

RESEARCH

Open Access



Distribution of the DNA transposon family, *Pokey* in the *Daphnia pulex* species complex

Shannon H. C. Eagle and Teresa J. Crease*

Abstract

Background: The *Pokey* family of DNA transposons consists of two putatively autonomous groups, *PokeyA* and *PokeyB*, and two groups of Miniature Inverted-repeat Transposable Elements (MITEs), *mPok1* and *mPok2*. This TE family is unusual as it inserts into a specific site in ribosomal (r)DNA, as well as other locations in *Daphnia* genomes. The goals of this study were to determine the distribution of the *Pokey* family in lineages of the *Daphnia pulex* species complex, and to test the hypothesis that unusually high *PokeyA* number in some isolates of *Daphnia pulex* is the result of recent transposition. To do this, we estimated the haploid number of *Pokey*, *mPok*, and rRNA genes in 45 isolates from five *Daphnia* lineages using quantitative PCR. We also cloned and sequenced partial copies of *PokeyA* from four isolates of *D. pulex*.

Results: Haploid *PokeyA* and *PokeyB* number is generally less than 20 and tends to be higher outside rDNA in four lineages. Conversely, the number of both groups is much higher outside rDNA (~120) in *D. pulex*, and *PokeyB* is also somewhat higher inside rDNA. *mPok1* was only detected in *D. pulex*. *mPok2* occurs both outside (~30) and inside rDNA (~6) in *D. pulex*, but was rare (≤ 2) outside rDNA in the other four lineages. There is no correlation between *Pokey* and rRNA gene number (mean = 240 across lineages) in any lineage. Variation among cloned partial *PokeyA* sequences is significantly higher in isolates with high number compared to isolates with an average number.

Conclusions: The high *Pokey* number outside rDNA in *D. pulex* and inside rDNA in some *D. pulex* isolates is consistent with a recent increase in transposition rate. The *D. pulex* increase may have been triggered by insertion of *PokeyA* into a region of transcriptionally active rDNA. The expansion in *D. pulex* (thought to be of hybrid origin) may be a consequence of release from epigenetic repression following hybridization. Previous work found *D. obtusa* to be very different from the *D. pulex* complex; mean *PokeyA* is higher in rDNA (~75), rDNA array size is nearly twice as large (415), and the two are positively correlated. The predominance of *Pokey* in only one location could be explained by purifying selection against ectopic recombination between elements inside and outside rDNA.

Keywords: *Daphnia*, Transposon, *Pokey*, Ribosomal DNA, MITE

Background

Transposable elements (TEs) are segments of DNA that can move around the genome. TEs are often detrimental to the host because they can interrupt gene function, their transposition has energy costs (e.g. the repair of double strand breaks), they can cause ectopic recombination [1], and the epigenetic mechanisms used to control them can shut down surrounding genes [2]. Usually, DNA transposons must undergo horizontal transfer to remain active

[3]. However, *Pokey*, a class II cut-and-paste DNA transposon in the *piggyBac* superfamily [4] has been vertically inherited in the subgenus, *Daphnia* [5]. *Pokey* inserts at TTAA sites and creates a target site duplication (TSD). It ranges in size from 4.5 to 10 kilobase pairs (kb); the length variation is due to repeat sequences at the 5' end derived from the ribosomal intergenic spacer (IGS) [4] and/or the ribosomal internal transcribed spacer (ITS) [6]. *Pokey* is the only DNA transposon known to insert in a specific location in ribosomal DNA (rDNA); other rDNA elements, such as R1 and R2, are non-Long Terminal Repeat (LTR) retrotransposons [7]. Nevertheless, Kojima and Jurka [8] recently reported the discovery of class II *Dada* TEs in

* Correspondence: tcrease@uoguelph.ca
Department of Integrative Biology, University of Guelph, Guelph, ON N1G 2W1, Canada

organisms as taxonomically diverse as fish, molluscs, annelids, and cladoceran crustaceans (*Daphnia*). Like *Pokey*, they show target-site specificity for particular multicopy RNA genes (small nuclear RNA, transfer RNA), encode a putatively cut-and-paste transposase, and create TSDs on insertion. However, unlike other elements that target multicopy RNA genes, *Pokey* also inserts into TTAA sites outside rDNA in the genomes of North American (NA) *D. pulex* [9] and NA *D. pulicaria* [10].

In eukaryotes, rDNA is a multigene family arranged in tandem arrays of units each consisting of the 18S, 5.8S, and 28S rRNA genes plus the ITS, the external transcribed spacer (ETS), and the IGS. The number of rDNA units per eukaryotic genome varies from less than 50 to over 25,000 [11] and is usually much higher than the number required for viability [7]. rDNA typically displays the phenomenon of concerted evolution, which means that intraspecific sequence divergence between rDNA copies is generally much lower than interspecific divergence [7]. The mechanisms of concerted evolution are thought to be gene conversion and unequal crossing over [7], which can cause the number of rDNA units to fluctuate. Due to its multicopy nature and the presence of copies above the number required for viability, some TEs, such as R1, R2, and *Pokey*, have persisted in rDNA for very long periods of time [7] even though the homogenizing mechanisms of concerted evolution are expected to continually remove them. In addition, TEs inserted in rRNA genes render the genes non-functional, so selection is expected to favor loss of these elements [7].

Daphnia are freshwater cladoceran crustaceans with a worldwide distribution [12]. Based on molecular analyses, the *D. pulex* species complex (subgenus *Daphnia*) is composed of eight lineages between which hybridization frequently occurs: North American (NA) *D. pulex*, *D. melanica*, *D. middendorffiana*, North American (NA) *D. pulicaria*, *D. tenebrosa*, European (EU) *D. pulex*, European (EU) *D. pulicaria* [13], and *D. arenata* [14]. Although these analyses suggest that the North American and European populations of *D. pulex* and *D. pulicaria* should be classified as separate species, they have not been officially described as such [13]. While most cladocerans reproduce by cyclical parthenogenesis (diapausing eggs are produced sexually and direct-developing eggs are produced apomictically), populations of obligate parthenogens (diapausing eggs are also produced apomictically) also occur in some lineages of this species complex [13].

Eagle and Crease [10] found that the total haploid number of *Pokey* in NA *D. pulex* and NA *D. pulicaria* is usually less than 20 copies, most of which are found outside rDNA. In addition, *Pokey* number in rDNA is not correlated with rRNA gene number. In contrast, LeRiche et al. [15] found *Pokey* number in another species in the same subgenus, *Daphnia obtusa*, to be as high as 154 (mean = 75), almost

all of which are located inside rDNA. Moreover, *Pokey* number in rDNA is strongly correlated with rRNA gene number.

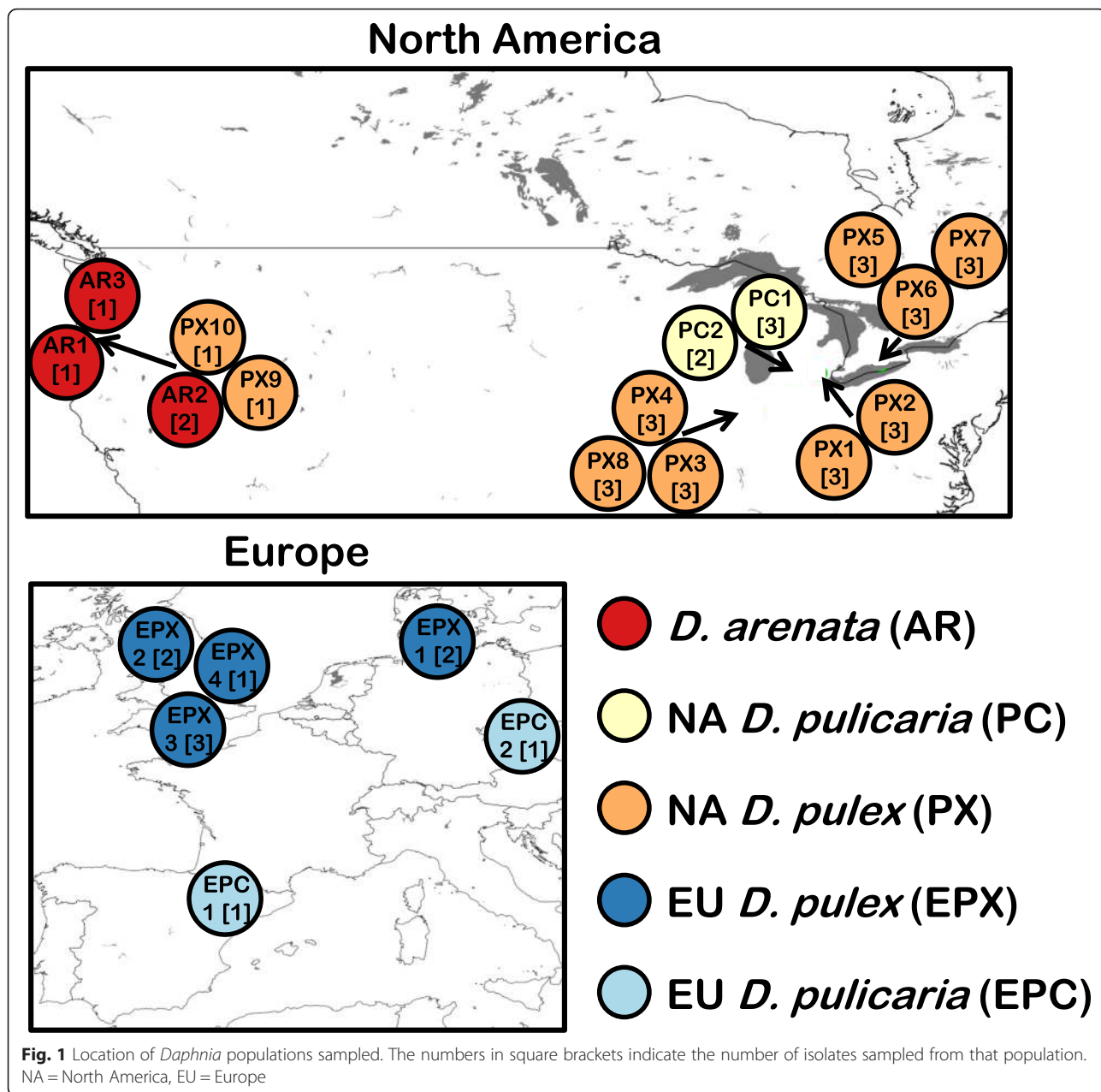
Both autonomous (encoding a functional transposase) and non-autonomous (the transposase gene is degraded or partially deleted) copies of *Pokey* have been found [6]. Further, Miniature Inverted Repeat Transposable Elements (MITEs), which are very short (usually less than 600 nucleotides) [16], non-autonomous TEs have been observed in *D. arenata* [6]. Overall, Elliott et al. [6] discovered four TEs in the *D. arenata Pokey* family: the original *Pokey* group discovered in *D. pulex*, now called *PokeyA*; a variant containing a transposase gene (complete or partial) that was initially seen in *D. obtusa* [5], now called *PokeyB*; a MITE with terminal inverted repeats (TIRs) similar to those of *PokeyA*, called *mPok1*, and a MITE with TIRs similar to those of *PokeyB*, called *mPok2*. The *mPok* elements average 760 base pairs (bp) in length [6]. *PokeyA* and *mPok1* have 16 bp imperfect TIRs, and *PokeyB* and *mPok2* have 12 bp imperfect TIRs [6]. Like *PokeyA*, *PokeyB* also contains an open reading frame (ORF) encoding a putative transposase and both ORFs contain an intron [6]. Mean sequence divergence between copies of *PokeyA* and *PokeyB* containing a complete or partial transposase gene from *D. arenata* is 40 %, suggesting that the two TE groups did not diverge from one another recently.

Three additional types of TEs have now been discovered in the *D. arenata Pokey* family, but we do not know if they occur in other lineages in the *D. pulex* complex. Our first goal was to determine the distribution of *Pokey* and *mPok* in the *D. pulex* complex. We did this using quantitative polymerase chain reaction (qPCR) to estimate gene number in 45 cyclically parthenogenetic isolates representing five lineages: NA *D. pulex* ($N = 26$), NA *D. pulicaria* ($N = 5$), *D. arenata* ($N = 4$), EU *D. pulex* ($N = 8$), and EU *D. pulicaria* ($N = 2$) (Fig. 1, Additional file 1: Table S1). Our second goal was to test the hypothesis of recent transposition of *PokeyA* in NA *D. pulicaria*, suggested by Eagle and Crease [10], by estimating divergence among partial *PokeyA* sequences from four isolates; two with unusually high *PokeyA* number and two with numbers close to the average for this species.

Results

Pokey and *mPok* number in the *D. pulex* complex

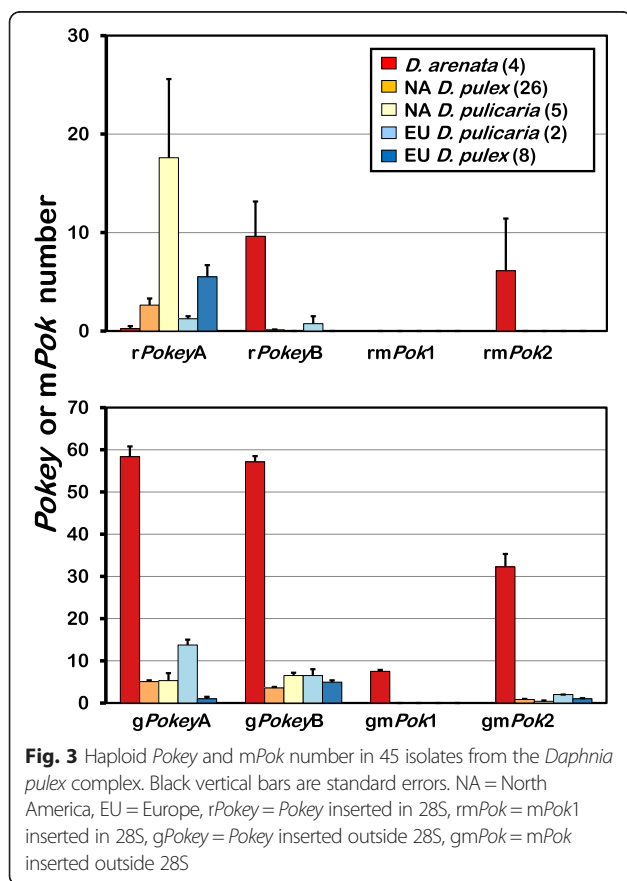
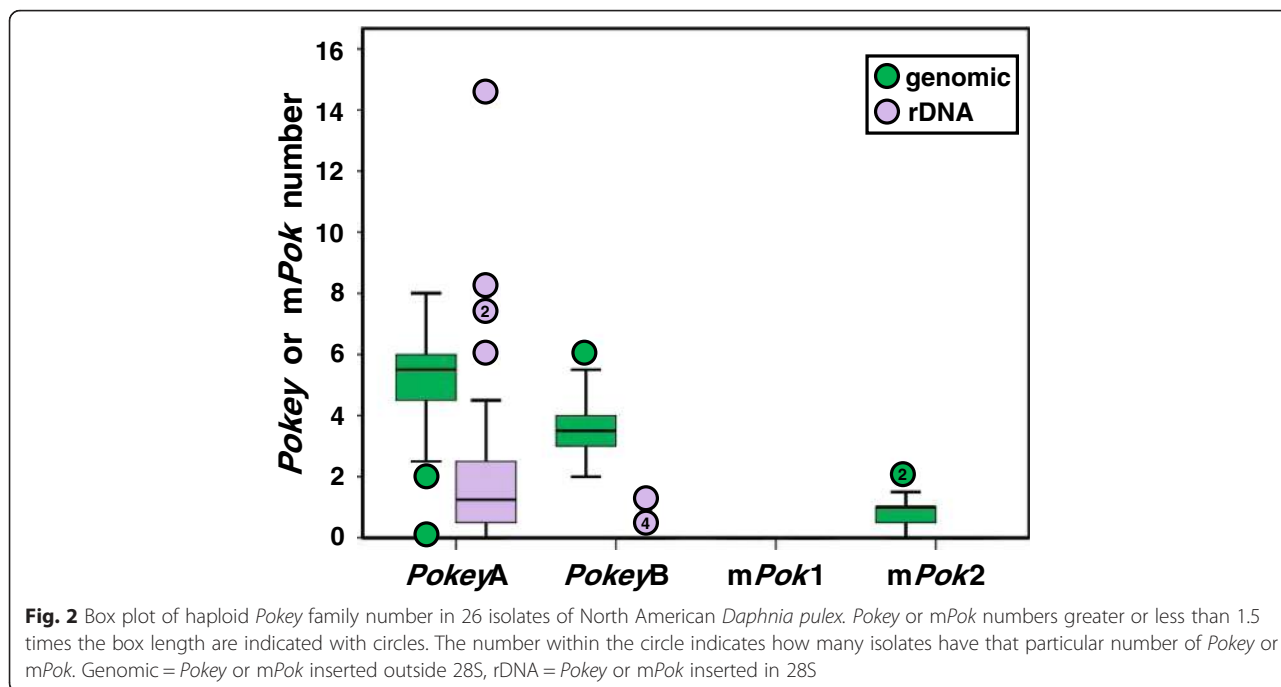
We used qPCR (Additional file 1: Table S2) to estimate the haploid number of the four TEs in the *Pokey* family (*PokeyA*, *PokeyB*, *mPok1*, *mPok2*) in the entire genome (*tPokey*, *tmPok*) and in rDNA (*rPokey*, *rmPok*) by comparing their PCR amplification rate to the PCR amplification rate of two single copy genes, *Gtp* (a member of the RAB subfamily of small GTPases) and *Tif* (a transcription initiation factor) [10], [17]. We calculated the number of *Pokey* and *mPok* outside rDNA (*gPokey*, *gmPok*) by subtracting



rPokey or *rmPok* from *tPokey* or *tmPok*. We also estimated the number of 18S and 28S rRNA genes in each isolate.

The mean haploid number of all *PokeyA* and *PokeyB* in NA *D. pulex* is 11.4 and never exceeds 19 copies per isolate (Additional file 1: Table S3). Mean *gPokeyA* plus *gPokeyB* number (8.6) exceeds mean *rPokeyA* plus *rPokeyB* number (2.8) and this is true for both *PokeyA* (5.3 vs 2.8) and *PokeyB* (3.6 vs 0.1, Fig. 2), individually. *PokeyA* number is higher than *PokeyB* number in NA *D. pulex*, inside and outside rDNA (Fig. 2). These patterns are similar in each of the 26 NA *D. pulex* isolates (Additional file 1: Table S3).

The mean total number of *PokeyA* plus *PokeyB* per haploid genome in three of the other four lineages (NA *D. pulicaria*, EU *D. pulex*, EU *D. pulicaria*) is similar to NA *D. pulex*, with means ranging from 11.4 to 29.4 (Fig. 3). In contrast, the mean total number in *D. arenata* is substantially higher (125.4) (Additional file 1: Table S3), and the mean number of *gPokeyA* plus *gPokeyB* is highest in *D. arenata* (Table 1, Fig. 3). Although the mean is lower compared to *D. arenata*, we still detected *gPokeyB* in all of the other 41 isolates, and *gPokeyA* in all but six of them (one NA *D. pulex*, one NA *D. pulicaria* and four EU *D. pulex*; Additional file 1: Table S3). The *gPokeyA* estimate in these



six isolates was set to zero because the estimate of *rPokeyA* is larger than the estimate of *tPokeyA* giving a negative value for *gPokeyA*. Thus, we cannot exclude the possibility that very low numbers of *gPokeyA* occur in these isolates.

The mean number of *rPokeyA* is lowest in *D. arenata* (Table 1, Fig. 3) and highest in NA *D. pulicaria*. The high mean in NA *D. pulicaria* is due to the unusually high number in two isolates (PC1.2, PC2.1), as previously described by Eagle and Crease [10]. These isolates were chosen to determine whether the high number of *rPokeyA* could be explained by recent transposition (see below). *rPokeyA* is absent in three of the four *D. arenata* isolates, but is only absent in three of the other 41 isolates (Additional file 1: Table S3). This pattern is reversed for *rPokeyB* in which the mean number is highest in *D. arenata* (Table 1, Fig. 3), but the element is absent from rDNA in 35 of the other 41 isolates (Additional file 1: Table S3).

The MITES, *mPok1* and *mPok2* are primarily found in *D. arenata* (Fig. 3). Indeed, the only MITE we detected in the other four lineages is *gmPok2*, but it is present in no more than 2 copies per haploid genome (Table 1), and absent in most isolates (Additional file 1: Table S3). *mPok2* was detected both inside (mean = 6) and outside (mean = 32) rDNA in *D. arenata* (Table 1, Fig. 3), but *mPok1* was only detected outside rDNA (mean = 7.5) (Table 1, Fig. 3).

rRNA gene number in the *D. pulex* complex

Haploid 18S number ranges from 98 to 460 and 28S number ranges from 88 to 577 in the 45 *Daphnia*

Table 1 Haploid number of *Pokey*, *mPok*, and rRNA genes in isolates from the *Daphnia pulex* complex

| | | NA ^c <i>D. pulex</i> | NA ^c <i>D. pulicaria</i> | <i>D. arenata</i> | EU ^c <i>D. pulex</i> | EU ^c <i>D. pulicaria</i> |
|-------------------|----------------------|---------------------------------|-------------------------------------|-------------------|---------------------------------|-------------------------------------|
| Gene ^a | N ^b | 26 | 5 | 4 | 8 | 2 |
| 18S | Mean | 216.3 | 310.4 | 303.8 | 194.5 | 366.3 |
| | Std Dev ^b | 85.3 | 112.7 | 38.3 | 93.2 | 35.7 |
| | Range | 112.5 to 440.5 | 176.0 to 460.0 | 248.0 to 332.5 | 97.5 to 339.5 | 341.0 to 391.5 |
| 28S | Mean | 230.9 | 405.9 | 309.3 | 201.5 | 347.8 |
| | Std Dev | 81.1 | 128.3 | 30.3 | 124.3 | 9.5 |
| | Range | 119.5 to 452.5 | 234.5 to 576.5 | 264.5 to 330.5 | 87.5 to 418.0 | 341.0 to 354.5 |
| <i>rPokeyA</i> | Mean | 2.6 | 17.6 | 0.3 | 5.5 | 1.3 |
| | Std Dev | 3.4 | 17.9 | 0.5 | 3.4 | 0.4 |
| | Range | 0 to 14.5 | 3.5 to 48.0 | 0 to 1.0 | 1.0 to 10.5 | 1.0 to 1.5 |
| <i>rPokeyB</i> | Mean | 0.1 | 0 | 9.6 | 0 | 0.8 |
| | Std Dev | 0.3 | | 7.0 | | 1.1 |
| | Range | 0 to 1.0 | | 5.0 to 20.0 | | 0 to 1.5 |
| <i>rmPok1</i> | Mean | 0 | 0 | 0 | 0 | 0 |
| | Std Dev | | | | | |
| | Range | | | | | |
| <i>rmPok2</i> | Mean | 0 | 0 | 6.1 | 0 | 0 |
| | Std Dev | | | 10.6 | | |
| | Range | | | 0 to 22.0 | | |
| <i>gPokeyA</i> | Mean | 5.1 | 5.3 | 58.4 | 1.0 | 13.8 |
| | Std Dev | 1.7 | 4.0 | 4.9 | 1.3 | 1.8 |
| | Range | 0.0 to 8.0 | 0.0 to 10.5 | 52.5 to 63.5 | 0.0 to 3.0 | 12.5 to 15.0 |
| <i>gPokeyB</i> | Mean | 3.6 | 6.5 | 57.1 | 4.9 | 6.5 |
| | Std Dev | 1.1 | 1.5 | 2.8 | 1.2 | 2.1 |
| | Range | 2.0 to 6.0 | 5.0 to 9.0 | 54.5 to 61.0 | 3.5 to 7.0 | 5.0 to 8.0 |
| <i>gmPok1</i> | Mean | 0 | 0 | 7.5 | 0 | 0 |
| | Std Dev | | | 0.7 | | |
| | Range | | | 7.0 to 8.5 | | |
| <i>gmPok2</i> | Mean | 0.8 | 0.4 | 32.3 | 1.0 | 2.0 |
| | Std Dev | 0.5 | 0.4 | 6.1 | 0.4 | 0.0 |
| | Range | 0 to 2.0 | 0 to 1.0 | 25.0 to 38.5 | 0.5 to 1.5 | 2.0 to 2.0 |

^a18S = 18S rRNA genes, 28S = 28S rRNA genes, *rPokey* = *Pokey* inserted in 28S, *gPokey* = *Pokey* inserted outside 28S, *rmPok* = *mPok* inserted in 28S, *gmPok* = *mPok* inserted outside 28S

^bN = number of isolates. Std Dev = Standard Deviation

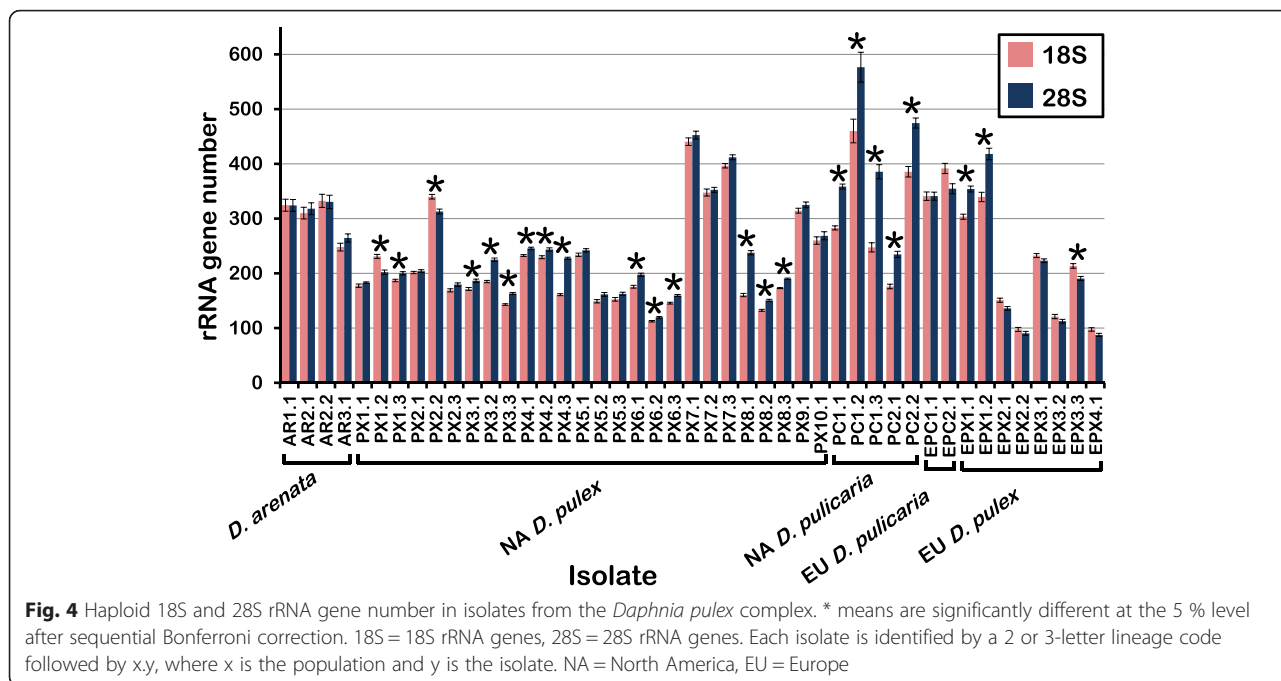
^cNA = North American, EU = European

isolates (Table 1). 28S number is larger than 18S number by 2.5 to 138 copies in 33 of the 45 isolates, and the difference is significant in 23 of these (Fig. 4, Additional file 1: Table S3). Despite the differences between the two genes, their numbers are significantly correlated in NA *D. pulex* with a slope of 0.92 and an R^2 of 0.937 ($p < 10^{-15}$, Table 2, Additional file 2: Figure S1). The slope of the correlation based on the other 19 isolates is 1.19 with an R^2 of 0.873 ($p = < 10^{-9}$, Table 2, Additional file 2: Figure S1). There is no correlation between 28S and total *rPokey* plus *rmPok* number in the 26 NA *D. pulex*

isolates ($R^2 = 0.04$, Table 2, Additional file 2: Figure S2). Similarly, there is no correlation between 28S and *rPokey* plus *rmPok* number in the 19 isolates from the other four lineages ($R^2 = 0.03$, Table 2, Additional file 2: Figure S2).

Variation in partial *PokeyA* sequences in North American *D. pulicaria*

An approximately 1500 bp segment, including 519 bp of the 5' end of the *PokeyA* transposase and the length-variable region upstream was PCR-amplified, cloned, and



sequenced from four isolates of NA *D. pulicaria*. We predicted that sequence divergence would be lower among copies from isolates with high *PokeyA* number if the increase was a consequence of recent transposition. Using data from Eagle and Crease [10], we chose two isolates, one with high and one with average *rPokeyA* number, from each of two populations. The number of other *Pokey* family elements is similar in all four isolates (Table 3). The number of 28S genes in these isolates ranges from 235 to 577 and the one with the lowest number (PC2.1) has the highest number of *rPokeyA* (48, Table 3).

On average, the mean number of differences between each cloned sequence and the isolate consensus sequence is lower in the isolate with a high *PokeyA* number compared to the isolate with an average number from the same population. Moreover, the difference between means is significant in both populations (Table 3). The Neighbor-joining dendrogram of these sequences (Fig. 5) shows two major clusters; cluster 1 contains 11 of 12 sequences from isolate High 1 (PC1.2), as well as sequences from both

Average isolates. Cluster 2 contains all 12 sequences from isolate High 2 (PC2.1), as well as sequences from both Average isolates. All but two of the sequences encode a putatively functional transposase (no insertions, deletions, or stop codons). The exceptions are High 1–19, which contains several stop codons and numerous nucleotide substitutions throughout, and Average 2–32, in which the first 29 bp of the coding sequence are deleted (gaps are not considered in calculations of sequence divergence so this sequence does not appear to be divergent from other sequences in cluster 2).

Discussion

Pokey* family distribution in NA and EU *D. pulex* and *D. pulicaria

In general, *Pokey* number is similar in four of the five lineages we examined; NA and EU *D. pulex* and NA and EU *D. pulicaria*. The mean number of *PokeyA* plus *PokeyB* is less than 20 and both TEs are primarily located outside of rDNA. Our *PokeyA* numbers (Table 1) are similar to

Table 2 Correlation between haploid *rPokey* family and rRNA gene number in isolates from five lineages of the *Daphnia pulex* complex

| lineage | X-axis ^c | Y-axis ^c | slope | y-intercept | R ² | p-value* | Figure |
|---------------------------------|---------------------|----------------------|--------|-------------|----------------|-------------------------------|------------------------------|
| NA <i>D. pulex</i> ^a | 18S | 28S | 0.920 | 31.88 | 0.937 | 6.3 x 10⁻¹⁶ | Additional file 2: Figure S1 |
| four lineages ^b | 18S | 28S | 1.19 | 23.39 | 0.873 | 4.9 x 10⁻⁹ | Additional file 2: Figure S1 |
| NA <i>D. pulex</i> | 28S | <i>rPokey</i> family | -0.008 | 4.60 | 0.035 | 0.361 | Additional file 2: Figure S2 |
| four lineages | 28S | <i>rPokey</i> family | 0.014 | 6.44 | 0.026 | 0.51 | Additional file 2: Figure S2 |

^aThis analysis is based on 26 isolates

^bThis analysis is based on 19 isolates including *D. arenata* (4), EU *D. pulex* (8), and NA (5) and EU (2) *D. pulicaria*

^c18S = 18S rRNA genes, 28S = 28S rRNA genes, *rPokey* family = *Pokey* plus *mPok* inserted in 28S = [*rPokeyA* + *rPokeyB* + *rmPok1* + *rmPok2*]

*p-values < 0.05 are indicated in bold font

Table 3 Haploid number of *Pokey*, *mPok*, and rRNA genes in four isolates of North American *Daphnia pulicaria*

| gene ^a | PC1.1 Average 1 ^b | PC1.2 High 1 | PC2.1 High 2 | PC2.2 Average 2 |
|--|---------------------------------|-----------------------|-----------------|--------------------|
| 18S | 283 | 460 | 176 | 385.5 |
| 28S | 358.5 | 576.5 | 234.5 | 474.5 |
| <i>rPokeyA</i> | 6 | 17 | 48 | 3.5 |
| <i>gPokeyA</i> | 3.5 | 7.5 | 0 | 5 |
| <i>gPokeyB</i> | 5 | 6.5 | 9 | 6 |
| <i>gmPok2</i> | 1 | 0 | 0.5 | 0.5 |
| mean number of differences from consensus ^c | 49.9 | 10.8 | 1.6 | 45.8 |
| standard deviation | 43.3 | 36.9 | 3.4 | 32.5 |
| range | 0 to 97 | 0 to 128 ^d | 0 to 6 | 12 to 78 |
| <i>p</i> -value ^e | | 0.027 | | 0.010 |

^a18S = 18S rRNA genes. 28S = 28S rRNA genes. *rPokeyB*, *rmPok2* and *mPok1* were not detected in these four isolates

^bAverage and High correspond to the sequence labels in Fig. 4

^cThe number of differences between the consensus and each of 12 cloned 1500 bp sequences was determined for each isolate. Gaps were included in the analysis. Consecutive gaps were counted as one variant nucleotide position

^dAll other values were 0 or 1 for this isolate

^eANOVA was used to compare the mean number of differences across all four isolates. This test was significant ($F = 6.62$, $d.f. = 3$, $p = 0.0009$). A *Post hoc* Tukey's HSD test was used to determine if means in the Average and High isolates from the same population were significantly different. They *p*-values refer to the Tukey's tests

estimates from Eagle and Crease [10] for NA *D. pulex* (mean *rPokeyA* = 2.1 and mean *gPokeyA* = 9.6 based on 43 isolates) and NA *D. pulicaria* (mean *rPokeyA* = 6.6 and mean *gPokeyA* = 9.5 based on 26 isolates). However, the NA *D. pulicaria* *rPokeyA* mean in the current study (17.6) is higher than the previous estimate because we only included five isolates, two of which have unusually high *rPokeyA* number, and were chosen for that reason.

Both the current and previous estimates of *gPokeyA* and *gPokeyB* are within the range of other DNA transposons found in the *Daphnia* genome sequence [18]. For example, there are approximately 31 copies per *Tc1/mariner* family (217 copies from 7 families) and approximately 6 copies per *hAT* family (33 copies from 6 families). In contrast, the mean numbers of *rPokeyA* and *rPokeyB* (Table 1) tend to be low compared to the mean number of rDNA-specific non-LTR retroelements, R1 (34) and R2 (12) in *D. melanogaster* [19].

Overall, *PokeyA* outnumbers *PokeyB*, which could be due to a higher transposition rate, weaker purifying selection at the level of the host, and/or lower rates of deletion of *PokeyA*. Due to the fact that they both insert into the same TTAA site in rDNA, it seems unlikely that the strength of purifying selection against these two TEs at the level of the host, or their rate of deletion differs. Therefore, we suggest that the higher number of *PokeyA* may be due to a higher transposition rate, which could be tested using yeast excision assays.

Although *Pokey* number and distribution is similar in the four *D. pulex* and *D. pulicaria* lineages, it was necessary to set *gPokeyA* to 0 in six isolates because *rPokeyA* exceeds *tPokeyA*. The *tPokeyA* primer pair is located at

the 5' end of the transposase gene, and can only amplify autonomous *PokeyA*. Conversely, it was necessary to locate the *rPokeyA* forward primer near the 3' end of the gene, even though it was not possible to design primers in that region to target only *rPokeyA* or *rmPok1*. Consequently, the *rPokeyA* forward primer amplifies both autonomous and non-autonomous *rPokeyA*, as well as *rmPok1*. However, the higher number of *rPokeyA* compared to *tPokeyA* cannot be due to the presence of *rmPok1* because no copies of *mPok1* were detected in these isolates by end-point PCR using internal, *mPok1*-specific primers. Thus, the higher number of *rPokeyA* compared to *tPokeyA* in some isolates is most likely explained by copies of *rPokeyA* that did not amplify with the *tPokeyA* primers (Table 1, Fig. 3), but further study would be required to confirm this.

The number of *mPok* we observed in NA and EU *D. pulex* and NA and EU *D. pulicaria* is very low (0 to 2 haploid copies per isolate) compared to other *piggyBac* MITEs, for which numbers range from 50 to 2×10^4 in metazoans including nematodes, insects, and vertebrates [20]. Indeed, the absence of any *mPok1* in the 41 isolates we examined suggests that this MITE may not occur in these four lineages. In addition, the absence of *mPok2* in rDNA and the low number outside rDNA (maximum of two copies) suggests that the activity of *mPok2* is very low in these lineages.

Pokey* family distribution in *D. arenata

While *mPok1* and *mPok2* were both detected in *D. arenata*, the former was only detected outside rDNA and in low numbers suggesting its activity is low. The mean *p*-distance

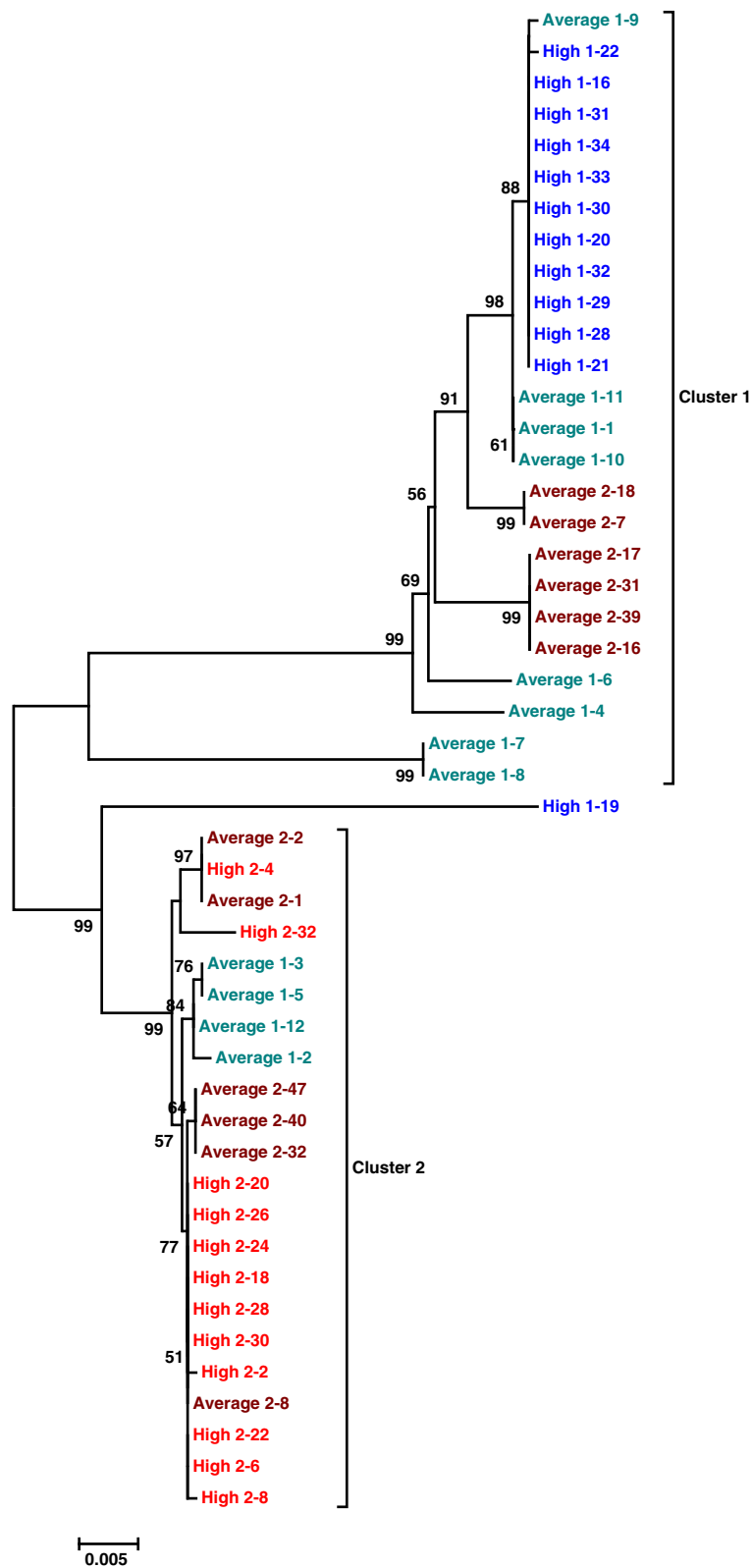


Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 Dendrogram of partial *PokeyA* sequences from four isolates of North American *Daphnia pulicaria*. The isolates were from two different populations, PC1 and PC2. Two isolates possessed High numbers of *PokeyA*, PC1.2 (blue) and PC2.1 (red); and two isolates possessed Average numbers of *PokeyA*, PC1.1 (teal) and PC2.2 (dark red). The sequences were approximately 1500 bp in length and included 519 bp of the 5' end of the *PokeyA* transposase and the length-variable region upstream. Bootstrap values >50 from 1000 replicates are provided on the nodes

between copies of *mPok1* and the *mPok1* consensus sequence is only 2 % in *D. arenata* isolate AR1.1 (Additional file 2: Figure S3). This, along with the fact that it does not occur in the other lineages, suggests that *mPok1* could be of recent origin. In contrast, *mPok2* occurred both inside and outside rDNA, with a mean haploid number of genomic copies greater than 30. This is at the top of the range for *gPokey* in the other four lineages (Fig. 3), suggesting that *mPok2* may be active in *D. arenata*. Recent activity is also consistent with a p-distance less than 3 % between the *mPok2* consensus sequence and 75 % of the *mPok2* in isolate AR1.1 (Additional file 2: Figure S3). However, the p-distance ranges from 4 to 22 % for the other copies of *mPok2* (Additional file 2: Figure S3). This, and the fact that *mPok2* is found in other lineages in the *D. pulex* complex, suggests that this MITE is not of recent origin [6]. Further work is required to determine the distribution of *mPok* in the *D. pulex* complex, and in other species in the subgenus *Daphnia*.

The number of *Pokey* and *mPok*, except *rPokeyA*, is substantially higher in *D. arenata* than the other four lineages, resulting in a very different distribution (Fig. 3). *Daphnia arenata* is endemic to western Oregon [13] and may have diverged from NA *D. pulex* during the Pleistocene glaciation, during which it was isolated from other refugial populations south of the glaciers by the Cordilleran ice sheet [21]. This restriction to a small geographic region could have resulted in a smaller effective population size, such that the fate of slightly deleterious mutations (such as some TE insertions) could have been primarily determined by genetic drift instead of purifying selection [22].

Another explanation for the high *Pokey* number in *D. arenata* is a much higher transposition rate compared to the other lineages. Such an increase could be the result of changes in *Pokey*, such as mutations in the transposase that increase transposition rate, or mutations in the terminal repeats that increase recognition by the transposase. These hypotheses could be tested using yeast excision, yeast one-hybrid, and/or electrophoretic mobility shift assays.

A third explanation for high *Pokey* number in *D. arenata* is a loss of epigenetic regulation by the host. The most well-known example of such a loss is the phenomenon of P-M hybrid dysgenesis in *Drosophila*, which occurs when a male from a strain with *P*-elements is mated to a female from a strain without *P*-elements [23]. The female usually provides the epigenetic regulation, but if she does not have *P*-elements, then the paternal *P*-elements are able to

transpose without regulation in the offspring [24], leading to effects such as male recombination, sterility, mutation, chromosomal aberrations, and nondisjunction [23].

Interspecific hybridization is also known to stimulate increased TE activity due to the disruption of host defenses, although this is not always the case (see Ungerer and Kawakami [25] for an example). Labrador et al. [26] found that the *Oswaldo* LTR retrotransposon has a higher transposition rate in hybrids of *Drosophila buzzatii* and *Drosophila koepferae* compared to non-hybrids. Similar increases in transposition rate in interspecific hybrids have been seen for other elements in rice and *Drosophila* (e.g. [27–31]).

Vergilino et al. [32] suggested that *D. arenata* may be a hybrid between NA *D. pulex* and *D. tenebrosa* or EU *D. pulicaria* because the position of *D. arenata* varies on phylogenetic trees generated from different nuclear loci [32–34]. NA *D. pulex* is likely the maternal parent based on analysis of mitochondrial DNA [13, 21]. If *D. arenata* does have a hybrid origin, release from host regulation could explain the relatively high number of *Pokey* and *mPok2* and possibly the occurrence of *mPok1*. Further work is required to establish the distribution of *mPok* in other *Daphnia* species, determine the origin of *mPok1*, confirm the hybrid origin of *D. arenata*, and test the hypothesis that the high *Pokey* load in *D. arenata* is a consequence of increased transposition in a hybrid lineage. If this is the case, we would also expect the number of active TEs from other superfamilies to be higher in *D. arenata* compared to other lineages in the *D. pulex* complex, but this remains to be tested.

Distribution of *Pokey* in the subgenus, *Daphnia*

Our results are very different from those obtained for *D. obtusa* in which *rPokey* number is much higher (mean = 75 for *PokeyA* and 9 for *PokeyB*) and *gPokey* number is much lower (mean = 0.5 for *PokeyA*) [15]. Moreover, *rPokeyA* is significantly positively correlated with 28S number in *D. obtusa* ($R^2 = 0.54$). The mean haploid rRNA gene number in lineages of the *D. pulex* complex is 240, with values ranging from 90 to nearly 580. Conversely, the mean haploid number in 21 isolates of *D. obtusa* is 415, with values ranging from 180 to 960. LeRiche et al. [15] suggested that *rPokeyA* number could be higher in *D. obtusa* due to the larger size of its rDNA array, which may be a consequence of higher rates of sister chromatid exchange [15]. A large rDNA array provides more *Pokey* insertion sites, and a high rate of sister chromatid exchange provides more

opportunity for the element to increase via recombination, even if the transposition rate is low.

Although the number of *Daphnia* species whose *Pokey* distribution has been studied is modest, the emerging pattern is that these elements tend to be concentrated in one genomic location (inside or outside of rDNA) and occur in low number in the other. In the *D. pulex* complex, *Pokey* is concentrated outside rDNA and in *D. obtusa* it is concentrated inside rDNA. Moreover, with the exception of *D. arenata*, *gPokey* in the *D. pulex* complex (less than 30 copies) tends to be much lower than *rPokey* in *D. obtusa* (up to 154 copies). Even so, the lower number of *gPokey* in the *D. pulex* complex is consistent with the number of other DNA elements observed in the *Daphnia* genome sequence as mentioned above [18].

One explanation for the tendency of *Pokey* to be concentrated in only one genomic location is purifying selection against chromosomal rearrangements caused by ectopic recombination between copies inside and outside rDNA. The ectopic recombination model [35] suggests that TEs should accumulate in genomic regions of low recombination due to the deleterious effects of exchange between elements at non-homologous sites. Although we do not know the location of *gPokey* in *Daphnia* genomes, their relatively low number suggests that their accumulation is resisted by selection. In contrast, Eickbush [36] suggested that ectopic recombination between TEs inserted in rDNA may not have severe effects because the outcome is similar to the effects of concerted evolution. On the other hand, ectopic recombination between *gPokey* and *rPokey* could severely impact genome organization and result in strong purifying selection, such that *Pokey* is generally not able to persist in substantial numbers in both locations in the same genome. *Daphnia arenata* is a notable exception to this general pattern as *PokeyB* and *mPok2* numbers are unusually high both inside and outside rDNA. If *Pokey* has been recently transposing in *D. arenata*, as suggested above, then there may not have been sufficient time for purifying selection to limit the expansion of *Pokey* in one or the other genomic location.

Variation among partial *PokeyA* sequences in North American *D. pulicaria*

Eagle and Crease [10] suggested that the transposition rate of *rPokeyA* may have recently increased in some populations of NA *D. pulicaria* because its number is unusually high in some isolates from populations PC1 and PC2. Sequence divergence between TE copies that have recently transposed is expected to be low as there has been little time for them to accumulate differences from their parent copy. Our results are consistent with this hypothesis as the sequences from isolates with high *PokeyA* number (PC1.2 and PC2.1) show significantly less deviation from their consensus sequence than do

sequences from isolates with average *PokeyA* numbers in the same population (PC1.1 and PC2.2, Table 3). Moreover, all but two of the sequences encode a potentially functional transposase. However, we cannot rule out the possibility that recent recombination within the rDNA array has also played a role in this *rPokeyA* expansion.

A recent increase in transposition rate is often a consequence of an element's ability to evade host silencing, which has been suggested to depend on the distribution of elements that occur in rDNA. For example, Eickbush et al. [37] found that rates of R2 transcription are higher when the elements are spread throughout the rDNA array than when they are clustered, and indirect evidence suggests that *Pokey* insertions are clustered in the rDNA of NA *D. pulex* [38]. Eickbush et al. [37] concluded that the largest continuous block of uninserted rDNA units tend to be transcribed, while other units remain silent. However, when R2 insertions are spread across the rDNA array, or have recently transposed into the midst of a large continuous block of uninserted units, they are more likely to be transcribed. Similarly, it is possible that recent insertion of *PokeyA* into a large continuous block of uninserted rDNA units in a NA *D. pulicaria* individual (either by recombination or transposition) triggered an increase in *PokeyA* transcription. The hypothesis that transcription rate is higher in individuals with high *PokeyA* number could be tested using nuclear run-on transcription assays and Reverse Transcription-PCR [37]. In addition, fiber-FISH assays could be used to determine the arrangement of *Pokey* within rDNA arrays. This technique has been used to show that *Pokey* is dispersed in isolate AR1.1 (Figure S11 in [18]).

Conclusions

The goals of this study were to determine the distribution of the *Pokey* family (*PokeyA*, *PokeyB*, *mPok1* and *mPok2*) inside and outside rDNA in lineages of the *D. pulex* complex, and to test the hypothesis that *rPokeyA* expansion in NA *D. pulicaria* is the result of recent transposition. We found the distribution to be similar in four of the lineages; *D. pulex* and *D. pulicaria* from North America (NA) and Europe (EU) and in general, *Pokey* is more common outside than inside rDNA. *PokeyA* expansion in NA *D. pulicaria* rDNA appears to be recent and we suggest it could have been triggered by a change in rDNA distribution that reduced the host's ability to regulate *Pokey* transcription.

The *Pokey* family distribution in *D. arenata* is very different from the other four lineages. In particular, the mean number of both *PokeyA* and *PokeyB* outside rDNA is five to six times higher in *D. arenata* and the two types of *mPok* are nearly exclusive to this lineage. We suggest that the proliferation of *Pokey* and *mPok* in *D. arenata* may be a consequence of release from epigenetic repression as a

result of interspecific hybridization, although a hybrid origin for this lineage requires confirmation.

The mean number of rRNA genes and *rPokeyA* is much larger in *D. obtusa* compared to the five *D. pulex* lineages. However, both *PokeyA* and *PokeyB* are absent or nearly so outside *D. obtusa* rDNA. Overall, these results suggest that *Pokey* primarily occupies only one or the other genomic location within a species. We suggest that this may be a consequence of purifying selection against ectopic recombination between elements in different genomic locations, which could cause severe genomic rearrangements.

Elliott et al. [6] suggested that *Pokey* originated as a genomic element and subsequently invaded *Daphnia* rDNA, which makes it unique compared to other DNA elements. Moreover, *Pokey* has been vertically inherited in the subgenus *Daphnia* since it diverged from the other *Daphnia* subgenera, *Ctenodaphnia* and *Hyalodaphnia* [10], which may have occurred as long ago as 100 million years [39]. It seems likely that the persistence of *Pokey* in *Daphnia* over such a long period of time is at least partially due to the potential for copies from one genomic location to reinvade the other if it is lost [6]. However, the factors that determine whether *Pokey* is primarily a genomic or an rDNA element within each species, and whether this configuration is stable over time are unknown and warrant further research.

Methods

Daphnia isolates and DNA extractions

We analyzed a total of 45 cyclically parthenogenetic *Daphnia* isolates from five lineages in the *D. pulex* complex (Fig. 1, Additional file 1: Table S1). Twenty-six of these isolates were NA *D. pulex*. One to three isolates were sampled from ten populations in the Midwest United States and southern Ontario. In addition, 19 isolates from the other lineages were included: four *D. arenata*, five NA *D. pulicaria*, eight EU *D. pulex*, and two EU *D. pulicaria*. *Daphnia* individuals were collected and clonally propagated (isolates) as described in Eagle and Crease [10]. DNA was extracted from multiple individuals from each isolate using phenol:chloroform or the Aquagenomics kit (MultiTarget Pharmaceuticals LLC, Salt Lake City, Utah, USA) as in Eagle and Crease [10]. DNA concentrations, which ranged from 13 to 4400 ng/ μ L (Additional file 1: Table S1), were measured using a NanoDrop[®] ND-8000 spectrophotometer (ThermoScientific).

Estimation of *Pokey* family and rRNA gene number

We used the SYBR green real-time qPCR and ΔC_T relative quantification method as described in Eagle and Crease [10] to estimate the haploid number of each type of gene. C_T , or cycle threshold, is the point at which the amplification curve crosses a set threshold. We estimated the number of the four TEs in the *Pokey* family (*PokeyA*, *PokeyB*, *mPok1*, and *mPok2*) in the entire genome (*tPokey*,

tmPok) and in rDNA (*rPokey*, *rmPok*), as well as the number of 18S and 28S rRNA genes in all 45 *Daphnia* isolates.

A total of 11 primer pairs were used (Table 4). The forward primers for *rPokey*/*rmPok* are located near the 3' end of the element and the reverse primer is located in the 28S rRNA gene downstream of the *Pokey* family insertion site. Standard curves were run to determine the percent amplification efficiency (PAE) for each primer pair (Table 4, Additional file 3).

qPCRs had a final volume of 20 μ L and contained 1X PerfeCTa SYBR Green FastMix, ROX (Quanta BioSciences) and 0.25 μ M of each primer. Reactions were run on a StepOne Plus instrument (Applied Biosystems) using the following protocol: 95 °C for 10 min; followed by 40 cycles of 95 °C for 5 s and 60 °C for 30 s. Melt curves were run by heating the amplicons at 95 °C for 15 s, decreasing the temperature to 60 °C for 1 min, then heating to 95 °C in 0.3 °C increments. The DNA template was either serial dilutions of amplicons generated from plasmid DNA (standard curve plates) or 10 ng of genomic DNA (experimental plates). Samples were run in triplicate with the exception of the *tmPok1* and *rmPok2* primer pairs for samples in which previous endpoint PCR produced no amplicon (Additional file 3). In such cases, only a single reaction was done. Negative control reactions were run for each primer pair on every standard curve plate. Negative control reactions were also run for each primer pair on 15 out of the 22 experimental plates (Additional file 3). The StepOne Software (Applied Biosystems) set the baseline for each reaction. The threshold was set based on amplicon size as in Eagle and Crease [10] (Table 4). A threshold of 0.2 was used for 50 bp amplicons, and the threshold for larger amplicons was determined using the formula, $0.2 \times 2^{[1-(50/\text{length in bp})]}$.

C_T values were used to estimate gene number, using the formula: $2^{-\Delta C_T}$ where ΔC_T is $([C_T \times \text{PAE}_{\text{multicopy gene}}] - [C_T \times \text{PAE}_{\text{single-copy gene}}])$, as in Eagle and Crease [10]. If the standard deviation for mean C_T for a gene was greater than 0.2, then the most extreme value was omitted (Additional file 1: Table S2) from further analysis. If there was no clear outlier among the three C_T values, then all three values were used. The two or three C_T values for each multicopy gene were compared to the two or three C_T values for both single copy genes producing 8 to 18 estimates of gene number. The *Tif-Gtp* ratio was 1.66 in isolate PX1.2 and 1.42 in isolate PX2.2; these two isolates likely have three copies of the *Tif* gene instead of the expected two. Therefore, the estimates of gene number using *Tif* as the single copy gene were multiplied by 1.5 in these isolates. In addition, both *tmPok1* and *rPokeyA* + *rmPok1* amplified in one isolate, AR3.1. To determine if the estimate of *rPokeyA* + *rmPok1* included *rPokeyA*, *rmPok1* or both, this isolate

Table 4 Primer pairs used for qPCR of the *Pokey* family and rRNA genes in the *Daphnia pulex* complex

| Gene ^a | Primer | Primer sequence | Amplicon Size (bp) | Source ^b | Threshold ^c | PAE ^d |
|-------------------|-------------|-----------------------------------|--------------------|-------------------------|------------------------|------------------|
| 18S | 18S 1864 F | 5'-ccg cgt gac agt gag caa ta | 50 | [17] | 0.200 | 0.947 |
| | 18S 1913 R | 5'-ccc agg aca tct aag ggc atc | | [17] | | |
| 28S | 28S 2508 F | 5'-gcc tgc tcg tac cga tat cc | 50 | | 0.200 | 0.937 |
| | 28S 2558 R | 5'-cta gag gct gtt cac ctt gga ga | | | | |
| tPokeyA | Pok 3720 F | 5'-cag ttc aaa gag tgg ctc ctc c | 50 | | 0.200 | 0.868 |
| | Pok 3770 R | 5'-cgg gtc tga ctt ctg gtt cg | | | | |
| tPokeyB | PokB 1375 F | 5'-aaa gag gag aag aat gac ccg g | 50 | | 0.200 | 0.897 |
| | PokB 1425 R | 5'-tca gaa gag cac cct acc ttg g | | | | |
| tmPok1 | mPok1 524 F | 5'-gga cac cta tgg cgg gat t | 50 | | 0.200 | 0.874 |
| | mPok1 574 R | 5'-cgc tga ggt ctg tcg gga | | | | |
| tmPok2 | mPok2 714 F | 5'-ggt cag ttg gct ccg aca a | 45 | | 0.185 | 0.911 |
| | mPok2 759 R | 5'-aa ccc ttt atc gac gcg aag a | | T. Elliott, unpublished | | |
| rPokeyA + rmPok1 | Pok 6561 F | 5'-caa tcg aat ccg acc atc g | 66 | | 0.237 | 0.914 |
| | 28S 3073 R | 5'-tga cga ggc att tgg cta cc | | | | |
| rPokeyB | PokB 4283 F | 5'-aat ttc agt caa gca cgg cc | 70 | | 0.244 | 0.903 |
| | 28S 3073 R | 5'-tga cga ggc att tgg cta cc | | | | |
| rmPok2 | mPok 714 F | 5'-ggt cag ttg gct ccg aca a | 68 | | 0.240 | 0.900 |
| | 28S 3073 R | 5'-tga cga ggc att tgg cta cc | | | | |
| Tif | TIF392F | 5'-gac atc atc ctg gtt ggc ct | 50 | [17] | 0.200 | 0.936 |
| | TIF442R | 5'-aac gtc agc ctt ggc atc tt | | [17] | | |
| Gtp | GTP385R | 5'-tat tca gca tgg aga gac ggc | 50 | [17] | 0.200 | 0.928 |
| | GTP435R | 5'-gat gtc gac tga cgc tgg aa | | [17] | | |

^a18S = 18S rRNA genes, 28S = total 28S rRNA genes, tPokey = total *Pokey*, tmPok = total mPok, rPokey = *Pokey* inserted in 28S, rmPok = mPok1 inserted in 28S

^bUnless otherwise indicated, primers were designed for this study

^cA threshold of 0.2 was used for 50 bp amplicons. The threshold for larger amplicons was determined using the formula, $0.2 \times 2^{[1-(50/\text{length in bp})]}$ as in Eagle and Crease [10]

^dPAE = primer amplification efficiency

was screened for rmPok1 with end-point PCR (Additional file 3), but no amplicon was detected.

gPokey/gmPok number was calculated as $[tPokey - rPokey / gmPok - rmPok]$. In six isolates (EPX1.1, EPX1.2, EPX2.1, EPX2.2, PC2.1, and PX6.3), the calculation of *gPokeyA* produced a negative number in which case *gPokeyA* was set to zero, and *tPokeyA* was assumed to be the same as *rPokeyA*.

We used paired t-tests for means, two-sample t-tests assuming unequal variances with the sequential Bonferroni correction [40], and linear regressions (Microsoft Excel, Richmond, Washington, USA) to examine the relationship between 18S and 28S rRNA gene number. We also estimated the correlation between rRNA genes and *Pokey* family number using regression analysis.

Sequencing partial *PokeyA* copies from North American *D. pulicaria*

We cloned and sequenced partial *PokeyA* elements from four NA *D. pulicaria* isolates from two populations. From each population, an isolate with high number of *PokeyA*

(PC1.2, PC2.1) and an isolate with average number of *PokeyA* (PC1.1, PC2.2) were selected (Table 3). The average number of *PokeyA* for NA *D. pulicaria* was based on the results of Eagle and Crease [10]. Partial *PokeyA* sequences were amplified from approximately 50 ng of genomic DNA in a 25 μ L reaction, which contained 1X Phusion HF Buffer (NEB), 0.4 mM dNTPs, 0.08 μ M each of the Pok-2904 F (5' ggg aca tag gtg tcc cgg) and Pok-6178R (5' tcg acc agg ggt ctt tcc agt c) primers, and 1 units of Phusion DNA Polymerase. Reactions were run on either a PTC-100 Thermocycler (MJ Research) or a T100 Thermocycler (Bio-Rad Laboratories, Inc.) using the following protocol: 3 min initial denaturation at 98 $^{\circ}$ C; 35 cycles of 10 s denaturation at 98 $^{\circ}$ C, 30 s annealing at 55 $^{\circ}$ C, and 2 min elongation at 72 $^{\circ}$ C; followed by a final elongation at 72 $^{\circ}$ C for 10 min. The 3.3 kb amplicons were verified by electrophoresis on a 1 % TAE agarose gel, stained with GelRedTM Nucleic Acid Gel Stain (Biotium, Inc.) and visualized under UV light. The amplicons were then cloned into the plasmid, pSC-B-amp/kan using the StrataClone Blunt PCR Cloning Kit (Agilent Technologies) as per the

manufacturer's protocol with a few modifications. The modifications were as follows: 2.5 μ L of PCR product and 0.5 μ L of StrataClone Blunt Vector Mix were used in the ligation; 25–50 μ L of StrataClone SoloPack cells was used for the transformation; an incubation of 30–45 min was used because the insert was large (3.3 kb); 500 μ L of Terrific Broth was used instead of 250 μ L of LB medium; and 50 μ L and 400 μ L were plated instead of 5 μ L and 100 μ L.

Positive colonies (white) were triple-streaked onto new plates and grown overnight. One of the three streaks was added with a toothpick to 10 μ L of water and incubated at 99.9 $^{\circ}$ C for 3 min. One microliter was amplified in a 25 μ L reaction with two sets of primers: Pok-2904 F with Pok-3811R (5' ccg tgt tac ttc acc atc gg) to generate a 910 bp fragment; and Pok-3720 F (5' cag ttc aaa gag tgg ctc c) with Pok-4488R (5' gaa tcg ctc gcg agt cat gg) to generate a 770 bp fragment. Reactions contained 1X GenScript buffer (GenScript USA Inc.), 0.04 mM dNTPs, 0.04 μ M of each primer, and 0.5 units of GenScript DNA polymerase (GenScript USA Inc.). Reactions were run on either a PTC-100 Thermocycler (MJ Research) or a T100 Thermocycler (Bio-Rad Laboratories, Inc.) using the following protocol: 2 min initial denaturation at 94 $^{\circ}$ C; 35 cycles of 30 s denaturation at 94 $^{\circ}$ C, 30 s annealing at 55 $^{\circ}$ C, and 1 min extension at 72 $^{\circ}$ C; followed by a final elongation at 72 $^{\circ}$ C for 5 min. Amplicons from 12 clones per isolate were sequenced with the primers Pok-2904 F and Pok-3720 F in 12 μ L reactions. The reactions contained 0.3 μ L BigDye[®] Terminator v3.1 Ready Reaction Mix (Applied Biosystems), 0.4X Sequencing Buffer (Applied Biosystems), 0.83 μ M of primer, and 2 μ L of amplicon. Sequencing reactions were run for 1 min at 96 $^{\circ}$ C followed by 30 cycles of a 20 s denaturation at 96 $^{\circ}$ C, a 20 s annealing at 55 $^{\circ}$ C, and a 4 min extension at 60 $^{\circ}$ C. Reactions were resolved on an ABI 3730 DNA Analyzer (Applied Biosystems) by the Genomics Facility at the University of Guelph. Sequences were analyzed using CLC Main Workbench software (CLC Bio). We used MEGA 5.0 [41] to generate a 1608 bp alignment containing 48 sequences from the four NA *D. pulicaria* isolates. The maximum composite likelihood model was used to estimate a matrix of pairwise sequence divergence from which a Neighbor-Joining dendrogram was generated. Gaps were excluded from this analysis using the pairwise deletion option. In addition, a consensus sequence was generated for each isolate from the 12 sequences. The number of differences between each sequence and its consensus was determined in MEGA. Gaps were included in this analysis, but consecutive gaps were coded as single nucleotide changes. ANOVA was used to test for heterogeneity among the mean number of differences in each of the four isolates. A *post hoc* Tukey's HSD test was used to compare the means in the average and high isolates from the same population.

Additional files

Additional file 1: Table S1. Details of *Daphnia* isolates analyzed in this study. **Table S2.** C_T values from qPCR analysis of *Pokey*, *mPok*, and rRNA gene number in 45 isolates from the *Daphnia pulex* complex. **Table S3.** Estimates of *Pokey*, *mPok*, and rRNA gene number in 45 isolates of the *Daphnia pulex* complex. (XLSX 51 kb)

Additional file 2: Figure S1. Correlation between 18S and 28S rRNA gene number *D. pulex* lineages. **Figure S2.** Correlation between 28S rRNA gene and *rPokey* family number in *D. pulex* lineages. **Figure S3.** Repeat landscape for *mPok* from *Daphnia arenata* isolate AR1.1. (PDF 34 kb)

Additional file 3: Details of qPCR methods. (PDF 232 kb)

Abbreviations

18S, 18S rRNA gene; 28S, 28S rRNA gene; bp, base pair; C_T , threshold cycle; ETS, external transcribed spacer; EU, European; *gmPok1*, *mPok1* found outside rDNA; *gmPok2*, *mPok2* found outside rDNA; *gPokey*, *Pokey* elements found outside rDNA; *gPokeyA*, *mPok1* found outside rDNA; *gPokeyB*, *mPok2* found outside rDNA; IGS, intergenic spacer; ITS, internal transcribed spacer; kb, kilo base pairs; LTR, long terminal repeats; MITE, Miniature Inverted-repeat Transposable Element; NA, North American; PAE, percent amplification efficiency; qPCR, quantitative Polymerase Chain Reaction; rDNA, ribosomal DNA; *rmPok1*, *mPok1* in the 28S gene; *rmPok2*, *mPok2* in the 28S gene; *rPokey*, *Pokey* elements found in the 28S gene; *rPokeyA*, *mPok1* in the 28S rRNA gene; *rPokeyB*, *mPok2* in the 28S rRNA gene; rRNA, ribosomal RNA; TE, transposable element; TIR, terminal inverted repeat; *tmPok1*, *mPok1* in the genome; *tmPok2*, *mPok2* in the genome; *tPokey*, *Pokey* elements found in the genome; *tPokeyA*, *PokeyA* in the genome; *tPokeyB*, *PokeyB* in the genome; TSD, target site duplication.

Acknowledgements

We thank Andrew Beckerman, Melania Cristescu, France Dufresne, Michael Lynch, Seanna McTaggart and Julia Reger for providing some of the *Daphnia* samples, and the Genomics Facility at the University of Guelph for assistance with the qPCR and sequencing of plasmid clones. Comments from Tyler Elliott and two anonymous reviewers greatly improved the manuscript.

Funding

This research was funded by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada to T.J.C. SHCE was partially supported by an Ontario Graduate Studies Science and Technology Scholarship. The funding agencies had no involvement in the design of the study, the collection, analysis, and interpretation of data, or writing the manuscript.

Availability of data and material

The sequence dataset supporting the conclusions of this article is available in GenBank, with accession numbers KX354776 to KX354823. All other data supporting the conclusions of this article are included within the article and its additional files.

Authors' contributions

SHCE participated in the design of the study, carried out the molecular genetic analyses, analyzed the data, and drafted the manuscript. TJC conceived of the study and participated in its design, and helped to draft the manuscript. Both authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable

Ethics approval and consent to participate

Not applicable

Received: 27 March 2016 Accepted: 13 June 2016

Published online: 17 June 2016

References

- Nuzhdin SV. Sure facts, speculations, and open questions about the evolution of transposable element copy number. *Genetica*. 1999;107:129–37.
- Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet*. 2007;8:272–85.
- Schaack S, Gilbert C, Feschotte C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol*. 2010;25:537–46.
- Penton EH, Sullender BW, Crease TJ. *Pokey*, a new DNA transposon in *Daphnia* (Cladocera: Crustacea). *J Mol Evol*. 2002;55:664–73.
- Penton EH, Crease TJ. Evolution of the transposable element *Pokey* in the ribosomal DNA of species in the subgenus *Daphnia* (Crustacea: Cladocera). *Mol Biol Evol*. 2004;21:1727–39.
- Elliott TA, Stage DE, Crease TJ, Eickbush TH. In and out of the rDNA genes: characterization of *Pokey* elements in the sequenced *Daphnia* genome. *Mobile DNA*. 2013;4:20.
- Eickbush TH, Eickbush DG. Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics*. 2007;175:477–85.
- Kojima KK, Jurka J. A superfamily of DNA Transposons targeting multicopy small RNA genes. *PLoS ONE*. 2013;8:68260.
- Valizadeh P, Crease TJ. The association between breeding system and transposable element dynamics in *Daphnia pulex*. *J Mol Evol*. 2008;66:643–54.
- Eagle SHC, Crease TJ. Copy number variation of ribosomal DNA and *Pokey* transposons in natural populations of *Daphnia*. *Mobile DNA*. 2012;3:4.
- Prokopowich CD, Gregory TR, Crease TJ. The correlation between rDNA copy number and genome size in eukaryotes. *Genome*. 2003;46:48–50.
- Hebert P. The population biology of *Daphnia* (Crustacea, Daphniidae). *Biol Rev*. 1978;53:387–426.
- Colbourne J, Crease T, Weider L, Hebert P, Dufresne F, Hobaek A. Phylogenetics and evolution of a circumarctic species complex (Cladocera: *Daphnia pulex*). *Biol J Linn Soc*. 1998;65:347–65.
- Hebert P. The *Daphnia* of North America: an illustrated fauna. CD-ROM. 1995. University of Guelph, Guelph, ON.
- LeRiche K, Eagle SHC, Crease TJ. Copy number of the transposon, *Pokey*, in rDNA is positively correlated with rDNA copy number in *Daphnia obtusa*. *PLoS ONE*. 2014;9:114773.
- Feschotte C, Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. *Ann Rev Genet*. 2007;41:331–68.
- McTaggart SJ, Dudycha JL, Omilian A, Crease TJ. Rates of recombination in the ribosomal DNA of apomictically propagated *Daphnia obtusa* lines. *Genetics*. 2007;175:311–20.
- Colbourne JK, Pfrender ME, Gilbert DG, Thomas WK, Tucker A, Oakley TH, et al. The ecoresponsive genome of *Daphnia pulex*. *Science*. 2011;331:555–61.
- Averbeck KT, Eickbush TH. Monitoring the mode and tempo of concerted evolution in the *Drosophila melanogaster* rDNA locus. *Genetics*. 2005;171:1837–46.
- Feschotte C, Zhang X, Wessler S. Miniature inverted-repeat transposable elements (MITEs) and their relationship with established DNA transposons. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. *Mobile DNA II*. Washington DC: ASM Press; 2002. p. 1147–58.
- Crease T, Lee S-K, Yu S-L, Spitze K, Lehman N, Lynch M. Allozyme and mtDNA variation in populations of the *Daphnia pulex* complex from both sides of the Rocky Mountains. *Heredity*. 1997;79:242–51.
- Lynch M, Conery JS. The origins of genome complexity. *Science*. 2003; 302:1401–4.
- Kidwell MG, Kidwell JF, Sved JA. Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics*. 1977;86:813–33.
- Malone CD, Hannon GJ. Small RNAs as guardians of the genome. *Cell*. 2009; 136:656–68.
- Ungerer MC, Kawakami T. Transcriptional dynamics of LTR retrotransposons in early generation and ancient sunflower hybrids. *Gen Biol Evol*. 2013;5:329–37.
- Labrador M, Farré M, Utzet F, Fontdevila A. 1999 Interspecific hybridization increases transposition rates of *Osvaldo*. *Mol Biol Evol*. 1999;16:931–7.
- Shan X, Liu Z, Dong Z, Wang Y, Chen Y, Lin X, Long L, Han F, Dong Y, Liu B. Mobilization of the active MITE transposons mPing and Pong in rice by introgression from wild rice (*Zizania latifolia* Griseb). *Mol Biol Evol*. 2005;22: 976–90.
- Wang H-Y, Tian Q, Ma Y-Q, Wu Y, Miao G-J, Ma Y, Cao D-H, Wang X-L, Lin C, Pang J, Liu B. Transpositional reactivation of two LTR retrotransposons in rice-*Zizania* recombinant inbred lines (RILs). *Heredity*. 2010;147:264–77.
- Wang N, Wang H, Wang H, Zhang D, Wu Y, Ou X, Liu S, Dong Z, Liu B. Transpositional reactivation of the *Dart* transposon family in rice lines derived from introgressive hybridization with *Zizania latifolia*. *BMC Plant Biol*. 2010;10:190.
- Vela D, Fontdevila A, Vieira C, García Guerreiro MP. A genome-wide survey of genetic instability by transposition in *Drosophila* hybrids. *PLoS ONE*. 2014;9:88992.
- García Guerreiro MP. Interspecific hybridization as a genomic stressor inducing mobilization of transposable elements in *Drosophila*. *Mob Genet Elements*. 2014;4:34394.
- Vergilino R, Markova S, Ventura M, Manca M, Dufresne F. Reticulate evolution of the *Daphnia pulex* complex as revealed by nuclear markers. *Mol Ecol*. 2011;20:1191–207.
- Omilian AR, Lynch M. Patterns of intraspecific DNA variation in the *Daphnia* nuclear genome. *Genetics*. 2009;182:325–36.
- Crease TJ, Floyd R, Cristescu ME, Innes D. Evolutionary factors affecting *Lactate dehydrogenase A* and *B* variation in the *Daphnia pulex* species complex. *BMC Evol Biol*. 2011;11:212.
- Montgomery E, Charlesworth B, Langley CH. A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. *Genet Res*. 1987;46:31–41.
- Eickbush TH. R2 and related site-specific non-long terminal repeat retrotransposons. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. *Mobile DNA II*. Washington DC: ASM Press; 2002. p. 813–35.
- Eickbush DG, Ye J, Zhang X, Burke WD, Eickbush TH. Epigenetic regulation of retrotransposons within the nucleolus of *Drosophila*. *Mol Cell Biol*. 2008;28:6452–61.
- Glass S, Moszczyńska A, Crease TJ. The effect of transposon *Pokey* insertions on sequence variation in the 28S rRNA gene of *Daphnia pulex*. *Genome*. 2008;51:988–1000.
- Kotov AA, Taylor DJ. Mesozoic fossils (>145 Mya) suggest the antiquity of the subgenera of *Daphnia* and their coevolution with chaoborid predators. *BMC Evol Biol*. 2011;11:129.
- Rice W. Analyzing tables of statistical tests. *Evolution*. 1989;43:223–5.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28:2731–9.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

