A Wolbachia pangenome reconstructed through genome-resolved metagenomics from individual Culex pipiens ovaries reveals viral differentiation

Julie Reveillaud, Sarah R. Bordenstein, Corinne Cruaud, Alon Shaiber, Ignace Rakotoarivony, Seth R. Bordenstein, A. Murat Eren

A Wolbachia pangenome reconstructed through genome-resolved metagenomics from individual *Culex pipiens* ovaries reveals viral differentiation

Julie Reveillaud[1,♦,*], Sarah R. Bordenstein[2,♦], Corinne Cruaud[3], Alon Shaiber[4,5], Ignace Rakotoarivony[1], Seth R. Bordenstein[2,6,7], A. Murat Eren[4,5,8,*]

[1] ASTRE, INRA, CIRAD, University of Montpellier, Montpellier, France; [2] Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA; [3] CEA / Institut de Biologie François Jacob / Génoscope, Evry, France; [4] Graduate Program in the Biophysical Sciences, University of Chicago, Chicago, IL 60637, USA; [5] Department of Medicine, University of Chicago, Chicago, IL, USA; [6] Department of Pathology, Microbiology, and Immunology, Vanderbilt University, Nashville, TN, USA; [7] Vanderbilt Institute for Infection, Immunology, and Inflammation, Vanderbilt University, Nashville, TN, USA; [8] Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA, USA

[*] Correspondence: reveillaud.j@gmail.com; a.murat.eren@gmail.com; [♦] These authors contributed equally to this work.

Introduction

Members of *Wolbachia*, a genus of obligate intracellular bacteria within the class Alphaproteobacteria, infect an estimated 40% of arthropods. Recent studies demonstrate that temperate phage WO of *Wolbachia* carry the *cifA* and *cifB* genes linked to cytoplasmic incompatibility, a common host reproductive manipulation that results in embryonic lethality. These findings suggest that phages could have a critical role in controlling the spread of arboviruses in the wild, and could represent key targets for vector control strategies. Here, we study four wild *Culex pipiens* individuals captured in Southern France from a single collect through shotgun metagenomics (Figure 1). Using state-of-the-art assembly and binning strategies, we were able to reconstruct near-complete *Wolbachia* genomes from each individual, along with their phage WO and additional viral variants.

Figure 1 *Culex pipiens* individuals (a) from a single collect (b-c-d) examined for taxonomy prior to dissection

Methods

DNA extraction, sequencing, filtration and assembly

We extracted total genomic DNA from each ovary sample using MoBio PowerFecal DNA Isolation Kit. We sequenced the library using a HiSeq4000 Illumina sequencer at the Genoscope (Evry, France). We used IDBA_UD v1.1.2 (Peng et al. 2012) to assemble paired-end reads into contigs and Bowtie2 v2.2.9 (Langmead and Salzberg 2012) for all read recruitment analyses. We processed our contigs that are longer than 1,000 nts and read recruitment results with anvi'o v5 (Eren et al. 2015) to recover metagenome-assembled genomes from *C. pipiens* metagenomes. Our binning strategy used sequence composition signatures and differential coverage statistics of contigs across samples. Pangenomes and metagenomes were linked as in Delmont and Eren (2018).

Results and Discussion

Assembly, Mapping and binning

Shotgun sequencing of total community DNA yielded 65 to 78 million high-quality sequences for the four *Culex pipiens* individuals (O03, O07, O11, O12). For each sample, we were able to reconstruct a highly-complete single bacterial metagenome-assembled genome (MAG) that resolved to *Wolbachia* (Table 1).

MAG	Percent completion (PC)	Percent redundancy (PR)	Number of contigs (N)	Number of genes (n)	Length (total number of nucleotides)	GC content (%)
O03	94.24	0.72	93	1091	1,213,072	33.83
O07	94.24	0.72	127	1181	1,317,313	33.78
O11	94.24	0	99	1085	1,208,099	33.84
O12	94.24	0	99	1113	1,237,800	33.95

Table 1 *Wolbachia* MAGs with completion and redundancy estimates, number of contigs (N), number of genes (n), total number of nucleotides and % GC.

SNVs

A relatively low number of single-nucleotide variants (0.07% to 0.14%) suggested that each *Wolbachia* MAG represented a nearly-monoclonal population of bacterial cells.

The Wolbachia pangenome

We performed a metapangenomic analysis of the four *Wolbachia* MAGs and *w*Pip in conjunction with the four metagenomes from individual mosquitoes to link phylogenetic relationships between genes and their abundances in *C. pipiens* metagenomes. The *Wolbachia* pangenome contained 1,166 genes, the majority of which were conserved across all five genomes (Figure 2).

wPip gene clusters not recovered in MAGs

*w*Pip accessory genes that were not recovered in our MAGs encoded functions including several transposases, bacteriophage capsid protein coding genes, and other phage-related sequences, most of which were associated with known *Wolbachia* prophages.

Gene clusters unique to our MAGS

Gene clusters unique to MAG O07 contigs (but which recruited reads from the three other metagenomes) included a gene coding for an ankyrin and tetratricopeptide repeats protein previously identified in phage WO from *Nasonia vitripennis* wasps (Bordenstein and Bordenstein 2016), a Retron-type reverse transcriptase genes coding for Transposases, a Transposase InsO and inactivated derivatives gene as well as many eukaryotic viral genes.

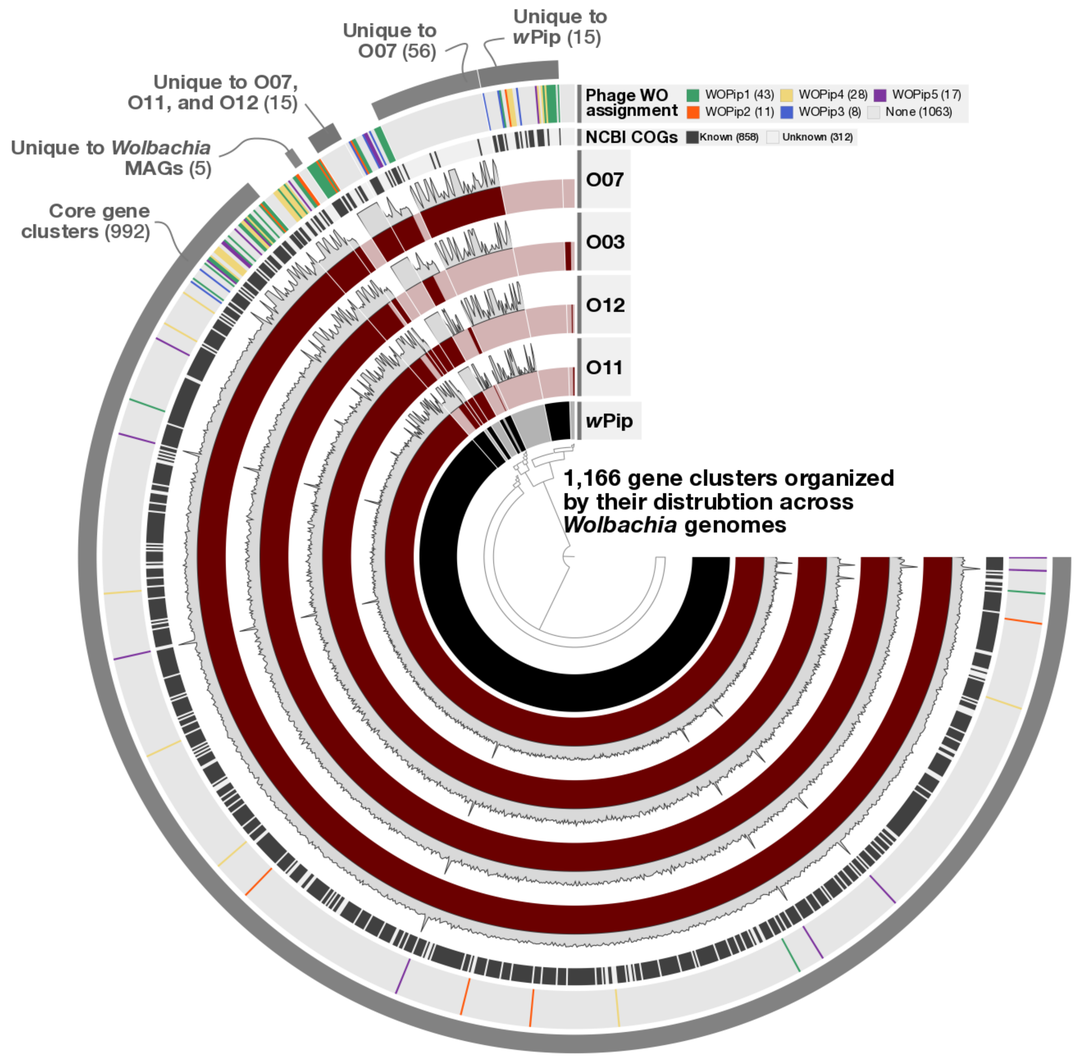

Figure 2 *Wolbachia* pangenome; the inner dendrogram organizes the 1,166 GC based on their distribution across wPip and the four MAGS genomes (black and red radial circles, respectively). A bar in the genome layer indicates that the genome had at least one gene that appeared in that gene cluster. The four MAGS-associated line graphs represent mean coverage values of GC in individual MAG. The following radial circles represent COGs annotations, and GC assigned to prophage WO (extracted from Klasson et al., 2008).

Conclusions

While our pangenomic analysis suggested high level of genomic conservation across the bacterial part of *Wolbachia* chromosomes, there was notable variation between individual mosquitoes due to differences in prophage WO and other viral genes. Overall, our findings revealed almost as many virus-associated genes as the ones that were previously recognized in the reference genome *w*Pip.

Acknowledgements

We thank Albane Marie, Gregory Lambert, Jean-Baptiste Panchau from EID (Entente Interdépartementale de Démoustication) for field Culex samples; Özcan C. Esen, Evan Kiefl, Richard Fox, and Frederick Gavory for their help with computational matters.

References

Bordenstein, Sarah R. and Seth R. Bordenstein. 2016. Eukaryotic Association Module in Phage WO Genomes from Wolbachia. *Nature Communications* 7:1–10.

Delmont, Tom O. and A. Murat Eren. 2018. Linking Pangenomes and Metagenomes: The *Prochlorococcus* Metapangenome. *PeerJ* 6:e4320.

Eren, A.M. et al., 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3, 1319 (2015).

Klasson, L. et al. 2009. The Mosaic Genome Structure of the Wolbachia WRi Strain Infecting Drosophila Simulans. *Proceedings of the National Academy of Sciences* 106(14): 5725–30.

Langmead, Ben and Steven L. Salzberg. 2012. Fast Gapped-Read Alignment with Bowtie 2. *Nature Methods* 9:357.

Peng, Yu, Henry C. M. Leung, S. M. Yiu, and Francis Y. L. Chin. 2012. IDBA-UD: A de Novo Assembler for Single-Cell and Metagenomic Sequencing Data with Highly Uneven Depth. *Bioinformatics* 28(11):1420–28.