Methods

*Generation of our 212.4Mb assembly by post-filtering of our raw 252Mb assembly.*

All predicted proteins from our 252Mb assembly were used as input for BLASTp against NCBI’s nr protein database. The top hit for each predicted protein was used as input into Galaxy tool’s Fetch taxonomic representation and Summarize taxonomy tools. Scaffolds with >10 genes all of bacterial origin were removed from our 252Mb assembly, resulting in a post-filtered assembly of 212.4Mb analyzed in (Boothby et al., 2015).

*HGT index analysis*

Gene Ids for metazoa, bacteria, plants, fungi, and archaea were download from NCBI’s protein database. Database aliases were generated using NCBI’s blastdb\_aliastool. Predicted protein sequences for each assembly were used as query for BLASTp (BLAST 2.3.0+ - with default parameters) searches against metazoan, bacterial, plant, fungi, and archaea databases. The top hit for each tardigrade sequence from each database were used to perform HGT index calculations. Only hits with an Evalue <=1e-5 were retained. As in Boschetti *et al*., 2012, sequences with no hits with Evalue <=1e-5 were excluded from further analysis. HGT calculations for each gene were performed by subtracting the maximum metazoan bitscore from the maximum non-metazoan (max non-metazoan bitscore – max metazoan bitscore). Genes with an HGT index score >=30 were considered foreign.

*Identification of genes with prokaryotic but not eukaryotic genes*

Output from HGT index analysis was parsed to identify genes with a hit to bacteria and/or archaea, but not to metazoan, plant, or fungal sequences in the nr protein database.

*Identification of Class C foreign genes*

Class C genes were identified according to the methods detailed in Crisp *et al.,* 2015. HGT index scores were calculated as detailed above and in Boschetti *et al.,* 2012. Predicted protein sequences for each assembly were used as query for BLASTp (BLAST 2.3.0+ - with default parameters) searches against metazoan, bacterial, plant, fungi, and archaea databases. The top hit for each tardigrade sequence from each database were used to perform HGT index calculations. Only alignments with an Evalue <=1e-5 were retained. As in Boschetti *et al*., 2012, sequences with no hits with Evalue <=1e-5 were excluded from further analysis. HGT calculations for each gene were performed by subtracting the maximum metazoan bitscore from the maximum non-metazoan (max non-metazoan bitscore – max metazoan bitscore). Only genes with at least one non-metazoan alignment with a bitscore >=100 and an HGT index of >=30 were retained (Class C genes).

*Anvi’o visualization of scaffold coverage*

Anvi’o 1.2.2 was used to visualize scaffold coverage using 14 next-generation read datasets (Eren et al., 2015). Unless otherwise noted all programs were used with default settings. Reads were mapped against assemblies using Bowtie2 with default settings to generate Sam files. Sam files were converted to Bam format using Samtools. Bam files were index using anvi-init-bam tool using default settings. Anvi’o databases were generated using anvi-gen-contigs-database with split length set to 100,000,000 to ensure no contigs were split. Indexed Bam files were used with Anvi’o databases and anvi-profile using –M 250. Anvi’o profiles were merged using anvi-merge. Visualization was carried out using anvi-interactive with clustering performed using both sequence composition and coverage.

*CEGMA analysis*

CEGMA analysis was performed as described in (Parra et al, 2007) using CEGMA\_v2.5.

*H. dujardini EST representation*

*Hypsibius dujardini* ESTs were downloaded from NCBI’s EST database. BLAT (v35) was used to query ESTs against assemblies. The baa.pl script (Ryan, 2013) and BLAT output were used to assess EST representation in various genome assemblies.

*Assembly size and N50 assessments*

Assembly size and N50 calculations were derived using Joseph Fass’s count\_fasta.pl (http://wiki.bioinformatics.ucdavis.edu/index.php/Count\_fasta.pl).