Correspondence

New names for old strains? *Wolbachia w*Sim is actually *w*Ri

Iñaki Iturbe-Ormaetxe, Markus Riegler and Scott L O'Neill

A response to **Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species** by SL Salzberg, JC Dunning Hotopp, AL Delcher, M Pop, DR Smith, MB Eisen and WC Nelson. *Genome Biology* 2005, **6:**R23

Address: School of Integrative Biology, The University of Queensland, St Lucia, Brisbane, QLD 4072, Australia.

Correspondence: Scott O'Neill. E-mail: scott.oneill@uq.edu.au

Published: 24 June 2005

Genome ***Biology*** 2005, **6:**401 (doi:10.1186/gb-2005-6-7-401)

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2005/6/7/401

A recent paper published by Salzberg *et al.* [1] reports the discovery, assembly and comparative analysis of three partial *Wolbachia* endosymbiont genomes. These data were retrieved from the Trace Archive [2] from sequencing projects that were focused on the endosymbiont hosts - *Drosophila simulans*, *D. ananassae* and *D. mojavensis* - using the fully sequenced *w*Mel *Wolbachia* genome [3] as a probe. Salzberg *et al.* refer to these partial genomes as belonging to *Wolbachia* strains *w*Sim, *w*Ana and *w*Moj respectively [1]. These strain names are new constructions and it appears that the annotated *w*Sim genome sequence is essentially identical to the previously described *w*Ri strain [4] and should be named accordingly.

There is a large body of previous work on the biology of *Wolbachia* infections of *D. simulans*. To date, five *Wolbachia* strains have been described from *D. simulans* (for a review see [5]), three of them belonging to group A, *w*Au [6], *w*Ri [7] and *w*Ha [4], and two belonging to group B, *w*No [8] and *w*Ma [9]. When the partial genome sequence of *w*Sim [1] is compared to previously published sequences of the different *D. simulans Wolbachia* strains, it is clear that *w*Sim is most likely to be the *w*Ri *Wolbachia* strain that has been extensively studied over the years. Blastn analysis of numerous *w*Ri sequences available at GenBank (accession numbers X61770, 16S rRNA; AB002288, groES and groEL; AB036661, bacteriophage WO gene for capsid protein; AF348330, *ubiA*, *rbfA*, *infB*, *nusA*, and *acrD* genes; AJ012073, *glnA* and *dnaA* genes and two genes encoding hypothetical proteins; and AJ580923, *wsp*B gene) reveals that the *w*Ri sequences are 99-100% identical to the partially assembled *w*Sim genome [1]. On the basis of the molecular data publicly available in National Center of Biotechnology Information (NCBI) databases it is apparent that the strain designated as *w*Sim by Salzberg *et al.* [1] is actually *w*Ri. This strain was first described phenotypically by Hoffmann *et al.* in 1986 [7] in *D. simulans* collected in Riverside, California. *w*Ri is characterized by the ability to induce high levels of cytoplasmic incompatibility (CI) in its native *D. simulans* host [7] and has the ability to spread quickly through host populations by the induction of CI [10,11]. Biogeographic studies have revealed that *w*Ri is currently the most abundant strain infecting continental populations of *D. simulans* [12].

Finally, the Trace Archive for *D. simulans* contains reads from various *D. simulans* lines [13] of different biogeographic origin: wsim501 and sim6, both North American and most likely infected by *w*Ri, and simNC48S from New Caledonia and potentially infected with *w*No and *w*Ha [12]. Therefore, it would be helpful if the authors could clarify which Trace data were used for the assembly of the *w*Sim genome, as it might be possible that the assembly reported is chimeric, containing predominantly sequences from *w*Ri and possibly some sequence from other *Wolbachia* strains.

While the discovery of these partial genome sequences in the Trace Archive is an exciting development, it is important that the finding is connected to the large established literature in this field if the data is to be of most value to the scientific community.

Julie Dunning Hotopp, William C Nelson and Steven L Salzberg respond:

We are aware that our newly discovered *Wolbachia* strain from the ongoing *D. simulans* sequencing project, which we have designated *w*Sim [1], might be the same as *w*Ri, as Iturbe-Ormaetxe *et al.* claim. Unfortunately, the evidence to support this claim, which is entirely based on

sequence similarity, fails to distinguish it from other hypotheses. Iturbe-Ormaetxe *et al*. searched *w*Sim against fragments of several *D. simulans Wolbachia* strains and found that *w*Ri was the best match; from this they conclude that *w*Sim and *w*Ri are the same. If one searches these same *w*Ri fragments against *w*Ana, however, one finds an even closer match to *w*Ana.

The small number of *w*Ri genomic fragments available in GenBank (representing less than 18 kilobases (kb), not 'numerous sequences' despite the contention of Iturbe-Ormaetxe *et al*.) are diverging too slowly to be used for definitive strain identification; in some cases even the *w*Ri and *w*Mel sequences cannot be differentiated. The *wsp* gene is simply missing from our *w*Sim assembly, but is 99.9% identical between *w*Ana and *w*Ri. The *w*Ri sequence of *wspB* is 99.2% identical over 788 base-pairs (bp) to *w*Ana and 98% identical over 226 bp to *w*Sim. The two longest genome fragments of *w*Ri, AF348330 (9,235 bp) and AJ012073 (4,838 bp), match *w*Sim and *w*Ana equally well. Clearly, *w*Ri, *w*Sim, and *w*Ana are closely related, as discussed in Table 2 of our paper [1], but if one uses sequence identity to assign strain designations, then *w*Ri looks more like *w*Ana than *w*Sim.

As should be apparent from this analysis, the assertion made by Iturbe-Ormaetxe *et al*. that *w*Sim = *w*Ri rests on a logical fallacy; that is, that if the best unidirectional BLAST matches of genome A (*w*Sim) correspond to genome B (*w*Ri), then A = B. This ignores that fact genome B might have a better match to genome C - in this case *w*Ana. Even more critical is the fact that only a tiny fraction of *w*Ri has been sequenced. The BLAST analysis shows only that *w*Sim and *w*Ri are highly similar across a few sequence fragments representing less than 1.5% of their genomes.

We are aware that *D. simulans* has been reported to carry the *w*Ri strain as well as the strain we designate *w*Sim, and that some of the sequenced *D. simulans* strains carry the white mutation [13,14]. It should be noted, however, that although the *D. simulans* sequencing project included a mixture of three *Drosophila* strains, virtually all (99.9%) of the *w*Sim sequences came from just one strain, sim6; thus both *w*Sim and *w*Ri were found in the California population of *D. simulans*. Neither this nor the BLAST alignments are, however, sufficient evidence to collapse the strains into one: *Wolbachia* species from closely related insect species often retain different strain identifiers [15-17] despite sharing some identical gene sequences. This is important because sometimes these *Wolbachia* infections result in different host phenotypes [16]. Less commonly, *Wolbachia* species with identical *wsp* genes isolated from the same insect species (for example, *D. simulans*) retain different strain designations [15].

This nomenclature is also common in other prokaryotes. Organisms with identical multi-locus sequencing typing (MLST) profiles isolated from the same geographical area will be given different strain designations to preserve information about their origin. This may be important if they have genomic rearrangements and single-nucleotide polymorphisms (SNPs) that confer different phenotypes. In *Wolbachia*, genomic rearrangements appear common [1,3], which may support the maintenance of separate strain designations to differentiate ancestry. In the absence of complete genome sequences, definitive genotyping assays, or phenotypic characterization of *w*Sim, resolving strain differences is clearly complicated and beyond the scope of our paper.

Julie Dunning Hotopp, William C Nelson and Steven L Salzberg

Correspondence should be sent to Steven L Salzberg: Center for Bioinformatics and Computational Biology, University of Maryland Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA. E-mail: salzberg@umd.edu

References

1. Salzberg SL, Dunning Hotopp JC, Delcher AL, Pop M, Smith DR, Eisen MB, Nelson WC: **Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species.** *Genome Biol* 2005, **6:**R23.
2. **Trace Archive v.3** [http://www.ncbi.nih.gov/Traces]
3. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N, *et al.*: **Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements.** *PLoS Biol* 2004, **2:**E69.
4. O'Neill SL, Karr TL: **Bidirectional incompatibility between conspecific populations of *Drosophila simulans*.** *Nature* 1990, **348:**178-180.
5. Merçot H, Charlat S: ***Wolbachia* infections in *Drosophila melanogaster* and *D. simulans*: polymorphism and levels of cytoplasmic incompatibility.** *Genetica* 2004, **120:**51-59.
6. Hoffmann AA, Clancy D, Duncan J: **A naturally-occurring *Wolbachia* infection in *Drosophila simulans* that does not cause cytoplasmic incompatibility.** *Heredity* 1996, **76:**1-8.
7. Hoffmann AA, Turelli M, Simmons GM: **Unidirectional incompatibility between populations of *Drosophila simulans*.** *Evolution* 1986, **40:**692-701.
8. Merçot H, Llorente B, Jacques M, Atlan A, Montchamp-Moreau C: **Variability within the Seychelles cytoplasmic incompatibility system in *Drosophila simulans*.** *Genetics* 1995, **141:**1015-1023.
9. Giordano R, O'Neill SL, Robertson HM: ***Wolbachia* infections and the expression of cytoplasmic incompatibility in *Drosophila sechellia* and *D. mauritiana*.** *Genetics* 1995, **140:**1307-1317.
10. Turelli M, Hoffmann AA: **Rapid spread of an inherited incompatibility factor in California *Drosophila*.** *Nature* 1991, **353:**440-442.
11. Turelli M, Hoffmann AA: **Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from natural populations.** *Genetics* 1995, **140:**1319-1338.
12. Ballard JWO: **Sequential evolution of a symbiont inferred from the host: *Wolbachia* and *Drosophila simulans*.** *Mol Biol Evol* 2004, **21:**428-442.
13. **Washington University in St Louis Genome Sequencing Center: *D. simulans*** [http://www.genome.wustl.edu/projects/simulans/index.php?species=1]
14. Hoffmann AA, Turelli M, Harshman LG: **Factors affecting the distribution of cytoplasmic incompatibility in *Drosophila simulans*.** *Genetics* 1990, **126:**933-948.
15. Dyer KA, Jaenike J: **Evolutionarily stable infection by a male-killing endosymbiont in *Drosophila innubila*: molecular evidence from the host and parasite genomes.** *Genetics* 2004, **168:**1443-1455.
16. Jiggins FM, Bentley JK, Majerus ME, Hurst GD: **Recent changes in phenotype and patterns of host specialization in *Wolbachia* bacteria.** *Mol Ecol* 2002, **11:**1275-1283.
17. Kikuchi Y, Fukatsu T: **Diversity of *Wolbachia* endosymbionts in heteropteran bugs.** *Appl Environ Microbiol* 2003, **69:**6082-6090.