### **Sample collection, DNA extraction, and whole genome sequencing**

NOTE: *Culex quinquefasciatus* female mosquitoes were collected from New Caledonia on 15th and 16th December 2022 using a BG sentinel trap.

* 1. Collect female mosquitoes, transport them alive to the laboratory, and process them as described in the protocol, yielding three midguts and three ovaries that must be stored in preservation buffer.
  2. Immediately before DNA extraction, filter the midgut samples through a sterile 5 µm filter as in1 to increase the proportion of bacterial DNA recovered by retaining the eukaryotic nuclei on the membrane.
  3. Extract DNA using the referenced kit (**Table of Materials**), with an overnight incubation at 56 °C.
  4. Store the DNA samples at -20 °C in DNAse-free water until library preparation and sequencing at the MGX sequencing platform.
     1. Prepare the libraries using the DNA prep kit and quantify on a Fragment Analyzer with the NGS kit and the Library quantification kit.
     2. Finally, perform sequencing in paired-end mode with a read length of 150 bp, with a target sequencing depth of 40,000 reads for the ovaries and 80,000 for the midguts.

### **Genome reconstruction**

* 1. Carry out quality filtration of the demultiplexed metagenomic raw reads with Illumina-Utils python scripts2 following recommendations by Minoche et al.3.
  2. Trim adapter sequences using bbduk4.
  3. Mask the *Culex quinquefasciatus* genome (GCF\_015732765.1 on NCBI) with bbmask4, using *Wolbachia* genome *w*PipPel8 as the reference, to prevent spurious removal of bacterial sequences.
  4. Run the metagenomic Snakemake workflow5 available with Anvi’o v7.16,7 to remove short reads mapping to the masked *Culex* *quinquefasciatus* genome and perform single assemblies on the remaining reads using Megahit v1.2.99 with preset meta-sensitive and a minimum contig length of 1,000 bp. Identify Open Reading Frames (ORFs) in the resulting contigs with Prodigal10 and obtain taxonomic annotations at the contig level using Centrifuge v1.0.411. Identify single-copy core genes from bacterial, archaeal, and protist collections available in Anvi’o using hmm profiles12. Map the *Culex*-filtered short reads to the assembled contigs with bowtie213, 14 and merge the read recruitment profiles into Anvi’o profile databases.
  5. Conduct automatic binning with CONCOCT15, restricting the number of clusters/bins produced to 3 to prevent fragmentation errors. Inspect the bins, refine them manually in Anvi’o’s interactive interface16, and estimate the completeness and redundancy based on the single-copy core gene collection17.
  6. Dereplicate the reconstructed Metagenome-Assembled Genomes (MAGs) of length over 1 Mbp based on pyANI with a minimum alignment fraction of 0.7 and a similarity threshold of 0.99, keeping the longest MAG as representative.
  7. Run these representative MAGs again through the metagenomic workflow in reference mode to obtain final coverage values. When detection (or breadth of coverage) is over 0.9, use mean coverage values as a proxy for relative abundance of the MAG in a sample; otherwise, set it to 0.

### **Estimating prokaryotic and eukaryotic read proportion using PhyloFlash**

* 1. Use PhyloFlash18 to extract and taxonomically assign short subunit (SSU) rRNA fragments from quality-filtered reads using Silva v138.1 database19 as reference, with default parameters and 100% clustering id.
  2. Visualize the data and generate figures using software of choice (here R v4.2.120).

### **Data availability**

Sequence data for the three *Culex quinquefasciatus* specimen is available in the European Nucleotide Archive with the following accession number: PRJEB77608. Code for the reproducible workflow is available at <https://jreveillaud.github.io/RosaBioinfo/>.

## **REFERENCES:**

1. Gottlieb, Y., Lalzar, I., Klasson, L. Distinctive Genome Reduction Rates Revealed by Genomic Analyses of Two *Coxiella-* Like Endosymbionts in Ticks. *Genome Biol. Evol*. **7** (6), 1779–1796, doi: 10.1093/gbe/evv108 (2015).

2. Eren, A.M., Vineis, J.H., Morrison, H.G., Sogin, M.L. A Filtering Method to Generate High Quality Short Reads Using Illumina Paired-End Technology. *PLoS ONE*. **8** (6), e66643, doi: 10.1371/journal.pone.0066643 (2013).

3. Minoche, A.E., Dohm, J.C., Himmelbauer, H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biol*. **12** (11), R112, doi: 10.1186/gb-2011-12-11-r112 (2011).

4. Bushnell, B. BBMap. at <sourceforge.net/projects/bbmap/> (2014).

5. Köster, J., Rahmann, S. Snakemake--a scalable bioinformatics workflow engine. *Bioinformatics*. **28** (19), 2520–2522, doi: 10.1093/bioinformatics/bts480 (2012).

6. Shaiber, A. *et al.* Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. *Genome Biol*. **21** (1), 292, doi: 10.1186/s13059-020-02195-w (2020).

7. Eren, A.M. Community-led, integrated, reproducible multi-omics with anvi’o. *Nat. Microbiol*. **6** (1), 3–6, doi: 10.1038/s41564-020-00834-3 (2021).

8. Klasson, L. *et al.* Genome Evolution of Wolbachia Strain wPip from the Culex pipiens Group. *Mol Biol Evol*. **25** (9), 1877–1887, doi: 10.1093/molbev/msn133 (2008).

9. Li, D., Liu, C.-M., Luo, R., Sadakane, K., Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. **31** (10), 1674–1676, doi: 10.1093/bioinformatics/btv033 (2015).

10. Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. **11** (1), 119, doi: 10.1186/1471-2105-11-119 (2010).

11. Kim, D., Song, L., Breitwieser, F.P., Salzberg, S.L. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res*. **26** (12), 1721–1729, doi: 10.1101/gr.210641.116 (2016).

12. Eddy, S.R. Accelerated Profile HMM Searches. *PLoS Comput Biol*. **7** (10), e1002195, doi: 10.1371/journal.pcbi.1002195 (2011).

13. Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. **10** (3), R25, doi: 10.1186/gb-2009-10-3-r25 (2009).

14. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience*. **10** (2), giab008, doi: 10.1093/gigascience/giab008 (2021).

15. Alneberg, J. *et al.* Binning metagenomic contigs by coverage and composition. *Nat. Methods*. **11** (11), 1144–1146, doi: 10.1038/nmeth.3103 (2014).

16. Eren, A.M. *et al.* Anvi’o: an advanced analysis and visualization platform for ‘omics data. *PeerJ*. **3**, e1319, doi: 10.7717/peerj.1319 (2015).

17. Lee, M.D. GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics*. **35** (20), 4162–4164, doi: 10.1093/bioinformatics/btz188 (2019).

18. Gruber-Vodicka, H.R., Seah, B.K.B., Pruesse, E. phyloFlash: Rapid Small-Subunit rRNA Profiling and Targeted Assembly from Metagenomes. *mSystems*. **5** (5), doi: 10.1128/mSystems.00920-20 (2020).

19. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. **41** (Database issue), D590–D596, doi: 10.1093/nar/gks1219 (2013).

20. R Core Team R: A language and environment for statistical computing. at <https://www.R-project.org/> (2021).