A lack of evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*

Georgios Koutsovoulos[1], Sujai Kumar[1], Dominik Laetsch[1], Jennifer Daub[1], Claire Conlon[1], Habib Maroon[1], Fran Thomas[1], Lewis Stevens[1], Aziz Aboobaker[2] and Mark Blaxter[1*]

1 Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3FL, UK

2 Department of Zoology, University of Oxford, South Parks Road, Oxford OX, UK.

Supplemental File 5: Library insert sizes

We estimated library insert sizes by mapping read pairs to the initial single-end CLC assembly. This works well for short insert libraries, but less well for mate-pairs, where many read pairs do not map to the same contig. This failure to map to the same contig means that the mate pair reads are under-counted.

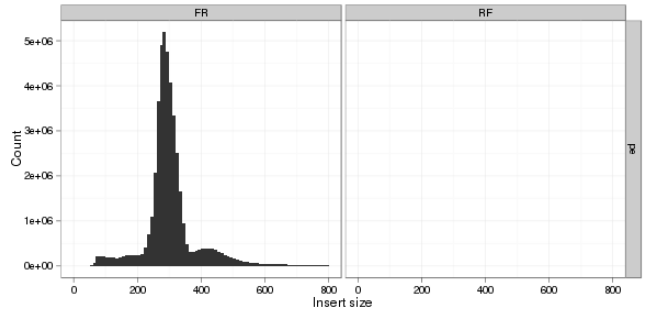
The insert size distribution of the "300 base" standard library on the preliminary SE assembly had a median of 292 bases (Standard Deviation (SD) 96 bases) (Figure S4A). The insert size distribution of the 4000 base mate pair library has a median of 1133 bases (1460 bases SD) (Figure S4B). There were many mate-pair fragments with very small virtual inserts in this library.

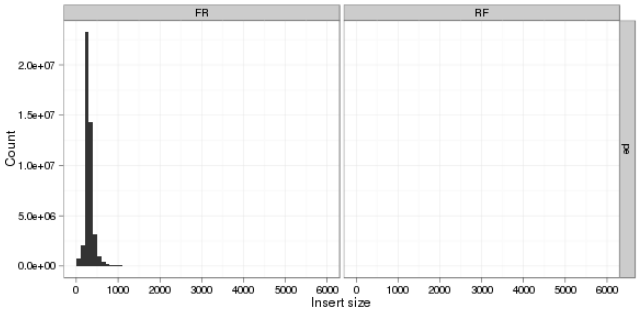
Figure S5 Insert size estimation for Illumina libraries

A Insert size distribution for the short-insert library. Left panel, left graph: read pairs mapping in the expected F-R orientation. Left panel, right graph: read pairs mapping in the unexpected R-F (mate) orientation. The right panel shows the same data with an expanded X-axis.

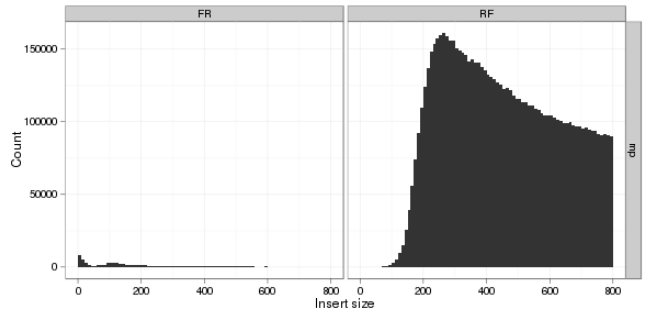
B Insert size distribution for the mate-pair library. Left panel, left graph: read pairs mapping in the unexpected F-R orientation. Left panel, right graph: read pairs mapping in the expected R-F (mate) orientation. The right panel shows the same data with an expanded X-axis.

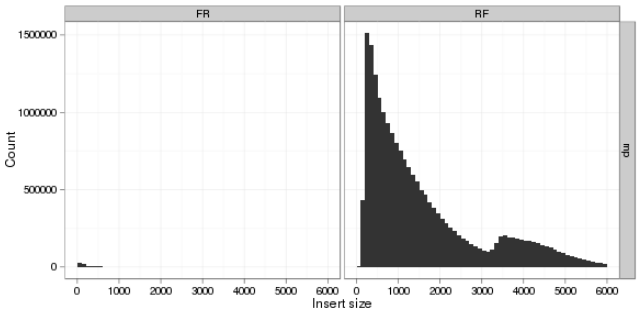
A


FR
RF
pe
Count
Insert size


FR
RF
pe
Count
Insert size

B


FR
RF
mp
Count
Insert size


FR
RF
mp
Count
Insert size