Massive Horizontal Gene Transfer in Bdelloid Rotifers

Citation

Published Version
http://dx.doi.org/10.1126/science.1156407

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:3120157

Terms of Use
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 http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Massive Horizontal Gene Transfer in Bdelloid Rotifers

Eugene A. Gladyshev,1 Matthew Meselson,1,2* Irina R. Arkhipova1,2*

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 (1) and in phagocytic and parasitic unicellular eukaryotes (2–4). 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 (5, 6). HGT may be facilitated by long-term association with organelles or with intracellular endosymbionts and parasites (7, 8), or it may involve transposable elements (TEs) (9, 10).

Bdelloid rotifers are small freshwater invertebrates that apparently lack sexual reproduction and can withstand desiccation at any life stage (11, 12). 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 (13, 14). We investigated TE distribution in bdelloids by sequencing clones from an Adineta vaga fosmid library hybridizing to Juno1 probes. Unexpectedly, in two Juno1 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 (Fig. 1A, Table 1, fig. S1A, and table S1). Half of these CDS have no metazoan orthologs, and three apparently bacterial CDS are interrupted by canonical spliceosomal introns, which are nonexistent in bacteria.

Fig. 1. Structural and functional analysis of the Av240 genomic region. (A) The 85-kb Av240B contig. The colinear 50-kb Av240A contig contains the Juno1 LTR (red triangle) and extends from TPR to LRR_RI (not shown). CDS (boxes) are colored according to their putative origin: metazoan, gray; bacterial, blue; fungal, purple; plant, green; indeterminate, striped; hypothetical, white. Intron positions are indicated. Pseudogenes are denoted by ψ, and defects in their reading frames appear as vertical lines. Scale bar, 1 kb. (B) RT-PCR analysis of Alr, Ddl, and UDP-glycosyltransferase genes. Unspliced and spliced RNA are visible in +RT lanes; −RT, no reverse transcriptase; gDNA, genomic DNA control. (C) 5′ RACE analysis of transcription start sites (arrows) for genes in Fig. 1B. (D) Expression in E. coli of the A. vaga Ddl cloned in pET45b to yield a protein fused to the 6×His tag. Left, SDS–polyacrylamide gel electrophoresis (PAGE) analysis of the Ddl protein after purification by metal-affinity chromatography; right, thin-layer chromatography after incubation of purified AvDdl with D-Ala; co-spot, control for D-Ala and D-Ala-D-Ala migration; pET, vector with no insert. (E) FISH of the Av240A probe, labeled with the red fluorophore, to A. vaga embryo nuclei. The arrow points to a signal in a nucleus with condensed chromosomes.
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 bdelloids in which chromosomes occur as co-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)—glycosyltransferase 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-glycosyltransferase mRNA is trans-spliced, as are numerous bdelloid genes (18), and that Alr and Ddl transcription initiates at oppositely oriented promoters located between them (Fig. 1C). Furthermore, the purified A. vaga AvDdl protein over-expressed 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 two different 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 nonmetazoans. We characterized the foreignness of such genes with an alien index (AI), which measures by 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 > 10 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 (20, 21). 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-glycosyltransferase 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 frame-shifts 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 as

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>AI</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>Nonribosomal 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; (Acidobacteria/ 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; (Acidobacteria/ Proteobacteria)</td>
<td>Prolyl oligopeptidase*</td>
</tr>
<tr>
<td>AV10104_Dur3</td>
<td>AvTeLB</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_Roma</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>Nonribosomal peptidase synthetase</td>
</tr>
<tr>
<td>AV10153_KgaA</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-xylanoglucanotransferase</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_β-Gal</td>
<td>AvTeLA</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-β-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>Amidohydrolase</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-β-alanine ligase</td>
</tr>
<tr>
<td>AV10140_PLDc</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>AvTeLB</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>
<tr>
<td>AV10011_FabG</td>
<td>Av212A</td>
<td>0</td>
<td>88</td>
<td>54</td>
<td>6.00E-67</td>
<td>2.00E-28</td>
<td>Bacteria</td>
<td>Short-chain dehydrogenase/reductase</td>
</tr>
<tr>
<td>AV10071_HAL</td>
<td>AvTeLK</td>
<td>0</td>
<td>77</td>
<td>48</td>
<td>2.00E-61</td>
<td>1.00E-27</td>
<td>Bacteria</td>
<td>Histidine ammonia-lyase</td>
</tr>
<tr>
<td>AV10095_GC5N</td>
<td>AvTeLA</td>
<td>0</td>
<td>59</td>
<td>35</td>
<td>2.00E-27</td>
<td>No hits</td>
<td>Bacteria; Proteobacteria</td>
<td>GC5N-related N-acetyltransferase*</td>
</tr>
<tr>
<td>AV10158_FabG</td>
<td>210B3</td>
<td>2</td>
<td>46</td>
<td>41</td>
<td>2.00E-39</td>
<td>2.00E-19</td>
<td>Bacteria</td>
<td>Short-chain dehydrogenase/reductase</td>
</tr>
</tbody>
</table>
as the metazoan long-chain acyl-coenzyme A synthetase (ACS) gene fragment on telomere K, which has an apparently intact 3′ 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 colinear 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.

We have not yet found a case of HGT in which the donor species could be identified by near-identity of a bdelloid gene to that of a putative donor, as in some other taxa (7, 8, 23), or in which the time of transfer could be estimated from the degree of synonymous difference, presumably because the transferred genes and their relatives are not yet represented in databases or because they have resided in bdelloids long enough to have substantially diverged. The similarity of the length distribution and sequence of introns in the putatively foreign genes, including genes of apparently bacterial origin, to that typical of other bdelloid genes (fig. S4) and the similarity of nucleotide composition and codon usage (table S1) suggest that some foreign genes were acquired long enough ago to conform to their host genomes, whereas other intronless genes such as MviM and XynB (fig. S2B) may have arrived more recently.

Our examination of ~1 Mb of bdelloid telomeric/TE-rich sequence (about 1% of the genome) revealed dozens of genes that are of foreign (bacterial, plant, or fungal) origin, and about twice as many indeterminate genes, some of which may be foreign (Figs. 2 and 3, Table 1, and table S1). The AI values for all the CDS in bdelloid telomeric/TE-rich regions indicate that a substantial fraction resulted from HGT, with about one-third having AI > 45 and the majority having AI > 0. In contrast, inspection of comparable DNA amounts from P. roseola and A. vaga gene-rich regions surrounding hsp82, histone, and Hox genes (16, 17, 19) shows that these regions contain mostly genes with known metazoan relatives and are virtually devoid of TEs (Fig. 3, A and B). Indeed, the AI distribution of CDS in these regions is indistinguishable from that for translated expressed sequence tags (ESTs) from the monogonont rotifer Brachionus plicatilis (24) and for CDS from Caenorhabditis elegans (including those near telomeres) and Drosophila melanogaster (Fig. 3C).

The concentration of foreign genes near telomeres may result from preferential incorporation at or near chromosome ends, perhaps at deprotected telomeres, or from selection against more proximal insertions that disrupt essential sequences or are associated with deleterious rearrangements caused by ectopic repair of double-strand breaks via capture of foreign DNA (15, 25, 26). The lack of orientation bias with respect to telomeres indicates that their acquisition does not involve an RNA intermediate copied directly into cDNA, as occurs during terminal retrotransposition. Among ORFs with AI > 45 (Table 1), we found 5 in various degrees of decay and 17 that appear to be intact. Most of the intact foreign genes code for simple enzymatic functions such as carbohydrate decomposition (Table 1) and are not parts of multicomponent pathways that might not function properly after transfer to a distant host (27).

The apparent HGT cannot reasonably be explained by vertical inheritance from the com-
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. Numbers correspond to the total count of entries from each category in a data set: genes (foreign; indeterminate, including repetitive proteins with AI > 45; metazoan; hypothetical from each category in a data set: genes (foreign; indeterminate, including repetitive proteins with AI > 45; metazoan; hypothetical from each category in a
cat e g o r y i nad a t as e t .

Numbers correspond to the total count of entries from each category in a data set: genes (foreign; indeterminate, including repetitive proteins with AI > 45; metazoan; hypothetical from each category in a
cat e g o r y i nad a t as e t .

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

William O. H. Hughes,1* Benjamin P. Oldroyd,2 Madeline Beekman,3 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
topotency. These results provide the first evidence that monogamy was critical in the evolution of eusociality, strongly supporting the prediction of inclusive fitness theory.

References and Notes

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
Supporting Online Material

12 February 2008; accepted 22 April 2008
10.1126/science.1156407

Eusocial behavior, exemplified by social insects, represents one of the pinnacles of sociality and is characterized by individuals altruistically helping to rear siblings rather than 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., helps 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