Evolutionary dynamics of transposable elements in bdelloid rotifers

Reuben W. Nowell[1,2,*], Christopher G. Wilson[1,2], Pedro Almeida[2,3], Philipp H. Schiffer[4], Diego Fontaneto[5], Lutz Becks[6,7], Fernando Rodriguez[8], Irina R. Arkhipova[8], and Timothy G. Barraclough[1,2,*]

[1]Department of Zoology, University of Oxford, 11a Mansfield Road, Oxford, United Kingdom. [2]Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot, Berkshire, United Kingdom.[3]Division of Biosciences, University College London, London, United Kingdom. [4]Institute of Zoology, Section Developmental Biology, Wormlab, University of Cologne, Köln, Germany. [5]National Research Council of Italy, Water Research Institute, Verbania Pallanza, Italy. [6]Community Dynamics Group, Department of Evolutionary Ecology, Max Planck Institute for Evolutionary Biology, Plön, Germany. [7]Aquatic Ecology and Evolution, University of Konstanz, Germany. [8]Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA

***Corresponding authors:** RWN: reubennowell@gmail.com; TGB: tim.barraclough@zoo.ox.ac.uk

ORCID numbers:

RWN: 0000-0001-7546-6495
CGW: 0000-0001-7649-5255
PA: 0000-0001-6790-8687
PHS: 0000-0001-6776-0934
DF: 0000-0002-5770-0353
LB: 0000-0002-3885-5253
FR: 0000-0003-4044-8734
IRA: 0000-0002-4805-1339
TGB: 0000-0002-8084-2640

Abstract

Transposable elements (TEs) are selfish genomic parasites whose ability to spread autonomously is facilitated by sexual reproduction in their hosts. If hosts become obligately asexual, TE frequencies and dynamics are predicted to change dramatically, but the long-term outcome is unclear. Here, we test current theory using whole-genome sequence data from eight species of bdelloid rotifers, a class of invertebrates where males are thus far unknown. Contrary to expectations, we find a diverse range of active TEs in bdelloid genomes, at an overall frequency within the range seen in sexual species. We find no evidence that TEs are spread by cryptic recombination or restrained by unusual DNA repair mechanisms, but we report that bdelloids share a large and unusual expansion of genes involved in RNAi-mediated TE suppression. This suggests that enhanced cellular defence mechanisms might mitigate the deleterious effects of active TEs and compensate for the consequences of long-term asexuality.

Introduction

Transposable elements (TEs) are repeated sequences of DNA that can mobilize and replicate themselves within genomes [1–3]. TEs are divided into two major categories: class I retrotransposons, which use a 'copy-and-paste' replication mechanism via a reverse-transcribed RNA intermediate, and class II DNA transposons, which use 'cut-and-paste' replication with a DNA intermediate. Both classes are ancient and diverse—retrotransposons are found in some

bacteria and nearly all eukaryotes, while DNA transposons are found across the tree of life [4–6]. Although TE replications are occasionally beneficial [7], the vast majority are deleterious for the host [8,9]. Costs include insertional mutations that disrupt genes [10], cellular costs of replicating and expressing excess DNA [11], and increased risk of chromosomal abnormalities due to ectopic recombination between homologous TE sequences interspersed through the genome [12,13]. Despite this, by replicating autonomously as selfish elements, TEs can accumulate to large numbers within genomes—for example, TEs comprise 46% of the human genome, including over 1 million (~11%) nonautonomous *Alu* retroelements [14,15]. TE numbers vary greatly, however, even between closely related species. In vertebrates, for example, TEs span an order of magnitude, from below 6% to over 50% of the genome [16], with similarly large variation observed within and between other groups such as arthropods [17], nematodes [18], and fungi [19]. Explaining this variation is vital to understanding the mechanisms affecting TE spread and control.

Sexual reproduction has long been thought to play a major role in TE dynamics within eukaryotes. On the one hand, sexual reproduction and outcrossing decouples the fate of TEs from other host genes, allowing them to jump into new genomic backgrounds and to behave as selfish genomic parasites [1,2,20]. On the other hand, sex enables the efficient removal of deleterious insertions from populations through recombination and segregation [21–23]. The risk of chromosome abnormalities due to ectopic recombination, arguably the main cost of high TE loads in eukaryotes [9,24], also occurs during chromosome pairing at meiosis. Sex therefore plays opposing roles—it permits spread and selfish behaviour of TEs, and yet it facilitates and strengthens selection against high loads. Variation in TE content among taxa might thus result from shifts in the balance of these different opposing forces.

By this logic, the loss of sexual reproduction should affect TE dynamics dramatically. Since asexual lineages generally arise from sexual species [25], it is likely that they initially harbor many active TEs [26,27]. All else being equal, the loss of recombination will limit the ability of selection to remove deleterious insertions from a fully linked host genome, and so the load of TEs should accumulate. At the same time, the fate of TEs is immediately coupled to that of the host genome, resulting in intensified selection for inactivation, excision or domestication of the elements [2,22,27–29]. The genomes of asexual lineages whose TEs continued to replicate unchecked would become overrun, potentially leading to extinction of the lineage and the TEs themselves. While some TEs could be maintained by horizontal transfer between species (especially class II DNA elements [6,30]) or by having beneficial effects (as in bacteria, [31,32]), other TEs—particularly the class I LINE-like (i.e. non–long terminal repeat [LTR]) retrotransposons—are thought to be transmitted almost exclusively vertically [5,33,34], and therefore depend strongly on sex for their persistence.

Models of the population genetics of vertically transmitted TEs in asexuals predict one of two outcomes: either TEs accumulate within lineages faster than they can be removed, overrunning each lineage in turn and driving the population extinct, or, conversely, TE removal outweighs proliferation and the population eventually purges itself entirely of deleterious TEs [27,35,36]. These predictions are difficult to test empirically, however, because the time required for a population to arrive at either extinction or redemption is expected to be on the order of millions of generations [27], too long to observe directly and beyond the lifespan of most asexual lineages [37,38].

Here, we test these ideas in a well-known group of asexual animals, the bdelloid rotifers. These microscopic invertebrates appear to have reproduced without males or meiosis for tens of

millions of years, diversifying into hundreds of species within limno-terrestrial and freshwater habitats globally [39,40]. Bdelloids sampled from nature (and those reared in the laboratory) consist entirely of parthenogenetic females, and neither males nor hermaphrodites are described for any species despite centuries of close observation by naturalists [39,41,42]. Genetic and genomic evidence for their proposed ancient and obligate asexuality remains uncertain, however. Initial evidence of long-term asexuality [43,44] has been refuted by later studies or confounded by alternative explanations [45–47]. Some recent studies have proposed alternative modes of inter-individual genetic exchange, but these suggestions would require exotic mechanisms unknown in other animals [44,48], or rates of sex that are difficult to reconcile with the lack of observed males [49]. While the precise nature of reproduction in bdelloids remains an open question, nonetheless they provide a unique test-case for models of TE evolution when conventional sex is absent or strikingly rare.

Initial PCR-based surveys of five bdelloid genomes found no evidence of class I retrotransposons from either the LTR or LINE-like superfamilies, but did reveal a diverse array of class II DNA transposons, mostly at low copy number [50]. The presence of class II TEs in bdelloids might be explained by horizontal transfer, which is thought to occur more frequently for class II TEs with DNA intermediates [6,30,33,34] (but see [51]). The apparent lack of retrotransposons contrasted sharply, however, with their near ubiquity in other taxa. At the time, the absence of class I TEs appeared consistent with the view that long-term asexual evolution in bdelloids had caused the loss of parasitic elements that depended on sexual transmission [22,27,50,52].

Another unusual aspect of bdelloid physiology was suggested to contribute to their seemingly low TE complement. In most bdelloid species (but not all), individuals can survive complete desiccation at any life stage via a process called anhydrobiosis ('life without water'). Desiccation causes double-strand breakages (DSBs) in DNA, but bdelloids are able to repair these and recover to an unusual degree [53–55]. It was proposed that anhydrobiosis might influence TE evolution in two ways [27,52,56]. First, DSB repair could aid TE removal, either via gene conversion from a homologous chromosome lacking the TE insertion, or excision of mis-paired regions. Second, the pairing of homologous chromosomes, if required during DSB-repair, could provide a context for ongoing selection against chromosomal abnormalities caused by ectopic recombination. In either case, anhydrobiosis would decrease the number of TEs, potentially helping to explain the low overall TE content encoded in bdelloid genomes.

These early ideas were transformed by more detailed studies of the model bdelloid species *Adineta vaga*, which used refined methods and genome-scale data to discover a variety of retrotransposon families. These include an endonuclease-deficient *Penelope*-like element (PLE) designated *Athena* [57,58], which is itself incorporated within much larger and highly unusual retroelements called *Terminons* [59]; another PLE that has retained its endonuclease [60], LTR retrotransposons (*Juno*, *Vesta*, *TelKA* and *Mag* [61,62]), and LINE-like retrotransposons (*R4*, *R9*, *Hebe*, *RTE*, *Tx1* and *Soliton* [44,63,64]). In total, TEs accounted for 2.2% of the 217 Mb genome (~4.8 Mb) [44], rising to ~4% on inclusion of the recently discovered giant *Terminon* elements [59]. The conclusion that bdelloids lack class I TEs therefore no longer holds, and the predicted effects of asexuality and anhydrobiosis on TE evolution remain open questions that are amenable to testing with comparative genomics. Specifically, the comparison of desiccating species with those few bdelloid rotifer lineages that are unable to survive desiccation would indicate whether anhydrobiosis does limit TE numbers as hypothesized. Also, comparisons within populations could shed light on the activity of TEs and whether bdelloids possess cryptic forms of recombination that could aid in TE removal or facilitate TE spread.

Here, we test the predicted effects of asexuality and anhydrobiosis on TE evolution by comparing 42 rotifer genomes belonging to 15 taxonomic species. Our sample includes both desiccating and nondesiccating bdelloids, and eight monogonont rotifers [65–67], a separate class that alternates sexual and asexual reproduction and cannot survive desiccation as adults. Results are set in context by comparison to published genomes from an acanthocephalan [68] (now classified with rotifers in the Phylum Syndermata) and a range of other animal phyla. We ask six questions raised by theory. (1) How does the abundance and diversity of TEs in bdelloids differ from that in other animals, including sexual rotifers? (2) Are TEs recently and currently active in all bdelloids? (3) Can we detect signatures of recombination affecting bdelloid TEs, as might be expected if bdelloids do have cryptic forms of sexual reproduction? (4) Do desiccating species contain fewer TEs than nondesiccating species, as previously theorized? (5) Are TEs in bdelloids under the same selective constraints as in other taxa? (6) Finally, do bdelloids possess alternative molecular mechanisms that might help to keep their TEs under control, as might be expected if the ability to control TEs by sexual reproduction is absent or reduced?

Results and Discussion

High-quality population genomics data for bdelloid rotifers. To quantify variation in repeat content within and between bdelloid species, we generated *de novo* whole-genome assemblies for 31 rotifer samples encompassing nine species (**Fig 1a**; **Table 1**; S1 Data). Three of these assemblies were generated using 10x Genomics linked-read data (for *Adineta steineri*, *Rotaria sordida*, and *Rotaria* sp. 'Silwood-1'), while 26 are from single-individual samples. In order to capture as many potential repeats as possible, we generated two assemblies for each sample: a 'reference' assembly, with a focus on quality and contiguity, and a 'maximum haplotype' (maxhap) assembly that included small or highly similar contigs that might be derived from recent TE duplications or other sources of copy number variation, at the expense of contiguity.

Reference genomes showed an expected scaffold size (AU, see Materials and methods) ranging from 21.1 kb (*Didymodactylos carnosus*) to 702.3 kb (*R.* sp. 'Silwood-1') and BUSCO scores that indicated 89–98% of 303 core eukaryote genes were completely recovered, increasing to 96–99% if fragmented copies are included (**Table 1**). General genome characteristics such as genome size (assembly span), the proportion of G + C nucleotides (GC%), the number of coding genes (CDS), and the level of homologous divergence (number of SNPs identified within CDS) were within the range expected from previous analyses of bdelloid genomes [44,46] (**Fig 1b–c**; **Table 1**; S1 Fig). Intragenomic collinearity and synonymous divergence of coding regions in the *A. steineri*, *R. sordida* and *R.* sp. 'Silwood-1' maximum haplotype assemblies reveal the characteristic signature of degenerate tetraploidy that has been found in all bdelloid species examined to date (**Fig 1d**).

Compared to the reference set, maxhap assemblies generally showed increased span (mean increase = 17.9 Mb ± 21.5 standard deviation [SD]) and were substantially more fragmented, as expected (S1 Table). Nonetheless, BUSCO completeness scores remained high, with 76% to 98% of genes completely recovered (increasing to 95–98% if fragmented copies are included), indicating that the majority of core genes are successfully captured (S1 Table). The BUSCO duplication metric ('D') does not increase greatly between reference and maxhap assemblies, which shows that the additional sequences retained in the maxhap assemblies do not contain complete extra copies of core genes. Thus, the maxhap assemblies are not fully haplotype-resolved representations of the genome, except in the case of the three 10x assemblies.

203 **Table 1.** Assembly statistics for one monogonont and 30 bdelloid rotifer reference assemblies presented in this study.
204

Sample ID	Species name	SZ (Mb)	NN	N50 (kb)	L50	AU (kb)	GC (%)	Gaps (kb)	Coverage (X)	Genome BUSCO score	CDS	Proteome BUSCO score	GenBank accession
Bc_PSC1	*Brachionus calyciflorus* (Monogonont)	116.7	14,869	18.5	1,692	26.6	25.6	78	186	C:96%[S:93%,D:3%],F:2%	24,404	C:98%[S:93%,D:5%],F:1%	GCA_904009505.1
Ar_ARIC003	*Adineta ricciae*	135.6	4,302	283.8	129	388	35.5	65	89	C:97%[S:58%,D:39%],F:2%	49,015	C:97%[S:52%,D:45%],F:1%	GCA_904047305.1
As_10x_p	*Adineta steineri*	171.1	8,257	200.1	163	394.5	29	206	198	C:95%[S:67%,D:33%],F:2%	50,321	C:97%[S:58%,D:38%],F:2%	
As_ASTE804	*Adineta steineri*	160.3	9,359	158.1	265	214.9	29.1	152	62	C:95%[S:74%,D:22%],F:2%	47,222	C:98%[S:74%,D:24%],F:2%	GCA_904047245.1
As_ASTE805	*Adineta steineri*	156.3	9,008	169.6	245	226.4	29.2	129	65	C:98%[S:77%,D:21%],F:1%	43,986	C:99%[S:72%,D:26%],F:1%	GCA_904047255.1
As_ASTE806	*Adineta steineri*	160.3	7,597	168.2	257	222.5	29.2	145	82	C:96%[S:72%,D:24%],F:2%	45,930	C:98%[S:74%,D:24%],F:2%	GCA_904047275.1
Dc_DCAR505	*Didymodactylos carnosus*	323.6	87,048	7.8	11,656	10.5	33.5	41	21	C:86%[S:69%,D:17%],F:8%	46,286	C:88%[S:71%,D:18%],F:9%	GCA_904047325.1
Dc_DCAR706	*Didymodactylos carnosus*	368.8	78,356	12	7,695	19.1	33.5	13	76	C:95%[S:70%,D:25%],F:2%	46,863	C:95%[S:71%,D:25%],F:2%	GCA_904053135.1
Rd_10x_p	*Rotaria sordida*	272.5	16,571	64.5	843	193.8	30.8	395	91	C:94%[S:77%,D:19%],F:2%	44,299	C:95%[S:69%,D:26%],F:2%	
Rd_RSOR408	*Rotaria sordida*	252.9	20,315	57.6	1,246	75.5	30.4	291	39	C:94%[S:76%,D:19%],F:3%	40,501	C:97%[S:73%,D:24%],F:2%	GCA_904053145.1
Rd_RSOR410	*Rotaria sordida*	252.6	19,518	60.9	1,179	80	30.4	252	42	C:95%[S:77%,D:18%],F:3%	40,474	C:98%[S:74%,D:24%],F:2%	GCA_904053495.1
Rd_RSOR504	*Rotaria sordida*	251.3	22,067	53.1	1,338	69.8	30.4	369	39	C:94%[S:78%,D:16%],F:3%	41,085	C:96%[S:73%,D:23%],F:3%	GCA_904047315.1
Rg_MAG1	*Rotaria magnacalcarata*	178.7	19,184	42	1,077	62.4	32	402	58	C:97%[S:81%,D:16%],F:1%	40,318	C:99%[S:76%,D:22%],F:1%	GCA_903995865.1
Rg_MAG2	*Rotaria magnacalcarata*	181.1	22,216	39.7	1,141	61	32	433	63	C:98%[S:81%,D:17%],F:1%	40,289	C:99%[S:74%,D:26%],F:0%	GCA_903995855.1
Rg_MAG3	*Rotaria magnacalcarata*	180.9	22,132	40.7	1,142	60	32	508	60	C:96%[S:80%,D:17%],F:1%	40,740	C:99%[S:77%,D:22%],F:0%	GCA_903995845.1
Rg_RM15	*Rotaria magnacalcarata*	174	18,391	46.5	966	67.3	32	430	55	C:96%[S:80%,D:16%],F:2%	38,283	C:99%[S:77%,D:22%],F:1%	GCA_903995885.1
Rg_RM9	*Rotaria magnacalcarata*	173.8	19,520	44	999	64.7	31.9	594	51	C:96%[S:80%,D:16%],F:1%	38,404	C:98%[S:76%,D:22%],F:1%	GCA_903995825.1
Rp_RPSE411	*Rotaria* sp. 'Silwood-2'	296.5	30,050	102.5	381	691.1	31	247	35	C:93%,[S:72%,D:21%],F:4%	48,378	C:95%[S:72%,D:23%],F:4%	GCA_904053095.1
Rp_RPSE503	*Rotaria* sp. 'Silwood-2'	285.6	33,174	78.3	449	627.2	31.3	446	34	C:91%,[S:75%,D:17%],F:5%	48,269	C:92%[S:72%,D:20%],F:7%	GCA_904053125.1
Rp_RPSE809	*Rotaria* sp. 'Silwood-2'	271.1	28,589	101.6	350	681.4	31	377	27	C:93%,[S:76%,D:17%],F:4%	47,010	C:95%[S:74%,D:22%],F:4%	GCA_904053085.1
Rp_RPSE812	*Rotaria* sp. 'Silwood-2'	264.1	34,498	80.8	403	616.2	31.1	428	27	C:89%,[S:74%,D:15%],F:8%	47,040	C:90%[S:73%,D:17%],F:8%	GCA_904053105.1
Rs_AK11	*Rotaria socialis*	149.2	6,303	111.3	370	150.2	31.8	442	39	C:97%,[S:80%,D:17%],F:1%	34,844	C:99%[S:75%,D:24%],F:1%	GCA_903995835.1
Rs_AK15	*Rotaria socialis*	147.4	5,030	134.7	305	177.6	31.8	423	37	C:96%,[S:79%,D:18%],F:1%	34,140	C:98%[S:76%,D:23%],F:1%	GCA_903995795.1
Rs_AK16	*Rotaria socialis*	147.4	4,720	139.5	296	180.3	31.8	332	43	C:97%,[S:80%,D:18%],F:0%	33,717	C:99%[S:76%,D:23%],F:1%	GCA_903995805.1
Rs_AK27	*Rotaria socialis*	149.9	5,952	123.7	343	159.8	31.8	458	36	C:97%,[S:80%,D:17%],F:0%	34,369	C:99%[S:75%,D:24%],F:1%	GCA_903995815.1
Rs_RS1	*Rotaria socialis*	151.1	6,254	124.9	334	166.2	31.8	490	40	C:97%,[S:80%,D:17%],F:0%	33,937	C:99%[S:77%,D:22%],F:1%	GCA_903995875.1

Rw_10x_p	*Rotaria* sp. 'Silwood-1'	310.4	16,995	211.8	211	126.2	31.1	534	53	C:95%,[S:76%,D:20%],F:1%	44,241	C:97%[S:73%,D:24%],F:1%	
Rw_RSIL801	*Rotaria* sp. 'Silwood-1'	268.4	28,548	136.5	288	687.3	30.8	472	45	C:94%,[S:77%,D:17%],F:4%	41,574	C:95%[S:75%,D:21%],F:5%	GCA_904053155.1
Rw_RSIL802	*Rotaria* sp. 'Silwood-1'	249.9	21,286	153.4	238	702.3	30.7	451	42	C:92%,[S:76%,D:16%],F:4%	39,577	C:94%[S:76%,D:18%],F:4%	GCA_904053505.1
Rw_RSIL804	*Rotaria* sp. 'Silwood-1'	247.6	25,643	118.3	287	660.4	30.8	667	34	C:94%,[S:78%,D:16%],F:3%	41,139	C:96%[S:78%,D:19%],F:3%	GCA_904053115.1
Rw_RSIL806	*Rotaria* sp. 'Silwood-1'	294.1	29,968	132.4	333	681.9	30.8	500	31	C:95%,[S:79%,D:16%],F:2%	48,259	C:97%[S:78%,D:19%],F:2%	GCA_904054495.1

Sequence statistics codes: **SZ**, total sequence length (Mb); **NN**, number of sequences; **N50**, N50 scaffold length (kb); **L50**, N50 index; **AU**, expected scaffold size (area under 'Nx' curve, kb).
Gaps: Undetermined bases (Ns) introduced during scaffolding.
BUSCO score based on eukaryote set (n = 303); BUSCO codes: **C**, complete; **S**, complete and single copy; **D**, complete and duplicated; **F**, fragmented.
Abbreviations: GC, guanine + cytosine; BUSCO, Benchmarking Universal Single-Copy Orthologs.

To these new data, we added published genomes for seven monogonont taxa (one from the *Brachionus calyciflorus* species complex [65] and six from four species of the *Brachionus plicatilis* species complex, namely *B. asplanchnoidis*, *B. plicatilis* sensu stricto (HYR1), *B. rotundiformis*, and *B.* sp. 'Tiscar' [66,67]) and four bdelloids (*A. vaga*, *Adineta ricciae*, *Rotaria magnacalcarata* and *Rotaria macrura* [44,46]), yielding a total of 42 rotifer genomes. Of these, 11 samples belong to nondesiccating bdelloid species (five individuals each from *R. magnacalcarata* and *Rotaria socialis*, and the previously published genome of *R. macrura*).

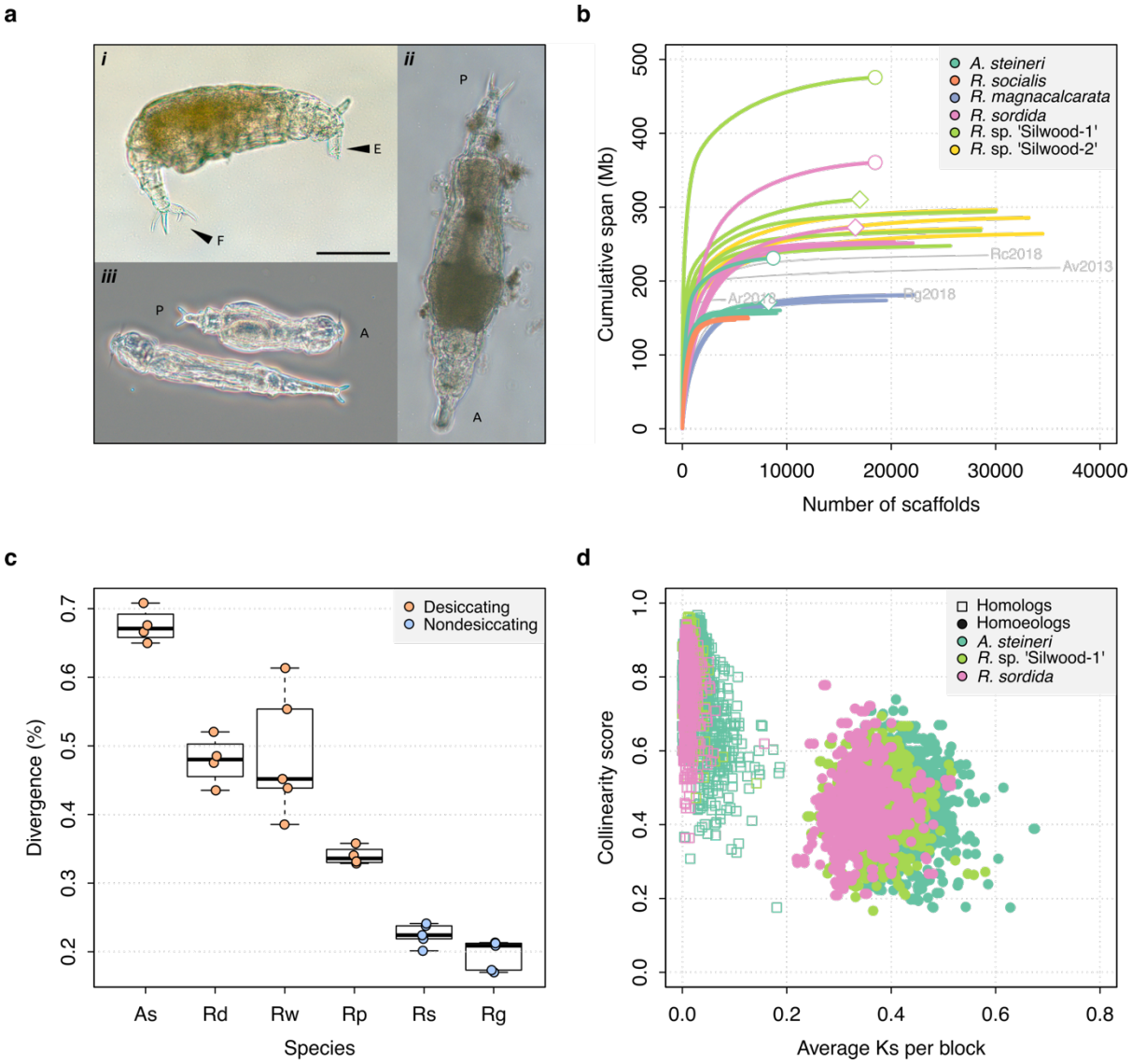

Fig 1. Genome properties of sequenced rotifers. (a) Bdelloid rotifer morphology; scale bar indicates 100 µm. (*i*) Individual from an undescribed species of *Rotaria* (*R.* sp. 'Silwood-1'), showing eyes (E) and foot (F) with two spurs and three toes. (*ii*) Further image of *R.* sp. 'Silwood-1' with anterior–posterior (A–P) axis marked. (*iii*) Two individuals of *A. steineri* in phase contrast. **(b)** Cumulative assembly span for six bdelloid species with population genomics data ($n > 2$). 10x Genomics haploid ('pseudohap') and diploid ('megabubbles') assemblies for *A. steineri*, *R. sordida* and *R.* sp. 'Silwood-1' are indicated with diamond and circle symbols, respectively. The four previously published genomes for *A. vaga* ('Av2013', GenBank accession GCA_000513175.1) and *A. ricciae* ('Ar2018', GCA_900240375.1), *R. macrura* ('Rc2018', GCA_900239685.1) and *R. magnacalcarata* ('Rg2018', GCA_900239745.1) are indicated in grey, for comparison. **(c)** Intragenomic divergence, measured as the number of SNPs detected in coding regions (CDS). Boxplots show the median (band), interquartile range (box) and minimum/maximum values

(whiskers). Underlying data are shown as jittered points. Desiccation-tolerant species are in orange, intolerant species in blue. Species abbreviations: **As**, *A. steineri*; **Rd**, *R. sordida*; **Rw**, *R.* sp. 'Silwood-1'; **Rp**, *Rotaria* sp. 'Silwood-2'; **Rg**, *R. magnacalcarata*; **Rs**, *R. socialis*. **(d)** Genome structure in *A. steineri*, *R. sordida* and *R.* sp. 'Silwood-1' haplotype-resolved ('megabubbles') assemblies. Each point represents a collinear block of genes, plotted by average pairwise synonymous (K_S, *X*-axis) and collinearity score (see Materials and methods and S1 note) on the *Y*-axis. Separation into two distinct clusters representing homologous (squares) and homoeologous (circles) relationships among gene copies is consistent with ancestral tetraploidy, with homoeologous copies derived from a putative ancient genome duplication.

Abundant and diverse TEs in bdelloid genomes. To ascertain the repeat content of bdelloid genomes relative to other taxa in a consistent manner, we used the RepeatModeler and RepeatMasker pipelines to identify and classify repeats across genomes. The total proportion of the genome classified as repetitive ranged from ~19% to 45% across bdelloid genera, with variation within and between species (**Fig 2a**; S2 Fig; S3 Fig; S2 Data; S3 Data). Most of these are simple or unclassified repeats that do not belong to major TE superfamilies. While the precise nature of these unclassified repeats is not elucidated, an appreciable fraction (~7–27%, mean = 17%) are also annotated as protein-coding and thus may be derived from gene expansions or other duplications, while a further small fraction (< 1%) are accounted for by an additional survey of small, nonautonomous class II TEs called miniature inverted-repeats (MITEs) (S2 Data). The proportion of the genome accounted for by known TEs was much smaller, ranging from 2.4% to 7.3% (mean = 4.9% ± 1.2 standard deviations [SD], median = 5.1%) in bdelloids. Broken down by class and superfamily, the mean values are: class I total = 2.09% ± 0.75 (PLEs = 0.59% ± 0.14; LTRs = 0.68% ± 0.26; and LINEs = 0.82% ± 0.47); class II total = 2.79 ± 0.8 (DNA transposons = 2.49% ± 0.77; rolling circles = 0.30 ± 0.11). These results are in broad agreement with previous estimates of TE content in bdelloids [44,46,47,59].

We first compare these values to closely related sexual lineages in the phylum Syndermata, the monogonont rotifers and acanthocephalans. The most striking difference relates to the recently published genome of the obligately sexual acanthocephalan *Pomphorhynchus laevis* [68], which encodes a substantially greater proportion of repeats than either bdelloids or monogononts. In agreement with Mauer et al., we find ~66% of the *P. laevis* genome to be composed of repeats. The large majority are class I retrotransposons (~71% of the total TE content) from the LINE (~52%) and PLE (~15%) superfamilies, and there are relatively few DNA transposons (~1.3%) (**Fig 2b**). There is increasing evidence that acanthocephalans may be the closest relatives to the Class Bdelloidea [69–71]. However, all members of the Acanthocephala (also known as the thorny-headed worms) are obligate endoparasites and are highly differentiated in both morphological and molecular terms from other syndermatans. It is unclear whether these features may contribute to the high repeat content and other unusual genome characteristics of *P. laevis* [68].

In monogononts, the mean values for class I TEs is 5.2% ± 1.5 SD and for class II TEs it is 2.5% ± 1.0 SD. Thus, monogononts are slightly more TE-rich than bdelloids, but also substantially more variable between species (**Fig 2a**; S3 Fig). Repeat content differs between bdelloids and monogononts in two main ways, both in regard to the composition of class I retrotransposons. First, monogononts encode substantially more LINE-like retroelements than bdelloids, making up (on average) approximately 50% and 16% of the total TE content in each clade respectively. The frequency of LINEs is of particular interest, because this class of TEs is thought to be least likely to undergo horizontal transfer and thus the most dependent on sex for transmission [5,33,34]. Second, *Penelope*-like elements (PLEs) have increased in proportion in all bdelloids relative to monogononts, from ~1% in monogononts to ~12% in bdelloids on average.

Interestingly, a high proportion of PLEs is also seen in the acanthocephalan *P. laevis*, while a high proportion of LINEs is found in both *D. carnosus* isolates (~35% of total TE content), a deeply branching lineage sister to all other bdelloid taxa included in the analyses. Thus, assuming that acanthocephalans are the closest relatives to bdelloids, the most parsimonious explanation for these broad-scale patterns is that the expansion of PLEs occurred in the ancestor to bdelloids and acanthocephalans, whereas the contraction of LINEs has occurred more recently, confined to a subset of bdelloid genera.

To put these differences in TE content among rotifers in a broader context, we applied the same repeat-finding pipeline to animals from a range of more distantly related protostome phyla: three insects, a nematode, two tardigrades, five molluscs, two annelids, a brachiopod, platyhelminth and orthonectid [72–83] (S2 Table). As expected, both the abundance and diversity of TEs varied widely across taxa (**Fig 2b**). Total TE content ranged from 0.8% (the insect *Polypedilum vanderplanki*) to ~24% (octopus and platyhelminth). Although the acanthocephalan genome appears to be particularly rich in TEs, bdelloids (except for *D. carnosus*) have modest amounts of TEs, including class I TEs specifically, while in monogononts these amounts vary greatly among species. All bdelloids encode relatively more TEs than both *Polypedilum* species but fewer than *Drosophila melanogaster*, *Caenorhabditis elegans*, annelid worms and some molluscs, and are intermediate with respect to other taxa. Note that TE proportion in molluscs, nematodes and flatworms is known to be highly variable (e.g. [18]), while bdelloids display much less variability barring the early-branching *D. carnosus*.

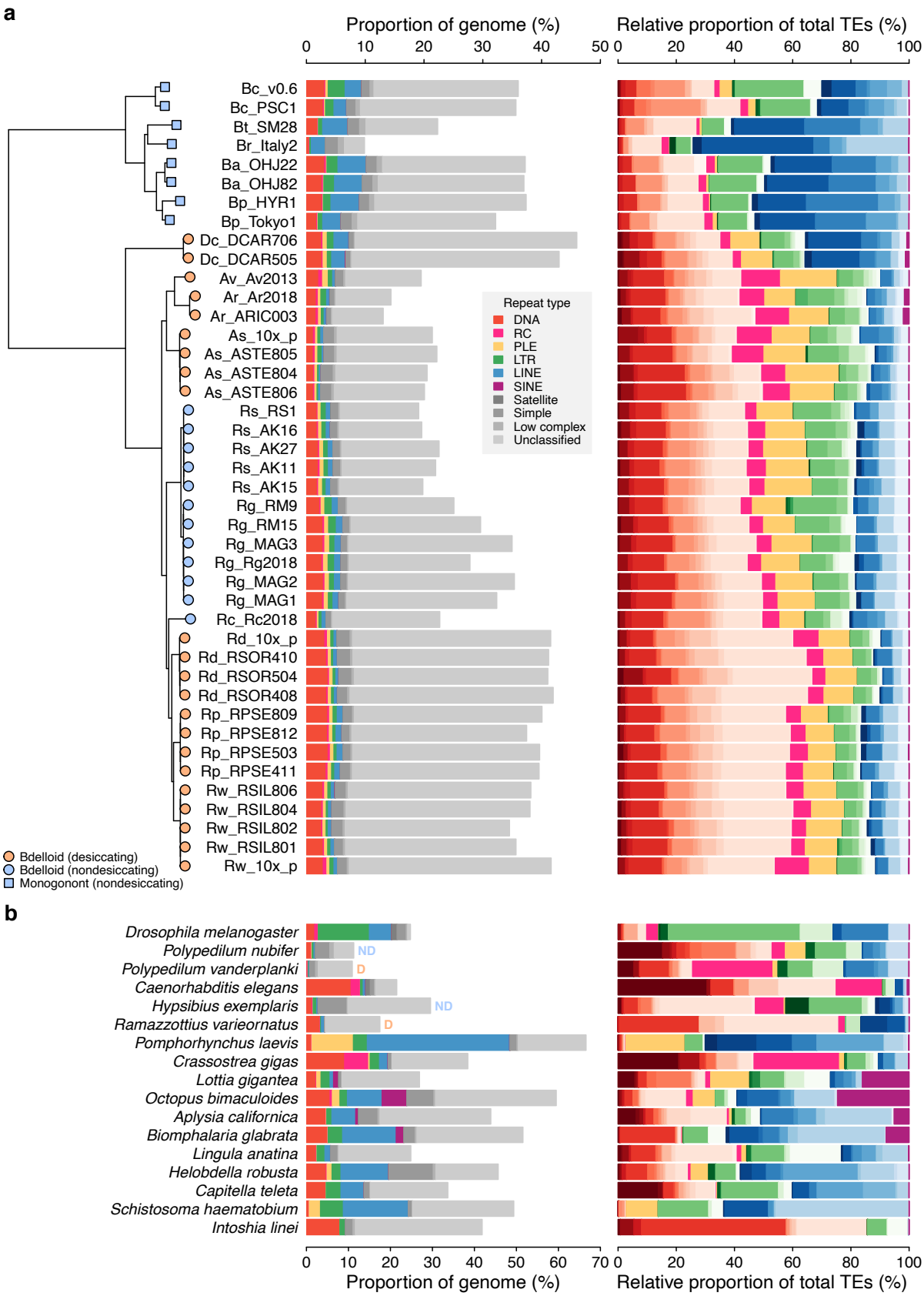

Fig 2. Repeat content and diversity in rotifer genomes. (a) Maximum likelihood phylogeny of eight monogonont (square symbols on tips) and 34 bdelloid (circles) genomes based on the concatenated alignment of a subset of core eukaryotic (BUSCO) genes. Orange and blue tip colours indicate desiccating and nondesiccating taxa, respectively. Species codes in tip names are: **Bc**, *Brachionus*

calyciflorus; **Br**, *B. rotundiformis*; **Bt**, *B.* sp. 'Tiscar'; **Bp**, *B. plicatilis* HYR1; **Ba**, *B. asplanchnoidis*; **Dc**, *Didymodactylos carnosus*; **Av**, *Adineta vaga*; **Ar**, *A. ricciae*; **As**, *A. steineri*; **Rs**, *Rotaria socialis*; **Rg**, *R. magnacalcarata*; **Rc**, *R. macrura*; **Rd**, *R. sordida*; **Rw**, *R.* sp. 'Silwood-1'; **Rp**, *R.* sp. 'Silwood-2'. Repeat content is shown as the genome proportion (%) broken down by TE superfamily (middle panel), and relative proportion (%) of total known (i.e., classified) TEs (right panel), where colours represent TE superfamilies (see legend) and shades of colour represent different TE families within each superfamily. **(b)** Equivalent repeat content analysis in 17 protostome animal genomes, including the model species *D. melanogaster* and *C. elegans*, the recently published acanthocephalan rotifer *P. laevis*, and selected other species from across the protostome group. Two further examples of desiccating (orange 'D') and nondesiccating (blue 'ND') species pairs are shown: the insects *P. nubifer* and *P. vanderplanki* and the tardigrades *H. exemplaris* and *R. varieornatus*.

These results show that bdelloid species encode an abundant diversity of both class I and II TEs, and, set against the repeat content of other rotifers and animals, do not appear particularly deficient or unusual with regard to the proportion of nucleotides encoding transposable elements. Although the bdelloids do have lower frequencies of class I TEs (~16% of total TE content) than the monogononts (~50%) or the acanthocephalan (~52%), as predicted by theory for elements dependent on vertical transmission, they still possess them in numbers that are comparable to sexual organisms (e.g. *C. elegans*). Furthermore, the basally divergent bdelloid *D. carnosus* did not show the same magnitude of decrease in LINEs (~35%), indicating that different dynamics may be at play in different lineages. Thus, it is clear that the most simplistic expectations of TE evolution under the hypothesis of long-term asexuality (i.e., either runaway proliferation or complete elimination) are not met, necessitating an evaluation of possible explanations that may align theory with observation.

TE transposition is recent and ongoing. One possible explanation is that TEs in bdelloid genomes do not replicate autonomously or are otherwise inactivated or 'fossilised' within their host genome. To investigate this, we first generated divergence 'landscapes' for identified TE copies within each genome, using the de novo RepeatMasker results. TE landscapes measure the amount of sequence divergence between each TE copy and a consensus derived from all copies in its family [84]. Histograms of the resulting Kimura distances (*K*-values [85]) provide insights into the evolutionary history of TE activity [16,86,87].

TE landscapes for the three diploid (10x Genomics) assemblies of *A. steineri*, *R.* sp. 'Silwood-1' and *R. sordida* show that TE divergence is bimodal but strongly zero-inflated (**Fig 3**). A large number of TE copies have very low or no divergence from the consensus (*K*-value ≤ 1%). Assuming a molecular clock for nucleotide substitutions within duplicated TEs, such elements represent recent duplications that are highly similar to their progenitor copy, consistent with recent transposition of an active element. In proportion, most of these belong to class II DNA transposon superfamilies (in red), but the spike of zero divergence is also present for class I retrotransposons (in blue and green). An older, broader mode is seen around a *K*-value of 20–30% that probably reflects historical TE transpositions and/or a signal from the tetraploid genome structure present in all bdelloids sequenced to date. The same pattern was observed in the haplotype-resolved assemblies of *A. vaga* [44] and *A. ricciae* [46] (**Fig 3b**; S4 Fig), and was generally present but less pronounced in the other 'maxhap' assemblies depending on the repeat pipeline applied (S4 Fig; S5 Fig).

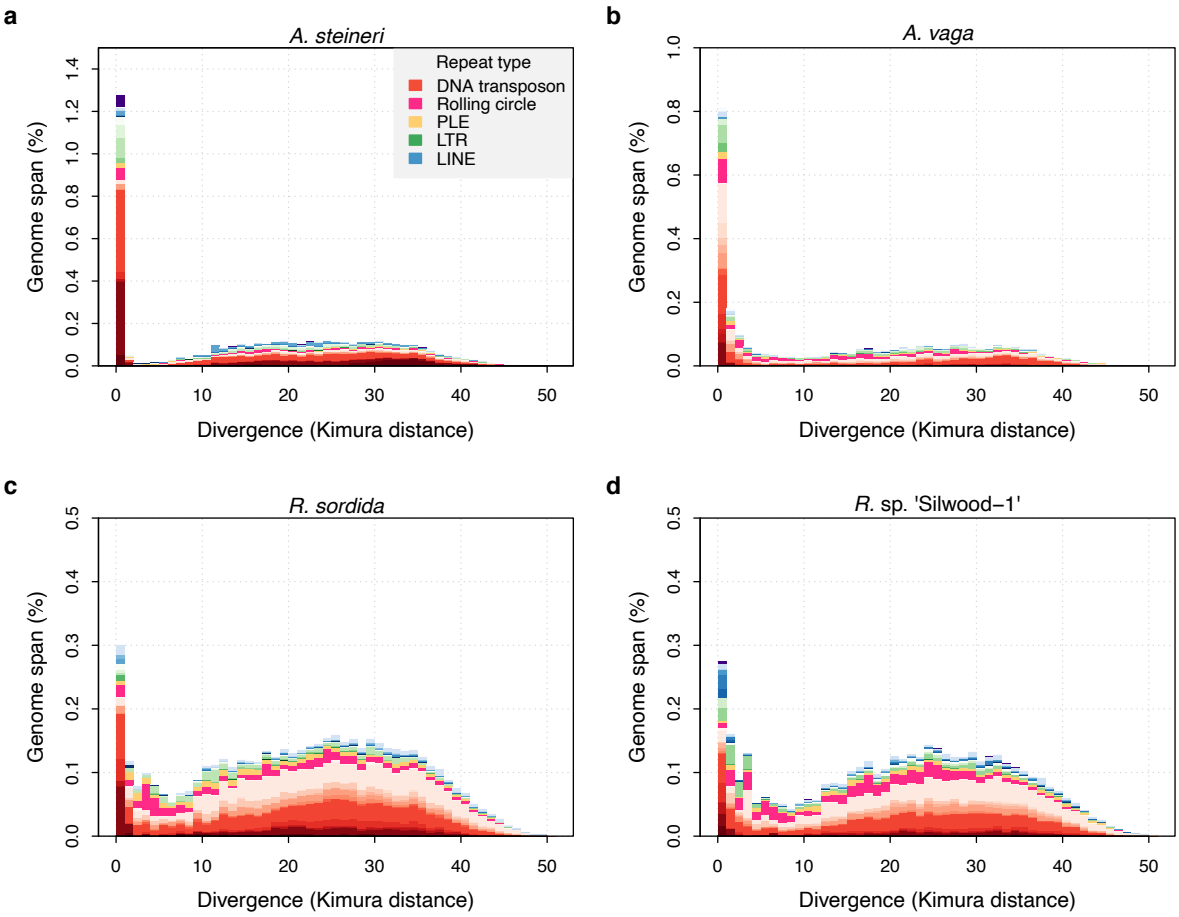

Fig 3. TE divergence landscapes for selected genomes. The *X*-axes show the level of divergence (Kimura substitution level, CpG adjusted) between each identified TE copy and the consensus sequence for that TE family (the inferred ancestral copy). Thus, if newly arising TE copies evolve neutrally, the amount of divergence is a proxy for the time since its duplication, with older copies accumulating more substitutions and appearing further to the right. The *Y*-axis shows the proportion of the genome occupied by each bin. Colours represent TE superfamilies (see legend) and shades of colour represent different TE families within each superfamily. Data are shown for the 10x Genomics diploid ('megabubbles') assemblies of *A. steineri*, *R.* sp. 'Silwood-1' and *R. sordida* compared to the published genome of *A. vaga*. Note different scales on some *Y*-axes.

To evaluate recent TE activity further, we developed a simple method to identify insertion sites for LTR retrotransposons (LTR-Rs) and assess their presence or absence in related individuals. Because most LTR-Rs insert into random genome locations [5,8], the neighbouring genome sequence provides a unique marker for a given insertion event [88]. We constructed a library of such insertion markers ('LTR-tags') for all 'full-length' LTR-Rs (i.e. those with long-terminal repeats present at both the 5' and 3' ends of the element) detected in our genomes, and then searched for their presence or absence in the other samples. For a given LTR-tag identified in genome *A*, the presence of a contiguous alignment in genome *B* indicates that the same insertion is shared between *A* and *B*.

For a set of 161 high-confidence and non-redundant LTR-Rs identified in the single-individual samples, alignment contiguity for each LTR-tag versus each of the other genomes was scored using a read-mapping approach (see Materials and methods), resulting in a pairwise matrix of presence/absence scores (**Fig 4a**; S4 Data). High scores for LTR insertion-site presence

correlated strongly with the phylogeny, resulting in an average score of ~0.9 within species compared to < 0.1 between species and a clear visual signal along the diagonal of Fig 4a. Very few LTR insertion sites were shared between bdelloid species. While some absences could reflect loss rather than gain, the restriction of nearly all LTR insertion sites to single species indicates that they have been gained during the separate evolutionary history of that species.

LTR-R insertions also vary between individuals within the same species, indicating recent transposition events and the potential for ongoing fitness consequences for the host. One case-study is illustrated for *R. magnacalcarata* (**Fig 4b**; S6 Fig). The individuals RM9 and RM15 share an LTR-R insertion that is not present in conspecifics. Aligning the regions of the genome assemblies containing these LTR-tags indicates that an 8.1 kb LTR-R has inserted into a protein-coding sequence in the lineage leading to RM9 and RM15. It has introduced a premature stop codon to a gene that encodes a protein (7,479 residues) of unknown function but with partial similarity to midasin, an ATPase essential to ribosome biosynthesis in several model eukaryotes [89,90]. In RM9 and RM15, the predicted product is substantially truncated (to 6,025 residues) by the element insertion. Despite the potential fitness consequences, RM9 and RM15 have evidently persisted for some time since, because they differ at approximately 0.5% of single-nucleotide sites across the 8.1 kb LTR element itself. A possible explanation is that both the RM9 and RM15 assemblies also contain a scaffold with an empty insertion site, indicating an intact version of the coding sequence spanned by the LTR insertion (represented in Fig. 4b by the partial matches on scaffolds 16719 and 07127, respectively). If the insertion is hemizygous, an uninterrupted homologous copy of the affected gene might mask or reduce the effect of the mutation. Given the degenerate tetraploid structure of bdelloid genomes, further homoeologous copies (i.e. derived from whole-genome duplication) might provide additional functional redundancy and help perform critical functions of interrupted genes.

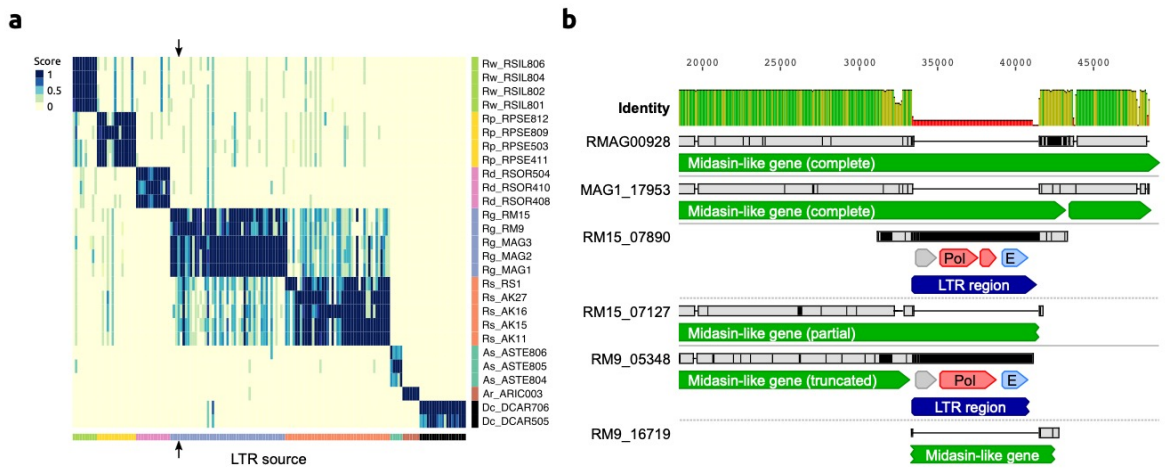

Fig 4. LTR insertion-site polymorphism in bdelloid species. (a) Columns represent 161 LTR-Rs identified across bdelloid samples, arranged by genome of origin (see colours at bottom and side). Support for the presence of a given LTR-R at a specific insertion site in each genome is scored from 0 (absent, yellow) to 1 (present, dark blue), where a score < 0.5 is strong evidence for absence (see Materials and methods for details). Arrows demark the location of the LTR-R example shown in b. **(b)** Nucleotide alignment of region around an LTR-R insertion (blue) identified in RM9 (scaffold 05348) and RM15 (scaffold 07890), alongside their putative homologous scaffolds (scaffolds 16719 and 07127 respectively) that do not show the insertion. Scaffolds from Rg2018 (RMAG00928) and MAG1 are also shown for comparison. Predicted CDS with similarity to Pol and Env proteins are shown in red and light blue. The LTR-R is most likely a member of the *TelKA* family, based on sequence similarity.

427
428
429 Thus, these data contradict the idea that bdelloid TEs are inactive. All TE superfamilies in other
430 bdelloids show a substantial fraction of copies at low-divergence, indicative of recent
431 proliferation. Moreover, there are multiple cases of insertion-site polymorphism within species,
432 and at least one case where a recent retroelement insertion into a protein-coding sequence seems
433 likely to have potential fitness consequences.
434
435 **No evidence that cryptic recombination helps to limit the spread of LTR-Rs**. Another
436 possible explanation for the apparent discrepancy between bdelloid TE profiles and theory is that
437 bdelloids in fact possess cryptic inter-individual recombination, either through undetected sex or
438 some alternative form of gene transfer. We therefore tested for a signature of recombination
439 among polymorphic LTR-R insertion sites within species. Under strict clonality, the pattern of
440 presence and absence across LTR-R loci should be nested and compatible with only mutational
441 gain and loss at each site. In contrast, in a sexual, outcrossing population, variation should be
442 shuffled among loci. LTR-Rs provide a powerful test of these predictions because random
443 insertion makes independent origins of the same LTR-R insertion site highly unlikely.
444
445 In every species with multiple samples, we found that variation in polymorphic TEs is perfectly
446 nested, with a consistency index in parsimony reconstruction of 1. Furthermore, in the two
447 species with multiple parsimony-informative characters, *R. socialis* and *R. magnacalcarata*, we
448 found a significantly positive index of association of presence and absences among LTR-R
449 insertion sites, as expected with clonal inheritance (S3 Table; S6 Fig). Approximate Bayesian
450 Computation with simulations of expected patterns under varying frequencies of sexual
451 reproduction showed that strictly clonal evolution could not be rejected (S7 Fig). While this test
452 uses a restricted set of markers, and so should not be viewed as a test of recombination for the
453 whole genome or species, it does support clonal inheritance of LTR-R loci and finds no evidence
454 that inter-individual recombination helps to limit their spread. Nevertheless, local LTR-LTR
455 recombination within genomes, leading to solo LTR formation, may act to bring the copy
456 number down [44].
457
458 **No evidence for lower TE loads in desiccating bdelloids.** The desiccation hypothesis posits
459 that TE numbers may be kept in check via the action of DSB-repair processes during recovery
460 from desiccation. Our study includes 11 nondesiccating bdelloid samples encompassing three
461 obligately aquatic species (*R. macrura*, *R. magnacalcarata* and *R. socialis*), while the remaining
462 samples were isolated from ephemeral ponds or moss and must undergo frequent cycles of
463 desiccation and rehydration to survive. Contrary to the prediction that TE load should be reduced
464 in desiccating species, there is little overall difference in TE proportions between desiccating and
465 nondesiccating lineages (mean = 4.8% ± 1.3% *vs.* 5.0% ± 0.9% respectively). Broken down by
466 TE superfamily, desiccating taxa have relatively more DNA transposons, simple, low complexity
467 and unclassified repeats, and relatively fewer PLE, LTR and LINE-like retroelements (S3 Fig;
468 S8a Fig), with the biggest differences seen between *Rotaria* lineages (S8b Fig). However,
469 perhaps unsurprisingly given only two independent shifts in desiccation ability within our sample
470 (see phylogeny in Fig 2a), results from a Bayesian mixed-effects modelling approach that
471 controlled for phylogenetic relationships showed no significant correlations between desiccation
472 ability and TE load, for either overall proportion or for any individual TE superfamily (S4
473 Table). For most TE superfamilies the strength of the phylogenetic signal (λ) was close to 1 (S9
474 Fig), consistent with a high fit of the data to the phylogeny under a Brownian motion model as
475 would be expected if TE load evolves neutrally along branches of the phylogeny.
476

Two further comparisons of desiccating versus nondesiccating species among our wider sample of animals also present contrasting results. In chironomid midges, the desiccation-tolerant *P. vanderplanki* encodes substantially fewer TEs than its nondesiccating sister species *P. nubifer*, as predicted (0.8% and 2.2% respectively, although this rises to ~11% in both species when all repeats are included). In tardigrades, however, the desiccation tolerant *Ramazzottius varieornatus* encodes a greater proportion of TEs than *Hypsibius exemplaris* (4.3% and 2.8%, respectively), which does not survive desiccation without extensive conditioning [81], although the trend is reversed when all repeats are included due to a large fraction of simple repeats in *H. exemplaris*. We therefore find no consistent evidence for the hypothesised link between anhydrobiosis and TE load in bdelloids or beyond. The overall effect of desiccation on TEs might be dual: while repair of a DSB within a TE via non-homologous end-joining would likely result in its inactivation (thus acting to reduce TE load), an efficient DSB repair system would enhance repair of DSBs that arise during transposition of cut-and-paste DNA TEs that leave a DSB behind upon excision (thus allowing an increased TE load).

Bdelloids experience similar selective constraints on TEs as do other species. A fourth hypothesis is that the selective environment for TEs is different in bdelloids than in other animals, thereby shifting their TE profiles compared to simple theory. For instance, bdelloids might tolerate insertions within genes unusually well, owing to redundancy arising from tetraploidy or multiple gene copies [45,91,92]. First, we explored the genomic 'environment' of TE insertions and their potential effects on genome function. Differences in the location of TE insertions might reveal differential costs and benefits compared to other taxa. To do this, we first compiled a high-confidence list of class I retrotransposons by searching for proteins with significant similarity to the reverse transcriptase (RT) domain found in all retrotransposons. Phylogenies of the resulting alignments showed a diverse array of RTs in all species, most of them full-length (in terms of conserved subdomain presence) and clustered within the three primary retrotransposon superfamilies of PLEs, LTRs and LINEs (**Fig 5a**; S10 Fig; S5 Data). Many (but not all) clustered within families previously identified in *A. vaga*. The elevated LINE content in *D. carnosus* in comparison to other bdelloids is mostly due to the expanded Soliton clade and to the presence of CR1-Zenon and Tad/I/Outcast clades, the latter being characterized by the presence of the RNase H domain.

We then characterised surrounding genome features for these TEs. In 50 kb windows surrounding each class I TE identified above, we counted the occurrence and span of three features of interest: other (non-TE) genes, other (non-focal) TEs, and the telomeric repeat "TGTGGG" (identified from *A. vaga* [58] and supported in other rotifers, see S2 Note). Relative to a set of core metazoan (BUSCO) genes, the regions surrounding PLE, LINE and LTR TEs all showed significant decreases in gene density, but significant increases in both TE and sub-telomeric repeat density (**Fig 5b–d**; S5 Table). In contrast, using a linear mixed-effects modelling approach (see Materials and methods), there were no significant differences between the monogonont *B. calyciflorus* and bdelloids, or between desiccating and nondesiccating bdelloid species (S5 Table). These results are consistent with previous findings that TEs are mainly confined to sub-telomeric regions of bdelloid genomes [56], a bias that is presumably due to either selection against insertions at or near functioning genes and/or strong specificity in TE insertion site. Thus, it appears that most TE insertions are costly in bdelloid rotifers, as in other taxa, and that selection leads to their concentration outside of gene-rich regions.

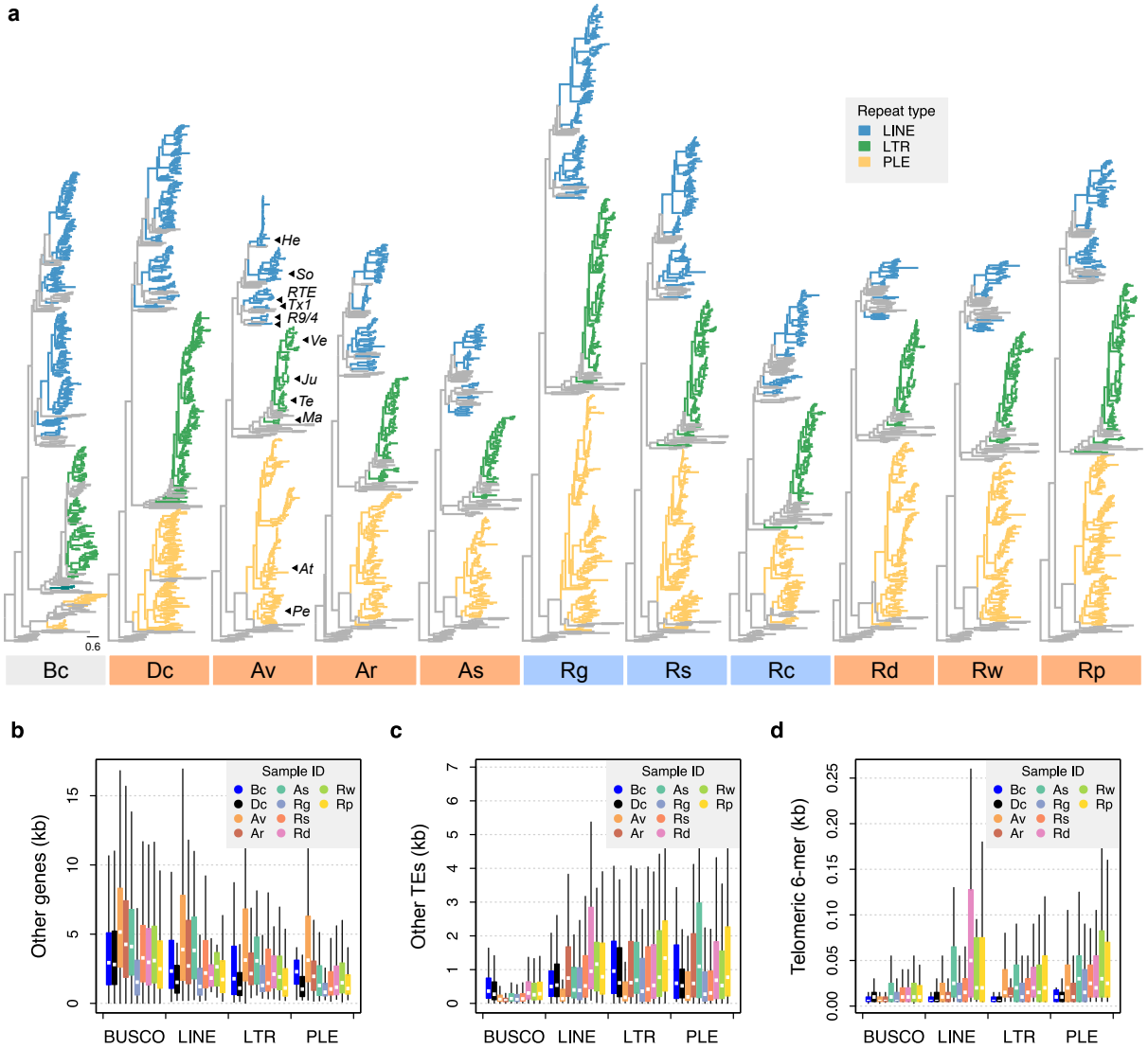

Fig 5. Phylogenetic diversity and genomic context of reverse transcriptase genes. (a) For each phylogeny, coloured branches represent identified rotifer-encoded RT copies and grey branches represent the core RT sequences from which the HMM was built (see Materials and methods and S9 Fig for core RT tree details). Colours indicated the major superfamilies. Previously characterised retrotransposons are indicated on the *A. vaga* tree (**He**, *Hebe*; **So**, *Soliton*; **RTE**, *RTE*; **Tx1**, *Tx1*; **R9/4**, *R9* and *R4*; **Ve**, *Vesta*; **Ju**, *Juno*; **Te**, *TelKA*; **Ma**, *Mag*; **At**, *Athena*; **Pe**, *Penelope*). All phylogenies are rooted on the branch separating the bacterial retrons. Scale bar represents 0.6 amino acid substitutions per site. Desiccating and nondesiccating species are indicated with orange and blue, as previously. Species codes: **Bc**, *B. calyciflorus* PSC1; **Dc**, *D. carnosus* DCAR706, **Av**, *A. vaga* Av2013; **Ar**, *A. ricciae* ARIC003; **As**, *A. steineri* ASTE805; **Rg**, *R. magnacalcarata* MAG3; **Rs**, *R. socialis* AK11; **Rc**, *R. macrura* Rc2018; **Rd**, *R. sordida* RSOR408; **Rw**, *R.* sp. 'Silwood-1' RSIL806; **Rp**, *R.* sp. 'Silwood-2' RPSE503. The genomic context in which RT genes reside is then described based on proximity to three other features: **(b)** 'other' genes (that do not overlap with any TE annotation), **(c)** other TEs, and **(d)** telomeric repeats ("TGTGGG"; that do not overlap with any coding region) as identified in *A. vaga*. For each plot, a 25 kb window is drawn around the focal RT gene and the total span (kb) of each feature within the window is counted, broken down per sample ID (coloured boxes) per TE superfamily (*X*-axis groups). Boxplots show the median (band), interquartile range (box) and minimum/maximum values (whiskers; outliers not plotted). The equivalent data for BUSCO genes (metazoan set) are also shown for comparison. The same representative samples are used in (b–d) as for (a).

As a second source of selective constraints, we tested for evidence of selection against ectopic recombination (ER). ER is argued to be a major cost of TEs in sexual taxa, but its effects derive from chromosomal abnormalities during meiosis, which should be lacking in bdelloids. Because the rate of ER increases with both the number of elements and their length [12], the strength of selection is expected to be strongest against longer TEs at higher copy number [9,24,93]. Two testable predictions arise: first, that bdelloids should have longer TEs than sexual taxa (under the hypothesis that ER is absent in bdelloids because of a lack of meiosis), and second, that nondesiccating bdelloids should have longer TEs than desiccating bdelloids (under the hypothesis that ER may still occur when chromosomes pair during the repair of DSBs). However, comparisons of TE length distributions provide no evidence for these predictions—there was no significant difference between monogononts and bdelloids, or between desiccating and nondesiccating bdelloids (**Fig 6a**; S6 Table; S6 Data). Thus, while the precise estimation of TE lengths will no doubt improve with increasing assembly contiguity, our current data do not show any indication of changes in TE length linked to asexuality (when compared to monogononts) or desiccation ability within bdelloids. It is interesting to note that this finding also applies to the PLEs, which include the *Athena* elements that comprise the unusually large and complex *Terminon* elements found at bdelloid telomeres [59,94]. In this case, it is possible that selection may still act on individual *Athena* elements but does not affect *Terminons* per se because their structural diversity and genomic location does not make them targets of ER.

A final prediction of selection against ER is that there should be a negative correlation between TE frequency and length, as is observed in *Drosophila* [24] and humans [93]. For both monogononts and bdelloids, the majority of identified TEs are short (< 1 kb), which are presumably partial matches or degraded copies. Nonetheless, we observe a sharp decline in copy number as mean TE length increases above ~0.5 kb, and a distinct lack of longer elements at higher copy numbers (**Fig 6b**). In vertebrates, previous work has suggested a lower threshold of ~0.6–1 kb under which ectopic recombination does not operate [93,95]. Thus, the observed patterns in rotifers are consistent with the hypothesis that longer elements above a certain length threshold are selected against more strongly due to the deleterious effects of ectopic recombination. This finding is contrary to a similar analysis performed on TEs in nematodes, which did not recover the expected relationship [18]. However, the pattern is the same in both desiccating and nondesiccating bdelloid representatives as well as the monogonont *B. calyciflorus*, and to some extent the acanthocephalan *P. laevis* (S11 Fig), suggesting that selection against longer TEs at higher copy number is a general feature in Syndermata, regardless of desiccation ability.

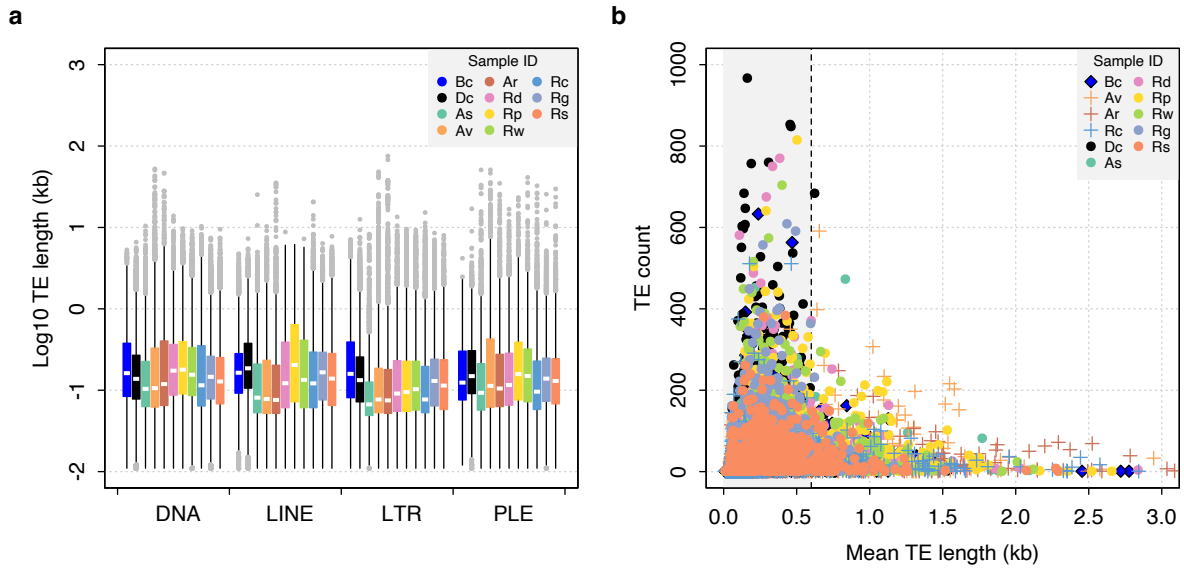

Fig 6. TE length dynamics. (a) Distribution of TE length for selected syndermatan samples decomposed into the major TE superfamilies (DNA transposons, LINE-like, LTR and PLE retrotransposons). Boxplots show the median (band), interquartile range (box) and minimum/maximum values (whiskers; outliers are shown in grey). Species codes: **Bc**, *B. calyciflorus* PSC1 (monogonont); **Dc**, *D. carnosus* DCAR706; **As**, *A. steineri* ASTE804; **Av**, *A. vaga* Av2013; **Ar**, *A. ricciae* Ar2018; **Rd**, *R. sordida* RSOR408; **Rp**, *R.* sp. 'Silwood-2' RPSE411; **Rw**, *R.* sp. 'Silwood-1' RSIL801 (desiccating bdelloids); **Rc**, *R. macrura* Rc2018; **Rg**, *R. magnacalcarata* MAG1; **Rs**, *R. socialis* AK11 (nondesiccating bdelloids). An equivalent plot including the acanthocephalan *P. laevis* is shown in S10 Fig. **(b)** Relationship between mean TE length per TE family (*X*-axis) and copy number (i.e., the number of TEs identified within each family; *Y*-axis). The same set of individuals are shown as for (a). A dashed line is drawn at 0.6 kb, given as the length threshold under which the rate of homologous ectopic recombination is negligible in mice.

Large expansion of TE silencing pathways in bdelloids. Having rejected a range of hypotheses to reconcile theory and observation of TE levels in bdelloid rotifers, we finally looked for expansions and/or diversifications in the molecular pathways that defend against TEs. We characterised copy number variation for three well-known gene families with direct roles in TE suppression via RNA interference (RNAi). (1) Argonaute proteins of both the Ago and Piwi subfamilies, the core effectors of RNAi gene-silencing that form complexes with various classes of small RNA [96,97]; (2) Dicer, an RNase III–family protein that cleaves double-stranded RNA (dsRNA) molecules from 'target' genes into shorter fragments that are subsequently incorporated into Argonaute complexes [98,99]; and (3) RNA-dependent RNA polymerase (RdRP), an RNA replicase that synthesises secondary small interfering RNAs (siRNAs) that amplify the silencing response [98,100].

Based on hidden Markov model (HMM) matches of key domains to the predicted proteomes of the Illumina maxhap assemblies (in which homologous copies are largely collapsed), we detect an average of 21.5 putative Argonaute copies, 3.9 Dicer copies and 37.3 RdRP copies in bdelloid genomes (**Fig 7a**; S4 Data). These expansions are substantially larger than previously uncovered from the diploid assembly of *A. vaga* (eight Dicers, 23 Ago/Piwi, 20 RdRP per diploid genome) [44], particularly after correcting for the assembly resolution (i.e. diploid vs haploid assembly; see Materials and methods), perhaps due to the increased sensitivity of the HMM-based approach, or to different degrees of pseudogenization and/or homolog collapse. Phylogenies of identified copies revealed a number of divergent clades, particularly in the Argonaute and RdRP families (**Fig 7b–d**; S12 Fig; S7 Data; S8 Data), that might indicate both expansion and

622 diversification of these proteins. Even accounting for degenerate tetraploidy in bdelloids, the
623 RdRP domain in particular is significantly expanded relative to monogononts and other
624 eukaryotes (**Fig 7e–g**; S9 Data). This expansion is not found in the monogononts *B. calyciflorus*
625 or *B. plicatilis* HYR1, nor is there evidence for it in the (unannotated) acanthocephalan genome
626 (S4 Data), suggesting that the expansions seen in Ago and (particularly) RdRP are unique to
627 bdelloids, including the basally divergent *D. carnosus*.
628
629 The expansion of bdelloid-specific Ago genes (denoted 'BDAGOs' in **Fig 7**) is comparable to
630 the 'worm-specific' Ago clade (WAGOs) found in nematodes [101–103]. Intriguingly, the
631 majority of nematode species (although not *C. elegans*) have apparently lost their Piwi-like
632 orthologue, the branch of the Argonaute family that is usually involved in TE suppression in
633 most other animals [104], but instead mediate TE silencing using a Dicer/RdRP pathway [105].
634 The primary function of RdRP is to amplify RNAi responses via the production of small
635 interfering RNAs (siRNAs) [98], but this family is found only rarely in animal genomes [100]. In
636 *C. elegans*, RdRP-generated siRNAs (known as '22G-RNAs') also play important roles in the
637 recognition of 'self' versus 'non-self' RNA and multigenerational (i.e., inherited) epigenetic
638 memory [103,106]. However, an analysis of TE variation across the nematode phylum found no
639 effect of RNAi pathway differences (in terms of presence/absence of RNAi genes) among
640 species, concluding that TE content is mediated primarily by the stochastic action of genetic drift
641 [18].
642
643 Why do bdelloids possess such a marked expansion of gene silencing machinery? One
644 explanation may be that it was required as part of the successful long-term transition to
645 asexuality, if other mechanisms usually acting in sexual populations were no longer operating.
646 Intriguingly, it has been shown in *A. vaga* that piwi-interacting small RNAs (piRNAs) target
647 both TEs and putatively 'foreign' genes (i.e. non-metazoan genes gained via horizontal gene
648 transfer [HGT]) [107], the latter of which are unusually frequent in bdelloid genomes
649 [44,46,108–110]. Thus, it may be the case that bdelloids have an enhanced RNAi system to
650 defend against invasion from horizontally transferred TEs, particularly if the level of exposure or
651 rate of import is higher relative to other animals. Furthermore, the ability of some RNAi
652 pathways to distinguish self from non-self may be needed to maintain genome integrity over
653 longer timescales, if such mechanisms are operating. Further work on the precise functions of the
654 divergent Ago and RdRP clades is required to explore these possibilities.
655
656

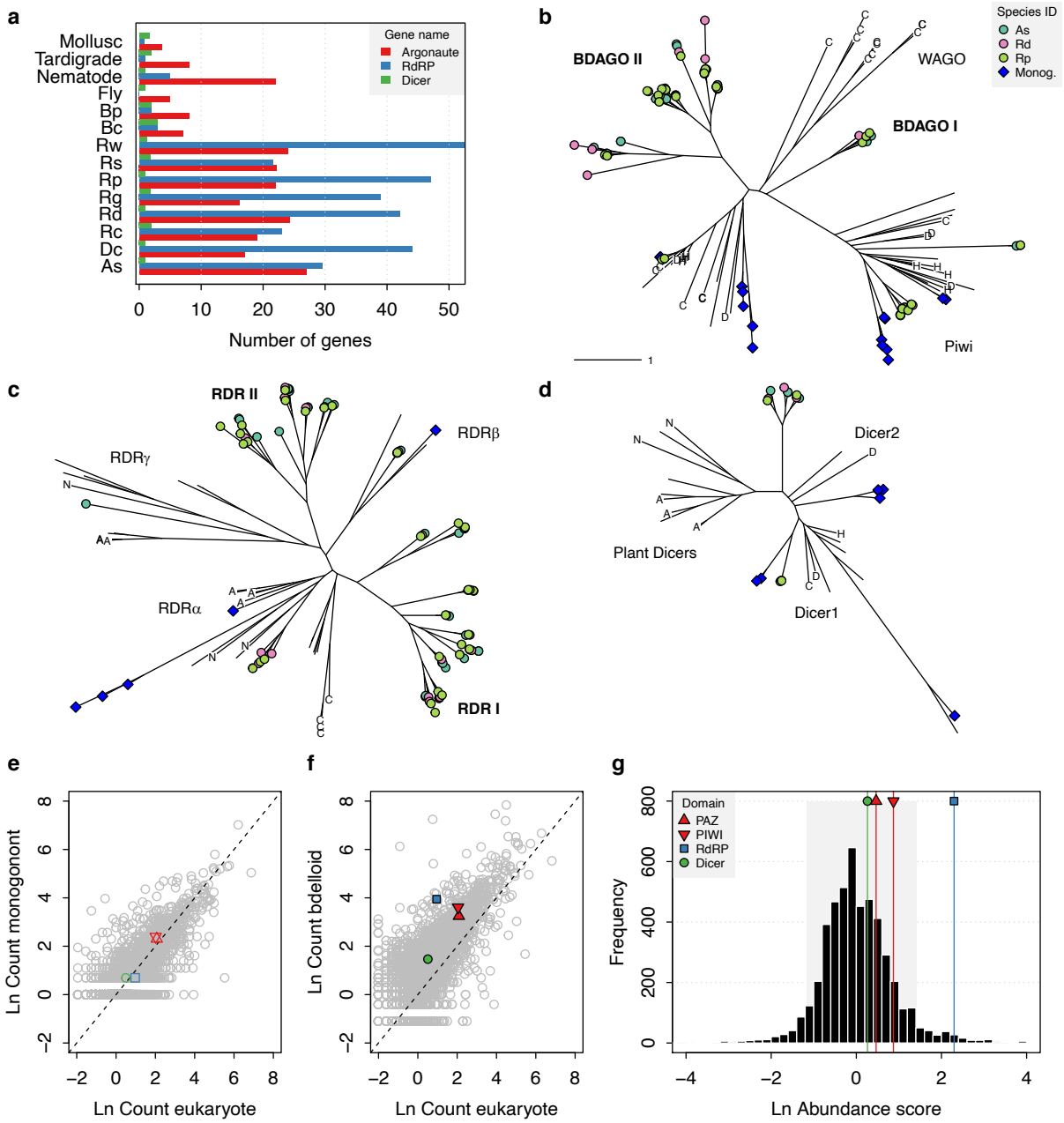

Fig 7. Expansion of TE silencing pathways in bdelloid rotifers. (a) Copy number variation for RNAi gene families Argonaute (Ago/Piwi, red), RNA-dependent RNA polymerase (RdRP, blue) and Dicer (green) in bdelloids compared to other protostome groups. Proteins are identified based on the presence of key identifying domains (see Materials and methods). Species codes for rotifers: **Bc**, *B. calyciflorus*; **Bp**, *B. plicatilis* HYR1; **Dc**, *D. carnosus*; **As**, *A. steineri*; **Rg**, *R. magnacalcarata*; **Rs**, *R. socialis*; **Rc**, *R. macrura*; **Rd**, *R. sordida*; **Rw**, *R.* sp. 'Silwood-1'; **Rp**, *R.* sp. 'Silwood-2'. Maximum likelihood unrooted phylogenies are then shown for **(b)** Argonaute, **(c)** RdRP and **(d)** Dicer gene copies identified in *A. steineri*, *R. sordida* and *R.* sp. 'Silwood-1' 10x haploid ('pseudohap') assemblies, aligned with orthologs from representative species from across the eukaryotes. Blue symbols indicate copies identified in the monogonont *B. plicatilis*, and letters on tips show selected reference species to aid visual orientation: 'C', *C. elegans*; 'H', human; 'D', *D. melanogaster*; 'N' *N. crassa*; 'A', *A. thaliana*. Some clade names are also shown where relevant; 'WAGO' indicates the worm-specific cluster of Ago genes in the Argonaute phylogeny. 'BDAGO' I and II and 'RDR' I and II indicate putative bdelloid-specific clades of Argonaute and RdRP proteins, respectively. **(e)** Comparative protein-domain abundance plot. Each point represents a Pfam domain ID, with (log) average abundance (i.e., count) in the reference eukaryote set shown on the *X*-axis and (log) abundance in the monogonont *B. plicatilis* on the *Y*-axis (see Materials and methods). The positions of the PAZ, PIWI (red up/down triangles), RdRP (blue square) and Dicer (green circle) domains are highlighted. Dashed line indicates the 1-1 relationship. **(f)** Equivalent plot for bdelloids,

where the *Y*-axis shows the (log) average abundance for the *A. steineri*, *R. sordida* and *R.* sp. 'Silwood-1' 10x haploid assemblies. Note that the average abundance for all Pfam entries is shifted above the 1-1 line due to the ancient genome duplication in all bdelloids, such that many genes are found in double copy (i.e., homoeologs) even in a 'haploid' representation. **(g)** Comparative protein–domain abundance plot for bdelloids versus eukaryotes (see Materials and methods). Entries to the right of the distribution are overrepresented in bdelloids with respect to eukaryotes. The shaded area represents the 5% and 95% quantiles of the distribution, and the scores for the PAZ, PIWI, Dicer and RdRP domains are indicated (see legend).

Conclusions

We show that all bdelloids encode a rich diversity of TEs from both class I (retroelements) and class II (DNA transposons), and thus reject the idea that bdelloids are deficient or unusual in their TE content or diversity. Moreover, a substantial fraction of these elements has been active relatively recently within populations. This finding is at odds with the original predictions of population genetic theory for TEs in asexuals. One possible resolution is that theory is missing some component or assumption. It is possible that parameter space exists that permits intermediate levels of TEs in an asexual population, perhaps sustained by high rates of horizontal transfer, which there is evidence for in bdelloids. The HGT idea has some support from the increased prevalence of class II DNA transposons in bdelloids, given their greater propensity for horizontal transfer, but we found that the basally divergent lineage leading to *D. carnosus* had a profile more like the monogonont rotifers with regard to non-LTR retrotransposon abundance, despite sharing other features with the rest of bdelloids. Alternatively, some TEs might have been co-opted to provide beneficial functions, which is hypothesised to explain the unusually large and complex *Terminon* repeats. Other TEs may have evolved strong site-specificity to neutral genome regions to mitigate negative effects of transposition. This idea is supported by the preference shown for insertions into gene-poor regions that are probably at or near the telomeres, although it seems unlikely that the full complement of bdelloid TEs have accumulated in this way.

We also ask what forces may be acting to suppress TE activity in bdelloid populations. A major finding is that an abundance and diversity of RNAi gene silencing pathways, characterised by a large expansion of Argonaute and RdRP genes, appears to be a unique feature of bdelloid genomes. The precise origins and functions of these divergent Ago and RdRP clades are yet to be elucidated, but it seems likely that such an extended arsenal of TE defence genes offers enhanced protection against the deleterious effects of TE activity, particularly if bdelloid populations cannot keep TEs in check through sexual processes but are still exposed to new invasions via HGT. This enhanced RNAi system may then provide a more deterministic action against TEs, as opposed to stochastic forces (such as genetic drift) that may predominate in other animal groups and may potentially explain the greater uniformity of TE proportions in bdelloids relative to monogononts.

An alternative resolution is that the assumption of no recombination and strict clonality is not met in bdelloids. Previous work, for example, proposed that intra-individual recombination during the repair of DSBs caused by desiccation could provide a mechanism to keep TE numbers in check. We found no evidence here that overall TE loads were lower in desiccating species than nondesiccating species, a finding not limited to bdelloids, or for differences in activity or rate of turnover between them. It remains possible that an equivalent mechanism, such as mitotic recombination, or unusual DNA repair mechanisms operating in nondesiccating species as well, could still act to limit TE proliferation and facilitate elimination of inserted copies, maintaining

728 an equilibrium between TE spread and removal by excision and selection. Finally, there could be
729 some hidden mechanism of inter-individual recombination that facilitates TE removal. We found
730 no evidence for its action here, but further work is needed for a final answer on the conundrum
731 of bdelloid asexuality.
732
733
734 ## Materials and methods
735
736 **Rotifer sampling and culture.** For most samples, individual rotifers were collected from
737 permanent and temporary freshwater habitats around Imperial College London's Silwood Park
738 campus, Ascot, UK between May 2015 and February 2019. Three samples (*R. magnacalcarata*
739 RM9 and RM15, and *R. socialis* RS1) were collected from a freshwater spring in Fontaneto
740 d'Agogna, Italy in 2016 (see S1 Data for details). Although we focused on the genera *Adineta*
741 and *Rotaria*, we also included two individuals from the desiccation-tolerant species
742 *Didymodactylos carnosus.* Preliminary phylogenetic data had identified this as a distant outgroup
743 to the focal genera, useful in rooting phylogenetic trees and as a further independent datapoint to
744 test the generality of conclusions about bdelloids. A total of 26 samples were submitted for
745 single-individual, whole genome sequencing; for these, DNA was extracted using either a Chelex
746 preparation (Bio-Rad InstaGene Matrix) or a QIAamp DNA Micro Kit (Qiagen), and whole-
747 genome amplified using a REPLI-g Single Cell kit (Qiagen) before sequencing on either
748 Illumina NextSeq500 at the Department of Biochemistry, University of Cambridge (Cambridge,
749 UK), or Illumina HiSeq X at Edinburgh Genomics, University of Edinburgh (Edinburgh, UK).
750 For *A. ricciae* ARIC003, DNA was extracted from ~200 animals descended from a single
751 individual before whole-genome amplification. For *B. calyciflorus* PSC1, individuals for DNA
752 extractions were derived from an individual isolate from a laboratory stock population previously
753 isolated from field-collected resting eggs [111]. DNA was extracted from ~5000 starved
754 individuals using a phenol-chloroform protocol and sequenced on the Illumina NextSeq500 at
755 the Max Planck Institute for Evolutionary Biology. Three 10x Genomics Chromium 'linked
756 reads' libraries were generated for *A. steineri*, *Rotaria* sp. 'Silwood-1' and *R. sordida*; for these,
757 high molecular weight DNA was extracted from thousands of animals reared clonally from a
758 single wild-caught animal, without whole-genome amplification, using the Chromium
759 Demonstrated Protocol "HMW gDNA Extraction from Single Insects"
760 (https://support.10xgenomics.com/permalink/7HBJeZucc80CwkMAmA4oQ2). Linked-read
761 libraries were constructed at the Centre for Genomics Research, Liverpool, UK, before
762 sequencing on the HiSeq X at Edinburgh Genomics. Further details on rotifer sampling, DNA
763 extraction and sequencing are provided in S1 Data.
764
765 **Biological replicates.** To check the repeatability of the whole-genome amplification (WGA),
766 sequencing, assembly and analysis pipelines, we included several samples that were either
767 biological replicates of the same rotifer clone, or where high-quality genomes were available for
768 the same clone from unamplified source material. Specifically, for *Rotaria* sp. 'Silwood-2' we
769 isolated two consecutive offspring from the same wild-caught mother and conducted WGA,
770 sequencing, assembly and analysis for these sisters independently (as Rp_RPSE411 and
771 Rp_RPSE503). From the same clonal laboratory line of *Rotaria* sp. 'Silwood-1' that was used
772 for 10x Genomics DNA preparation, we isolated two more individuals and processed each
773 independently using the WGA workflow (as Rw_RSIL801 and Rw_RSIL802). Finally, we
774 applied the WGA method to DNA from *A. ricciae*, for which a previous assembly was available
775 from unamplified DNA [46] on the same clonal culture and included this replicate in
776 downstream analyses alongside the earlier reference assembly.
777

Data filtering and genome assembly. We generated two assembly versions for each of the single-individual rotifer samples. The 'reference' assemblies were scaffolded and polished to result in haploid assemblies with improved contiguity. The 'maximum haplotig' ('maxhap') assemblies instead retained highly similar contigs that might otherwise be removed during assembly polishing. Our pipeline is outlined as follows.

For the Illumina libraries, raw sequence data were filtered for low quality bases and adapter sequence using BBTools v38.73 'bbduk' (https://sourceforge.net/projects/bbmap/), and error corrected using BBTools 'tadpole'. Data quality was inspected manually using FastQC v0.11.5 [112] aided by MultiQC [113] visualisation. For the *A. steineri*, *R.* sp. 'Silwood-1' and *R. sordida* linked-read libraries, data were assembled into haploid ('pseudohap') and diploid ('megabubbles') genome representations using the 10x Genomics proprietary software Supernova v2.1.1 [114] and further scaffolded with ARKS v1.0.4 [115]. All raw sequencing data are deposited in the relevant International Nucleotide Sequence Database Collaboration (INSDC) databases under the Study ID PRJEB39843 (see S1 Data for run accessions and counts for raw and filtered data).

For the single-individual samples, an initial assembly was generated using SPAdes v3.13.0 [116] with default settings. Contaminating reads from non–target organisms, identified based on aberrant GC content, read coverage and/or taxonomic annotation, were then identified and removed using BlobTools v1.1.1 [117,118]. For *R. magnacalcarata* and *R. socialis* samples, resultant haplotigs were then collapsed using Redundans [119] with default settings before scaffolding and gap filling with SSPACE v3.0 and GapCloser v1.12 respectively [120,121]. For *A. steineri*, *R.* sp. 'Silwood-1' and *R. sordida* single-individual samples, the scaffolding step was performed with RaGOO v1.1 [122,123], using the matching 10x Genomics 'pseudohap' assembly as a reference (contigs from *R.* sp. 'Silwood-2' were scaffolded using the *R.* sp. 'Silwood-1' 10x reference), specifying the '-C' parameter to prevent concatenation of unaligned contigs. Scaffolded assemblies were subjected to further rounds of BlobTools to remove any additional sequences derived from non-target organisms. These assemblies were designated the reference set described above.

For the maxhap assemblies, filtered fastq files were first generated by mapping the original (trimmed and error-corrected) sequencing reads to each reference genome, using the 'outm=filtered_R#.fq' functionality of BBTools 'bbmap', and then reassembled with SPAdes, increasing the final kmer value to 121. Assembly metrics were summarised using 'calN50.js' (https://github.com/lh3/calN50), which reports the 'expected scaffold size' (AU) as an alternative metric of assembly contiguity that is less biased than N50 (defined as the area under the cumulative genome span versus contig length graph, equivalent to the expected scaffold size for a randomly chosen assembly location [124]). Gene-completeness scores for core eukaryotic (n = 303) and metazoan (n = 978) genes were calculated for all assemblies using BUSCO v3.0.2 [125] with default settings. Reference and maxhap assemblies for *B. calyciflorus* PSC1 and *D. carnosus* are the same, due to a lack of appropriate data for scaffolding.

Gene prediction. Gene prediction was performed on reference assemblies using one of three approaches, depending on the availability of RNA-seq data. For *B. calyciflorus*, *A. ricciae*, and all *R. magnacalcarata*, *R. socialis*, and *R. sordida* assemblies, published RNA-seq data [109,110,126] were downloaded from NCBI Sequence Read Archive (SRA), quality-trimmed using BBTools 'bbduk' with default settings and aligned to the genomic scaffolds using STAR v2.7.3a [127] with the option '--twoPassMode Basic'. Aligned BAM files were then provided to BRAKER v2.1.2 [128–132] with default settings for gene prediction. For *A. steineri*, *R.* sp.

828 ‘Silwood-1’ and *R.* sp. ‘Silwood-2’ assemblies, RNA-seq data from a related species (*A. ricciae*
829 and *R. magnacalcarata* respectively) were used instead, aligned using BBTools ‘bbmap’ with the
830 options ‘maxindel=200k minid=0.5’, before gene prediction with BRAKER as above. Finally,
831 for the distantly related *D. carnosus*, BRAKER was run using gene-model parameters estimated
832 from BUSCO analysis of the genomic scaffolds. The quality of predicted proteins was assessed
833 using BUSCO in protein mode. Intragenomic divergence between homologous gene copies and
834 collinearity was calculated as for Nowell *et al.* (2018) and described in S1 Note. Genome
835 assemblies and gene predictions were converted to EMBL format using EMBLmyGFF3 v2
836 [133], and are deposited at DDBJ/ENA/GenBank under the Study ID PRJEB39843 (see **Table 1**
837 and S1 Table for individual GenBank accessions).
838
839 **Rotifer phylogeny.** Evolutionary relationships among new genomes and published genomes of
840 rotifers were determined using a core-genome phylogenomics approach based on the BUSCO
841 eukaryotic gene set. For genomes from species with very high intragenomic homologous
842 divergence (*A. ricciae* and *A. vaga*), redundancy was removed by selecting the copy with the
843 highest BUSCO score for all BUSCO genes with multiple copies, using the script
844 BUSCO_collapse_multicopy.pl’ (https://github.com/reubwn/scripts). One-to-one co-orthologs
845 found in at least 95% of the samples were then identified using the script
846 ‘BUSCO_phylogenomics.py’ (https://github.com/jamiemcg/BUSCO_phylogenomics). Protein
847 sequences were aligned using Clustalo [134] and concatenated in Geneious R9 [135]. The full
848 alignment was checked by eye and sections with ambiguous alignment within the bdelloid clade
849 were removed across all sequences to avoid aligning potential paralogs or homoeologs.
850 Translation errors arising from annotation issues in specific bdelloid genomes were identified by
851 obvious mismatches to the consensus of closely related genomes, and the affected residues were
852 deleted in the affected genome only. Potential alignment issues within the monogonont clade
853 were less obvious owing to the substantial genetic divergence from bdelloids and the smaller
854 number of genomes and replicates, so corrections were less stringent. A maximum-likelihood
855 phylogeny was then estimated using IQ-TREE v1.6.12, with automatic model selection
856 (VT+F+I+G4) [136,137]. Branching support was assessed using SH-aLRT and ultrafast
857 bootstrap sampling (‘-alrt 1000 -bb 5000’) [138,139].
858
859 **Repeat annotation and TE dynamics.** TEs and other repeats were identified using the
860 RepeatModeler and RepeatMasker pipelines. For each sample, a *de novo* repeat library was
861 generated directly from the assembled nucleotides using RepeatModeler2 [140] and combined
862 with a database of 12,662 protostome repeats from Repbase v23.08 [141] and 278 additional TEs
863 manually curated from the *A. vaga* genome [44]. Repeats and TEs were then detected and
864 classified using RepeatMasker v4.1.0 [84], and resultant outputs were post-processed using the
865 ‘One code to find them all’ Perl script [142]. The breakdown of TE superfamilies in the final
866 database was 4,145 DNA transposons (including 300 rolling circles), 5,523 LTRs, 2,583 LINEs
867 (including SINEs), 227 PLEs, and 165 simple or low-complexity repeats. TE content (expressed
868 as a proportion of genome size) was mapped onto the phylogeny using ‘contMap’ in the Phytools
869 v0.6-99 package in R v3.6.0 [143,144]. There is no module for the detection of class II MITEs in
870 RepeatMasker; for these, the separate program Generic Repeat Finder (GRF) was run using
871 default parameters. TE dynamics were investigated by constructing Kimura 2-parameter
872 divergence [85] landscapes using the utility scripts in the RepeatMasker package and plotted
873 using custom scripts (see below). Selected assemblies were also submitted to the REPET v2.5
874 ‘TEdenovo’ [145,146] TE detection and annotation pipeline with default parameters, for
875 comparison. In addition, for *D. carnosus* and *R. sordida* (using 10x Genomics) reference
876 assemblies, we increased the parameter ‘minNbSeqPerGroup’ from 3 to 5 to evaluate
877 contribution from tetraploid genes, which was judged to be negligible. Although

878 REPET*denovo* TE consensus sequences are automatically classified using Wicker's TE
879 classification [147], RepeatMasker was additionally applied for further TE classification,
880 detection and landscape divergence plot building.
881
882 The presence or absence of specific LTR retrotransposon (LTR-R) insertions in our population
883 data was inferred using a read-mapping approach. Specifically, the presence of a given insertion
884 was scored based on the alignment score of the 'best' read that mapped continuously and
885 contiguously across the LTR-genome boundary. First, full-length LTR-Rs (i.e. those with
886 annotated 5' and 3' LTR regions) were identified from each reference assembly using
887 LTR_retriever v2.8 [148]. Three filters were then applied to remove false positives. Candidates
888 that showed an overlap with a predicted gene in the 5' or 3' LTR itself or an 'N' base within 150
889 bases upstream or downstream of its genomic location that might indicate local mis-assembly
890 were removed. Candidates also required supporting evidence of LTR homology from a separate
891 RepeatMasker annotation of the reference assembly. For each remaining LTR-R, a library of
892 'LTR-tags' was then generated by extracting a 100 bp sequence that spanned 50 bases into the
893 genomic (i.e., non-TE) region of the insertion site from both the 5' and 3' terminal repeated
894 regions. Thus, each pair of 'LTR-tags' represents an insertion of a particular LTR into a specific
895 location in the focal genome, and a score is calculated based on the alignment information
896 contained in the CIGAR string of the 'best' read (i.e. with the highest number of alignment
897 matches) from the SAM mapping file: $S_i = ((M_{Li} - X_{Li}) + (M_{Ri} - X_{Ri}))/200$, where M_{Li} is the
898 number of alignment matches for the left-hand tag for LTR i, penalised by the number of
899 mismatches X_{Li}, with equivalent scoring for the right-hand tag. Since tag length is 100 bases, the
900 maximum score for a perfect alignment is 200, or 1 after normalisation. The number of mapped
901 reads is also recorded to provide an estimate of coverage (but note that S_i is taken from the best
902 read only). Sequencing reads from all single-individual rotifer samples were aligned to the
903 filtered LTR-tag set using BBTools 'bbmap' with the parameters 'minid=0.5 local=t' and scored
904 using the above system. Because orthologous LTR-Rs may be identified from searches started in
905 different genomes, we identified these cases by reconstructing the phylogeny of the LTR-tags
906 and any with pairwise sequence divergence less than 0.1 were collapsed to yield a condensed
907 final matrix.
908
909 The LTR-tag case-study in Fig. 4b was selected for closer investigation in the draft assemblies
910 after consideration of several examples, because it illustrates variability for an element insertion
911 site within a species and indicates that Class I TEs can insert in coding regions, with potential
912 fitness consequences. The LTR-tags were mapped to the RM15 draft assembly using Geneious
913 Prime v2020.1.2 [135], and were found to match an element annotated by LTR_retriever,
914 containing four predicted genes. In the RM9 draft assembly, only the left-hand tag was mapped,
915 as the scaffold ended before the inserted element was fully assembled. For the same reason, the
916 element in RM9 had not been annotated as such by LTR_retriever, but the sequence is nearly
917 identical (99.7%) to the insertion in RM15 along its aligned length (except that the annotations
918 predicted three element-associated genes rather than four). The scaffolds were trimmed to the
919 focal gene and aligned, and the region was used as a BLASTn query against local databases for
920 two other *R. magnacalcarata* reference genomes where the LTR-tag was absent: MAG1 and
921 Rg2018. In each case, this provided the location of a closely similar but uninterrupted copy of
922 the focal gene, although the annotations of the gene's structure differed slightly among genomes.
923 These scaffolds were trimmed and aligned against the copies from RM9 and RM15, using the
924 Geneious alignment tool with default settings, except that the gap extension penalty was reduced
925 from 3 to 0.2 to enable the algorithm to handle the element insertion. Local features were
926 manually reannotated to illustrate the interpretation provided in the text. To investigate the
927 potential function of the interrupted gene, the copy from MAG1 (g37061) was translated and

928 used as a BLASTp query against the NCBI RefSeq Protein Database [149]. A region of
929 approximately 1000 residues was found to have weak similarity (~25% pairwise identity) to
930 proteins annotated as midasins, from a range of eukaryotes. As a final step, the intact gene from
931 MAG1 was used as a BLASTn query against the full draft genomes of RM9 and RM15, which
932 revealed a separate scaffold in each case, containing a partial copy of the gene in which the
933 coding sequence was intact across the junction spanned by the LTR-tag, and the element
934 insertion was absent.
935
936 **Recombination analyses.** We tested sexual versus clonal patterns of variation in LTR presence
937 and absences. First, we calculated consistency indices (CI) with parsimony reconstruction of the
938 binary matrix. LTR-tags with scores > 0.875 were coded as present (i.e. no more than half of the
939 genome context or LTR region from both left and right LTR-tags was missing) and < 0.875
940 coded as absent (alternative thresholds led to the same qualitative results). A CI = 1 indicates
941 perfect nesting with no homoplasy, whereas a score less than 1 is expected if variation is shuffled
942 among loci and not tree-like. Next, we calculated the index of association and ran permutations
943 to test for significant linkage disequilibrium of the LTR-tag data relative to a null model of
944 random shuffling (expected in a fully outcrossing sexual population). We used the modified
945 index of association by Agapow and Burt (2001) [150] that corrects for an effect of the number
946 of loci on the index, and ran permutations using the 'ia' function in the Poppr v2.8.5 library
947 [151] in R. Data were coded as diploid and codominant presence/absence data (because of the
948 lack of diploid assemblies in the population-level data). Finally, for *R. magnacalcarata* and *R.*
949 *socialis* we ran simulations with the FacSexCoalescent simulator of Hartfield et al. (2018) [152]
950 to generate 50000 datasets with the same number of individuals and sampled binary loci as
951 observed, but with frequencies of sexual versus asexual reproduction within the populations
952 varying from 10^{-7} (i.e. negligible) to 1 (i.e. obligate sexual). We estimate the posterior
953 distribution of the frequency of sex for our observed samples using Approximate Bayesian
954 Computation on the simulated datasets implemented in the abc package in R.
955
956 **Reverse transcriptase survey.** A hidden Markov model (HMM) approach was used to survey
957 the predicted rotifer proteomes for proteins encoding the reverse transcriptase (RT) domain
958 (Pfam ID PF00078). First, a HMM was constructed from an alignment of 51 RT domains from
959 across the tree of life [57], supplemented with 67 bdelloid-specific retroelements
960 [44,57,58,61,63,64] (S9 Fig). Alternative transcripts were first removed from predicted
961 proteomes and proteins with a significant match (*E*-value ≤ 1e−5) were identified and inserted
962 into the core RT alignment using HMMER v3.2.1 'hmmsearch' and 'hmmalign', respectively
963 (http://hmmer.org/). Maximum likelihood phylogenies were then constructed using IQ-TREE as
964 above, specifying the root of the phylogeny to be on the branch leading to the bacterial retrons
965 [57]. Trees were manipulated using FigTree v1.4.4 (https://github.com/rambaut/figtree),
966 colouring the identified RT-encoding rotifer proteins based on their phylogenetic position. The
967 span accounted for by genome 'features' (other genes, other TEs and telomeric repeats) in a 25
968 kb window around each identified RT-containing protein was summarised using BEDTools
969 v2.29.2 'intersect' and 'groupby' [153]. Other genes were counted as predicted coding regions
970 that did not overlap with any TE annotation. Genomic locations of the telomeric hexamer
971 'TGTGGG' [58] were identified using EMBOSS 'fuzznuc' [154], excluding any hexamer that
972 overlapped with a predicted CDS. Note that the telomeric repeat for *Brachionus* is not known,
973 but the sequence above was among the most frequent G-rich hexamers identified in the PSC1
974 genome (see S2 Note).
975
976 **TE silencing machinery survey.** A similar HMM based approach was used to characterise gene
977 copy evolution of three key pathways involved in RNAi gene-silencing. Putative Argonaute

proteins were identified based on the presence of both the PAZ and PIWI domains (Pfam IDs PF02170 and PF02171 respectively), putative Dicer proteins were identified based on the presence of both PAZ and Dicer (PF03368) domains, and putative RdRP proteins were identified based on the presence of the RdRP domain (PF05183). Stockholm files were downloaded from Pfam [155] and aligned to the proteomes using HMMER (*E*-value ≤ 1e−5) as above. Reference proteomes from a selection of eukaryotic species to represent the diversity and distribution of Argonaute, Dicer and RdRP proteins were downloaded (June 2020) from UniProt and subjected to the same procedure: *Arabidopsis thaliana* (UP000006548), *Oryza sativa* (UP000007015), *Neurospora crassa* (UP000001805), *Schizosaccharomyces pombe* (UP000002485), *Laccaria bicolor* (UP000001194), *Dictyostelium discoideum* (UP000002195), *D. melanogaster* (UP000000803), *C. elegans* (UP000001940), *H. exemplaris* (UP000192578), *H. robusta* (UP000015101), *L. gigantea* (UP000030746), *S. haematobium* (UP000054474), *B. plicatilis* (UP000276133), *Branchiostoma floridae* (UP000001554) and *Homo sapiens* UP000005640. Proteins were aligned using either 'hmmalign' from the HMMER package or Clustalo, and ML phylogenies were constructed using IQ-TREE as above.

Comparative protein–domain abundance plots were constructed using counts of Pfam entries parsed directly from InterProScan5 [156] annotation of predicted proteomes. The 'abundance score' was computed as the (log) ratio of domain counts in bdelloids divided by the domain counts in eukaryotes, corrected for inflation in bdelloids due to the ancient whole-genome duplication by dividing the former by two. This correction is likely to be conservative, since many loci have lost one branch of the ancient duplication (i.e. tetraploidy is degenerate). To check that the putative RdRP expansion was indeed eukaryotic in origin, rather than viral, the HMMs for four viral RdRP families (PF00680, PF00978, PF00998 and PF02123) were downloaded from Pfam and submitted to the same search protocol, with zero hits to bdelloid proteomes recorded.

Statistical analyses. To assess differences in TE content between desiccating and non-desiccating rotifer species, we ran Bayesian linear mixed-effects models of TE content (as a percentage of genome span) including desiccation as a two-level fixed factor and sample ID as a random intercept term. The BUSCO gene phylogeny was used to account for non-independence among species. A separate model was run for each TE superfamily (including 'unclassified' TEs) as well as for all TEs combined. Inverse-Wishart priors were used for the random and residual variances, and models were run for 42,0000 iterations with a burn-in of 20,000 and a thinning interval of 200. This resulted in 2,000 stored samples of the posterior with minimal autocorrelation in all cases (< 0.2) [157]. Models were run using the MCMCglmm v2.29 [158] package in R. The phylogenetic signal, defined as the proportion of the total variance in TE content attributable to the phylogeny [159], was estimated from the MCMCglmm model output using the formula: $\lambda = \sigma P^2/(\sigma P^2 + \sigma R^2)$.

The density of genomic features surrounding BUSCO genes versus RT-containing genes were compared using linear mixed effects models with species as a random effect and gene class (BUSCO, PLE, LINE or LTR) as a fixed effect, run using lme4 [160] v1.1-21 in R. TE length distributions were compared using the same approach, with species as a random effect and binary factors specifying monogononts versus bdelloid, and desiccating versus nondesiccating species.

Code availability. All TE analysis scripts used in this study are available at https://github.com/reubwn/te-evolution.

Acknowledgements

Genome sequencing was performed by the UK Natural Environment Research Council (NERC) Biomolecular Analysis Facility at the Centre for Genomic Research (CGR) at the University of Liverpool (NBAF-Liverpool) and the DNA Sequencing Facility in the Biochemistry Department at the University of Cambridge. The authors wish to thank the following: Christiane Hertz-Fowler, Pia Koldkjær and John Kenny (CGR), Shilo Dickens and Nataliya Scott (Cambridge). Matthew Arno and Colin Sharp (Edinburgh), and Steven Van Belleghem (University of Puerto Rico) for support with the planning and execution of various aspects of genome sequencing and/or assembly, Tom Smith and Anita Kristiansen for rotifer sampling, and Mike Tristem for helpful discussions on detecting LTR polymorphisms.

References

1. Orgel LE, Crick FH. Selfish DNA: the ultimate parasite. Nature. 1980;284: 604–607. doi:10.1038/284604a0
2. Hickey DA. Selfish DNA: a sexually-transmitted nuclear parasite. Genetics. 1982;101: 519–531.
3. Charlesworth B, Charlesworth D. The population dynamics of transposable elements. Genet Res. 1983;42: 1–27. doi:10.1017/S0016672300021455
4. Doolittle RF, Feng D-F, Johnson MS, McClure MA. Origins and evolutionary relationships of retroviruses. Q Rev Biol. 1989;64: 1–30.
5. Eickbush TH, Malik HS. Origins and Evolution of Retrotransposons. In: Craig NL, Lambowitz AM, Craigie R, Gellert M, editors. Mobile DNA II. American Society of Microbiology; 2002. pp. 1111–1144. doi:10.1128/9781555817954.ch49
6. Robertson HM. Evolution of DNA Transposons in Eukaryotes. In: Craig NL, Lambowitz AM, Craigie R, Gellert M, editors. Mobile DNA II. American Society of Microbiology; 2002. pp. 1093–1110. doi:10.1128/9781555817954.ch48
7. Capy P, Gasperi G, Biémont C, Bazin C. Stress and transposable elements: co-evolution or useful parasites? Heredity. 2000;85 (Pt 2): 101–106. doi:10.1046/j.1365-2540.2000.00751.x
8. Burt A, Trivers R. Genes in Conflict: The Biology of Selfish Genetic Elements. Harvard University Press; 2009.
9. Bourgeois Y, Boissinot S. On the Population Dynamics of Junk: A Review on the Population Genomics of Transposable Elements. Genes. 2019;10. doi:10.3390/genes10060419
10. Finnegan DJ. Transposable elements. Curr Opin Genet Dev. 1992;2: 861–867.
11. Nuzhdin SV. Sure facts, speculations, and open questions about the evolution of transposable element copy number. Genetica. 1999;107: 129–137.
12. Montgomery E, Charlesworth B, Langley CH. A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. Genet Res . 1987;49: 31–41. doi:10.1017/s0016672308009634
13. Langley CH, Montgomery E, Hudson R, Kaplan N, Charlesworth B. On the role of unequal exchange in the containment of transposable element copy number. Genet Res. 1988;52: 223–235. doi:10.1017/s0016672300027695
14. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409: 860–921. doi:10.1038/35057062
15. Craig NL. A Moveable Feast: An Introduction to Mobile DNA. In: Craig NL, Chandler M, Gellert M, Lambowitz AM, Rice PA, Sandmeyer S, editors. Mobile DNA III. American Society of Microbiology; 2015. pp. 3–39. doi:10.1128/microbiolspec.MDNA3-0062-2014
16. Chalopin D, Naville M, Plard F, Galiana D, Volff J-N. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. Genome Biol Evol. 2015;7: 567–580. doi:10.1093/gbe/evv005
17. Petersen M, Armisén D, Gibbs RA, Hering L, Khila A, Mayer G, et al. Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. BMC Evol Biol. 2019;19: 11. doi:10.1186/s12862-018-1324-9
18. Szitenberg A, Cha S, Opperman CH, Bird DM, Blaxter ML, Lunt DH. Genetic drift, not life history or RNAi, determine long-term evolution of transposable elements. Genome Biol Evol. 2016;8: 2964–2978. doi:10.1093/gbe/evw208

19. Castanera R, López-Varas L, Borgognone A, LaButti K, Lapidus A, Schmutz J, et al. Transposable elements versus the fungal genome: Impact on whole-genome architecture and transcriptional profiles. PLoS Genet. 2016;12: e1006108. doi:10.1371/journal.pgen.1006108
20. Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. Nature. 1980;284: 601–603. doi:10.1038/284601a0
21. Charlesworth B, Langley CH. The population genetics of *Drosophila* transposable elements. Annu Rev Genet. 1989;23: 251–287. doi:10.1146/annurev.ge.23.120189.001343
22. Wright S, Finnegan D. Genome evolution: sex and the transposable element. Curr Biol. 2001;11: R296-9. doi:10.1016/S0960-9822(01)00168-3
23. Schaack S, Choi E, Lynch M, Pritham EJ. DNA transposons and the role of recombination in mutation accumulation in *Daphnia pulex*. Genome Biol. 2010;11: R46. doi:10.1186/gb-2010-11-4-r46
24. Petrov DA, Aminetzach YT, Davis JC, Bensasson D, Hirsh AE. Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. Mol Biol Evol. 2003;20: 880–892. doi:10.1093/molbev/msg102
25. Schwander T, Crespi BJ. Twigs on the tree of life? Neutral and selective models for integrating macroevolutionary patterns with microevolutionary processes in the analysis of asexuality. Mol Ecol. 2009;18: 28–42. doi:10.1111/j.1365-294X.2008.03992.x
26. Charlesworth B, Langley CH. The evolution of self-regulated transposition of transposable elements. Genetics. 1986;112: 359–383.
27. Dolgin ES, Charlesworth B. The fate of transposable elements in asexual populations. Genetics. 2006;174: 817–827. doi:10.1534/genetics.106.060434
28. Bast J, Jaron KS, Schuseil D, Roze D, Schwander T. Asexual reproduction reduces transposable element load in experimental yeast populations. Elife. 2019;8. doi:10.7554/eLife.48548
29. Fujita MK, Singhal S, Brunes TO, Maldonado JA. Evolutionary Dynamics and Consequences of Parthenogenesis in Vertebrates. Annu Rev Ecol Evol Syst. 2020. doi:10.1146/annurev-ecolsys-011720-114900
30. Schaack S, Gilbert C, Feschotte C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol. 2010;25: 537–546. doi:10.1016/j.tree.2010.06.001
31. Basten CJ, Moody ME. A branching-process model for the evolution of transposable elements incorporating selection. J Math Biol. 1991;29: 743–761. doi:10.1007/bf00160190
32. Edwards RJ, Brookfield JFY. Transiently beneficial insertions could maintain mobile DNA sequences in variable environments. Mol Biol Evol. 2003;20: 30–37. doi:10.1093/molbev/msg001
33. Silva JC, Loreto EL, Clark JB. Factors that affect the horizontal transfer of transposable elements. Curr Issues Mol Biol. 2004;6: 57–71.
34. Peccoud J, Loiseau V, Cordaux R, Gilbert C. Massive horizontal transfer of transposable elements in insects. Proc Natl Acad Sci U S A. 2017. doi:10.1073/pnas.1621178114
35. Boutin TS, Le Rouzic A, Capy P. How does selfing affect the dynamics of selfish transposable elements? Mob DNA. 2012;3: 5. doi:10.1186/1759-8753-3-5
36. Startek M, Le Rouzic A, Capy P, Grzebelus D, Gambin A. Genomic parasites or symbionts? Modeling the effects of environmental pressure on transposition activity in asexual populations. Theor Popul Biol. 2013;90: 145–151. doi:10.1016/j.tpb.2013.07.004
37. Neiman M, Meirmans S, Meirmans PG. What can asexual lineage age tell us about the maintenance of sex? Ann N Y Acad Sci. 2009;1168: 185–200. doi:10.1111/j.1749-6632.2009.04572.x
38. Jaron KS, Bast J, Nowell RW, Rhyker Ranallo-Benavidez T, Robinson-Rechavi M, Schwander T. Genomic features of asexual animals. bioRxiv. 2019. p. 497495. doi:10.1101/497495
39. 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. doi:10.1007/978-90-481-2770-2_13
40. 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. doi:10.1073/pnas.1012678108
41. Hudson CT, Gosse PH. The Rotifera or wheel-animalcules. Longmans, Green; 1886.
42. Donner J. Ordnung Bdelloidea. Akademie Verlag, Berlin; 1965. p. 297.
43. Mark Welch D, Meselson M. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science. 2000;288: 1211–1215. doi:10.1126/science.288.5469.1211
44. 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. doi:10.1038/nature12326
45. 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. doi:10.1073/pnas.0800972105
46. Nowell RW, Almeida P, Wilson CG, Smith TP, Fontaneto D, Crisp A, et al. Comparative genomics of bdelloid rotifers: Insights from desiccating and nondesiccating species. PLoS Biol. 2018;16: e2004830.

1145 doi:10.1371/journal.pbio.2004830
1146 47. Simion P, Narayan J, Houtain A, Derzelle A, Baudry L, Nicolas E, et al. Homologous chromosomes in
1147 asexual rotifer *Adineta vaga* suggest automixis. bioRxiv. 2020. p. 2020.06.16.155473.
1148 doi:10.1101/2020.06.16.155473
1149 48. Signorovitch A, Hur J, Gladyshev E, Meselson M. Allele sharing and evidence for sexuality in a
1150 mitochondrial clade of bdelloid rotifers. Genetics. 2015;200: 581–590. doi:10.1534/genetics.115.176719
1151 49. Vakhrusheva OA, Mnatsakanova EA, Galimov YR, Neretina TV, Gerasimov ES, Ozerova SG, et al.
1152 Recombination in a natural population of the bdelloid rotifer *Adineta vaga*. bioRxiv. 2018. p. 489393.
1153 doi:10.1101/489393
1154 50. Arkhipova I, Meselson M. Transposable elements in sexual and ancient asexual taxa. Proc Natl Acad Sci U S
1155 A. 2000;97: 14473–14477. doi:10.1073/pnas.97.26.14473
1156 51. Zhang H-H, Peccoud J, Xu M-R-X, Zhang X-G, Gilbert C. Horizontal transfer and evolution of transposable
1157 elements in vertebrates. Nat Commun. 2020;11: 1362. doi:10.1038/s41467-020-15149-4
1158 52. Arkhipova I, Meselson M. Deleterious transposable elements and the extinction of asexuals. Bioessays.
1159 2005;27: 76–85. doi:10.1002/bies.20159
1160 53. Ricci C. Anhydrobiotic capabilities of bdelloid rotifers. Hydrobiologia. 1998;387–388: 321–326.
1161 doi:10.1023/A:1017086425934
1162 54. Gladyshev E, Meselson M. Extreme resistance of bdelloid rotifers to ionizing radiation. Proc Natl Acad Sci U
1163 S A. 2008;105: 5139–5144. doi:10.1073/pnas.0800966105
1164 55. Hespeels B, Knapen M, Hanot-Mambres D, Heuskin A-C, Pineux F, Lucas S, et al. Gateway to genetic
1165 exchange? DNA double-strand breaks in the bdelloid rotifer *Adineta vaga* submitted to desiccation. J Evol
1166 Biol. 2014;27: 1334–1345. doi:10.1111/jeb.12326
1167 56. Gladyshev EA, Arkhipova IR. Genome structure of bdelloid rotifers: shaped by asexuality or desiccation? J
1168 Hered. 2010;101 Suppl 1: S85-93. doi:10.1093/jhered/esq008
1169 57. Arkhipova IR, Pyatkov KI, Meselson M, Evgen'ev MB. Retroelements containing introns in diverse
1170 invertebrate taxa. Nat Genet. 2003;33: 123–124. doi:10.1038/ng1074
1171 58. Gladyshev EA, Arkhipova IR. Telomere-associated endonuclease-deficient *Penelope*-like retroelements in
1172 diverse eukaryotes. Proc Natl Acad Sci U S A. 2007;104: 9352–9357. doi:10.1073/pnas.0702741104
1173 59. Arkhipova IR, Yushenova IA, Rodriguez F. Giant reverse transcriptase-encoding transposable elements at
1174 telomeres. Mol Biol Evol. 2017. doi:10.1093/molbev/msx159
1175 60. Arkhipova IR, Yushenova IA, Rodriguez F. Endonuclease-containing *Penelope* retrotransposons in the
1176 bdelloid rotifer *Adineta vaga* exhibit unusual structural features and play a role in expansion of host gene
1177 families. Mob DNA. 2013;4: 19. doi:10.1186/1759-8753-4-19
1178 61. Gladyshev EA, Meselson M, Arkhipova IR. A deep-branching clade of retrovirus-like retrotransposons in
1179 bdelloid rotifers. Gene. 2007;390: 136–145. doi:10.1016/j.gene.2006.09.025
1180 62. Rodriguez F, Kenefick AW, Arkhipova IR. LTR-retrotransposons from bdelloid rotifers capture additional
1181 ORFs shared between highly diverse retroelement types. Viruses. 2017;9. doi:10.3390/v9040078
1182 63. Gladyshev EA, Arkhipova IR. Rotifer rDNA-specific R9 retrotransposable elements generate an exceptionally
1183 long target site duplication upon insertion. Gene. 2009;448: 145–150. doi:10.1016/j.gene.2009.08.016
1184 64. Gladyshev EA, Arkhipova IR. A subtelomeric non-LTR retrotransposon *Hebe* in the bdelloid rotifer *Adineta*
1185 *vaga* is subject to inactivation by deletions but not 5' truncations. Mob DNA. 2010;1: 12. doi:10.1186/1759-
1186 8753-1-12
1187 65. Kim H-S, Lee B-Y, Han J, Jeong C-B, Hwang D-S, Lee M-C, et al. The genome of the freshwater
1188 monogonont rotifer *Brachionus calyciflorus*. Mol Ecol Resour. 2018;18: 646–655. doi:10.1111/1755-
1189 0998.12768
1190 66. Han J, Park JC, Choi B-S, Kim M-S, Kim H-S, Hagiwara A, et al. The genome of the marine monogonont
1191 rotifer *Brachionus plicatilis*: Genome-wide expression profiles of 28 cytochrome P450 genes in response to
1192 chlorpyrifos and 2-ethyl-phenanthrene. Aquat Toxicol. 2019;214: 105230. doi:10.1016/j.aquatox.2019.105230
1193 67. Blommaert J, Riss S, Hecox-Lea B, Mark Welch DB, Stelzer CP. Small, but surprisingly repetitive genomes:
1194 transposon expansion and not polyploidy has driven a doubling in genome size in a metazoan species
1195 complex. BMC Genomics. 2019;20: 466. doi:10.1186/s12864-019-5859-y
1196 68. Mauer K, Hellmann SL, Groth M, Fröbius AC, Zischler H, Hankeln T, et al. The genome, transcriptome, and
1197 proteome of the fish parasite *Pomphorhynchus laevis* (Acanthocephala). PLoS One. 2020;15: e0232973.
1198 doi:10.1371/journal.pone.0232973
1199 69. Wey-Fabrizius AR, Herlyn H, Rieger B, Rosenkranz D, Witek A, Welch DBM, et al. Transcriptome data
1200 reveal Syndermatan relationships and suggest the evolution of endoparasitism in Acanthocephala via an
1201 epizoic stage. PLoS One. 2014;9: e88618. doi:10.1371/journal.pone.0088618
1202 70. Sielaff M, Schmidt H, Struck TH, Rosenkranz D, Mark Welch DB, Hankeln T, et al. Phylogeny of
1203 Syndermata (syn. Rotifera): Mitochondrial gene order verifies epizoic Seisonidea as sister to endoparasitic
1204 Acanthocephala within monophyletic Hemirotifera. Mol Phylogenet Evol. 2016;96: 79–92.

1205 doi:10.1016/j.ympev.2015.11.017
1206 71. Laumer CE, Fernández R, Lemer S, Combosch D, Kocot KM, Riesgo A, et al. Revisiting metazoan phylogeny
1207 with genomic sampling of all phyla. Proc Biol Sci. 2019;286: 20190831. doi:10.1098/rspb.2019.0831
1208 72. C. elegans Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for
1209 investigating biology. Science. 1998;282: 2012–2018. doi:10.1126/science.282.5396.2012
1210 73. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, et al. The genome sequence of
1211 *Drosophila melanogaster*. Science. 2000;287: 2185–2195. doi:10.1126/science.287.5461.2185
1212 74. Young ND, Jex AR, Li B, Liu S, Yang L, Xiong Z, et al. Whole-genome sequence of *Schistosoma*
1213 *haematobium*. Nat Genet. 2012;44: 221–225. doi:10.1038/ng.1065
1214 75. Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, et al. The oyster genome reveals stress adaptation and
1215 complexity of shell formation. Nature. 2012;490: 49–54. doi:10.1038/nature11413
1216 76. Simakov O, Marletaz F, Cho S-J, Edsinger-Gonzales E, Havlak P, Hellsten U, et al. Insights into bilaterian
1217 evolution from three spiralian genomes. Nature. 2013;493: 526–531. doi:10.1038/nature11696
1218 77. Gusev O, Suetsugu Y, Cornette R, Kawashima T, Logacheva MD, Kondrashov AS, et al. Comparative
1219 genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge.
1220 Nat Commun. 2014;5: 4784. doi:10.1038/ncomms5784
1221 78. Albertin CB, Simakov O, Mitros T, Wang ZY, Pungor JR, Edsinger-Gonzales E, et al. The octopus genome
1222 and the evolution of cephalopod neural and morphological novelties. Nature. 2015;524: 220–224.
1223 doi:10.1038/nature14668
1224 79. Luo Y-J, Takeuchi T, Koyanagi R, Yamada L, Kanda M, Khalturina M, et al. The *Lingula* genome provides
1225 insights into brachiopod evolution and the origin of phosphate biomineralization. Nat Commun. 2015;6: 8301.
1226 doi:10.1038/ncomms9301
1227 80. Mikhailov KV, Slyusarev GS, Nikitin MA, Logacheva MD, Penin AA, Aleoshin VV, et al. The genome of
1228 *Intoshia linei* affirms orthonectids as highly simplified spiralians. Curr Biol. 2016;26: 1768–1774.
1229 doi:10.1016/j.cub.2016.05.007
1230 81. Hashimoto T, Horikawa DD, Saito Y, Kuwahara H, Kozuka-Hata H, Shin-I T, et al. Extremotolerant
1231 tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. Nat
1232 Commun. 2016;7: 12808. doi:10.1038/ncomms12808
1233 82. Yoshida Y, Koutsovoulos G, Laetsch DR, Stevens L, Kumar S, Horikawa DD, et al. Comparative genomics of
1234 the tardigrades *Hypsibius dujardini* and *Ramazzottius varieornatus*. PLoS Biol. 2017;15: e2002266.
1235 doi:10.1371/journal.pbio.2002266
1236 83. Adema CM, Hillier LW, Jones CS, Loker ES, Knight M, Minx P, et al. Whole genome analysis of a
1237 schistosomiasis-transmitting freshwater snail. Nat Commun. 2017;8: 15451. doi:10.1038/ncomms15451
1238 84. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013-2015. Available: http://www.repeatmasker.org
1239 85. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative
1240 studies of nucleotide sequences. J Mol Evol. 1980;16: 111–120. doi:10.1007/BF01731581
1241 86. Kapusta A, Suh A. Evolution of bird genomes-a transposon's-eye view. Ann N Y Acad Sci. 2017;1389: 164–
1242 185. doi:10.1111/nyas.13295
1243 87. Shao F, Han M, Peng Z. Evolution and diversity of transposable elements in fish genomes. Sci Rep. 2019;9:
1244 15399. doi:10.1038/s41598-019-51888-1
1245 88. Belshaw R, Pereira V, Katzourakis A, Talbot G, Paces J, Burt A, et al. Long-term reinfection of the human
1246 genome by endogenous retroviruses. Proc Natl Acad Sci U S A. 2004;101: 4894–4899.
1247 doi:10.1073/pnas.0307800101
1248 89. Garbarino JE, Gibbons IR. Expression and genomic analysis of midasin, a novel and highly conserved AAA
1249 protein distantly related to dynein. BMC Genomics. 2002;3: 18. doi:10.1186/1471-2164-3-18
1250 90. Li P-C, Li K, Wang J, Zhao C-Z, Zhao S-Z, Hou L, et al. The AAA-ATPase MIDASIN 1 functions in
1251 ribosome biogenesis and is essential for embryo and root development. Plant Physiol. 2019;180: 289–304.
1252 doi:10.1104/pp.18.01225
1253 91. Hur JH, Van Doninck K, Mandigo ML, Meselson M. Degenerate tetraploidy was established before bdelloid
1254 rotifer families diverged. Mol Biol Evol. 2009;26: 375–383. doi:10.1093/molbev/msn260
1255 92. Eyres I, Frangedakis E, Fontaneto D, Herniou EA, Boschetti C, Carr A, et al. Multiple functionally divergent
1256 and conserved copies of alpha tubulin in bdelloid rotifers. BMC Evol Biol. 2012;12: 148. doi:10.1186/1471-
1257 2148-12-148
1258 93. Song M, Boissinot S. Selection against LINE-1 retrotransposons results principally from their ability to
1259 mediate ectopic recombination. Gene. 2007;390: 206–213. doi:10.1016/j.gene.2006.09.033
1260 94. Arkhipova IR, Yushenova IA. Giant transposons in eukaryotes: Is bigger better? Genome Biol Evol. 2019;11:
1261 906–918. doi:10.1093/gbe/evz041
1262 95. Cooper DM, Schimenti KJ, Schimenti JC. Factors affecting ectopic gene conversion in mice. Mamm Genome.
1263 1998;9: 355–360. doi:10.1007/s003359900769
1264 96. Höck J, Meister G. The Argonaute protein family. Genome Biol. 2008;9: 210. doi:10.1186/gb-2008-9-2-210

1265 97. Juliano C, Wang J, Lin H. Uniting germline and stem cells: the function of Piwi proteins and the piRNA
1266 pathway in diverse organisms. Annu Rev Genet. 2011;45: 447–469. doi:10.1146/annurev-genet-110410-
1267 132541
1268 98. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev Genet. 2009;10: 94–108.
1269 doi:10.1038/nrg2504
1270 99. de Jong D, Eitel M, Jakob W, Osigus H-J, Hadrys H, Desalle R, et al. Multiple dicer genes in the early-
1271 diverging metazoa. Mol Biol Evol. 2009;26: 1333–1340. doi:10.1093/molbev/msp042
1272 100. Zong J, Yao X, Yin J, Zhang D, Ma H. Evolution of the RNA-dependent RNA polymerase (RdRP) genes:
1273 duplications and possible losses before and after the divergence of major eukaryotic groups. Gene. 2009;447:
1274 29–39. doi:10.1016/j.gene.2009.07.004
1275 101. Yigit E, Batista PJ, Bei Y, Pang KM, Chen C-CG, Tolia NH, et al. Analysis of the *C. elegans* Argonaute
1276 family reveals that distinct Argonautes act sequentially during RNAi. Cell. 2006;127: 747–757.
1277 doi:10.1016/j.cell.2006.09.033
1278 102. Shi Z, Montgomery TA, Qi Y, Ruvkun G. High-throughput sequencing reveals extraordinary fluidity of
1279 miRNA, piRNA, and siRNA pathways in nematodes. Genome Res. 2013;23: 497–508.
1280 doi:10.1101/gr.149112.112
1281 103. Buck AH, Blaxter M. Functional diversification of Argonautes in nematodes: an expanding universe. Biochem
1282 Soc Trans. 2013;41: 881–886. doi:10.1042/BST20130086
1283 104. Aravin AA, Hannon GJ, Brennecke J. The Piwi-piRNA pathway provides an adaptive defense in the
1284 transposon arms race. Science. 2007;318: 761–764. doi:10.1126/science.1146484
1285 105. Sarkies P, Selkirk ME, Jones JT, Blok V, Boothby T, Goldstein B, et al. Ancient and novel small RNA
1286 pathways compensate for the loss of piRNAs in multiple independent nematode lineages. PLoS Biol. 2015;13:
1287 e1002061. doi:10.1371/journal.pbio.1002061
1288 106. Shirayama M, Seth M, Lee H-C, Gu W, Ishidate T, Conte D Jr, et al. piRNAs initiate an epigenetic memory of
1289 nonself RNA in the C. elegans germline. Cell. 2012;150: 65–77. doi:10.1016/j.cell.2012.06.015
1290 107. Rodriguez F, Arkhipova IR. Multitasking of the piRNA silencing machinery: targeting transposable elements
1291 and foreign genes in the bdelloid rotifer *Adineta vaga*. Genetics. 2016;203: 255–268.
1292 doi:10.1534/genetics.116.186734
1293 108. Gladyshev EA, Meselson M, Arkhipova IR. Massive horizontal gene transfer in bdelloid rotifers. Science.
1294 2008;320: 1210–1213. doi:10.1126/science.1156407
1295 109. Boschetti C, Carr A, Crisp A, Eyres I, Wang-Koh Y, Lubzens E, et al. Biochemical diversification through
1296 foreign gene expression in bdelloid rotifers. PLoS Genet. 2012;8: e1003035.
1297 doi:10.1371/journal.pgen.1003035
1298 110. Eyres I, Boschetti C, Crisp A, Smith TP, Fontaneto D, Tunnacliffe A, et al. Horizontal gene transfer in
1299 bdelloid rotifers is ancient, ongoing and more frequent in species from desiccating habitats. BMC Biol.
1300 2015;13: 90. doi:10.1186/s12915-015-0202-9
1301 111. Becks L, Agrawal AF. The effect of sex on the mean and variance of fitness in facultatively sexual rotifers. J
1302 Evol Biol. 2011;24: 656–664. doi:10.1111/j.1420-9101.2010.02199.x
1303 112. Andrews S. FastQC: A quality-control tool for high-throughput sequence data. 2015. Available:
1304 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
1305 113. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and
1306 samples in a single report. Bioinformatics. 2016;32: 3047–3048. doi:10.1093/bioinformatics/btw354
1307 114. Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. Direct determination of diploid genome sequences.
1308 Genome Res. 2017;27: 757–767. doi:10.1101/gr.214874.116
1309 115. Coombe L, Zhang J, Vandervalk BP, Chu J, Jackman SD, Birol I, et al. ARKS: chromosome-scale scaffolding
1310 of human genome drafts with linked read kmers. BMC Bioinformatics. 2018;19: 234. doi:10.1186/s12859-
1311 018-2243-x
1312 116. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome
1313 assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19: 455–477.
1314 doi:10.1089/cmb.2012.0021
1315 117. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods.
1316 2015;12: 59–60. doi:10.1038/nmeth.3176
1317 118. Laetsch DR, Blaxter ML. BlobTools: Interrogation of genome assemblies. F1000Res. 2017;6.
1318 doi:10.12688/f1000research.12232.1
1319 119. Pryszcz LP, Gabaldón T. Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids
1320 Res. 2016;44: e113. doi:10.1093/nar/gkw294
1321 120. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE.
1322 Bioinformatics. 2011;27: 578–579. doi:10.1093/bioinformatics/btq683
1323 121. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-
1324 efficient short-read de novo assembler. Gigascience. 2012;1: 18. doi:10.1186/2047-217X-1-18

122. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34: 3094–3100. doi:10.1093/bioinformatics/bty191
123. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, et al. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 2019;20: 224. doi:10.1186/s13059-019-1829-6
124. Salzberg SL, Phillippy AM, Zimin A, Puiu D, Magoc T, Koren S, et al. GAGE: A critical evaluation of genome assemblies and assembly algorithms. Genome Res. 2012;22: 557–567. doi:10.1101/gr.131383.111
125. 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. doi:10.1093/bioinformatics/btv351
126. Hanson SJ, Stelzer C-P, Welch DBM, Logsdon JM Jr. Comparative transcriptome analysis of obligately asexual and cyclically sexual rotifers reveals genes with putative functions in sexual reproduction, dormancy, and asexual egg production. BMC Genomics. 2013;14: 412. doi:10.1186/1471-2164-14-412
127. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29: 15–21. doi:10.1093/bioinformatics/bts635
128. Stanke M, Schöffmann O, Morgenstern B, Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 2006;7: 62. doi:10.1186/1471-2105-7-62
129. 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. doi:10.1093/bioinformatics/btn013
130. Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics. 2011;27: 1691–1692. doi:10.1093/bioinformatics/btr174
131. 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. doi:10.1093/bioinformatics/btv661
132. Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. Whole-Genome Annotation with BRAKER. Methods Mol Biol. 2019;1962: 65–95. doi:10.1007/978-1-4939-9173-0_5
133. Norling M, Jareborg N, Dainat J. EMBLmyGFF3: a converter facilitating genome annotation submission to European Nucleotide Archive. BMC Res Notes. 2018;11: 584. doi:10.1186/s13104-018-3686-x
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. doi:10.1038/msb.2011.75
135. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28: 1647–1649. doi:10.1093/bioinformatics/bts199
136. 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. doi:10.1093/molbev/msu300
137. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14: 587–589. doi:10.1038/nmeth.4285
138. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59: 307–321. doi:10.1093/sysbio/syq010
139. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol. 2018;35: 518–522. doi:10.1093/molbev/msx281
140. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117: 9451–9457. doi:10.1073/pnas.1921046117
141. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6: 11. doi:10.1186/s13100-015-0041-9
142. Bailly-Bechet M, Haudry A, Lerat E. “One code to find them all”: a perl tool to conveniently parse RepeatMasker output files. Mob DNA. 2014;5: 13. doi:10.1186/1759-8753-5-13
143. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol. 2012;3: 217–223. doi:10.1111/j.2041-210X.2011.00169.x
144. R Core Team. R: A language and environment for statistical computing. 2016. Available: https://www.R-project.org/
145. Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, et al. Combined evidence annotation of transposable elements in genome sequences. PLoS Comput Biol. 2005;1: 166–175. doi:10.1371/journal.pcbi.0010022
146. Flutre T, Duprat E, Feuillet C, Quesneville H. Considering transposable element diversification in de novo annotation approaches. PLoS One. 2011;6: e16526. doi:10.1371/journal.pone.0016526
147. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for

eukaryotic transposable elements. Nat Rev Genet. 2007;8: 973–982. doi:10.1038/nrg2165
148. Ou S, Jiang N. LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. Plant Physiol. 2018;176: 1410–1422. doi:10.1104/pp.17.01310
149. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 2007;35: D61-5. doi:10.1093/nar/gkl842
150. Agapow P-M, Burt A. Indices of multilocus linkage disequilibrium. Mol Ecol Notes. 2001;1: 101–102. doi:10.1046/j.1471-8278.2000.00014.x
151. Kamvar ZN, Tabima JF, Grünwald NJ. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. 2014;2: e281. doi:10.7717/peerj.281
152. Hartfield M, Wright SI, Agrawal AF. Coalescence and Linkage Disequilibrium in Facultatively Sexual Diploids. Genetics. 2018;210: 683–701. doi:10.1534/genetics.118.301244
153. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26: 841–842. doi:10.1093/bioinformatics/btq033
154. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000;16: 276–277. doi:10.1016/s0168-9525(00)02024-2
155. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. Nucleic Acids Res. 2019;47: D427–D432. doi:10.1093/nar/gky995
156. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30: 1236–1240. doi:10.1093/bioinformatics/btu031
157. Garamszegi LZ, editor. Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice. Springer, Berlin, Heidelberg; 2014. doi:10.1007/978-3-662-43550-2
158. Hadfield JD. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. J Stat Softw. 2010;33. doi:10.18637/jss.v033.i02
159. de Villemereuil P, Nakagawa S. General Quantitative Genetic Methods for Comparative Biology. In: Garamszegi LZ, editor. Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. pp. 287–303. doi:10.1007/978-3-662-43550-2_11
160. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, Articles. 2015;67: 1–48. doi:10.18637/jss.v067.i01