Research Article

Evidence of Extensive Homologous Recombination in the Core Genome of *Rickettsia*

Jinyu Wu,[1] Tonghai Yu,[1] Qiyu Bao,[1] and Fangqing Zhao[1, 2]

[1] *Institute of Biomedical Informatics/Zhejiang Provincial Key Laboratory of Medical Genetics, Wenzhou Medical College, Wenzhou 325035, China*

[2] *Center for Comparative Genomics and Bioinformatics, Department of Biochemistry and Molecular Biology, Pennsylvania State University, PA 16802, USA*

Correspondence should be addressed to Qiyu Bao, baoqywzmc@yahoo.com.cn and Fangqing Zhao, fuz3@psu.edu

The important role of homologous recombination has been extensively demonstrated to be fundamental for genetic variation in bacterial genomes. In contrast to extracellular or facultative intracellular bacteria, obligate intracellular bacteria are considered to be less prone to recombination, especially for their core genomes. In *Rickettsia*, only antigen-related genes were identified to have experienced homologous recombination. In this study, we employed evolutionary genomic approaches to investigate the impact of recombination on the core genome of *Rickettsia*. Phylogenetic network and phylogenetic compatibility matrix analyses are clearly consistent with the hypothesis that recombination has occurred frequently during *Rickettsia* evolution. 28% of *Rickettsia* core genes (194 out of 690) are found to present the evidence of recombination under four independent statistical methods. Further functional classification shows that these recombination events occur across all functional categories, with a significant overrepresentation in the cell wall/membrane/envelope biogenesis, which may provide a molecular basis for the parasite adaptation to host immunity. This evolutionary genomic analysis provides insight into the substantial role of recombination in the evolution of the intracellular pathogenic bacteria *Rickettsia*.

1. Introduction

Rickettsia spp. are obligate intracellular bacteria that are classified as gram-negative bacteria and belong to alpha subgroup of proteobacteria. They often live in arthropods, such as ticks, mites, louse, and fleas and can stably coexist within the host population. Some *Rickettsia* species can infect humans through bites or feces of the vectors, and then cause mild to serious diseases. It has been reported that various kinds of diseases caused by *Rickettsia* can be found worldwide. Based on phenotypic and phylogenetic evidence [1], *Rickettsia* have been classified into three different groups, including the spotted fever group (e.g., *R. africae*, *R. conorii*, and *R. massiliae*), the typhus group (e.g., *R. prowazekii* and *R. typhi*), and the unclassified group (e.g., *R. bellii*). Characterizing the evolutionary mechanism of *Rickettsia* is a key step towards understanding its adaptation to different hosts and the design of more effective drugs [2].

Homologous recombination, involved in transferring of a specific DNA fragment from one strain into the homologous region of another strain, is considered to be fundamental for generating genetic variation and to be one of the most important evolutionary mechanisms in bacterial evolution [3]. Meanwhile, it was believed that recombination is largely restricted to extracellular or facultative intracellular bacteria, such as *Escherichia coli* [4], *Mycobacterium* [5], *Helicobacter pylori* [6], and *Streptococcus* [7]. By contrast, obligate intracellular bacteria are often considered to be less prone to recombination [8, 9], and recombination cases are rarely reported. Especially, their core genomes are thought to be free of recombination. Recently, however, a completely different scenario was addressed in *Wolbachia* genome [10]. Based on four housekeeping genes (*gltA*, *dnaA*, *ftsZ*, and *groEL*) as well as other previously reported cases of *Wolbachia* recombination, Baldo et al. suggested that *Wolbachia* could be subject to widespread recombination. More recently,
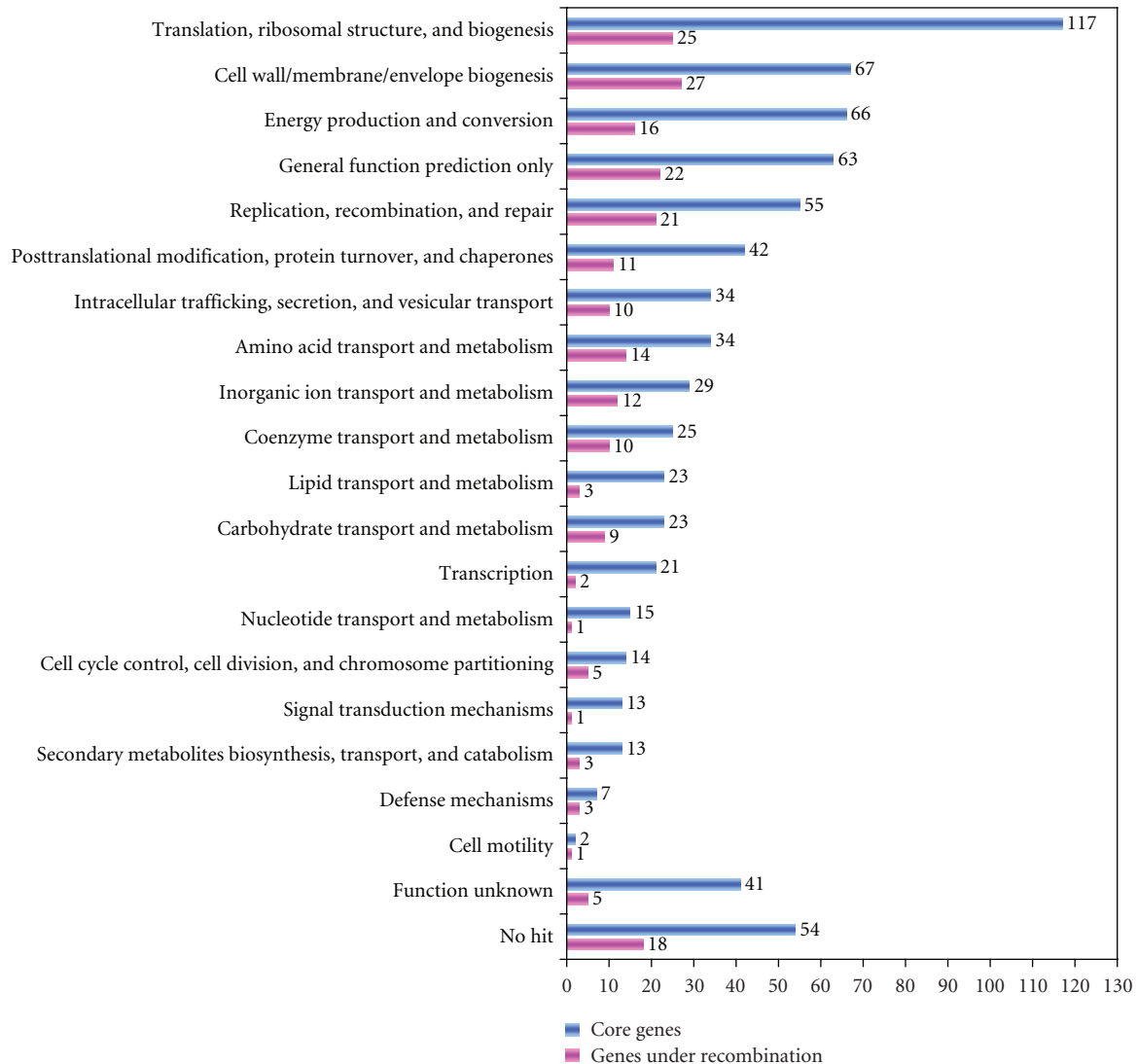
FIGURE 1: Functional classification of the *Rickettsia*'s core genes and the genes identified to be under recombination. It is notable that some proteins may be classified into multiple categories.

by investigating of multiple loci dispersed throughout the *Chlamydia trachomatis* chromosome, Gomes et al. [11] showed that recombination is widely spread in *Chlamydia trachomatis*. However, *Rickettsia*, as a group of important obligate intracellular bacteria, are often considered to have evolved with little impact of recombination [12]. Till now, only the antigens have been reported to present the evidence of recombination [13].

In this study, a number of evolutionary genomic methods have been employed to reveal the evidence of recombination in the evolution of *Rickettsia*. As a result, significant evidence for recombination is found among *Rickettsia* core genomes through phylogenetic network reconstruction and phylogenetic compatibility matrix analyses. We also found that 28% of the core genes may have been under recombination, which is unexpected for intracellular bacteria. Functional classification showed that recombination events occurred across all functional categories but with a significant bias in the

genes involved in cell wall/membrane/envelope biogenesis. This evolutionary genomic analysis provided insight into the important role of recombination in *Rickettsia*.

2. Materials and Methods

12 *Rickettsia* genome sequences (*R. africae* ESF-5, *R. sibirica* 246, *R. conorii* Malish7, *R. rickettsii*, *R. massiliae* MTU5, *R. akari* Hartford, *R. felis* URRWXCal2, *R. prowazekii* Madrid E, *R. typhi* Wilmington, *R. canadensis* McKie, *R. bellii* OSU 85–389, and *R. bellii* RML369-C) were obtained from the PATRIC database (http://patric.vbi.vt.edu/organism/genomeList.php?organismId=3&dm=Tabular). Orthologous groups of genes from different *Rickettsia* genomes were identified using OrthoMCL with the default parameters [14], which has been proven to be a very powerful tool in identifying orthologous families from multiple genomes,
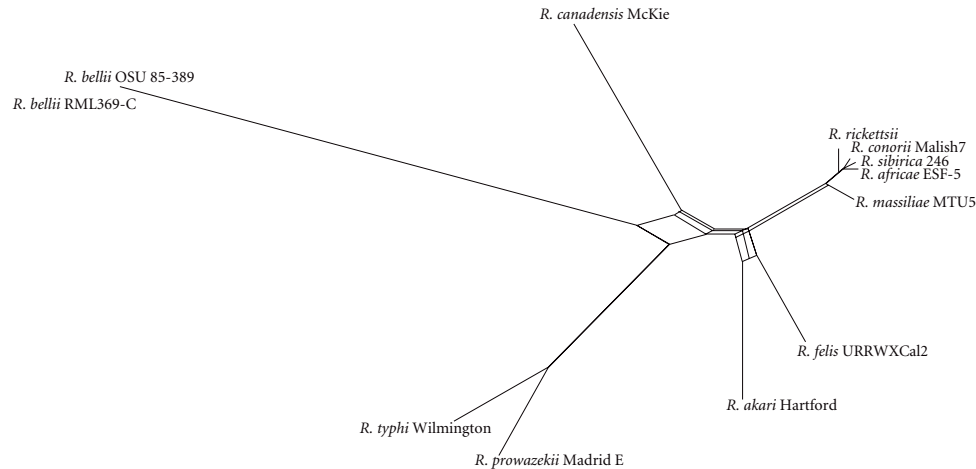
FIGURE 2: Phylogenetic network of *Rickettsia*'s core genes based on the concatenated nucleotide sequences. The phylogenetic network is constructed using the Neighbor-net method algorithm [20] and shows several parallel paths, indicating occurrence of recombination in the core genome of *Rickettsia*.
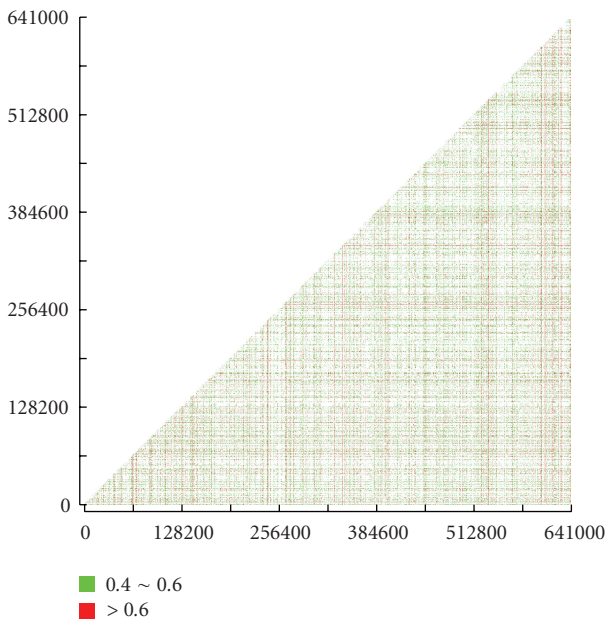

FIGURE 3: The phylogenetic compatibility matrix of *Rickettsia*'s core genes. The matrix is plotted based on the concatenated alignment of core genes with a 300-bp window and a 100-bp step. Phylogenetic trees are generated for each region using the neighbor-joining algorithm with a 70% bootstrap value. Phylogenetic violations of different topologies are calculated by comparison of the violations of the branching orders in different phylogenetic trees. Frequency is colored to indicate the number of phylogenetic violations per sequences.

and it has been widely adopted in recent studies [15, 16]. Among the identified families, only the family with one-to-one orthologous relationship was considered for further analyses. The functional category of each orthologous family was obtained by BLASTing to the COG database (http://www.ncbi.nih.gov/COG/) with an E-value of 10^{-5}.

Multiple sequence alignment was performed at the amino acid level using the ClustalW program with the default settings [17]. The corresponding nucleotide sequence alignment was generated according to the protein alignment using the tranalign program implemented in the EMBOSS package [18]. To construct a phylogenetic compatibility matrix, all the orthologous gene sets were firstly concatenated into a single alignment. Then, the position of sequences in trees generated for each region was compared with that for all the other regions to produce a phylogenetic compatibility matrix using the TreeOrderScan program [19]. The phylogenetic network of the concatenated DNA sequences was constructed using the Neighbor-net method [20] implemented in the SplitsTree version 4.8 [21]. Identification of potential recombination evidence from each aligned orthologous group (at a nucleotide level) was performed using the RDP [22], MaxChi [23], Chimaera [24], and GENECONV [25] methods implemented in the RDP3 program [26]. These four programs have been indicated to be powerful in detecting recombination event and have been widely used by many studies [10, 27].

3. Results and Discussion

A core genome is a set of orthologous genes, which are shared by all members of a group of bacteria. Using the Markov Clustering algorithm [14], a total of 690 orthologous sets are found in the 12 *Rickettsia* genomes (See supplementary Table S1 in supplementary material available online at doi: 10.1155/2009/510270). This core set represents a large proportion (690/835) of the total number of ORFs in the smallest genome (*R. prowazekii* Madrid E) and nearly a half (690/1512) in the largest genome (*R. felis* URRWXCal2). In spite of the differences on the use of orthologous identification method and the number of genomes, the core set size of *Rickettsia* estimated here is quite similar to previous report [12]. Functional classification of these core sets of genes

based on the COG database shows that they are present in all primary functions (Figure 1); the majority of which encode components of information-processing systems involved in translation, ribosomal structure, and biogenesis. However, there are also ~100 genes being classified into the class of "Functionally unknown" or "No hit."

A strategy for recombination detection is to construct a phylogenetic network for the target genes, which can generalize the phylogenetic tree through allowing the representation of conflicting signals or alternative phylogenetic histories [28]. To observe the phylogenetic incongruence among different gene fragments, all 690 core-set genes were concatenated into a single alignment. Using the Neighbor-net method [20], we found that multiple recombination events have occurred in the *Rickettsia* lineage (Figure 2). Meanwhile, based on the Phi test [29], it also strongly evidenced the presence of recombination in the *Rickettsia*'s core genes ($P \ll .00\,001$). Another useful strategy for detection of recombination is to generate phylogenetic trees from different parts of a gene [19]. If adjacent regions yield significantly different tree topologies, then recombination may have occurred. To visualize phylogenetic incongruence of different parts of a gene, a phylogenetic compatibility matrix was generated. As shown in Figure 3, the matrix shows a considerable number of phylogeny violations among *Rickettsia*, indicating the wide spread of recombination. Among them, about 3.4% of the phylogeny violation frequency is above 0.6 and 11.7% from 0.4 to 0.6.

To detect individual recombination event within the orthologous gene set, four recombination detection methods were used, including RDP [22], MaxChi [23], Chimaera [24], and GENECONV [25]. To ensure the reliability of our estimated results, only the recombination events that are supported by at least two independent tests were taken into account. As a result, a total of 194 core genes were found to be prone to recombination (supplementary Table S2), of which 60 gene sets were supported by four approaches and 45 gene sets could be identified by three methods. Hence, it is suggested that recombination should be prevalent than ever expected in *Rickettsia*'s core genes. Functional classification of these recombined genes clearly shows that none of the functional categories can be free from recombination (Figure 1), including the categories of translation ribosomal structure and biogenesis category. Interestingly, recombination preferentially occurs in genes involved in cell wall/membrane/envelope biogenesis. Such an elevated rate of recombination may be related with the coevolutionary arms race between *Rickettsia* and their host cells and may confer a selective advantage to survive in different niches.

Taken together, our results strongly suggest that recombination may have played an important role in *Rickettsia* evolution. Since this study is only designed to detect the evidence of recombination events during the evolution of *Rickettsia*, the recombination breakpoints, parental and daughter sequences, as well as the possible reasons of widespread recombination in the core genome of *Rickettsia* are not considered. Combined with previous reports on *Wolbachia* [10] and *Chlamydia trachomatis* [11], we

may have to update the evolutionary scenario of obligate intracellular bacteria, of which they can also be subject to extensive recombination events during evolution. The identification of the genes involved in recombination may prompt further investigation into their biological functions.

Acknowledgments

References

[1] L. Vitorino, I. M. Chelo, F. Bacellar, and L. Zé-Zé, "Rickettsiae phylogeny: a multigenic approach," *Microbiology*, vol. 153, part 1, pp. 160–168, 2007.

[2] M. J. Pallen and B. W. Wren, "Bacterial pathogenomics," *Nature*, vol. 449, no. 7164, pp. 835–842, 2007.

[3] C. Fraser, W. P. Hanage, and B. G. Spratt, "Recombination and the nature of bacterial speciation," *Science*, vol. 315, no. 5811, pp. 476–480, 2007.

[4] B. Mau, J. D. Glasner, A. E. Darling, and N. T. Perna, "Genome-wide detection and analysis of homologous recombination among sequenced strains of *Escherichia coli*," *Genome Biology*, vol. 7, no. 5, article R44, pp. 1–12, 2006.

[5] E. Krzywinska, J. Krzywinski, and J. S. Schorey, "Naturally occurring horizontal gene transfer and homologous recombination in *Mycobacterium*," *Microbiology*, vol. 150, part 6, pp. 1707–1712, 2004.

[6] S. Suerbaum, J. M. Smith, K. Bapumia, et al., "Free recombination within *Helicobacter pylori*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 21, pp. 12619–12624, 1998.

[7] T. Lefébure and M. J. Stanhope, "Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition," *Genome Biology*, vol. 8, no. 5, article R71, pp. 1–17, 2007.

[8] V. Daubin, N. A. Moran, and H. Ochman, "Phylogenetics and the cohesion of bacterial genomes," *Science*, vol. 301, no. 5634, pp. 829–832, 2003.

[9] E. Lerat, V. Daubin, H. Ochman, and N. A. Moran, "Evolutionary origins of genomic repertoires in bacteria," *PLoS Biology*, vol. 3, no. 5, article e130, pp. 1–8, 2005.

[10] L. Baldo, S. Bordenstein, J. J. Wernegreen, and J. H. Werren, "Widespread recombination throughout *Wolbachia* genomes," *Molecular Biology and Evolution*, vol. 23, no. 2, pp. 437–449, 2006.

[11] J. P. Gomes, W. J. Bruno, A. Nunes, et al., "Evolution of *Chlamydia trachomatis* diversity occurs by widespread interstrain recombination involving hotspots," *Genome Research*, vol. 17, no. 1, pp. 50–60, 2007.

[12] G. Blanc, H. Ogata, C. Robert, et al., "Reductive genome evolution from the mother of *Rickettsia*," *PLoS Genetics*, vol. 3, no. 1, article e14, pp. 1–12, 2007.

[13] F. M. Jiggins, "Adaptive evolution and recombination of *Rickettsia* antigens," *Journal of Molecular Evolution*, vol. 62, no. 1, pp. 99–110, 2006.

[14] F. Chen, A. J. Mackey, J. K. Vermunt, and D. S. Roos, "Assessing performance of orthology detection strategies applied to eukaryotic genomes," *PLoS ONE*, vol. 2, no. 4, article e383, pp. 1–12, 2007.

[15] X. Xu, J. Wu, J. Xiao, et al., "PlasmoGF: an integrated system for comparative genomics and phylogenetic analysis of *Plasmodium* gene families," *Bioinformatics*, vol. 24, no. 9, pp. 1217–1220, 2008.

[16] J. Wu, S. Wang, J. Bai, et al., "ArchaeaTF: an integrated database of putative transcription factors in Archaea," *Genomics*, vol. 91, no. 1, pp. 102–107, 2008.

[17] J. D. Thompson, D. G. Higgins, and T. J. Gibson, "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice," *Nucleic Acids Research*, vol. 22, no. 22, pp. 4673–4680, 1994.

[18] S. A. Olson, "EMBOSS opens up sequence analysis. European Molecular Biology Open Software Suite," *Briefings in Bioinformatics*, vol. 3, no. 1, pp. 87–91, 2002.

[19] P. Simmonds and S. Midgley, "Recombination in the genesis and evolution of hepatitis B virus genotypes," *Journal of Virology*, vol. 79, no. 24, pp. 15467–15476, 2005.

[20] D. Bryant and V. Moulton, "Neighbor-net: an agglomerative method for the construction of phylogenetic networks," *Molecular Biology and Evolution*, vol. 21, no. 2, pp. 255–265, 2004.

[21] T. H. Kloepper and D. H. Huson, "Drawing explicit phylogenetic networks and their integration into SplitsTree," *BMC Evolutionary Biology*, vol. 8, article 22, pp. 1–7, 2008.

[22] D. Martin and E. Rybicki, "RDP: detection of recombination amongst aligned sequences," *Bioinformatics*, vol. 16, no. 6, pp. 562–563, 2000.

[23] J. M. Smith, "Analyzing the mosaic structure of genes," *Journal of Molecular Evolution*, vol. 34, no. 2, pp. 126–129, 1992.

[24] D. Posada, "Evaluation of methods for detecting recombination from DNA sequences: empirical data," *Molecular Biology and Evolution*, vol. 19, no. 5, pp. 708–717, 2002.

[25] S. Sawyer, "Statistical tests for detecting gene conversion," *Molecular Biology and Evolution*, vol. 6, no. 5, pp. 526–538, 1989.

[26] D. P. Martin, C. Williamson, and D. Posada, "RDP2: recombination detection and analysis from sequence alignments," *Bioinformatics*, vol. 21, no. 2, pp. 260–262, 2005.

[27] A. D. Tsaousis, D. P. Martin, E. D. Ladoukakis, D. Posada, and E. Zouros, "Widespread recombination in published animal mtDNA sequences," *Molecular Biology and Evolution*, vol. 22, no. 4, pp. 925–933, 2005.

[28] L. Wang, K. Zhang, and L. Zhang, "Perfect phylogenetic networks with recombination," *Journal of Computational Biology*, vol. 8, no. 1, pp. 69–78, 2001.

[29] T. C. Bruen, H. Philippe, and D. Bryant, "A simple and robust statistical test for detecting the presence of recombination," *Genetics*, vol. 172, no. 4, pp. 2665–2681, 2006.