

Supplementary materials for: Heteroplasmy, structure and evolution of an atypical mitochondrial genome untangled by PacBio sequencing

Jean Peccoud, Mohamed Amine Chebbi, Alexandre Cormier, Bouziane Moumen, Clément Gilbert, Isabelle Marcadé, Christopher Chandler, Richard Cordaux

Supplementary text

[Absence of detection of circular monomers in oniscid mtDNA](#)

It has been proposed that linear mtDNA monomers in oniscids species could replicate themselves by transitioning to a rare circular form resembling a “standard” animal mtDNA molecule (Doublet, et al. 2013). We searched for such forms using our high-depth short read data. This search was restricted to lineage *A. vulgare* BF because it had both the highest sequencing depth of short reads (fig. 3) and the highest frequency of monomers devoid of non-complementary bases (fig. 7B), which may be produced by monomer replication.

The circular monomer form joins the “head” and “tail” of a genome unit, linking the Cytochrome b gene to the 12S rRNA gene (Doublet, et al. 2013). We constructed this sequence by placing two genome units in tandem fashion. The second unit was pasted just after the junction between genome units in the dimer. We verified that the transition between units was similar to the sequence obtained by Doublet, et al. (2013).

We mapped Illumina paired reads on this construction, using the same protocol as described in the method section. We looked for pairs of reads mapping on both sides of the junction. We imposed that no read of the pair covered the junction, as such read may partially align only because it was sequenced from the end of a genomic unit of a dimer. We indeed saw that reads covering the junction aligned properly on only one of its flanking region.

No pair of reads passed our criteria. As a verification, we applied similar criteria for hypothetical junctions located 300bp upstream and downstream the real junction. These returned 14441 and 6180 read pairs, respectively. We conclude that no molecule of the circular monomeric configuration was sequenced.

mtDNA monomer replication and frequency of dimers with asymmetric haplotypes

If monomers with non-complementary bases replicate into dimers, these dimers should harbor asymmetric haplotypes (fig. 7A). Let p_0 be the proportion of dimers carrying asymmetric haplotypes at generation 0. A generation is defined as a replication cycle. Let q be the proportion of linear monomers that carry non-complementary bases, which is assumed to be constant. We define the fraction of dimers arising from the replication of other dimers as f , $1-f$ being the fraction of dimers arising from the replication of monomers.

Assuming that asymmetry or non-complementary of bases do not affect replication rates, the proportion of dimers with asymmetric haplotypes at generation 1 (p_1) is

$$p_1 = p_0 \cdot f + q(1 - f),$$

hence

$$p_1 = f(p_0 - q) + q \quad (\text{equation 1}).$$

At generation 2,

$$p_2 = f(p_1 - q) + q.$$

Replacing p_1 according to equation 1 gives

$$p_2 = f(f(p_0 - q) + q - q) + q,$$

which reduces to

$$p_2 = f^2(p_0 - q) + q.$$

At generation n ,

$$p_n = f^n(p_0 - q) + q.$$

At large n , $p_n \approx q$ if $f < 1$. In other words, if some monomers replicate into dimers, the fraction of dimers with asymmetric haplotype should be similar to the fraction of monomers with non-complementary bases.

Sequencing errors cannot explain fragments carrying both complementary and non-complementary bases

For a site of a DNA fragment to be considered as carrying complementary or non-complementary bases (with a complementarity index i of 1 or -1, respectively, see methods), at least two sequencing reads among those from a given DNA strand must support the right base, and none should support the other dominant base. For sequencing errors to produce a DNA fragment with a mixture of sites with complementary and non-complementary bases (“mixed” DNA fragment), these errors have to replace the original base by the same new base in all the reads of a strand. We tested whether this could have occurred in *A. vulgare* WXf,

which comprises the vast majority of mixed DNA fragments. For each of its six heteroplasmic sites, we randomly picked a base that was not among the dominant ones (table 2), as its presence at this site most likely results from a sequencing error. We then looked for fragments in which this base was found in at least two sequencing reads of the same strand, and which lack both dominant bases among reads of this strand. We found no such fragment. We conclude that sequencing errors cannot have produced the number of mixed DNA fragments we observed in this lineage.

Reference

Doublet V, Helleu Q, Raimond R, Souty-Grosset C, Marcade I. 2013. Inverted Repeats and Genome Architecture Conversions of Terrestrial Isopods Mitochondrial DNA. *J. Mol. Evol.* 77:107-118.

Figures S1-S3

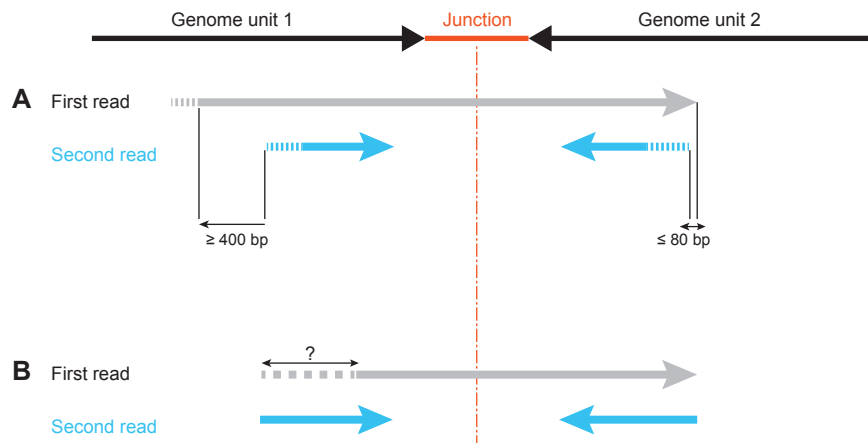


Figure S1. Inference of mapping orientation of a read from its mapping coordinates. If a read does not span the junction between genome units (whose middle is represented by a red dashed line), it can map equally well on either genome unit in opposite orientations. Mapping orientation is shown by the direction of arrows. However, the second read of a polymerase read should have its start position aligned near the edge of the region mapped by the previous read (see fig. 4A, main text). A) If one alignment places the start of a read (considering a possible unaligned “clipped” part shown as a dashed segment) at least 400 bp within the region covered by the previous read (not considering its clipped parts), while the other places this start coordinate at less than 80 bp from the edge of the region mapped by the previous read, we considered the orientation of the latter alignment to be the correct one. B) *A priori*, if one alignment places the start coordinate of the second read well outside the region covered by the previous read, its orientation should not be correct. However, first reads may not always start just after the SMRT bell and may be incomplete. Uncertainty about read length is represented by the question mark. We did not infer a more likely orientation in the case described in B).

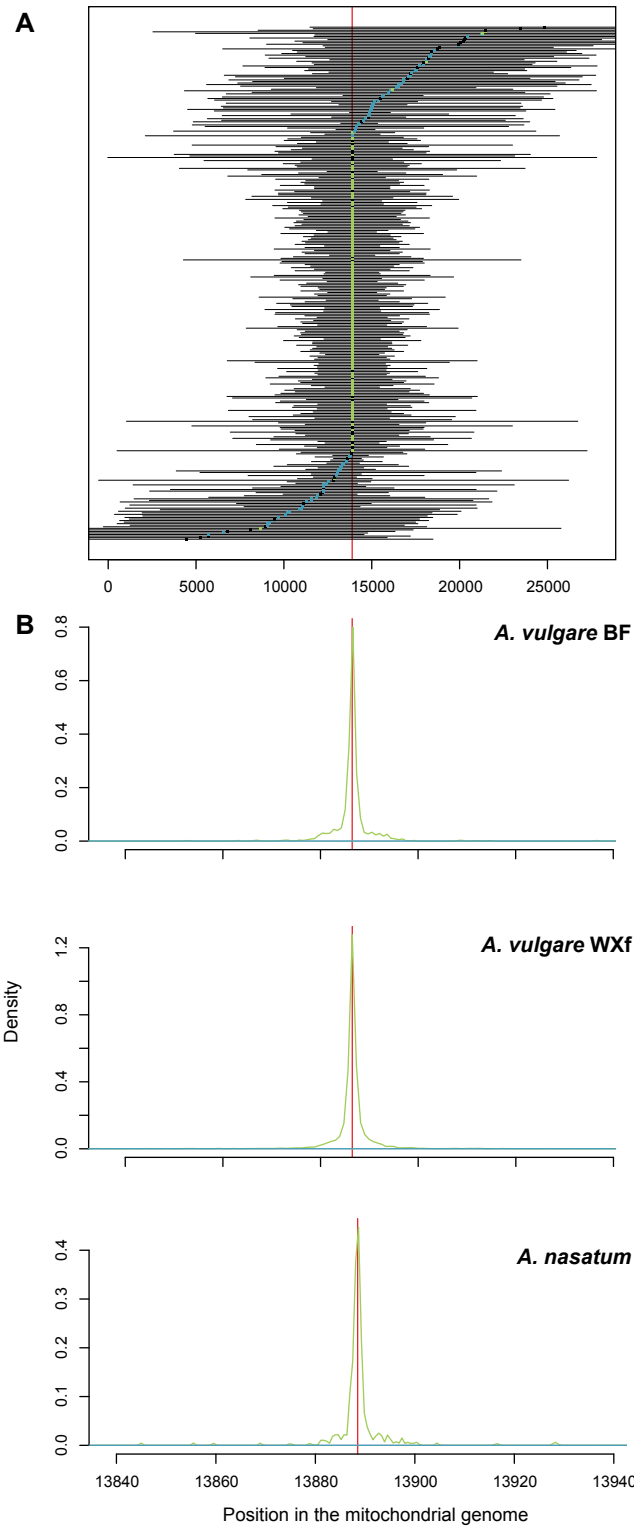


Figure S2. A) Representation of the mapping of SMRT reads from *A. vulgare* BF on its dimeric mitochondrial genome. Only reads covering the junction between genome units (represented as a red line) are shown. The region covered by a read (including its clipped parts) is drawn as a horizontal segment, the middle of which is represented by a square-shaped point. Green points represent molecules (polymerase reads) whose reads all aligned in

the same orientation on the reference, likely representing linear monomers. Blue points represent molecules whose reads aligned in alternate orientations, likely representing dimers. Black points represent molecules where mapping orientation could not be reliably determined for more than one read. For each molecule, only the read whose center mapped more closely to the middle of the junction is represented. B) Density distribution of the middle of regions mapped by reads obtained from different lineages. Colors represent different types of molecules as in A). Red vertical lines are drawn at the midpoints of junctions between “heads” of genome units.

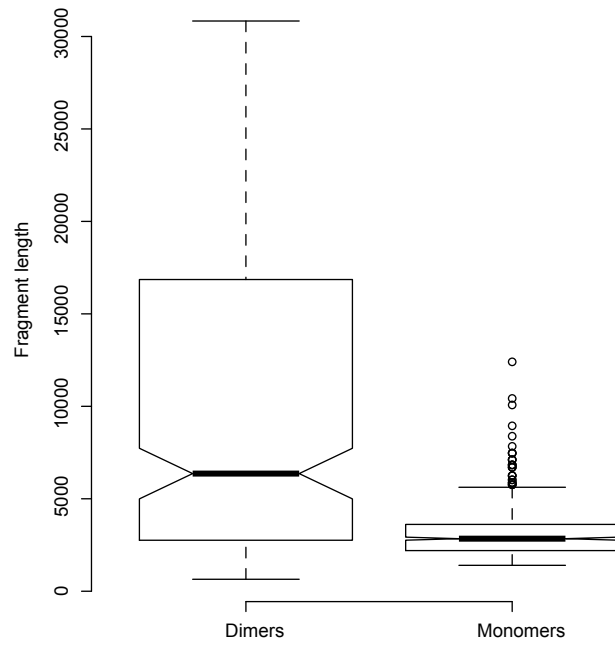


Figure S3. Length of SMRT-sequenced mtDNA fragments assigned to monomers or dimers, for *A. vulgare* BF. The length of fragments assigned to monomers was taken as half the length of reads (see fig. 4B, main text). Fragments sequenced in the other *Armadillidium* lineages present similar length patterns (not shown).