Phylogenomic analysis of the intracellular symbiont *Wolbachia* in arthropods reveals genome evolution and interclade recombination events

Xiao Xiong[1,2*], Xiaozhu Wang[1*], Wenqi Cao[1], Chao Zhang[2], John H. Werren[3†] and Xu Wang[1,4,5†]

1-Dept of Pathobiology, Auburn University, 2-Tongji University, Shanghai, China
3-Biology, University of Rochester, 4-Alabama Agricultural Experiment Station
5-HudsonAlpha Institute for Biotechnology, Huntsville, AL
*co-first authors; †co-corresponding author. Contact: xzw0070@auburn.edu

Introduction

Wolbachia are widespread obligate intracellular bacteria that commonly infect arthropods and filarial nematodes[1-3]. In particular, more than half of the arthropod species are infected by *Wolbachia*[4,5]. *Wolbacia* can mediate many important biological processes, such as male-killing, parthenogenetic induction, and cytoplasmic incompatibility[2,6]. Most of the previous research on recombination has focused on five MLST genes, *Wolbachia* surface protein (wsp), and 16S rRNA, or for a subset of genomes from the A-D and F supergroups[7]. Therefore, whole-genome analyses in a large number of *Wolbachia* strains of all supergroups are needed to identify additional homologous recombination events among *Wolbachia* across the different supergroups.

In this study, we performed phylogenomic analyses on 33 annotated *Wolbachia* genomes in the A, B, C, D, E and F supergroups and analyzed the individual gene trees to identify conserved core gene set and detect potential recombination events across the supergroups.

Methods

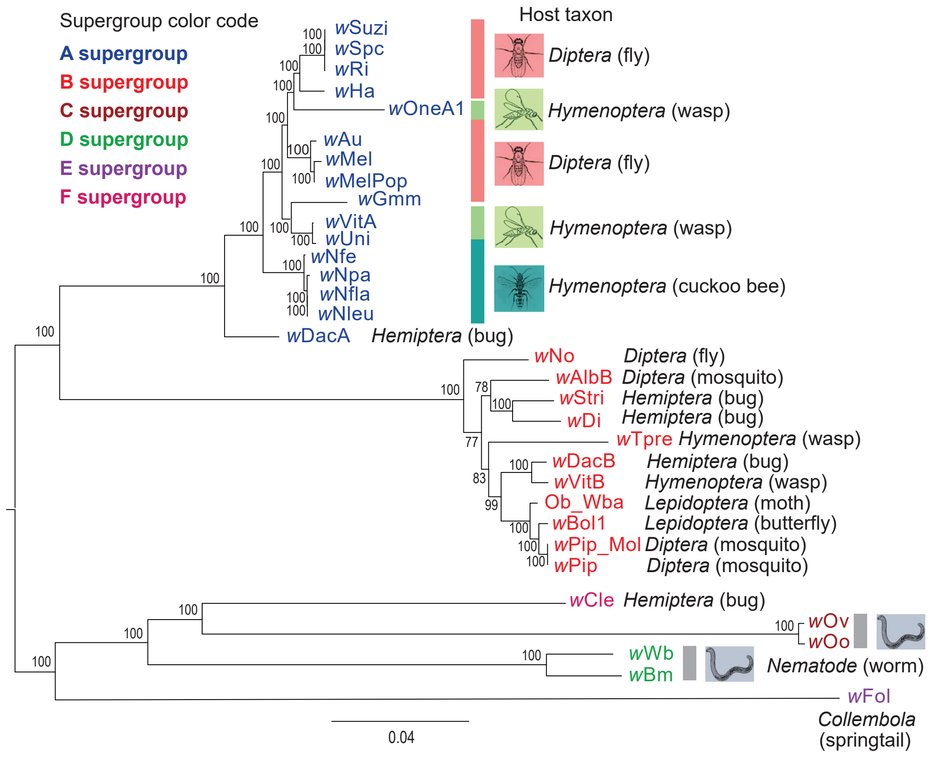
We conducted phylogenomic analysis using 33 annotated *Wolbachia* genomes. Homologous genes and ortholog clusters were determined by using OrthoFinder v1.1.8[8] with default settings. A Maximum Likelihood (ML) tree was constructed with the GTRGAMMA model and 1,000 bootstrap replicates by RAxMLv8.2[9] using the concatenated nucleotide sequence alignment of the core gene set.

The interclade recombination screening method was developed to detect candidate recombination events between supergroups. The single gene ML trees for genes with recombination events were constructed with their corresponding nucleotide sequence alignments by RAxML v8.2[9]. We also constructed protein trees for these identified genes with their corresponding protein sequence alignments and the core genes concatenated protein sequence alignment using the best fit protein model detected by ProtTest 3[10] and 1,000 rapid bootstrap replicates by RAxML v8.2[9].

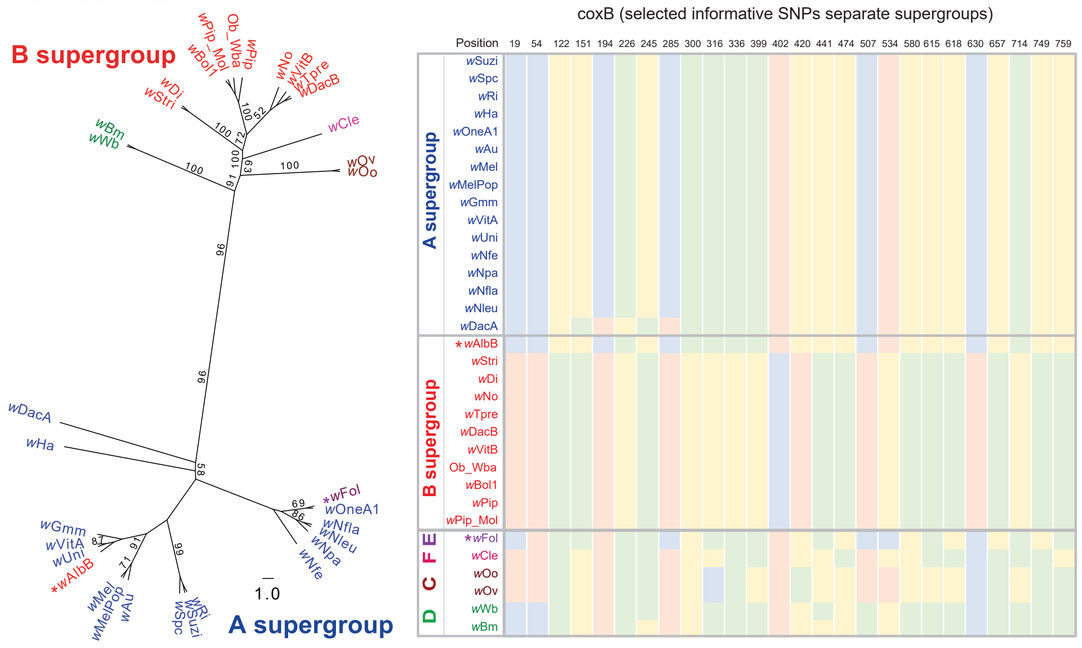
The five MLST (Multi Locus Sequence Typing) genes(gatB, fbpA, hcpA, coxA, ftsZ)[11,12] were examined to further characterize the phylogenetic relationships of *Wolbachia* strains in *Nasonia*. The pairwise evolutionary divergence distances between 33 Wolbachia species were estimated with both the core gene set identified in this study, five MLST genes and the concatenated sequence of these five MLST genes in 33 Wolbachia species by using the Maximum Composite Likelihood model[13] in MEGA7[14]. The Pearson correlation coefficient of estimated evolutionary divergences with the identified core genes and the MLST gene set was calculated with Hmisc package[15] in R.

Results

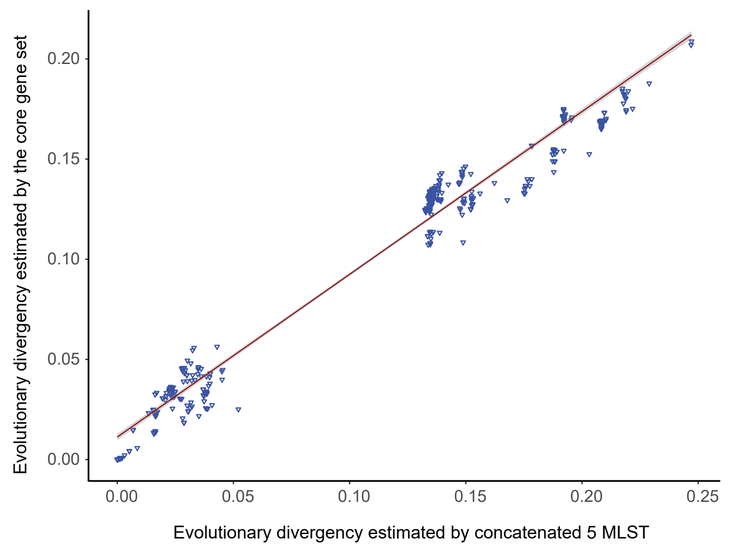
Based on the concatenated coding nucleotide and protein sequences of the core gene set, ML phylogenetic trees of 33 *Wolbachia* genomes confirmed the separation of different supergroups A, B, C, D, E and F with 100% bootstrap support. As expected, our genomic analyses support extensive horizontal movement of *Wolbachia* strains between divergent host species. This pattern was previously observed using MLST genes in *Wolbachia*, but is now supported by a much larger data set.



A total of 5 genes (2.4%) with 9 recombination events were identified between A and B supergroups, including B-supergroup genes ftsH and rplU in A-supergroup strains *w*Au and *w*DacA respectively, and 7 A-in-B recombination events in coxB , WONE_04820 and argS. The coxB gene from an A-*Wolbachia* was transferred to B-*Wolbachia* *w*AlbB (IR =92). For the E supergroup there is only one released genome (*w*Fol), we found *w*Fol strain cluster with A-*Wolbachia* in coxB. From the selected informative SNPs, wAlbB from B supergroup in coxB is the same as A supergroup strains. In addition, there is an apparent A-B recombinant event in the coxB gene of *w*DacA based on a stretch of 5 A-B diagnostic SNPs, this is the evidence for intragenic recombination.



The MLST performed very well in both identifying closely related strains and in genetic divergence among strains compared to the genome wide data set. The Pearson correlation coefficient of estimated evolutionary divergence with core gene set and the concatenated MLST set is 0.98 with P-value < 2.2 x 10^{-16}. Eventually, whole genome data sets will supplant the MLST system.



Conclusions

This finding suggests a relatively low frequency (2.4%) of intergroup recombination in *Wolbachia*. Maintenance of such selective transfers between supergroups suggests possible roles in *Wolbachia* infection related functions. Concordance of MLST and whole genome divergence indicates that MLST is a reliable method for identifying related strains and the identified core gene set is informative for strain identification. The interclade recombination screening method developed in our study will serve as a valuable foundation for investigation of recombination and genome evolution in *Wolbachia*.

References and Acknowledgements

[1] Fenn K *et al*. ***Trends Parasitol*** (2006)
[2] Werren IH *et al*. ***Nat Rev Microbiol*** (2008)
[3] Werren JH *et al*. ***Annu Rev Entomol*** (1997)
[4] Hilgenboecker K *et al*. ***FEMS Microbiol Lett*** (2008)
[5] Zug R *et al*. ***PLoS One*** (2012)
[6] Stouthamer R *et al*. ***Annu Rev Microbiol*** (1999)
[7] Lindsey AR *et al*. ***G3*** (2016)
[8] Emms DM *et al*. ***Genome Biol*** (2015)
[9] Stamatakis A *et al*. ***Bioinformatics*** (2014)
[10] Darriba D *et al*. ***Bioinformatics*** (2011)
[11] Baldo L *et al*. ***Appl Environ Microbiol*** (2006)
[12] Jolley KA *et al*. ***BMC Bioinformatics*** (2010)
[13] Tamura K *et al*. ***Proceedings of the National Academy of Sciences of the United States of America*** (2004)
[14] Kumar S *et al*. ***Mol Biol Evol*** (2016)
[15] Harrell Jr, FE *et al*. ***CRAN2018*** (2019)

This project is supported by an AUIGP Award to X.W. (AUIGP-180271) and an USDA NIFA Hatch project 1018100. X.W. is supported by NSF EPSCoR RII Track-4 Research Fellowship (OIA 1928770), an AAES Enabling Grant, as well as a laboratory start-up fund from AU College of Veterinary Medicine. X.X. is supported by AU Presidential Graduate Research Fellowship and AUCVM Dean's Fellowship. Contributions of J.H.W. were supported by NSF IOS 1456233 and the Nathaniel and Helen Wisch Professorship. S Cheng is thanked for conducting tests for directional and purifying selection on recombined gene lineages.