Current Biology

Cross-Contamination Explains "Inter and Intraspecific Horizontal Genetic Transfers" between Asexual Bdelloid Rotifers

Highlights

- Clear evidence that findings in a 2016 article are artifacts of experimental error

- A new method to analyze Sanger chromatograms reveals contamination between samples

- No credible evidence for horizontal DNA exchange within or between bdelloid species

- The evolutionary scandal of the asexual rotifers continues to defy resolution

Authors

Christopher G. Wilson,
Reuben W. Nowell,
Timothy G. Barraclough

Correspondence

chris.wilson@imperial.ac.uk

In Brief

Wilson et al. re-examine raw data from a 2016 article that had reported lateral transfer of DNA between bdelloid rotifers in lieu of sex. Samples that were supposed to contain DNA from a single animal are found to have included DNA from two or three different rotifers, revealing that data and conclusions from the earlier publication are unreliable.

CellPress

Cross-Contamination Explains "Inter and Intraspecific Horizontal Genetic Transfers" between Asexual Bdelloid Rotifers

Christopher G. Wilson,[1,2,*] Reuben W. Nowell,[1] and Timothy G. Barraclough[1]
[1]Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot, Berkshire SL5 7PY, UK
[2]Lead Contact
*Correspondence: chris.wilson@imperial.ac.uk
https://doi.org/10.1016/j.cub.2018.05.070

SUMMARY

A few metazoan lineages are thought to have persisted for millions of years without sexual reproduction. If so, they would offer important clues to the evolutionary paradox of sex itself [1, 2]. Most "ancient asexuals" are subject to ongoing doubt because extant populations continue to invest in males [3–9]. However, males are famously unknown in bdelloid rotifers, a class of microscopic invertebrates comprising hundreds of species [10–12]. Bdelloid genomes have acquired an unusually high proportion of genes from non-metazoans via horizontal transfer [13–17]. This well-substantiated finding has invited speculation [13] that homologous horizontal transfer between bdelloid individuals also may occur, perhaps even "replacing" sex [14]. In 2016, *Current Biology* published an article claiming to supply evidence for this idea. Debortoli et al. [18] sampled rotifers from natural populations and sequenced one mitochondrial and four nuclear loci. Species assignments were incongruent among loci for several samples, which was interpreted as evidence of "interspecific horizontal genetic transfers." Here, we use sequencing chromatograms supplied by the authors to demonstrate that samples treated as individuals actually contained two or more highly divergent mitochondrial and ribosomal sequences, revealing cross-contamination with DNA from multiple animals of different species. Other chromatograms indicate contamination with DNA from conspecific animals, explaining genetic and genomic evidence for "intraspecific horizontal exchanges" reported in the same study. Given the clear evidence of contamination, the data and findings of Debortoli et al. [18] provide no reliable support for their conclusions that DNA is transferred horizontally between or within bdelloid species.

RESULTS AND DISCUSSION

Several Patterns in the Data of Debortoli et al. Are Unexpected Given Their Hypothesis

Debortoli et al. [18] collected lichen from five trees in a small area and moved bdelloid rotifers of the cosmopolitan genus *Adineta* [10] to 576 Eppendorf tubes for DNA extraction. They amplified a 0.6-kb region of mitochondrial cytochrome oxidase I (mtCO1) by PCR and used this common molecular barcode [19] to delineate six molecular taxa [20], which they call "*Adineta vaga* species A–F." They applied whole-genome amplification (WGA) to a subset of 82 samples representing a range of mtCO1 haplotypes from these "cryptic species." Four nuclear marker loci were then amplified again by PCR and directly Sanger sequenced. Most samples yielded sequences characteristic of a single species at all five loci. However, for six samples (7.3%), sequences at different loci matched two or even three different species. The authors interpret this incongruence as "strong evidence" of "interspecific horizontal genetic transfers" from "donor species" to "recipient individuals" that "may be mediated by DNA repair through homologous recombination." They claim to have discovered "an unexpected (and possibly unique)… ameiotic strategy of genetic exchange and recombination among asexual, morphologically female organisms," which they "propose here to call 'sapphomixis' (from the name of the Greek lesbian poetess Sappho and *mixis* 'mingling')."

Given this interpretation, several features of the results are puzzling. In every case of interspecific incongruence, the inferred donor species was found on the same maple or plane tree as the putative recipient at the time of sampling (Table S3 in [18]). Rotifers of the genus *Adineta* have high dispersal potential [21, 22] and vast cryptic diversity ("*A. vaga*" alone comprises at least 36 independently evolving entities [23, 24]). If horizontal genetic transfer (HGT) occurs so extensively among such diverse and mobile animals, it is remarkable that every donor species happened to be sampled in the same small area (300 m^2) as its respective recipient, at the same time. Furthermore, every case of incongruence involved a "transferred" DNA sequence whose exact counterpart was recovered natively in one of the other 81 rotifers sampled, enabling the authors to construct a perfectly self-contained circular representation of the "interspecific transfers" (Figure 4 in [18]). This means the transfers must all be so recent that, at the time of the study, every recipient was still living alongside a donor species bearing an identical sequence, which seems to require extremely high frequencies of homologous recombination (HR) between molecules from different species.

The divergence between sequences involved in the proposed transfer events is also noteworthy. As Debortoli et al. [18] recognize, it is well established that the frequency of HR "is strongly correlated with the degree of identity between the recombining DNA fragments and dramatically declines as the sequences diverge." This means that species pairs involved in interspecific

Table 1. Sequence Divergence between DNA Molecules for "Interspecific Horizontal Genetic Transfers" Reported by Debortoli et al.

Sample	Interpreted As	Marker	Inferred Replacement via HR (GenBank IDs)	Pairwise Identity (%)	GC Content Difference (%)
B11	Ind 21 [A]	Nu1054	Hap22 [E] (KU861059) rpl. Hap2 [A] (KU861060)	71.0	−21.2
B11	Ind 21 [A]	EPIC25	Hap35 [E] (KU860812) rpl. Hap1 [A] (KU860872)	68.6	−17.7
B11	Ind 21 [A]	EPIC63	Hap16 [E] (KU860935) rpl. Hap1 [A] (KU860973)	68.9	−13.9
B14	Ind 1 [A]	EPIC25	Hap9&10 [C] (KU860813-14) rpl. Hap1 [A] (KU860815)	76.1	+4.8
B14	Ind 1 [A]	EPIC63	Hap16&20 [E] (KU860936-37) rpl. Hap1 [A] (KU860938)	68.2	−13.7
B22	Ind 58 [C]	EPIC25	Hap37 [E] (KU860816) rpl. Hap10 [C] (KU860907)	68.4	−22.3
B39	Ind 66 [E]	mtCO1	Hap10 [C] (KU860601) rpl. Hap31 [E] (KU860644)	88.3	0.0
B3B1	Ind 23 [A]	EPIC25	Hap35 [E] (KU860823) rpl. Hap1 [A] (KU860872)	68.6	−17.7
D14	Ind 5 [A]	EPIC25	Hap10 [C] (KU860854) rpl. Hap4 [A] (KU860878)	76.0	+4.7
Mean (median)	–	–	–	72.7 (68.9)	−10.8 (−13.9)

The divergence values shown here appear incompatible with current understanding of homologous recombination (HR). Frequent horizontal exchange between such divergent molecules also appears inconsistent with the tetraploid structure of the *A. vaga* genome, where separate gene copies evolve independently despite sharing a median pairwise homology of 75% [14]. High divergence values are linked to substantial guanine-cytosine (GC) content differences among species A, C, and E. Ind, individual; Hap, haplotype; rpl., replaces. See also Figure S1.

recombination ought to have very high sequence identities at the "exchanged" loci. We checked this by aligning each putatively transferred marker against the native sequence it was supposed to have replaced (Table 1). The mean pairwise identity was just 72.7% (median 68.9%). Similarly low values were obtained when we examined wider genomic regions surrounding the markers, at multiple scales (Figure S1). According to established understanding from bacteria [25–29], protists [30], fungi [31], plants [32, 33], and animals [34, 35], these molecules are too divergent to undergo HR at measurable rates, let alone so frequently and across horizontal barriers. The observed patterns therefore seem incompatible with the mechanism of HR proposed by the authors and with genomic features of *Adineta* (Figure S1). Although we cannot rule out some unknown alternative mechanism, these unexpected features seem to suggest a simpler hypothesis.

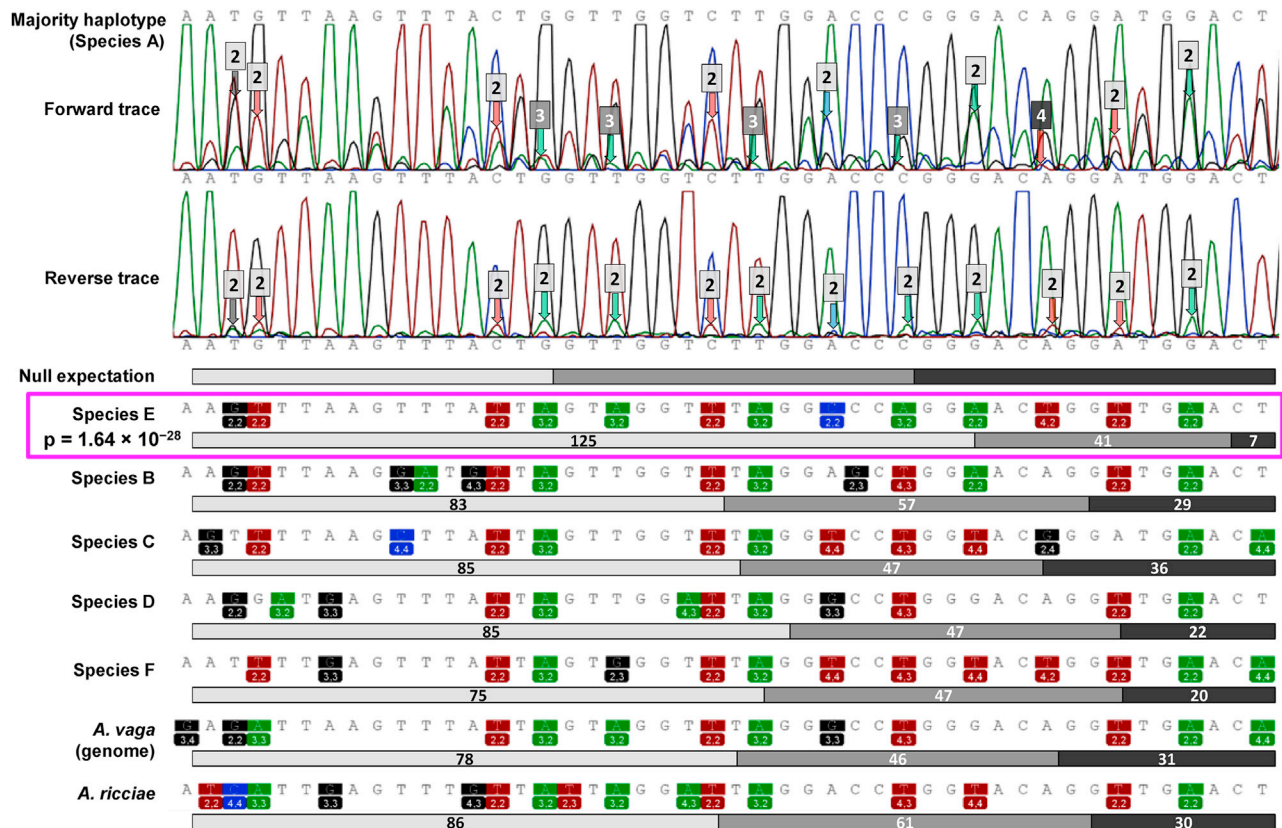## Cross-Contamination with Heterospecific DNA Explains "Interspecific Genetic Exchanges"

We investigated whether evidence interpreted as HGT might instead have resulted from accidental cross-contamination between tubes during preparation of samples from natural material. It is technically challenging to isolate *Adineta* individuals. The animals are small even when extended; when disturbed, they contract rapidly into tiny, motionless, transparent spheroids that stick tenaciously to plasticware and refract light similarly. If, on occasion, more than one animal (or loose DNA from extra animals) were added to a single tube, it would give the misleading appearance of inter- and intraspecific genetic exchanges. This would explain the puzzling features discussed above, without reference to extraordinary new genetic mechanisms.

If a tube contains DNA from two or more animals of different species, two substantially divergent sequences ought to be evident as extra peaks in chromatograms from direct Sanger sequencing of PCR amplicons. In particular, prior sequencing of thousands of individual rotifers indicates that each animal has only one mtCO1 haplotype (e.g., [36–38]); therefore, a pattern of double peaks in mtCO1 chromatograms would indicate contamination. We checked this prediction by deliberately

adding two animals of different species to tubes and generating chromatograms for mtCO1 using the methods of Debortoli et al. [18]. One sequence tended to become overrepresented during PCR, but the second sequence was visible as small secondary peaks at sites where the two species differ. We developed and validated a simple quantitative method, which we call ConTAMPR (contingency table analysis of minority peak ranks), to detect the second sequence in our chromatograms and identify it from a panel of candidates, even against considerable background noise (Figures S2 and S3; Table S1). Contamination is revealed by a statistically significant excess of minority fluorescence peaks that match the proposed contaminant sequence and rank second or third in height, relative to either the null expectation (equal numbers of second-, third-, and fourth-ranked peaks) or to control sequences (see STAR Methods).

We requested original raw chromatogram files generated by Debortoli et al. [18] and applied ConTAMPR to investigate the six samples where interspecific horizontal genetic transfers were diagnosed based on incongruent species assignments. We aligned the mtCO1 chromatograms bidirectionally against the sequences reported by the authors and analyzed the minority peaks with reference to other mtCO1 sequences in their study, including those of the supposed HGT donor species (Figures 1 and S2; see STAR Methods).

This analysis revealed highly significant evidence of additional sequences (Table 2; Figures S2 and S3). Three samples (B11, B22, and B39) each contained two clearly identifiable mtCO1 haplotypes, so highly divergent that they are only consistent with DNA from two animals [39]. In each case, the second animal belonged to the supposed donor species (e.g., Figure 1), thus resolving the apparent incongruence. The probabilities of observing such marked patterns of minority peaks given the null expectation are vanishingly small (e.g., $p = 1.64 \times 10^{-28}$ for Figure 1). Two further samples (B14 and B3B1) showed evidence of nuclear DNA originating in three different rotifers (from species A, C, and E). The mtCO1 chromatograms for these samples were extreme outliers in Phred quality scores (Figures S3J and S3K), with too much noise to narrow down just one candidate contaminant (Table 2; Figures S3F and S3G).

Figure 1. An Example of the ConTAMPR Method: Minority Peaks in Sanger Sequencing Chromatograms for mtCO1 Reveal Heterospecific Cross-Contamination in Sample B11 of Debortoli et al.

Sample B11 was interpreted as an individual rotifer belonging to "*Adineta vaga* species A," whose genome contained nuclear DNA acquired horizontally from species E. A single mtCO1 haplotype was reported for species A (Hap6). Here, a short section of this majority haplotype is aligned with the original chromatograms and with other mtCO1 haplotypes reported in the study. Where a base differs from the majority signal, two small numbers in a colored box indicate the height ranks of the corresponding minority peaks in the forward and reverse chromatograms. Minority peaks matching species E (Hap31) are indicated with arrows and height ranks in gray boxes on the traces. Horizontal gray-shaded bars indicate the proportion of second-, third-, and fourth-ranked peak heights for the null expectation (top) and for all non-identical sites across each aligned haplotype. If these peaks resulted from random noise, the rank distribution for each aligned haplotype would follow the null expectation (equal proportions of ranks). For species E, however, there is a highly significant excess of second-ranked peaks relative to the null expectation ($\chi^2 = 127.95$; d.f. = 2; p = 1.64×10^{-28}). Species E also is a significantly better fit ($\chi^2 = 39.2$; d.f. = 12; p = 9.73×10^{-5}) than the other four species reported by Debortoli et al. [18] and two *Adineta* reference clones (*A. vaga* and *A. ricciae*). This indicates contamination with DNA from an animal belonging to species E. See also Figures S2 and S3 and Data S1.

More contaminating sequences were detected when we applied the same technique to examine chromatograms for nuclear markers (see STAR Methods). Samples B11, B14, and B22 contained extra 28S ribosomal DNA sequences, which again matched the supposed donor species (Table 2; Figures S3A, S3C, and S3H). Ribosomal DNA undergoes concerted evolution [40] that ought to preclude the maintenance of substantially divergent copies in a single animal. Ribosomal DNA has been amplified from hundreds of rotifers without finding intragenomic copies that differ by more than a handful of bases (e.g., [41]). Debortoli et al. [18] report no more than three base differences between 28S copies within any of 82 animals (Figure 3A in [18]). In contrast, the two sequences discovered using ConTAMPR in B11, B14, and B22 differed at over 40 positions in the first 700 bp alone, indicating multiple animals from different species. Owing to minority peaks from additional sequences, quality scores for 28S chromatograms were significantly lower in the subset of samples where HGT was claimed

(Figure S3K; Mann-Whitney test: N = 122; W = 373.5; p = 0.015). This analysis revealed a seventh contaminated sample (A3B1), which contained mtDNA from species C and also from a reference strain of *A. vaga* that is maintained in long-term culture for genetic work [14] in the laboratory used by Debortoli et al. [18] but was not found at any of their natural sampling sites (Figures S3I and S3K). For samples B14 and B3B1, ConTAMPR also revealed extra copies of the EPIC25 (exon-primed intron-crossing) nuclear marker, matching different species (Table 2).

For samples B39 and D14, ConTAMPR revealed the "native" sequences that were supposed to have been replaced by horizontal acquisitions at the mtCO1 and EPIC25 loci, respectively (Figure S3E; Data S1). All loci in these samples are now brought into congruence: each sample represents an animal with a full complement of markers typical of its species, whereas the incongruent sequences correspond to DNA fragments originating from different animals (see STAR Methods).

Table 2. Summary of ConTAMPR Results for Six Samples with Incongruent Species Assignments among Marker Loci

Sample	Interpreted As	mtCO1	28S	EPIC25	EPIC63	Nu1054	Additional Sequences Indicated by ConTAMPR	As Seen Also In	p (Fit) Given Null Distribution	p (Fit) versus {Control Sequences}
B11	Ind 21 [A]	Hap6 [A]	Hap1 [A]	Hap35 [E]	Hap16 [E]	Hap22 [E]	mtCO1 cf. Hap31 [E]	Ind 81 [E]	1.64×10^{-28}	9.73×10^{-5} {B;C;D;F;Ar;Av}
							28S cf. Hap16 [E]	Ind 81 [E]	2.30×10^{-19}	0.0085 {B&C;F}
B14[a]	Ind 1 [A]	Hap6 [A]	Hap1 [A]	Hap9–Hap10 [C]	Hap16–Hap20 [E]	missing[b]	mtCO1 cf. Hap10 [C]	Ind 50 [C]	1.94×10^{-5}	0.187 {B;D;E;F;Ar;Av} n.s.
							28S cf. Hap16 [E]	Ind 81 [E]	2.41×10^{-20}	3.0×10^{-4} {B;C;F;Ar}
							EPIC25 cf. Hap30 [E]	Ind 81 [E]	0.00374	NA (only one direction available)
B22	Ind 58 [C]	Hap10 [C]	Hap5–Hap6 [C]	Hap37 [E]	missing[b]	Hap16 [C]	mtCO1 cf. Hap36 [E]	Ind 80 [E]	2.63×10^{-38}	3.99×10^{-10} {A;B;D;F;Ar;Av}
							28S cf. Hap16 [E]	Ind 72 [E]	3.69×10^{-10}	0.029 {A&B&F}
							EPIC25 ≠ Hap37 [E]	Ind 72 [E]	NA	NA (>7 clear polymorphisms)
B3B1[a]	Ind 23 [A]	Hap1 [A]	Hap1–Hap2 [A]	Hap35 [E]	Hap1 [A]	missing[b]	mtCO1 cf. Hap11 [C]	Ind 33 [C]	0.00158	0.99 {B;C;D;E;F;Ar;Av} n.s.
							EPIC25 cf. Hap10 [C] (and cf. Hap6 [C]?)	Ind 47 [C]	2.84×10^{-24}	NA (no other species would align)
B39	Ind 66 [E]	Hap10 [C]	Hap11–Hap16 [E]	Hap30–Hap36 [E]	Hap16 [E]	Hap19 [E]	mtCO1 cf. Hap31 [E]	Ind 81 [E]	2.82×10^{-18}	6.55×10^{-12} {A;B;D;F;Av;Ar}; 4.27×10^{-8} {Hap35,36 [E]}; 0.045 {Hap29,34 [E]}
D14	Ind 5 [A]	Hap3 [A]	Hap1 [A]	Hap10 [C]	Hap1 [A]	Hap1 [A]	EPIC25 Hap4 [A]	Ind 6 [A]	3.19×10^{-17}	NA (no other sequence would align)

Debortoli et al. [18] interpreted these samples as *Adineta* individuals whose genomes had acquired DNA horizontally from a supposed donor species at one or more of five sequenced loci. However, raw chromatograms revealed minority peaks that closely matched sequences seen in other animals in the dataset. ConTAMPR was used to calculate the probability of obtaining a fit at least this good given random sequencing noise. Where possible, the equivalent fit was calculated for control species or sequences, along with the probability that the fit of the focal sequence shared the same distribution. The additional sequences matched the putative donor species, or the "native" sequence that was supposed to have been replaced, indicating that incongruence was caused by cross-contamination. [A]–[F], *Adineta* species A–F; Ar, *A. ricciae*; Av, *A. vaga* (reference genome); n.s., not significant; NA, not applicable. See also Figure S3, Data S1, and STAR Methods.
[a]Samples where sequences from three different species were detected.
[b]Cases where sequences were missing from the work of Debortoli et al. [18], owing to amplification or sequencing failure.
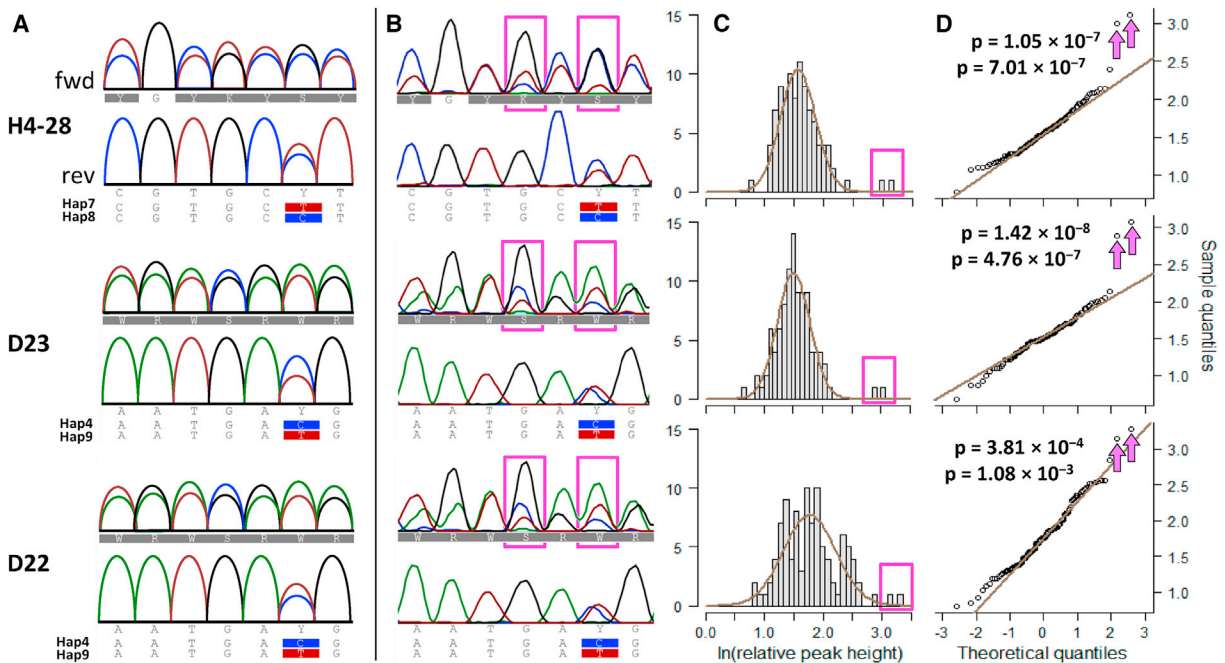
Figure 2. Conspecific Contamination Is Indicated by Triple Peaks in EPIC63 Chromatograms for Samples H4-28, D23, and D22, Revealing Four Alleles where Only Two Are Expected for a Single Animal

(A) Predicted appearance of overlapping chromatogram peaks for the two alleles reported by Debortoli et al. [18], which differ by a C/T SNP and a 2 bp indel earlier in the sequence that has shifted their forward reads 2 bp out of phase.

(B) In each case, the data show two unexpected triple peaks (pink boxes), indicating that C and T variants each occur as both long and short alleles. Four different copies therefore are present (long/C, long/T, short/C, and short/T), whereas a single animal would only have two. This is hidden in the reverse reads (and at other loci) because all copies are in phase at the SNP site.

(C) The relative heights of 100 other third-ranked peaks in the surrounding chromatograms fit a normal frequency distribution, as expected for background noise.

(D) The probabilities are extremely low that sampling from these noise distributions would yield a triple peak as pronounced as those highlighted (p values). The aggregate probability that six such peaks would arise randomly at known SNP sites is less than 2.05×10^{-34}. In contrast, the expected probability of conspecific cross-contamination in at least three tubes in species C exceeds 99% prior to considering these data.

See also Table S2.

Contamination with Conspecific DNA Explains "Intraspecific Horizontal Exchanges"

Given that at least seven samples interpreted as individuals by Debortoli et al. [18] were contaminated with DNA from a second animal belonging to a different species, it seems prudent to consider whether cases of "intraspecific DNA exchange" reported from the same dataset might correspond to samples that were contaminated with DNA from a second animal belonging to the same species. Debortoli et al. [18] describe three trios of "heterozygous individuals" in species A and C that seem to share alleles "in a cyclic fashion… (a||b), (b||c), and (c||a)." According to the authors, this pattern "can only be explained by recombination between individuals." However, an alternative hypothesis is that one or more tubes per trio contained DNA from two conspecific individuals.

Unlike heterospecific contamination, conspecific contamination cannot be tested by applying ConTAMPR to mtCO1 and 28S. Within species A and within species C, all samples share the same one or two 28S copies, and identical mtCO1 sequences are seen in samples with different nuclear alleles, so that DNA from a "heterozygous individual" or two homozygotes would look the same. Instead, we can apply probability theory. Evidence for inter- and intraspecific transfer relies on the same

576 tubes prepared from the same natural material. Given the relative abundances of *Adineta* species in that material and the established number of heterospecific contamination events, we calculate that the number of tubes expected to be contaminated with conspecific DNA is eight out of 82 (95% confidence interval [CI]: 4–15), even using highly conservative assumptions (see STAR Methods; Table S2).

Given its high relative abundance, species C is by far the most likely to be involved in conspecific contamination. The prior probability that one or more tubes contained DNA from two animals belonging to species C was calculated to exceed 99.9% (see STAR Methods), and the most likely number of such tubes is 7 (95% CI: 4–14). Debortoli et al. [18] present partial genomic data for three samples interpreted as individuals from species C, selected for sequencing post hoc because they seemed to form a conspecific "allele-sharing" trio at EPIC25. To test the hypothesis that one or more of these tubes (A112; D23; H4-28) was contaminated with conspecific DNA, we exploited a property of the EPIC63 nuclear marker. The two EPIC63 copies in a single heterozygous animal sometimes differ in length, owing to a 2 bp insertion-deletion polymorphism (indel), and thus run out of phase in chromatograms, producing double peaks from which each copy can be phased [42]. No matter how they are shifted,

though, two alleles cannot yield a triple peak. To find triple peaks demonstrates the presence of three or more alleles and is therefore inconsistent with DNA from a single rotifer (see STAR Methods). We tested this prediction for samples H4-28 and D23 (and D22, whose genotype appeared identical to D23). In each sample, we found two triple peaks matching known SNPs, revealing the four different alleles expected for DNA from at least two animals (Figure 2). Conspecific contamination in samples H4-28 and D23 explains not only the EPIC25 "allele-sharing" trio, but also genomic patterns at other loci that Debortoli et al. [18] attributed to "intraspecific horizontal exchange."

The problem of HR across unprecedented genetic distances does not affect purported intraspecific transfers, because the mean homology within conspecific trios is 99.1%. Instead, the opposite problem applies: when sequences differ at just 3–5 sites, apparent allele sharing "may be explained by mutations and gene conversions alone," as Debortoli et al. [18] remark. We tested this convergent substitution hypothesis for the remaining two trios (in species A) by analyzing shared polymorphisms (Figure S4). In both cases, a single point mutation and a single gene conversion event are sufficient to explain the whole cycle (see STAR Methods). Thus, intraspecific transfers are not required to explain these trios, no matter whether they are artifacts of conspecific contamination or not.

Conclusions and Prospects

Debortoli et al. [18] claimed to supply "strong evidence that inter- and intraspecific DNA exchanges occur within the bdelloid rotifer genus *Adineta*." However, chromatograms provided by the authors reveal that samples treated as individuals were accidentally contaminated with DNA from multiple animals, sometimes from different species and elsewhere from the same species. Contamination parsimoniously explains all of the genetic and genomic evidence offered for "inter- and intraspecific DNA exchanges," where any explanation is needed beyond mutation and gene conversion alone. These analyses constitute clear evidence that the data and findings of Debortoli et al. [18] are unreliable. The interesting hypothesis that "individuals of the genus *Adineta* exchange DNA within and between species" may be true or false, but the work supplies no credible evidence to address that question.

The study appears to have met with wide acceptance [43–47], perhaps because bdelloid asexuality has long been considered problematic [48], and the results seemed to confirm the enticing idea that sex is "replaced" by HGT [13, 14]. This view has a superficial appeal because bdelloid genomes encode a high proportion of non-metazoan genes relative to other animals, a finding that is in not in question here [13–17]. However, current formulations equating HGT with sex are simplistic and imprecise, and a clearer theoretical framework is needed to assess evolutionary plausibility. Sex has genome-wide consequences for every generation via crossing-over, segregation, independent assortment, outcrossing, and sexual selection. These underpin a plethora of formal models for the maintenance of amphimixis [49]. Fragmentary horizontal transfer (homologous or otherwise) has quite different population genetic consequences [50–53],

even if it could occur between bdelloids in the rapid manner proposed rather than on the much longer timescales estimated for acquisitions of non-metazoan genes [16].

The work of Debortoli et al. [18] was criticized by Signorovitch et al. [54], not on the grounds discussed here but in relation to their own prior claim of "a striking pattern of allele sharing consistent with sexual reproduction and with meiosis of an atypical sort" in the bdelloid *Macrotrachela quadricornifera* [55]. Debortoli et al. had argued that "our observations do not support the hypothesis of an *Oenothera*-like meiosis" in *Adineta*, because "for four nuclear markers, no trio of individuals presented congruent patterns of shared sequences" [56]. This lack of evidence for congruent allele sharing remains pertinent even in light of our findings, because cross-contamination would not obscure such a striking and widespread pattern as Signorovitch et al. [55] described. Aspects of that study seem to require empirical clarification, and we join the authors in "awaiting full genome sequencing of the allele-sharing individuals of *M. quadricornifera*" [54], which will shed further light.

Schwander [57] commented on the work of Signorovitch et al. [55] and Debortoli et al. [18], under the title "*The End of an Ancient Asexual Scandal*." Schwander asserts that "these two studies show beyond doubt that genetic exchange between individuals occurs in different bdelloid species." Schwander takes the position that "even small amounts of recombination and genetic exchange between individuals appear to be enough to provide all the benefits of sex," and therefore judges that "bdelloids should no longer be considered as asexuals." With the advantage of hindsight, we suggest that some of these rather strong assertions might now be considered premature.

In the words of Maynard Smith [48], bdelloid rotifers "remain something of an evolutionary scandal." Molecular investigations have revealed some remarkable characteristics in these tiny and unassuming creatures [13, 15, 17, 58–60], inviting speculation about links to their unusual reproductive mode. It is tempting to seek confirmation of these exciting possibilities and to expect further extraordinary discoveries. However, that approach opens the door to well-known biases [61], both when interpreting data and when evaluating the work of colleagues. Given concerns around contamination and misinterpretation, we join others in recommending that heightened scrutiny be applied to future claims of "non-canonical" genetic exchange in microscopic animals [62, 63] and to work on the molecular genetics of bdelloid rotifers in particular. A more cautious and incremental program will better facilitate firm progress and help exploit the unusual leverage that bdelloids offer on fundamental evolutionary and genetic questions.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - *Adineta vaga* (genome reference clone AD008)
  - *Adineta* sp. (AD006)

- **METHOD DETAILS**
  - Calculating distances for interspecific HR
  - Experimentally simulated cross-contamination
  - ConTAMPR method: validation and details
  - Chromatogram quality score analysis
  - Interspecific transfer: case-specific methods
  - Analysis of "intraspecific DNA exchanges"
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
- **DATA AND SOFTWARE AVAILABILITY**
- **ADDITIONAL RESOURCES**

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures, two tables, and one data file and can be found with this article online at https://doi.org/10.1016/j.cub.2018.05.070.

AUTHOR CONTRIBUTIONS

Conceptualization, C.G.W.; Methodology, C.G.W. and R.W.N.; Software, R.W.N.; Formal Analysis, C.G.W. and T.G.B.; Investigation, C.G.W.; Data Curation, C.G.W.; Writing – Original Draft, C.G.W.; Writing – Review & Editing, C.G.W., R.W.N., and T.G.B.; Visualization, C.G.W. and R.W.N.; Supervision, C.G.W. and T.G.B.; Project Administration, C.G.W. and T.G.B.; Funding Acquisition, C.G.W. and T.G.B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Judson, O.P., and Normark, B.B. (1996). Ancient asexual scandals. Trends Ecol. Evol. *11*, 41–46.

2. Normark, B.B., Judson, O.P., and Moran, N.A. (2003). Genomic signatures of ancient asexual lineages. Biol. J. Linn. Soc. Lond. *79*, 69–84.

3. Palmer, S.C., and Norton, R.A. (1991). Taxonomic, geographic and seasonal distribution of thelytokous parthenogenesis in the Desmonomata (Acari: Oribatida). Exp. Appl. Acarol. *12*, 67–81.

4. Smith, R.J., Kamiya, T., and Horne, D.J. (2006). Living males of the 'ancient asexual' Darwinulidae (Ostracoda: Crustacea). Proc. Biol. Sci. *273*, 1569–1578.

5. Schwander, T., Crespi, B.J., Gries, R., and Gries, G. (2013). Neutral and selection-driven decay of sexual traits in asexual stick insects. Proc. Biol. Sci. *280*, 20130823.

6. Wahab, A. (1962). Untersuchungen über Nematoden in den Drüsen des Kopfes der Ameisen (Formicidae). Zeitschrift für Morphol. und Ökologie der Tiere *52*, 33–92.

7. Snyder, D.W., Opperman, C.H., and Bird, D.M. (2006). A method for generating *Meloidogyne incognita* males. J. Nematol. *38*, 192–194.

8. Fradin, H., Kiontke, K., Zegar, C., Gutwein, M., Lucas, J., Kovtun, M., Corcoran, D.L., Baugh, L.R., Fitch, D.H.A., Piano, F., and Gunsalus, K.C. (2017). Genome architecture and evolution of a unichromosomal asexual nematode. Curr. Biol. *27*, 2928–2939.e6.

9. Hubbs, C., Drewry, G.E., and Warburton, B. (1959). Occurrence and morphology of a phenotypic male of a gynogenetic fish. Science *129*, 1227–1229.

10. Hudson, C.T., and Gosse, P.H. (1886). The Rotifera; or Wheel-Animalcules (Longmans, Green & Co.).

11. Mark Welch, D.B., Ricci, C., and Meselson, M. (2009). Bdelloid rotifers: progress in understanding the success of an evolutionary scandal. In Lost Sex, I. Schön, K. Martens, and P. Dijk, eds. (Springer), pp. 259–279.

12. Birky, C.W., Jr. (2010). Positively negative evidence for asexuality. J. Hered. *101* (Suppl 1), S42–S45.

13. Gladyshev, E.A., Meselson, M., and Arkhipova, I.R. (2008). Massive horizontal gene transfer in bdelloid rotifers. Science *320*, 1210–1213.

14. Flot, J.F., Hespeels, B., Li, X., Noel, B., Arkhipova, I., Danchin, E.G.J., Hejnol, A., Henrissat, B., Koszul, R., Aury, J.-M., et al. (2013). Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. Nature *500*, 453–457.

15. Boschetti, C., Carr, A., Crisp, A., Eyres, I., Wang-Koh, Y., Lubzens, E., Barraclough, T.G., Micklem, G., and Tunnacliffe, A. (2012). Biochemical diversification through foreign gene expression in bdelloid rotifers. PLoS Genet. *8*, e1003035.

16. Eyres, I., Boschetti, C., Crisp, A., Smith, T.P., Fontaneto, D., Tunnacliffe, A., and Barraclough, T.G. (2015). Horizontal gene transfer in bdelloid rotifers is ancient, ongoing and more frequent in species from desiccating habitats. BMC Biol. *13*, 90.

17. Nowell, R.W., Almeida, P., Wilson, C.G., Smith, T.P., Fontaneto, D., Crisp, A., Micklem, G., Tunnacliffe, A., Boschetti, C., and Barraclough, T.G. (2018). Comparative genomics of bdelloid rotifers: Insights from desiccating and nondesiccating species. PLoS Biol. *16*, e2004830.

18. Debortoli, N., Li, X., Eyres, I., Fontaneto, D., Hespeels, B., Tang, C.Q., Flot, J.F., and Van Doninck, K. (2016). Genetic exchange among bdelloid rotifers is more likely due to horizontal gene transfer than to meiotic sex. Curr. Biol. *26*, 723–732.

19. Hebert, P.D.N., Cywinska, A., Ball, S.L., and deWaard, J.R. (2003). Biological identifications through DNA barcodes. Proc. Biol. Sci. *270*, 313–321.

20. Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D., and Vogler, A.P. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst. Biol. *55*, 595–609.

21. Fontaneto, D., Barraclough, T.G., Chen, K., Ricci, C., and Herniou, E.A. (2008). Molecular evidence for broad-scale distributions in bdelloid rotifers: everything is not everywhere but most things are very widespread. Mol. Ecol. *17*, 3136–3146.

22. Wilson, C.G., and Sherman, P.W. (2013). Spatial and temporal escape from fungal parasitism in natural communities of anciently asexual bdelloid rotifers. Proc. Biol. Sci. *280*, 20131255.

23. Fontaneto, D., Iakovenko, N., Eyres, I., Kaya, M., Wyman, M., and Barraclough, T.G. (2011). Cryptic diversity in the genus *Adineta* Hudson & Gosse, 1886 (Rotifera: Bdelloidea: Adinetidae): a DNA taxonomy approach. Hydrobiologia *662*, 27–33.

24. Robeson, M.S., King, A.J., Freeman, K.R., Birky, C.W., Jr., Martin, A.P., and Schmidt, S.K. (2011). Soil rotifer communities are extremely diverse globally but spatially autocorrelated locally. Proc. Natl. Acad. Sci. USA *108*, 4406–4410.

25. Watt, V.M., Ingles, C.J., Urdea, M.S., and Rutter, W.J. (1985). Homology requirements for recombination in *Escherichia coli*. Proc. Natl. Acad. Sci. USA *82*, 4768–4772.

26. Zawadzki, P., Roberts, M.S., and Cohan, F.M. (1995). The log-linear relationship between sexual isolation and sequence divergence in *Bacillus* transformation is robust. Genetics *140*, 917–932.

27. Vulić, M., Dionisio, F., Taddei, F., and Radman, M. (1997). Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. Proc. Natl. Acad. Sci. USA *94*, 9763–9767.

28. Majewski, J. (2001). Sexual isolation in bacteria. FEMS Microbiol. Lett. *199*, 161–169.

29. Thomas, C.M., and Nielsen, K.M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat. Rev. Microbiol. *3*, 711–721.

30. Bell, J.S., and McCulloch, R. (2003). Mismatch repair regulates homologous recombination, but has little influence on antigenic variation, in *Trypanosoma brucei*. J. Biol. Chem. *278*, 45182–45188.

31. Datta, A., Hendrix, M., Lipsitch, M., and Jinks-Robertson, S. (1997). Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. Proc. Natl. Acad. Sci. USA *94*, 9757–9762.

32. Opperman, R., Emmanuel, E., and Levy, A.A. (2004). The effect of sequence divergence on recombination between direct repeats in *Arabidopsis*. Genetics *168*, 2207–2215.

33. Li, L., Jean, M., and Belzile, F. (2006). The impact of sequence divergence and DNA mismatch repair on homeologous recombination in *Arabidopsis*. Plant J. *45*, 908–916.

34. Larocque, J.R., and Jasin, M. (2010). Mechanisms of recombination between diverged sequences in wild-type and BLM-deficient mouse and human cells. Mol. Cell. Biol. *30*, 1887–1897.

35. Do, A.T., and LaRocque, J.R. (2015). The role of *Drosophila* mismatch repair in suppressing recombination between diverged sequences. Sci. Rep. *5*, 17601.

36. Fontaneto, D., Herniou, E.A., Boschetti, C., Caprioli, M., Melone, G., Ricci, C., and Barraclough, T.G. (2007). Independently evolving species in asexual bdelloid rotifers. PLoS Biol. *5*, e87.

37. Fontaneto, D., Kaya, M., Herniou, E.A., and Barraclough, T.G. (2009). Extreme levels of hidden diversity in microscopic animals (Rotifera) revealed by DNA taxonomy. Mol. Phylogenet. Evol. *53*, 182–189.

38. Iakovenko, N.S., Smykla, J., Convey, P., Kašparová, E., Kozeretska, I.A., Trokhymets, V., Dykyy, I., Plewka, M., Devetter, M., Duriš, Z., et al. (2015). Antarctic bdelloid rotifers: diversity, endemism and evolution. Hydrobiologia *761*, 5–43.

39. Barr, C.M., Neiman, M., and Taylor, D.R. (2005). Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. New Phytol. *168*, 39–50.

40. Liao, D. (2000). Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA genes in bacteria and archaea. J. Mol. Evol. *51*, 305–317.

41. Tang, C.Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T.G., and Fontaneto, D. (2012). The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. Proc. Natl. Acad. Sci. USA *109*, 16208–16212.

42. Flot, J.F. (2007). CHAMPURU 1.0: a computer software for unraveling mixtures of two DNA sequences of unequal lengths. Mol. Ecol. Notes *7*, 974–977.

43. Tilquin, A., and Kokko, H. (2016). What does the geography of parthenogenesis teach us about sex? Philos. Trans. R. Soc. Lond. B Biol. Sci. *371*, 20150538.

44. Krause, A., Pachl, P., Schulz, G., Lehmitz, R., Seniczak, A., Schaefer, I., Scheu, S., and Maraun, M. (2016). Convergent evolution of aquatic life by sexual and parthenogenetic oribatid mites. Exp. Appl. Acarol. *70*, 439–453.

45. Ram, Y., and Hadany, L. (2016). Condition-dependent sex: who does it, when and why? Philos. Trans. R. Soc. Lond. B Biol. Sci. *371*, 20150539.

46. Sharp, N.P., and Otto, S.P. (2016). Evolution of sex: using experimental genomics to select among competing theories. BioEssays *38*, 751–757.

47. Warren, W.C., García-Pérez, R., Xu, S., Lampert, K.P., Chalopin, D., Stöck, M., Loewe, L., Lu, Y., Kuderna, L., Minx, P., et al. (2018). Clonal polymor-phism and high heterozygosity in the celibate genome of the Amazon molly. Nat. Ecol. Evol. *2*, 669–679.

48. Maynard Smith, J. (1986). Evolution: contemplating life without sex. Nature *324*, 300–301.

49. Kondrashov, A.S. (1993). Classification of hypotheses on the advantage of amphimixis. J. Hered. *84*, 372–387.

50. Redfield, R.J. (2001). Do bacteria have sex? Nat. Rev. Genet. *2*, 634–639.

51. Narra, H.P., and Ochman, H. (2006). Of what use is sex to bacteria? Curr. Biol. *16*, R705–R710.

52. Agrawal, A.F. (2009). Spatial heterogeneity and the evolution of sex in diploids. Am. Nat. *174 (Suppl 1)*, S54–S70.

53. Croucher, N.J., Mostowy, R., Wymant, C., Turner, P., Bentley, S.D., and Fraser, C. (2016). Horizontal DNA transfer mechanisms of bacteria as weapons of intragenomic conflict. PLoS Biol. *14*, e1002394.

54. Signorovitch, A., Hur, J., Gladyshev, E., and Meselson, M. (2016). Evidence for meiotic sex in bdelloid rotifers. Curr. Biol. *26*, R754–R755.

55. Signorovitch, A., Hur, J., Gladyshev, E., and Meselson, M. (2015). Allele sharing and evidence for sexuality in a mitochondrial clade of bdelloid rotifers. Genetics *200*, 581–590.

56. Flot, J.F., Debortoli, N., Hallet, B., and Van Doninck, K. (2016). Response to Signorovitch et al. Curr. Biol. *26*, R755.

57. Schwander, T. (2016). Evolution: the end of an ancient asexual scandal. Curr. Biol. *26*, R233–R235.

58. Arkhipova, I., and Meselson, M. (2000). Transposable elements in sexual and ancient asexual taxa. Proc. Natl. Acad. Sci. USA *97*, 14473–14477.

59. Mark Welch, D.B., Mark Welch, J.L., and Meselson, M. (2008). Evidence for degenerate tetraploidy in bdelloid rotifers. Proc. Natl. Acad. Sci. USA *105*, 5145–5149.

60. Arkhipova, I.R., Yushenova, I.A., and Rodriguez, F. (2017). Giant reverse transcriptase-encoding transposable elements at telomeres. Mol. Biol. Evol. *34*, 2245–2257.

61. Nickerson, R.S. (1998). Confirmation bias: a ubiquitous phenomenon in many guises. Rev. Gen. Psychol. *2*, 175–220.

62. Richards, T.A., and Monier, A. (2016). A tale of two tardigrades. Proc. Natl. Acad. Sci. USA *113*, 4892–4894.

63. Koutsovoulos, G., Kumar, S., Laetsch, D.R., Stevens, L., Daub, J., Conlon, C., Maroon, H., Thomas, F., Aboobaker, A.A., and Blaxter, M. (2016). No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*. Proc. Natl. Acad. Sci. USA *113*, 5053–5058.

64. Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. *3*, 294–299.

65. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.

66. Mark Welch, J.L., and Meselson, M. (1998). Karyotypes of bdelloid rotifers from three families. Hydrobiologia *387*, 403–407.

67. Wilson, C.G., and Sherman, P.W. (2010). Anciently asexual bdelloid rotifers escape lethal fungal parasites by drying up and blowing away. Science *327*, 574–576.

68. Barron, G. (2004). Fungal parasites and predators of rotifers, nematodes, and other invertebrates. In Biodiversity of Fungi: Inventory and Monitoring Methods, G.M. Mueller, G.F. Bills, and M.S. Foster, eds. (Elsevier Academic Press), pp. 435–450.

69. Katoh, K., Asimenos, G., and Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT. Methods Mol. Biol. *537*, 39–64.

70. Hur, J.H., Van Doninck, K., Mandigo, M.L., and Meselson, M. (2009). Degenerate tetraploidy was established before bdelloid rotifer families diverged. Mol. Biol. Evol. *26*, 375–383.

71. Eyres, I., Frangedakis, E., Fontaneto, D., Herniou, E.A., Boschetti, C., Carr, A., Micklem, G., Tunnacliffe, A., and Barraclough, T.G. (2012). Multiple

functionally divergent and conserved copies of alpha tubulin in bdelloid rotifers. BMC Evol. Biol. *12*, 148.

72. McVey, M., and Lee, S.E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet. *24*, 529–538.

73. Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. (2016). BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics *32*, 767–769.

74. Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. *40*, e49.

75. Nei, M., and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. *3*, 418–426.

76. Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigian, C., Fuellen, G., Gilbert, J.G.R., Korf, I., Lapp, H., et al. (2002). The Bioperl toolkit: Perl modules for the life sciences. Genome Res. *12*, 1611–1618.

77. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

78. Da Ines, O., Gallego, M.E., and White, C.I. (2014). Recombination-independent mechanisms and pairing of homologous chromosomes during meiosis in plants. Mol. Plant *7*, 492–501.

79. Gladyshev, E., and Kleckner, N. (2016). Recombination-independent recognition of DNA homology for repeat-induced point mutation (RIP) is modulated by the underlying nucleotide sequence. PLoS Genet. *12*, e1006015.

80. Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using *phred*. II. Error probabilities. Genome Res. *8*, 186–194.

81. Kronick, M.N. (1997). Heterozygote determination using automated DNA sequencing technology. In Laboratory Methods for the Detection of Mutations and Polymorphisms in DNA, G.R. Taylor, ed. (CRC Press), pp. 175–190.

82. Mullis, K.B., Ferre, F., and Gibbs, R.A. (1994). The Polymerase Chain Reaction (Birkhäuser).

83. Segers, H., and Shiel, R.J. (2005). Tale of a sleeping beauty: a new and easily cultured model organism for experimental studies on bdelloid rotifers. Hydrobiologia *546*, 141–145.

84. Agresti, A. (2007). An Introduction to Categorical Data Analysis, Second Edition (John Wiley & Sons).

85. Bangalore, S.S., Wang, J., and Allison, D.B. (2009). How accurate are the extremely small P-values used in genomic research: an evaluation of numerical libraries. Comput. Stat. Data Anal. *53*, 2446–2452.

86. Han, T., Chang, C.-W., Kwekel, J.C., Chen, Y., Ge, Y., Martinez-Murillo, F., Roscoe, D., Težak, Z., Philip, R., Bijwaard, K., and Fuscoe, J.C. (2012). Characterization of whole genome amplified (WGA) DNA for use in genotyping assay development. BMC Genomics *13*, 217.

87. Polz, M.F., and Cavanaugh, C.M. (1998). Bias in template-to-product ratios in multitemplate PCR. Appl. Environ. Microbiol. *64*, 3724–3730.

88. Yim, E., O'Connell, K.E., St Charles, J., and Petes, T.D. (2014). High-resolution mapping of two types of spontaneous mitotic gene conversion events in *Saccharomyces cerevisiae*. Genetics *198*, 181–192.

89. Do, A.T., Brooks, J.T., Le Neveu, M.K., and LaRocque, J.R. (2014). Double-strand break repair assays determine pathway choice and structure of gene conversion events in *Drosophila melanogaster*. G3 (Bethesda) *4*, 425–432.

90. Williams, A.L., Genovese, G., Dyer, T., Altemose, N., Truax, K., Jun, G., Patterson, N., Myers, S.R., Curran, J.E., Duggirala, R., et al.; T2D-GENES Consortium (2015). Non-crossover gene conversions show strong GC bias and unexpected clustering in humans. eLife *4*, 1–21.

91. Miles, A., Iqbal, Z., Vauterin, P., Pearson, R., Campino, S., Theron, M., Gould, K., Mead, D., Drury, E., O'Brien, J., et al. (2016). Indels, structural variation, and recombination drive genomic diversity in *Plasmodium falciparum*. Genome Res. *26*, 1288–1299.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
Escherichia coli OP50	Caenorhabditis Genetics Center	OP50
Critical Commercial Assays		
illustra PuReTaq Ready-To-Go PCR Beads	GE Healthcare	Cat#27955702
Macrogen Europe ABI Sequencing Service	Macrogen Europe	https://dna.macrogen.com/
Deposited Data		
mtCO1 chromatograms for ConTAMPR validation	This paper	https://doi.org/10.17632/fsjpwrnv4t.1
Original sequencing chromatograms generated by Debortoli et al. [18]	Debortoli et al. [18]	https://github.com/jflot/Debortoli2016CurrentBiology
DNA sequences inferred by Debortoli et al. [18]	Debortoli et al. [18]	GenBank: KU860573–KU861170
Reference DNA sequences for *A. vaga* and *A.* sp. AD006	Flot et al. [14]; this paper	GenBank: GQ398061, JX184001, EF173187, KM043216, KM043183
A. vaga reference genome data	Flot et al. [14]	GenBank: GCA_000513175.1
A. vaga RNA-Seq data	Flot et al. [14]	SRA: ERR260376
Experimental Models: Organisms/Strains		
Adineta vaga (genome reference clone)	Gift of K. Van Doninck	AD008
Adineta sp. (clone AD006)	51° 24' 32.06" N 0° 38' 41.71" W in moss	AD006
Saccharomyces cerevisiae (strain S288c)	Gift of V. Koufopanou	S288c
Oligonucleotides		
LCO1-1490: 5'-GGTCAACAAATCATAAAGATATTGG-3' Folmer et al. [64]	Thermo Fisher Scientific	LCO1
HCO1-2198, 5'-TAAACTTCAGGGTGACCAAAAAATCA-3' Folmer et al. [64]	Thermo Fisher Scientific	HCO1
Software and Algorithms		
R version 3.3.1: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.	R Core Team, 2016	https://www.r-project.org/
Geneious (v. 8.1.9, Biomatters Ltd, Auckland, New Zealand)	Kearse et al. [65]	https://www.geneious.com/
Scripts for regional and genomic analysis of microhomology	This paper	https://github.com/reubwn/microhomology
CodonCode Aligner (v. 7.0.1)	CodonCode Corporation	http://www.codoncode.com/
Other		
Protocol for needle isolation of bdelloid rotifer individuals	This paper	https://www.biorxiv.org/highwire/filestream/45474/field_highwire_adjunct_files/1/150490-2.pdf
Two preprint versions of this paper, with public comments	http://www.bioRxiv.org, operated by Cold Spring Harbor Laboratory	https://doi.org/10.1101/150490

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Christopher G. Wilson (chris.wilson@imperial.ac.uk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Adineta vaga (genome reference clone AD008)

For validation of ConTAMPR we used cultures of *Adineta vaga* belonging to the clone that supplied DNA for a reference genome [14], kindly provided to us by K. Van Doninck on 2013-04-11. We call this clone "*A. vaga* (AD008)" or "*A. vaga* (genome)." We think it was collected originally in Italy prior to 1998 [66]. All individuals used were reproductive females. This clone was cultured using methods described previously [67]. Every two to three weeks, populations were moved to fresh dishes of sterile distilled water over Czapek-Dox 0% agar [68]. Cultures were fed with a standardized inoculum of *Escherichia coli* (OP50), and *Saccharomyces cerevisiae* (S288c), and were maintained at 20°C in an illuminated incubator (LMS, Kent, UK) with a 12 hr light: 12 hr dark cycle.

Adineta sp. (AD006)

During validation of ConTAMPR, we used cultures of a second bdelloid rotifer clone in the genus *Adineta* that was isolated from *Brachythecium rutabulum* (Hedwig), growing on *Quercus* sp. at Silwood Park, Ascot, UK (51° 24' 32.06" N 0° 38' 41.71" W), and kept in continuous culture since 2012-01-09. We call this clone "*A.* sp. (AD006)." It is maintained as described above for *A. vaga*. All individuals used were reproductive females.

METHOD DETAILS

Calculating distances for interspecific HR
Establishing reference values

Debortoli et al. [18] remark that interspecific horizontal transfers "may be mediated by double-strand break repair through homologous recombination (HR)." In their view, "this hypothesis is reinforced by the observation that… the transferred sequences replaced the original copies." In the context of HGT, Thomas and Nielsen [29] define HR as "recombination that depends on extensive segments of high sequence similarity between two DNA molecules." A log-linear decline in HR with decreasing sequence identity is well established in bacteria [25–29]; protists [30]; fungi [31]; plants [32, 33] and animals [34, 35]. Datta et al. [31] found that a single mismatched base (99.7% identity) reduced HR rates fourfold in yeast, and reductions to 99% or 90% identity reduced HR rates by one and two orders of magnitude respectively. At these distances, the relationship is largely governed by active mismatch repair systems; HR does not decline so dramatically if these are abolished. However, when sequence identity falls below 90%, rates of HR decline exponentially even in mutants lacking mismatch repair, indicating that the machinery of HR itself fails to engage when so many bases are mispaired. At 83.5% identity, HR is effectively absent in yeast (three recombinant cells per billion [31]), and rates were too low to measure even for otherwise promiscuous *Bacillus* species [26].

Determining interspecific homology values

To quantify pairwise homology between sequences involved in "interspecific recombination," we downloaded each incongruent marker sequence from GenBank (KU860573–KU861170), together with an orthologous sequence for the "recipient" species whose native copies were supposed to have been overwritten. These pairs of copies were imported to Geneious (v. 8.1.9, Biomatters Ltd, Auckland, New Zealand) [65] and aligned using the MAFFT v. 7.017 algorithm [69], implemented via the Geneious plugin (v. 1.3.3) with default settings. This recreates the genetic divergences across which Debortoli et al. [18] posit multiple, recent HR events. We recorded the pairwise identity for each alignment and the GC content of each sequence, as reported in the Geneious "Statistics" interface (Table 1). Where two sequences were "transferred" together, we took the mean of both pairs of contrasts. The mean pairwise identity across all markers was 72.7% (median 68.9%).

Validating distances at wider genomic scales

The genetic distances in Table 1 are estimated for short (< 1kb) markers that Debortoli et al. [18] selected for easy amplification. However, horizontal transfer events were claimed to extend several kilobases beyond the markers they encompass (Figure 5 in Debortoli et al. [18]). We considered the possibility that genomic regions beyond the focal markers might show substantially greater sequence homology between species, thereby decreasing the mechanistic implausibility of HGT. For example, the EPIC25 marker (ca. 400 bp) spans the first intron of a gene approximately 4.3 kb long, encoding a product with similarity to the vertebrate metastasis suppressor protein 1 (MTSS1). In the *A. vaga* reference genome assembly [14], the next-closest gene is approximately 2kb from the marker in a 5′ direction and encodes a product with similarity to vertebrate trifunctional enzyme subunit beta (HADBH). As introns often are highly variable, pairwise identity at the EPIC25 marker might underestimate the homology between two species for the shared genes in this region.

   We tested this hypothesis for the "interspecific recombination" event inferred from Sample B22 ("Individual 58"), in which EPIC25 Hap10 [C] was putatively replaced by Hap37 [E]. At the marker itself, these sequences only shared 68.4% genetic identity. We wanted to determine whether the value might be higher if we considered the wider region. While investigating putative intraspecific exchange, Debortoli et al. [18] sequenced a longer (10.8kb) region surrounding EPIC25 Hap10 for another sample from Species C (A112). The region containing the genes *MTSS1* and *HADBH* was syntenic with the *A. vaga* reference genome. To determine homology, it would be ideal to align this extended region with its counterpart in Species E, but no genome sequence for E is available. However, the reference genome clone of *A. vaga* happens to be very closely related to Species E. In fact, homology to Hap10 [C] at the EPIC25 marker is identical (68.4%) for Hap37 [E] and the *A. vaga* reference. We compared the equivalent identities at all five available loci, and in each case, the homology to Species C was nearly identical for Species E and the *A. vaga* reference genome (Figure S1A).

This coincidence enabled us to use the *A. vaga* reference genome as a surrogate for Species E, with some confidence that the results would reflect the relationship with Species C.

We aligned the sequenced EPIC25 region from Individual 42 [C] (GenBank: KU861136.1) against the matching region from the reference genome (Assembly GCA_000513175.1) using MAFFT as implemented in Geneious, with default settings. For clarity, we delineated a focal region from the stop codon of *MTSS1* to the stop codon of *HADBH* (8949 bp, approximately centered on EPIC25). This included exons and introns from both complete genes, and the intergenic region. We then measured pairwise identity between the species for a variety of subregions (Figure S1B). The homology between these species for the whole region is 62.3%, which is lower than the estimate based on the EPIC25 marker. The intergenic regions have very low identity (52.4%), but even if we only compare genes, the values are either lower than the estimate from the EPIC25 marker (for *HADBH*), or identical to it (for *MTSS1*). Thus, we can reject the hypothesis that the marker-based identities in Table 1 underestimate the homology between species for broader genomic regions.

Conflicts with A. vaga genome structure

Given the distances involved, the proposed interspecific transfers appear to conflict not only with established understanding of HR from model systems, but with the genome structure of *A. vaga* itself. Bdelloid rotifers are ancestrally tetraploid, and thousands of genes occur as two divergent pairs of copies, in parallel lineages that appear to arise from an ancient duplication [17, 59, 70]. By convention, each of these genes has a closely related 'homolog', and a pair of distantly related 'ohnologs' [14]. Homologous copies within pairs "are on average 96.2% identical at the nucleotide level (median = 98.6%)" [14], which indicates concerted evolution via gene conversion, intragenomic recombination and perhaps other mechanisms [17]. In contrast, ohnologs are so divergent that homologous exchange between the two lineages appears minimal [70]. The mean identity between ohnologs is "73.6% (median = 75.1%)" [14].

This feature of *A. vaga* does not seem compatible with the hypothesis that genes are subject to very frequent horizontal substitution by heterospecific DNA fragments whose identity is 72.7% or lower. If that were so, ohnologs whose identity exceeded this value could not stably diverge and coexist in the same genome, nor could multi-copy gene families (e.g., [71]). Each copy would be subject to regular horizontal replacement by exogenous DNA fragments containing the others, either from different cells in the same animal or other animals from the same clone. On average, conspecific DNA fragments must be more abundant and accessible to rotifers than heterospecific DNA, and the copies have even higher homology. For identity values exceeding those in Table 1, copies would be drawn into concerted evolution via whatever mechanisms mediate horizontal exchange. Pairs of ohnologs would then only evolve independently if their identities fell substantially below 72.7%. These predictions are rejected because the median identity value for *A. vaga* ohnologs is 75.1%, with considerable variation.

Interspecific and intragenomic microhomology

Although mechanisms of interspecific recombination based on overall homology seem to be excluded, we also considered alternative mechanisms with less stringent identity requirements than HR (N. Debortoli, J.-F. Flot, K. Van Doninck, pers. comm.). One example is homology-facilitated illegitimate recombination (HFID), in which "single regions of high nucleotide-sequence similarity (∼200 bp in length)… initiate recombination events that lead to the additive integration of >1000-bp-long heterologous DNA fragments" [29]. Another is microhomology-mediated end joining (MMEJ), whose "foremost distinguishing property… is the use of 5-25bp microhomologous sequences during the alignment of broken ends before joining" [72]. If interspecifically transferred sequences share longer or more frequent mismatch-free blocks than genomic ohnologs [27], then HFID or MMEJ may explain how they undergo such frequent horizontal exchange whereas thousands of ohnologous pairs with higher global homology evolve independently.

To test this hypothesis, we estimated the length and frequency of microhomologous blocks for genes surrounding the putative transfer event seen in Sample B22, between EPIC25 Hap10 [C] and Hap37 [E]. As described above, we used the reference genome for *A. vaga* as a surrogate for Species E (Figure S1A). At every scale from 1-40 bp, we compared interspecific microhomology for *MTSS1-A* and *HADBH* with intragenomic microhomology for 7650 ohnologous pairs of genes in the *A. vaga* reference genome. As an important point of comparison, we highlighted the microhomology between *MTSS1-A* and its own genomic ohnolog (*MTSS1-B*), which contains the EPIC63 marker. Debortoli et al. [18] implicitly assume there is no exchange between these highly divergent genes, which were treated as "independent nuclear markers" and sequenced with "ohnologue-specific" primers.

Gene models for *A. vaga* were constructed using BRAKER [73], with RNASeq as evidence (SRA accession: ERR260376 [14]) Collinear regions were identified using MCScanX [74], setting the maximum number of allowed gaps between collinear genes to 10. Between all pairs of collinear genes, synonymous (K_S) and nonsynonymous (K_A) divergences were calculated using the method of Nei and Gojobori [75], implemented in BioPerl [76]. Ohnologs were defined as pairs of genes within collinear regions with $K_S \geq 0.5$ (n = 7,650). Ohnologous regions (comprising exons plus introns) were extracted and aligned using MAFFT [77] with default settings. Introns were included because interspecific recombination was claimed to involve intronic markers, and inter-ohnolog recombination facilitated by microhomology would involve unspliced genomic DNA. Microhomology across all 7,650 alignments was calculated with a custom Perl script, using a sliding window along each alignment (sliding one base each iteration), from a size of one to 40 bp, counting any window of exact identity as a match. The number of identical blocks was scaled relative to alignment length to account for gene size, and multiplied by 1,000 to give a per-kb estimate. All scripts are available at https://github.com/reubwn/microhomology. The same method was used to measure microhomology between Species C and *A. vaga* for the alignments of *MTSS1-A* and *HADBH* discussed above. Finally, we highlighted the specific microhomology curve in *A. vaga* corresponding to the pairing between *MTSS1-A* and its ohnolog, *MTSS1-B*.

Gene copies involved in putative interspecific recombination did not share significantly more or longer blocks of microhomology than independently evolving ohnologs in the same genomes, at any scale from one to 40 bp (Figure S1C). At most scales they even shared less microhomology with the supposed donor species than with their own ohnologs. It seems most implausible that micro-homology-based mechanisms such as MMEJ and HFIR would be able to facilitate interspecific recombination with distantly related DNA fragments, yet fail to permit horizontal exchange between conspecific ohnologous copies with even more extensive microho-mology, especially because conspecific DNA fragments must be more abundant and available than heterospecific ones.

Other mechanisms of homology recognition seem even less applicable. For instance, pairing of chromosomes in early meiosis appears to be independent of recombination in some cases [78], but this involves chromosome-scale architecture such as centromeres and telomeres, which are not shared by loose DNA fragments. Pairing must still be stabilized by recombination, via sequence-dependent pathways. Another recognition mechanism involves homologous trinucleotide repeats interspersed at a specific periodicity within otherwise divergent sequences [79], but there is no evidence for this distinctive architecture in the putatively exchanged sequences we examined, and that pathway is not linked to recombination. It seems improbable *prima facie* that any mechanism could enable ready exchange of heterospecific DNA fragments, while simultaneously precluding exchange between less divergent ohnologs via conspecific DNA fragments. We suggest that any argument to the contrary must bear the burden of proof.

Integration of DNA over large genetic distances must be possible at least occasionally in bdelloid rotifers, since genes with little or no homology to metazoan sequences are incorporated [13]. However, the mechanism in these cases would be additive rather than substitutive; it would not involve HR and would not prevent ohnologs evolving independently, nor would it distinguish between exogenous DNA from rotifers versus non-rotifer sources. It may involve a combination of unlikely events, which would explain why absolute rates are estimated to be quite low: on the order of ten events per million years [16].

Experimentally simulated cross-contamination
Design and replication
We conducted an experiment to determine the expected pattern of mtCO1 chromatogram peaks when two rotifers are known to be present in one tube. According to Debortoli et al. [18], the most frequent signatures of "interspecific recombination" were observed between Species A and E, and between Species C and E. These pairs of species share 86.1% and 86.4% sequence identity respectively at the mtCO1 marker. The identity between our cultured *Adineta* clones AD008 and AD006 at the mtCO1 marker is 86.5%, and thus commensurate with the species pairs involved in potential cross-contamination. We therefore prepared tubes to experimentally simulate cross-contamination involving individuals of these two species.

We prepared 12 tubes, divided into three groups. Biological replicates refer to different tubes, technical replicates refer to repeated PCR and sequencing using the same tube of template DNA. Tubes in Group "1X6" each contained a single animal from clone AD006, with three biological replicates, one of which was also assigned three technical replicates. Tubes in group "1X8" each contained a single animal from clone AD008, with two biological replicates (a third was lost). One tube was assigned three technical replicates. Tubes in Group "2X6-8" each contained a single animal from clone AD006 and a single animal from clone AD008, with six biological replicates. One tube was assigned three technical replicates and two were assigned two technical replicates.

Rotifer isolation
For our experiment, it was critical to be certain of the exact number of rotifers in each tube. The methods and references provided by Debortoli et al. [18] do not describe the technique used to isolate animals. We requested a protocol from the authors, and were given the following summary:

> "Our procedure is simple, we collect the lichen/grass patch and put it in Spa® water overnight. The next day, we isolate the individuals identified as A. vaga by pipetting and washing them in clean water drops (serial dilutions). We then carefully checked under the binocular each tube to make sure that only one individual was present." (N. Debortoli and K. Van Doninck, personal communication)

This description raised some technical concerns for us. Prior to 2014, we had employed a similar approach, but we found that pipettes were unsuitable for systematic isolation of individuals from nature, and that quality control procedures requiring visual inspection of Eppendorf tubes were inherently unreliable. Bdelloid rotifers often rest motionless and invisible under the distortion of a meniscus or against the plastic base of a tube. They attach themselves to the inside of pipette tips, where they become very difficult to dislodge or even to see. In addition to wasting time and consumables, loss of specimens in this way will bias the subset of animals successfully retrieved by the protocol. We occasionally experienced contamination via the following events. An individual is serially washed and deposited into a tube by pipette. The tube is carefully checked under the binocular microscope, but no animal is detected. It is assumed that the specimen is stuck in the pipette tip, but this cannot be verified directly. The tip is changed and a second rotifer is serially washed and placed in the tube. This time, when the tube is carefully checked, a rotifer is visually confirmed, and the tube is sealed for DNA extraction. In fact, the first animal also entered the tube but was hidden; thus, two rotifers now share one tube. This is not a particular problem when isolating animals from a clonal culture, but becomes a critical technical issue when isolating genetically different animals from nature. Opportunities for contamination increase when large numbers of animals must be isolated.

To address these problems, we now use a needle-based protocol when isolating individuals (see Additional Resources). This was used to prepare samples to simulate cross-contamination. Into each of the 12 experimental tubes, 8μL of sterile Milli-Q water was pipetted. To each tube in groups 1X6 and 2X6-8, the needle protocol was used to move a single rotifer from a stock culture of *A.* sp.

(AD006), via a wash droplet of 1mL sterile Milli-Q water. To each tube in groups 1X8 and 2X6-8, a single rotifer was moved from a stock culture of *A. vaga* (AD008). The tubes in group 2X6-8 therefore contained two rotifers, one from each species.

DNA extraction, sequencing and replicability

We extracted DNA from these samples and amplified the mitochondrial cytochrome oxidase I (mtCO1) marker by PCR using the methods described by Debortoli et al. [18]. We used the same primers (LCO1 and HCO1 [64]) at the same concentrations, and the same concentration of template in the same reaction volume (25µL). Amplifications were performed using GE Healthcare illustra PuReTaq Ready-To-Go PCR Beads. PCR products were purified, and sequenced in both directions with the same primers using an ABI 3730xl DNA Analyzer (Applied Biosystems), via a commercial Sanger sequencing service (Macrogen Europe, Amsterdam, the Netherlands).

The Phred quality scores of the chromatogram files [80] were only slightly and not significantly lower for tubes with two animals versus one (mean Q20: 90% versus 92%, N = 38, t = 1.13, p = 0.26). Even when two animals were present, the vast majority of base calls (97.9%–99.6%) matched a single species (Table S1). The additional animal did not manifest as obvious double peaks at the expected polymorphic sites, but as small, subtle minority peaks, typically hidden in baseline noise [81] and sometimes missing entirely. Perhaps this is unsurprising: double peaks are seldom equal in height even when amplifying alleles from diploid heterozygotes [81]. PCR is exponential, and if two animals contribute divergent templates, large biases in final amplicon representation may arise from small initial differences in numbers of cells or mitochondria, or the efficiency of lysis, DNA extraction, primer binding, denaturation, etc [82].

Technical replicates within the 2X6-8 group were concordant and the same clone always dominated the amplicon pool, but biological replicates showed different dominant clones (Table S1). For three samples, *A. vaga* (AD008) supplied the majority mitochondrial haplotype (99.5%, 99.3% and 99.6% of bases called); for the other three samples, *A.* sp. (AD006) was in the majority (99.6%, 99.7%, 97.9%). Concordance among technical replicates suggests that small differences in efficiency of lysis or DNA extraction are at least as important as differences during PCR in determining which of the templates is amplified. Consequently, it may not be surprising to see a consistent majority sequence when amplifying repeatedly from the same sample, even if it contained multiple animals. The discordance among biological replicates suggests that the direction of the bias may be inconsistent between samples, even when the same two species are involved.

ConTAMPR method: validation and details

Overview and experimental validation

We developed a simple quantitative method, called ConTAMPR (Contingency Table Analysis of Minority Peak Ranks), to test whether the identity of the additional animal in deliberately contaminated samples could be recovered from the pattern of minority peaks. Briefly, chromatograms were aligned bidirectionally with the known sequences of both *Adineta* clones. At each site where the two reference sequences differed, we manually ranked the heights of the fluorescence trace lines corresponding to the three minority nucleotide bases, and recorded the rank for the base fitting the known contaminant sequence. Using contingency table analysis, we tested whether the distribution of peak height ranks differed from the expectation if chromatogram noise were unrelated to the query sequence. For additional rigor, we also measured and tested the distribution of minority peaks fitted to a control species (*A. ricciae* [83]) that was not present in any sample, but whose sequence identity to both AD006 and AD008 at mtCO1 was equal (87.5%).

In an uncontaminated single-animal sample containing *A.* sp. (AD006), rank distributions of minority peaks matching *A. vaga* (AD008) and *A. ricciae* did not differ significantly from the null expectation or from each other (Figure S2A; $\chi^2 = 3.32$, d.f. = 4, p = 0.51). In a deliberately contaminated sample, however, minority peaks fitted the known contaminant species significantly better than the null expectation (Figure S2B; $\chi^2 = 174.54$, d.f. = 2, p = 1.26×10^{-38}) or the control species ($\chi^2 = 22.54$, d.f. = 2, p = 1.27×10^{-5}). ConTAMPR correctly detected that this tube contained two animals, identified the second sequence and differentiated it from another candidate. The results in the left bar graphs are for Sample 1X6_01a, which only contained A. sp. (AD006), and Sample 2X6-8_03a, a deliberately contaminated sample where *A.* sp. (AD006) was in the majority and *A. vaga* (AD008) in the minority. To check robustness in the reciprocal case, we repeated the analysis for Samples 1X8_02c and 2X6-8-05a, where *A. vaga* (AD008) was in the majority. The outcomes were as expected (Figure S3, right bar graphs). Again, the minority peaks for AD006 often were hidden within noise associated with polymerase slippage and other errors, but the contaminant was recovered via ConTAMPR. For comparison, Figure S2C re-plots equivalent data for the same section of Sample B11 that is shown in Figure 1.

This method works because PCR amplicons were directly sequenced, leaving visual evidence of minority haplotypes. An approach using post-PCR cloning might have missed minority sequences entirely [82]. The ConTAMPR method was therefore considered suitable to test the directly sequenced amplicons of Debortoli et al. [18] for contamination. It may also have applications for detecting putative contaminants in other datasets where direct Sanger sequencing has been applied.

Multiple sequence alignment procedures

To test for the signatures described above, we requested raw chromatogram files from Debortoli and colleagues in 2016. We promptly received mtCO1 files for the six key samples where interspecific transfers had been claimed. We were sent 36 mtCO1 files (3.1% of the 1152 mtCO1 chromatograms generated in the study), 483 out of 656 chromatograms for 28S ribosomal DNA (73.6%) and 133 out of 158 files for the EPIC25 marker (84.2%). We did not initially receive files for EPIC63 or Nu1054. In 2017 we received more chromatograms, taking the totals to 1102 out of 1152 expected files for mtCO1 (95.7%); 534 out of 656 for 28S (81.4%); 140 out of 158 for EPIC25 (88.6%); 143 out of 162 for EPIC63 (88.3%) and 120 out of 130 files for Nu1054 (92.3%). The files are now available

to download at: https://github.com/jflot/Debortoli2016CurrentBiology. We are grateful to the authors for their open and collaborative sharing of data.

All sequences delimited by Debortoli et al. [18] were retrieved from GenBank (KU860573–KU861170), along with relevant sequences for the *A. vaga* reference clone (GQ398061; JX184001), *A. ricciae* (EF173187; KM043216) and A. sp. AD006 (KM043183). Multiple alignment of chromatograms and candidate sequences was performed using MAFFT v. 7.017 [69], implemented in Geneious using the MAFFT plugin (v. 1.3.3). For some sequences, particularly at mtCO1, this algorithm alone was sufficient to bring peaks for the majority and putative minority sequences into alignment with the GenBank references. This is because there are no insertion-deletion polymorphisms (indels) between species for the mtCO1 marker, so peaks corresponding to amplicons from different templates are superimposed ('in phase'), though we typically saw slight displacement of minority peaks by less than a base-width in one direction or another relative to the majority peak. This is illustrated in Data S1, using annotated screenshots from the visual interface of Geneious.

The 28S ribosomal marker was amplified by Debortoli et al. [18] in four overlapping fragments. This locus is highly conserved among even distantly related species. We restricted all our analyses to the first fragment, amplified with the primers 28S0FCT and 28S1RCT. Among the six *Adineta* species reported by Debortoli et al. [18], this fragment has 70 variable sites in 700 bp (10%), whereas the remaining 1610 bp includes only 24 variable sites (1.5%). Multiple alignment of all unique sequences for *A. vaga* Species A-F revealed three indels of 1-2 bp within a single 100 bp region of the focal first fragment. This occasionally created challenges in testing whether two sequences from different species were present in an amplicon population. For instance, Sample B14 was predicted to include molecules from Species A and E. There is a single 1 bp indel between these species, which means the minority peaks in a forward chromatogram are predicted to run approximately in phase with the majority peaks until the indel, then become misaligned by 1 bp, whereas the minority peaks in a reverse chromatogram will show the opposite pattern [42]. The contaminant might thus be mistaken for a polymerase slippage artifact [82]. To analyze such a pattern, it is necessary to manually shift the alignment of each chromatogram by 1 bp around the indel, which means the majority peaks are out of phase with the aligned majority sequence for part of its length. This alignment shift is illustrated in Data S1. For clarity, we point out every peak corresponding to the minority Species E sequence in both directions. The pattern of multiple peaks changes as predicted at the site of the 1 bp indel between Species A and Species E, and this feature represents further, sequence-independent evidence for the additional gene copy.

EPIC25 is a highly variable intronic marker with multiple indels of up to 12 bp within and 23 bp between *A. vaga* Species A-F. Superimposed sequences from different species are therefore displaced even further and more frequently than at 28S, which makes it challenging to align even one candidate allele to the minority peaks. Multiple manual adjustments to an initial MAFFT alignment were necessary to follow the minority allele after each indel. It would not be feasible to attempt to align sequences from more than two candidate species at once against these chromatograms, since the peaks and variable sites would almost never be predicted to coincide with both candidates. However, it was not necessary to align every species individually to test for the EPIC25 sequences listed in Table 2, because the predicted matches were so clear relative to the null expectation, and the alignment constraints made a match to any alternative sequence so improbable. Data S1 illustrates the alignment shifts necessary to follow a minority Species A allele running alongside a Species C sequence with at least 5 dispersed indels of varying lengths. Each peak we interpret as corresponding to the minority sequence is individually highlighted for the avoidance of ambiguity.

Peak rank assignments

After bidirectional pairs of chromatograms had been aligned with candidate sequences, we manually scored the relative heights of minority fluorescence peaks at each site where a candidate sequence differed from the majority reference. Where the trace lines formed clear peaks, we assigned ranks 2, 3 and 4 based on their relative heights, and recorded the rank corresponding to the base required by each candidate sequence. Absolute heights were not considered because we experimentally determined that peaks corresponding to a known second animal may be very small, inconsistent or even absent (Figure S2B). Where two peaks appeared equal in height, we increased the chromatogram zoom using the Geneious interface until a difference, however small, became clear. If no such difference was apparent, the available ranks were randomly assigned to the two peaks.

When the minority sequence was 1 bp out of phase with the majority reference, minority bases sometimes did not correspond to distinct peaks, but produced either a trailing or a leading 'tail' attached to the next or previous peak. In these cases, there was usually a local maximum or at least an inflection point that we took as the height of the peak. If a base did not correspond to a clear peak, local maximum or inflection point, we estimated the mean relative heights of the trace lines at that site, regardless of shape. Where the trace line for a nucleotide showed absolutely no signal at a site, we treated it as a fourth-ranked peak call for the purpose of analysis. Very rarely, the trace lines for two nucleotides both were flat, with no fluorescence signal. In these cases, the two missing nucleotides were split equally between ranks 3 and 4 for analysis.

At a locus like mtCO1, where all sequences are 'in phase', the majority peak cannot usually be called to represent a minority variant. However, Data S1 illustrates a complication that occurs when majority and minority peaks are not aligned in phase, as at 28S and EPIC25. Sometimes, the expected minority peak at a variable site will happen to match the out-of-phase majority peak. This is only a coincidence, but it prevents an assessment of how high the predicted secondary peak would have been otherwise. Where the "minority" sequence happened to match a shifted majority peak, we took a conservative approach and excluded these first-ranked "minority" calls from contingency table analyses.

Statistical analysis of contingency tables

If minority peaks are a consequence of noise, non-rotifer contaminants, polymerase slippage or other sequencing artifacts, then all else being equal we would predict that the peak ranks corresponding to any given rotifer sequence would not differ significantly from

an equal distribution (i.e., a 1:1:1 ratio for second-, third- and fourth-ranked peaks). We validated this prediction using chromatograms from our "1X" experimental groups (Figure S2). Alternatively, if a second variant is present, the set of minority peaks corresponding to that sequence ought to differ significantly from the null distribution, showing a significant bias in favor of second- rather than third-ranked peaks, and third- rather than fourth-ranked peaks. For each candidate sequence, we tested whether the peak rank distribution differed significantly from the null hypothesis of a 1:1:1 ratio, using Pearson's Chi-square test for count data [84], implemented in R (v. 3.3.1, R Core Team) via the "chisq.test" function. Very small probabilities are printed by R as "p value < 2.2e-16," reflecting accuracy considerations given floating-point constraints. However, a more precise probability is stored, which appears to be accurate across the range of χ^2 values we obtained [85]. To report this value, we invoked the "str" command on the test object.

In many cases, more than one alternative sequence produced a significant deviation from the null distribution. This is because at many variable sites an alternative base is shared by more than one rotifer species; therefore, the true matching sequence distorts the rank distribution of related species that share bases. For example, in Figure S2B, the *A. ricciae* control sequence deviates from the null expectation (χ^2 = 64.9, d.f. = 2, p = 8.07 × 10^{-15}), but only because it shares some variants with the true contaminant: *A.* sp. (AD006). Excluding these shared sites abolishes the fit (χ^2 = 3.2, d.f. = 2, p = 0.202). To distinguish the primary match, we compared the degree of fit not only against the null distribution, but among different candidate species. We typically used the Chi-square test of independence for a 3 x n contingency table, where n is the number of candidate species or sequences (Figure S3). If the initial table included cell counts too small to meet the assumptions of the test, two or more control species or two peak rank classes were pooled (e.g., Figures S3C and S3E). Table 2 indicates the species or sequences that were used for each comparison; distributions that were pooled are indicated with "&." In some cases, further pairwise contrasts are reported. To correct for the problem of multiple comparisons, α was adjusted using the Bonferroni correction, but all hypothesis tests remained significant even using this highly conservative approach.

Chromatogram quality score analysis
Chromatogram data sources and software
Sequencing chromatograms for our experimental samples were returned by Macrogen Europe in the ABIF format, which included Phred quality scores [80]. Chromatogram files generated by Debortoli et al. [18] were provided to us in the .s*cf* file format. We used CodonCode Aligner (v. 7.0.1, CodonCode Corporation) to assign Phred quality scores to these chromatograms, which were imported for further processing in Geneious.
mtCO1 Phred scores
Chromatograms for mtCO1 were uniformly trimmed to 605 bp to avoid sequencing artifacts near the priming sites, and Phred Q20 quality scores are reported for this section. The scores for 36 mtCO1 chromatograms initially provided by Debortoli and colleagues were plotted (Figure S3J), using R with the default setting for whisker length, and annotated manually. This revealed four obvious outliers, and the distributions were assessed to be non-normal (Shapiro-Wilk test: W = 0.75, p = 2.48 × 10^{-6}). The distributions for samples with and without "interspecific recombination" were compared statistically using the Mann-Whitney test, implemented in R through the "wilcox.test" function. The two groups do not have significantly different distributions (Mann-Whitney Test, N = 35, W = 127.5, p = 0.46), which matches the outcome we saw in experimental samples deliberately contaminated with two animals. The equivalent distributions for our newly generated mtCO1 data were less strongly non-normal (Shapiro-Wilk test: W = 0.909, p = 0.007), but the difference between groups was not significant whether analyzed using a parametric (N = 38, t = 1.13, p = 0.26) or nonparametric approach (N = 38, Mann-Whitney W = 234, p = 0.1186).
28S Phred scores
Chromatograms for the focal 28S ribosomal DNA fragment were trimmed to a uniform length of 659 bp (28S0FCT) or 666 bp (28S1RCT) and Phred quality scores were calculated for this informative region. We used the Q40 Phred score because the overall quality of 28S chromatograms was higher than for mtCO1; the whole-genome amplification step may have increased the representation of a single template and reduced the magnitude of the competing signal, especially after further PCR. The two distributions are plotted separately because they were statistically different (Figure S3K). The distribution was significantly lower for samples where HGT was claimed (Mann-Whitney Test: N = 122, W = 373.5, p = 0.015). One datapoint for Sample B14 has a very prominent second sequence, but the two groups still differ significantly even if this is removed from the analysis (Mann-Whitney Test: N = 121, W = 373.5, p = 0.04). Both the median and second-highest quartile of the "HGT" samples fall within the second-lowest quartile of the "non-HGT" samples. The HGT group also had significantly more files with scores below 80% (Fisher's Exact Test, 4 in 12 versus 2 in 110, p = 0.0004; or if we treat paired chromatograms from the same sample as non-independent, 2 in 6 versus 1 in 55, p = 0.012). These analyses highlighted Sample A3B1 as a clear outlier in the "non-HGT" group, with unusually low 28S Q40 scores. ConTAMPR later revealed that A3B1 contained mtCO1 sequences from two species.
Quality analysis at other loci
No attempt was made to assess quality scores for EPIC25 or EPIC63 chromatograms, because even in the case of DNA from a single animal, these amplicons represent a pair of homologous intronic regions that frequently differ by at least one indel and multiple SNPs (Figure 2A), which means two sets of peaks are superimposed in some regions [42], rendering quality scores uninformative.

Interspecific transfer: case-specific methods
Sample B11

Sample B11 was interpreted by Debortoli et al. [18] as a single animal belonging to *A. vaga* Species A, with one mtCO1 haplotype (Hap6 [A]) and one 28S ribosomal sequence (Hap1 [A]). At other nuclear loci it had supposedly acquired DNA horizontally from Species E (EPIC25 Hap35, EPIC63 Hap16 and Nu1054 Hap22). All three of these "transferred" haplotypes also were seen 'natively' in Species E animals sampled during the study. In fact, they all occurred together in a single animal: Individual 81 [E] (Table S3 in [18]). Contamination with DNA from an animal like this would explain the apparently incongruent signal in Sample B11. If so, the mtCO1 chromatograms for Sample B11 ought to show evidence of a second sequence comparable to Hap31 [E], which is found in Individual 81.

We tested this prediction using ConTAMPR (Figures 1 and S2C). The fit of minority peaks to Hap31 [E] was significantly better than the null expectation ($\chi^2 = 127.95$, d.f. = 2, p = 1.64×10^{-28}) or the fit to *A. ricciae* ($\chi^2 = 25.39$, d.f. = 2, p = 3.07×10^{-6}). We attempted to match the minority peaks to multiple control sequences, including the *A. vaga* reference genome and the other species that Debortoli et al. [18] delimited (B, C, D, F). Hap31 [E] remained a significantly better fit (Figure 1, $\chi^2 = 39.2$, d.f. = 12, p = 9.73×10^{-5}). Peak rank distributions for the six control species were statistically indistinguishable from each other ($\chi^2 = 6.66$, d.f. = 10, p = 0.76) and only differed from the null expectation because each happened to share some bases with Hap31 [E]. For example, if we exclude polymorphisms *A. ricciae* shares with Hap31 [E], it no longer differs significantly from an equal ratio of second, third and fourth peaks (30:40:26; $\chi^2 = 3.25$, d.f. = 2, p = 0.197).

Like other animals, single rotifers are not known to harbor two mitochondrial populations that differ at 15% of nucleotide sites (e.g., [23]). A review by Barr et al. [39] concludes that even minor site heteroplasmy is "rare" and "transient" for animal mitochondria: "[t]he apparent scarcity of site heteroplasmy may be linked to the quick return to homoplasmy during the bottlenecks that accompany transmission of mitochondria from parent to offspring… there is a variety of empirical evidence suggesting that the lineage sorting and drift that occur during uniparental mitochondrial transmission will quickly eliminate heteroplasmy." When heteroplasmy is detected in animals, Barr et al. [39] note that it is limited to a handful of sites across the mitogenome: "there are substantial logistical challenges to detecting heteroplasmy when the variant genomes differ by no more than a few base pairs." That divergence is on the order of 0.1%, whereas the different haplotypes detected in the samples of Debortoli et al. [18] are over 10% divergent. The most parsimonious inference is that Sample B11 contained DNA from a second animal belonging to Species E, very similar to Individual 81 [E]. This rotifer supplied the incongruent Species E sequences at the nuclear loci where interspecific genetic exchanges were claimed.

Although Debortoli et al. [18] amplified mtCO1 directly from genomic DNA, they performed PCR for the 28S, EPIC25, EPIC63 and Nu1054 nuclear markers only after whole-genome amplification (WGA) of the samples with an Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare, Amersham Biosciences). When we used PCR to amplify DNA from two animals, the two mtCO1 haplotypes were represented very unequally (Figure S2B). Interposing another nonlinear amplification step introduces further opportunities for bias: some sequences might be dropped entirely, or a majority of amplicons may be generated from a minority template. For these reasons, it was less clear whether nuclear loci would show corresponding evidence of a second animal. However, we attempted to conduct the same analysis, first at the ribosomal 28S marker. The putative second animal was predicted to show a 28S sequence characteristic of Species E. Given that Individual 81 [E] shares all the nuclear sequences that feature incongruently in Sample B11, we aligned the B11 chromatograms to the 28S variant found in that individual: Hap13 [E]. The choice of control sequences to represent each of the other species was random.

As predicted, the fit of minority peaks to Species E was significantly better than the null expectation (Figure S3A), and also superior to Species F ($\chi^2 = 22.45$, d.f. = 2, p = 1.3×10^{-5}), B ($\chi^2 = 12.75$, d.f. = 2, p = 0.0017), C ($\chi^2 = 12.24$, d.f. = 2, p = 0.0022) or *A. ricciae* ($\chi^2 = 11.13$, d.f. = 2, p = 0.0038). These results remain significant even after applying a conservative Bonferroni correction ($\alpha = 0.0083$). Species E was a better fit than Species D and the *A. vaga* reference genome too, but not significantly so, as the three species are nearly identical at this conserved marker. As discussed above, a contaminant will tend to distort the apparent fit of related species owing to shared bases. This effect is illustrated in Figure S3A, where the fit to minority peaks recapitulates a phylogenetic tree based on 28S: species more closely related to E fit better. The tree was constructed using the neighbor-joining method implemented in the Geneious Tree Builder tool, with the default settings and 100 bootstrap replicates. The evidence at 28S therefore reinforces the mtCO1 results, indicating a second animal belonging to Species E. Because Debortoli et al. [18] already reported Species E sequences at EPIC25, EPIC63 and Nu1054, evidence from all five sequenced loci is now brought into congruence, as predicted if Sample B11 contained a second animal belonging to Species E, with a similar genotype to Individual 81.

Although ConTAMPR revealed additional mtCO1 and 28S variants for Sample B11 (Figure 1; Table 2; Figures S2 and S3), the corresponding chromatograms are not obvious outliers in terms of Phred quality scores (Figures S3J and S3K). At 28S, for instance, both chromatograms for Sample B11 lie within the interquartile range for samples where no HGT is claimed. This is consistent with the results of our experiments: even when multiple animals are present, we do not necessarily see obvious differences in chromatogram quality, since minority peaks may be very small, or absent.

If Sample B11 contained two animals, it is interesting that the majority of amplicons for mtCO1 and 28S were from Species A, whereas those at EPIC25, EPIC63 and Nu1054 were from Species E. This may simply reflect the chance outcome of two consecutive nonlinear amplifications (WGA and PCR). However, the guanine-cytosine (GC) content of the Species A molecules expected at EPIC25, EPIC63 and Nu1054 (Table 1) is much higher than the corresponding Species E molecules that Debortoli et al. [18] recovered (+17.7%, +13.6%, +21.2% respectively), whereas GC differences were much lower at mtCO1 and 28S (+1.9%). The whole-genome

amplification kit used by Debortoli et al. [18] has been found to show a bias in favor of templates with lower GC content [86], and multi-template PCR also is sensitive to this parameter [87]. Substantial differences in GC content may have helped to skew amplification of the competing templates, effectively masking Species A even when ConTAMPR was applied. At mtCO1 and 28S, GC was not substantially different, and the identities of both animals were recovered via ConTAMPR, though their representation was far from equal. At EPIC25, EPIC63 and Nu1054, we found no evidence of Species A minority sequences, which would have had substantially elevated GC contents. Owing to effects like this, absence of evidence is not evidence of absence when considering amplicons from potentially contaminated samples. Even if two animals are present, one may be masked by the other at certain loci or in certain amplifications. The hypothesis of contamination cannot be excluded on the basis of apparently clean chromatograms, but it is immediately telling to discover an extra haplotype even at a single locus.

Sample B22

The predicted donor of the incongruent EPIC25 sequence in "Individual 58" belonged to Species E (Table 2). According to Table S3 in Debortoli et al. [18], EPIC25 Hap37 [E] occurs natively in animals that have mtCO1 Hap29 [E] and 28S Hap16 [E]. For ConTAMPR, we therefore aligned these candidates against the B22 mtCO1 and 28S chromatograms, along with control sequences from other species. Again, we found evidence for the predicted alleles (Table 2, Figures S3B and S3C). The extra 28S copy in Sample B22 was sufficiently prominent that the Phred quality scores for these chromatograms fell outside the range of values for samples where no HGT was claimed (Figure S3K).

We examined chromatograms for Sample B22 at the EPIC25 marker itself, where the authors reported a single, incongruent allele: Hap37 [E]. We did not find evidence for an expected 'native' Species C sequence. We suggest this allele was lost either during PCR or WGA, perhaps because its GC content would have been approximately 22% higher than Hap37 [E].

Debortoli et al. [18] characterized "Individual 58" as "homozygous" at EPIC25. The terms "homozygous" and "heterozygous" are debatable for a genome without segregation, but we have followed this usage for simplicity. In Sample B22, more than one Species E variant clearly is present. The trace files show hundreds of double and several triple peaks of comparable heights (Figure S3D). As discussed elsewhere, these represent homologous copies running slightly out of phase following an indel. There are at least 7 SNPs between Hap37 [E] and the other variant in Sample B22, along with indels. Triple peaks indicate more than two alleles, as expected if Sample B22 was contaminated with DNA from two different Species E individuals. Many of the SNPs correspond to standing polymorphisms shared by other animals in the Species E population. The presence of multiple divergent "transferred" alleles in a putative HGT recipient is not consistent with the scenario posited by Debortoli et al. [18], in which interspecific recombination replaced the original DNA at one site, and "gene conversion promptly copied the integrated DNA on its homologous region." That would produce two identical sequences in the recipient. It would not preserve various SNPs and indels found in a heterozygous donor. On the other hand, this pattern is predicted if the variants arose from contaminating DNA from a Species E heterozygote with combinations at EPIC25 similar to those seen in Individuals 81 (Hprim14a/b) and 78 (B39a/b). This is an independent line of evidence for the conclusions supported by ConTAMPR at 28S and mtCO1.

Sample B39

Debortoli et al. [18] interpreted Sample B39 as "Individual 66" [E]. A Species E mtCO1 haplotype had been replaced, in their view, by one imported from Species C. An alternative hypothesis is that Sample B39 was contaminated with mtDNA from Species C, which happened to be amplified by the mtCO1 primers instead of the native sequence. This hypothesis predicts that a Species E haplotype congruent with the other loci will occur among the minority mtCO1 amplicons. Specifically, Table S3 in Debortoli et al. [18] indicates that Individual 66 ought to have mtCO1 Hap31 [E], as seen in Individual 81. Like Individual 66, that animal had Hap16 [E] at 28S, Hap16 [E] at EPIC63, Hap19 [E] at Nu1054 and Hap30 [E] at EPIC25. No other individual had such a combination.

We used ConTAMPR to test whether minority peaks corresponded to Species E rather than any other species, and to Hap31 [E] rather than any other Species E mtCO1 haplotype. We aligned B39 chromatograms not only to other *Adineta* species, but to seven different haplotypes from Species E. We found strong and specific evidence for the 'missing' native mtCO1 haplotype predicted for Individual 66 (Figure S3E). This result brings all five loci into concordance, without the need to invoke interspecific horizontal genetic transfer or transformation of mtDNA. The only incongruence in Sample B39 is the presence of mtCO1 Hap10 [C] among the amplicons. We attribute this to chance amplification of loose Species C mtDNA associated with the surface or gut of Individual 66, or perhaps to a second animal whose nuclear sequences were dropped or outcompeted during WGA or PCR, as seen with samples B11 and B22.

Samples B14, B3B1 and A3B1

For samples B14 and B3B1, the minority peaks in mtCO1 chromatograms were a significantly better match to several other bdelloid mtCO1 haplotypes than the null expectation. This is consistent with the presence of additional mtCO1 sequences, and therefore DNA from additional animals in these samples (Figures S3F and S3G). However, the peak rank distributions for several candidate species could not be distinguished statistically when compared with each other.

Looking at other loci, we noticed that B14 and B3B1 were the only samples to show unambiguous evidence of molecules originating from at least three different species (in each case, A, C and E; Table 2). We guessed that we were unable to identify a single consistent secondary sequence at mtCO1 because three animals from quite different species had all contributed DNA to these samples. This hypothesis was supported by the observation that mtCO1 chromatograms from these two samples were extreme outliers in quality (Figure S3J). The presence of DNA from three animals would explain why additional mtCO1 haplotypes cannot be narrowed down to a single candidate. Relative to the subtle minority peaks produced when we added just one extra animal (Figure S2B), the

noise is such that almost any *Adineta* haplotype could be present. The hypothesis of "interspecific horizontal genetic transfers" supplies no obvious explanation for the anomalous features of the mtCO1 amplicons obtained from these specific samples.

For Sample B14, we found evidence of an additional variant at 28S, consistent with Species E and a better match by a significant margin than more distantly related species (Figure S3H). This suggests that one of the additional animals belonged to Species E, which would explain the incongruent Species E sequences Debortoli et al. [18] reported at EPIC63 (Table 2). Peaks corresponding to this additional variant also explain the unusually low Phred quality scores for the 28S chromatograms for Sample B14 (Figure S3K). These peaks are annotated fully in Data S1. The absolute heights of secondary peaks were often very low, as we saw at mtCO1 when we deliberately added two animals. At 28S, this may reflect preferential nonlinear amplification of one copy during both WGA and PCR. This would explain why some of the other contaminated samples (e.g., B11) did not show obviously anomalous 28S quality scores, or sometimes any detectable second 28S variant (e.g., B3B1).

Conversely, no horizontal exchange was claimed for Sample A3B1 ("Individual 56" [C]), yet it was another clear outlier in Figure S3K, which suggested that DNA from multiple species might be present. The 28S files for Sample A3B1 were of poor quality and could not be used to identify a specific contaminant. We therefore analyzed mtCO1 files instead, using the same panel of candidates as for Sample B22 [C]. Surprisingly, the minority peaks matched the *A. vaga* reference genome significantly better than the null expectation ($\chi^2 = 80.82$, d.f. = 2, p = 2.82×10^{-18}), and better than any of the wild-caught *Adineta* species ($\chi^2 = 29.30$, d.f. = 12, p = 0.00356). This clone (AD008) has been in culture for over 20 years [66] and is used extensively for genetic work in the laboratory where Debortoli et al. [18] prepared DNA [14]. It has not been reported from Belgium, to our knowledge, and was not recorded in any of the 576 natural samples examined by Debortoli et al. [18]. The discovery of its mtDNA mixed with that of wild-caught rotifers provides further evidence that contamination occurred during laboratory processing of animals, DNA or PCR products. Any natural explanation with origins outside the authors' laboratory can be excluded even more decisively in this case. Sample A3B1 was not claimed as a case of incongruence because Species C DNA happened to be in the majority at all loci. This suggests the true prevalence of contamination is likely to extend beyond the six cases of incongruence initially reported.

For Sample B14, we found evidence of an additional EPIC25 sequence consistent with Species E, even though two different EPIC25 copies from Species C had already been reported by Debortoli et al. [18]. The match was significantly better than the null expectation ($\chi^2 = 11.18$, d.f. = 2, p = 0.00374). The presence of a third EPIC25 variant in a single sample is inconsistent with a single animal regardless of its identity, since there are only two copies of EPIC25 in sequenced *Adineta* genomes [14, 17], and Debortoli et al. [18] reported no animal with more than two copies. Only one chromatogram was provided (for the primer EPIC25F), and without the guidance of bidirectional reads for this variable and indel-rich marker, we did not attempt to align and match sequences from other species.

Debortoli et al. [18] reported just one EPIC25 sequence for Sample B3B1, matching Species E. However, we found evidence of at least one and possibly two further EPIC25 variants uniquely matching Species C (Table 2). The ratio of fourth-, third- and second-ranked minority peaks (10:27:105) departed significantly from the null expectation ($\chi^2 = 108.44$, d.f. = 2, p = 2.84×10^{-24}). Given the indel issues discussed above, no other species could be aligned to the minority peaks. The model of interspecific recombination and gene conversion posited by Debortoli et al. [18] has difficulty accommodating two different non-native sequences at one locus, in addition to the more serious obstacles we report elsewhere. The 'native' EPIC25 sequence we would predict for Species A was not recovered; we suggest it was dropped during WGA or PCR. At the noisy mtCO1 locus, Species C again showed the strongest evidence of a fit to minority peaks (Figure S3G, $\chi^2 = 12.9$, d.f. = 2, p = 0.00158), though the fit to Species E also differed significantly from the null expectation ($\chi^2 = 7.35$, d.f. = 2, p = 0.0254). This evidence points to Species C as one of the contaminants. Our interpretation is that Sample B3B1 contained animals or loose DNA belonging to Species A, C and E.

Sample D14
Debortoli et al. [18] interpreted Sample D14 as "Individual 5," and inferred that a Species A EPIC25 sequence had been replaced by Hap10 from Species C. If the incongruent sequence instead came from a second animal or contaminant DNA from Species C, then a native Species A sequence is expected among the amplicons. According to Table S3 in Debortoli et al. [18], this ought to be either Hap1 or Hap4 [A]. In the chromatograms, Hap10 [C] is in the majority, but in both directions a second sequence with peaks of almost equal height runs several base pairs out of phase, as expected if indels are present. It corresponds exactly to Hap4 [A], with a specific number of GAA tandem repeats to distinguish it from Hap1 [A]. Data S1 shows these aligned chromatograms, pointing out peaks matching the second sequence, and highlighting sites where it differs from the Hap 10 [C] interpretation. It was not necessary to attempt to align all the other *Adineta* species to these chromatograms and compare peak heights, as EPIC25 is so variable that no other species would match. Almost all minority base calls for Hap4 [A] use second-ranked peaks (N = 52), and hardly any third- (N = 9) or fourth-ranked peaks (N = 0) are required. The probability of this happening by chance is negligible ($\chi^2 = 75.97$, d.f. = 2, p = 3.19×10^{-17}). The most parsimonious interpretation is that Individual 5 had concordant Species A sequences at all loci in the expected combinations, but its sample tube (D14) was contaminated by DNA from Species C, bearing an EPIC25 copy that was amplified by WGA and PCR along with the native sequence.

Debortoli and colleagues supplied us with two separate bidirectional pairs of EPIC25 chromatograms for Sample D14. Each pair represented an independent PCR amplification from the same DNA sample following WGA (N. Debortoli, pers. comm.). The predicted 'native' Species A sequence was unambiguous from the pair of files described above, and almost equivalent in peak strength to the heterospecific sequence. However, these extra peaks were completely absent from the second pair of chromatograms, which represented amplicons from a second PCR using the same tube of template. Even on close scrutiny of minor secondary peaks, only the incongruent sequence from Species C was visible (Data S1). Since the DNA in the template tube presumably could not have changed

from one PCR to the next, the inconsistency must have arisen from complete loss of an allele during the capricious dynamics of multi-template PCR. If the locus had only been amplified once, there might have been no evidence of the native sequence, and the case for interspecific recombination would have been harder to reject. The lack of minority peaks after a replicated PCR gives important guidance when interpreting the dataset as a whole: even if a sample contains two different sequences, one may be dropped entirely during any given WGA or PCR. A significant pattern of minority peaks is thus sufficient but not necessary to indicate a contaminating sequence, whereas the absence of such a pattern is necessary but not sufficient to exclude it.

This evidence also helps to explain another observation. We did not detect multiple sequences for the nuclear loci EPIC63 or Nu1054 in the seven contaminated samples, as we had done at mtCO1, 28S and EPIC25. In part, this is explained by chance loss of alleles during serial amplifications, as discussed above, but there may be another contributing factor. At these loci, Debortoli et al. [18] report that "amplification and/or sequencing did not work" for a number of samples, whose sequences were therefore missing from their publication (Table 2). One hypothesis is that contaminated samples sometimes did yield the predicted combination of heterospecific sequences at these loci, but these untidy traces (similar to the EPIC25 chromatograms in Data S1) were understandably treated as sequencing failures and either excluded or repeated. As with Sample D14, sometimes a second PCR amplification would yield a single signal, thus removing the signature of contamination. In other cases, if repeated PCRs still did not yield a "clean" sequence (or if the highly divergent templates interfered with the success of amplification itself) then the locus was recorded as missing. As this hypothesis predicts, the incidence of missing sequences at EPIC63 and Nu1054 is significantly higher for the seven samples where interspecific contamination was detected at other loci. Four sequences out of 14 (29%) are missing because "amplification and/or sequencing… failed," compared to 14 missing sequences out of 150 (9%) for the remaining samples (Fisher's Exact Test, p = 0.03). Some of the sequencing failures in this latter "control" group might themselves result from interspecific contamination, and so this comparison is conservative.

Analysis of "intraspecific DNA exchanges"
Probabilities of conspecific contamination
Debortoli et al. [18] report three intraspecific "haplotype trios" at nuclear PCR markers, which they interpret as "signatures of genetic exchange": two in Species A and one in Species C. Each involves three tubes interpreted to contain "heterozygous" individuals that share alleles "in a cyclic fashion… (a||b), (b||c), and (c||a)." The authors claim that this pattern "can only be explained by recombination between individuals," but there are many ways to generate such a pattern when tubes are contaminated with DNA from two conspecific animals. For example, two tubes might each contain one heterozygote: (a||b) and (b||c), while a third contains DNA from two animals homozygous at the focal locus: (c||c) and (a||a). As the authors rightly remark, this requires no intraspecific horizontal exchange, since "the observation of three distinct haplotypes a, b, and c in heterozygous individuals with genotypes (a||b) and (b||c) may be explained by mutations and gene conversions alone." As discussed for "interspecific" samples, contamination may either be evident at multiple loci (presumably when it involves genomic DNA from two animals, as with Sample B11), or only at some loci (presumably via amplification of a contaminating DNA fragment, as with Sample D14).

The number of tubes expected to contain DNA from two conspecific animals can be calculated using two pieces of information. First, we need to know the proportion of tubes containing DNA from two animals of different species. ConTAMPR revealed interspecific contamination in seven tubes: the six identified in Table 2, and A3B1 (Figures S3I and S3K). Some tubes contained DNA from three different animals and thus represent two separate contamination events each, but we conservatively treat these in the same way as tubes with a single contaminant. Second, we need to know the relative frequencies of animals from each of the focal species in the sampled material. These are retrieved from Table S2 in Debortoli et al. [18], based on a suitably large sample (576 animals). All putative inter- and intraspecific genetic exchanges involved Species A, C and E, so we can omit Species B, D and F from consideration here. We conservatively omit from consideration the single case (A3B1) where contaminating DNA matched a laboratory-cultured reference clone, since this was not associated with a claim of genetic exchange and the frequency of such an event is not easily estimated. We therefore take the heterospecific contamination rate to be 6 tubes out of 72, and the relative ratio of animals from Species A, C and E to be 130: 396: 21.

Let t be the probability that DNA from two individual animals was accidentally introduced to any given tube. The expected number of contaminated tubes in the A/C/E dataset is therefore $72t$. Let us assume the two animals in each tube-sharing pair are drawn at random from the population as a whole. This conservative assumption is discussed later, but it enables us to populate a table of probabilities from the species ratios above (Table S2). With this table, we can compute the expected number of conspecific contamination events given that six heterospecific contamination events are observed, as well as t itself. Of all contaminations, the proportion that are conspecific is 0.58206 and t is 0.199; thus, the most likely number of tubes showing conspecific contamination is 8. The binomial distribution can be applied to calculate 95% confidence intervals for this estimate (4 - 15 tubes). The probability that at least three tubes have conspecific cross-contamination is 0.9926.

To estimate the probability that conspecific contamination occurred in one or more tubes assigned to Species C, we used the binomial distribution and Table S2 to calculate the cumulative probability that one or more of the 72 tubes experienced conspecific cross-contamination at all (0.99987). We then calculated the probability that there was exactly one such event, involving two animals belonging to Species A (0.00126 × 0.09728 = 0.00012), or exactly two such events, both involving animals belonging to Species A (0.00126 × 0.09728 × 0.09728 = 0.000012). We deducted these probabilities from the initial value (0.99987 - 0.00012 - 0.000012 = 0.999738). The case of one or two events involving two animals from Species E (p = 7.5 × 10^{-7}) is sufficiently unlikely to be ignored, along with the case of three or more conspecific contamination events exclusively involving Species A or E pairs

(p < 7.4 × 10^{-6}). This leaves a 99.97% probability that at least one sample interpreted as a single Species C rotifer in fact contained DNA from two such animals. The most likely number of cases is 7 (95% C.I. 4-14).

These calculations make the highly conservative assumption that the species identities of any pair of rotifers in a tube are drawn randomly from the sampled population. In fact, Debortoli et al. [18] explicitly falsify this assumption empirically (Table S4 in [18]), by demonstrating far lower odds that a pair of rotifers sampled from the same patch belong to the same molecular species versus different species. Thus, conspecific contamination is even more common than we estimate here. Debortoli et al. [18] calculate an odds ratio of 11 (95% C.I. 8-16; p < 2.2 × 10^{-16}), which would suggest we have dramatically underestimated the rate of conspecific cross-contamination and the total fraction of contaminated tubes (t). This consideration places the reliability of the 2016 dataset even further into doubt. Much of the "heterozygosity" in *Adineta* "individuals" may well arise from two conspecific animals or their DNA sharing a sample tube. Computational phasing of chromatograms is unreliable if two or more animals (and thus four or more alleles) are present in a substantial fraction of tubes.

Triple peaks in Samples H4-28, D23 and D22

The high expected rate of conspecific contamination for Species C suggests a parsimonious explanation for the EPIC25 "allele-sharing" trio (KU861135-40) featured in Figures 3 and 5 in Debortoli et al. [18]. This comprises Samples A112 ("Individual 42"), H4-28 ("Individual 31") and D23 ("Individual 51"). The first link in the "cycle" is EPIC25 Hap6-Hap10, which is seen not only in A112, but in 12 other samples (Table S3 in Debortoli et al. [18]), so that it is very unlikely to be an artifact and need not be investigated further. In contrast, the Hap6-Hap14 genotype from Sample H4-28 is found nowhere else in the study, and therefore is the pairing most likely to be an artifact of DNA from two animals. The final link in the trio is EPIC25 Hap10-Hap 14, which is seen only in Samples D22 and D23. It seems plausible that both of these samples were contaminated, given that at least 7 contaminations are expected in Species C overall and they were prepared consecutively. Conspecific contamination in D23, H4-28 or both would also explain why genomic data for the same Species C trio was reported to show "signatures of inter-individual recombination" in the same cyclic manner at unlinked loci (EPIC63, *hsp82*) even though such concordance is not predicted if DNA fragments at different loci undergo recombination and gene conversion frequently and independently.

ConTAMPR cannot be applied to test for intraspecific cross-contamination in these samples because all the mtCO1 and 28S sequences involved would be either identical or nearly so, and could not be distinguished statistically. At nuclear protein-coding markers, even a single animal has two homologous copies. If two conspecific animals were present, differences between their sequences would simply have been interpreted as differences between alleles; they would have been phased into two copies and interpreted as a single "heterozygous" animal. Contamination could perhaps be detected if the two animals differed at a site with a triallelic SNP. If so, a sample might show three or more different bases at a single site, and this triple-peak in the chromatograms could not be phased without invoking three or more alleles. However, among all the Nu1054, EPIC63 and EPIC25 alleles reported for Species C there is no triallelic SNP; therefore we would not predict any triple peaks on this basis, however high the likelihood of contamination. A few areas of certain chromatograms may appear to have more than two peaks, but these are almost certainly sequencing noise rather than cryptic contaminants showing novel triallelic SNPs. The quality of those traces is consistently poor and the peaks occur all over the files rather than in positions corresponding to known SNPs.

A more powerful test can be applied in one specific situation. When the alleles in a sample differ in length owing to an indel, their corresponding chromatogram reads are displaced by a few bases, either after the indel position (in the forward read), or before it (in the reverse read). This phase-shift creates a set of double peaks across the length of the amplicon, and provides separate information about the longer and the shorter allele. Although the chromatograms look untidy, in fact the predictable displacement in the forward and reverse reads reveals which base variants at SNP sites belong to which allele. Indeed, this is the basis of a computational phasing algorithm [42] that Debortoli et al. [18] employed. When alleles of different lengths are separated in this way, the power to detect that more than two sequences are present is greatly increased. A triallelic SNP is no longer required to generate a triple peak; it only requires two alleles of the same length that differ for a biallelic SNP. In 50% of cases, this will yield a triple peak (in combination with the out-of-phase base from the other allele), revealing that at least three copies must be present. In the other 50% of cases, no triple peak will be evident because the out-of-phase allele will happen to have the same base as one of the two variants, leaving only a double peak. The absence of triple peaks therefore is not especially informative, but the presence of clear triple peaks is telling: a single animal is not predicted to produce a triple peak at a locus with only two copies.

In testing this prediction, the Nu1054 marker is not informative, as it is a section of a single-exon gene with no indels. Even if the Nu1054 chromatograms for Species C had been of high quality (and in fact they are noisy), it would not be informative to look for triple peaks. However, the EPIC63 marker crosses an intron, and alignments for Species C indicate a 2 bp indel near the forward priming site. We looked for cases where a single sample has alleles of different length, whose peaks would therefore be shifted out of phase. The only samples in Species C where this test is possible are precisely those whose haplotypes complete the "allele-sharing" trio: H4-28, D23 and D22. We therefore were able to test the prediction outlined above, using the three samples of greatest interest.

Two triple peaks were found at a known SNP site in the forward reads for all three samples (Figure 2), which indicates that both the long and short alleles in a single sample had variants differing by a C-T substitution. The third peaks were especially pronounced because the chromatograms had very little noise. To quantify this noise, we identified 100 double peaks in the forward chromatograms surrounding the focal SNPs, and manually measured the relative height of the third-ranked peak beneath each one using a transparent plastic ruler. The natural logarithms of these third-peak heights were normally distributed (Figure 2; Shapiro-Wilk test H4-28: W = 0.99, p = 0.49; D23: W = 0.99, p = 0.68; D22: W = 0.98, p = 0.21). Each of the focal peaks was so pronounced that the corresponding probability density under the respective noise distribution was essentially zero. In aggregate, the probability

that six such marked peaks would arise from noise at a known SNP site is negligible (p = 2.05 × 10^{-34}), and it is still less likely that they would each match a known variant.

These triple peaks clearly demonstrate the presence of at least four EPIC63 alleles in Samples H4-28, D22 and D23, as expected if all three samples were contaminated with DNA from at least two different animals. Samples D22 and D23 were prepared consecutively from the same patch of lichen and appear nearly identical; it is likely they contained the same mixture of DNA from two closely related clones, producing the same pattern of contamination. The EPIC25 "allele-sharing" trio can now be explained parsimoniously: two of the three samples contained DNA from two animals. It is unclear whether "EPIC25 Hap14" existed at all or was a phasing artifact, since it was only inferred from samples with at least four alleles. This evidence substantiates the inference from our earlier probability calculations that phasing is unreliable throughout the dataset.

The genomic data and inferences presented by Debortoli et al. [18] for Samples D23 and H4-28 are unreliable for the same reason. Conspecific contamination would have been difficult to exclude empirically even using high-quality genome data, much less the "extremely fragmented" partial assemblies that Debortoli et al. [18] report. It is perhaps unsurprising that "the posterior probabilities of our haplotypes reconstructed using PHASE were rather low" given that a substantial fraction of both the genomic and population reference samples would have contained DNA from multiple animals, and therefore more than two homologous sequences. The authors write that they "did not detect any additional copy of the EPIC25, EPIC63, and Nu1054 markers, nor any additional 28S rDNA sequence." This would not have been surprising even given conspecific contamination, for the reasons discussed above. In fact, however, the authors did detect four different copies of the 300 bp EPIC63 marker in Sample D23, and deposited the phased sequences in GenBank (KU860980.1; KU860981.1; KU861143.1; KU861144.1). These are the same four EPIC63 alleles we inferred from our own triple peak analysis (Hap9, Hap4, an un-named copy and Hap7, respectively). Debortoli et al. [18] therefore supply further evidence for conspecific contamination, since a single animal ("Individual 51") would only have two copies.

Haplotype trios: mutation and gene conversion

Debortoli et al. [18] discuss two "haplotype trios" involving Species A, at the Nu1054 and EPIC25 markers. These are interpreted as evidence of "intraspecific horizontal exchange" that "can only be explained by recombination between individuals." For these two trios, we need not even reach the hypothesis of conspecific contamination. The authors rightly note that patterns of allele sharing involving closely related clones and sequences "may be explained by mutations and gene conversions alone," but they do not explicitly or quantitatively test this hypothesis for either of the trios discussed before reaching the conclusion that recombination must have occurred. To supply the missing test, we obtained the relevant sequences of the samples comprising each trio from GenBank (Figure S4). The identities of the samples are not clear from the haplotype networks the authors present in their Figure 3, but can be inferred by examining codes in Table S3 in Debortoli et al. [18] and the sequences themselves. For the purpose of this analysis, we assume Debortoli et al. [18] phased the homologs correctly, setting aside the concerns about phasing described above.

To visualize each trio explicitly, we aligned the six sequences and identified every polymorphic site differentiating the three variants (Figure S4). For the EPIC25 trio, Debortoli et al. [18] state that "flanking regions were sequenced… yielding five additional SNPs between haplotypes." This seems to be an error, as examination of the flanking sequences reveals only two additional C/T SNPs, giving a total of 3 SNPs and one tandem repeat polymorphism. From each alignment, we extracted and concatenated the polymorphic sites defining each variant, and matched them to the samples comprising the trio. This gives a pair of sequences for each sample, interpreted as a single animal with two homologous gene copies. These were plotted to identify ways in which each trio "may be explained by mutations and gene conversions alone." We began with two copies from a single animal, and attempted to identify the most parsimonious pathway to generate the other two clones by substitution.

For the Nu1054 trio, the simplest path identified was a single point mutation in one descendant of the initial clone, and a single gene conversion event in a different descendant, with a tract length of 144 bp (Figure S4). Evidence of both substitution events is seen commonly in other individuals from the population sample. For the EPIC25 trio, the simplest hypothesis is a single point mutation in one descendant of the focal clone, and a single gene conversion in another, with a tract length of 274 bp (Figure S4). These hypotheses are highly parsimonious; in Species A, all but one of the Nu1054 and EPIC25 variants (93%) are separated from a second allele by a single substitution, indicating that point mutations are common and frequently persist. As Debortoli et al. [18] remark, gene conversion occurs even more commonly in *A. vaga*: "namely, 25 times more often than point mutation." The lengths of the two hypothetical gene conversion events are well within the mean tract length determined in model systems, including *Plasmodium*, yeast, humans and *Drosophila* [88–91]. Therefore, both trios are "explained by mutations and gene conversions alone" [18], without reference to "intraspecific horizontal exchange" or even to conspecific cross-contamination.

Future attempts to indirectly infer recombination among presumed asexuals ought to take explicit and quantitative account of alternative hypotheses for apparent allele sharing —including cross-contamination; multiple gene copies; phasing uncertainty; sequencing error; convergent substitution (via mutation, gene conversion or selection)— and the probability of falsely (or correctly) detecting haplotype trios or other patterns post hoc, given the numbers of isolates or genotypes examined in an initial sample pool. One way to exclude alternative hypotheses would be a direct demonstration of recombination or sexual crossing between individuals or clones in the laboratory.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was performed using R (v. 3.3.1, R Core Team) and all statistical details of experiments and analyses can be found in the Method Details, including formal tests for normality where appropriate and use of nonparametric alternatives. Significance was

defined using $\alpha = 0.05$, where α is the probability of rejecting the null hypothesis when it is true. As explained in Method Details, the probabilities generated via ConTAMPR and shown in Table 2 correspond to tests of two distinct null hypotheses. The first hypothesis is that an observed set of minority peak ranks is drawn from an equal distribution of second, third and fourth peaks, as expected given random noise. The second hypothesis is that a panel of candidate haplotypes share the same proportion of second-, third- and fourth-ranked minority peaks and therefore fit a pair of chromatograms equally well.

DATA AND SOFTWARE AVAILABILITY

Chromatogram data newly generated in this study are available to download from: https://doi.org/10.17632/fsjpwrnv4t.1. Chromatogram data generated by Debortoli et al. [18] and re-analyzed here are available to download from: https://github.com/jflot/Debortoli2016CurrentBiology. Scripts used to analyze interspecific microhomology with reference to the *Adineta vaga* genome are available at: https://github.com/reubwn/microhomology. Sequences delimited by Debortoli et al. [18] have been deposited in GenBank under ID codes KU860573–KU861170. Relevant sequences for laboratory-cultured *Adineta* clones have been deposited in GenBank under ID codes GQ398061 and JX184001 (*A. vaga* reference clone); EF173187 and KM043216 (*A. ricciae*); and KM043183 (*A.* sp. AD006). Reference genome data for *A. vaga* [14] have been deposited in GenBank under ID code GCA_000513175.1. RNA-Seq data have been deposited in the Sequence Read Archive under ID code ERR260376.

ADDITIONAL RESOURCES

A version of this paper was made available via the pre-print server bioRxiv.org on June 20th 2017 and revised on September 12th 2017 (https://doi.org/10.1101/150490). A protocol for isolating bdelloid rotifers with needles is mentioned in the Method Details and can be downloaded at: https://www.biorxiv.org/highwire/filestream/45474/field_highwire_adjunct_files/1/150490-2.pdf.