Massive Horizontal Gene Transfer in Bdelloid Rotifers

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td><a href="http://dx.doi.org/10.1126/science.1156407">http://dx.doi.org/10.1126/science.1156407</a></td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:3120157">http://nrs.harvard.edu/urn-3:HUL.InstRepos:3120157</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Massive Horizontal Gene Transfer in Bdelloid Rotifers

Eugene A. Gladyshev, Matthew Meselson, Irina R. Arkhipova

Horizontal gene transfer in metazoans has been documented in only a few species and is usually associated with endosymbiosis or parasitism. By contrast, in bdelloid rotifers we found many genes that appear to have originated in bacteria, fungi, and plants, concentrated in telomeric regions along with diverse mobile genetic elements. Bdelloid proximal gene-rich regions, however, appeared to lack foreign genes, thereby resembling those of model metazoan organisms. Some of the foreign genes were defective, whereas others were intact and transcribed; some of the latter contained functional spliceosomal introns. One such gene, apparently of bacterial origin, was overexpressed in Escherichia coli and yielded an active enzyme. The capture and functional assimilation of exogenous genes may represent an important force in bdelloid evolution.

HORIZONTAL gene transfer (HGT), the movement of genes from one organism to another by means other than direct descent (vertical inheritance), has been documented in prokaryotes and in phagocytic and parasitic unicellular eukaryotes. Despite the large number of sequenced genomes, documented HGT is rare in metazoans, at least in part because of the sequestration of the germ line. HGT may be facilitated by long-term association with organelles or with intracellular endosymbionts and parasites, or it may involve transposable elements (TEs).

Bdelloid rotifers are small freshwater invertebrates that apparently lack sexual reproduction and can withstand desiccation at any life stage. Their genomes contain diverse TEs, including DNA transposons and retrovirus-like env-containing retrotransposons, such as Juno and Vesta, possibly acquired from exogenous sources and concentrated near telomeres. We investigated TE distribution in bdelloids by sequencing clones from an Adineta vaga fosmid library hybridizing to Juno probes. Unexpectedly, in two Juno long terminal repeat (LTR)–containing clones (contigs Av240A and Av212A), we found 10 protein-coding sequences (CDS) yielding strong database hits (BLAST E-values of 8E-102 to 0.0) to bacterial and fungal genes. Half of these CDS have no metazoan orthologs, and three apparently bacterial CDS are interrupted by canonical spliceosomal introns, which are nonexistent in bacteria.
Fluorescent in situ hybridization (FISH) with a probe from Av240A confirmed that these CDS reside in the *A. vaga* genome (Fig. 1E). Their hybridization pattern resembles that of known telomeric fosmids (15), suggesting proximity to telomeres. The appearance of two hybridizing sites in some nuclei is consistent with the genome structure of bdelloid genomes in which chromosomes occur as linear pairs (16, 17). Indeed, we found colinear partners of both Av240A and Av212A (Av240B and Av212B) with overall pairwise divergence (4%) similar to that in other regions of bdelloid genomes.

The cluster of foreign genes near the *Juno1* LTR in Av240 includes two divergently oriented genes for enzymes involved in bacterial cell wall peptidoglycan biosynthesis—*Alr*, encoding alanine racemase, and *Ddl*, encoding d-Ala-d-Ala ligase—adjacent to a uridine diphosphate (UDP)—glycoseutransfer gene apparently of plant origin (Fig. 1A). Reverse transcription polymerase chain reaction (RT-PCR) shows that all three genes are transcribed and that their introns are properly spliced (Fig. 1B). 5' RACE (rapid amplification of cDNA ends) demonstrates that the UDP-glycoseutransfer mRNA is trans-spliced, as are numerous bdelloid genes (18), and that *Alr* and *Ddl* transcription initiates at opposite oriented promoters located between them (Fig. 1C). Furthermore, the purified *A. vaga* AvDdl protein overexpressed in *E. coli* catalyzes the synthesis of the d-Ala-d-Ala dipeptide from d-Ala in vitro (Fig. 1D).

In addition to ubiquitous bacterial genes, such as *Alr* and *Ddl*, we identified genes occurring in only a few species of bacteria and fungi. The *XynB*-like gene (figs. S1A and S2B) apparently represents a fusion between a two divergent conserved domains and is found in only 10 bacterial species. Next to the *Alr-Ddl* pseudo- operon, there is a *HemK*-like methyltransferase and a putative adenosine triphosphatase. These two genes are also rare: They are found in only four genera of proteobacteria and three genera of filamentous fungi. In the bacterium *Methylobacillus flagellatus* and in the fungus *Phaeosphaeria nodorum* they are adjacent and in the same relative orientation as in *A. vaga*, indicating that they might have been acquired from a single source.

We also found genes with similarity to those present in both metazoans and nanotomases. We characterized the foreignness of such genes with an alien index (AI), which measures how many orders of magnitude the BLAST E-value for the best metazoan hit differs from that for the best nonmetazoan hit (Table 1 and table S1) (19). We tested the AI validity by phylogenetic analyses of all CDS with AI > 0 yielding metazoan hits, excluding those with repetitive sequences. We found that AI < 45, corresponding to a difference of ≥20 orders of magnitude between the best nonmetazoan and metazoan hits, was a good indicator of foreign origin, as judged by phylogenetic assignment to bacterial, plant, or fungal clades (Fig. 2 and fig. S2A). Genes with 0 < AI < 45 were designated indeterminate, because their phylogenetic analysis may or may not confidently support a foreign origin. Four FabG-like genes for short-chain dehydrogenases/reductases (SDH), from two different SDH subfamilies, are most likely of bacterial origin (AI = 98/92/88/45; Fig. 2A). The *A. vaga* galacturonidase (AI = 212) appears to be of fungal origin (Fig. 2B), and the UDP-glucosyltransferase in Av240 (AI = 28, indeterminate) belongs to a plant clade (Fig. 2C). Two genes, *XynB* and *MviM*, had sufficient nucleotide sequence similarity (~70%) to bacterial homologs for phylogenetic reconstruction and determination of nonsynonymous and synonymous divergence (fig. S2B).

We extended our search for foreign genes to two pairs of contigs ending with telomeric repeats (telomeres K and L) (15) (fig. S1, B and C). In addition to various TEs, telomeric repeats, and *Athena* retroelements characteristic of bdelloid telomeric regions, we found additional examples of foreign genes, including a pseudogene of fungal origin (putative urea transporter; Table 1) with three frameshifts and two in-frame stop codons in one of the two colinear homologs of telomere L. Additionally, we identified foreign genes sandwiched between short stretches of telomeric repeats, suggesting their addition to deprotected chromosome ends (fig. S1C).

We also observed examples of possible loss of genes and TEs from telomeres (fig. S3A), such

---

**Table 1.** Representative bdelloid CDS of foreign origin homologous to genes with known function. For a complete list and additional information on each CDS, see table S1. Data are from BLASTP similarity searches, as described in (19). Asterisks indicate putative pseudogenes.

<table>
<thead>
<tr>
<th>Gene ID, name</th>
<th>Contig ID</th>
<th>Introns</th>
<th>Al %</th>
<th>Identity to best hit</th>
<th>Best hit, E-value</th>
<th>Best hit, metazoan</th>
<th>Best hit, taxonomy</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AV10027_XynB</td>
<td>Av212A</td>
<td>0</td>
<td>460</td>
<td>63</td>
<td>0.00E+00</td>
<td>No hits</td>
<td>Bacteria; Bacteroidetes</td>
<td>Xylosidase/arabinosidase</td>
</tr>
<tr>
<td>AV10001_NRPS</td>
<td>Av110A</td>
<td>10</td>
<td>460</td>
<td>32</td>
<td>0.00E+00</td>
<td>No hits</td>
<td>Bacteria; Proteobacteria; Cyanobacteria</td>
<td>Nonribosominal peptide synthetase</td>
</tr>
<tr>
<td>AV10134_PheA</td>
<td>161F07</td>
<td>0</td>
<td>400</td>
<td>61</td>
<td>1.00E-174</td>
<td>No hits</td>
<td>(Fungi; Bacteria)</td>
<td>Monooxygenase, FAD dependent</td>
</tr>
<tr>
<td>AV10002_TrkA</td>
<td>Av110A</td>
<td>0</td>
<td>379</td>
<td>54</td>
<td>1.00E-175</td>
<td>4.00E-11</td>
<td>Bacteria; Proteobacteria</td>
<td>Monooxygenase, NAD dependent</td>
</tr>
<tr>
<td>PR10002_MviM</td>
<td>182F10</td>
<td>0</td>
<td>327</td>
<td>67</td>
<td>1.00E-149</td>
<td>2.00E-07</td>
<td>Bacteria; (Achromobacter/ Chloroflexi)</td>
<td>Oxidoreductase</td>
</tr>
<tr>
<td>PR10010_DAP2</td>
<td>182F10</td>
<td>0</td>
<td>310</td>
<td>27</td>
<td>1.00E-140</td>
<td>1.00E-05</td>
<td>Bacteria; (Achromobacter/ Proteobacteria)</td>
<td>Prolyl oligopeptidase*</td>
</tr>
<tr>
<td>AV10104_Dur3</td>
<td>AvTelLB</td>
<td>1</td>
<td>243</td>
<td>44</td>
<td>1.00E-132</td>
<td>4.00E-27</td>
<td>Eukaryota; Fungi</td>
<td>Urea active transporter*</td>
</tr>
<tr>
<td>PR10012_Rama</td>
<td>182J17</td>
<td>0</td>
<td>246</td>
<td>31</td>
<td>1.00E-107</td>
<td>No hits</td>
<td>(Bacteria; Fungi)</td>
<td>α-L-Rhamnosidase</td>
</tr>
<tr>
<td>AV10121_NRPS</td>
<td>99G07</td>
<td>4</td>
<td>237</td>
<td>30</td>
<td>1.00E-103</td>
<td>No hits</td>
<td>Bacteria; Cyanobacteria</td>
<td>Nonribosominal peptide synthetase</td>
</tr>
<tr>
<td>AV10153_2gHa</td>
<td>210B3</td>
<td>0</td>
<td>212</td>
<td>50</td>
<td>1.00E-108</td>
<td>2.00E-16</td>
<td>Eukaryota; Fungi</td>
<td>Endo-xylolagacturonan hydrolase</td>
</tr>
<tr>
<td>AV10042_HemK</td>
<td>Av240B</td>
<td>1</td>
<td>199</td>
<td>56</td>
<td>2.00E-91</td>
<td>1.00E-04</td>
<td>Bacteria; Proteobacteria</td>
<td>HemK-like methyltransferase</td>
</tr>
<tr>
<td>AV10092_beta-Gal</td>
<td>AvTelLA</td>
<td>0</td>
<td>153</td>
<td>33</td>
<td>1.00E-105</td>
<td>4.00E-39</td>
<td>Eukaryota; Viridiplantae</td>
<td>β-D-Galactosidase</td>
</tr>
<tr>
<td>AV10044_Alr</td>
<td>Av240B</td>
<td>1</td>
<td>152</td>
<td>38</td>
<td>1.00E-67</td>
<td>No hits</td>
<td>Bacteria; Bacteroidetes</td>
<td>Alanine racemase</td>
</tr>
<tr>
<td>AV10025_AMH</td>
<td>Av212A</td>
<td>1</td>
<td>150</td>
<td>52</td>
<td>8.00E-77</td>
<td>2.00E-11</td>
<td>Eukaryota; Fungi</td>
<td>Amidoligohydrolase</td>
</tr>
<tr>
<td>AV10045_Ddl</td>
<td>Av240B</td>
<td>1</td>
<td>138</td>
<td>40</td>
<td>1.00E-60</td>
<td>No hits</td>
<td>Bacteria; Bacteroidetes</td>
<td>β-Alanine-o-alanine ligase</td>
</tr>
<tr>
<td>AV10140_PDLC</td>
<td>193E18</td>
<td>2</td>
<td>126</td>
<td>31</td>
<td>1.00E-55</td>
<td>No hits</td>
<td>Eukaryota; Fungi</td>
<td>Phospholipase-D active site motif protein*</td>
</tr>
<tr>
<td>AV10016_FabG</td>
<td>Av212A</td>
<td>0</td>
<td>98</td>
<td>58</td>
<td>1.00E-74</td>
<td>8.00E-32</td>
<td>Bacteria</td>
<td>Short-chain dehydrogenase/reductase</td>
</tr>
<tr>
<td>AV10109_FabG</td>
<td>AvTelLB</td>
<td>0</td>
<td>92</td>
<td>57</td>
<td>4.00E-73</td>
<td>5.00E-33</td>
<td>Bacteria</td>
<td>Short-chain dehydrogenase/reductase*</td>
</tr>
</tbody>
</table>
| AV10011_FabG | Av212A    | 0      | 88   | 54                   | 6.00E-67         | 2.00E-28             | Bacteria | Short-chain dehydrogenase/reductase*
| AV10071_HAL | AvTelKA   | 0      | 77   | 48                   | 2.00E-61         | 1.00E-27             | Bacteria | Histidine ammonia-lyase |
| AV10095_GCNS | AvTelLA   | 0      | 59   | 35                   | 2.00E-27         | No hits              | Bacteria; Proteobacteria | GCNS-related N-acetyltansferase* |
| AV10158_FabG | 210B3     | 2      | 46   | 41                   | 2.00E-39         | 2.00E-19             | Bacteria | Short-chain dehydrogenase/reductase |

---

Downloaded from www.sciencemag.org on June 2, 2008
as the metazoan long-chain acyl-coenzyme A synthetase (ACS) gene fragment on telomere K, which has an apparently intact 5' sequence but is 3'-truncated by telomeric repeat addition to exon 2 (fig. S3A). Single-telomere length analysis PCR (STELA-PCR) (20) verified its telomeric localization (fig. S3B). No collinear partner of ACS was found, and the lack of RT-PCR product suggests that it is transcriptionally inactive or that its transcript is unstable (fig. S3C).

Two other contigs containing telomeric repeats (Table 1 and fig. S1D) had apparently intact genes for two nonribosomal peptide synthetases (NRPSs), large enzymes responsible for synthesis in bacteria and fungi of biologically active peptides including antibiotics and toxins (21). This finding suggests that bdelloid biosynthetic activity includes the production of secondary metabolites.

We also examined sequences in the vicinity of retrovirus-like elements in Philodina roseola, a bdelloid that separated from A. vaga tens of millions of years ago (22). The P. roseola telomere P (15) contains a gag gene from a retrovirus-like element similar to A. vaga Juno1, named Juno2, which is 3'-truncated by P. roseola telomeric repeats (fig. S3A). We probed a P. roseola genomic library with this gag fragment and found that two of five Juno2 copies are surrounded by foreign genes (Table 1 and fig. S1E). Thus, extensive HGT probably represents a general feature of bdelloid rotifers.

![Fig. 2. Bayesian analysis of A. vaga CDS from three different kingdoms. Clades with different taxonomic affiliation are color-coded as in Fig. 1; CDS from A. vaga are in red. (A) Two subfamilies of FabG short-chain dehydrogenases/reductases, with four A. vaga FabG-like genes of bacterial origin (one contains an intron; three are intronless and could have entered independently or undergone duplication). (B) Galacturonidases and the corresponding A. vaga gene of fungal origin. (C) UDP-glycosyltransferases and the corresponding A. vaga gene of plant origin. Bayesian posterior probabilities are shown; for branches leading to A. vaga, neighbor-joining bootstrap support values are also indicated. For alignments, see table S11. Scale bars, 0.1 amino acid substitution per site.](image-url)
Fig. 3. Comparison of bdelloid telomeric/TE-rich regions (A) with bdelloid gene-rich regions (B) and with other model invertebrates (C). Pie charts were built with 921,903 base pairs (bp) (A) and 661,316 bp (B) of genomic DNA, respectively (excluding one of the two colinear partners). For TE count, ~1.3 Mb of genomic DNA in each data set was analyzed (including both colinear partners). The size of each sector corresponds to the percentage of the total length (bp) occupied by each category in a data set: genes (foreign; indeterminate, including repetitive proteins with AI > 45; metazoa; hypothetical ORFs) and TEs (DNA TEs; retrovirus-like LTR retrotransposons; telomere-associated Athena retroelements). The numbers of telomeric repeat stretches in data sets A and B are 67 and 1, respectively. (C) Histogram showing the distribution of the AI value in bins of 50 for CDS from the bdelloid telomeric/TE-rich and gene-rich data sets and from C. elegans, D. melanogaster, C. elegans telomeres, and B. pictilis EST (24).

Ancestral Monogamy Shows Kin Selection Is Key to the Evolution of Eusociality

William O. H. Hughes,2* Benjamin P. Oldroyd,2 Madeleine Beekman,2 Francis L. W. Ratnieks3

Close relatedness has long been considered crucial to the evolution of eusociality. However, it has recently been suggested that close relatedness may be a consequence, rather than a cause, of eusociality. We tested this idea with a comparative analysis of female mating frequencies in 267 species of eusocial bees, wasps, and ants. We found that mating with a single male, which maximizes relatedness, is ancestral for all eight independent eusocial lineages that we investigated. Mating with multiple males is always derived. Furthermore, we found that high polyandry (>2 effective mates) occurs only in lineages whose workers have lost reproductive altruism (i.e., helpers increase the number of their own offspring (1)). The established paradigm, based on inclusive fitness (kin selection) theory, is that eusociality evolves because of a combination of the direct benefits of altruism (i.e., helpers increase the number of individuals reared) and close relatedness between group members, such that the inclusive fitness of helpers exceeds that achievable through a solitary lifestyle (2–5). High relatedness, arising from the delayed dispersal of offspring [possibly

References and Notes
8. J. C. Dunning Hotopp et al., Science 317, 1753 (2007); published online 30 August 2007 (10.1126/ science.1142490).

19. See supporting material on Science Online.
31. We thank W. Reznikoff, M. Belfort, and D. Mark Welch for comments and J. Mark Welch, K. van Doninck, and J. Hur for communicating results before publication. Supported by NSF grant MCB-0614142 (M.M. and I.R.A.) and NIH grant GM072708 (M.M.). Sequences obtained in this study were deposited in GenBank (accession numbers EU643473 to EU643504).

Supporting Online Material
www.sciencemag.org/cgi/content/full/320/5880/1210/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 to S12
References
12 February 2008; accepted 22 April 2008
10.1126/science.1156407