, H., Azad, R.K. and Choudhary, M. (2019) Multiple Chromosomes in Bacteria: Low Level of Evolutionary Constraint Drives the Rapid Genetic Divergence of Chromosome II. *Advances in Microbiology*, **9**, 656-677.
https://doi.org/10.4236/aim.2019.97041

Received: June 10, 2019
Accepted: July 28, 2019
Published: July 31, 2019

Abstract

Multiple chromosomes in bacteria are designated as a larger primary chromosome (CI) and smaller accessory chromosomes (CII and CIII). Although previous studies examined multiple chromosomes in several bacterial species, the evolutionary mechanisms for the origin of CIIs still remain unclear. In this study, the four following hypotheses were tested. 1) CIIs exhibit lower sequence conservation and sequence divergence compared to their corresponding CIs across species of *Proteobacteria*. 2) The differential sequence divergence of CI and CII depends on pathogenic and non-pathogenic lifestyles. 3) CIIs harbor a higher level of horizontal gene transfers (HGTs) than CIs. 4) Orthologs located on CIIs experience less purifying selection than their corresponding orthologs on CIs. Results reveal a higher level of sequence conservation of CIs than the sequence conservation of CIIs. There is no significant difference in HGT estimates between CIs and CIIs. A majority of orthologous genes of CIs and CIIs experience purifying selection; however, genes on CIIs were significantly less constrained than the corresponding ones on CIs. This finding is true for both pathogenic and non-pathogenic bacteria, but the selective constraints for non-pathogenic bacteria are relatively less constrained. It was concluded that the differential selective constraint is a potent driving force for the rapid evolution of CII. Therefore, gene expression analysis at the transcriptome and proteome levels may shed light on the gene regulation mechanisms that might affect the sequence divergence between CI and CII.

Keywords

Multiple Chromosomes, *Proteobacteria*, Horizontal Gene Transfer, Selective Constraint, Pathogenic and Non-pathogenic Lifestyles

1. Introduction

There is a long-held paradigm that the bacterial genome is comprised of a single circular chromosome and additional dispensable non-essential plasmid(s). This view has been revised because of the existence of multiple chromosomes in several bacterial species including *Rhodobacter sphaeroides* 2.4.1 [1] [2], *Brucella melitensis* 16M [3], *Agrobacterium tumefaciens* C58 [4], *Leptospira interrogans* Verdun and RZ11 [5], *Paracoccus denitrificans* Pd1222 [6], and *Vibrio cholerae* AI1837 [7]. The explosion and impact of genome sequencing technology [8] [9] further illuminated the presence of accessory chromosomes in ten percent of the total sequenced bacterial genomes in the NCBI database, and these species were widely distributed throughout different lineages [10]. Therefore, the existence of multiple chromosomes is now an accepted paradigm of the bacterial genome structure.

Three strains of *R. sphaeroides* (2.4.1, ATCC 17025, and ATCC 17029) have been previously examined for their sequence divergences and the result indicated that genes on CIIs have lower nucleotide identity, which suggests the rapid evolution of CIIs in *R. sphaeroides* [11]. The study was further expanded among strains of both within-species and between-species of *Proteobacteria,* including species of *α*-, *β*-, and *γ*-*Proteobacteria* [12]. Out of the ten within-species comparisons, eight organisms displayed higher levels of CII-specific DNA sequence divergences. Also, trans-genera pairwise comparisons indicated an average of ~42% nucleotide identity among CIs and a significantly lower nucleotide identity of ~27% among CIIs. The previous studies further demonstrated a high divergence and rapid evolution of CII-specific sequences compared with CI-specific sequences in a majority of bacterial species [10].

The current study tests the following four hypotheses. 1) CIIs exhibit lower sequence conservation and sequence divergence compared to their corresponding CIs across species of *Proteobacteria*. 2) The differential sequence divergence of CI and CII depends on pathogenic and non-pathogenic lifestyles. 3) CIIs harbor a higher level of HGTs than CIs; thus they diverge more rapidly. 4) Orthologs located on CIIs experience less purifying selection than their corresponding those on CIs.

A total of 83 strains were analyzed within 16 individual species comparisons to test the above hypotheses. This study employs a variety of genomic and bioinformatics tools and methods, which include Mauve [13], Island Viewer [14], GEMINI [15], BLAST [16] [17], KaKs_Calculator [18], MUSCLE [19], and PAL2NAL [20].

2. Materials and Methods

2.1. Comparative Genome Alignment

One hundred bacterial strains (see Supplementary **Table S1**) with multipartite genomes were identified using the National Center for Biotechnology Information (NCBI) database, of which 83 of the strains belong to the *Proteobacterial* phylum. A total of 16 species that contain two or more strains and represent *α*-, *β*-, *γ*-subgroups of *Proteobacteria* were selected for whole genome alignment. They include 11 pathogenic and 5 non-pathogenic species (see Supplementary **Table S2**).

DNA sequence files (in FASTA format) for CI and CII from 50 strains were downloaded from the NCBI database. Mauve 2.3.1 was utilized to obtain whole genome alignments of the 16 *Proteobacterial* species CIs and CIIs, separately. Both percentage conservation from the local collinear blocks (LCBs) and percentage nucleotide identity from the entire conserved regions were extracted. A paired t-test was performed to assess the significant difference in percent conservation and percent nucleotide identity between CIs and CIIs. These results were further classified by subgroup and lifestyle. Again, a paired t-test was used to estimate the significant difference in percent conservation and percent nucleotide identity between CIs and CIIs.

2.2. Estimation of HGT

HGTs in CIs and CIIs were estimated using Island Viewer [14] which identifies genomic islands (GI) using three methods. Two of the three are sequence composition-based methods, namely, SIGI-HMM and Island Path-DIMOB. The third one, Island Pick, is a comparative genomics method that assesses putative islands by performing alignment of closely related genomes using Mauve [13]. SIGI-HMM, based on a hidden Markov model (HMM), exploits codon usage bias to discriminate between the native and GI sequences. Island Path-DIMOB recognizes GIs by common characteristics, such as mobility genes (e.g. integrases, transposases) and anomalous sequence composition. Island Viewer takes an integrated approach to increase the accuracy; it also allows access to results from each component method.

A recently published genome-mining tool, called GEMINI [15], based on a recursive segmentation and clustering procedure, was utilized to quantify HGTs. The framework recursively segments a genome sequence into two parts at each position where the compositional difference between the resulting sequence segments is maximized in terms of information-theoretic measure. The tool separates the horizontally acquired DNA segments from the vertically inherited DNA segments in a genome, performs a hypothesis testing of whole genome segmentation, and recursively applies the process for each resulting segment, until each segment becomes compositionally homogeneous. CIs and CIIs in the 83 *Proteobacterial* species were evaluated independently for the HGT segments and a paired t-test was performed to assess any significant difference in the HGT

estimates between CIs and CIIs.

2.3. Selective Constraints on Genes in Multipartite Genomes

To quantify the evolutionary divergence of orthologs shared by multiple strains of the same species, BLASTP was employed to compare genes of CI and CII separately within and across species. A total of the 50 strains representing the 16 species within α-, β-, and γ-subgroups were analyzed for further determination of the selective pressures on CI and CII, separately. Each comparison began by pre-processing which gives specific instruction for file extraction, preparation, processing, and storage. The query and target sequences in GenBank flat file format (.gbk) were downloaded from the NCBI database and converted to FASTA (.faa) format. The target files in FASTA format were organized into a local database for sequence comparison. Then, BLASTP was initiated and the query sequence was compared to all sequences in the target local database. Post-processing converted aligned results from the previous step into files available for further analysis. Next, orthologs for CIs and CIIs were identified with BLAST E-values < 10^{-20} and aminoacid identity > 30%. Orthologs satisfying these criteria were used to determine whether CI and CII have different modes of selection.

The rates of non-synonymous substitution (*Ka*), synonymous substitution (*Ks*), and the selective constraints (ω = *Ka/Ks*) were calculated using KaKs_Calculator [18], which requires DNA sequences and computes *Ka*, *Ks*, and ω, considering biases in transition to transversion ratio and the codon redundancy. The amino acid sequences corresponding to their nucleic acid sequences of orthologs were aligned using MUSCLE [19] and then reverse translated to DNA sequences using PAL2NAL [20]. The resulting nucleic acid sequences were used as an input to the KaKs_Calculator, and then *Ka*, *Ks*, and ω were calculated. Negative selection is inferred if $\omega < 0.3$, implying that the orthologs likely have maintained identical function. Neutral selection is inferred if $0.3 < \omega < 3$, suggesting orthologues are nearly equivalent and more likely to perform similar functions. Positive selection is inferred if $\omega > 3$, allowing new gene functions to evolve.

3. Results

3.1. Multipartite Genome Evolution in Prokaryotes

Chromosome size, gene count, and gene density (genes/kb) on CIs, CIIs, and CIIIs of the 100 bacterial strains are shown in Supplemental **Table S1**. The α-proteobacterial subgroup has 27 strains representing 17 species of eight different genera. The β-subgroup has 35 strains distributed over 19 species within four genera. The γ-subgroup has 21 strains consisting of 13 species within three genera. Of the 83 strains in the three subgroups of *Proteobacteria*, 50 strains, including 35 pathogenic strains representing 11 species and 15 non-pathogenic strains representing five species, are listed in Supplemental **Table S2(a)** and

Supplemental Table S2(b).

Whole genome alignments for the related strains of the species within each *α*-, *β*-, and *γ*-subgroup in *Proteobacteria* were examined. As an example of visual representation, the whole genome alignment of CIs and CIIs for the four strains of *R. sphaeroides* is depicted in Figure 1. The colored blocks represent local collinear blocks (LCBs), which are areas of conserved regions between four strains and the vertical lines link each matching LCB displaying large-scale chromosomal alterations, such as large deletions, insertions, translocations, and inversions. The alignment results indicate 80% conservation of CIs, while 27% sequence conservation of CIIs among all the four strains of *R. sphaeroides*. Also, inversions, deletions, and translocations are more abundant on CIIs, possibly contributing to the rapid divergence of CIIs in *R. sphaeroides.*

The whole genome alignment results, including the percent conserved regions and percent nucleotide identity for CIs and CIIs in the 50 strains within 16 species, are depicted in Figure 2(a) and Figure 2(b). Also, the data associated with

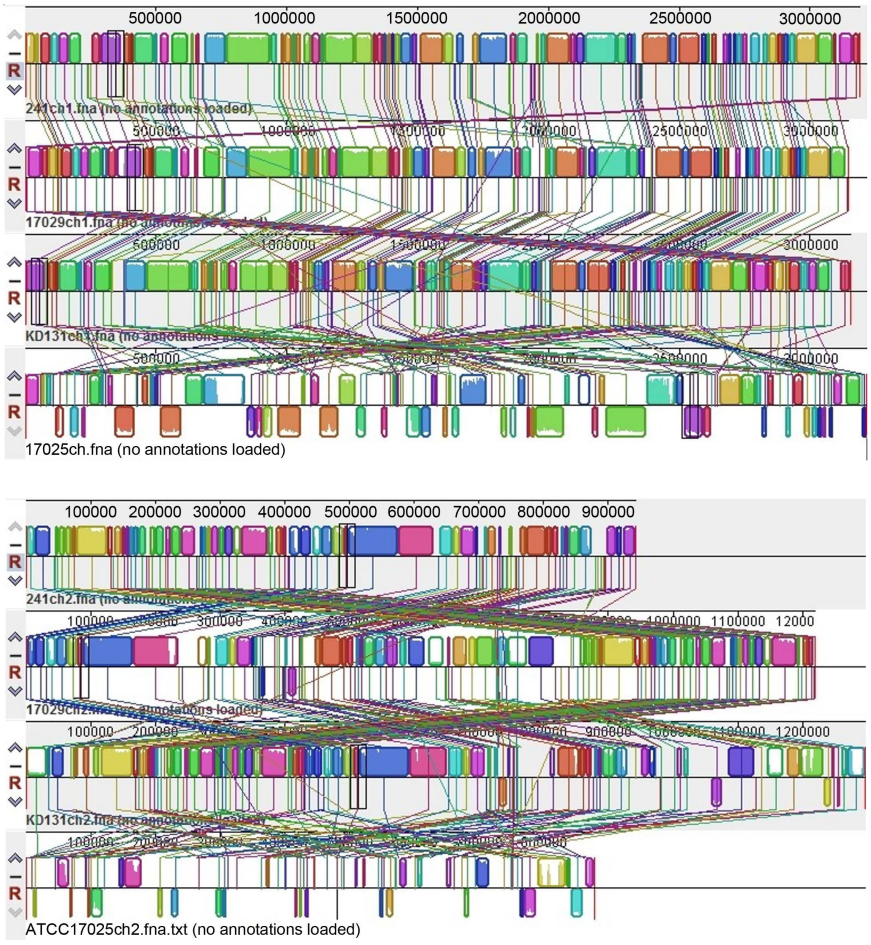

Figure 1. A visual representation of CIs (top panel) and CIIs (bottom panel) alignment of the four strains (2.4.1, ATCC 17029, KD131, and ATCC 17025) of *R. sphaeroides*. Conserved local collinear blocks (LCBs) are displayed as colored boxes among the four strains and the vertical lines link each matching LCB displaying large-scale deletions, insertions, translocations, and inversions.

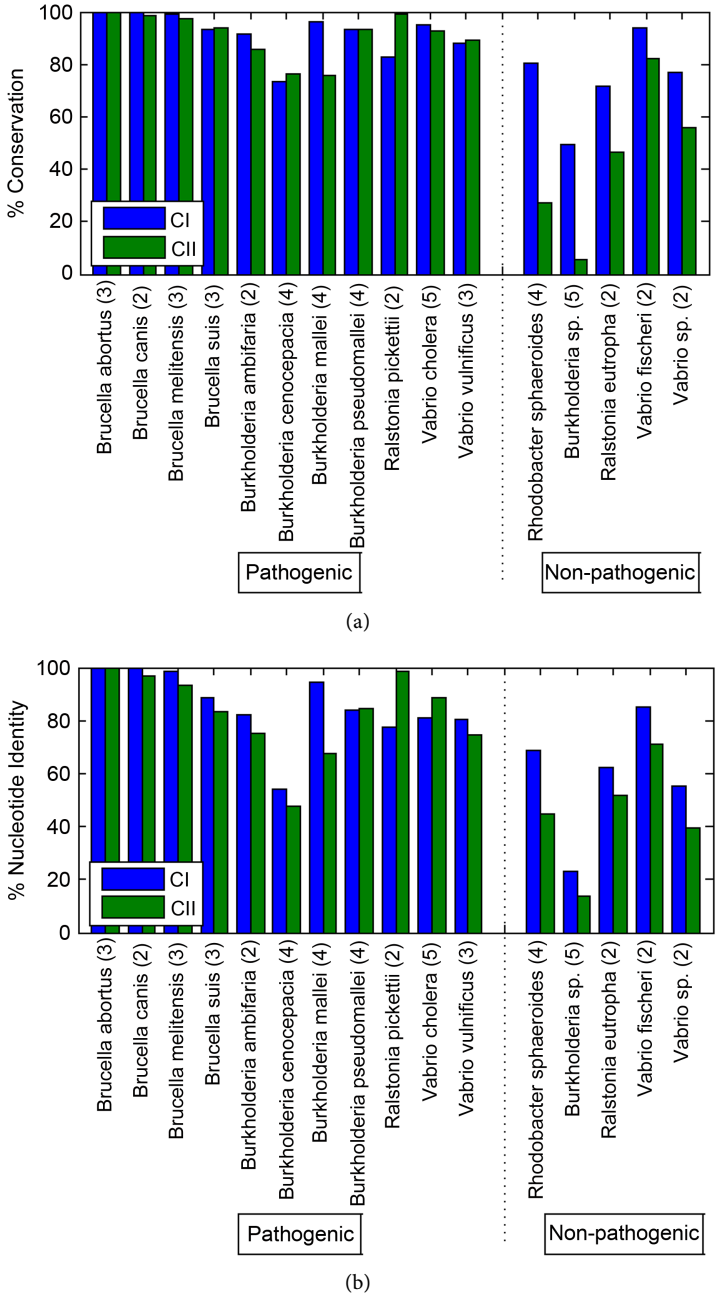

Figure 2. (a) Percent conservation of CI (in blue) and CII (in green) among the 11 pathogenic species (left) and the 5 non-pathogenic species (right). The whole genome alignment for the pathogenic strains was completed using 35 sequenced multipartite genomes within 4 genera of α-, β-, and γ-subgroups of *Proteobacteria*. Also, the whole genome alignment for the pathogenic strains was completed using 15 sequenced multipartite genomes within 4 genera. The number of strains used in each species is specified in parentheses.; (b) Percent nucleotide identity of CI and CII within the total conserved regions (Local Collinear Blocks, LCBs) among the 11 pathogenic species (left) and the 5 non-pathogenic species (right). The whole genome alignment for the pathogenic strains was completed using 35 sequenced multipartite genomes within 4 genera of α-, β-, and γ-subgroups of *Proteobacteria*. Also, the whole genome alignment for the pathogenic strains was completed using 15 sequenced multipartite genomes within 4 genera. The number of strains used in each species is specified in parentheses.

the figures are given in Supplemental Table S2(a) and Table S2(b). Figure 2 exhibits that of the 16 whole genome alignments, 11 CI alignments show higher % conservation than CII. To be more specific, 4 out of 5 species in α-subgroup have higher % conservation in CI than in CII; 4 out of 7 species in β-subgroup have higher % conservation in CI than in CII, and 3 out of 4 species in γ-subgroup have higher % conservation on CI than CII. A paired t-test was performed on all the 16 alignments, and there was a significant difference (P-Value = 0.037) in the total conserved regions on CIs compared to those on CIIs.

3.2. Sequence Conservation in Nucleotide Identity between Pathogenic and Non-Pathogenic Life-Styles

The impact of lifestyle on sequence conservation was also examined by classifying the species as pathogens and non-pathogens. The 11 pathogenic strains in Figure 2(a) show that sequence conservation between CIs and CIIs was not significantly different (Paired t-test, P-Value = 0.690). However, among the five non-pathogenic strains also shown in Figure 2(a), sequence conservation for CIs were significantly higher than that for CIIs (Paired t-test, P-Value = 0.016). These results indicate that CIIs in non-pathogenic strains within α-, β-, and γ-subgroups in *Proteobacteria* have diverged more rapidly than CIs.

All the 16 whole genome alignments also revealed that the percent nucleotide identity of the conserved regions in CII was significantly lower than that of the CI conserved regions (Paired t-test, P-Value = 0.036), as shown in Figure 2(b). The percent nucleotide identity between pathogenic and non-pathogenic species was shown in Figure 2(b). In the pathogenic bacterial strains, percent nucleotide identity for CIs is not significantly different from that for CIIs (Paired t-test, P-Value = 0.437). However, percent nucleotide identity for CIs is significantly higher than that for CIIs in the non-pathogenic bacteria (Paired t-test, P-Value = 0.005).

3.3. Horizontal Gene Transfer Estimates for CIs and CIIs

Island Viewer estimates of HGT among the 83 strains within the α-, β-, and γ-subgroups of *Proteobacteria* are listed in Supplemental Table S3. There is no significant difference in HGT estimates between CIs and CIIs (Paired t-test, P-Value = 0.790). In each subgroup, the following number of strains had higher HGT estimates on CI compared to CII: 17 out of 27 strains in α-subgroup, 19 out of 35 strains in β-subgroup, and 12 out of 21 strains in γ-subgroup. The result of paired t-test indicated no significant difference in the HGT levels between CIs and CIIs within every subgroup as follows: α-subgroup (P-Value = 0.707), β-subgroup (P-Value = 0.793), and γ-subgroup (P-Value = 0.857). For both α- and β-subgroups, the average HGT estimates are slightly higher in CIs than in CIIs. However, in γ-subgroup, the average HGT estimate for CIs is slightly lower than that for CIIs, but not significantly different. Furthermore, HGT estimates for the 83 strains were compared between the two subcategories: pathogens and

non-pathogens. There was no significant difference in the HGT estimate between CIs and CIIs for the pathogenic strains (P-Value = 0.407) and the non-pathogenic strains (P-Value = 0.544).

The HGT estimates for CI and CII of *R. sphaeroides* were obtained from a recently proposed integrated segmentation and clustering method, GEMINI and they are schematically represented in Figure 3. Furthermore, the different HGT estimates of α-, β-, and γ-subgroups of *Proteobacteria* are shown in Figure 4. A total of 48 out of 76 strains had higher HGT estimates (% nucleotides labeled "alien") on CIs than those on CIIs as follows: α-subgroup (16 out of 24 strains), β-subgroup (21 out of 34 strains), and γ-subgroup (11 out of 18 strains). A paired t-test result indicated no significant difference in HGT estimates between CI and CII (P-Value = 0.180), although this difference was more pronounced than that observed using Island Viewer (P-Value = 0.790). As observed with Island Viewer, the difference within every subgroup was also not significant as P-values for α-subgroup, β-subgroup and γ-subgroup were 0.4231, 0.4064 and 0.4354, respectively. Together, these results suggest that HGT events are not a significant player in the rapid evolution of accessory chromosomes in *Proteobacteria.*

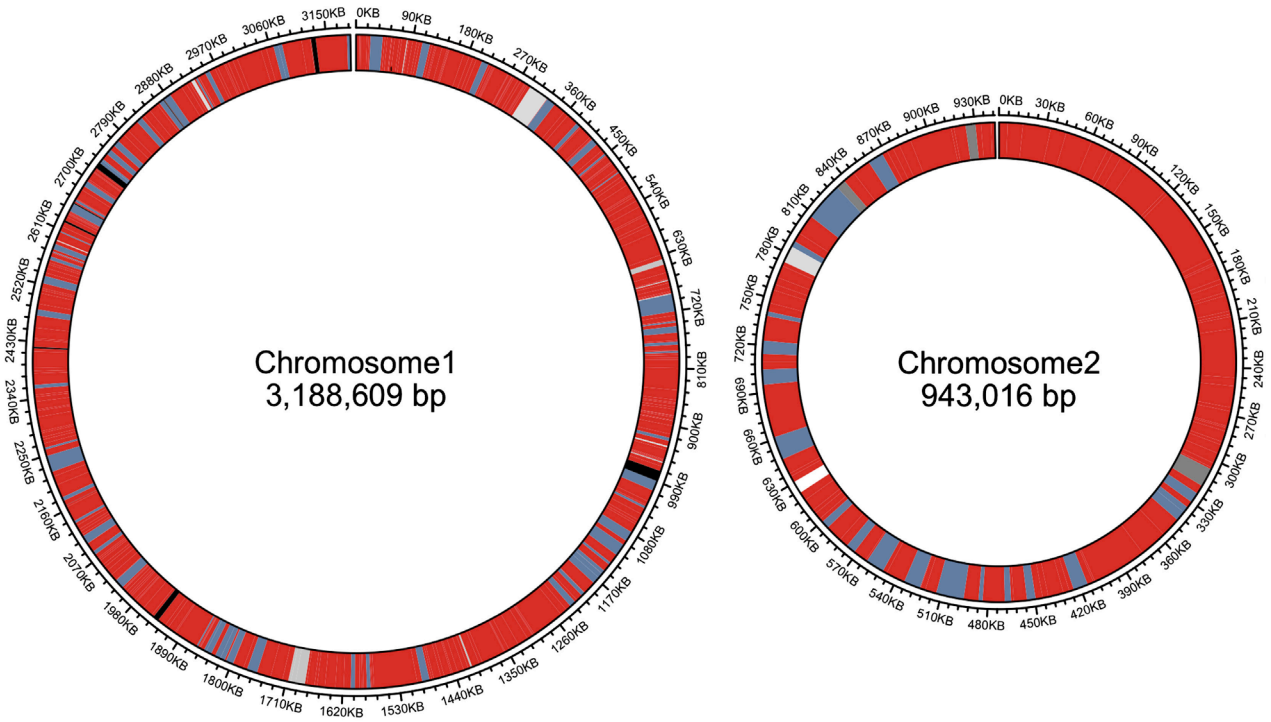

Figure 3. A schematic representation of HGT estimates within chromosomes I (CI) and II (CII) of *Rhodobacter sphaeroides* 2.4.1. HGT estimates are obtained by employing an integrated segmentation and clustering method. Each chromosome contains both native sequence segment in red color and alien sequence segments in blue, green, and yellow colors. Gene clusters are organized with their corresponding coordinates in CI and CII. CI contains 153 segments, of which 67 segments represented a single large native cluster in red color and the remaining 86 segments represented 16 different gene clusters in different colors. CII contains 45 segments, of which 22 segments are in a single large native gene cluster in red color and the remaining 23 segments representing four alien gene clusters in different colors.

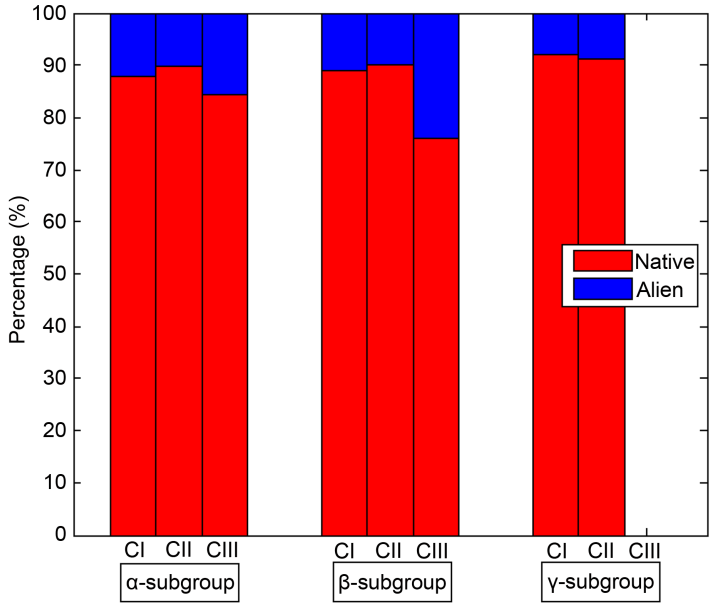

Figure 4. Proportion of vertically inherited ("Native") and horizontally acquired ("Alien") DNAs within the chromosomes of *α*-, *β*-, and *γ*-subgroup *Proteobacteria*, deciphered using an integrative method of segmentation and clustering. Both segmentation and clustering were performed within a statistical hypothesis framework with significance levels for segmentation and two-stage clustering set at 10^{-5}, 10^{-6} and 10^{-7}, respectively. Both segmentation and clustering procedures employed Markovian Jensen-Shannon divergence measure for quantifying difference between DNA sequences. Homogeneous Markov models of order 2 were used for analysis.

3.4. Selective Constraints of CI and CII

Orthologs present on CI and CII within the 50 strains were identified and selective constraint analysis of CIs and CII of the 16 species was performed. The number of orthologs between each species, their multi-way comparisons, and their average *Ka*, *Ks*, and ω values are presented in Supplemental **Table S4**. In all the 16 species, the majority of the orthologs on CI and CII are under negative selection as shown in **Figure 5**; however, the average selective constraint value for CI ($\omega = 0.124$) was lower than that for CII ($\omega = 0.196$). T-test indicated a significant difference between CI and CII selective pressures among the 16 species (P-Value = 0.038).

4. Discussions

4.1. Differential Evolution of CI and CII in Prokaryotes

Mauve analysis revealed that sequence conservation and DNA sequence divergence levels for CIs and CIIs are significantly different in within-species comparisons. Thirteen out of sixteen such comparisons revealed that CIIs have a higher nucleotide divergence than their corresponding CIs, and thus CIIs evolved faster than CIs. However, in species *Ralstonia pickettii* CII is highly conserved and CII might have originated more recently in this species and may not have had enough time to diverge. Alternatively, genes on CII in this species

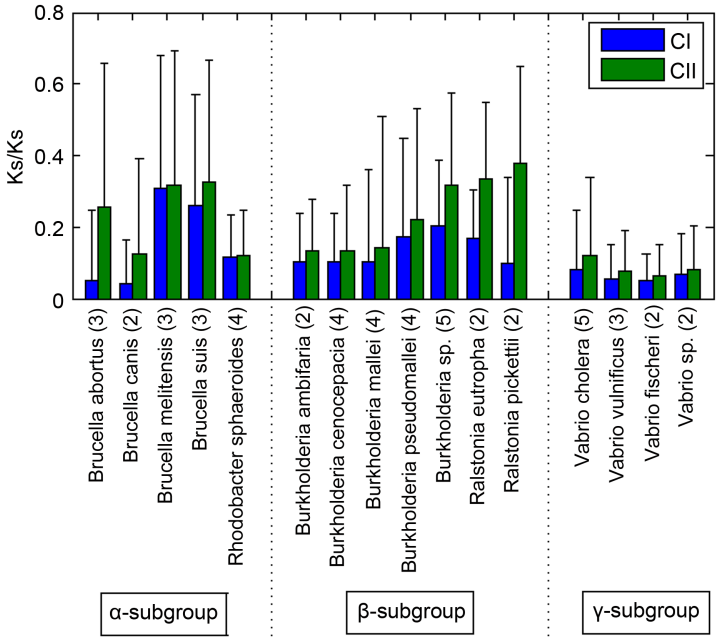

Figure 5. Average selective constraint (ω) value of CI (in blue) and CII (in green) in the α-, β-, and γ-subgroups of *Proteobacteria.* The number of strains used in each species comparison is specified in parentheses. Pairwise comparisons of orthologs located on the corresponding CI and CII across species compute the rates of non-synonymous substitution (*Ka*), synonymous substitution (*Ks*). The selective constraints ($\omega = Ka/Ks$) are averaged.

are functionally constrained and possibly involved with the species' pathogenic lifestyle. Also, no significant differences in CI and CII sequence divergence exist in other pathogenic strains examined in this study. This finding suggests that a higher level of purifying selection operates on the genomes of pathogenic strains, allowing organisms to retain CII encoded essential functions that aid in pathogenicity.

In contrast, there is a significant difference in genetic divergences between CI and CII in non-pathogenic bacteria. The rapid evolution of CII in the majority of bacterial species, especially in non-pathogenic strains, is possibly due to relaxed selection on CII compared to CI and/or more frequent HGTs on CII than on CI.

4.2. Similar HGT Estimates in CI and CII

HGT plays an important role in the evolution of bacterial genomes [21], metabolic functions [22], and re-patterning of the regulation mechanisms [23]. Since CIs are longer than CIIs, we presume that CIs may have higher HGT estimates than CIIs. Alternatively, as CIs possess more essential genes than CIIs, CIs are evolutionarily more constrained than CIIs. Therefore, in spite of the greater length of CIs, smaller CIIs may have a higher number of HGT estimates. Results of this study do not exhibit a significant difference of HGT estimates between CIs and CIIs across the bacterial species examined. Therefore, HGT may not

appear to be a major mechanism contributing to the rapid evolution of CII in *Proteobacteria*. Even when the genomes were examined separately by pathogenic and non-pathogenic group, the results still indicated no difference in HGT levels between CI and CII. The role of HGT in the initial phase of origin and development of the accessory chromosome may be important; however, it may not contribute much through remaining evolutionary processes.

Since HGT is a continuing process, the amount of HGTs present in replicon is possibly indicative of the relative evolutionary history. The high HGT estimates in certain CIIs could suggest a recent origin of the accessory chromosome, while similar HGT estimates on CI and CII may suggest ancient origin of CII. In the case of *R. sphaeroides*, CI and CII have similar HGT estimates. However, three out of the four strains have slightly higher HGT estimates on CII. The oldest strain ATCC 17025 has higher HGT estimates on CI, which supports the theory of the ancient origin of CII [10] [11], while the other three strains that emerged later display higher HGT estimates on CII than on CI. The older CIIs may have lost a substantial fraction of earlier acquired alien DNAs. It is also possible that the old alien genes may have ameliorated their composition to that of the native genome and therefore may not be detected by Island Viewer. On the other hand, foreign genes in recently originated CIIs might not have substantially ameliorated yet and therefore they may be identified by Island Viewer. Furthermore, among all the three subgroups, α-subgroup displayed slightly higher HGT estimate for CIthan CII.

4.3. CI and CII Are under Purifying Selection

Selective constraint (ω) is a measure of the ratio of the synonymous substitution rate (K_s) and nonsynonymous substitution rate (K_a). It signifies the mode of selective pressures [24]: negative, neutral, or positive selection. The majority of orthologous gene pairs located on both chromosomes have selective constraint (ω) values < 0.3, which demonstrates that regardless of their locations, these orthologs were maintained under purifying selection. Although both chromosomes are under purifying selection, selective constraints on CII are more relaxed than that on CI. The differential selective pressure on multipartite genomes supports the rapid divergence of CII. As genes on the accessory chromosomes experience weaker negative selection because they are possibly semi-essential, not always expressed, redundantly connected within gene network, and/or more susceptible to mutation [25]; thus, they diverge more rapidly than their corresponding primary chromosomes. Previous studies have demonstrated that the synonymous codon usage orderliness (SCUO) was significantly less on CII than on CI [26]. This codon usage bias is reflective of the decreased purifying selection on CII. Additionally, codon usage bias varied greatly among genomes and was highly dependent on their G + C composition. Furthermore, as the reduced SCUO is an inherent attribute of genes on CII, they experience reduced selection for translation efficiency because of either their reduced expression or greater protein dispensability [27]. CIIs of most multipartite ge-

nomes harbor more nonessential genes including genes of hypothetical protein and unknown functions. This suggests that the nonessential genes possibly evolve faster and therefore they are much more suited for biodiversity including in the emergence of new strains and species.

4.4. Evolutionary Mechanisms for the Origin of CIIs

Two different models have been previously invoked to explain the origin of CIIs in bacteria [10]. In the first model, when a native or a newly captured plasmid from another species secures some essential genes via intra- or inter-genomic gene transfer from CI, this plasmid becomes non-dispensable for bacterial survival an growth under all growth conditions. In this scenario, the presence of a plasmid type of replication-origin and some essential genes for the cell survival will be preserved in the cell. In the second model, CI in the ancestral cell breaks apart to form two chromosomes of unequal sizes. As the two newly formed chromosomes still maintain a chromosomal origin of replication and many essential or important gene functions, they are stably retained in the cell. For example, the origin of second linear chromosome of *Agrobacterium tumefaciens* C58 as well as CII (pSymA) and CIII (pSymB) of *Sinorhizobium meliloti* is supported by the plasmid origin model, while the second circular chromosome of *Brucella melitensis* 16M and *Rhodobacter sphaeroides* 2.4.1 revealed that CIIs originated from their ancestral primary chromosomes.

4.5. Evolutionary Role of CIIs

The multipartite genome structure in bacteria has three advantages, which include shortening the genome replication time, specializing the CII for a specific or pathogenic environment, and evolving new metabolic functions for generating species diversity. The first advantage is to reduce the duration of the genome replication and cell doubling or cell cycle, when a large primary chromosome splits into two chromosomes; therefore, the multipartite genome structure helps bacteria to cope with the nutrient-rich environments. The second advantage is to differentially regulate the expression of genes on different chromosomes under a specific host or free-living condition. It was validated that in *Vibrio cholera*, a pathogenic bacterium, CII-specific genes involved in various metabolisms and pathogenicity are highly expressed under pathogenic conditions, while CI-specific genes are expressed at similar levels under both free-living and pathogenic conditions [10]. The current study has shown that the orthologs located on CII compared to the corresponding orthologs present on CI are structurally and functionally less constrained. Therefore, CII-specific genes may evolve more rapidly and thus generate high level of strain or species diversity within a specific group of bacteria.

5. Conclusion

This study concludes that CII-specific sequences have higher sequence diver-

gence than their corresponding CI-specific sequences, and therefore CII evolved rapidly. The results of previous and current studies suggest that both differential selective constraint and pathogenic/non-pathogenic lifestyle seem to be the major driving force for the genetic divergence between the primary and the accessory chromosomes. Although HGTs do not seem to be a major factor for the differential sequence divergences of CIs and CIIs, the role of HGT for genome evolution should not be underestimated based on the limited number of species examined. Future studies focused on transcriptome and proteome analyses will help establish if gene orthologs located on primary and accessory chromosomes are differentially expressed for adaptation in specific host-environments or free-living growth conditions.

Acknowledgements

We thank the Department of Biological Sciences of Sam Houston State University for financial support. We also thank Lauren Flores for proofreading the manuscript.

Authors Contribution

Conceptualization, M. C., H. C., and R. K. A.; methodology and investigation, C. T., R. S. P., H. C., M. C., and R. K. A.; formal analysis, C. T., H.C., M. C., and R. K. A.; manuscript preparation, M. C., H. C., U. S., and R. K. A

Conflicts of Interest

The authors declare no conflict of interest regarding the publication of this paper.

References

[1] Suwanto, A. and Kaplan, S. (1989) Physical and Genetic Mapping of the *Rhodobacter sphaeroides* 2.4.1 Genome: Genome Size, Fragment Identification, and Gene Localization. *Journal of Bacteriology*, **171**, 5840-5849. https://doi.org/10.1128/jb.171.11.5840-5849.1989

[2] Suwanto, A. and Kaplan, S. (1989) Physical and Genetic Mapping of the *Rhodobacter sphaeroides* 2.4.1 Genome: Presence of Two Unique Circular Chromosomes. *Journal of Bacteriology*, **171**, 5850-5859. https://doi.org/10.1128/jb.171.11.5850-5859.1989

[3] 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. *Journal of Bacteriology*, **175**, 701-705. https://doi.org/10.1128/jb.175.3.701-705.1993

[4] 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. *Journal of Bacteriology*, **175**, 7869-7874. https://doi.org/10.1128/jb.175.24.7869-7874.1993

[5] Zuerner, R.L., Herrmann, J.L. and Girons, I.S. (1993) Comparison of Genetic Maps for Two *Leptospira interrogans serovars* Provides Evidence for Two Chromosomes

and Intraspecies Heterogeneity. *Journal of Bacteriology*, **175**, 5445-5451. https://doi.org/10.1128/jb.175.17.5445-5451.1993

[6] Winterstein, C. and Ludwig, B. (1998) Genes Coding for Respiratory Complexes Map on All Three Chromosomes of the *Paracoccus denitrificans* Genome. *Archives of Microbiology*, **169**, 275-281. https://doi.org/10.1007/s002030050572

[7] Trucksis, M., Michalski, J., Deng, Y.K. and Kaper, J.B. (1998) The *Vibrio cholerae* Genome Contains Two Unique Circular Chromosomes. *PNAS*, **95**, 14464-14469. https://doi.org/10.1073/pnas.95.24.14464

[8] Schuster, S.C. (2007) Next-Generation Sequencing Transforms Today's Biology. *Nature Methods*, **5**, 16-18. https://doi.org/10.1038/nmeth1156

[9] Mardis, E.R. (2008) The Impact of Next-Generation Sequencing Technology on Genetics. *Trends in Genetics*, **24**, 142-149. https://doi.org/10.1016/j.tig.2007.12.007

[10] Choudhary, M., Cho, H., Bavishi, A., Trahan, C. and Myagmarjav, B. (2012) Evolution of Multipartite Genomes in Prokaryotes. In: Pontarotti, P., Ed., *Evolutionary Biology*: *Mechanisms and Trends*, Springer, New York, 301-323. https://doi.org/10.1007/978-3-642-30425-5_17

[11] 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. *Journal of Bacteriology*, **189**, 1914-1921. https://doi.org/10.1128/JB.01498-06

[12] Bavishi, A., Abhishek, A., Lin, L. and Choudhary, M. (2010) Complex Prokaryotic Genome Structure: Rapid Evolution of Chromosome II. *Genome*, **53**, 675-687. https://doi.org/10.1139/G10-046

[13] Darling, A.C.E., Mau, B., Blattner, F.R. and Perna, N.T. (2004) Mauve: Multiple Alignment of Conserved Genomic Sequence with Rearrangements. *Genome Research*, **14**, 1394-1403. https://doi.org/10.1101/gr.2289704

[14] Langille, M.G. and Brinkman, F.S. (2009) Island Viewer: An Integrated Interface for Computational Identification and Visualization of Genomic Islands. *Bioinformatics*, **25**, 664-665. https://doi.org/10.1093/bioinformatics/btp030

[15] Azad, R.K., Jani, M. and Mathee, K. (2016) Identification of Novel Genomic Islands in Liverpool Epidemic Strain of *Pseudomonas aeruginosa* Using Segmentation and Clustering. *Frontiers in Microbiology*, **7**, 1210. https://doi.org/10.3389/fmicb.2016.01210

[16] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Research*, **25**, 3389-3402. https://doi.org/10.1093/nar/25.17.3389

[17] Zhou, Y. and Landweber, L.F. (2007) BLASTO: A Tool for Searching Orthologous Groups. *Nucleic Acids Research*, **35**, W678-W682. https://doi.org/10.1093/nar/gkm278

[18] Zhang, Z., Li, J., Zhao, X., Wang, J., Ka-ShuWong, G. and Yu, J. (2006) KaKs Calculator: Calculating Ka and Ks through Model Selection and Model Averaging. *Genomics*, *Proteomics*, *& Bioinformatics*, **4**, 259-263. https://doi.org/10.1016/S1672-0229(07)60007-2

[19] Edgar, R.C. (2004) MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput**.** *Nucleic Acids Research*, **32**, 1792-1797. https://doi.org/10.1093/nar/gkh340

[20] Suyama, M., Torrents, D. and Bork, P. (2006) PAL2NAL: Robust Conversion of

Protein Sequence Alignments into the Corresponding Codon Alignments. *Nucleic Acids Research*, **34**, W609-W612. https://doi.org/10.1093/nar/gkl315

[21] Jain, R., Rivera, M.C., Moore, J.E. and Lake, J.A. (2003) Horizontal Gene Transfer Accelerates Genome Innovation and Evolution. *Molecular Biology and Evolution*, **20**, 1598-1602. https://doi.org/10.1073/pnas.96.7.3801

[22] Caro-Quintero, A. and Konstantinidis, K.T. (2014) Inter-Phylum HGT Has Shaped the Metabolism of Many Mesophilic and Anaerobic Bacteria. *The ISME Journal*, **9**, 958-967. https://doi.org/10.1038/ismej.2014.193

[23] Price, M.N., Dehal, P.S. and Arkin, A.P. (2008) Horizontal Gene Transfer and the Evolution of Transcriptional Regulation in *Escherichia coli*. *Genome Biology*, **9**, R4. https://doi.org/10.1186/gb-2008-9-1-r4

[24] Nei, M. (1988) Relative Roles of Mutation and Selection in the Maintenance of Genetic Variability. *Philosophical Transactions of the Royal Society B*, **319**, 615-629. https://doi.org/10.1098/rstb.1988.0069

[25] Pál, C., Papp, B. and Lercher, M.J. (2006) An Integrated View of Protein Evolution. *Nature Reviews Genetics*, **7**, 337-348. https://doi.org/10.1038/nrg1838

[26] Cooper, V.S., Vohr, S.H., Wrocklage, S.C. and Hatcher, P.J. (2010) Why Genes Evolve Faster on Secondary Chromosomes in Bacteria. *PLOS Computational Biology*, **6**, e1000732. https://doi.org/10.1371/journal.pcbi.1000732

[27] Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarranton, G., Stephens, P., Millican, A., Eaton, M. and Humphreys, G. (1984) Codon Usage Can Affect Efficiency of Translation of Genes in *Escherichia coli*. *Nucleic Acids Research*, **12**, 6663-6671. https://doi.org/10.1093/nar/14.19.7737

Supplementary

Table S1. Chromosomes size, gene count, and gene density (genes/kb) on CI, CII, and CIII of 100 bacterial strains. All chromosomes are circular in structure unless denoted with an (L) which are linear in structure.

Species	CI			CII			CIII		
	Size (bp)	Gene Count	Genes/Kb	Size (bp)	Gene Count	Genes/Kb	Size (bp)	Gene Count	Genes/Kb
Acidobacteria									
Candidatus Chloracidobacterium thermophilum B	2,683,362	2273	0.847	1,012,010	837	0.827			
Bacteroidetes									
Prevotella intermedia 17	2,119,790	1947	0.918	579,647	513	0.885			
Prevotella melaninogenica ATCC 25845	1,796,408	1409	0.784	1,371,874	992	0.723			
Chloroflexi									
Sphaerobacter thermophilus DSM 20745	2,741,033	2517	0.918	1,252,731	1,068	0.853			
Cynobacteria									
Cyanothece sp. ATCC 51142	4,934,271	4821	0.977	429701 (L)	451	1.050			
Deinococcus-Thermus									
Deinococcus radiodurans R1	2,648,638	2687	1.014	412,348	369	0.895			
Firmicutes									
Butyrivibrio proteoclasticus B316	3,554,804	2998	0.843	302,358	259	0.857			
Clostridium difficile BI1	4,118,573	3690	0.896	300,869	281	0.934			
α-proteobacteria									
Agrobacterium radiobacter K84	4,005,130	3852	0.960	2,650,913	2382	0.900			
Agrobacterium sp H13-13	2,823,930	2799	0.990	2,148,289 (L)	1967	0.920			
Agrobacterium tumefaciens C58	2,841,490	2835	1.000	2,075,560 (L)	1899	0.910			
Agrobacterium vitis S4	3,726,375	3313	0.890	1,283,187	1077	0.840			
Asticcacaulis excentricus CB 48	2,588,221	2409	0.930	1,315,949	1177	0.890			
Brucella abortus A13334	2,123,773	2231	1.050	1,162,259	1170	1.010			
Brucella abortus S19	2,122,487	2056	0.970	1,161,449	1077	0.930			
Brucella abortus biovar 1 9-941	2,124,241	2199	1.040	1,162,204	1156	0.990			
Brucella canis ATCC 23365	2,105,969	2200	1.040	1,206,800	1218	1.010			
Brucella canis HSK A52141	2,107,023	2174	1.030	1,170,489	1168	1.000			
Brucella melitensis ATCC 23457	2,125,701	2181	1.030	1,185,518	1166	0.980			
Brucella melitensis biovar Abortus 2308	2,121,359	2236	1.050	1,156,948	1182	1.020			
Brucella melitensis 16M	2,117,144	2107	1.000	1,177,787	1157	0.980			
Brucella microti CCM 4915	2,117,050	2201	1.040	1,220,319	1209	0.990			
Brucella ovis ATCC 25840	2,111,370	2094	0.990	1,164,220	1104	0.950			
Brucella suis 1330	2,107,794	2230	1.060	1,207,381	1217	1.010			
Brucella suis ATCC 23445	1,923,763	2008	1.040	1,400,844	1418	1.010			
Brucella suis VBI22	2,108,637	2228	1.060	1,207,451	1210	1.000			
Ochrobactrum anthropi ATCC 49188	2,887,297	2803	0.970	1,895,911	1725	0.910			
Paracoccus denitrificans PD1222	2,852,282	2874	1.010	1,730,097	1683	0.970			

Continued

Rhodobacter sphaeroides 2.4.1	3,188,609	3092	0.970	943,016	866	0.920			
Rhodobacter sphaeroides ATCC 17025	3,217,726	3181	0.990	877,879	849	0.970			
Rhodobacter sphaeroides ATCC 17029	3,147,721	3036	0.960	1,219,053	1100	0.900			
Rhodobacter sphaeroides KD131	3,152,792	3145	1.000	1,297,642	1249	0.960			
Sinorhizobium meliloti AK83	3,820,344	3762	0.980	1,680,879	1578	0.940	1,312,480	1335	1.017
Sphingobium chlorophenolicum L-1	3,080,818	2970	0.960	1,368,670	1171	0.860			
Sphingobium japonicum UT26S	3,514,822	3586	1.020	681,892	599	0.880			
β-proteobacteria									
Burkholderia ambifaria AMMD	3,556,545	3287	0.920	2,646,969	2364	0.890	1,281,472	1021	0.797
Burkholderia ambifaria MC40 6	3,443,585	3173	0.920	2,769,416	2422	0.870	1,127,947	990	0.878
Burkholderia cenocepacia AU 1054	3,294,563	3054	0.930	2,788,459	2504	0.900	1,196,094	1074	0.898
Burkholderia cenocepacia HI2424	3,483,902	3239	0.930	2,998,664	2707	0.900	1,055,417	927	0.878
Burkholderia cenocepacia J2315	3,870,082	3544	0.920	3,217,062	2816	0.880	875,977	779	0.889
Burkholderia cenocepacia MC0 3	3,532,883	3260	0.920	3,213,911	2837	0.880	1,224,595	1070	0.874
Burkholderia gladioli BSR3	4,413,616	3964	0.900	3,700,833	3006	0.810			
Burkholderia glumae BGR1	3,906,529	3493	0.890	2,827,355	2286	0.810			
Burkholderia mallei ATCC 23344	3,510,148	3393	0.970	2,325,379	2115	0.910			
Burkholderia mallei NCTC 10229	3,458,208	3408	0.990	2,284,095	2215	0.970			
Burkholderia mallei NCTC 10247	3,495,687	3537	1.010	2,352,693	2388	1.020			
Burkholderia mallei SAVP1	3,497,479	3531	1.010	1,734,922	1763	1.020			
Burkholderia multivorans ATCC 17616	3,448,466	3224	0.930	2,472,928	2169	0.880	919,805	810	0.881
Burkholderia phymatum STM815	3,479,187	3167	0.910	2,697,374	2450	0.910			
Burkholderia phytofirmans PsJN	4,467,537	4051	0.910	3,625,999	3265	0.900			
Burkholderia pseudomallei 1106a	3,988,455	4081	1.020	3,100,794	3174	1.020			
Burkholderia pseudomallei 1710b	4,126,292	3798	0.920	3,181,762	2621	0.820			
Burkholderia pseudomallei 668	3,912,947	4003	1.020	3,127,456	3191	1.020			
Burkholderia pseudomallei K96243	4,074,542	3529	0.870	3,173,005	2406	0.760			
Burkholderia sp 383	3,694,126	3414	0.920	3,587,082	3188	0.890	1,395,069	1221	0.875
Burkholderia sp CCGE1001	4,063,449	3669	0.900	2,770,302	2489	0.900			
Burkholderia sp CCGE1002	3,518,940	3254	0.920	2,593,966	2404	0.930	1,282,816	1204	0.939
Burkholderia sp CCGE1003	4,077,097	3596	0.880	2,966,498	2593	0.870			
Burkholderia sp KJ006	3,145,156	2998	0.953	2,356,985	2151	0.913	1,082,410	949	0.877
Burkholderia sp Y123	3,131,280	2837	0.910	1,773,019	1546	0.870	1,569,570	1370	0.873
Burkholderia rhizoxinica HKI 454	2,755,309	2949	1.070	822,304	784	0.950			
Burkholderia thailandensis E264	3,809,201	3343	0.880	2,914,771	2369	0.810			
Burkholderia vietnamiensis G4	3,652,814	3387	0.930	2,411,759	2148	0.890	1,241,007	1171	0.944
Burkholderia xenovorans LB400	4,895,836	4608	0.940	3,363,523	3047	0.910	1,471,779	1388	0.943
Cupriavidus necator N-1	3,872,936	3701	0.960	2,684,606	2468	0.920			
Cupriavidus taiwanensis LMG 19424z	3,416,911	3220	0.940	2,502,411	2268	0.910			
Ralstonia eutropha H16	4,052,032	3711	0.920	2,912,490	2568	0.880			
Ralstonia eutropha JMP134	3,806,533	3529	0.930	2,726,152	2442	0.900			

Continued

Ralstonia pickettii 12D	3,647,724	3521	0.970	1,323,321	1212	0.920
Ralstonia pickettii 12J	3,942,557	3817	0.970	1,302,238	1192	0.920
Variovorax paradoxus S110	5,626,353	5358	0.950	1,128,644	1095	0.970
γ-proteobacteria						
Aliivibrio salmonicida LFI1238	3,325,165	3199	0.960	1,206,461	1115	0.920
Photobacterium profundum SS9	4,085,304	3606	0.880	2,237,943	2029	0.910
Pseudoalteromonas haloplanktis TAC125	3,214,944	3074	0.960	635,328	546	0.860
Pseudoalteromonas sp. SM9913	3,332,787	3179	0.950	704,884	637	0.900
Vibrio anguillarum 775	3,063,912	2916	0.950	988,135	922	0.930
Vibrio cholerae M66-2	2,892,523	2784	0.960	1,046,382	1055	1.010
Vibrio cholerae MJ-1236	3,149,584	2897	0.920	1,086,784	1013	0.930
Vibrio cholerae O1 biovar El Tor str. N16961	2,961,149	2888	0.980	1,072,315	1119	1.040
Vibrio cholerae O1 str. 2010EL-1786	3,031,375	2902	0.960	1,046,365	1044	1.000
Vibrio cholerae O395	3,024,069	2861	0.950	1,108,250	1137	1.030
Vibrio fischeri ES114	2,897,536	2738	0.940	1,330,333	1192	0.900
Vibrio fischeri MJ11	2,905,029	2712	0.930	1,418,848	1268	0.890
Vibrio furnissii NCTC 11218	3,294,546	3124	0.950	1,621,862	1454	0.900
Vibrio harveyi ATCC BAA-1116	3,765,351	3708	0.980	2,204,018	2411	1.090
Vibrio parahaemolyticus RIMD 2210633	3,288,558	3223	0.980	1,877,212	1769	0.940
Vibrio sp EJY3	3,478,307	3210	0.920	1,974,339	1727	0.870
Vibrio sp. Ex25	3,259,580	3060	0.940	1,829,445	1624	0.890
Vibrio splendidus LGP32	3,299,303	3089	0.940	1,675,515	1514	0.900
Vibrio vulnificus CMCP6	3,281,944	3019	0.920	1,844,853	1553	0.840
Vibrio vulnificus MO6-24/O	3,194,232	3105	0.970	1,813,536	1596	0.880
Vibrio vulnificus YJ016	3,354,505	3387	1.010	1,857,073	1711	0.920
Spirochaetes						
Leptospira biflexa serovar Patoc strain 'Patoc 1 (Ames)'	3,603,977	3351	0.930	277,995	267	0.960
Leptospira biflexa serovar Patoc strain 'Patoc 1 (Paris)'	3,599,677	3440	0.956	277,655	276	0.994
Leptospira borgpetersenii serovar Hardjo-bovis L550	3,614,446	3003	0.831	317,336	270	0.851
Leptospira borgpetersenii serovar Hardjo-bovis JB197	3,576,473	2980	0.833	299,762	262	0.874
Leptospira interrogans serovar Lai str. 56601	4,338,762	3457	0.797	359,372	293	0.815
Leptospira interrogans serovar Lai str. IPAV	4,349,158	3466	0.797	359,372	293	0.815
Leptospira interrogans serovar Copenhageni str. Fiocruz L1-130	4,277,185	3486	0.815	350,181	276	0.788
Thermobaculum.						
Thermobaculum terrenum ATCC BAA-798	2,026,947	1937	0.956	1,074,634	996	0.927

Table S2. (a) Percent conservation and percent nucleotide identy of CI and CII among pathogenic species; (b) Percent conservation and percent nucleotide identy of CI and CII among non-pathogenic species.

(a)

Subgroup/Species	Chromosome I				Chromosome II			
	LCB Number	Length of common LCBs	% Conservation	% Nucleotide Identity	LCB Number	Length of common LCBs	% Conservation	% Nucleotide Identity
α-proteobacteria								
Brucella abortus (3)	2	2,124,461	99.93	99.88	2	1,162,418	99.94	99.89
Brucella canis (2)	2	2,107,122	99.97	99.94	2	1,206,868	98.47	96.97
Brucella melitensis (3)	8	2,133,234	99.24	98.46	8	1,218,919	97.20	93.35
Brucella suis (3)	7	2,142,622	92.99	88.59	1	1,429,644	93.72	83.25
β-proteobacteria								
Burkholderia ambifaria (2)	24	3,781,665	91.51	82.44	47	2,961,148	85.42	75.39
Burkholderia cenocepacia (4)	16	4,538,164	73.57	54.21	9	4,573,723	76.31	47.74
Burkholderia mallei (4)	54	3,556,645	96.11	94.25	17	2,439,998	75.67	67.40
Burkholderia pseudomallei (4)	24	4,429,541	93.05	83.77	34	3,435,954	93.17	84.35
Ralstonia pickettii (2)	134	3,807,888	83.04	77.76	6	1,318,093	98.97	98.49
γ-proteobacteria								
Vibrio cholerae (5)	11	3,502,052	94.96	81.19	36	1,106,524	92.54	88.69
Vibrio vulnificus (3)	44	3,507,518	87.90	80.19	2	2,138,463	89.30	74.41

(b)

Subgroup/Species	Chromosome I				Chromosome II			
	LCB Number	Length of common LCBs	% Conservation	% Nucleotide Identity	LCB Number	Length of common LCBs	% Conservation	% Nucleotide Identity
α-proteobacteria								
Rhodobacter sphaeroides (4)	97	3,163,523	80.49	68.73	45	554,956	27.36	44.45
β-proteobacteria								
Burkholderia sp (5)	82	5,230,447	49.33	23.05	24	707,711	5.52	13.81
Ralstonia eutropha (2)	134	3,795,742	71.52	62.46	240	2,043,894	46.46	51.83
γ-proteobacteria								
Vibrio fischeri (2)	4	3,073,948	94.02	85.17	62	1,512,354	82.42	70.92
Vibrio sp (2)	39	3,779,449	76.63	55.37	34	2,089,504	55.83	39.27

Table S3. HGT estimates of CI and CII in 83 bacterial strains.

	Length of HGT Estimates			%HGT		
	CI	CII	CIII	CI	CII	CIII
α-proteobacteria						
Agrobacterium radiobacter K84	162,442	105,515		4.06	3.98	
Agrobacterium sp H13-13	78,338	86,870		2.77	4.04	
Agrobacterium tumefaciens C58	52,915	58,171		1.86	2.80	

Continued

Agrobacterium vitis S4	186,473	0		5.00	0.00	
Asticcacaulis excentricus CB 48	87,351	27,492		3.37	2.09	
Brucella abortus A13334	68,594	4949		0.03	0.00	
Brucella abortus S19	109,017	0		5.14	0.00	
Brucella abortus biovar 1 9-941	126,341	4132		5.95	0.36	
Brucella canis ATCC 23365	93,757	41,535		4.45	3.44	
Brucella canis HSK A52141	70,574	9286		3.00	1.00	
Brucella melitensis ATCC 23457	121,903	21,397		5.73	1.80	
Brucella melitensis biovar Abortus 2308	109,142	7665		5.14	0.66	
Brucella melitensis 16M	75,981	37,545		3.59	3.19	
Brucella microti CCM 4915	97,354	49,823		4.60	4.08	
Brucella ovis ATCC 25840	86,470	57,648		4.10	4.95	
Brucella suis 1330	114,178	43,292		5.42	3.59	
Brucella suis ATCC 23445	99,076	202,471		5.15	14.45	
Brucella suis VBI22	113,133	34,107		0.05	0.03	
Ochrobactrum anthropi ATCC 49188	188,800	115,548		6.54	6.09	
Paracoccus denitrificans PD1222	276,746	93,176		9.70	5.39	
Rhodobacter sphaeroides 2.4.1	91,200	57,248		2.86	6.07	
Rhodobacter sphaeroides ATCC 17025	388,956	68,515		12.09	7.80	
Rhodobacter sphaeroides ATCC 17029	90,190	124,545		2.87	10.22	
Rhodobacter sphaeroides KD131	49,174	76,120		1.56	5.87	
Sinorhizobium meliloti AK83	111,386	52,351	29,694	2.92	3.11	2.26
Sphingobium chlorophenolicum L1	116,096	135,247		3.77	9.88	
Sphingobium japonicum UT26S	264,985	33,060		7.54	4.85	
β-proteobacteria						
Burkholderia ambifaria AMMD	158,884	115,478	0	4.47	4.36	0.00
Burkholderia ambifaria MC40 6	70,010	190,834	115,076	2.03	6.89	10.20
Burkholderia cenocepacia AU 1054	253,091	215,906	26,067	7.68	7.74	2.18
Burkholderia cenocepacia HI2424	178,253	182,704	34,619	5.12	6.09	3.28
Burkholderia cenocepacia J2315	491,969	195,340	156,952	12.71	6.07	17.92
Burkholderia cenocepacia MC0 3	197,640	419,442	126,230	5.59	13.05	10.31
Burkholderia gladioli BSR3	217,492	105,733		4.93	2.86	
Burkholderia glumae BGR1	223,501	227,947		5.72	8.06	
Burkholderia mallei ATCC 23344	301,424	315,337				
Burkholderia mallei NCTC 10229	184,667	232,937				
Burkholderia mallei NCTC 10247	167,189	100,048		4.78	4.25	
Burkholderia mallei SAVP1	135,220	60,583		3.87	3.49	

Continued

Burkholderia multivorans ATCC 17616	274,808	160,746	108,001	7.97	6.50	11.74
Burkholderia phymatum STM815	126,442	275,386		3.63	10.21	
Burkholderia phytofirmans PsJN	429,932	284,261		9.62	7.84	
Burkholderia pseudomallei 1106a	339,151	204,832		8.50	6.61	
Burkholderia pseudomallei 1710b	313,937	252,734		7.61	7.94	
Burkholderia pseudomallei 668	285,966	145,678		7.31	4.66	
Burkholderia pseudomallei K96243	382,881	177,945		9.40	5.61	
Burkholderia sp 383	203,924	227,666	119,124	5.52	6.35	8.54
Burkholderia sp CCGE1001	408,397	221,542		10.05	8.00	
Burkholderia sp CCGE1002	210,557	144,399	22,124	5.98	5.57	1.72
Burkholderia sp CCGE1003	346,028	270,049		8.49	9.10	
Burkholderia sp KJ006			141,685			13.09
Burkholderia sp Y123	236,350	163,928	155,593			
Burkholderia rhizoxinica HKI 454	168,291	30,150		6.11	3.67	
Burkholderia thailandensis E264	309,135	254,283		8.12	8.72	
Burkholderia vietnamiensis G4	561,091	263,638	229,818	15.36	10.93	18.52
Burkholderia xenovorans LB400	676,505	228,537	78,180	13.82	6.79	5.31
Cupriavidus necator N-1	272,688	292,670		7.04	10.90	
Cupriavidus taiwanensis LMG 19424z	148,803	136,454		4.35	5.45	
Ralstonia eutropha H16	425,326	356,972		10.50	12.26	
Ralstonia eutropha JMP134	386,502	75,049		10.15	2.75	
Ralstonia pickettii 12D	314,048	69,788		8.61	5.27	
Ralstonia pickettii 12J	297,802	83,768		7.55	6.43	
Variovorax paradoxus S110	53,423	90,424		0.95	8.01	
γ-proteobacteria						
Aliivibrio salmonicida LFI1238	198,336	90,817		5.96	7.53	
Photobacterium profundum SS9	186,103	169,486		4.56	7.57	
Pseudoalteromonas haloplanktis TAC125	50,841	0		1.58	0.00	
Pseudoalteromonas sp. SM9913	113,350	18,255		3.40	2.59	
Vibrio anguillarum 775	214,064	74,339		6.99	7.52	
Vibrio cholerae M66-2	113,954	120,306		3.94	11.50	
Vibrio cholerae MJ-1236	219,312	42,604		6.96	3.92	
Vibrio cholerae O1 biovar El Tor str. N16961	133,868	148,186		4.52	13.82	
Vibrio cholerae O1 str. 2010EL-1786	168,009	113,071		5.54	10.81	
Vibrio cholerae O395	241,561	151,585		5.01	13.68	
Vibrio fischeri ES114	21,858	0		0.75	0.00	
Vibrio fischeri MJ11	0	5087		0.00	0.36	
Vibrio furnissii NCTC 11218	138,812	0		0.04	0.00	

Continued

Vibrio harveyi ATCC BAA-1116	413,361	94,267	10.98	4.28
Vibrio parahaemolyticus RIMD 2210633	287,292	104,618	8.74	5.57
Vibrio sp EJY3	539,693	67,368	2.73	1.94
Vibrio sp. Ex25	111,733	0	3.43	0.00
Vibrio splendidus LGP32	230,679	41,688	6.99	2.49
Vibrio vulnificus CMCP6	326,328	100,421	9.94	5.44
Vibrio vulnificus MO6-24/O	85,686	84,173	2.68	4.64
Vibrio vulnificus YJ016	254,485	45,811	7.59	2.47

Table S4. Average Ka, Ks, and ω values of CI and CII in16 species within *α*-, *β*-, *γ*-subgroup of Proteobacteria. The number of strains used in each comparison is indicated in parentheses.

	CI							CII						
	Orthologs	Average *Ka*	St Dev *Ka*	Average *Ks*	St Dev *Ks*	Average *Ka/Ks*	Avg *Ka/Ks* St Dev	Orthologs	Average *Ka*	St Dev *Ka*	Average *Ks*	St Dev *Ks*	Average *Ka/Ks*	Avg *Ka/Ks* St Dev
α-proteobacteria														
Brucella abortus (3)	1784	0.012	0.057	0.021	0.074	0.051	0.197	935	0.048	0.147	0.086	0.318	0.255	0.404
Brucella canis (2)	1892	0.002	0.011	0.008	0.022	0.041	0.121	1,005	0.016	0.051	0.026	0.067	0.126	0.265
Brucella mellintensis (3)	1736	0.006	0.041	0.013	0.061	0.307	0.372	883	0.005	0.021	0.011	0.027	0.316	0.379
Brucella suis (3)	1780	0.004	0.042	0.011	0.064	0.258	0.313	1,052	0.004	0.022	0.009	0.027	0.324	0.345
Rhodobacter sphaeroides (4)	2461	0.041	0.067	0.272	0.264	0.117	0.116	333	0.045	0.082	0.280	0.269	0.121	0.126
β-proteobacteria														
Burkholderia ambifairia (2)	2853	0.013	0.068	0.081	0.155	0.102	0.137	2,018	0.025	0.108	0.111	0.227	0.132	0.147
Burkholderia cenocepacia (4)	2338	0.016	0.068	0.102	0.171	0.104	0.135	1,971	0.021	0.090	0.117	0.130	0.135	0.181
Burkholderia mallei (4)	2444	0.011	0.082	0.020	0.117	0.101	0.261	1,086	0.005	0.020	0.011	0.030	0.140	0.369
Burkholderia pseudomallei (4)	2938	0.002	0.008	0.012	0.024	0.174	0.274	2,023	0.004	0.074	0.015	0.069	0.219	0.311
Burkholderia sp (5)	2005	0.144	0.147	0.643	0.245	0.204	0.184	453	0.275	0.267	0.782	0.279	0.317	0.258
Ralstonia eutropha (2)	2722	0.118	0.154	0.603	0.292	0.168	0.135	1,483	0.230	0.245	0.786	0.333	0.333	0.215
Ralstonia picketti (2)	2936	0.031	0.109	0.234	0.248	0.098	0.242	1,135	0.299	0.192	1.032	0.659	0.380	0.269
γ-proteobacteria														
Vibrio cholerae (5)	2411	0.005	0.045	0.055	0.137	0.082	0.165	826	0.020	0.121	0.098	0.338	0.119	0.219
Vibrio fischeri (2)	2414	0.016	0.080	0.186	0.241	0.057	0.094	986	0.027	0.101	0.265	0.312	0.077	0.115
Vibrio sp (2)	2346	0.066	0.117	1.462	0.935	0.049	0.074	1,056	0.127	0.211	1.949	0.987	0.064	0.087
Vibrio vulnificus (3)	2505	0.007	0.045	0.066	0.164	0.067	0.115	1,325	0.006	0.025	0.071	0.126	0.082	0.123