Evidence for degenerate tetraploidy in bdelloid rotifers

David B. Mark Welch*, Jessica L. Mark Welch*, and Matthew Meselson*†‡

*Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA 02543; and †Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138

Contributed by Matthew Meselson, January 31, 2008 (sent for review November 9, 2007)

Rotifers of class Bdelloidea have evolved for millions of years apparently without sexual reproduction. We have sequenced 45- to 70-kb regions surrounding the four copies of the hsp82 gene of the bdelloid rotifer Philodina roseola, each of which is on a separate chromosome. The four regions comprise two colinear gene-rich pairs with gene content, order, and orientation conserved within each pair. Only a minority of genes are common to both pairs, also in the same orientation and order, but separated by gene-rich segments present in only one or the other pair. The pattern is consistent with degenerate tetraploidy with numerous segmental deletions, some in one pair of colinear chromosomes and some in the other. Divergence in 1,000-bp windows varies along an alignment of a colinear pair, from zero to as much as 20% in a pattern consistent with gene conversion associated with recombinational repair of DNA double-strand breaks. Although pairs of colinear chromosomes are a characteristic of sexually reproducing diploids and polyploids, a quite different explanation for their presence in bdelloids is suggested by the recent finding that bdelloid rotifers can recover and resume reproduction after suffering hundreds of radiation-induced DNA double-strand breaks per oocyte nucleus. Because bdelloid primary oocytes are in G1 and therefore lack sister chromatids, we propose that bdelloid breaks per oocyte nucleus. Because bdelloid primary oocytes are in G1 and therefore lack sister chromatids, we propose that bdelloid colinear chromosome pairs are maintained as templates for the repair of DNA double-strand breaks caused by the frequent desiccation and rehydration characteristic of bdelloid habitats.

axial reproduction | genome structure | evolution of sex | gene conversion | anhydrobiosis

There are numerous hypotheses and theoretical models but no general agreement about what selective factors and underlying mechanisms are responsible for the nearly universal occurrence of sexual reproduction among animals and plants and the relatively early extinction of lineages that abandon it (1–3). Ancient taxax in which sexual reproduction is unknown have therefore attracted attention as systems whose study may shed light on what has allowed them to avoid extinction and successfully evolve. Of these, the group for which evidence of ancient asexuality is strongest is the class Bdelloidea of the phylum Rotifera (4).

Bdelloid rotifers are common invertebrates a few tenths of a millimeter long characteristically found in the water films of mosses and lichens, in temporary freshwater pools, and in other ephemeral aquatic habitats. They are able to thrive in such settings because they can survive desiccation at any stage of their life cycle by entering a metabolically quiescent state of anhydrobiosis (5). Distinguished by their ciliated head structure, bilateral ovaries, and jaw-like mastax, they have ganglia; muscles; digestive and secretory systems; photosensitive and tactile sensory organs; and structures for feeding, crawling, and swimming (6, 7). The class includes 461 described species classified in four families and 19 genera (8). The identification to the family level of bdelloid remains in 35- to 40-million-year-old amber, the high synonymous site difference between bdelloid species of different families, and the molecular clock estimate that bdelloids separated from their sister class, the facultatively sexual monogonont rotifers, at least 100 million years ago all suggest that the class is of Mesozoic origin (9–13).

Despite much observation and study of bdelloid rotifers in the field and in laboratory culture since they were first observed >300 years ago, males, hermaphrodites, vestigial male structures, and meiosis have never been documented within the class (14–16). Eggs are produced from primary oocytes by two mitotic divisions, without chromosome pairing and without reduction in chromosome number (17, 18).

One of the genes that has been characterized most extensively in bdelloids is the heat-shock gene hsp82 of Philodina roseola, of which there are four copies, each on a separate chromosome (19, 20). Here we describe the sequence of the 45- to 70-kb region surrounding each copy of hsp82 in P. roseola. The four regions are found to consist of two colinear, gene-rich pairs. The two pairs have only a minority of genes in common, and synonymous divergence between these genes is consistently much higher than that within a pair. The genes common to all four sequenced regions are in the same order and orientation and are separated by segments present in only one pair or the other. The arrangement therefore suggests that the genome is that of a degenerate tetraploid. Our findings are discussed in relation to bdelloid genome structure and its evolution, the adaptation of bdelloids to ephemeral aquatic habitats, and their putative ancient asexuality.

Results

The 45- to 70-kb regions surrounding each of the four copies of hsp82 in P. roseola are depicted in Fig. 1, and annotations and divergence values are listed in Table 1. It is seen that (i) the four contigs may be grouped in a quartet of two colinear pairs, with gene content, order, and orientation conserved within each pair. The two pairs are designated A and B, respectively, corresponding to the A and B pairs of hsp82 genes we previously identified (19, 20). (ii) The average divergence between members of a pair in 1-kb sliding windows is 3.3 ± 2.6% and 5.3 ± 3.8% for the A and B pairs, respectively, and varies widely along their length: there are regions as long as 1 kb that differ by as much as 20% and tracts of identity or near-identity up to several kilobases in length. (iii) Of the 16 genes identified in the region bracketed by the two outside genes common to both pairs (plasma membrane calcium ATPase and SAICAR synthetase-AIR羧酰酶), only two other genes are present on both pairs (myosin light chain and hsp82). The four common genes are in the same order and orientation on each of the four contigs and are separated by genes that are present in one pair but absent in the other. (iv) Coding sequences of identified genes make up ∼50% of the sequenced regions. (v) Synonymous site

Author contributions: D.B.M.W. and J.L.M.W. designed research; D.B.M.W. and J.L.M.W. performed research; D.B.M.W. and M.M. analyzed data; and D.B.M.W. and M.M. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accesion nos. EU432546–EU432549).

To whom correspondence should be addressed. E-mail: msm@wjh.harvard.edu.

© 2008 by The National Academy of Sciences of the USA

PNAS | April 1, 2008 | vol. 105 | no. 13 | 5145–5149
divergence (Ks) of genes within a pair displays a wide range of values, from zero to \( \sim 20\% \). The nonweighted average Ks is 7.4\%, more than three standard deviations greater than the synonymous and silent site diversity of 2.7\% averaged over diverse regions in a large number of invertebrate species and greater than the average Ks recorded for any one species (21).

Additional findings, not depicted in the figure, are that (vi) introns are numerous and typically 50–60 bp long, with conserved GT-AG boundaries. An intron in copy 2 of methionine-sulfoxide reductase B is absent in copy 1 of the gene, and in plasma membrane calcium ATPase the A pair and the B pair each share an intron absent in the other pair. (vii) Except for a possible foldback element in one of the B contigs (22), there is no evidence of intact or relict transposons or pseudogenes in any of the four contigs. (viii) Within both colinear pairs there are numerous small indels, a few as large as 500 bp. (ix) The A sequences cannot be aligned with the B sequences except in the portions corresponding to the four genes that they share. (x) Ks between A and B copies of these four genes is 112\%, 68\%, 67\%, and 122\%, respectively. (xi) There are no tracts of identity or near identity between A and B sequences. (xii) There are no significant differences between contigs in overall or third position GC content or in codon bias. (xiii) No identified gene contains frameshifts or internal stop codons. (xiv) Ka/Ks ranges from 0.03 to 0.59, indicating the operation of purifying selection on amino acid sequence.

**Discussion**

**Bdelloid Genome Structure.** We find that the sequences around the four copies of hsp82 in *P. roseola* occur as two colinear pairs, a quartet in which the two pairs share a limited number of genes in the same order and orientation but with genes present in each pair that are absent from the other. That the quartet structure of the hsp82 region of *P. roseola* is conserved in Bdelloidea is supported by the finding of an hsp82-containing quartet with a gene content and structure closely similar to that depicted in Fig. 1 in the bdelloid *Adineta vaga* (23), a species belonging to a family that separated from that of *P. roseola* \( \sim 60 \) million years ago (12, 19, 23). Moreover, the existence of such quartet structures along entire chromosomes is suggested by the finding in *P. roseola* of a quartet organization of the regions surrounding a Hox gene and by the observation by FISH that the Hox and hsp82 quartets appear to be far apart on the same four mid-size chromosomes (J.L.M.W., unpublished observations).

That bdelloid genomes are generally organized as quartets is also consistent with the karyotypes of *P. roseola* and *A. vaga*. The karyotype of *A. vaga* consists of 12 mid-size chromosomes, probably comprising three quartets, whereas that of *P. roseola* comprises 10 mid-size chromosomes, a single chromosome of approximately twice that length, and two dots of unequal size (18, 24). The large chromosome appears to be an isochromosome, because probes that hybridize to it do so at pairs of sites, each approximately equidistant from its center (ref. 25 and unpublished data). The 13 chromosomes of *P. roseola* may therefore consist of two quartets of mid-size chromosomes, with the two remaining mid-size chromosomes and the two arms of the isochromosome making up the third quartet. The two dots may be B chromosomes, because one or the other is often missing in karyotypes (J.L.M.W., unpublished observations).

Thus, the evidence suggests that bdelloids are degenerate
tetraploids. The lack of any discernable homology between pairs outside of protein coding regions and the similarity of the hsp82-containing quartet of *P. roseola* to that of *A. vag* indicate that this tetraploidy originated long ago, probably before the radiation of modern bdelloids. Tetraploidization could have resulted from whole-genome duplication (26, 27) followed by extensive gene loss. Although it cannot be said whether any particular segment present in one colinear pair and not in the other has simply been lost through decay and deletion or is instead the result of translocation to or from a site elsewhere in the genome, rampant translocation of small segments would have destroyed evidence of initial linkage. In that case, except for the hsp82-containing segment on which our selection of cosmids for sequencing was based, we would not expect to find segments common to both colinear pairs. In fact, as may be seen in Fig. 1, there are three such additional segments, in the same order and orientation on all four contigs. In addition, the evidence from FISH that hsp82 and a Hox gene are on the same four chromosomes suggests that there are few translocations.

Comparison with other degenerate tetraploids such as *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, and the teleost Tetraodon nigroviridis reveals both similarities and apparent differences from what we find in *P. roseola* (28–31). In all of these cases, consistent with the pattern we see in bdelloids, most genes have been lost from one or the other of the genomes of the tetraploid, mainly as the result of deletions covering one or a few genes. Although deletions greatly predominate over rearrangements in *S. cerevisiae*, *A. thaliana*, and *T. nigroviridis*, many rearrangements have nevertheless occurred. Individual chromosomes in these species are mosaics of numerous blocks, each containing a minority of its genes common to a block on a different chromosome, most of them in the same order and orientation in both blocks, so that each such pair of blocks resembles the A and B contigs depicted in Fig. 1.

Table 1. Genes and divergence between gene copies

<table>
<thead>
<tr>
<th>No.</th>
<th>Gene</th>
<th>A</th>
<th>B</th>
<th>AB</th>
<th>A</th>
<th>B</th>
<th>AB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>U1-like Zn-finger-containing protein</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>ORF of unknown function</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>Gene of unknown function containing DUF288</td>
<td>8.5</td>
<td>1.3</td>
<td>1.3</td>
<td>8.5</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>4</td>
<td>Methionine-sulfoxide reductase B</td>
<td>21.1</td>
<td>0.7</td>
<td>0.7</td>
<td>21.1</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>5</td>
<td>Strabismus</td>
<td>5.0</td>
<td>0.6</td>
<td>0.6</td>
<td>5.0</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>6</td>
<td>Serine/threonine protein kinase</td>
<td>12.0</td>
<td>1.2</td>
<td>1.2</td>
<td>12.0</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>7</td>
<td>Plasma membrane calcium ATPase</td>
<td>9.6</td>
<td>8.5</td>
<td>111.7</td>
<td>0.9</td>
<td>0.8</td>
<td>18.9</td>
</tr>
<tr>
<td>8</td>
<td>7-transmembrane receptor A</td>
<td>13.0</td>
<td>1.9</td>
<td>1.9</td>
<td>13.0</td>
<td>1.9</td>
<td>1.9</td>
</tr>
<tr>
<td>9</td>
<td>7-transmembrane receptor B</td>
<td>13.5</td>
<td>2.8</td>
<td>2.8</td>
<td>13.5</td>
<td>2.8</td>
<td>2.8</td>
</tr>
<tr>
<td>10</td>
<td>7-transmembrane receptor C</td>
<td>9.9</td>
<td>1.1</td>
<td>1.1</td>
<td>9.9</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>11</td>
<td>7-transmembrane receptor D</td>
<td>7.6</td>
<td>1.5</td>
<td>1.5</td>
<td>7.6</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>12</td>
<td>Myosin light chain</td>
<td>1.9</td>
<td>68.2</td>
<td>68.2</td>
<td>1.9</td>
<td>68.2</td>
<td>68.2</td>
</tr>
<tr>
<td>13</td>
<td>Cysteine-rich gene of unknown function A</td>
<td>2.4</td>
<td>0.7</td>
<td>0.7</td>
<td>2.4</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>14</td>
<td>Cysteine-rich gene of unknown function B</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>15</td>
<td>Cysteine-rich gene of unknown function C</td>
<td>7.7</td>
<td>2.1</td>
<td>2.1</td>
<td>7.7</td>
<td>2.1</td>
<td>2.1</td>
</tr>
<tr>
<td>16</td>
<td>82-kDa heat-shock protein</td>
<td>1.3</td>
<td>67.4</td>
<td>67.4</td>
<td>1.3</td>
<td>67.4</td>
<td>67.4</td>
</tr>
<tr>
<td>17</td>
<td>F-box leucine-rich repeat</td>
<td>5.7</td>
<td>2.6</td>
<td>2.6</td>
<td>5.7</td>
<td>2.6</td>
<td>2.6</td>
</tr>
<tr>
<td>18</td>
<td>TPR domain-containing protein</td>
<td>8.7</td>
<td>4.4</td>
<td>4.4</td>
<td>8.7</td>
<td>4.4</td>
<td>4.4</td>
</tr>
<tr>
<td>19</td>
<td>Gene of unknown function containing LicD</td>
<td>9.2</td>
<td>2.2</td>
<td>2.2</td>
<td>9.2</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>20</td>
<td>ORF of unknown function</td>
<td>5.6</td>
<td>1.9</td>
<td>1.9</td>
<td>5.6</td>
<td>1.9</td>
<td>1.9</td>
</tr>
<tr>
<td>21</td>
<td>EGF/LDL receptor</td>
<td>7.8</td>
<td>3.9</td>
<td>3.9</td>
<td>7.8</td>
<td>3.9</td>
<td>3.9</td>
</tr>
<tr>
<td>22</td>
<td>Zn carboxypeptidase</td>
<td>7.9</td>
<td>0.8</td>
<td>0.8</td>
<td>7.9</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>23</td>
<td>SAICAR synthetase-air carboxylase</td>
<td>12.1</td>
<td>67.1</td>
<td>67.1</td>
<td>12.1</td>
<td>67.1</td>
<td>67.1</td>
</tr>
<tr>
<td>24</td>
<td>ORF of unknown function</td>
<td>5.8</td>
<td>1.1</td>
<td>1.1</td>
<td>5.8</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>25</td>
<td>Lung-like 7 transmembrane receptor</td>
<td>9.0</td>
<td>0.3</td>
<td>0.3</td>
<td>9.0</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>26</td>
<td>Phosphoglycerate mutase</td>
<td>9.9</td>
<td>2.5</td>
<td>2.5</td>
<td>9.9</td>
<td>2.5</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Colinear Chromosome Pairs and DNA Repair. As described in the companion article (32), bdelloids are extraordinarily resistant to ionizing radiation, being able to continue reproducing after exposure to a dose causing hundreds of DNA double-strand breaks (DSBs) per genome. As discussed there, the extraordinary resistance of bdelloid rotifers to radiation and desiccation, like that of the bacterium *Deinococcus radiodurans* (33), is almost certainly the result of evolutionary adaptation to repair DNA breaks and to prevent or repair other damage caused by the episodes of desiccation to which bdelloids are exposed in their ephemeral aquatic habitats.

The ability of bdelloid rotifers to remain fertile after experiencing high levels of radiation-induced DSBs and the implication that frequent DNA breakage and repair is a characteristic of the bdelloid lifestyle offer an explanation for why bdelloid genomes are made up of colinear chromosome pairs. Bdelloid primary oocyte nuclei are in G1 and are therefore without sister chromatids (34–36). The colinear chromosomes are therefore the only templates available for the accurate repair of DSBs. Sequence homology between the members of a pair could be maintained by conversion and mitotic crossing-over associated with DSB repair, particularly that occurring in association with desiccation and rehydration, and by selection against clones in which sequence difference reaches levels that substantially reduce the efficiency or accuracy of repair.

The occurrence of homogenizing events is indicated by the sizeable tracts of identity or near-identity between colinear contigs depicted in Fig. 1. Such tracts cannot be chance occurrences. The probability of their chance occurrence may be estimated by using Stephens’s runs test (37) under the assumption that nucleotide differences are distributed at random within a region of given length and overall sequence divergence. For example, the probability of the 470-bp tract of identity in the A
lineage contigs occurring anywhere in the region of their overlap is $3.0 \times 10^{-6}$, and that of the 400-bp tract in the B lineage is $2.8 \times 10^{-7}$. Inspection of the colinear pairs reveals additional long tracts that are nearly identical. Such tracts of identity and near-identity appear to have arisen recently, possibly associated with DSB repair, which may also produce tracts too short to be recognized as such. The correlation of divergence values between adjacent windows is very strong ($r \approx 0.7$) for windows up to $\sim 2$ kb and becomes insignificant by about twice that length (data not shown). This pattern, seen over a wide range of divergence values, would be expected for gene conversion tracts usually of more than a few kilobases occurring over millions of years. The contigs depicted in Fig. 1 exhibit no evidence of recent conversion between A and B lineages, and these are now probably too divergent to function as templates for interlineage homologous repair.

In addition to satisfying the requirement for templates for accurate DNA repair in G1, colinear chromosome pairs may be of additional benefit to bdelloid rotifers by allowing the accumulation of heterotic interactions. Because the average synonymous divergence we find for gene pairs on colinear chromosomes is $\sim 7\%$, such interactions could accumulate and persist for appreciable times between homogenization events. Also, the presence of colinear chromosome pairs and the occurrence of gene conversion may produce clones with combinations of mutations within a gene that are beneficial only in cis more rapidly than could be achieved if each chromosome were represented only once. Furthermore, given the requirement for colinear chromosome pairs, conversion and mitotic crossing-over in clones heterozygous for recessive beneficial mutations may hasten the appearance of better-adapted clones that otherwise could arise only by further mutation in the clone in which they occur, analogous to the effect of allelic segregation and syngamy in sexual species (38, 39).

**Implications of Colinear Pairs for Bdelloid Asexuality.** Our findings call for a reevaluation of the significance of the highly diverged pairs of genes we have found in individual bdelloid genomes (19). Previously, although not rejecting the possibility that bdelloids descend from an ancient polyploid ancestor and engage in some rare or elusive form of sexual reproduction, we thought it more plausible, consistent with other indications of bdelloid asexuality and the presumed rarity of polyploidy in animals (40), to interpret the highly diverged copies of *hsp*82 we found in diverse bdelloid species as descendants of alleles that had ceased segregation and had not undergone homogenization by conversion or mitotic crossing-over since the separation of bdelloid families and therefore as evidence for ancient asexuality (19). The present evidence that bdelloids are degenerate tetraploids suggests that the highly diverged gene pairs are the result of ancient whole-genome duplication, perhaps involving a hybridization between closely related species, rather than descendants of former alleles and raises the question of whether the colinear chromosomes are meiotic homologs. Although it is not entirely excluded that bdelloids engage in rare sexual reproduction, as explained above an entirely different explanation for the maintenance of colinear chromosome pairs follows from the discovery of the extraordinary resistance of bdelloid rotifers to ionizing radiation (32) and the implication that colinear chromosome pairs are required in order for bdelloids to repair DNA breakage associated with their ephemeral aquatic habitats.

**Materials and Methods**

For each of the four copies of *hsp*82, two cosmids with overlapping inserts were selected from a genomic library prepared from a single-egg culture of *P. rosea* as previously described (20). Each cosmid was digested to completion with two or three six-base recognition restriction endonucleases and cloned in pBSKII (Stratagene). Plasmid DNA was extracted manually (41) or with a MiniPrep 24 (MacConnell Research) or PrepRepl (Gene Machines) system. Sequences were obtained by using BigDye Terminator sequencing reagents and ABI automated capillary sequencers or the SequiTherm EXCEL II DNA sequencing kit (Epicentre) and a 4,000-bp DNA scanner (LI-COR). Sequences of the overlapping cosmids were assembled into single contigs with six times average coverage using Sequencher (Gene Codes). Sequence gaps were filled by sequencing directly from 1 μg of cosmid DNA using specifically designed primers. The identity of overlapping regions of cosmids from each contig indicated that the cosmid inserts are not chimeric. Contigs were aligned by using the EMBOSS program stretcher.

Genome regions were identified by BLASTX searches of Drosophila melanogaster, Caenorhabditis elegans, and the National Center for Biotechnology Information nonredundant databases followed by BLASTX and BLASTN searches of putative intergenic regions, by hidden Markov model searches of Pfam full-length and fragment domain models using HMMER (42), and by searches for ORFs. Exons and introns were mapped by comparison to homologous amino acid sequences using genewise (43). Where homology to annotated reference sequences was poor (particularly near exon boundaries), aligned contigs were scanned in all six reading frames for regions in which the ratio of nonsynonymous to synonymous difference is <1.

Divergence at synonymous sites (Ka) and at nonsynonymous sites (Ks) was measured with the diverge program of the Wisconsin Package (Accelyr), which uses the method of Pamilo and Bianchi (44). Sliding window analysis of total divergence was done with DnaSP 4.0. The runs test of Stephens (37) was implemented by using Sterling’s approximation for factorials.

**ACKNOWLEDGMENTS.** We thank the Eukaryotic Genetics Program of the U.S. National Science Foundation for support and the Bay Paul Center W. M. Keck Ecological and Evolutionary Genomics Sequencing Facility.