Methods

Postfiltering of our 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. This subtraction resulted in a postfiltered assembly of 212.3Mb analyzed in (1).

Genome assemblies, sequencing datasets, and annotations
Bemm *et al.* assemblies and annotation files were obtained from:
https://github.com/greatfireball/hypsibius_genome_revised

Datasets, assemblies, and annotations from the Blaxter group's project were obtained from:
http://bang.bio.ed.ac.uk/hypsibius_dujardini/

Datasets provided by Dr. Arakawa were obtained from:
http://www.g-language.org/data/gaou/pnas/

RNAseq datasets generated by Dr. Itai Yanai were obtained from NCBI's SRA database:
http://www.ncbi.nlm.nih.gov/sra?linkname=bioproject_sra_all&from_uid=272543

HGT index analysis
Gene IDs for metazoa, bacteria, plants, fungi, and archaea were downloaded from NCBI's protein database. Database aliases were generated using NCBI's blastdb_aliastool. Predicted protein sequences for each assembly were used as queries 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 was used to perform HGT index calculations. Only hits with an Evalue ≤1e-5 were retained. As in Boschetti *et al*. (2), 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 bitscore (max non-metazoan bitscore – max metazoan bitscore). Genes with an HGT index score ≥30 were considered foreign. Genes failing the HGT index test were considered tardigrade (non-foreign) genes.

Identification of genes with prokaryotic but not eukaryotic matches
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. For *D. melanogaster* HGT index analysis was performed as above using a metazoan database excluding *D. melanogaster* sequences.

Identification of Class C foreign genes
Class C genes were identified according to the methods detailed in (3). HGT index scores were calculated as detailed above and in (2). Predicted protein sequences for each assembly were

used as queries 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 was used to perform HGT index calculations. Only alignments with an Evalue ≤1e-5 were retained. As in (2), 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 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 (4). Unless otherwise noted all programs were used with default settings. Reads were mapped against assemblies using Bowtie2 (5) with default settings to generate Sam files. Sam files were converted to Bam format using Samtools (6). Bam files were indexed using Anvi'o's anvi-init-bam tool with 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 (–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 (7) 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 (8) and BLAT output were used to assess EST representation in various genome assemblies.

Assembly size and N50 assessments

Assembly size and N50 calculations were calculated using Joseph Fass's count_fasta.pl script (https://github.com/guyleonard/random_scripts/blob/master/count_fasta.pl).

Identification of Chitinophagaceae containing scaffolds

We obtained the same Chitinophagaceae protein dataset used in Bemm *et al.,* (30,844 Chitinophagaceae proteins downloaded from UniProtKB) and together with protein predictions form our raw assembly performed reciprocal best BLAST hit analysis with an Evalue cutoff of $\leq 1*10^{-10}$. The 10 scaffolds from our unfiltered assembly containing the highest number of Chitinophagaceae genes are shown in Fig. 10.

Assessment of saturation in short-read datasets

Saturation for each short-read dataset from our original paper (1) was assessed using bbmap's bbcountunique.sh script (Version 35.82 - http://sourceforge. net/projects/bbmap) with default settings. Raw saturation data was input into Prism (V 6.0g) and fit for visualization using nonlinear regression.

SNP identification

Genomic read datasets were mapped against our postfiltered assembly using Bowtie2 (5) with default parameters. Sam output from Bowtie2 mapping was converted to Bam format using Samtools (6) and Bam files were indexed using the anvi-init-bam function of Anvi'o (4). IGV v2.3 (9) was used to visualize read mapping to our postfiltered assembly.

1. 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.

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

3. 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.

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

5. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.

6. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

7. Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23(9):1061–1067.

8. Ryan JF (2013) Baa. pl: A tool to evaluate de novo genome assemblies with RNA transcripts. *ArXiv Prepr ArXiv13092087*. Available at: http://arxiv.org/abs/1309.2087 [Accessed April 2, 2016].

9. Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14(2):178–192.