Figure 1 – Identification of foreign sequences in independent *H. dujardini* assemblies by different methods.

(A), Foreign genes in independent *H. dujardini* genome assemblies were identified using the HGT index (1). The percent of all foreign genes in each dataset is shown. Tardigrade data is labeled in blue. Rotifer (animals with high levels of HGT) data is shown in brown. Data for other invertebrates (*C. elegans* and *D. melanogaster*) is labeled in green. Rotifer, *C. elegans,* and *D. melanogaster* data was obtained from (1).

(B), Foreign genes were identified by selecting only those genes with BLAST hits to prokaryotes, but not eukaryotes (Evalue cutoff 1e-5). Note that this is a significantly more stringent approach for identifying foreign genes than the HGT index, and excludes identification of prokaryotic genes with metazoan homologs, non-metazoan eukaryotic genes, or metazoan genes that have been horizontally transferred. The raw number of all foreign genes in each dataset is shown. Tardigrade data is labeled in blue. Rotifer (animals with high levels of HGT) data is shown in brown. Data for other invertebrates (*C. elegans* and *D. melanogaster*) is labeled in green. Rotifer and *C. elegans* data was derived from (1). *D. melanogaster* data was obtained by performing HGT index analysis (see methods) and the applying selection criteria detailed above.

(C), Class C genes were identified in various datasets using the parameters detailed in (2). Data for chordate and ecdysozoan animals was obtained from (2) and the source of other datasets is noted parenthetically on the X-axis. Plotted on the Y-axis is the percent of total genes that are classified as Class C foreign genes according the methods used in (2). Data for rotifer species are colored brown. Data from various tardigrade assemblies are colored blue. Data for other ecdysozoan animals (the group of animals to which tardigrades belong) is colored in green. Data for chordate animals is colored purple. The numbers above bars denote the percent of all genes classified as Class C foreign genes.

Figure 2 – Foreign genes are contained in scaffolds confirmed by multiple datasets.

14 next-generation sequencing read datasets, originating from 4 independent sequencing projects, were mapped against **(A)** Bemm *et al*.'s (3) and **(B)** Koutsovoulos *et al*.'s (4) assembly and visualized using Anvi'o (5). Tracks showing coverage of moleculo long-read (LR), Pacbio, and short-insert library reads (UNC 300, 500, & 800) from our original study (6) are colored blue. Tracks showing coverage by genomic reads from (4) are colored purple. Tracks showing coverage by genomic and pooled RNAseq reads generated by Dr. Kazuharu Arakawa are colored green. The track showing coverage by pooled RNAseq reads generated by Dr. Itai Yanai is colored yellow. Black tick-marks in the outer rings indicate scaffolds that contain foreign sequences identified using the method indicated in parentheses. Highlighted in red are scaffolds that are covered by genomic reads originating from only 1 group's sequencing effort. Highlighted in orange are scaffolds covered by 2 of 3 groups' genomic reads.

Figure 3 – The majority of foreign genes reside on scaffolds with tardigrade genes.

For each foreign gene (identified using the HGT index) we asked if that gene resides on a scaffold that also contains a tardigrade gene. Shown in the graph are the percentages of foreign genes on scaffolds with tardigrade genes in different *H. dujardini* assemblies.

Figure 4 – The distribution and proportion of best BLAST hits for foreign and tardigrade genes across various *H. dujardini* assemblies.
All protein sequences from one assembly were compared to all protein sequences from another assembly using BLASTp. The graphs show the distribution of each protein (represented by a blue dot) as a function of its query coverage (how much of the protein was covered by its best BLAST hit) and percent identity to its best hit in the other assembly. Distributions of foreign (identified using the HGT index - right column) and tardigrade (left column) genes were compared and do not display gross differences in distribution. The number in the upper right-hand corner of each graph shows the percent of genes that align with both a query coverage and percent identity of ≥99% to their best hit in the other genome assembly.

Figure 5 – The effect of removing scaffolds containing foreign genes on the completeness, size, and N50 of assemblies.
For each assembly, scaffolds containing foreign genes (identified using the HGT index) were removed to generate filtered assemblies. The completeness of the original and filtered assembly in terms of representation of core eukaryotic genes (CEGMA) and *H. dujardini* ESTs is shown in **(A)**. **(B)** the size of each original and filtered assembly. **(C)** the scaffold N50 for each original and filtered assembly.

Figure 6 – K-mer analysis of short-read datasets shows signs of heterozygosity within individuals and populations of *H. dujardini* tardigrades.
Modified version of Figure 1 from (7) demonstrating how heterozygosity introduces multiple peaks (one with half the multiplicity of the other) within a k-mer distribution. Homozygous loci accumulate k-mer counts that are equivalent to the average sequencing depth (in this case 50X). Heterozygous loci, for which there are 2 equally represented SNPs, will accumulate k-mer counts that are approximately half the total sequencing coverage (in this case 25X). Thus, a dataset derived from a mixed population of heterozygous specimens should show a k-mer distribution with 2 peaks, one with half the multiplicity of the other.

Figure 7 – SNPs within the *H. dujardini* genome.
Representative portion of Scaffold374 from our postfiltered assembly. Reads from our 300, 500, and 800 short read datasets (generated from a population) were mapped against our assembly, along with paired end reads from Koutsovoulos *et al.,* 2016 (8) (generated from a population) and Arakawa, 2016 (7) (generated from an individual tardigrade) and visualized in IGV. Grey bars represent homozygous loci, whereas colored bar represent SNPs. Blue boxes highlight SNPs detectable in mixed populations while the red box highlights a SNP represented in the individual tardigrade sequenced by Arakawa.

Figure 8 – Saturation of short-read datasets.
Short-read datasets used by Bemm *et al.* (3) for k-mer selection were assessed for saturation using bbmap's bbcountunique.sh script (Version 35.82) with default settings. Raw saturation data was input into Prism (V 6.0g) and fit for visualization using nonlinear regression. The graph shows the percent of non-redundant reads sampled (Y-axis) as a function of total reads sampled (X-axis). Even in the best cases (500bp insert set 2 and 800bp set 1) the number of

non-redundant reads samples is above 75%. A fully (or nearly) saturated dataset should approach 0% non-redundant reads.

Figure 9 – 10 scaffolds with the highest number of Chitinophagaceae genes in our unfiltered (252Mb) assembly.
Genes with reciprocal best BLAST hits to genes from Chitinophagaceae bacteria were identified and enumerated by scaffold. The graph shows the number of Chitinophagaceae genes on the 10 scaffolds containing the highest number of Chitinophagaceae genes. In red are scaffolds that were removed from our raw assembly. Scaffolds retained in our post-filtered assembly are shown in grey. Genes on postfiltered scaffolds were not used for HGT analysis in our original manuscript.

1. Boschetti C, et al. (2012) Biochemical Diversification through Foreign Gene Expression in Bdelloid Rotifers. *PLoS Genet* 8(11):e1003035.

2. Crisp A, Boschetti C, Perry M, Tunnacliffe A, Micklem G (2015) Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol* 16(1):50.

3. Bemm FM, Weiß CL, Schultz J, Förster F The genome of a tardigrade - Horizontal gene transfer or bacterial contamination? *Proc Natl Acad Sci*.

4. Koutsovoulos G, et al. (2015) The genome of the tardigrade Hypsibius dujardini. *bioRxiv*:033464.

5. Eren AM, et al. (2015) Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3:e1319.

6. Boothby TC, et al. (2015) Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. *Proc Natl Acad Sci* 112(52):15976–15981.

7. Arakawa K (2016) No evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. *Proc Natl Acad Sci*.

8. Koutsovoulos G, et al. (2016) No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*. *Proc Natl Acad Sci*:201600338.