Correction

GENETICS

Correction for "Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade," by Thomas C. Boothby, Jennifer R. Tenlen, Frank W. Smith, Jeremy R. Wang, Kiera A. Patanella, Erin Osborne Nishimura, Sophia C. Tintori, Qing Li, Corbin D. Jones, Mark Yandell, David N. Messina, Jarret Glasscock, and Bob Goldstein, which appeared in issue 52, December 29, 2015, of *Proc Natl Acad Sci USA* (112:15976–15981; first published November 23, 2015; 10.1073/pnas.1510461112).

The authors wish to note the following: "The data deposition footnote included in our published manuscript provides a link to an outdated version of our genome assembly. The version of the assembly that was used in our paper has been uploaded to GitHub and is available at the following link: https://github.com/Hd-tg-genome/PNAS_response/tree/master/Postfiltered_assembly."

www.pnas.org/cgi/doi/10.1073/pnas.1613046113

Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade

Thomas C. Boothby[a,1], Jennifer R. Tenlen[a,2], Frank W. Smith[a], Jeremy R. Wang[a,b], Kiera A. Patanella[a], Erin Osborne Nishimura[a], Sophia C. Tintori[a], Qing Li[c], Corbin D. Jones[a], Mark Yandell[c], David N. Messina[d], Jarret Glasscock[d], and Bob Goldstein[a]

[a]Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; [b]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; [c]Eccles Institute of Human Genetics, University of Utah, Salt Lake City, UT 84112; and [d]Cofactor Genomics, St. Louis, MO 63110

Edited by W. Ford Doolittle, Dalhousie University, Halifax, Canada, and approved September 28, 2015 (received for review May 28, 2015)

Horizontal gene transfer (HGT), or the transfer of genes between species, has been recognized recently as more pervasive than previously suspected. Here, we report evidence for an unprecedented degree of HGT into an animal genome, based on a draft genome of a tardigrade, *Hypsibius dujardini*. Tardigrades are microscopic eight-legged animals that are famous for their ability to survive extreme conditions. Genome sequencing, direct confirmation of physical linkage, and phylogenetic analysis revealed that a large fraction of the *H. dujardini* genome is derived from diverse bacteria as well as plants, fungi, and Archaea. We estimate that approximately one-sixth of tardigrade genes entered by HGT, nearly double the fraction found in the most extreme cases of HGT into animals known to date. Foreign genes have supplemented, expanded, and even replaced some metazoan gene families within the tardigrade genome. Our results demonstrate that an unexpectedly large fraction of an animal genome can be derived from foreign sources. We speculate that animals that can survive extremes may be particularly prone to acquiring foreign genes.

horizontal gene transfer | lateral gene transfer | genome | tardigrade | stress tolerance

Tardigrades (also known as water bears; Fig. 1*A*) are animals that are known for anhydrobiosis (surviving life without water). Tardigrades survive several additional stresses normally thought to be incompatible with life (1, 2), including extreme temperatures (–272 to 151 °C) (3, 4), radiation intensities orders of magnitude greater than humans can withstand (5, 6), incubation in organic solvents (7), and extremes of pressure (8). Tardigrades are the only animal known to survive exposure to the vacuum of space (8).

Water bears comprise their own phylum, the phylum Tardigrada, which, along with the phyla Arthropoda and Nematoda, is contained within the group Ecdysozoa. The phylogenetic position of tardigrades near two of the best-studied invertebrate model systems, *Caenorhabditis elegans* (a nematode) and *Drosophila melanogaster* (an arthropod), makes them attractive models for studying the evolution of molecular and developmental mechanisms (9).

Despite the interest in and value of tardigrades as a model system, there are minimal sequencing resources for this phylum of animals to date. To address this deficit, we have sequenced the genome of the tardigrade *Hypsibius dujardini*. We find that tardigrades have been taking up and incorporating foreign DNA into their genomes to a degree unprecedented in animals.

Results and Discussion

We extracted DNA from cultures of the tardigrade *H. dujardini* that were founded by a single parthenogenic animal, and sequenced the genome to high average coverage (126-fold) using a combination of Illumina Moleculo long reads and short insert mate pair libraries (Fig. S1*A*). A total of 26.8 Gb of raw reads resulted in an assembly 212.3 Mb in length containing 38,145 predicted genes (Fig. S1 *B* and *C*). The draft genome is comparable to existing eukaryotic reference genomes in terms of the percentage of guanine-cytosine content, number of exons per gene, exon size, and length of coding sequences (Fig. S1 *D*–*G*). The draft genome assembly has a contig N50 of 15.2 kb and a scaffold N50 only slightly greater (15.9 kb), despite the strong depth of coverage and the combination of paired-end reads and Moleculo long reads. Both Pacific Biosciences (PacBio) and Illumina scaffold analysis suggests a large degree of repetitive sequence is present in the *H. dujardini* genome and is a likely cause of limited scaffold assembly (*SI Text*). Despite the limited degree of scaffold assembly, the draft genome assembly appears nearly complete, because it contains 95.16% of core eukaryotic genes, a proportion that is similar to genomes of *C. elegans* and other model organisms that were used to build the core eukaryotic gene set (10, 11) (Fig. 1*B* and Fig. S1 *H* and *I*). In further support of the degree of completeness, our assembly contains 95.9% of all publicly available *H. dujardini* EST sequences. Of the remaining 216 missing EST sequences, only 48 are from metazoan sources, with the remaining having no hits, hits with no taxonomic information, or hits to nonmetazoan sequences, suggesting that the majority of missing ESTs are contaminants or sequencing artifact.

Preliminary BLAST analysis showed that an unexpectedly large proportion of the genes present in the *H. dujardini* genome had a top hit to sequences from nonmetazoan sources, suggesting that a large fraction of *H. dujardini* genes might have been acquired

Significance

Despite fascinating scientists for over 200 years, little at the molecular level is known about tardigrades, microscopic animals resistant to extreme stresses. We present the genome of a tardigrade. Approximately one-sixth of the genes in the tardigrade genome were found to have been acquired through horizontal transfer, a proportion nearly double the proportion of previous known cases of extreme horizontal gene transfer (HGT) in animals. Foreign genes have impacted the composition of the tardigrade genome: supplementing, expanding, and replacing endogenous gene families, including those families implicated in stress tolerance. Our results extend recent findings that HGT is more prevalent in animals than previously suspected, and they suggest that organisms that survive extreme stresses might be predisposed to acquiring foreign genes.

Author contributions: T.C.B. and B.G. designed research; T.C.B., J.R.T., K.A.P., E.O.N., S.C.T., D.N.M., and J.G. performed research; T.C.B., F.W.S., J.R.W., Q.L., C.D.J., and M.Y. analyzed data; J.R.T., E.O.N., and S.C.T. isolated material for sequencing; J.R.W. and C.D.J. performed Pacific Biosciences sequencing and assembly; Q.L. and M.Y. performed genome annotation; D.N.M. and J.G. performed Illumina sequencing and assembly; and T.C.B. and B.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: This Whole Genome Shotgun project has been deposited in the DDBJ/EMBL/GenBank database (accession no. LMYF00000000). The version described in this paper is version LMYF01000000.

[1]To whom correspondence should be addressed. Email: tboothby@gmail.com.

[2]Present address: Department of Biology, Seattle Pacific University, Seattle, WA 98119.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1510461112/-/DCSupplemental.

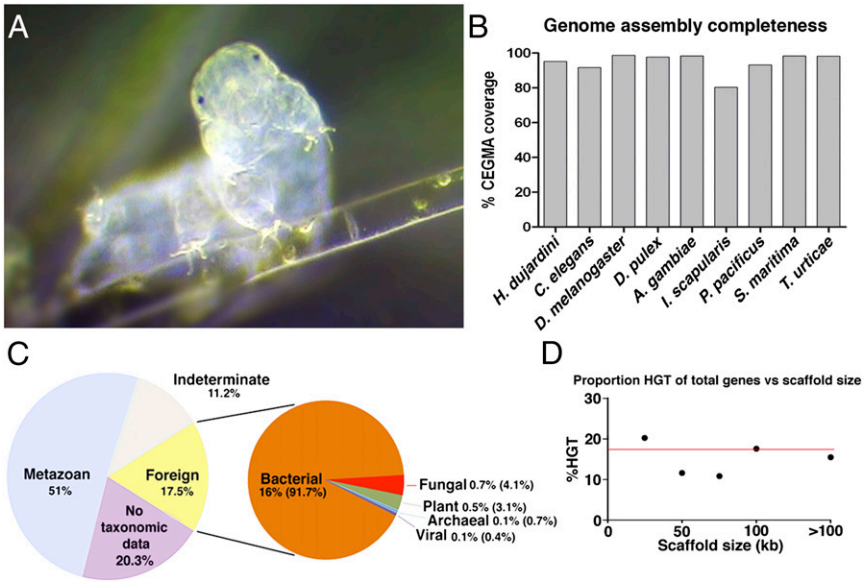

Fig. 1. Genome of the tardigrade *H. dujardini*. (*A*) Light micrograph of a tardigrade specimen. Image courtesy of S. Stammers, used with permission. (*B*) Percentage of coverage (complete + partial) for core eukaryotic genes in our *H. dujardini* genome, as well as genome assemblies from recently sequenced and model organisms. *A. gambiae*, *Anopheles gambiae*; CEGMA, core eukaryotic genes mapping approach; *D. pulex*, *Daphnia pulex*; *I. scapularis*, *Ixodes scapularis*; *P. pacificus*, *Pristionchus pacificus*; *S. maritima*, *Strigamia* maritima; *T. urticae*, *Tetranychus urticae*. (*C*) Source of genes in the *H. dujardini* genome as determined by HGT index calculations following Galaxy tools taxonomy extraction. (*D*) Proportion of horizontally transferred genes vs. total number of genes by scaffold size in the *H. dujardini* genome. The red line indicates the proportion of HGT genes in the total assembly (17.5%).

through horizontal gene transfer (HGT), since *H. dujardini* split from other animals with sequenced genomes. However, BLAST analysis alone can overestimate the degree of HGT (12, 13). Therefore, we took several steps to estimate the overall degree of HGT reliably, further testing whether genes that appear to derive from HGT were indeed horizontally transferred and not simply contaminants or results of convergent evolution.

The rotifer *Adineta ricciae* is the animal with the highest known degree of foreign origin to date, for which an estimated 9.6% of genes are of foreign origin (14). We used the same metric as was used previously for *A. ricciae* to calculate an HGT index value for each gene. The HGT index has been validated by phylogenetics, with maximum likelihood analysis as a useful metric for HGT (14). Using the HGT indexing method, we found that 6,663 *H. dujardini* genes (17.5% of all genes) had HGT index values greater than or equal to a previously established threshold, suggesting that they are likely to be derived from nonmetazoan sources (14) (Dataset S1). Increasing the stringency of the threshold continued to support a high degree of HGT (Fig. S2). The 6,663 genes showed closest sequence similarity to genes from bacteria, Archaea, fungi, plants, and viruses, with the majority (91.7%) most closely matching sequences from diverse bacteria (1,361 bacterial species, 40 archaeal species, 91 fungal species, 45 plant species, and 6 different viruses; Fig. 1*C* and Dataset S1). Our tardigrade cultures are fed algae, not bacteria, and although our algal cultures are not axenic, we would expect little to no bacterial contamination in our sequencing data. To ensure bacterial contamination is not a major issue in our assembly, we performed several independent analyses addressing this issue. First, we examined the coverage of all genes and scaffolds within our assembly. If foreign-looking sequences were contaminants, one would expect their abundance to be significantly less than the abundance of true tardigrade genes. Although there was variation in coverage found between genes and scaffolds, this variation was not biased but rather appears to be systematically present in both metazoan and foreign sequences. For example, the SD in coverage for a gene of metazoan origin was 138 compared with 108 for genes of foreign origin. Additionally, many genes of foreign origin were found to have higher sequencing coverage than genes of metazoan origin, and vice versa (Fig. S2 *C* and *D*). Second, all available bacterial rRNA sequences were downloaded from the Ribosomal Database Project (rdp.cme.msu.edu/) and used to perform reciprocal best-hit BLAST analysis against scaffolds in our genome assembly. Only four scaffolds in our assembly encoded a putative bacterial rRNA. These scaffolds were small, ranging in size from 4,331–14,016 bases and containing only nine (eight HGT and one non-HGT of unknown function) genes, suggesting that potential bacterial contamination is minimal and does not account for the disproportionately large number of horizontally acquired genes. Furthermore, general contamination in our assembly appears to be minimal. For example, we find no human contaminating sequence (Dataset S1). Together with codon optimization, intronization, and PCR analysis (discussed below), our analysis of gene coverage and rRNA contamination suggests that widespread contamination is not a major issue within our assembly.

No one bacterial species accounted for more than 3% of horizontally transferred genes. The proportion of genes with an HGT index exceeding the threshold was similar among scaffolds of various lengths (Fig. 1*D*, Fig. S2, and *SI Text*), suggesting that the genes predicted to be horizontally transferred were not concentrated among the least well-assembled sequences and that they are not all clustered within one long scaffold but are instead proportionally distributed throughout the assembly and the genome.

To analyze more directly whether 17.5% is a reliable estimate of the proportion of genes derived from foreign sources, we constructed and analyzed gene trees for a sample of 107 genes using both maximum likelihood and Bayesian analysis (Fig. 2*B* and Dataset S2). These sequences comprised 100 genes selected at random and seven hand-picked genes of interest, including predicted bacterial-, fungal-, archaeal-, plant-, and viral-derived genes. Together, the randomly selected genes represented a diverse set of gene characteristics: Percentage of identity with top nonmetazoan sources ranged from 88% to 25%, and intron counts (discussed further below) ranged from 20 to 0. A total of 103 of the 107 genes had sufficient sequence similarity to make gene trees. We used the same scoring system as was used for rotifers (14) and found that 101 of 103 of the resulting gene trees supported the HGT index-based conclusion by clustering the *H. dujardini* gene with a single nonmetazoan gene taxon (e.g., the *H. dujardini* gene fell within a clade of bacterial genes) or strongly rejecting monophyly with metazoan sequences, with the remaining two trees being ambiguous. None of our trees showed strong support for a predicted foreign gene, being instead monophyletic with metazoan sequences (Dataset S2). The tree topologies found, together with maximum likelihood and Bayesian analysis results and the lack of predicted U2-spliceosomal introns in many bacterial-like genes (42.9% of these genes, as discussed below), indicate that alternative explanations (sequence convergence or loss of genes in multiple independent metazoan clades) cannot easily explain the observed similarity to foreign sequences. These results confirm that the HGT index (14) is an effective tool for rapidly predicting the taxonomic origins of genes in large datasets, and they support

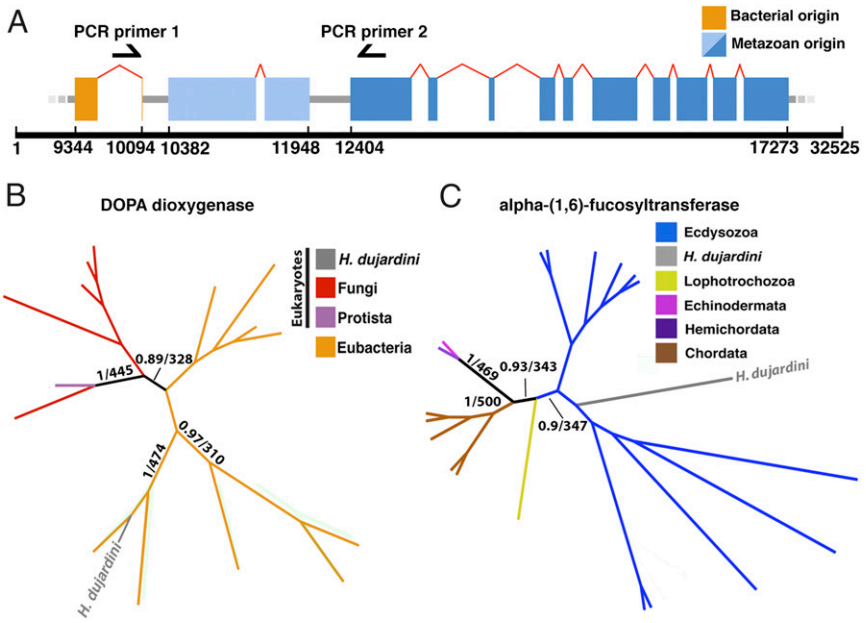

Fig. 2. HGT in the genome of *H. dujardini*. (*A*) Schematic representation of a portion of scaffold 962. Boxes represent exons, red arches represent U2 spliceosomal introns, and black arrows represent primer sites. (*B*) Gene 0.37 (orange, bases 9,344–10,094) encodes a putative bacterially derived DOPA dioxygenase with no identifiable animal homologs. (*C*) Gene 0.3 (dark blue, bases 12,404–17,273) encodes a tardigrade (metazoan) FUT8 gene. Supports on trees in *B* and *C* are given as Bayesian/bootstrap (500) supports. Trees for the remaining 102 genes and associated information are available in Dataset S2.

the conclusion that a large proportion of genes in the draft *H. dujardini* genome assembly are of nonmetazoan origin.

Despite similarities in coverage between metazoan and foreign sequences, assembly artifacts might result in contaminant sequences assembling next to tardigrade genes. Therefore, we looked for signs of contamination and tested directly the accuracy of assembly for a sample of genes by two methods. First, we performed PCR to test physical linkage of pairs of genes that were found on the same genomic scaffolds (Fig. 2*A*, Fig. S3, and Dataset S2). For PCR experiments, we used the same randomly selected set of genes as used for tree building above. For 104 of 107 genes, we recovered amplified products of the appropriate size, confirming that foreign genes are physically present within the *H. dujardini* genome (Fig. S3). Second, we used low-depth PacBio single molecule real-time (SMRT) sequencing for comparison with our initial assembly (*SI Text*). PacBio sequencing uses real-time imaging of synthesis of single DNA molecules to produce long sequencing reads. Because each long read comes from a single DNA molecule, the resulting contigs cannot be assembly artifacts; therefore, these data allowed us to test whether our initial assembly produced chimeric contigs. Our two assemblies were highly congruent. Synteny was preserved between contigs, and assemblies were 98.53% concordant per base (Fig. S4). Therefore, both PCR and resequencing provide support for the accuracy of our assembly.

Fifty-nine of the gene pairs we tested above contained a predicted nonmetazoan member near a metazoan member. Of these gene pairs, 58 (98.3%) gave PCR products of the correct size, suggesting that genes confirmed by our gene trees to be of foreign origin are physically linked to genes of metazoan origin (Fig. S3). For example, these tests included a metazoan alpha-(1,6)-fucosyltransferase (FUT8) gene for which our gene tree confirmed a close relationship with sequences from animals, and specifically ecdysozoans (Fig. 2*C*). This FUT8 gene is located on a scaffold near a gene that we determined by a gene tree using maximum likelihood and Bayesian analysis to be of bacterial origin, and that, to date, is found in no other animal (Fig. 2*B*). PCR performed with primers bridging these genes resulted in the correct product being amplified. Because nearly all genes in our sample could be confirmed by gene trees to be of foreign origin and by PCR to be accurately assembled, we consider approximately one-sixth to be a conservative estimate of the proportion of genes horizontally transferred into the *H. dujardini* genome since the origin of the tardigrades.

To begin to assess what happens to foreign genes after they have been horizontally transferred, we looked for diagnostic features of metazoan genes within the foreign genes found in the *H. dujardini* genome. Codon use of bacterially derived genes appears to have evolved metazoan bias, with codon use in horizontally acquired genes in *Hypsibius* being more similar to metazoan *Hypsibius* homologs than to their closest bacterial homolog (Fig. 3*A*). Additionally, average codon use in horizontally transferred genes of bacterial origin was more similar to codon use in genes of metazoan origin in the *H. dujardini* genome than it was to codon use in genes from highly represented bacterial species (Fig. 3*A* and Dataset S3). In addition, 57.1% of genes of bacterial origin contained at least one predicted U2-spliceosomal intron, a feature not found in bacterial genes (15) (Fig. 3*B* and Dataset S3). Consistent with most other eukaryotes, nearly all (99.0%) genes of metazoan origin in the *H. dujardini* genome contained predicted U2 spliceosomal introns, with the acceptor and donor consensus sequences being essentially identical between introns in genes of bacterial vs. metazoan origin (Fig. 3*B*). Taken together, our codon use and intron results suggest that horizontally transferred genes evolved properties characteristic of animal genes, and more specifically of the *H. dujardini* genome, further supporting that they are not artifacts generated by sequence from contaminating sources. Our results suggest that foreign genes have been assimilated into the genome of *H. dujardini* to a degree that is unprecedented to date in animals.

We next analyzed the degree to which the genes of foreign origin have expanded, supplemented, or replaced metazoan gene families. Sixty-seven percent of the genes in the assembly had known Pfam annotations, a proportion similar to the proportion of well-annotated reference genomes of model organisms (*Homo sapiens*, *Mus musculus*, *D. melanogaster*, *C. elegans*, *Arabidopsis thaliana*, and *Saccharomyces cerevisiae*) (16). A total of 99.2% of all protein-coding genes were represented only once in the assembly, with the average number of scaffolds that a single gene was found on being 1.008. This lack of redundancy suggests that any large apparent increases in numbers of protein domains and families within our assembly are due primarily to true expansions and not to assembly redundancy.

To identify these expansions, we compared protein families and domain counts from the *H. dujardini* genome with protein families and domain counts of the most closely related species with well-annotated genomes [*D. melanogaster* (*Dm*) and *C. elegans* (*Ce*)]. We identified numerous and diverse expansions: The *H. dujardini*

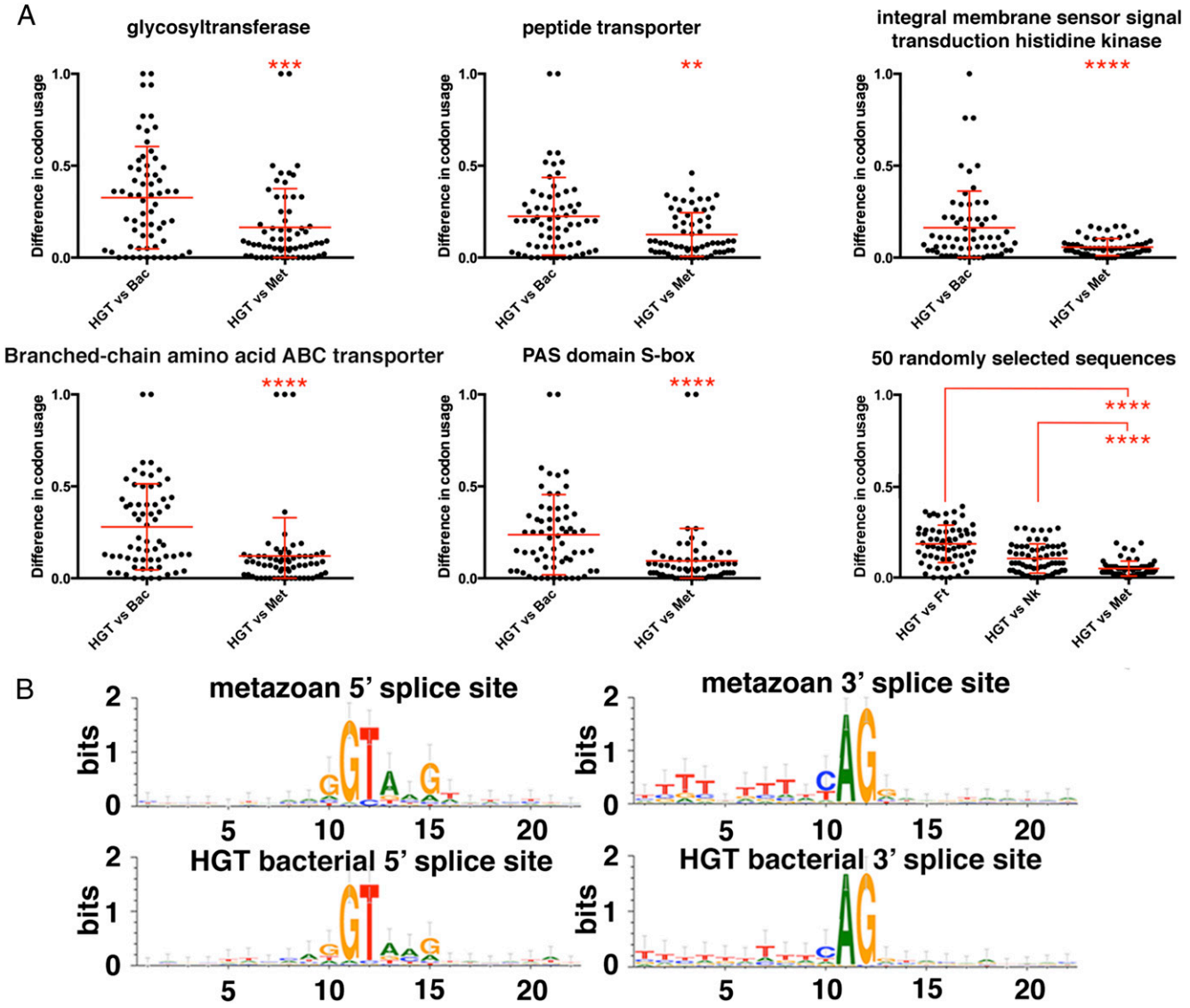

Fig. 3. Horizontally transferred genes have acquired characteristics of the metazoan genes. (*A*) For a particular horizontally transferred gene from bacteria, its closest metazoan homolog (Met) in the *H. dujardini* genome and its closest bacterial homolog (Bac) were found, and codon use statistics for each codon were calculated and compared. (*Lower Right*) Additionally, for each of the 64 codons, the difference in codon use between genes of foreign origin in the *H. dujardini* genome and metazoan genes from the *H. dujardini* genome, the bacterium *Niastella koreensis*, and the bacterium *Fluviicola taffensis* (bacterial species with the highest representation in the *H. dujardini* genome) was calculated. Horizontal lines represent the average difference between the codon use in *Hypsibius* genes of foreign origin and each other corresponding dataset. Unpaired *t* test: $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$. (*B*) Sequence logos generated for U2 intron 5′ and 3′ splice sites for genes of metazoan (*Top*) and bacterial (*Bottom*) origin.

genome has increased numbers of proteins containing domains involved in classic stress response, including heat shock (3.5-fold vs. *Dm* and sevenfold vs. *Ce*) and other chaperone proteins (fourfold vs. *Dm* and eightfold vs. *Ce*), DNA damage repair enzymes (3.6-fold vs. *Dm* and 7.3-fold vs. *Ce*), and antioxidant pathway members (2.7-fold vs. *Dm* and 4.7-fold vs. *Ce*), as well as enzymes involved in carbohydrate (fivefold vs. *Dm* and 2.8-fold vs. *Ce*) and lipid metabolism (1.9-fold vs. *Dm* and 1.5-fold vs. *Ce*) (Dataset S4).

Like typical gene family expansions, many gene family expansions in the *H. dujardini* genome have arisen through apparent gene duplication events. Other expansions are the result of acquiring homologs by HGT (Fig. 4*A* and Dataset S4). Some of the most expanded families include those families involved in sensing, responding, and adapting to environmental conditions and stress (Dataset S4). Surprisingly, 13.2% of unique protein domains are represented exclusively by genes of foreign origin, and 47.6% of *H. dujardini* protein families have been supplemented by at least one foreign member (Dataset S4).

The two most abundant domains that are contributed exclusively by genes acquired through HGT are catalase immune-responsive domain (IPR010582) and catalase core domain (IPR011614) (Fig. 4*B* and Dataset S4). Catalases are antioxidant enzymes involved in neutralizing oxidative stress, a major source of damage in cells under various stress conditions (17). Organisms closely related to tardigrades (e.g., *C. elegans*, *D. melanogaster*) encode metazoan versions of catalase genes, but *H. dujardini* appears to have replaced these genes with catalases acquired from foreign sources (Fig. 4*B*).

Tardigrades accumulate extensive DNA double-strand breaks with prolonged desiccation and when exposed to high levels of radiation (6, 18, 19). A number of protein families involved in DNA repair have been expanded in tardigrades (Dataset S4), and this expansion is due to the acquisition of foreign DNA in many cases. For example, our assembly contains eight genes encoding

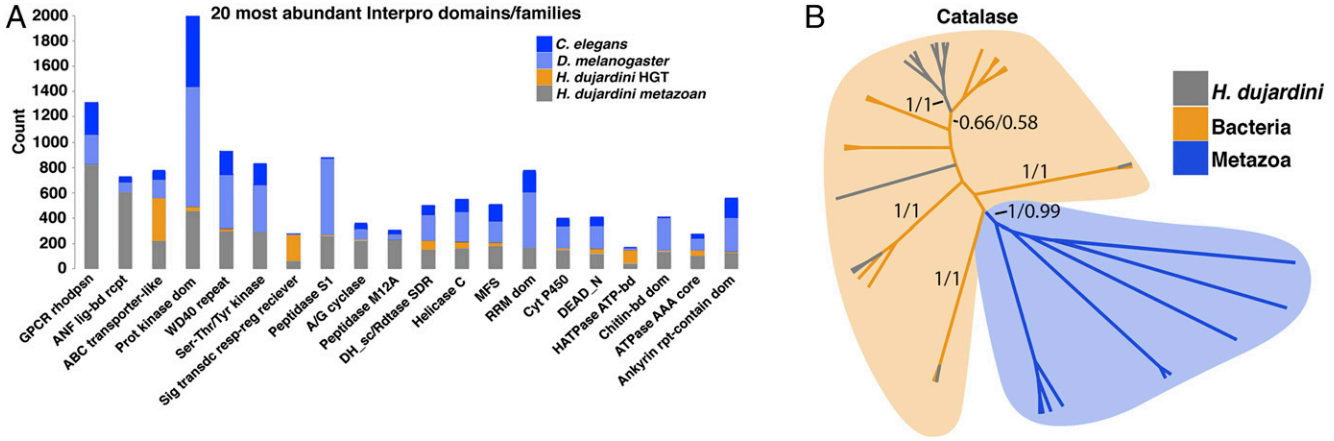

Fig. 4. Protein families in the *H. dujardini* genome have been expanded, supplemented, and replaced by genes acquired through horizontal transfer. (*A*) Raw counts for the 20 most abundant InterPro domains and families represented by *H. dujardini* genes of metazoan origin (blue), along with comparative data from two closely related species, *D. melanogaster* (dark gray) and *C. elegans* (light gray). The count for each InterPro domain contributed by genes of foreign origin is also shown (orange). ABC, ATP-binding cassette transporter-like; ANF lig-bd rcpt, ligand binding receptor region; Ankyrin rpt-contain dom, Ankyrin repeat-containing domain; Chitin-bd dom, chitin binding domain; Cyt P450, Cytochrome P450; DEAD_N, DEAD-box N-terminal domain; DH_sc/Rdtase SDR, short-chain dehydrogenases/reductases family; GPCR rhodpsn, rhodopsin-like G protein-coupled receptor domain; HATPase ATP-bd, histidine kinase-like ATPase C-terminal domain; MFS, major facilitator superfamily; Prot kinase dom, Protein kinase domain; RRM dom, RNA recognition motif; Sig transdc resp-reg receiver, signal transduction response regulator receiver domain. (*B*) Cladogram showing evolutionary relationships between foreign *H. dujardini* (gray), bacterial (orange), and metazoan (blue) genes implicated in stress tolerance. Numbers on branches indicate Bayesian followed by bootstrap (500) supports.

proteins with a Ku beta-barrel domain (IPR006164). Ku proteins are involved in double-strand break repair, and of the eight putative Ku genes found in our assembly, two are of metazoan origin and six are of foreign origin (Fig. S5 and Dataset S4). Similarly, our assembly contains nine metazoan UV mutant C (umuC) encoding genes and 14 foreign umuC encoding genes. UmuC is a DNA repair protein involved in translesion synthesis, which allows for replication to continue in the presence of broken DNA that would typically stall the replication machinery (20) (Fig. S5). Interestingly, our assembly also includes a gene of foreign origin encoding an adaptive response (Ada) DNA repair protein. Ada proteins are common bacterial enzymes that help cope with stress induced by alkylating agents (Fig. S5). Additionally, 6 of 13 DNA recombination and repair A (recA; Rad51 in eukaryotes) genes in our assembly are predicted to come from bacterial sources (Fig. S5).

Genes encoding enzymes involved in the biosynthesis of polyamines (small polycationic molecules known to be up-regulated during stress in plants that, among other roles, help protect membranes) have also been supplemented through acquisitions of foreign DNA. The polyamine spermidine is produced by the enzyme spermidine synthase. There are 15 spermidine synthase-encoding genes in the *H. dujardini* genome: 10 metazoan and five horizontally acquired (Fig. S5).

Heat shock proteins are molecular chaperones implicated in a number of stress responses (21). Heat shock protein 70 (hsp70) has been implicated in a number of stress responses specifically in tardigrades, including desiccation, radiation exposure, and heat stress (22). Of the 66 genes with hsp70 domains, 9 are from foreign sources (Fig. S5).

The various stresses tardigrades are known to tolerate cause the accumulation of reactive oxygen species, damaged DNA, denaturation of proteins, and disruption of membrane integrity. In sum, our data suggest that foreign genes, whose classic role(s) are the tolerance of these stresses, have been transferred into the genome of *H. dujardini* from bacteria, expanding or, in some cases, replacing endogenous gene families.

Our results demonstrate that an animal genome can be composed of a much greater proportion of horizontally transferred genes than was expected. Genes of foreign origin within the tardigrade genome are physically linked to animal genes (Fig. 2*A* and Fig. S3), have acquired introns (Fig. 3*B*), and have undergone codon optimization closely mirroring the codon use of metazoan genes in the tardigrade genome (Fig. 3*A*). Our evidence for the distribution of foreign genes throughout the genome assembly (Fig. 1*D*) and the diversity of apparent sources (Dataset S1) lead us to speculate that these horizontal transfers did not take place en masse but rather accumulated over evolutionary time.

Before this study, the animal with the highest known degree of HGT was *A. ricciae*, a bdelloid rotifer in which 8–9% of genes appear to have been acquired from foreign sources (14). Tardigrades and rotifers also comprise two of the only five animal clades known to survive desiccation, along with some arthropods, nematodes, and the eggs of certain flatworms (23). Comparison of all 39,532 predicted *H. dujardini* protein sequences with all 28,937 available *A. ricciae* rotifer mRNA sequences revealed 4,197 putative orthologs, reciprocal best BLAST hits with an excepted value (e-value) less than or equal to 1E-10. For just 3.5% (148) of these gene pairs, both genes appear to have been horizontally transferred into *H. dujardini* and *A. ricciae* (based on HGT index; Dataset S5). This overlap comprised genes encoding putative antioxidant enzymes, including a superoxide dismutase and GST as well as a putative Ku DNA damage repair protein and a putative hsp70 in each organism (Dataset S5). Such stress-related genes shared between *H. dujardini* and *A. ricciae* are likely to have been acquired through separate HGT events, because the majority (139 of 148) of genes were more closely related to genes from distinct nonmetazoan species than to each other (Dataset S5). For example, a gene predicted to encode a glutathione synthase in tardigrades appears to be derived from bacteria, whereas the glutathione synthase gene in rotifers is derived from fungi. It appears that rotifers and tardigrades have independently acquired and retained at least some similar genes that may contribute to stress tolerance.

It has been proposed that bdelloid rotifers are ancient asexual and ameiotic organisms, and that HGT serves as a source of genetic diversity in the absence of sex (24, 25). Although our laboratory cultures of *H. dujardini* are, to our knowledge, completely parthenogenic, consisting exclusively of females, males of the species and the genus *Hypsibius* have been reported (26) and meiosis occurs in *H. dujardini* (27), suggesting that a high degree of HGT may be a common feature of animals that survive desiccation rather than animals that lack sex and meiosis.

It has recently been proposed that in typical eukaryotes, horizontal transfer of genes is mediated, in large part, through endosymbiosis (28). Desiccation-tolerant organisms might be

unusually susceptible to taking up and incorporating foreign DNA into their genomes from their environment rather than exclusively from endosymbionts (Fig. S6): When desiccated membranes are rehydrated, they become transiently leaky, making the uptake of large macromolecules possible (29), which has been exploited previously to introduce large nucleic acids and drugs into the cytoplasm of rehydrating anhydrobiotic cells (30–32). In addition, when tardigrades, rotifers, and other anhydrobionts desiccate, genomic double-stranded breakages and other damage are induced (19, 33, 34), and there appear to be robust mechanisms for repairing this damage (19, 34). We speculate that desiccation and associated membrane leakiness and DNA breakages might predispose these animals to take up and incorporate foreign material in their genomes (Fig. S6).

Experimental Procedures

Genome Sequencing, Assembly, and Annotation. Short insert mate pair and Moleculo long-read libraries were generated from DNA extracted from *H. dujardini*. Assembly was performed with the Celera assembler, version 8.1 (35), largely following the method of McCoy et al. (36). Annotations for the *H. dujardini* genome assembly were generated using the automated genome annotation pipeline MAKER (16, 37, 38). Details can be found in *SI Experimental Procedures*.

HGT Index Calculation and Validation. HGT index scores were calculated as described by Boschetti et al. (14). Maximum likelihood and Bayesian supports were calculated for gene trees to complement HGT index score predictions. Details can be found in *SI Experimental Procedures*.

ACKNOWLEDGMENTS. We thank M. Chalfie, W. Marshall, D. Weisblat, S. Ahmed, G. Wray, M. Srivastava, and members of the B.G. laboratory for comments. This work was funded by National Science Foundation Grant IOS 1257320 (to B.G.), National Aeronautics and Space Administration Grant 14-14SF_Step2-0019 (to T.C.B.), National Institute of General Medical Science Grant K12GM000678 (to J.R.T.), and North Carolina Biotechnology Center Grant 2013-MRG-1110 (to C.D.J.), and by contributions from the University of North Carolina at Chapel Hill's Integrative Program for Biological and Genome Sciences, Carolina Center for Genome Sciences, and Office of the Vice Chancellor for Research.

1. Kinchin IM (1994) *The Biology of Tardigrades* (Portland Press, London).
2. Wright JC (2001) Cryptobiosis 300 years on from van Leuwenhoek: What have we learned about tardigrades? *Zool Anz* 240(3-4):563–582.
3. Becquerel P (1950) La suspension de la vie au dessous de 1/20 K absolu par demagnetization adiabatique de l'alun de fer dans le vide les plus elève. *CR Hebd Seances Acad Sci* 231:261–263. French.
4. Rahm PG (1921) Biologische und physiologische Beiträge zur Kenntnis der Moosfauna. *Zeitschrift Allg Physiologie* 20:1–35. German.
5. Jönsson KI, Harms-Ringdahl M, Torudd J (2005) Radiation tolerance in the eutardigrade Richtersius coronifer. *Int J Radiat Biol* 81(9):649–656.
6. Horikawa DD, et al. (2013) Analysis of DNA repair and protection in the Tardigrade Ramazzottius varieornatus and Hypsibius dujardini after exposure to UVC radiation. *PLoS One* 8(6):e64793.
7. Ramløv H, Westh P (2001) Cryptobiosis in the Eutardigrade Adorybiotus (Richtersius) coronifer: Tolerance to alcohols, temperature and de novo protein synthesis. *Zool Anz* 240(3-4):517–523.
8. Jönsson KI, Rabbow E, Schill RO, Harms-Ringdahl M, Rettberg P (2008) Tardigrades survive exposure to space in low Earth orbit. *Curr Biol* 18(17):R729–R731.
9. Gabriel WN, et al. (2007) The tardigrade Hypsibius dujardini, a new model for studying the evolution of development. *Dev Biol* 312(2):545–559.
10. Parra G, Bradnam K, Korf I (2007) CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23(9):1061–1067.
11. Wang C, Grohme MA, Mali B, Schill RO, Frohme M (2014) Towards decrypting cryptobiosis–analyzing anhydrobiosis in the tardigrade Milnesium tardigradum using transcriptome sequencing. *PLoS One* 9(3):e92663.
12. Salzberg SL, White O, Peterson J, Eisen JA (2001) Microbial genes in the human genome: Lateral transfer or gene loss? *Science* 292(5523):1903–1906.
13. Stanhope MJ, et al. (2001) Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. *Nature* 411(6840):940–944.
14. Boschetti C, et al. (2012) Biochemical diversification through foreign gene expression in bdelloid rotifers. *PLoS Genet* 8(11):e1003035.
15. Roy SW, Gilbert W (2006) The evolution of spliceosomal introns: Patterns, puzzles and progress. *Nat Rev Genet* 7(3):211–221.
16. Holt C, Yandell M (2011) MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12(1):491.
17. França MB, Panek AD, Eleutherio ECA (2007) Oxidative stress and its effects during dehydration. *Comp Biochem Physiol A Mol Integr Physiol* 146(4):621–631.
18. Rebecchi L, Cesari M, Altiero T, Frigieri A, Guidetti R (2009) Survival and DNA degradation in anhydrobiotic tardigrades. *J Exp Biol* 212(Pt 24):4033–4039.
19. Neumann S, Reuner A, Brümmer F, Schill RO (2009) DNA damage in storage cells of anhydrobiotic tardigrades. *Comp Biochem Physiol A Mol Integr Physiol* 153(4):425–429.
20. Goodman MF, Woodgate R (2013) Translesion DNA polymerases. *Cold Spring Harb Perspect Biol* 5(10):a010363.
21. Parsell DA, Lindquist S (1993) The function of heat-shock proteins in stress tolerance: Degradation and reactivation of damaged proteins. *Annu Rev Genet* 27(1):437–496.
22. Jönsson KI, Schill RO (2007) Induction of Hsp70 by desiccation, ionising radiation and heat-shock in the eutardigrade Richtersius coronifer. *Comp Biochem Physiol B Biochem Mol Biol* 146(4):456–460.
23. Watanabe M (2006) Anhydrobiosis in invertebrates. *Appl Entomol Zool (Jpn)* 41(1):15–31.
24. Flot JF, et al. (2013) Genomic evidence for ameiotic evolution in the bdelloid rotifer Adineta vaga. *Nature* 500(7463):453–457.
25. Gladyshev EA, Meselson M, Arkhipova IR (2008) Massive horizontal gene transfer in bdelloid rotifers. *Science* 320(5880):1210–1213.
26. Guidetti R, Colavita C, Altiero T, Bertolani R, Rebecchi L (2007) Energy allocation in two species of Eutardigrada. *J Limnol* 66(Suppl 1):111–118.
27. Ammerman D (1967) The cytology of parthenogenesis in the tardigrade Hypsibius dujardini. *Chromosoma* 23(2):203–213.
28. Ku C, et al. (2015) Endosymbiotic origin and differential loss of eukaryotic genes. *Nature* 524(7566):427–432.
29. Leopold AC, Musgrave ME, Williams KM (1981) Solute leakage resulting from leaf desiccation. *Plant Physiol* 68(6):1222–1225.
30. Boothby TC, Wolniak SM (2011) Masked mRNA is stored with aggregated nuclear speckles and its asymmetric redistribution requires a homolog of Mago nashi. *BMC Cell Biol* 12(1):45.
31. Boothby TC, Zipper RS, van der Weele CM, Wolniak SM (2013) Removal of retained introns regulates translation in the rapidly developing gametophyte of Marsilea vestita. *Dev Cell* 24(5):517–529.
32. Wolniak SM, Boothby TC, van der Weele CM (2015) Posttranscriptional control over rapid development and ciliogenesis in *Marsilea*. *Methods Cell Biol* 127:403–444.
33. Rebecchi L, et al. (2009) Tardigrade Resistance to Space Effects: First results of experiments on the LIFE-TARSE mission on FOTON-M3 (September 2007). *Astrobiology* 9(6):581–591.
34. Gladyshev E, Meselson M (2008) Extreme resistance of bdelloid rotifers to ionizing radiation. *Proc Natl Acad Sci USA* 105(13):5139–5144.
35. Myers EW, et al. (2000) A whole-genome assembly of Drosophila. *Science* 287(5461):2196–2204.
36. McCoy RC, Garud NR, Kelley JL, Boggs CL, Petrov DA (2014) Genomic inference accurately predicts the timing and severity of a recent bottleneck in a nonmodel insect population. *Mol Ecol* 23(1):136–150.
37. Cantarel BL, et al. (2008) MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* 18(1):188–196.
38. Yandell M, Ence D (2012) A beginner's guide to eukaryotic genome annotation. *Nat Rev Genet* 13(5):329–342.
39. Benson G (1999) Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res* 27(2):573–580.
40. Ryan JF (2013) Baa.pl: A tool to evaluate de novo genome assemblies with RNA transcripts. arXiv:1309.2087.
41. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2011) Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27(4):578–579.
42. Smit AFA, Hubley R (2010) RepeatModeler Open-1.0. Available at www.repeatmasker.org/. Accessed July 31, 2015.
43. O'Donovan C, et al. (2002) High-quality protein knowledge resource: SWISS-PROT and TrEMBL. *Brief Bioinform* 3(3):275–284.
44. UniProt Consortium (2010) The universal protein resource (UniProt) in 2010. *Nucleic Acids Res* 38(Database issue):D142–D148.
45. Pruitt KD, Tatusova T, Klimke W, Maglott DR (2009) NCBI Reference Sequences: Current status, policy and new initiatives. *Nucleic Acids Res* 37(Database issue):D32–D36.
46. Korf I (2004) Gene finding in novel genomes. *BMC Bioinformatics* 5(1):59.
47. Stanke M, Steinkamp R, Waack S, Morgenstern B (2004) AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Res* 32(Web Server issue):W309–W312.
48. Hunter S, et al. (2009) InterPro: The integrative protein signature database. *Nucleic Acids Res* 37(Database issue):D211–D215.
49. Eilbeck K, Moore B, Holt C, Yandell M (2009) Quantitative measures for the management and comparison of annotated genomes. *BMC Bioinformatics* 10(1):67.
50. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.
51. Giardine B, et al. (2005) Galaxy: A platform for interactive large-scale genome analysis. *Genome Res* 15(10):1451–1455.
52. Goecks J, Nekrutenko A, Taylor J; Galaxy Team (2010) Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 11(8):R86.
53. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.
54. Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56(4):564–577.
55. Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* 25(7):1307–1320.
56. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307–321.
57. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17(8):754–755.
58. Stothard P (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28(6):1102–1104.