Horizontal or lateral gene transfer (HGT) involves the transmission of genetic material between separate genomes. It is a major driver of evolutionary innovation in archaea and bacteria (1–4), but the role of HGT in eukaryotes, especially multicellular organisms, remains controversial. The genome sequence of Hypsibius dujardini, a member of the tardigrades, a group of animals that can survive a range of extreme environments, was reported to have undergone extensive HGT, amounting to 17.5% (5) of the genes identified. Even though Boothby et al. performed numerous checks to support their claim, reanalysis using new approaches to identify genome contamination combined with additional sequencing suggests only ~0.4% of the H. dujardini gene repertoire can confidently be identified as derived by HGT (6). These are fascinating animals and no doubt there will be further genome studies on these organisms before the last word on this topic emerges.

There are clear examples of HGT in animal lineages (e.g., refs. 7–10), yet HGT is thought to be infrequent in multicellular organisms (11–13). This is because multicellular organisms often separate their germ-line cells from their somatic cells, so the reproductive genome is in part isolated from contact with foreign sources of DNA. There are exceptions to this rule, such as when intracellular bacteria are present in the reproductive cells of an animal host (14, 15). Recent work has also demonstrated that bona fide HGT, separate from gene transfer arising from the progenitors of endosymbiotic organelles, appears to be rare in 55 genomes spanning the known diversity of the eukaryotes (16). Although more analysis is very much needed on a diversity of different eukaryotes, this finding is consistent with other work suggesting that eukaryotic groups that have shared environments and ecological interactions for over 400 million y (17) have undergone very little gene transfer (18).

Boothby et al. (5) published the first genome of a tardigrade animal, with the primary conclusion that ~one-sixth of the gene content is of HGT ancestry. Using a comparative similarity index approach (19), Boothby et al. (5) identified that 6,663 of the 39,532 predicted genes had sequence similarity profiles, suggesting they are not vertically inherited. Because of the scarcity of HGTs in animals, the extent of transfer identified in the tardigrade genome was quite a surprise. Boothby et al. set out to test this observation, performing several checks to validate the transfers identified. First, they performed additional phylogenetic analyses of a subsample of 107 largely randomly selected candidate transfers (a test HGT dataset), finding phylogenetic support for HGT in 101 cases. To exclude assembly artifacts, they then confirmed their genome assembly using long-read sequencing technologies and used PCR to confirm the genomic scaffold of 104 of the 107-test HGT dataset. Of these PCR amplifications, 59 targeted a section of a genome contig that bridged a gap between a putative HGT gene and a native (vertically derived) gene. Of these “bridging” PCRs, 58 seemed to validate the synteny recovered in their genome assembly, suggesting these candidate HGTs were integrated into the native tardigrade genome. Finally, 57% of all of the putatively transferred genes reported also contained a predicted spliceosomal intron. Spliceosomal introns are a hallmark of eukaryotic encoded genes, supporting Boothby et al.’s initial claim that the prokaryotic genes have been integrated into the tardigrade genome.

Shortly after this publication, a manuscript emerged reporting a second H. dujardini genome sequence (6). Flow cytometry analysis demonstrated a genome size of 110 Mb (6), over 100 Mb smaller than the previously published (5) genome assembly. Preliminary assembly of this second genome amounted to 186 Mb. Using GC content vs. read coverage plots (20), Koutsovoulos et al. (6) identified extensive traces of contamination in both genome assemblies. This tool was then used to clean the second genome assembly, resulting in a draft genome of 135 Mb and encoding 23,021 putative proteins. Koutsovoulos et al. then retested the provenance of the “test HGT dataset” selected by Boothby et al. (5), identifying 51 as informative for HGT analysis. Further analysis also demonstrated that the majority of the HGTs, inferred from the similarity index approach, were likely derived from contamination (6). Finally, analysis of the cleaned genome identified 94 strong HGT candidates, 0.4% of the total repertoire of genes identified: 49 derived from other eukaryotes and 55 from

**A tale of two tardigrades**

Thomas A. Richards\(^a\)\(^b\)\(^1\) and Adam Monier\(^c\)

---

\(^a\)Department of Biosciences, University of Exeter, Exeter EX4 4QD, United Kingdom; \(^b\)Integrated Microbial Biodiversity Program, Canadian Institute for Advanced Research, Toronto, ON, Canada M5G 1Z8; and \(^c\)Québec Océan, Université Laval, Québec, QC, Canada G1V 0A6

Author contributions: T.A.R. and A.M. wrote the paper.

The authors declare no conflict of interest.

See companion article on page 5053.

\(^1\)To whom correspondence should be addressed. Email: t.a.richards@exeter.ac.uk.
Although rotifers and the red alga Galdieria sulphuraria, which can both survive in “extreme” environments, may still represent exceptions to this rule (19, 21), this trend suggests that barriers to HGT integration do exist in eukaryotes and therefore transfer in eukaryotes is of a different form and frequency than that observed in prokaryotes (16). However, there are clear cases where HGTs have added important functions to the recipient lineage (e.g., refs. 7–10, 22, 23). Consequently, identification of transferred genes is likely to inform our understanding of how biological systems evolve and function.

With the use of next-generation technologies, genome sequencing is rapidly becoming the default approach of a range of research fields. As sequencing becomes accessible, it is clear that reliable genome assembly is not so straightforward. The conflation between the challenge of genome assembly and the desire to seek HGTs of importance to the evolution of biological function is generating friction. As this field progresses we should look to develop a set of guidelines for identifying HGT. Here we list a set of tentative guidelines for informing HGT analysis.

First, contamination is unavoidable in de novo genome sequencing. As such, de novo genome sequences should be viewed as metagenomes. We must therefore implement cutting edge bioinformatics based on database search similarity patterns/GC/word use/coverage comparisons to identify and remove contamination (6, 24).

Second, during the preparation of DNA samples for genome sequencing, DNA should be tested for contamination using environmental small subunit rDNA-based protocols (e.g., ref. 25), preferably targeting multiple gene-sequence markers (i.e., using both 18S and 16S small subunit PCR primers). This will provide preliminary data for assessing contamination in genome assemblies.

Third, where possible, multiple genome assemblies from different representatives of the same species and from related groups should be sequenced to validate HGTs and polarize the point of gene acquisition.

Fourth, all putative HGTs should be validated using phylogenetic analysis. Phylogenetic trees that demonstrate topological support (bootstrap and alternative topology analysis) for a putative recipient lineage branching with and within a donor group can be used to provide support for HGT (see figure 1 in ref. 26).

Fifth, synteny analysis should be used to confirm that candidate HGTs are nested on native sections of chromosomal DNA (i.e., adjacent to vertically derived genes).

Sixth, transcription of putative HGTs should be validated to confirm intron architecture and test if the gene is transcriptionally functional in the recipient genome.

These guidelines can be modified and added to as more datasets emerge and more cases of HGT are investigated. However, we predict that adherence to these rules will confirm that HGT into eukaryotes does occur but at a lower frequency compared with prokaryotes. Furthermore, we also predict that such criteria will help the field to focus on the HGTs that represent important gains-of-function in the recipient lineages. For us, identifying cases where transfer has amended or added function in the recipient lineage, however rare, represents one of the most exciting prospects of this field of research. Researchers are therefore correct to look for HGT in organisms that are capable of “extraordinary” biological functions, for example colonizing extreme or variant environments in contrast to related taxa (27). However, caution is needed in these endeavors.


Richards and Monier

PNAS | May 3, 2016 | vol. 113 | no. 18 | 4893