

The Complexities and Nuances of Analyzing the Genome of *Drosophila ananassae* and Its *Wolbachia* Endosymbiont

Julie C. Dunning Hotopp*¹ and Lisa Klasson[†]

*Institute for Genome Sciences, Department of Microbiology and Immunology, Greenebaum Cancer Center, University of Maryland School of Medicine, Baltimore, Maryland 21201 and [†]Molecular Evolution, Department of Cell and Molecular Biology, Uppsala University, 75124 Sweden

ORCID IDs: 0000-0003-3862-986X (J.C.D.H.); 0000-0002-5874-7153 (L.K.)

ABSTRACT In “Retrotransposons Are the Major Contributors to the Expansion of the *Drosophila ananassae* Muller F Element,” Leung *et al.* (2017) improved contigs attributed to the Muller F element from the original CAF1 assembly, and used them to conclude that most of the sequence expansion of the fourth chromosome of *D. ananassae* is due to a higher transposon load than previously thought, but is not due to *Wolbachia* DNA integrations. While we do not disagree with the first conclusion, the authors base their second conclusion on the lack of homology detected between their improved CAF1 genome assembly attributed to *D. ananassae* and reference *Wolbachia* genomes. While the consensus CAF1 genome assembly lacks any sequence similarity to the reference genome of the *Wolbachia* endosymbiont of *Drosophila melanogaster* (wMel), numerous studies from multiple laboratories provide experimental support for a large lateral/horizontal gene transfer (LGT) of a *Wolbachia* genome into this *D. ananassae* line. As such, we strongly suspect that the original whole genome assembly was either constructed after the removal of all *Wolbachia* reads, or that *Wolbachia* sequences were directly removed from the contigs in the CAF1 assembly. Hence, Leung *et al.* (2017) could not have identified the *Wolbachia* LGT using the CAF1 assembly. This manuscript by Leung *et al.* (2017) highlights that an assembly of the *Wolbachia* sequence reads and their mate pairs was erroneously attributed solely to the *Wolbachia* endosymbiont, albeit before we understood the extent of LGT in *D. ananassae*. As such, we recommend that the sequences deposited at the National Center for Biotechnology Information (NCBI) under PRJNA13365 should not be attributed to *Wolbachia* endosymbiont of *D. ananassae*, but should have their taxonomy reclassified by NCBI as “Unclassified sequences.” As our knowledge about genome biology improves, we need to reconsider and reanalyze earlier genomes removing the prejudice introduced from now defunct paradigms.

We were interested to read the recent paper by Leung *et al.* (2017) entitled “Retrotransposons Are the Major Contributors to the Expansion of the *Drosophila ananassae* Muller F Element.” Leung *et al.* (2017) use contigs attributed to the Muller F element from the original CAF1

assembly (Zimin *et al.* 2008), as well as improvements they made, to conclude that most of the sequence expansion of the fourth chromosome of *D. ananassae* is due to a higher transposon load than previously thought, but is not due to *Wolbachia* DNA integrations. Although we do not disagree with the first conclusion, we were surprised to see that the authors stated that the *Wolbachia* sequences integrated into the *D. ananassae* genome are a minor contributor to expansion of the Muller F Element. The authors base their conclusions on the lack of homology detected between their improvements to the CAF1 genome assembly attributed to *D. ananassae* and reference *Wolbachia* genomes. However, the CAF1 assembly was undertaken at a time when the dogma was that animal genomes did not contain lateral/horizontal gene transfer (LGT) from bacteria. As such, we strongly suspect that the original whole

Copyright © 2018 Dunning Hotopp and Klasson

doi: <https://doi.org/10.1534/g3.117.300164>

Manuscript received August 17, 2017; accepted for publication October 11, 2017

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

¹Corresponding author: Institute for Genome Sciences, 801 W. Baltimore St, Suite 600, Baltimore, MD 21201. E-mail: jdhotopp@som.umaryland.edu

genome assembly was either constructed after the removal of all of the reads matching the closed/complete genome of the *Wolbachia* endosymbiont of *Drosophila melanogaster* (*wMel*), the only *Wolbachia* genome available at the time, or that contigs in the assembly matching the closed/complete genome of the *Wolbachia* endosymbiont of *D. melanogaster* (*wMel*) were removed from the two assemblies used to construct the CAF1 assembly. Despite our best efforts to clarify this by contacting as many of the assembly experts involved at that time as we could find, we cannot say definitively. However, this deduction is supported by the copious numbers of raw sequence reads with homology to *Wolbachia*, and that the only portions of *Wolbachia* sequence in the CAF1 assembly are those regions that do not share homology with the *wMel* genome, as discussed previously (Klasson *et al.* 2009). Given the dogma at that time, it is reasonable that either of these approaches were undertaken, but unfortunately went unreported. Hence, Leung *et al.* (2017) could not have identified the *Wolbachia* LGT on the fourth or any chromosome of *D. ananassae* using the CAF1 assembly, as most of the *Wolbachia* sequences have been removed.

Furthermore, given that the original whole genome sequencing project on *D. ananassae* (Drosophila 12 Genomes Consortium *et al.* 2007) did not rely on genomic DNA prepared from embryos of an antibiotic-treated line to remove the *Wolbachia* endosymbionts (T. Markow, personal communication), we now understand that the “*Wolbachia*” sequence reads are a mixture of *Drosophila* and *Wolbachia* sequences. Unfortunately, given the very high similarity between the LGT and the residing bacterium, it is not possible to assign the reads to the *Drosophila* genome or the *Wolbachia* genome. This collective work on *D. ananassae* genomics, including this manuscript by Leung *et al.* (2017), highlights that an assembly of these sequences was erroneously attributed solely to the *Wolbachia* endosymbiont (Salzberg *et al.* 2005), albeit before we understood the extent of LGT that occurs between *Wolbachia* and its hosts. We now know that there are multiple copies of the *Wolbachia* genome integrated into the *D. ananassae* genome with insertional mutagenesis of the LGT by retrotransposons active in *D. ananassae* (Klasson *et al.* 2014; Dunning Hotopp *et al.* 2007), making it nearly impossible to resolve the sequence and organization with next generation sequencing techniques or bacterial artificial chromosomes. Therefore, we used fluorescence *in situ* hybridization and microscopy to demonstrate the likely location is the Muller F element (Klasson *et al.* 2014). The methods Leung *et al.* (2017) used to make the improvements to the contigs attributed to the Muller F element would not be sufficient to assemble the massive LGT from *Wolbachia* into *D. ananassae*, and, thus, are not sufficient to contradict this result.

This highlights the complexity of genome sequencing projects and their interpretation. While genomes are often presented and considered as final, static, and definitive objects, the experiments undertaken to obtain these sequences have nuances and/or assumptions that often need to be considered for proper interpretation of subsequent results. This also highlights that, as our knowledge about genome biology improves, we may need to reconsider and reanalyze earlier genomes removing the prejudice introduced from now defunct paradigms. As such, we recommend that the sequences deposited at the NCBI under PRJNA13365 should not be attributed to *Wolbachia* endosymbiont of *D. ananassae*, but should have their taxonomy reclassified by NCBI as “Unclassified sequences.”

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health through an NIH Director’s Transformative Research Award (1-R01-CA206188) and the National Science Foundation Advances in Biological Informatics (ABI-1457957) to J.C.D.H.

Note added in proof: See Leung and Elgin in this issue for a related work.

LITERATURE CITED

- Drosophila 12 Genomes Consortium Clark, A. G., M. B. Eisen, D. R. Smith, C. M. Bergman *et al.*, 2007 Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450: 203–218.
- Dunning Hotopp, J. C., M. E. Clark, D. C. Oliveira, J. M. Foster, P. Fischer *et al.*, 2007 Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 317: 1753–1756.
- Klasson, L., J. Westberg, P. Sapountzis, K. Näslund, Y. Lutnaes *et al.*, 2009 The mosaic genome structure of the *Wolbachia* *wRi* strain infecting *Drosophila simulans*. *Proc. Natl. Acad. Sci. USA* 106: 5725–5730.
- Klasson, L., N. Kumar, R. Bromley, K. Sieber, M. Flowers *et al.*, 2014 Extensive duplication of the *Wolbachia* DNA in chromosome four of *Drosophila ananassae*. *BMC Genomics* 15: 1097.
- Leung, W., C. D. Shaffer, E. J. Chen, T. J. Quisenberry, K. Ko *et al.*, 2017 Retrotransposons are the major contributors to the expansion of the *Drosophila ananassae* Muller F element. *G3* 7: 2439–2460.
- Salzberg, S. L., J. C. Dunning Hotopp, A. L. Delcher, M. Pop, D. R. Smith *et al.*, 2005 Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. *Genome Biol.* 6: R23.
- Zimin, A. V., D. R. Smith, G. Sutton, and J. A. Yorke, 2008 Assembly reconciliation. *Bioinformatics* 24: 42–45.

Communicating editor: B. Andrews