University of Plymouth

PEARL **https://pearl.plymouth.ac.uk**

Faculty of Science and Engineering School of Biological and Marine Sciences

2018-04-24

Comparative genomics of bdelloid rotifers: insights from desiccating and non-desiccating species

Nowell, R

http://hdl.handle.net/10026.1/11307

10.1371/journal.pbio.2004830
PLoS Biology
Public Library of Science

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

PLOS BIOLOGY

RESEARCH ARTICLE

Comparative genomics of bdelloid rotifers: Insights from desiccating and nondesiccating species

Reuben W. Nowell[1]*, Pedro Almeida[1¤a], Christopher G. Wilson[1], Thomas P. Smith[1], Diego Fontaneto[2], Alastair Crisp[3¤b], Gos Micklem[4], Alan Tunnacliffe[3], Chiara Boschetti[3,5☯]*, Timothy G. Barraclough[1☯]*

1 Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot, Berkshire, United Kingdom, **2** National Research Council of Italy, Institute of Ecosystem Study, Verbania Pallanza, Italy, **3** Department of Chemical Engineering and Biotechnology, West Cambridge Site, University of Cambridge, Cambridge, United Kingdom, **4** Department of Genetics, Cambridge Systems Biology Centre, Downing Site, University of Cambridge, Cambridge, United Kingdom, **5** School of Biological and Marine Sciences, Plymouth University, Portland Square Building, Plymouth, United Kingdom

☯ These authors contributed equally to this work.
¤a Current address: Department of Genetics, Evolution and Environment, University College London, London, United Kingdom
¤b Current address: MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge, United Kingdom
* r.nowell@imperial.ac.uk (RWN); chiara.boschetti@plymouth.ac.uk (CB); t.barraclough@imperial.ac.uk (TGB)

Check for updates

OPEN ACCESS

Citation: Nowell RW, Almeida P, Wilson CG, Smith TP, Fontaneto D, Crisp A, et al. (2018) Comparative genomics of bdelloid rotifers: Insights from desiccating and nondesiccating species. PLoS Biol 16(4): e2004830. https://doi.org/10.1371/journal.pbio.2004830

Academic Editor: Chris Tyler-Smith, The Wellcome Trust Sanger Institute, United Kingdom of Great Britain and Northern Ireland

Received: November 19, 2017

Accepted: March 19, 2018

Published: April 24, 2018

Copyright: © 2018 Nowell et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All raw sequence data have been deposited in the relevant INSDC databases under the accession IDs ERR2135445–55 (Supplementary S1 Table). Reference and maximum haplotype genome assemblies for *Adineta ricciae*, *Adineta vaga*, *Rotaria macrura*, and *Rotaria magnacalcarata* have been deposited at DDBJ/ENA/GenBank under the Project accession ID PRJEB23547.

Abstract

Bdelloid rotifers are a class of microscopic invertebrates that have existed for millions of years apparently without sex or meiosis. They inhabit a variety of temporary and permanent freshwater habitats globally, and many species are remarkably tolerant of desiccation. Bdelloids offer an opportunity to better understand the evolution of sex and recombination, but previous work has emphasised desiccation as the cause of several unusual genomic features in this group. Here, we present high-quality whole-genome sequences of 3 bdelloid species: *Rotaria macrura* and *R. magnacalcarata*, which are both desiccation intolerant, and *Adineta ricciae*, which is desiccation tolerant. In combination with the published assembly of *A. vaga*, which is also desiccation tolerant, we apply a comparative genomics approach to evaluate the potential effects of desiccation tolerance and asexuality on genome evolution in bdelloids. We find that ancestral tetraploidy is conserved among all 4 bdelloid species, but homologous divergence in obligately aquatic *Rotaria* genomes is unexpectedly low. This finding is contrary to current models regarding the role of desiccation in shaping bdelloid genomes. In addition, we find that homologous regions in *A. ricciae* are largely collinear and do not form palindromic repeats as observed in the published *A. vaga* assembly. Consequently, several features interpreted as genomic evidence for long-term ameiotic evolution are not general to all bdelloid species, even within the same genus. Finally, we substantiate previous findings of high levels of horizontally transferred nonmetazoan genes in both desiccating and nondesiccating bdelloid species and show that this unusual feature is not shared by other animal phyla, even those with desiccation-tolerant

Funding: European Research Council (ERC) Advanced Investigator Grant https://erc.europa.eu/ (grant number 233232) to AT. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Natural Environment Research Council (NERC) http://www.nerc.ac.uk/ (grant number NE/M01651X/1) to TGB. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Biotechnology and Biological Sciences Research Council (BBSRC) https://www.bbsrc.ac.uk/ (grant number BB/F020856/1) to AT, CB and TGB. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: BUSCO, Benchmarking Universal Single-Copy Orthologs; CDS, coding sequence; CEGMA, Core Eukaryotic Gene Mapping Approach; CGR, Centre for Genomic Research; CHS, Consensus Hit Support; DSB, double-strand break; GC, guanine-cytosine; HGT, horizontal gene transfer; HGT_C, HGT candidate; INSDC, International Nucleotide Sequence Database Collaboration; LINE, long interspersed element; LTR, long terminal repeat; MAF, minor allele frequency; RNASeq, RNA sequencing; SEG, single-exon gene; SINE, short interspersed element; SNP, single-nucleotide polymorphism; SRA, Sequence Read Archive; TE, transposable element; WGA, whole-genome amplification.

representatives. These comparisons call into question the proposed role of desiccation in mediating horizontal genetic transfer.

Author summary

Bdelloid rotifers are microscopic animals that live in freshwater habitats throughout the world. Two life history characteristics distinguish these common invertebrates as extraordinary. First, they have existed for millions of years apparently without sex: males have not been reported, and females produce genetically identical daughters via parthenogenesis. Second, most bdelloid species are highly resilient to desiccation and can survive without water for extended periods of time (a process known as anhydrobiosis). These 2 attributes have been predicted to leave signatures in bdelloid genomes. Here, we present new draft genomes for 3 bdelloid species and employ comparative genomics to explore the potential impacts of bdelloids' unusual lifestyle on genome structure and content. We find that many proposed genomic consequences of asexuality and desiccation tolerance do not hold true for all species. The genomes of bdelloids may be more similar to those of other animals than previously thought, though a remarkable exception is the high proportion of genes acquired horizontally from nonmetazoan taxa. Our findings necessitate a reevaluation of the effects of asexuality and desiccation on genome evolution in bdelloid rotifers.

Introduction

The bdelloid rotifers are a class of microscopic invertebrates found in freshwater habitats worldwide. Two life history characteristics make these soft-bodied filter feeders unusual among animals. First, bdelloids famously lack males [1] or cytological evidence of meiosis [2,3] and are only known to reproduce via mitotic parthenogenesis. They are therefore one of the best-substantiated examples of a eukaryotic taxon that has evolved apparently without sex or meiosis for tens of millions of years [1,2,4,5]. Famously labelled 'an evolutionary scandal' [6], bdelloids have diversified into over 500 species [7,8] defying the usual fate of asexual lineages [9–11]. Their persistence has implications for theories of the evolution of sex and recombination, a fundamental puzzle in biology [12–15]. A second key feature is that most bdelloid species are remarkably tolerant of desiccation and can survive the loss of almost all cellular water at any stage in their life cycle, including as adults [16,17]. As water evaporates, animals contract their bodies into flat, ellipsoid 'tuns' and enter a dormant state called anhydrobiosis, during which all metabolic activities associated with life are suspended [5,16,18]. Individuals can remain in this condition for long periods, usually days or weeks but occasionally several years [19,20]. The return of water restores metabolism and reproduction, with little evidence of negative fitness consequences for survivors [21]. Species that live in limnoterrestrial habitats such as puddles, leaf litter, and moss are subject to rapid and repeated cycles of drying. The ability to survive desiccation has been proposed to play a key role in bdelloid evolution [5,22].

Early marker-based analyses of bdelloid genomes recovered highly divergent gene copies that were interpreted as nonrecombining descendants of ancient former alleles [4]. Along with the low copy number of vertically inherited transposable elements (TEs) [23], this result was considered positive genetic evidence of long-term asexual evolution. However, subsequent investigations of larger genomic regions revealed evidence of tetraploidy, probably arising

from an ancient hybridisation or genome duplication event affecting diploid ancestors prior to the diversification of bdelloid families [5,24,25]. Genes generally have up to 4 copies, arranged as 2 pairs, with greater divergence between pairs ('ohnologs', also known as homeologs in other polyploid systems) than within pairs ('homologs') [5,24,25]. Another extraordinary feature was that a remarkably high proportion of bdelloid genes show similarity to nonmetazoan orthologs, mostly from bacteria but also fungi and plants, suggesting a rate of horizontal gene transfer (HGT) into bdelloid genomes at least an order of magnitude greater than that observed in other eukaryotes [26]. Many genes originating by HGT from nonmetazoans are expressed and functional [26,27].

The first whole-genome sequence for a bdelloid substantiated many of these findings [28]. The tetraploid genome of *Adineta vaga* comprises both homologous regions with low divergence (median 1.4% for protein-coding genes) and conserved gene order (i.e., high collinearity) as well as ohnologous regions with much higher divergence (median 24.9%) and degeneration of gene order (i.e., low collinearity). The genome encodes remarkably few TEs (approximately 3% of the genome) but a high proportion of foreign genes (approximately 8%), many of which occur as quartets and were therefore presumably acquired prior to tetraploidisation. Unusual structural features were also reported, including a large number of breaks in collinearity between homologous regions and linkage of homologs on the same assembly scaffold, often arranged as genomic palindromes. This assembly cannot, therefore, be decomposed into haploid sets, a finding that was interpreted as further evidence of long-term ameiotic evolution in *A. vaga* [28].

To what extent are the genomic characteristics of bdelloid rotifers explained by their unusual biology and ecology? An important discovery was that bdelloids can survive doses of ionising radiation that would be lethal to nearly any other animal, owing to their ability to repair the resulting DNA double-strand breaks (DSBs) and recover from extensive genome fragmentation [29,30]. Experiments in *A. vaga* showed that comparable genome fragmentation also occurs during desiccation, and this led to the view that bdelloid genomes may be shaped by the need for recovering animals to repair DSBs arising from repeated desiccation [31]. One hypothesis is that homologous gene copies are used as reciprocal templates for repairing DSBs, a process that would act to homogenise homologous regions periodically via gene conversion and select against individuals with excessive divergence because template mismatches would disrupt DNA repair [24,28–31]. In this scenario, the molecular consequences of desiccation are directly linked to the patterns of intragenomic divergence observed in bdelloid genomes, via breakage and repair of DNA. However, the link between desiccation and DSBs is not unequivocal, and evidence from other anhydrobiotic taxa is mixed. For example, DNA integrity is largely maintained in desiccating tardigrades [32–34], but not in the chironomid midge *Polypedilum vanderplanki* [35]. In bdelloids, a key prediction is that species that undergo desiccation more frequently should experience higher rates of DSB repair, resulting in more opportunities for gene conversion and thus a lower level of homologous divergence.

A related hypothesis is that foreign DNA present in the environment may become incorporated into bdelloid genomes via nonhomologous recombination during DSB repair following desiccation, resulting in a higher rate of HGT than is experienced by other eukaryotes [26,28,30,31]. Evidence for high levels of nonhomologous transfer inspired further suggestions that DSB repair might even facilitate homologous horizontal transfer and genetic exchange between individual animals [26,28,30,31]. These ideas remain controversial, however, and recent claims of evidence for DNA transfer between individuals of *A. vaga* [36] have subsequently been identified as artefacts of experimental cross-contamination [37]. A separate recent study reported a striking pattern of allele sharing among 3 individuals of another bdelloid species, *Macrotrachela quadricornifera*, which was interpreted as evidence of sexual

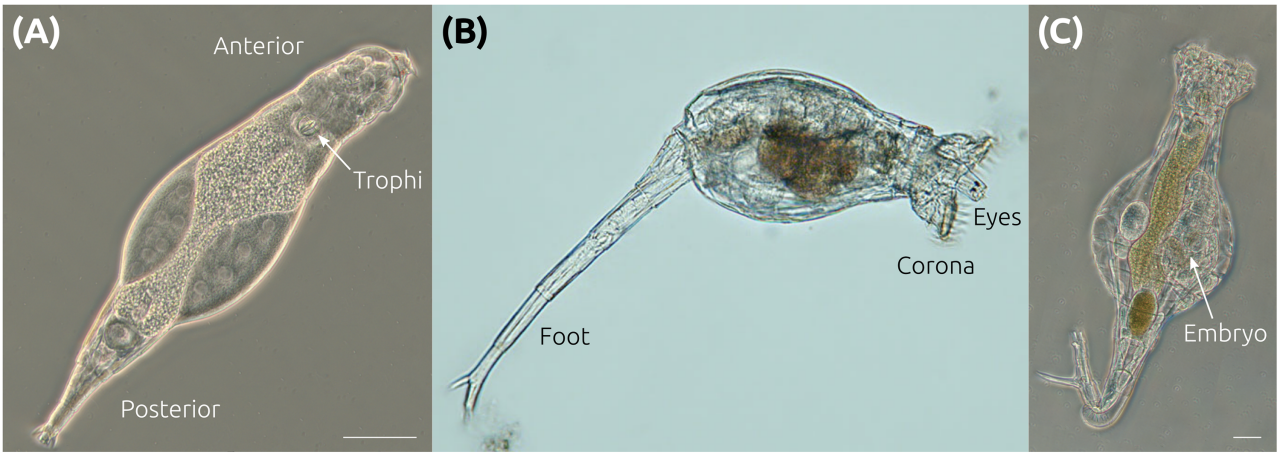

Fig 1. Bdelloid rotifer morphology. (A) *A. ricciae* individual in phase contrast, showing anterior–posterior axis and trophi (jaws). This genus does not possess the ciliated wheel-like organs (corona) on the head that distinguish other rotifers. (B) *R. macrura* showing eyes, extended corona used for feeding and locomotion, and a foot used for substrate attachment. (C) *R. magnacalcarata* with embryo developing internally. Scale bars indicate 50 μm; note: *R. macrura* is approximately the same size as *R. magnacalcarata*. Panels A and C courtesy of C. G. Wilson; panel B courtesy of M. Plewka (www.plingfactory.de).

https://doi.org/10.1371/journal.pbio.2004830.g001

reproduction via an unusual form of meiosis (similar to that of plants in the genus *Oenothera*) [38,39]. However, no evidence for such a mechanism was apparent in a larger study of the genus *Adineta* [40]. In the absence of clear evidence for either occasional sex or horizontal genetic transfer between individual bdelloids (but without discounting either possibility), the nature of bdelloid recombination remains an open question.

Showing both extensive anhydrobiotic capabilities and putatively ancient asexuality, bdelloid rotifers sit at a unique junction in animal evolution. To better understand the relative contributions of these features to bdelloid genome evolution, we have taken advantage of natural variation in the capacity of species to survive desiccation by sampling and comparing whole genomes from multiple taxa. In particular, many species in the genus *Rotaria* live in permanent water bodies and do not survive desiccation in the laboratory [17,41]. Here, we present high-coverage, high-quality genome sequences for 3 species from 2 genera: the desiccation-tolerant species *A. ricciae* and the obligately aquatic, nondesiccating species *R. macrura* and *R. magnacalcarata* (Fig 1). These are compared with the published genome of *A. vaga*. Using a range of assembly approaches, we confirm the conservation of ancestral tetraploidy in all species but demonstrate substantial variation in genome size among species. We then test predictions regarding the effects of desiccation tolerance on intragenomic homologous divergence and investigate genome architecture within species to ask whether the unusual genomic structures observed in *A. vaga* are a general feature across bdelloids. Finally, we contrast a range of genome characteristics, including homologous divergence, HGT content, and repeat abundance across a wider range of animal taxa, allowing us to place some of the unique features of bdelloid genomes in a wider metazoan context.

Results and discussion

Reference genome assembly and annotation

Reference genome sequences for *A. ricciae*, *R. macrura*, and *R. magnacalcarata* were assembled using a combination of long- and short-read sequencing technologies (S1 Table). Kmer spectra of raw and filtered sequencing reads indicated high (>100x) but variable coverage across sites

in each genome (S1 Fig). In addition, a large proportion of low-coverage kmers indicated substantial polymorphism in the *R. magnacalcarata* raw data, most likely corresponding to population variation in the multi-individual DNA sample collected for this species (see Materials and methods). Contaminating reads from non–target organisms were excluded by scrutinising initial draft assemblies for contigs showing abnormal guanine-cytosine (GC) content, coverage, or taxonomic annotations, removing 3.1% of reads from the *R. macrura* dataset, 2.1% from *R. magnacalcarata*, and 6.3% from *A. ricciae* (including approximately 9 Mb sequences annotated as *Pseudomonas* spp.) (S2 Fig). Given the complex patterns of intragenomic divergence and gene copy number observed in other bdelloid species [24,25,28], we adopted 2 assembly strategies. First, reference assemblies were generated with a focus on high assembly contiguity. Reference sequences were constructed using the Platanus assembler [42] and improved using Redundans [43] (S3 and S4 Figs). Second, 'maximum haplotype' assemblies were generated with a focus on maximum separation and resolution of homologous regions, even if this reduced assembly contiguity. This was intended to minimise the confounding effects of assembly 'collapse', a phenomenon whereby homologous regions with no or low divergence are assembled as a single contig with 2-fold coverage relative to separately assembled regions. Further assembly metrics are provided in S1 Data. All assemblies have been submitted to DDBJ/ENA/GenBank under the project accession ID PRJEB23547.

Genome metrics for *A. ricciae*, *R. macrura*, and *R. magnacalcarata* reference assemblies are shown in Table 1, alongside those for the published assembly of *A. vaga* (accession GCA_000513175.1, hereafter referred to as the '2013' assembly) [28]. The reference assembly of *A. ricciae* spanned 174.5 Mb, comprising 4,125 scaffolds with an N50 (length-weighted median) scaffold length of 276.8 kb (Fig 2A). The reference assemblies for *R. macrura* and *R. magnacalcarata* spanned 234.7 and 180.5 Mb over 29,255 and 20,900 scaffolds, respectively, with N50 scaffold lengths of 73.2 and 53.3 kb. The proportion of undetermined bases (i.e., gaps denoted as Ns) was low in all cases, accounting for 2.1%, 0.3%, and 0.8% of the *A. ricciae*, *R. macrura*, and *R. magnacalcarata* reference assemblies, respectively. The GC content was 35.6% for *A. ricciae*, 32.6% for *R. macrura*, and 31.9% for *R. magnacalcarata*; therefore, GC content in *A. ricciae* is higher relative to *A. vaga* (30.8%) and both *Rotaria* species. Gene completeness was assessed by comparing sets of core eukaryotic genes to each reference assembly using Core Eukaryotic Gene Mapping Approach (CEGMA) and Benchmarking Universal Single-Copy Orthologs (BUSCO) [44,45]. Recovery of full-length CEGMA genes ($n = 248$) was 98%, 94%, and 98% for *A. ricciae*, *R. macrura*, and *R. magnacalcarata*, respectively, and gene duplication (average copy number per CEGMA gene) was 2.9, 1.6, and 1.7, respectively. The equivalent recovery of a larger set of BUSCO core metazoan genes ($n = 978$) was 90% for all assemblies, with duplication scores of 2.0, 1.2, and 1.2, respectively. The equivalent completeness and duplication scores for the *A. vaga* 2013 assembly were 96% and 3.0 for CEGMA, and 91% and 2.0 for BUSCO.

The number of genes predicted from each reference assembly varied considerably among species. Gene prediction was performed using BRAKER [47] if RNA sequencing (RNASeq) data were available or MAKER/Augustus [48,49] if not, giving initial estimates of 55,801, 26,284, and 36,377 protein-coding genes for *A. ricciae*, *R. macrura*, and *R. magnacalcarata*, respectively (Table 1). Genes with BLAST matches to TEs (*E*-value $\leq 1 \times 10^{-5}$) and short genes with no matches to UniProt90 (i.e., likely spurious gene models) were removed, resulting in 'high-quality' sets of 49,857, 24,594, and 29,359 protein-coding genes for downstream analyses (Table 1; S2 Data). Reannotation of the *A. vaga* 2013 assembly using MAKER/Augustus resulted in 67,364 predicted genes, reducing to 57,431 after quality control (S2 Data). Therefore, the reference genomes of *R. macrura* and *R. magnacalcarata* appear to encode approximately half the number of genes observed in *A. ricciae* and *A. vaga*. Correspondingly, the

Table 1. Genome assembly and annotation metrics.

Assembly				
Species	***A. ricciae***	***A. vaga***[a]	***R. macrura***	***R. magnacalcarata***
Accession (assembly name)	GCA_900240375.1 (nAr.v1.8)	GCA_000513175.1 ('2013')	GCA_900239685.1 (nRc.v1.3)	GCA_900239745.1 (nRg.v1.8)
Coverage[b] **(mean)**	134.2	Not calculated	210.0	76.3
Span (Mb)	174.5	217.9	234.7	180.5
No. contigs	10,421	36,335	31,105	29,824
Contig N50[c] **(kb)**	58.9	96.7	58.3	24.7
No. scaffolds	4,125	36,167	29,255	20,900
Scaffold N50 (kb) (#N50[d]**)**	276.8 (#176)	260.3 (#240)	73.2 (#828)	53.3 (#890)
Scaffold N90 (kb) (#N90)	34.5 (#868)	7.8 (#1,470)	7.4 (#4,527)	5.3 (#4,560)
Scaffold longest (kb)	1,675.3	1,087.3	692.9	531.4
Gaps (Ns) span (kb) (% genome)	3,668.1 (2.1%)	4,136.3 (1.9%)	640.4 (0.3%)	1,351.6 (0.8%)
% GC	35.6	30.8	32.6	31.9
CEGMA[e] **(*n* = 248)**	C: 98%; F: 99%; D: 2.9	C: 96%; F: 97%; D: 3.0	C: 94%; F: 98%; D: 1.6	C: 98%; F: 99%; D: 1.7
$BUSCO_{EUK}$ (*n* = 303)	C: 97%; F: 98%; D: 2.1	C: 97%; F: 98%; D: 2.0	C: 95%; F: 98%; D: 1.2	C: 96%; F: 97%; D: 1.2
$BUSCO_{MET}$ (*n* = 978)	C: 90%; F: 92%; D: 2.0	C: 91%; F: 92%; D: 2.0	C: 90%; F: 91%; D: 1.2	C: 90%; F: 91%; D: 1.2
Reads mapped (raw) (%)	85.7	Not calculated	98.5	93.4
Transcripts mapped[f] **(%)**	99.1	Not calculated	NA	98.4
Annotation				
Species	***A. ricciae***	***A. vaga***	***R. macrura***	***R. magnacalcarata***
Method	BRAKER	MAKER/Augustus	MAKER/Augustus	BRAKER
No. genes	55,801	67,364	26,284	36,377
No. CDS (includes isoforms)	58,489	67,364	26,284	37,380
No. CDS hit to UniProt90 (Diamond *E*-value $\leq 1 \times 10^{-5}$)	44,974 (77%)	49,915 (74%)	21,014 (80%)	26,542 (71%)
CDS mean length (median) (bp)	1,426.0 (1,059)	1,290.4 (945)	1,596.5 (1,197)	1,260.7 (948)
No. HQ CDS[g]	49,857	57,431	24,594	29,359
Intron mean frequency (introns/gene)	4.8	4.1	5.3	4.0
Intron mean length (median) (bp)	104.1 (55)	118.2 (58)	362.3 (67)	208.4 (61)
Gene density (per Mb)	319.8	300.3	112.0	201.5
Mean intergenic distance (median) (kb)	1.2 (0.6)	1.4 (0.8)	3.9 (2.7)	2.2 (1.4)
Exon span (Mb) (%)	83.4 (47.8%)	86.9 (39.9%)	42.0 (17.9%)	47.1 (26.1%)
Intron span (Mb) (%)	29.0 (16.6%)	33.4 (15.3%)	51.6 (22.0%)	31.5 (17.5%)
No. SEG[h] **(%)**	12,250 (20.9%)	16,504 (24.5%)	1,149 (4.3%)	9,481 (25.3%)
Transcript % GC	38.2	32.8	35.0	34.3
$BUSCO_{EUK}$ (*n* = 303)	C: 96%; F: 99%; D: 2.4	C: 97%; F: 98%; D: 2.3	C: 96%; F: 98%; D: 1.2	C: 99%; F: 100%; D:1.2
$BUSCO_{MET}$ (*n* = 978)	C: 92%; F: 94%; D: 2.4	C: 92%; F: 93%; D: 2.3	C: 91%; F: 93%; D: 1.3	C: 94%; F: 96%; D: 1.3

[a]Assembly metrics for *A. vaga* are based on assembly GCA_000513175.1 by Flot et al. (2013) [28].

[b]Average read coverage based on trimmed and filtered data.

[c]Length-weighted median scaffold length; i.e., the length of the scaffold/contig at which 50% of the total genome span is reached.

[d]The number of scaffolds required to reach N50.

[e]CEGMA/BUSCO notation: C, proportion of genes completely recovered; F, proportion of genes at least partially recovered (includes complete genes); D, duplication rate (average number of copies recovered); *n*, number of genes tested.

[f]Proportion of transcripts with a significant BLAT [46] alignment to the genome.

[g]After removal of short genes (<30 bases) with either no hit to UniRef90 or similarity to known TEs.

[h]Genes without introns.

Abbreviations: BLAT, BLAST-like alignment tool; BUSCO, Benchmarking Universal Single-Copy Orthologs; CDS, coding sequence; CEGMA, Core Eukaryotic Gene Mapping Approach; EUK, Eukaryota; GC, guanine-cytosine; HQ, high-quality; MET, Metazoa; SEG, single-exon gene; TE, transposable element.

https://doi.org/10.1371/journal.pbio.2004830.t001

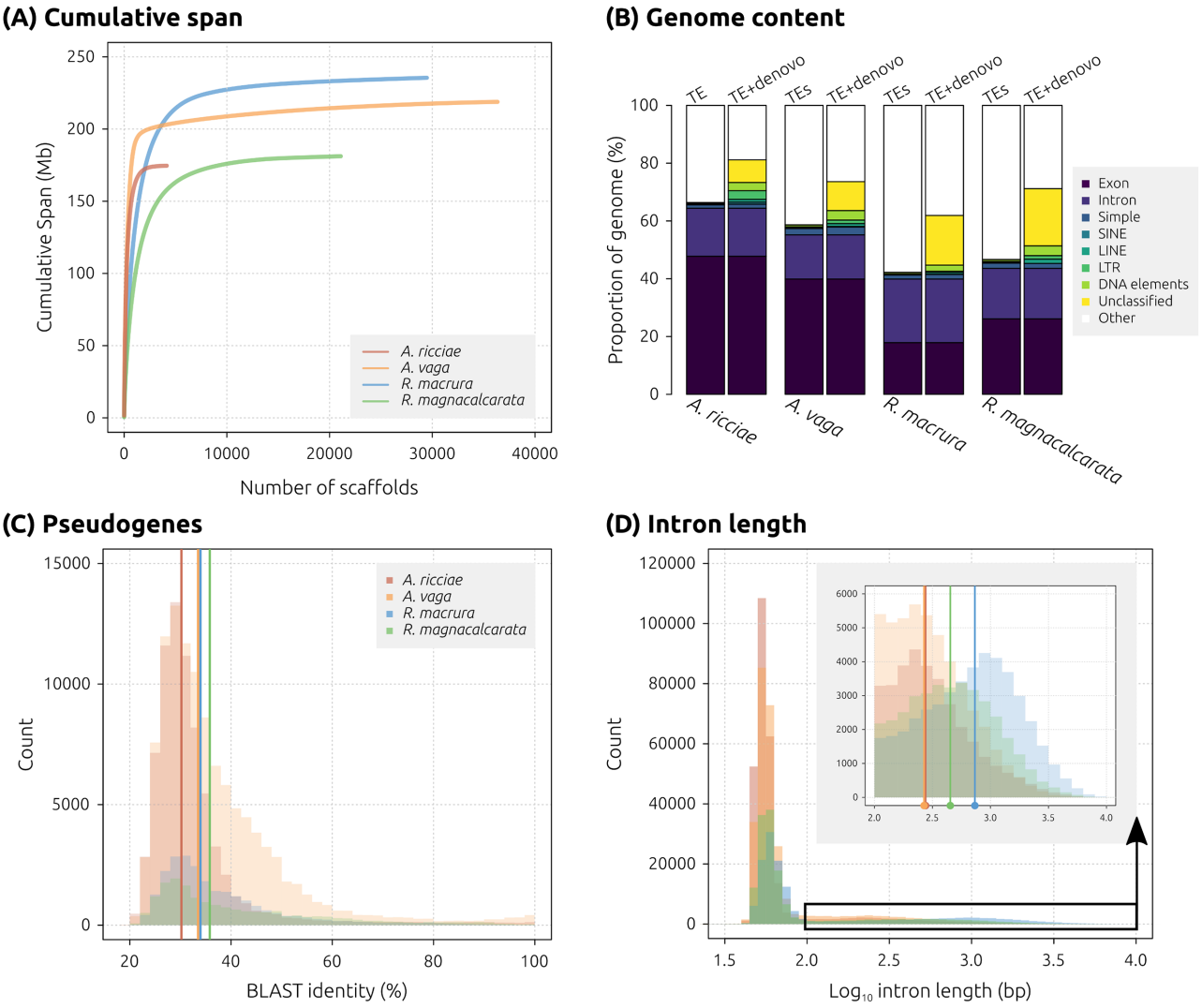

Fig 2. Genome properties of bdelloid rotifers. (A) Assembly contiguity. Cumulative assembly span for *A. ricciae* (red), *R. macrura* (blue), and *R. magnacalcarata* (green) reference assemblies, compared to the published sequence of *A. vaga* (orange). Scaffolds are arranged in descending length order along the x-axis, with cumulative span plotted along the y-axis. More contiguous assemblies achieve their total span with a smaller number of scaffolds, represented by a steeper line with a smaller tail. (B) Genome content. Proportion of each reference assembly covered by exons, introns, and identified repeat elements based on known metazoan TEs only (left-hand column) and TEs plus unclassified repeats detected using RepeatModeler (right-hand column). This shows that the inclusion of ab initio repeats results in substantially greater repeat content in all species, particularly in *Rotaria*. (C) Pseudogene detection. Distribution of percent identity for TBLASTN alignments (*E*-value $\leq 1 \times 10^{-20}$) of predicted proteins to their own genome, discounting hits that overlapped with any existing predicted gene model. Only hits with a query coverage $\geq$95% are plotted. Vertical coloured bars indicate median values. (D) Intron length distributions for predicted genes (species coloured as previously). The inset shows detail of the upper tail of the main distribution (black box, truncated at $\geq$2.0). Vertical bars indicate median values for each species (of truncated distributions). Note $\log_{10}$ scale on y-axis. LINE, long interspersed element; LTR, long terminal repeat; SINE, short interspersed element; TE, transposable element.

https://doi.org/10.1371/journal.pbio.2004830.g002

mean intergenic distance was higher in *R. macrura* (mean 3.9 kb) and *R. magnacalcarata* (2.2 kb) than in *A. vaga* (1.4 kb) or *A. ricciae* (1.2 kb) (S5 Fig).

We checked for misannotations by comparing each set of predicted proteins to the corresponding assembly using TBLASTN (*E*-value $\leq 1 \times 10^{-20}$). This did not reveal any highly similar matches (discounting hits that overlapped with existing gene models), indicating that 'missing' *Rotaria* genes were not the result of poor gene prediction. However, the assemblies of all 4 species showed a large number of matches at lower similarity (30%–35% median identity

at the amino acid level) (Fig 2C). These protein hits to putative noncoding regions may indicate pseudogenes, resulting either from degradation of coding regions following ancestral tetraploidisation or more recent duplications that have subsequently decayed and no longer encode functional proteins.

The structure of predicted genes also varied among species. The average intron length was 104 and 108 bp for *A*. *ricciae* and *A*. *vaga*, respectively, but up to 3 times longer in *R*. *macrura* and *R*. *magnacalcarata* (362 and 208 bp, respectively) (Fig 2D, Table 1; S6 Fig). Distributions of intron lengths showed 2 distinct classes, with the majority of introns falling in the range 30 to 100 bp but a substantial minority showing a higher variance around a much larger mean (inset of Fig 2D). The proportion of single-exon genes (SEGs) was substantially lower for *R*. *macrura* relative to other species, likely reflecting the lack of RNASeq guidance during annotation (Table 1).

The repeat content of bdelloid assemblies was measured following 2 approaches: (1) comparisons to known metazoan repeats, sampled from Repbase, and (2) comparisons to Repbase plus an additional library modelled ab initio from each assembly, using RepeatModeler (see Materials and methods). For (1), the relative abundances of TEs were low for all species, with the total proportion of interspersed repeats accounting for 1.2% of the assembly span for both *A*. *vaga* and *R*. *magnacalcarata*, 0.9% for *R*. *macrura*, and 0.8% for *A*. *ricciae* (S3 Data). Including simple and low-complexity repeats resulted only in modest increases, to 2.0%, 3.4%, 2.2%, and 3.0% for *A*. *ricciae*, *A*. *vaga*, *R*. *macrura*, and *R*. *magnacalcarata*, respectively. For (2), however, the inclusion of ab initio repeats resulted in considerably increased repeat content for all species but to a greater extent in *Rotaria* (16.8% for *A*. *ricciae*, 18.4% for *A*. *vaga*, 22.0% for *R*. *macrura*, and 27.6% for *R*. *magnacalcarata*). A large proportion of ab initio repeats were marked as 'unclassified', and their nature is yet to be determined (Fig 2B; S3 Data). The composition of bdelloid genomes with respect to genome size evolution is considered further below.

Marked differences in intragenomic divergence among bdelloid species

Our assembly results show an apparent 2-fold difference in the number of genes encoded by *Adineta* species relative to *Rotaria* species, suggesting substantial differences in either ploidy or divergence patterns between bdelloid genera. To investigate the evolutionary relationships among genes within each species, we estimated nucleotide divergence and collinearity among separately assembled gene copies using MCScanX [50]. This analysis identifies collinear blocks of genes, defined as pairs of genomic regions that show conserved gene order (see Materials and methods). We plotted the average synonymous divergence (K_S) between genes within each collinear block against a 'collinearity index', defined as the number of collinear genes divided by the total number of genes within a given block (following [28]). Both the *A*. *ricciae* reference and the (reannotated) *A*. *vaga* 2013 assembly showed a clear delineation of genes into both homologs (low K_S and high collinearity) and ohnologs (high K_S and low collinearity) (Fig 3A; S4 Data), as has been observed previously for *A*. *vaga* [28]. The number of *A*. *ricciae* genes that form homologous collinear blocks is 36,593 (73.4% total genes); for *A*. *vaga*, it is 37,061 (64.5%) (S2 Table). Comparisons between *Adineta* species show approximately half as many homologous collinear blocks in *A*. *ricciae* relative to *A*. *vaga* (475 versus 905), but these contain twice as many genes (median 24 versus 11). Therefore, the extent to which we have successfully captured homologous gene copies in *A*. *ricciae* appears to be at least equivalent to that for *A*. *vaga*. Strikingly, however, only ohnologous relationships are inferred in *Rotaria* genomes: collinear blocks composed of homologous genes are not observed (Fig 3A). Comparisons of ohnologous blocks across all species also suggest that the extent of ohnologous

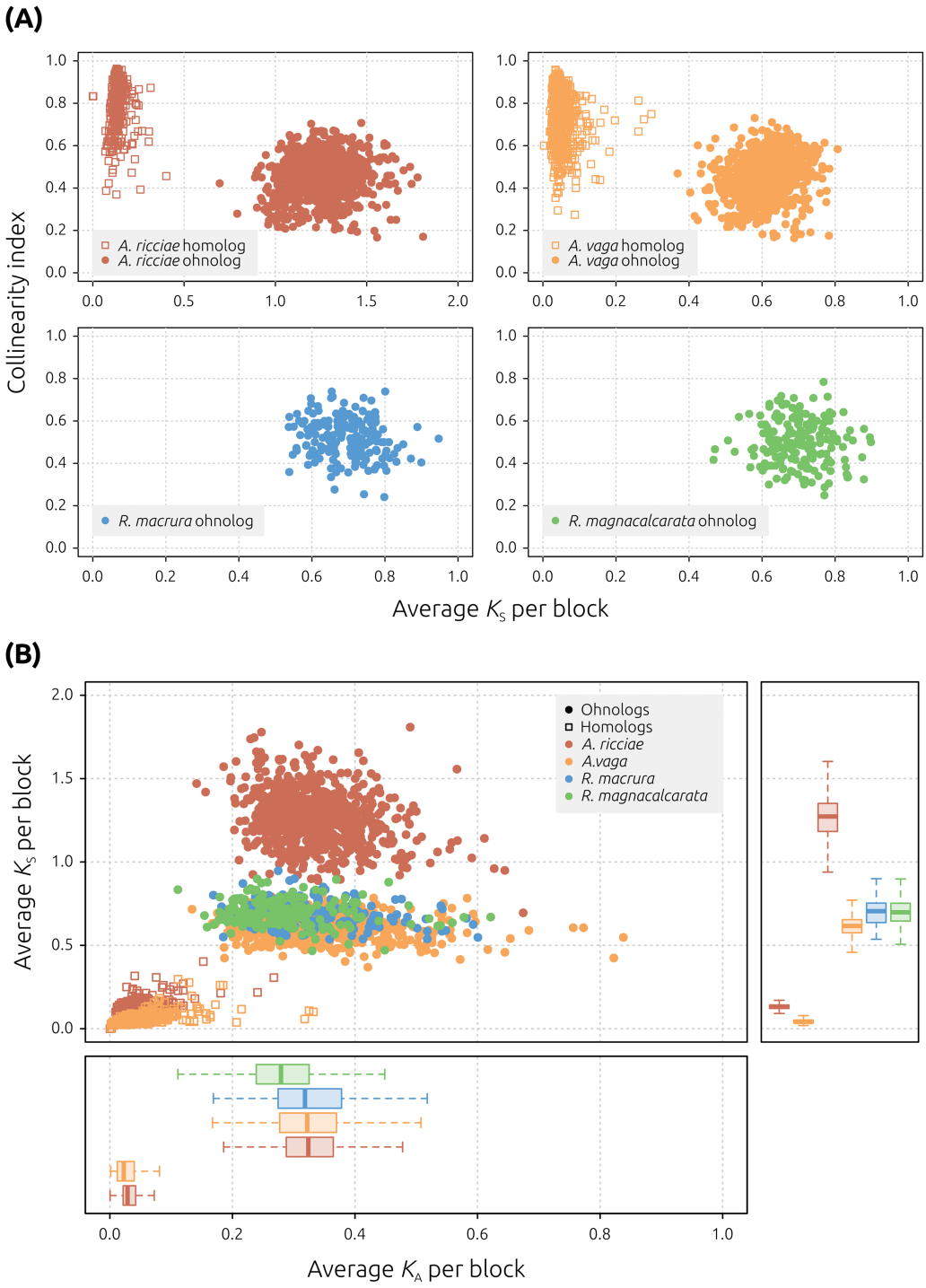

Fig 3. Collinearity in bdelloid genomes. (A) Collinearity versus synonymous divergence. Points represent collinear blocks of genes, plotted based on the average pairwise K_S between pairs of genes within the block (x-axis) and a 'collinearity index', defined as the number of collinear genes divided by the total number of genes within the genomic boundaries of that block (y-axis). Genes within *Adineta* species are clearly differentiated into 2 groups: homologs (low K_S and high collinearity; open squares) and ohnologs (high K_S and low collinearity; filled circles). In *Rotaria* species, however, genes within identified collinear blocks only show high K_S and low collinearity, equivalent to that observed between ohnologs in *Adineta*. Note the different x-axis limit for *A. ricciae*, reflecting a higher synonymous divergence between homologs in this species. (B) Synonymous versus nonsynonymous divergence per collinear block. Genes within homologous blocks (open squares) show a low rate of synonymous (K_S) and nonsynonymous (K_A) substitution. Homologous gene copies are not resolved in *Rotaria* assemblies. Genes within ohnologous blocks (filled circles, found in

all species) show relatively higher rates of both K_S and K_A. Right-hand panel shows elevated mean K_S in both *A. ricciae* homologs (0.14 ± 0.036 [SD] versus *A. vaga* = 0.05 ± 0.026) and ohnologs (1.27 ± 0.146 versus *A. vaga* = 0.61 ± 0.062; *R. macrura* = 0.70 ± 0.080; and *R. magnacalcarata* = 0.70 ± 0.078). This elevation is not observed for K_A (lower panel). Boxplots span the median (thick line), 50% of the values (box), and 95% of the values (whiskers).

https://doi.org/10.1371/journal.pbio.2004830.g003

collinearity is higher in *Adineta* species than in *Rotaria* species (more ohnologous blocks comprising a greater proportion of genes), notwithstanding confounding factors such as differences in the level of assembly fragmentation (S2 Table).

Assuming that the ancestor of extant *Rotaria* lineages was also tetraploid [25], the apparent 'loss' of homologous copies in *Rotaria* species may be caused by either (1) the genuine loss of homologous gene copies from *Rotaria* genomes—resulting in a shift from tetraploidy to highly diverged diploidy—or (2) extremely low levels of divergence between *Rotaria* homologs, such that the majority of homologous sites are identical and cannot be separately assembled (i.e., are collapsed). To differentiate between these hypotheses, we characterised patterns of nucleotide polymorphism and read coverage across each genome, as has been used to investigate the genomes of other polyploid or asexual species [51–55]. Widespread assembly collapse of homologous regions should result in single-nucleotide polymorphisms (SNPs) with a frequency around 50% and a total coverage (read depth of reference plus alternative bases) that is approximately equal to the genome-wide average, analogous to collapsed heterozygous sites in a segregating diploid genome (e.g., see Fig 2A of [42]). These patterns are not predicted under the hypothesis of gene loss in an uncollapsed assembly (i.e., all haplotypes separately assembled), where SNPs may arise in repetitive regions (TEs, tRNAs, low-complexity regions, etc.) but are unlikely to show a frequency of 50% or consistent read depth.

Reads were mapped to the reference assembly of each species, using single-clone (*A. ricciae* and *A. vaga*) or single-individual (*R. macrura* and *R. magnacalcarata*, whole genome amplified [WGA]) libraries. High-quality biallelic SNPs showing a minor allele frequency (MAF) distributed around 50% were detected in all assemblies, indicating at least partial collapse in all cases (Fig 4A, S5 and S6 Data). The relative platykurtosis observed in *Rotaria* species may be an artefact of WGA (inflation of low-frequency SNPs) or lower coverage in general (S7 Fig). In *A. vaga*, the majority of sites (approximately 76%) show coverage around 90x, representing separately assembled regions, with a minor peak at 180x, representing collapsed regions (Fig 4B; S8 Fig). The majority of SNPs occur in regions of 180x coverage, as would be expected under the scenario of localised assembly collapse [28]. For both *R. macrura* and *R. magnacalcarata*, however, read depth at SNP sites (reference plus alternative alleles) varied in concert with the genome-wide coverage distribution (Fig 4B). These patterns indicate that the majority of SNPs occur in collapsed regions, supporting the hypothesis of widespread assembly collapse in *Rotaria* species.

A different pattern is observed in *A. ricciae*, however. Here, a small proportion of sites (11%) are distributed around a peak at 75x coverage, which presumably represents the 1-fold coverage value, but the majority of sites (81%) show 150x coverage and are thus presumably 2-fold covered (i.e., present in double copy) (Fig 4B; S9 Fig). Furthermore, SNP depth is unimodal and is centred on the 150x coverage peak, indicating that the majority of variant sites occur in regions of putative 2-fold coverage. Given the successful capture of the majority of homologous gene copies in *A. ricciae* (approximately 73%, S2 Table), we infer that these conflicting signals are likely derived from another source of coverage heterogeneity that is unrelated to homologous collapse. This is unlikely to be due to an additional whole-genome duplication in *A. ricciae*, given that both *Adineta* species have 12 chromosomes [56,57], but may be caused by other phenomena that affect DNA stoichiometry at the level of either the

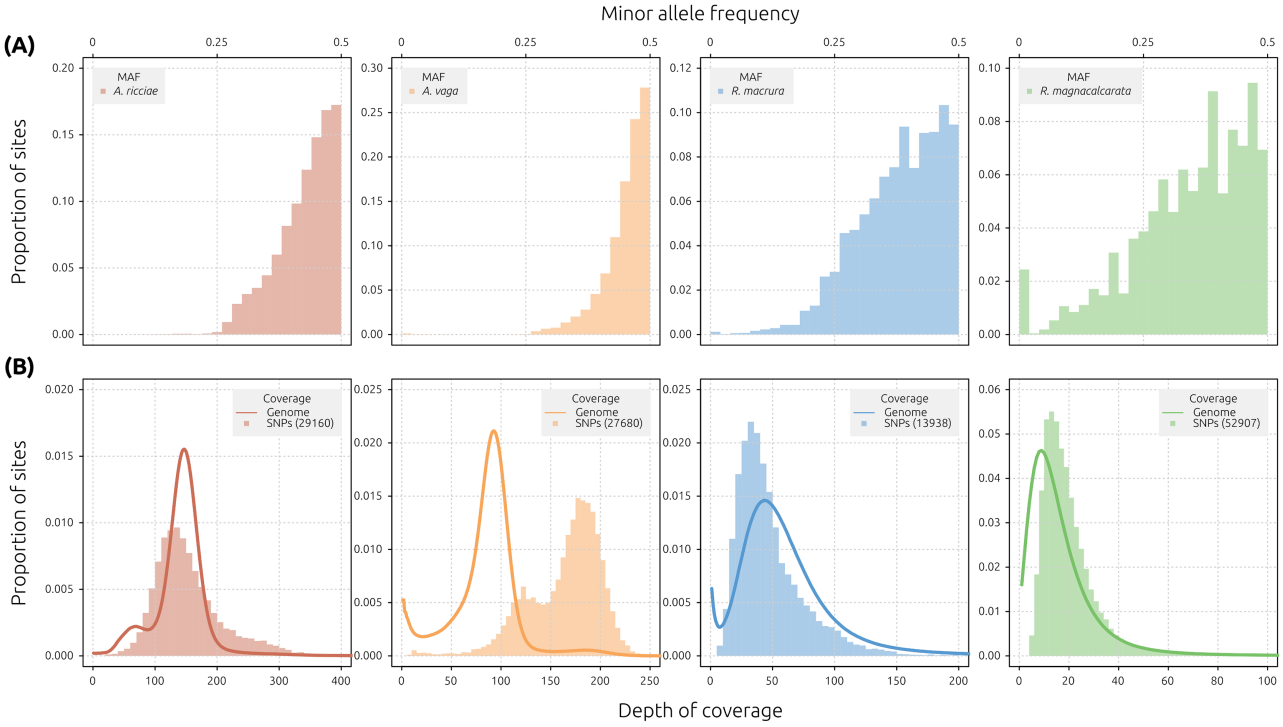

Fig 4. Distributions of MAF and read coverage. (A) Folded MAF spectra for detected SNPs in the reference genomes of *A. ricciae* (red), *A. vaga* (orange), *R. macrura* (blue), and *R. magnacalcarata* (green) are distributed around a mode of 0.5 in all species. (B) Read and SNP coverage distributions. In each plot, the bar histogram represents the distribution of read coverage at SNP sites only, while the overlayed line shows the distribution of read coverage across all sites in the genome. The y-axes indicate proportion sites with given depth of each category (i.e., peak heights are relative to each category). The number of SNPs contributing to the bar histogram is indicated in parentheses (see legend). The cause of the secondary peak in SNP depth (at approximately 125x) for *A. vaga* (library ERR321927) is unknown. MAF, minor allele frequency; SNP, single-nucleotide polymorphism.

https://doi.org/10.1371/journal.pbio.2004830.g004

genome (e.g., segmental or partial genome duplications) or the sample itself (e.g., endopolyploidy [58] or cryptic population structure) (S1 Text, S10 Fig). Further investigations of the *A. ricciae* genome are required to test these hypotheses.

A. ricciae displayed a further difference from other bdelloid genomes: a clear elevation in K_S, both for homologs (compared to *A. vaga*; mean $K_S^{Ar} = 0.135$ versus $K_S^{Av} = 0.05$; $t = 47$, $P < 0.01$) and for ohnologs (e.g., mean $K_S^{Ar} = 1.267$ versus $K_S^{Av} = 0.613$; $t = 124$, $P < 0.01$) (Fig 3B; S3 Table). No such elevation was observed in the rate of nonsynonymous substitution in *A. ricciae*, compared to the other species. However, the *A. ricciae* genome also shows the highest GC content of the 4 species (approximately 5% higher than *A. vaga*, and 3% to 4% higher than either *Rotaria* species). Therefore, one explanation for the increase in K_S may be selection for increased GC content in *A. ricciae*, with continued purifying selection at nonsynonymous sites [59].

Homologous divergence in *Rotaria* species is lower than allelic divergence in most sexual species

In the *A. ricciae* reference and *A. vaga* 2013 assemblies, the majority of homologs were separately assembled, allowing for the identification of homologous gene copies and the estimation of their sequence divergence using a BLAST-based approach [51,55]. The median divergence between separately assembled homologous gene copies was 4.55% (mode = 3.75%) in *A. ricciae* and 1.42% (mode = 1.25%) in *A. vaga* (in agreement with [28]) (Fig 5A). However, these

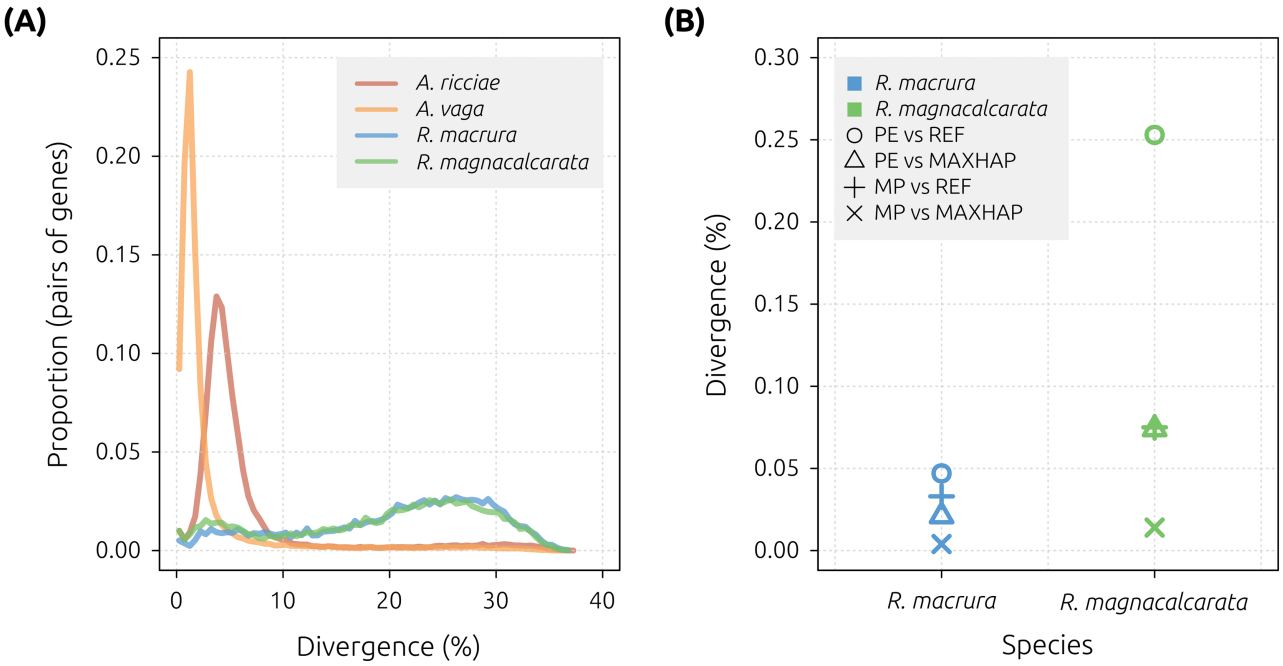

Fig 5. Estimates of intragenomic divergence. (A) Distribution of sequence identity (expressed as a divergence, i.e., proportion of nonidentical sites) for the top non-self BLAST hits from intragenomic comparisons for each species, showing highly similar gene copies within *Adineta* genomes but not within *Rotaria* genomes. Median values are 4.55% (*A. ricciae*, red), 1.42% (*A. vaga*, orange), 22.5% (*R. macrura*, blue), and 21.7% (*R. magnacalcarata*, green). (B) Homologous divergence estimated from SNP counts in coding regions in REF and MAXHAP assemblies for *R. macrura* and *R. magnacalcarata*. Mean estimates are 0.026% for *R. macrura* and 0.104% for *R. magnacalcarata*. MAXHAP, maximum haplotype; MP, mate pair; PE, paired end; REF, reference; SNP, single-nucleotide polymorphism.

https://doi.org/10.1371/journal.pbio.2004830.g005

estimates may be inflated because they fail to consider homologous regions with low divergences that are collapsed. Alternative estimates of homologous divergence based on SNPs detected in the collapsed *A. vaga* assembly were correspondingly lower (0.955% and 0.788%, based on alignment of libraries ERR321927 and SRR801084, respectively) (S5 Data).

Exploration of alternative assembly strategies, which aimed to minimise as much as possible the phenomenon of assembly collapse, did not result in the separate assembly of *Rotaria* homologous regions (S1 Data), indicating substantially lowered homologous divergence relative to *Adineta*. To estimate the divergence between collapsed *Rotaria* homologous gene copies, we instead counted the number of SNPs occurring in coding regions in each assembly. Based on single-individual, WGA mate-pair libraries aligned to the *R. macrura* and *R. magnacalcarata* reference assemblies, a total of 13,115 and 36,594 SNPs were detected across 40.0 and 40.3 Mb of coding sequences (CDSs), respectively (S5 and S6 Data). Assuming that all detected SNPs are the result of homologous collapse, an upper limit for the divergence between homologs is estimated at 0.033% and 0.075% for *R. macrura* and *R. magnacalcarata*, respectively (Fig 5B).

These results indicate that homologous divergence in nondesiccating *Rotaria* species is at least an order of magnitude lower than that observed in anhydrobiotic *Adineta* species. This contradicts hypotheses that emphasise the role of desiccation in shaping patterns of divergence in bdelloid genomes. If the rate of desiccation-induced DSB repair is positively correlated with the rate of gene conversion, a lower level of homologous divergence is expected in species with higher rates of desiccation. In fact, we observe the opposite: divergence between homologs in nondesiccating *Rotaria* species is considerably lower than in *A. ricciae* (median 4.6%), *A. vaga*

(1.4%) (here and [28]), and *Philodina roseola* (3% to 5%) [24], all of which are capable of anhydrobiosis. Across sexual eukaryotes, estimates of allelic divergence range from about 0.01% to 8% [60]. Therefore, none of these bdelloid species is beyond the range of observed values for sexual taxa, although some fall near the extremes of this distribution (e.g., *Rotaria* are towards the lower end, in contrast to *A*. *ricciae* and *P*. *roseola*).

How can we reconcile theory with these observations? The simplest explanation is that homologous divergence is unlinked to desiccation and that the observed differences are instead reflective of underlying phylogeny. Alternatively, it may be that homogenisation between homologous gene copies in *Rotaria* is not caused by desiccation-induced DSB repair but by gene conversion arising during a different process, such as mitotic crossing over [61–63]. In the yeast *Saccharomyces cerevisiae*, for example, various forms of mitotic recombination can produce tracts of gene conversion many kilobases long, often initiated from DNA nicks that are subsequently processed into DSBs [64–66]. Such processes may be especially pronounced and irreversible in asexuals: rapid loss of heterozygosity is observed in recent asexual lineages of the water flea *Daphnia pulex*, in which high rates of initial heterozygosity in hybrid asexual lineages are rapidly eroded via gene conversion and hemizygous deletion, which may ultimately limit their longevity [67].

However, this begs the question: why should the same homogenising mechanisms not operate in desiccation-tolerant species? One possibility is that desiccation-tolerant species have low or negligible background rates of gene conversion while hydrated, thanks to selection for highly effective DNA repair and error-checking mechanisms imposed by environments that desiccate on a regular basis. Such a repair system might faithfully prevent loss of diversity in the context of mitosis, even while occasional gene conversion remains an unavoidable consequence of the more demanding repairs required after desiccation. Alternatively, perhaps similar homogenising forces do indeed operate in desiccation-tolerant rotifers but are counteracted by mutations generated during repair of desiccation-induced DSBs, whose net effect is to sustain high rates of homologous divergence [68]. Positive evidence for a link between desiccation and DSBs in bdelloids is currently limited to experiments in a single species, *A*. *vaga* [31], and evidence from other anhydrobiotic taxa is mixed [32,33,35]. Further work on DNA integrity and genome evolution in bdelloids is needed to address these divergent predictions. A final explanation that cannot be entirely excluded is that low homologous divergence in *Rotaria* genomes results from cryptic sexual reproduction, constrained by small population sizes (i.e., inbreeding), although no males have so far been detected in *Rotaria* or any other bdelloid [69].

Architectural signals of long-term ameiotic evolution are lacking in *A. ricciae*

The *A*. *vaga* 2013 assembly (accession GCA_000513175.1) showed a number of unusual structural genomic features, including breaks in homologous collinearity, physical linkage of homologous genes (i.e., encoded on the same scaffold), and genomic palindromes of the form $g1_A$, $g2_A$, $g3_A$. . .$g3_B$, $g2_B$, $g1_B$, where A and B denote homologous copies of genes g1, g2, and g3. Such features would result in chromosomes that cannot be decomposed into haploid sets and thus imply a genome architecture that is incompatible with conventional meiotic pairing and segregation, as might be predicted under the hypothesis of long-term asexuality [28].

To test for such structures in other bdelloid genomes, we first analysed the reannotated *A*. *vaga* 2013 assembly. A total of 298 breaks in collinearity (32.9% of 905 homologous blocks) were detected (an example is shown for scaffold AVAG00001 in Fig 6A). In addition, 25 homologous blocks were encoded on the same genomic scaffold, 2 as tandem arrays and 23 as

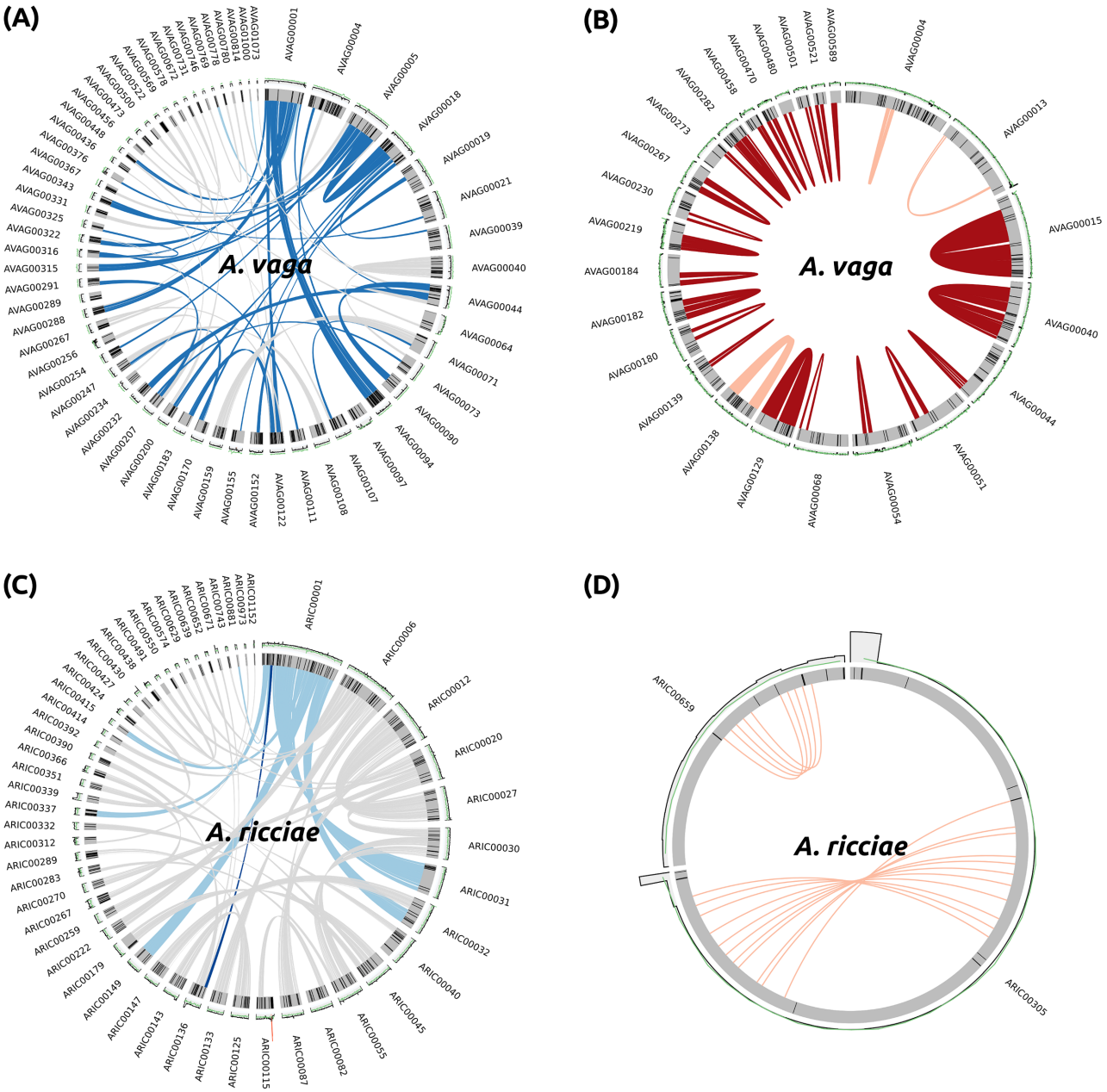

Fig 6. Unusual genomic features are not detected in *A. ricciae*. Genomic scaffolds are shown as grey bars, GC content (green line), and read coverage (grey histogram) averaged over 5-kb bins are shown above each scaffold. Black lines within scaffolds represent gaps (Ns) introduced during scaffolding. (A) Example of collinearity breaks in *A. vaga* 2013 assembly scaffold AVAG00001. Links between homologous blocks are shown in light blue; downstream connections are shown in grey. Collinearity breaks are shown in dark blue. (B) The majority of homologous blocks encoded on the same scaffold in the *A. vaga* 2013 assembly are palindromes (dark red); tandem repeats are shown in pink. (C) Example of a single collinearity break on *A. ricciae* scaffold ARIC00001. (D) No genomic palindromes detected in *A. ricciae*. Only 2 homologous blocks were found on the same scaffold (ARIC00305 and ARIC00659), both arranged as tandem repeats. GC, guanine-cytosine.

https://doi.org/10.1371/journal.pbio.2004830.g006

palindromes (Fig 6B). Thus, our detection methods were able to recover the same signals of ameiotic evolution reported by Flot et al. (2013) [28], for the same *A. vaga* assembly. The method of Flot et al. (2013) was to construct contigs from Roche 454 Titanium and GS-FLX data using the MIRA assembler [70], followed by correction and scaffolding using high-

coverage Illumina paired-end and mate-pair data (section C1 of [28] supplement). We attempted to reassemble the *A*. *vaga* paired-end Illumina data independently (incorporating both mate-pair and 454 data for scaffolding) using a variety of established short-read assemblers. However, this consistently resulted in highly fragmented assemblies (e.g., N50 of approximately 1 to 2 kb), except when allowing for the collapse of homologous regions (S1 Data). The lack of contiguity in *A*. *vaga* maximum haplotype assemblies precluded us from using alternative assembly approaches to investigate the features detected in the 2013 assembly. A lack of separately assembled homologous gene copies in either *Rotaria* species similarly precluded structural analysis.

The closely related species *A*. *ricciae*, however, showed both high assembly contiguity and a majority of separately assembled homologous gene copies. For this species, we detected only 8 collinearity breaks (1.7% of 466 homologous blocks) between homologs (an example is shown in Fig 6C), in contrast with hundreds inferred from the *A*. *vaga* 2013 assembly. Assuming that the *A*. *ricciae* and *A*. *vaga* 2013 assemblies are structurally accurate and that the homologous gene copies captured in both assemblies reflect the ancestral tetraploidisation common to all bdelloids, collinearity appears to be markedly more conserved in *A*. *ricciae* relative to *A*. *vaga*. However, many of the detected breaks in both species span regions separated by scaffold gaps (Ns introduced during the joining of contigs), suggesting that at least some detected collinearity breaks may be the result of scaffolding errors despite requisite care during assembly (Fig 6A and 6C). For example, an unscaffolded *A*. *ricciae* assembly showed only a single break in collinearity, although the increased fragmentation of this assembly (N50 = 18.7 kb) may limit our ability to detect such breaks. We did not detect any cases of homologous genes arranged as palindromes in *A*. *ricciae*: only 2 cases of linked homologous blocks were detected, both tandem repeats (Fig 6D).

Overall, these results suggest that certain unusual genomic features, previously interpreted as positive signatures of long-term ameiotic evolution in *A*. *vaga* [28], are largely absent from the closely related *A*. *ricciae* (and remain untested in other bdelloids). These patterns may reflect true differences between *Adineta* species, although no marked dissimilarity in karyotype is evident [56,57]. Alternatively, they may be either false-positive or false-negative artefacts of applying alternative assembly methodologies to complex genomes with different patterns of intragenomic divergence. Evidence from other taxa is limited. For example, similar features have been reported in the recently assembled genomes of the parthenogenetic springtail *Folsomia candida* [71] and in certain apomictic species of *Meloidogyne* root-knot nematodes [72]. However, these involved only small proportions of each respective genome, and the latter study used the same assembly approach that was applied to *A*. *vaga* [72]. Furthermore, recent investigations of genome evolution in asexual, nondesiccating *Diploscapter* nematodes have revealed a high degree of collinearity among homologous genes, despite high levels of divergence (approximately 4%, thus similar to *A*. *ricciae*) [54,73]. These data suggest that transitions to asexuality do not necessarily lead to the erosion of collinearity. Similar structural variants are also detected in many sexual organisms—for example, humans [74,75], cichlid fishes [76], and cows [77]—and may involve translocations or duplications that are many kilobases in length. Therefore, further work is required to improve and validate assembly contiguity of bdelloid genomes and to ascertain the evolutionary significance of these features.

Variation in genome size driven by expansion of noncoding elements in *Rotaria* species

Although variation in genome size among bdelloids has been inferred previously based on cytofluorometry of oocytes [78], there are some inconsistencies between reported values that

indicate possible errors (see S2 Text for details). Based on our assembly results, estimations of global genome properties such as total span and gene number can be estimated bioinformatically, using both kmer- and assembly-based approaches. The maximum haplotype assembly for *A*. *ricciae* was approximately 201 Mb in length encoding 63,000 genes, while reannotation of the partially collapsed 217-Mb *A*. *vaga* 2013 assembly showed approximately 67,000 genes. While maximum haplotype assemblies for *A*. *vaga* were highly fragmented (and therefore poorly annotated), a collapsed *A*. *vaga* assembly, reduced to 109 Mb, encoded 31,600 genes. Notwithstanding potential complexities of coverage heterogeneity in the *A*. *ricciae* data (discussed above), these values suggest that the full complement of genes in *A*. *ricciae* may be in the region of 60,000 to 65,000 genes across a total span of about 200 Mb (S11 Fig). The genome of *A*. *vaga* is likely to be of approximately equivalent size [28]. The largely collapsed reference assemblies of *R*. *macrura* and *R*. *magnacalcarata* showed about 25,000 and 35,000 genes, respectively, and thus are in broad agreement with observations from *Adineta*. This also implies that the total genome size for *Rotaria* is in the region of 400 to 500 Mb, assuming the majority of sites are in double copy (S11 Fig), and indicates that the genomes of *Rotaria* species may be considerably larger relative to *Adineta* species.

What mechanisms might explain these observed differences? Based on comparisons to known metazoan TEs from Repbase, the abundance of TEs and low-complexity repeats was low in all species, suggesting that expansions of known TEs or simple repeats in the *Rotaria* lineage is unlikely to be a major driver. However, the inclusion of the largely unclassified ab initio repeats did result in a marked increase in total repetitive sequences for all species (17%, 18%, 22%, and 28% for *A*. *ricciae*, *A*. *vaga* 2013, *R*. *macrura*, and *R*. *magnacalcarata*, respectively). The relative increase is greatest in the *Rotaria* species, suggesting that a substantial fraction of the *R*. *macrura* and *R*. *magnacalcarata* reference assemblies are covered by repeats whose exact nature remains to be elucidated. In addition, average intron sizes in *Rotaria* genes are longer (by at least 100%), driven primarily by an increase in the number of long introns. Intriguingly, a similar association between desiccation tolerance and genome 'compaction' has been observed in tardigrades: *Hypsibius dujardini* has a genome size of 104 Mb and only survives desiccation under certain conditions, whereas *Ramazzottius varieornatus* has a much smaller genome size of 56 Mb and is capable of rapid anhydrobiosis [34,79]. Future sampling of more phylogenetically independent comparisons of desiccating and nondesiccating species is needed to test these ideas.

Genome content analysis substantiates high levels of HGT, low numbers of TEs, and presence of meiosis genes

To better understand how bdelloid genomes compare to those of other metazoans, we characterised an additional 13 species from across the Protostomia (S4 Table) based on genome size, gene density, patterns of orthologous gene clustering, HGT content, and repetitive sequence content (Fig 7, S12 Fig, S5 and S6 Tables). Our comparison included a broad taxonomic range of species from different ecological niches, including molluscs [80–83], annelids [81], the platyhelminth *Schistosoma haematobium* [84], the desiccation-tolerant tardigrade *R*. *varieornatus* [34], the orthonectid intracellular parasite *Intoshia linei* [85], and chromosomal-level reference genomes for *Caenorhabditis elegans* [86] and *Drosophila melanogaster* [87]. Phylogenetic relationships among species were not estimated directly but inferred from the literature [85,88,89].

We assessed the extent of horizontal transfer into protostome genomes using both sequence comparison and phylogenetic approaches. The extent to which HGT contributes to the genomes of multicellular eukaryotes is controversial. For example, a recent claim of 17%

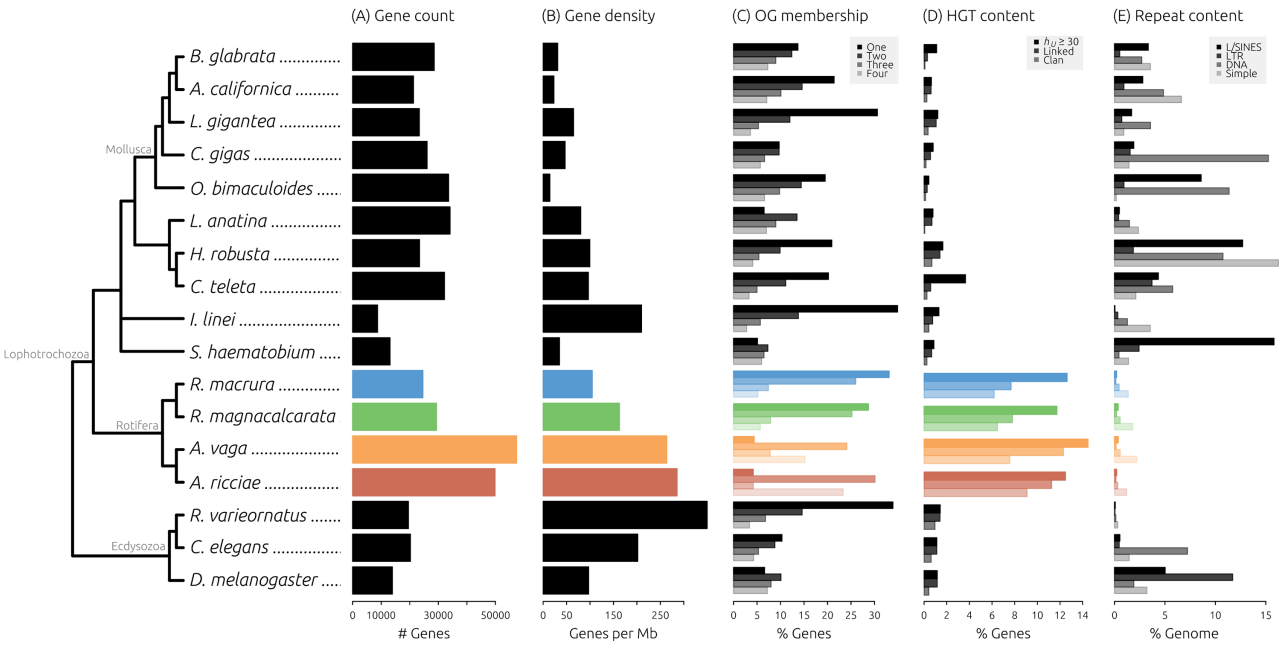

Fig 7. Bdelloid genome characteristics compared to other metazoans. Phylogenetic relationships among 17 protostome species are shown on the left, major groups are marked in grey. (A) Gene count, defined as the number of protein-coding genes in each genome (taken from Ensembl Metazoa [90] or from the relevant literature). (B) Gene density, defined as the number of genes divided by the genome span (inferred from assembly). (C) Same-species OG membership, with bars from top to bottom representing 1-, 2-, 3-, and 4-member clusters. (D) HGT content, with bars from top to bottom representing the proportion of genes with $h_U \geq 30$ for each species (HGT_C), in physical linkage with a known metazoan gene ('Linked'), and with clan membership to nonmetazoan genes from phylogenetic analyses ('Clan'). (E) Repeat content, with bars from top to bottom representing the proportion of each genome covered by LINE/SINEs, LTR elements, DNA elements, and simple/low-complexity repeats. HGT, horizontal gene transfer; HGT_C, HGT candidates; LINE, long interspersed element; LTR, long terminal repeat; OG, orthologous group; SINE, short interspersed element.

https://doi.org/10.1371/journal.pbio.2004830.g007

nonmetazoan genes encoded in the genome of the tardigrade *H. dujardini* was later shown to be derived mostly from contaminating non–target organisms [91–94]. Nonetheless, a high proportion of genes from a variety of nonmetazoan sources has consistently been inferred in bdelloid genomes from a range of independent data, including fosmid sequences [26], transcriptomes [27,41], and whole-genome data [28]. To measure the level of horizontal transfer, we developed an HGT assessment pipeline that uses both sequence comparison and phylogenetic approaches to build a body of evidence for the foreignness of each predicted gene. Our goal was not to unequivocally assert the evolutionary history of individual genes but rather to apply these tests consistently across the set of animal genomes as a fair comparison for estimating HGT.

Our initial screen identified 6,221 (12.5%), 8,312 (14.5%), 3,104 (12.6%), and 3,443 (11.7%) genes from *A. ricciae*, *A. vaga*, *R. macrura*, and *R. magnacalcarata*, respectively, as HGT candidates (HGT_C) (Fig 7D, S7 Data). These values are substantially higher than the proportion of HGT_C observed in any other protostome species included in this analysis, using the same pipeline and thresholds (the highest proportion of HGT_C for a non-bdelloid was 3.6%, for the annelid worm *Capitella teleta*). This is also noticeably higher than estimates based solely on the Alien Index. For each HGT_C, we then assessed (1) the presence of predicted introns, (2) scaffold linkage to another gene of unambiguous metazoan origin, (3) presence on a scaffold that encodes a high HGT_C proportion that might indicate contamination, (4) membership within a 'clan' of nonmetazoan orthologs, and (5) monophyly of the HGT_C with all present nonmetazoan orthologs to the exclusion of all metazoan orthologs (see Materials and methods). Testing for clan membership with nonmetazoan orthologs reduced the proportion of HGT_C to

9.1%, 7.6%, 6.2%, and 6.5% for the 4 bdelloids, compared with <1% for all other species (S7 Data). The final test was not applicable for the majority of HGT_C because metazoan orthologs were often not detected; thus, the number of genes that additionally showed evidence for monophyly with nonmetazoan orthologs was 189, 190, 111, and 82 for *A. ricciae*, *A. vaga*, *R. macrura*, and *R. magnacalcarata*, respectively, and was reduced to a handful or 0 in all other species. Sequential BLAST analysis of HGT_C 'clan' genes showed that the majority (approximately 80%) were found in all 4 bdelloids, suggesting that most detected HGT_C genes are of ancient origin (S13 Fig). Correspondingly, many HGT_C genes were also found as either pairs or quartets (for *A. ricciae* and *A. vaga*) or as singletons or pairs (for *R. macrura* and *R. magnacalcarata*) (S13 Fig).

These comparisons support previous findings of a high proportion of nonmetazoan genes in bdelloid genomes [26,28,41,95]. Compared to other metazoans, and at all levels of scrutiny, the 4 bdelloid genomes analysed here showed a substantially greater proportion of genes from nonmetazoan sources than do any other species in our comparison. Our results confirm a substantial proportion of foreign genes in the nondesiccating *Rotaria* genomes, in agreement with recent findings based on transcriptomes [41]. Our assessment also showed very low levels of HGT (approximately 1%) into the genome of the anhydrobiotic tardigrade *R. varieornatus*, in agreement with recent estimates [34,79]. In addition, recent genome investigations of the anhydrobiotic chironomid insect *P. vanderplanki*, which experiences a large number of DNA breakages during desiccation [35], also did not reveal an elevated rate of HGT [96]. Taken together, these findings bring into question the association between anhydrobiosis and elevated rates of HGT that previously has been suggested for bdelloids [26,28,30,31,95]. One explanation may be that differences in HGT content reflect species-specific differences in the mechanism of anhydrobiosis, in combination with particular ecological properties of each species. Further comparative work is thus required to elucidate any relationship between anhydrobiosis and horizontal transfer.

An alternative possibility is that HGT content in bdelloids does not reflect a deviation in the rate of import of foreign genes but an increased rate of retention arising from their putative longstanding asexuality. Based on transcriptome data from *Rotaria* species, Eyres et al. (2015) estimated the rate of gain to be low in absolute terms, on the order of approximately 10 HGT gains per lineage per million years [41]. Perhaps this is a typical background rate of import for organisms with similar ecological and physiological properties to bdelloid rotifers, but acquired genes are able to persist for longer in an ameiotic background given the lack of mechanisms such as segregation and unequal crossing over that would otherwise remove them. If so, foreign genes incorporated by asexuals, even if initially deleterious, might persist over the extended timescales necessary for domestication. The high proportion of nonmetazoan genes accumulated in bdelloid genomes may therefore owe more to a long-term lack of meiotic sex than to anhydrobiosis.

We also quantified the abundance of TEs and low-complexity repeats in each animal genome. We chose to focus on the quantification of known repeats and thus did not perform ab initio repeat modelling for non-bdelloid species. There was considerable variation in TE abundance among species (Fig 7E), with the total proportion of genome covered by interspersed repeats varying from 0.3% in the tardigrade *R. varieornatus* to 27.5% in the oyster *Crassostrea gigas* (S3 Data). The relative abundance of different classes of repeats, including long interspersed elements (LINEs), short interspersed elements (SINEs), long terminal repeats (LTRs), and DNA elements, also differed greatly among taxa, as did the amount of simple and low-complexity repeats. The proportion of total repeats (TEs plus low-complexity repeats) ranged from 0.6% in *R. varieornatus* to 42% in the annelid worm *Helobdella robusta*. All 4 bdelloid species display a low abundance of TEs, in agreement with previous findings

[23–25,28,97]. However, 2 other species also show low levels of TEs: *I. linei*, an intracellular parasite of marine invertebrates with a highly reduced genome (42 Mb) [85], and *R. varieornatus*, also with a relatively small genome (56 Mb) [34]. In fact, *R. varieornatus* encodes the fewest TEs of the species analysed here (0.6% as a proportion of assembly span), followed by the 4 bdelloids (2%–3%). These estimates of TE abundances in *I. linei* and *R. varieornatus* are substantially lower than the total repeat content of these genomes (28% and 20%, respectively), which includes a high proportion of ab initio repeats (inferred directly from the assembled nucleotides) marked as 'unclassified' (accounting for approximately 18% and approximately 19% total repeats, respectively [79,85]), matching our finding of higher ab initio repeat content in bdelloids. Additional work is required to elucidate the nature of these unclassified repeats in bdelloids and in other taxa.

What evolutionary forces may explain the low abundance of TEs in these species? Asexuality and anhydrobiosis have both previously been posited as factors contributing to the low number of TEs in bdelloid rotifers. For example, under long-term asexual evolution, TEs may proliferate freely within a genome and thus drive that lineage to extinction (an extension of Muller's ratchet) or become lost, domesticated, or otherwise silenced [30,98–100]. Frequent cycles of desiccation and rehydration may also favour the evolution of reduced repeat content, via selection against deleterious chromosomal rearrangements brought about by ectopic recombination of TEs during the repair of DSBs [29,30].

Our comparisons did not detect any substantial variation in the abundance of known TEs between desiccating (1.2% and 0.8% for *A. ricciae* and *A. vaga*, respectively) and nondesiccating (0.9% and 1.2% for *R. macrura* and *R. magnacalcarata*, respectively) species, despite a considerable increase in the inferred genome size of *Rotaria* species. Moreover, the proposed mechanism involving desiccation relies on DSB repair during rehydration, a process which is presumably limited in the aquatic species *R. macrura* and *R. magnacalcarata* and may also not apply in the case of *R. varieornatus*, whose DNA is protected during anhydrobiosis [34]. However, the vast majority of bdelloid rotifers are resistant to desiccation, suggesting that anhydrobiosis was probably the ancestral state [17]. Therefore, it may be that TEs and other repeats were already largely eradicated in the most recent common ancestor to nondesiccating *Rotaria* species, prior to their adaptation to a fully aquatic lifestyle and loss of anhydrobiosis.

Finally, we also tested for the presence of a suite of 41 sex-related genes [101] in bdelloids using both TBLASTN (comparing to the genome) and HMMER (comparing to the proteome). Tested genes included 11 associated with meiosis, 19 involved in recombinational repair, 6 involved in DNA damage detection, 4 involved in DSB repair via nonhomologous end-joining, and 1 involved in bouquet formation (S8 Data). A positive match using TBLASTN and/or HMMER was recorded in at least 1 bdelloid species for all tested genes (40 of 41; 98%) with the exception of *RED1*, which is involved in crossover regulation and was not detected in any bdelloid at any significance threshold (S8 Data). However, *RED1* was not detected in *D. melanogaster* and only as a poor match in *C. elegans* and thus may represent an ancestral loss that predates the bdelloids. Overall, these findings suggest that bdelloids do encode the majority of genes involved in meiosis and sex-related functions. However, the presence of these genes does not necessarily indicate the presence of sex or meiosis because they are likely to be retained for other functions related to homologous recombination and DSB repair [102].

Concluding remarks

The bdelloid rotifers have drawn attention because 2 features of their life history are remarkable among metazoans: their apparent ancient asexuality and their ability to withstand desiccation at any life stage. In this work, we have generated whole-genome sequence data for 3

additional bdelloid species with the overall aim of assessing hypotheses regarding the contributions of asexuality and anhydrobiosis to their genome evolution.

We find that both desiccating and nondesiccating species are ancestrally tetraploid, in agreement with previous work, but that homologous divergence in nondesiccating *Rotaria* species is substantially lower than that observed in anhydrobiotic *Adineta* species and may be low even compared to estimates of allelic heterozygosity from sexual eukaryotes. This finding runs counter to predictions based on current hypotheses regarding the genomic effects of desiccation and thus requires a reevaluation of the causes and consequences of intragenomic interactions between bdelloid homologs. Comparisons of genome architecture revealed that a number of unusual genome features posited as evidence of long-term ameiotic evolution in *A. vaga* were largely absent from the closely related species *A. ricciae*, for which a comparable assembly is now available. In addition, we find that bdelloids encode the majority of genes that are required for meiosis and syngamy in sexual taxa but emphasise that the precise function of these genes in bdelloids is currently unknown.

We reconfirm previous reports that bdelloids encode a high proportion of nonmetazoan genes. Here too, a role for desiccation tolerance had been hypothesised. We find that high HGT content is a potentially unique feature of bdelloid genomes among animals, but comparisons to other desiccation-tolerant taxa raise questions about the role of anhydrobiosis. Our extensive assembly results also allow for a refinement of the global parameters of bdelloid genomes and suggest substantial genome size differences between genera. The phylogenetic nonindependence of our comparative analysis currently precludes any certainty in linking these observed trends to desiccation tolerance. Further elucidation will be possible when data for anhydrobiotic species within *Rotaria* become available in the future. Overall, we conclude that many features of the bdelloid genomes analysed here are not markedly inconsistent with those found in sexual taxa, except for the remarkably high prevalence of HGT.

Finally, we hope that our approach may offer useful guidance for future studies involving the de novo assembly of non–model organisms with complicated genome characteristics from complex raw data. Our goal was to explore the assembly parameter space for each dataset, taking into consideration a number of potential confounding factors including polyploidy, intragenomic divergence, and sample polymorphisms. Our assembly results showed good contiguity and gene-completeness metrics, indicating a high level of overall quality. Nonetheless, we reiterate the caution that a full understanding of genome architecture and evolution in bdelloid rotifers will be possible only with highly contiguous, chromosome-level assemblies, towards which future efforts will be directed.

Materials and methods

Rotifer culture and sampling

Clonal cultures of *A. ricciae* [103] rotifers were grown as previously described [27,104–106]. Briefly, rotifers were grown in T75 tissue culture flasks (Nunc) with 15 to 25 ml ddH_2O and fed twice a week with 10 μl of either bacteria (*Escherichia coli* TOP10 [ThermoFisher] in water) or a solution of yeast extract and peptone (2.5% w/v each). Approximately 50,000 rotifers were starved overnight before collection and harvested by centrifugation at 10,000 g for 5 minutes before treatment according to the relevant DNA or RNA extraction protocol. A starter culture for *R. macrura* was generated from approximately 100 wild-caught animals isolated from a small pond near Lake Orta, Italy. Populations were grown in sterile distilled water and fed with autoclaved and filter-sterilised organic lettuce extract. Prior to DNA extraction, animals were washed twice in sterile distilled water and starved overnight (approximately 16 hours) before being washed again with HyPure molecular-grade water. Genomic DNA from

approximately 420 animals (260 derived from a single founding animal; the remainder derived from a subpopulation of approximately 10 wild-caught founders) was extracted using the DNeasy Blood & Tissue kit (Qiagen) following the standard protocol. DNA was extracted in batches and pooled to generate sufficient material. Paired-end data for *R. magnacalcarata* have been described previously [41]. Both *R. macrura* and *R. magnacalcarata* PE libraries are derived from multiple individual samples. For mate-pair library construction for both *R. macrura* and *R. magnacalcarata*, DNA was extracted from a single individual and subjected to WGA using the Repli-G Single Cell kit (Qiagen), following the manufacturer's protocol. DNA concentration and quality were ascertained using a Qubit (Invitrogen) and a NanoDrop spectrophotometer (Thermo Scientific). Desiccation tolerance of *R. macrura* and *R. magnacalcarata* were tested using protocols as previously described [17,41] (see S3 Text for further details).

Sequencing

For *A. ricciae*, a short-insert library with an insert size of 250 bp was prepared using Illumina Nextera reagents and sequenced (100 bases paired-end) on an Illumina HiSeq 2000 at the Eastern Sequence and Informatics Hub (Cambridge, UK). Two long-insert (mate-pair) libraries both with inserts of 3 kb were also sequenced (51 bases paired-end) at GATC Biotech (London, UK). In addition, a PacBio (Pacific Biosciences) long-read library with an insert of 10 kb was sequenced using 3 SMRT Cells on a PacBio RS II (The Genome Analysis Centre, Norwich, UK). An RNASeq library with an insert size of 250 bp was sequenced (150 bases paired-end) on an Illumina NextSeq500 at the Department of Biochemistry, University of Cambridge (Cambridge, UK). A short-insert library (500-bp insert) for *R. macrura* was prepared using Illumina TruSeq reagents at the Centre for Genomic Research (CGR) at the University of Liverpool (Liverpool, UK). Mate-pair libraries with 2-kb inserts were also prepared at CGR using Nextera reagents, and all libraries were sequenced (150 bases paired-end) over 3 lanes of an Illumina HiSeq4000 at CGR. Short-insert data for *R. magnacalcarata* have been described previously [41]. All raw data have been submitted to the Sequence Read Archive (SRA), an International Nucleotide Sequence Database Collaboration (INSDC), under the accession IDs ERR2135445–55 (S1 Table).

Genome assembly

For *A. ricciae*, *R. macrura*, and *R. magnacalcarata* data, adapter sequences and low-quality bases were removed from Illumina data using Skewer v0.2.2 [107], and data quality was manually assessed using FastQC v0.11.5 [108]. Genome coverage was estimated by generating kmer distributions using BBMap 'kmercountexact' v36.02 [109], and library insert sizes—along with initial genome size estimates—were calculated using SGA 'preqc' [110]. Error correction of reads was performed using BBMap 'tadpole' ($k = 31$), discarding any pairs of reads containing unique kmers.

Contaminant reads derived from non–target organisms were filtered using BlobTools v0.9.19 [111]. Briefly, trimmed and error-corrected paired-end data were digitally normalised to approximately 100x using BBMap 'bbnorm' [109], and a preliminary draft assembly was generated using Velvet v1.2.10 [112], setting a kmer length of 75. Taxonomic annotations for all contigs were determined by comparing contigs against the NCBI nucleotide database (nt) and a custom database containing recently published whole-genome sequences of metazoans within the Spiralia (Lophotrochzoa) group (S4 Table) using BLAST 'megablast' (*E*-value $\leq 1 \times 10^{-25}$) [113], and the UniRef90 database using Diamond 'blastx' [114]. Finally, read coverage for each contig was estimated by mapping non-normalised reads to each draft assembly using BWA 'mem' v0.7.12 [115]. Taxon-annotated GC coverage plots ('blobplots') [111,116]

were generated using BlobTools (default parameters) and inspected manually. Putative contaminant sequences were identified as contigs showing atypical GC content, read coverage, and/or taxonomic classification. Given the a priori expectation that a substantial number of bdelloid genes may derive from nonmetazoan sources, we did not exclude any contigs based on taxonomy alone. Paired reads were excluded from further analysis only if both mapped to an identified contaminant contig or if one of the pair mapped to a contaminant while the other was unmapped. Additional rounds of filtering were performed if previously unassembled contaminant sequences became evident upon reassembly.

Filtered reads were assembled into contigs using the Platanus assembler v1.2.4 [42] with default parameters. Mate-pair libraries were filtered to remove contaminating FR-orientated reads (i.e., reads originating from short fragments) by excluding reads that mapped within $\leq$500 bases from the terminus of a contig. Contigs were scaffolded using SSPACE v3.0 [117], and undetermined bases were filled using GapFiller v1.10 [118]. The *A. ricciae* assembly was further scaffolded with the PacBio library using SSPACE-LongRead v1.1 [119]. RNASeq reads for *A. ricciae* were assembled de novo using Trinity v2.2.0 [120] (default parameters) and used for additional scaffolding with L_RNA_Scaffolder [121] and SCUBAT v2 [122]. An available transcriptome for *R. magnacalcarata* [41] was similarly utilised. A final round of assembly 'polishing' was performed using Redundans v0.12b [43], and scaffolds less than 200 bases in length were discarded. Assembly completeness was evaluated using the CEGMA v2.5 [44] and BUSCO v3.0.0 [45] gene sets, choosing the Eukaryota (n = 303) and Metazoa (n = 978) databases in the latter case and increasing the search limit to 8. Alternative assemblies were also generated using Velvet [112], SPAdes [123], and dipSPAdes [124] for comparison.

The reference assembly pipeline above was designed to maximise assembly contiguity but may lead to assembly collapse, the extent of which is undetermined a priori. Therefore, maximum haplotype assemblies were also generated for each species for comparison, defined as assemblies with minimal reduction due to assembly collapse. Maximum haplotype assemblies were generated using either dipSPAdes (default settings) or Platanus with the 'bubble crush' reduction parameter set to 0. Details of assembly parameters trialled are given in S1 Data. Collapsed and maximum haplotype (re)assemblies for *A. vaga* were also generated following the same procedures, using Illumina short-insert libraries (accession IDs SRR801084 and ERR321927) for contig building as well as mate-pair (accession ID ERR321928) and 454 data for scaffolding (see [28] for details).

Gene prediction

Repetitive regions were masked prior to gene prediction. Repeats were modelled ab initio using RepeatModeler v1.0.5 [125]. Repeats arising from duplicated genes or recent gene family expansions (e.g., alpha-tubulin in *R. magnacalcarata* [126]) were removed from the custom repeat library by comparing each repeat library to the SwissProt database (BLASTX, *E*-value $\leq 1 \times 10^{-5}$) and retaining only those sequences with descriptions for known repeat elements. The filtered RepeatModeler library was merged with known Rotifera repeats from Repbase v22.02 [127] (accessed using the command 'queryRepeatDatabase.pl -species 'rotifera') and compared to each assembly using RepeatMasker v4.0.7 [128]. Low-complexity regions and simple repeats were additionally soft-masked.

Gene prediction was then performed using BRAKER v1.9 [47] where RNASeq data was available (*A. ricciae* and *R. magnacalcarata*). Briefly, RNASeq reads were aligned to the masked assembly using STAR, specifying the 'twoPassMode Basic' parameter to improve splice junction annotation. The resultant alignment BAM file was then input to the BRAKER pipeline with default settings. For *R. macrura*, an initial set of gene models was constructed using

MAKER v3.00 [48], using evidence from SNAP [129] and GeneMark-ES v4.3 [130]. MAKER-derived gene models were then passed to Augustus v3.2.1 [49] for final refinement. Transfer and ribosomal RNA genes were predicted using tRNAscan-SE v1.3.1 [131] and RNAmmer [132], respectively (S9 Data). The *A. vaga* 2013 assembly (GCA_000513175.1) was also reannotated for consistency with these results, using both approaches outlined above (in conjunction with RNASeq library accession ERR260376).

To test if CDSs had been inadvertently missed during gene prediction, we compared proteins to the source nucleotide sequences from which they had been predicted using TBLASTN (*E*-value $\leq 1 \times 10^{-20}$). Matches to existing gene models were discounted by removing alignments that showed any overlap with gene regions (BEDtools 'intersect' [133] with the '-v' option), leaving only hits to regions of the genome that had not already been annotated as a gene.

Collinearity analyses

Syntenic regions within and between genomes were identified using MCScanX [50], calling collinear 'blocks' regions with at least 5 homologous genes and fewer than 10 'gaps' (i.e., missing genes). Rates of synonymous (K_S) and nonsynonymous (K_A) substitution between pairs of collinear genes were estimated by aligning proteins with Clustal Omega [134] and back-translating to nucleotides before calculating K_A and K_S values using BioPerl [135]. The collinearity of each block was calculated by dividing the number of collinear genes in a block by the total number of genes in the same region [28]. We also counted the number of collinearity breakpoints between adjacent homologous blocks across each genome, defining a breakpoint as an occurrence in which homologous blocks cannot be aligned without rearrangement. Collinearity plots were generated using the Circos software [136] in conjunction with the circosviz.pl program from the mmgenome toolkit [137]. Collinearity analysis scripts are available at https://github.com/reubwn/collinearity.

Orthologous clustering and SNP finding

Orthologous relationships among proteins from the same set of protostomes as above were inferred using OrthoFinder v1.1.4 [138] with default settings. All genomic, GFF, and protein sequence datasets were downloaded from NCBI GenBank no later than May 2017. For SNP finding, data were mapped using Bowtie2 v2.2.6 [139] with the '--very-sensitive' preset to minimise mismapped reads, and SNPs and indels were called using Platypus v0.8.1 [140], setting a minimum mapping quality of 30, a minimum base quality of 20, filter duplicates to 1, and a minimum read depth to approximately 25% of the average coverage of each individual library. VCF manipulation and SNP statistics were calculated using VCFlib v1.0.0-rc1 [141]. For *A. vaga*, SNPs were called based on the Illumina dataset ERR321927 mapped to the published genome sequence [28]. For *R. macrura* and *R. magnacalcarata*, SNPs were called based on WGA mate-pair libraries mapped as single-end because paired-end data for these samples were composed of multiple nonclonal lineages.

HGT analyses

We assessed the extent of horizontal transfer into bdelloid genomes using a combination of sequence comparison and phylogenetics-based approaches and applied the same tests to a set of 13 publicly available proteomes from species across the Protostomia (S4 Table). Protein sequences were first compared to the UniRef90 database [142] (downloaded November 29, 2016) using Diamond 'blastp' [114] (*E*-value $\leq 1 \times 10^{-5}$; maximum target sequences = 500). To avoid potential bias from bdelloid sequences already submitted to GenBank, all hits to the

phylum Rotifera (NCBI taxonomy ID 10190) were omitted from further analysis. For each query, 2 HGT metrics were then calculated: (1) HGT Index (h_U [27]), defined as $B_{OUT} - B_{IN}$, where B_{IN} is the best (highest) Diamond bitscore from comparisons to 'ingroup' taxa and B_{OUT} is the corresponding score for hits to 'outgroup' taxa; and (2) Consensus Hit Support (CHS), defined as the proportion of all hits that support a given query's ingroup/outgroup classification, itself inferred from the highest sum of bitscores to ingroup or outgroup across all hits [94]. The CHS score therefore takes into account the taxonomic distribution of all hits for each query and militates against misclassifications based on h_U scores alone. We defined the ingroup as 'Metazoa' and the outgroup as 'non-Metazoa' and marked all proteins with an $h_U \geq 30$ and $CHS_{OUT} \geq 90\%$ as putative HGT_C. We then looked at the distribution of all HGT_C across the genome and discarded any candidate found on a scaffold encoding ≥95% of genes of putative foreign origin (i.e., 'HGT-heavy' scaffolds that may be derived from contaminant sequences that were not removed during assembly). For each HGT_C, physical linkage (i.e., presence on the same scaffold) to a gene with good evidence for metazoan origin ($h_U \leq 0$, $CHS_{IN} \geq 90\%$) and the number of predicted introns were also recorded. Finally, phylogenetic support for HGT was then assessed: for each HGT_C, the sequences of 15 metazoan and 15 non-metazoan UniRef90 hits (when present) were extracted and aligned using MAFFT v7.309 [143] with default parameters, and a maximum likelihood phylogeny was constructed using IQ-TREE v1.5.3 [144], specifying automatic model selection and 1,000 ultrafast bootstrap replicates. The functionality of GNU Parallel [145] was used to compute multiple trees simultaneously, and clusters with fewer than 4 taxa were not analysed. Branching patterns of resultant trees were then assessed using a custom script written in R v3.3.1 [146], utilising functions from the 'ape' v4.1 package [147]. HGT analysis scripts are available at https://github.com/reubwn/hgt.

TE analyses

The abundance of known TEs was assessed for the same set of protostomes using RepeatMasker, except using a Repbase (v22.02) repeat library specific to the Metazoa (i.e., 'queryRepeatDatabase.pl -species 'metazoa"). Custom species-specific repeat libraries (e.g., using RepeatModeler) were not generated for this analysis; only known repeats from Repbase were compared. The total span of LINEs/SINEs, LTR elements, DNA elements, and simple repeats relative to the assembly span for each species was then computed from the RepeatMasker results. We also calculated a genome density metric, defined as the number of protein-coding genes per Mb of haploid genome, i.e., accounting for variation in ploidy among species.

Meiosis genes

The presence of meiosis- and other sex-related genes was assessed following the approach of Tekle et al. [101]. A total of 41 orthologous groups were downloaded from the OrthoMCL database (v5) (http://orthomcl.org/orthomcl/; accessed September 2017) (S8 Data). Searches were conducted using both TBLASTN (E-value $\leq 1 \times 10^{-5}$) against the reference assemblies or HMMER3 (http://hmmer.org/) against the predicted protein sets, after alignment with Clustal Omega [134]. Presence was recorded if any query within each orthologous group showed a TBLASTN alignment with ≥50% identity over ≥50% query length and/or if HMMER reported an alignment above the default significance threshold. Multiple hits to the same location (caused by paralogy or hits to similar domains) were recorded if top hits overlapped among queries. The genomes and proteomes of *D. melanogaster* and *C. elegans* were also searched for comparison.

Supporting information

S1 Data. Assembly metrics for reference and maximum haplotype assemblies.
(XLSX)

S2 Data. Augustus GFF and protein FASTA files (filtered).
(TGZ)

S3 Data. Repeat content.
(XLSX)

S4 Data. MCScanX and collinearity analyses.
(TGZ)

S5 Data. SNP information. SNP, single-nucleotide polymorphism.
(XLSX)

S6 Data. Platypus VCF files.
(TGZ)

S7 Data. HGT metrics. HGT, horizontal gene transfer.
(XLSX)

S8 Data. Meiosis- and sex-related gene inventory.
(XLSX)

S9 Data. Transfer and ribosomal RNA gene annotation.
(TGZ)

S1 Fig. Kmer spectra for raw and filtered sequence data.
(PDF)

S2 Fig. Taxon-annotated GC coverage plots for initial and final assemblies. GC, guanine-cytosine.
(PDF)

S3 Fig. Effect of bubble crush on Platanus assembly.
(PDF)

S4 Fig. Effect of Redundans reduction.
(PDF)

S5 Fig. Distribution of intergenic distances between genes.
(PDF)

S6 Fig. Distribution of intron length.
(PDF)

S7 Fig. MAF distributions for paired-end *Rotaria* sequence data. MAF, minor allele frequency.
(PDF)

S8 Fig. Coverage profiles for reference and alternative assemblies for *A. vaga*.
(PDF)

S9 Fig. Coverage profiles for reference and alternative *A. ricciae* assemblies.
(PDF)

S10 Fig. Coverage profile and SNP density distribution for *A. ricciae* homologous genes. SNP, single-nucleotide polymorphism.
(PDF)

S11 Fig. Genome size estimates based on kmer spectra.
(PDF)

S12 Fig. Orthologous clustering within bdelloid genomes.
(PDF)

S13 Fig. Evidence for ancient origin for most HGT_C genes in bdelloids. HGT_C, horizontal gene transfer candidates.
(PDF)

S1 Table. Data counts and accession numbers for sequence data used in this study.
(PDF)

S2 Table. MCScanX collinearity metrics within genomes.
(PDF)

S3 Table. Homologous and ohnologous K_A and K_S.
(PDF)

S4 Table. Protostome species included in comparative analysis.
(PDF)

S5 Table. OrthoFinder clustering metrics.
(PDF)

S6 Table. OrthoFinder clustering metrics per species.
(PDF)

S1 Text. Signals of duplication and collapse in the *A. ricciae* sequence data.
(PDF)

S2 Text. Note on bdelloid genome size estimates.
(PDF)

S3 Text. Test of desiccation tolerance for *Rotaria* species.
(PDF)

Acknowledgments

The authors wish to thank the laboratory of Mark Blaxter, Georgios Koutsovoulos, David Mark Welch, Matthew Meselson, Stephen Richards, and Philipp Schiffer for comments on a previous version of this manuscript. We also thank Shilo Dickens at the Department of Biochemistry at the University of Cambridge and the Babraham Institute for help with additional Illumina sequencing, and Michael Plewka for bdelloid photography. PacBio next-generation sequencing and library construction were delivered via the BBSRC National Capability in Genomics (BB/CCG1720/1) at Earlham Institute (previously The Genome Analysis Centre) by members of the Genomics Pipelines Group.

Author Contributions

Conceptualization: Alan Tunnacliffe, Timothy G. Barraclough.

Data curation: Reuben W. Nowell, Pedro Almeida, Timothy G. Barraclough.

Formal analysis: Reuben W. Nowell, Christopher G. Wilson, Timothy G. Barraclough.

Funding acquisition: Alan Tunnacliffe, Timothy G. Barraclough.

Investigation: Reuben W. Nowell, Pedro Almeida, Christopher G. Wilson, Thomas P. Smith, Diego Fontaneto, Alastair Crisp, Gos Micklem, Alan Tunnacliffe, Chiara Boschetti, Timothy G. Barraclough.

Methodology: Reuben W. Nowell, Pedro Almeida, Christopher G. Wilson, Alan Tunnacliffe, Chiara Boschetti, Timothy G. Barraclough.

Project administration: Alan Tunnacliffe, Timothy G. Barraclough.

Resources: Christopher G. Wilson, Diego Fontaneto, Alan Tunnacliffe, Chiara Boschetti, Timothy G. Barraclough.

Software: Reuben W. Nowell, Timothy G. Barraclough.

Supervision: Alan Tunnacliffe, Chiara Boschetti, Timothy G. Barraclough.

Validation: Reuben W. Nowell, Timothy G. Barraclough.

Visualization: Reuben W. Nowell.

Writing – original draft: Reuben W. Nowell, Timothy G. Barraclough.

Writing – review & editing: Reuben W. Nowell, Pedro Almeida, Christopher G. Wilson, Thomas P. Smith, Diego Fontaneto, Alastair Crisp, Gos Micklem, Alan Tunnacliffe, Chiara Boschetti, Timothy G. Barraclough.

References

1. Hudson CT, Gosse PH. The Rotifera or wheel-animalcules. Longmans, Green; 1886.
2. Hsu WS. Oogenesis in the Bdelloidea rotifer *Philodina roseola* Ehrenberg. Cellule. 1956; 57: 283–296.
3. Hsu WS. Oogenesis in *Habrotrocha tridens* (Milne). Biol Bull. 1956; 111: 364–374.
4. Mark Welch D, Meselson M. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science. 2000; 288: 1211–1215. PMID: 10817991
5. Mark Welch DB, Ricci C, Meselson M. Bdelloid Rotifers: Progress in understanding the success of an evolutionary scandal. Lost Sex. Springer, Dordrecht; 2009. pp. 259–279.
6. Maynard Smith J. Evolution: contemplating life without sex. Nature. 1986; 324: 300–301. PMID: 3785401
7. Segers H. Annotated checklist of the rotifers (Phylum Rotifera), with notes on nomenclature, taxonomy and distribution. Zootaxa. 2007; 1564: 1–104.
8. Robeson MS, King AJ, Freeman KR, Birky CW Jr, Martin AP, Schmidt SK. Soil rotifer communities are extremely diverse globally but spatially autocorrelated locally. Proc Natl Acad Sci U S A. 2011; 108: 4406–4410. https://doi.org/10.1073/pnas.1012678108 PMID: 21368117
9. Burt A. Perspective: Sex, recombination, and the efficacy of selection—was Weismann right? Evolution. 2000; 54: 337–351. PMID: 10937212
10. Simon J-C, Delmotte F, Rispe C, Crease T. Phylogenetic relationships between parthenogens and their sexual relatives: the possible routes to parthenogenesis in animals. Biol J Linn Soc Lond. 2003; 79: 151–163.
11. Hartfield M, Keightley PD. Current hypotheses for the evolution of sex and recombination. Integr Zool. 2012; 7: 192–209. https://doi.org/10.1111/j.1749-4877.2012.00284.x PMID: 22691203
12. Williams GC. Sex and Evolution. Princeton University Press, Princeton; 1975.
13. Maynard Smith J. The Evolution of Sex. Cambridge University Press, Cambridge, UK; 1978.
14. Bell G. The Masterpiece of Nature: The Evolution and Genetics of Sexuality. University of California Press, Berkeley; 1982.

15. Otto SP. The evolutionary enigma of sex. Am Nat. 2009; 174 Suppl 1: S1–S14.
16. Ricci CN. Ecology of bdelloids: how to be successful. Hydrobiologia. 1987; 147: 117–127.
17. Ricci C. Anhydrobiotic capabilities of bdelloid rotifers. Hydrobiologia. 1998; 387–388: 321–326.
18. Marotta R, Leasi F, Uggetti A, Ricci C, Melone G. Dry and survive: morphological changes during anhydrobiosis in a bdelloid rotifer. J Struct Biol. 2010; 171: 11–17. https://doi.org/10.1016/j.jsb.2010.04.003 PMID: 20382230
19. Ricci C, Caprioli M. Anhydrobiosis in bdelloid species, populations and individuals. Integr Comp Biol. 2005; 45: 759–763. https://doi.org/10.1093/icb/45.5.759 PMID: 21676827
20. Fontaneto D, Bunnefeld N, Westberg M. Long-term survival of microscopic animals under desiccation is not so long. Astrobiology. 2012; 12: 863–869. https://doi.org/10.1089/ast.2012.0828 PMID: 22924877
21. Ricci C, Covino C. Anhydrobiosis of *Adineta ricciae*: costs and benefits. Hydrobiologia. 2005; 546: 307–314.
22. Wilson CG, Sherman PW. Anciently asexual bdelloid rotifers escape lethal fungal parasites by drying up and blowing away. Science. 2010; 327: 574–576. https://doi.org/10.1126/science.1179252 PMID: 20110504
23. Arkhipova I, Meselson M. Transposable elements in sexual and ancient asexual taxa. Proc Natl Acad Sci U S A. 2000; 97: 14473–14477. https://doi.org/10.1073/pnas.97.26.14473 PMID: 11121049
24. Mark Welch DB, Mark Welch JL, Meselson M. Evidence for degenerate tetraploidy in bdelloid rotifers. Proc Natl Acad Sci U S A. 2008; 105: 5145–5149. https://doi.org/10.1073/pnas.0800972105 PMID: 18362354
25. Hur JH, Van Doninck K, Mandigo ML, Meselson M. Degenerate tetraploidy was established before bdelloid rotifer families diverged. Mol Biol Evol. 2009; 26: 375–383. https://doi.org/10.1093/molbev/msn260 PMID: 18996928
26. Gladyshev EA, Meselson M, Arkhipova IR. Massive horizontal gene transfer in bdelloid rotifers. Science. 2008; 320: 1210–1213. https://doi.org/10.1126/science.1156407 PMID: 18511688
27. Boschetti C, Carr A, Crisp A, Eyres I, Wang-Koh Y, Lubzens E, et al. Biochemical diversification through foreign gene expression in bdelloid rotifers. PLoS Genet. 2012; 8: e1003035. https://doi.org/10.1371/journal.pgen.1003035 PMID: 23166508
28. Flot J-F, Hespeels B, Li X, Noel B, Arkhipova I, Danchin EGJ, et al. Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. Nature. 2013; 500: 453–457. https://doi.org/10.1038/nature12326 PMID: 23873043
29. Gladyshev E, Meselson M. Extreme resistance of bdelloid rotifers to ionizing radiation. Proc Natl Acad Sci U S A. 2008; 105: 5139–5144. https://doi.org/10.1073/pnas.0800966105 PMID: 18362355
30. Gladyshev EA, Arkhipova IR. Genome structure of bdelloid rotifers: shaped by asexuality or desiccation? J Hered. 2010; 101 Suppl 1: S85–93.
31. Hespeels B, Knapen M, Hanot-Mambres D, Heuskin A-C, Pineux F, Lucas S, et al. Gateway to genetic exchange? DNA double-strand breaks in the bdelloid rotifer *Adineta vaga* submitted to desiccation. J Evol Biol. 2014; 27: 1334–1345. PMID: 25105197
32. Neumann S, Reuner A, Brümmer F, Schill RO. DNA damage in storage cells of anhydrobiotic tardigrades. Comp Biochem Physiol A Mol Integr Physiol. 2009; 153: 425–429. https://doi.org/10.1016/j.cbpa.2009.04.611 PMID: 19361569
33. Rebecchi L, Cesari M, Altiero T, Frigieri A, Guidetti R. Survival and DNA degradation in anhydrobiotic tardigrades. J Exp Biol. 2009; 212: 4033–4039. https://doi.org/10.1242/jeb.033266 PMID: 19946082
34. Hashimoto T, Horikawa DD, Saito Y, Kuwahara H, Kozuka-Hata H, Shin-I T, et al. Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. Nat Commun. 2016; 7: 12808. https://doi.org/10.1038/ncomms12808 PMID: 27649274
35. Gusev O, Nakahara Y, Vanyagina V, Malutina L, Cornette R, Sakashita T, et al. Anhydrobiosis-associated nuclear DNA damage and repair in the sleeping chironomid: linkage with radioresistance. PLoS ONE. 2010; 5: e14008. https://doi.org/10.1371/journal.pone.0014008 PMID: 21103355
36. Debortoli N, Li X, Eyres I, Fontaneto D, Hespeels B, Tang CQ, et al. Genetic exchange among bdelloid rotifers is more likely due to horizontal gene transfer than to meiotic sex. Curr Biol. 2016; 26: 723–732. https://doi.org/10.1016/j.cub.2016.01.031 PMID: 26948882
37. Wilson CG, Nowell RW, Barraclough TG. Evidence for "inter- and intraspecific horizontal genetic transfers" between anciently asexual bdelloid rotifers is explained by cross-contamination. bioRxiv 150490 [Preprint]. 2017 [Cited 2017 May 1]. Available from: https://www.biorxiv.org/content/early/2017/09/12/150490. https://doi.org/10.1101/150490

38. Signorovitch A, Hur J, Gladyshev E, Meselson M. Allele sharing and evidence for sexuality in a mitochondrial clade of bdelloid rotifers. Genetics. 2015; 200: 581–590. https://doi.org/10.1534/genetics.115.176719 PMID: 25977472

39. Signorovitch A, Hur J, Gladyshev E, Meselson M. Evidence for meiotic sex in bdelloid rotifers. Curr Biol. 2016; 26: R754–5. https://doi.org/10.1016/j.cub.2016.06.051 PMID: 27554650

40. Flot J-F, Debortoli N, Hallet B, Van Doninck K. Response to Signorovitch et al. Curr Biol. 2016; 26: R755. https://doi.org/10.1016/j.cub.2016.06.052 PMID: 27554651

41. Eyres I, Boschetti C, Crisp A, Smith TP, Fontaneto D, Tunnacliffe A, et al. Horizontal gene transfer in bdelloid rotifers is ancient, ongoing and more frequent in species from desiccating habitats. BMC Biol. 2015; 13: 90. https://doi.org/10.1186/s12915-015-0202-9 PMID: 26537913

42. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. 2014; 24: 1384–1395. https://doi.org/10.1101/gr.170720.113 PMID: 24755901

43. Pryszcz LP, Gabaldón T. Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Res. 2016; 44: e113. https://doi.org/10.1093/nar/gkw294 PMID: 27131372

44. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 2007; 23: 1061–1067. https://doi.org/10.1093/bioinformatics/btm071 PMID: 17332020

45. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015; 31: 3210–3212. https://doi.org/10.1093/bioinformatics/btv351 PMID: 26059717

46. Kent WJ. BLAT—The BLAST-Like Alignment Tool. Genome Res. 2002; 12: 656–664. https://doi.org/10.1101/gr.229202 PMID: 11932250

47. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. Bioinformatics. 2016; 32: 767–769. https://doi.org/10.1093/bioinformatics/btv661 PMID: 26559507

48. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 2011; 12: 491. https://doi.org/10.1186/1471-2105-12-491 PMID: 22192575

49. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 2008; 24: 637–644. https://doi.org/10.1093/bioinformatics/btn013 PMID: 18218656

50. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012; 40: e49. https://doi.org/10.1093/nar/gkr1293 PMID: 22217600

51. Lunt DH, Kumar S, Koutsovoulos G, Blaxter ML. The complex hybrid origins of the root knot nematodes revealed through comparative genomics. PeerJ. 2014; 2: e356. https://doi.org/10.7717/peerj.356 PMID: 24860695

52. Ament-Velásquez SL, Figuet E, Ballenghien M, Zattara EE, Norenburg JL, Fernández-Álvarez FA, et al. Population genomics of sexual and asexual lineages in fissiparous ribbon worms (Lineus, Nemertea): hybridization, polyploidy and the Meselson effect. Mol Ecol. 2016; 25: 3356–3369. https://doi.org/10.1111/mec.13717 PMID: 27286413

53. Schiffer PH, Danchin E, Burnell AM, Schiffer A-M, Creevey C, Wong S, et al. Signatures of the evolution of parthenogenesis and cryptobiosis in the genomes of panagrolaimid nematodes. bioRxiv 159152 [Preprint]. 2017 [Cited 2017 May 1]. Available from: https://www.biorxiv.org/content/early/2017/07/03/159152. https://doi.org/10.1101/159152

54. Fradin H, Kiontke K, Zegar C, Gutwein M, Lucas J, Kovtun M, et al. Genome architecture and evolution of a unichromosomal asexual nematode. Curr Biol. 2017; 27: 2928–2939.e6. https://doi.org/10.1016/j.cub.2017.08.038 PMID: 28943090

55. Szitenberg A, Salazar-Jaramillo L, Blok VC, Laetsch DR, Joseph S, Williamson VM, et al. Comparative genomics of apomictic root-knot nematodes: hybridization, ploidy, and dynamic genome change. Genome Biol Evol. 2017; 9: 2844–2861. https://doi.org/10.1093/gbe/evx201 PMID: 29036290

56. Mark Welch JL, Meselson M. Karyotypes of bdelloid rotifers from three families. Rotifera VIII: A Comparative Approach. Springer, Dordrecht; 1998. pp. 403–407.

57. Pouchkina-Stantcheva NN, McGee BM, Boschetti C, Tolleter D, Chakrabortee S, Popova AV, et al. Functional divergence of former alleles in an ancient asexual invertebrate. Science. 2007; 318: 268–271. https://doi.org/10.1126/science.1144363 PMID: 17932297

58. Jones PA, Gilbert JJ. Polymorphism and polyploidy in the rotifer *Asplanchna sieboldi*: Relative nuclear DNA contents in tissues of saccate and campanulate females. J Exp Zool. 1977; 201: 163–168.

59. Hildebrand F, Meyer A, Eyre-Walker A. Evidence of selection upon genomic GC-content in bacteria. PLoS Genet. 2010; 6: e1001107. https://doi.org/10.1371/journal.pgen.1001107 PMID: 20838593

60. Leffler EM, Bullaughey K, Matute DR, Meyer WK, Ségurel L, Venkat A, et al. Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biol. 2012; 10: e1001388. https://doi.org/10.1371/journal.pbio.1001388 PMID: 22984349

61. Stern C. Somatic Crossing over and segregation in *Drosophila melanogaster*. Genetics. 1936; 21: 625–730. PMID: 17246815

62. Tsang PW, Cao B, Siu PY, Wang J. Loss of heterozygosity, by mitotic gene conversion and crossing over, causes strain-specific adenine mutants in constitutive diploid *Candida albicans*. Microbiology. 1999; 145 (Pt 7): 1623–1629.

63. Ayaki T, Fujikawa K, Ryo H, Itoh T, Kondo S. Induced rates of mitotic crossing over and possible mitotic gene conversion per wing anlage cell in *Drosophila melanogaster* by X rays and fission neutrons. Genetics. 1990; 126: 157–166. PMID: 2121593

64. LaFave MC, Sekelsky J. Mitotic recombination: Why? When? How? Where? PLoS Genet. 2009; 5: e1000411. https://doi.org/10.1371/journal.pgen.1000411 PMID: 19282976

65. Lee PS, Greenwell PW, Dominska M, Gawel M, Hamilton M, Petes TD. A fine-structure map of spontaneous mitotic crossovers in the yeast *Saccharomyces cerevisiae*. PLoS Genet. 2009; 5: e1000410. https://doi.org/10.1371/journal.pgen.1000410 PMID: 19282969

66. Symington LS, Rothstein R, Lisby M. Mechanisms and regulation of mitotic recombination in *Saccharomyces cerevisiae*. Genetics. 2014; 198: 795–835. https://doi.org/10.1534/genetics.114.166140 PMID: 25381364

67. Tucker AE, Ackerman MS, Eads BD, Xu S, Lynch M. Population-genomic insights into the evolutionary origin and fate of obligately asexual *Daphnia pulex*. Proc Natl Acad Sci U S A. 2013; 110: 15740–15745. https://doi.org/10.1073/pnas.1313388110 PMID: 23959868

68. Pfeiffer P, Goedecke W, Obe G. Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations. Mutagenesis. 2000; 15: 289–302. PMID: 10887207

69. Birky CW Jr. Positively negative evidence for asexuality. J Hered. 2010; 101 Suppl 1: S42–5.

70. Chevreux B, Wetter T, Suhai S, Others. Genome sequence assembly using trace signals and additional sequence information. German conference on bioinformatics. Hanover, Germany; 1999. pp. 45–56.

71. Faddeeva-Vakhrusheva A, Kraaijeveld K, Derks MFL, Anvar SY, Agamennone V, Suring W, et al. Coping with living in the soil: the genome of the parthenogenetic springtail *Folsomia candida*. BMC Genomics. 2017; 18: 493. https://doi.org/10.1186/s12864-017-3852-x PMID: 28659179

72. Blanc-Mathieu R, Perfus-Barbeoch L, Aury J-M, Da Rocha M, Gouzy J, Sallet E, et al. Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. PLoS Genet. 2017; 13: e1006777. https://doi.org/10.1371/journal.pgen.1006777 PMID: 28594822

73. Hiraki H, Kagoshima H, Kraus C, Schiffer PH, Ueta Y, Kroiher M, et al. Genome analysis of *Diploscapter coronatus*: insights into molecular peculiarities of a nematode with parthenogenetic reproduction. BMC Genomics. 2017; 18: 478. https://doi.org/10.1186/s12864-017-3860-x PMID: 28646875

74. Bruder CEG, Piotrowski A, Gijsbers AACJ, Andersson R, Erickson S, Diaz de Ståhl T, et al. Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. Am J Hum Genet. 2008; 82: 763–771. https://doi.org/10.1016/j.ajhg.2007.12.011 PMID: 18304490

75. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, et al. Mapping copy number variation by population-scale genome sequencing. Nature. 2011; 470: 59–65. https://doi.org/10.1038/nature09708 PMID: 21293372

76. Fan S, Meyer A. Evolution of genomic structural variation and genomic architecture in the adaptive radiations of African cichlid fishes. Front Genet. 2014; 5: 163. https://doi.org/10.3389/fgene.2014.00163 PMID: 24917883

77. Chen L, Chamberlain AJ, Reich CM, Daetwyler HD, Hayes BJ. Detection and validation of structural variations in bovine whole-genome sequence data. Genet Sel Evol. 2017; 49: 13. https://doi.org/10.1186/s12711-017-0286-5 PMID: 28122487

78. Mark Welch DB, Meselson M. Oocyte nuclear DNA content and GC proportion in rotifers of the anciently asexual Class Bdelloidea. Biol J Linn Soc Lond. 2003; 79: 85–91.

79. Yoshida Y, Koutsovoulos G, Laetsch DR, Stevens L, Kumar S, Horikawa DD, et al. Comparative genomics of the tardigrades *Hypsibius dujardini* and *Ramazzottius varieornatus*. PLoS Biol. 2017; 15: e2002266. https://doi.org/10.1371/journal.pbio.2002266 PMID: 28749982

80. Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, et al. The oyster genome reveals stress adaptation and complexity of shell formation. Nature. 2012; 490: 49–54. https://doi.org/10.1038/nature11413 PMID: 22992520

81. Simakov O, Marletaz F, Cho S-J, Edsinger-Gonzales E, Havlak P, Hellsten U, et al. Insights into bilaterian evolution from three spiralian genomes. Nature. 2013; 493: 526–531. https://doi.org/10.1038/nature11696 PMID: 23254933

82. Albertin CB, Simakov O, Mitros T, Wang ZY, Pungor JR, Edsinger-Gonzales E, et al. The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature. 2015; 524: 220–224. https://doi.org/10.1038/nature14668 PMID: 26268193

83. Adema CM, Hillier LW, Jones CS, Loker ES, Knight M, Minx P, et al. Whole genome analysis of a schistosomiasis-transmitting freshwater snail. Nat Commun. 2017; 8: 15451. https://doi.org/10.1038/ncomms15451 PMID: 28508897

84. Young ND, Jex AR, Li B, Liu S, Yang L, Xiong Z, et al. Whole-genome sequence of *Schistosoma haematobium*. Nat Genet. 2012; 44: 221–225. https://doi.org/10.1038/ng.1065 PMID: 22246508

85. Mikhailov KV, Slyusarev GS, Nikitin MA, Logacheva MD, Penin AA, Aleoshin VV, et al. The genome of *Intoshia linei* affirms orthonectids as highly simplified spiralians. Curr Biol. 2016; 26: 1768–1774. https://doi.org/10.1016/j.cub.2016.05.007 PMID: 27374341

86. C. elegans Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science. 1998; 282: 2012–2018. PMID: 9851916

87. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, et al. The genome sequence of *Drosophila melanogaster*. Science. 2000; 287: 2185–2195. PMID: 10731132

88. Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, et al. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature. 2008; 452: 745–749. https://doi.org/10.1038/nature06614 PMID: 18322464

89. Laumer CE, Bekkouche N, Kerbl A, Goetz F, Neves RC, Sørensen MV, et al. Spiralian phylogeny informs the evolution of microscopic lineages. Curr Biol. 2015; 25: 2000–2006. https://doi.org/10.1016/j.cub.2015.06.068 PMID: 26212884

90. Kersey PJ, Allen JE, Allot A, Barba M, Boddu S, Bolt BJ, et al. Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. Nucleic Acids Res. 2018; 46: D802–D808. https://doi.org/10.1093/nar/gkx1011 PMID: 29092050

91. Arakawa K. No evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. Proc Natl Acad Sci U S A. 2016; 113: E3057. https://doi.org/10.1073/pnas.1602711113 PMID: 27173901

92. Bemm F, Weiß CL, Schultz J, Förster F. Genome of a tardigrade: Horizontal gene transfer or bacterial contamination? Proc Natl Acad Sci U S A. 2016; 113: E3054–6. https://doi.org/10.1073/pnas.1525116113 PMID: 27173902

93. Delmont TO, Eren AM. Identifying contamination with advanced visualization and analysis practices: metagenomic approaches for eukaryotic genome assemblies. PeerJ. 2016; 4: e1839. https://doi.org/10.7717/peerj.1839 PMID: 27069789

94. Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C, et al. No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*. Proc Natl Acad Sci U S A. 2016; 113: 5053–5058. https://doi.org/10.1073/pnas.1600338113 PMID: 27035985

95. Crisp A, Boschetti C, Perry M, Tunnacliffe A, Micklem G. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. Genome Biol. 2015; 16: 50. https://doi.org/10.1186/s13059-015-0607-3 PMID: 25785303

96. Gusev O, Suetsugu Y, Cornette R, Kawashima T, Logacheva MD, Kondrashov AS, et al. Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge. Nat Commun. 2014; 5: 4784. https://doi.org/10.1038/ncomms5784 PMID: 25216354

97. Arkhipova IR, Meselson M. Diverse DNA transposons in rotifers of the class Bdelloidea. Proc Natl Acad Sci U S A. 2005; 102: 11781–11786. https://doi.org/10.1073/pnas.0505333102 PMID: 16081532

98. Arkhipova I, Meselson M. Deleterious transposable elements and the extinction of asexuals. Bioessays. 2005; 27: 76–85. https://doi.org/10.1002/bies.20159 PMID: 15612027

99. Dolgin ES, Charlesworth B. The effects of recombination rate on the distribution and abundance of transposable elements. Genetics. 2008; 178: 2169–2177. https://doi.org/10.1534/genetics.107.082743 PMID: 18430942

100. Bast J, Schaefer I, Schwander T, Maraun M, Scheu S, Kraaijeveld K. No accumulation of transposable elements in asexual arthropods. Mol Biol Evol. 2016; 33: 697–706. https://doi.org/10.1093/molbev/msv261 PMID: 26560353

101. Tekle YI, Wood FC, Katz LA, Cerón-Romero MA, Gorfu LA. Amoebozoans are secretly but ancestrally sexual: evidence for sex genes and potential novel crossover pathways in diverse groups of amoebae. Genome Biol Evol. 2017; 9:375–387. https://doi.org/10.1093/gbe/evx002 PMID: 28087686

102. Hanson SJ, Schurko AM, Hecox-Lea B, Welch DBM, Stelzer C-P, Logsdon JM Jr. Inventory and phylogenetic analysis of meiotic genes in monogonont rotifers. J Hered. 2013; 104: 357–370. https://doi.org/10.1093/jhered/est011 PMID: 23487324

103. Segers H, Shiel RJ. Tale of a Sleeping Beauty: a new and easily cultured model organism for experimental studies on bdelloid rotifers. Hydrobiologia. 2005; 546: 141–145.

104. Boschetti C, Pouchkina-Stantcheva N, Hoffmann P, Tunnacliffe A. Foreign genes and novel hydrophilic protein genes participate in the desiccation response of the bdelloid rotifer *Adineta ricciae*. J Exp Biol. 2011; 214: 59–68. https://doi.org/10.1242/jeb.050328 PMID: 21147969

105. Szydlowski L, Boschetti C, Crisp A, Barbosa EGG, Tunnacliffe A. Multiple horizontally acquired genes from fungal and prokaryotic donors encode cellulolytic enzymes in the bdelloid rotifer *Adineta ricciae*. Gene. 2015; 566: 125–137. https://doi.org/10.1016/j.gene.2015.04.007 PMID: 25863176

106. Barbosa EGG, Crisp A, Broadbent SE, Carrillo M, Boschetti C, Tunnacliffe A. A functional difference between native and horizontally acquired genes in bdelloid rotifers. Gene. 2016; 590: 186–191. https://doi.org/10.1016/j.gene.2016.06.019 PMID: 27312952

107. Jiang H, Lei R, Ding S-W, Zhu S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics. 2014; 15: 182. https://doi.org/10.1186/1471-2105-15-182 PMID: 24925680

108. Andrews S. FastQC a quality-control tool for high-throughput sequence data. Version 0.11.5 [software]. 2015. [Cited 1 May 2017]. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

109. Bushnell B. BBMap. Version 36.02 [software]. 2016. [Cited 1 May 2017]. https://sourceforge.net/projects/bbmap/.

110. Simpson JT, Durbin R. Efficient construction of an assembly string graph using the FM-index. Bioinformatics. 2010; 26: i367–73. https://doi.org/10.1093/bioinformatics/btq217 PMID: 20529929

111. Laetsch DR, Blaxter ML. BlobTools: Interrogation of genome assemblies. F1000Res. 2017; 6. https://doi.org/10.12688/f1000research.12232.1

112. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008; 18: 821–829. https://doi.org/10.1101/gr.074492.107 PMID: 18349386

113. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25: 3389–3402. PMID: 9254694

114. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015; 12: 59–60. https://doi.org/10.1038/nmeth.3176 PMID: 25402007

115. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168

116. Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M. Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. Front Genet. 2013; 4: 237. https://doi.org/10.3389/fgene.2013.00237 PMID: 24348509

117. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011; 27: 578–579. https://doi.org/10.1093/bioinformatics/btq683 PMID: 21149342

118. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. Genome Biol. 2012; 13: R56. https://doi.org/10.1186/gb-2012-13-6-r56 PMID: 22731987

119. Boetzer M, Pirovano W. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinformatics. 2014; 15: 211. https://doi.org/10.1186/1471-2105-15-211 PMID: 24950923

120. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011; 29: 644–652. https://doi.org/10.1038/nbt.1883 PMID: 21572440

121. Xue W, Li J-T, Zhu Y-P, Hou G-Y, Kong X-F, Kuang Y-Y, et al. L_RNA_scaffolder: scaffolding genomes with transcripts. BMC Genomics. 2013; 14: 604. https://doi.org/10.1186/1471-2164-14-604 PMID: 24010822

122. Koutsovoulos G. SCUBAT. Version 2 [software]. 2016. [Cited 1 May 2017]. https://github.com/GDKO/SCUBAT2.

123. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012; 19: 455–477. https://doi.org/10.1089/cmb.2012.0021 PMID: 22506599

124. Safonova Y, Bankevich A, Pevzner PA. dipSPAdes: Assembler for Highly Polymorphic Diploid Genomes. J Comput Biol. 2015; 22: 528–545. https://doi.org/10.1089/cmb.2014.0153 PMID: 25734602

125. Smit AFA, Hubley R. RepeatModeler Open-1.0. Version 1.0 [software]. 2015. [Cited 1 May 2017]. http://www.repeatmasker.org.

126. Eyres I, Frangedakis E, Fontaneto D, Herniou EA, Boschetti C, Carr A, et al. Multiple functionally divergent and conserved copies of alpha tubulin in bdelloid rotifers. BMC Evol Biol. 2012; 12: 148. https://doi.org/10.1186/1471-2148-12-148 PMID: 22901238

127. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015; 6: 11. https://doi.org/10.1186/s13100-015-0041-9 PMID: 26045719

128. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. Version 4.0 [software]. 2015. [Cited 1 May 2017]. http://www.repeatmasker.org.

129. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004; 5: 59. https://doi.org/10.1186/1471-2105-5-59 PMID: 15144565

130. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. 2008; 18: 1979–1990. https://doi.org/10.1101/gr.081612.108 PMID: 18757608

131. Lowe TM, Eddy SR. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997; 25: 955–964. PMID: 9023104

132. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007; 35: 3100–3108. https://doi.org/10.1093/nar/gkm160 PMID: 17452365

133. Quinlan AR. BEDTools: The Swiss-army tool for genome feature analysis. Curr Protoc Bioinformatics. 2014; 47: 11.12.1–34.

134. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011; 7: 539. https://doi.org/10.1038/msb.2011.75 PMID: 21988835

135. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, et al. The Bioperl toolkit: Perl modules for the life sciences. Genome Res. 2002; 12: 1611–1618. https://doi.org/10.1101/gr.361602 PMID: 12368254

136. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009; 19: 1639–1645. https://doi.org/10.1101/gr.092759.109 PMID: 19541911

137. Karst SM, Kirkegaard RH, Albertsen M. mmgenome: a toolbox for reproducible genome extraction from metagenomes. bioRxiv 059121 [Preprint]. 2016 [Cited 1 May 2017]. Available from: https://www.biorxiv.org/content/early/2016/06/15/059121. https://doi.org/10.1101/059121

138. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015; 16: 157. https://doi.org/10.1186/s13059-015-0721-2 PMID: 26243257

139. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9: 357–359. https://doi.org/10.1038/nmeth.1923 PMID: 22388286

140. Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, WGS500 Consortium, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet. 2014; 46: 912–918. https://doi.org/10.1038/ng.3036 PMID: 25017105

141. Garrison E. VCFlib. Version 1.0.0-rcl [software]. 2016. [Cited 1 May 2017]. https://github.com/vcflib/vcflib#vcflib.

142. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, UniProt Consortium. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics. 2015; 31: 926–932. https://doi.org/10.1093/bioinformatics/btu739 PMID: 25398609

143. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30: 772–780. https://doi.org/10.1093/molbev/mst010 PMID: 23329690

144. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015; 32: 268–274. https://doi.org/10.1093/molbev/msu300 PMID: 25371430

145. Tange O. GNU Parallel: The Command-Line Power Tool.; login: The USENIX Magazine. 2011; 36: 42–47.

146. R Core Team. R: A language and environment for statistical computing. Version 3.3.1 [software]. 2016. [Cited 1 May 2017]. https://www.R-project.org/.

147. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004; 20: 289–290. PMID: 14734327