Insertion Polymorphisms of Mobile Genetic Elements in Sexual and Asexual Populations of *Daphnia pulex*

Xiaoqian Jiang[1,*], Haixu Tang[2,*], Zhiqiang Ye[1], and Michael Lynch[1]

[1]Department of Biology, Indiana University, Bloomington, Indiana

[2]School of Informatics and Computing, Indiana University, Bloomington, Indiana

*Corresponding authors: E-mails: jiangxi@indiana.edu; hatang@indiana.edu.

Accepted: December 29, 2016

Abstract

Transposable elements (TEs) constitute a substantial portion of many eukaryotic genomes, and can in principle contribute to evolutionary innovation as well as genomic deterioration. *Daphnia pulex* serves as a useful model for studying TE dynamics as a potential cause and/or consequence of asexuality. We analyzed insertion polymorphisms of TEs in 20 sexual and 20 asexual isolates of *D. pulex* across North American from their available whole-genome sequencing data. Our results show that the total fraction of the derived sequences of TEs is not substantially different between asexual and sexual *D. pulex* isolates. However, in general, sexual clones contain fewer fixed TE insertions but more total insertion polymorphisms than asexual clones, supporting the hypothesis that sexual reproduction facilitates the spread and elimination of TEs. We identified nine asexual-specific fixed TE insertions, eight long terminal repeat retrotransposons, and one DNA transposon. By comparison, no sexual-specific fixed TE insertions were observed in our analysis. Furthermore, except one TE insertion located on a contig from chromosome 7, the other eight asexual-specific insertion sites are located on contigs from chromosome 9 that is known to be associated with obligate asexuality in *D. pulex*. We found that all nine asexual-specific fixed TE insertions can also be detected in some Daphnia *pulicaria* isolates, indicating that a substantial number of TE insertions in asexual *D. pulex* have been directly inherited from *D. pulicaria* during the origin of obligate asexuals.

Key words: transposable elements, TE insertions, *Daphnia pulex*, *Daphnia pulicaria*, asexuality.

Introduction

Transposable elements (TEs) are mobile DNA elements that can replicate from one location to another within a genome or between genomes, and are believed to be a major factor in large-scale genome reorganization. TEs constitute a large portion of many eukaryotic genomes, although the proportions of TEs vary among species, for example, TEs constitute ~4% of the *Drosophila* genome, ~6.5% of the *Caenorhabditis elegans* genome, ~45% of the human genome, ~60% of the soybean genome, and up to ~80% of the maize genome (SanMiguel et al. 1996; Duret et al. 2000; Lander et al. 2001; Kaminker et al. 2002). Mobilization of TEs is capable of introducing gene inactivation or illegitimate recombination, and the insertion of TEs typically reduces the viability of the host genome (Hurst and Werren 2001). Therefore, active TEs are a significant source of deleterious mutations. For example, TEs cause ~10 and ~50% of deleterious mutations in mammals and *Drosophila*, respectively (Finnegan 1992; Kazazian 1998). In a few cases, TE insertions may provide an adaptive advantage (González and Petrov 2009; Biémont 2010; de Souza et al. 2013), a premium example being the *Het-A* elements that comprise the telomeric DNA of *Drosophila* (Zemojtel et al. 2009).

TEs are generally divided into two major classes based on their transposition mechanisms: retrotransposons or class I elements, spreading in a "copy-and-paste" manner, copying through an RNA template prior to inserting into the genome, and DNA transposons or class II elements, spreading through a "cut-and-paste" manner, enabling movement and integration into a new location in the genome without copying the parent allele (Wicker et al. 2007). Retrotransposons can be further subdivided into two subclasses: long terminal repeat (LTR) retrotransposons, and non-LTR retrotransposons. LTR retrotransposons are composed of a coding region with a pair of LTRs, containing opening reading frames (ORFs) for *Gag* and *Pol* coding regions. The *Pol* ORF includes four key protein domains required for the replication and transposition of TEs into host genome: aspartic protease (AP), RNase H (RH),

reverse transcriptase (RT), and integrase (INT), among which RT is the most conserved protein domain. Intact non-LTR retrotransposons contain a protein region encoding an RT domain, which has been used as the main signal to identify and classify these elements (Rho and Tang 2009). Most autonomous DNA transposons are composed of a transposase coding region flanked by terminal inverted repeats (TIRs), except for the newly identified clade of *helitrons* (Jiang 2013a). Retrotransposons outnumber DNA transposons in most species partially due to their intrinsically higher proliferative potential. DNA transposons and a small subset of LTR retrotransposons are capable of horizontal transmission (Arkhipova and Meselson 2005). In contrast, non-LTR retrotransposons can only be transmitted vertically, and phylogenetic analysis of RT and RH protein domains indicates that non-LTR retrotransposons are much more ancient than LTR retrotransposons (Eickbush and Malik 2002; Arkhipova and Meselson 2005). The evolutionary factors that affect TE abundances have been studied extensively (Wright et al. 2003; Le Rouzic and Capy 2005; Le Rouzic et al. 2007), including the intrinsic mechanisms of mobilization, population structure, and the interaction between TEs and their host, among which the reproductive manner of the host is thought to play an important role in determining the fate of TEs (Charlesworth 1994; Wright and Finnegan 2001; Lynch 2007).

Although rare, asexual lineages arise from sexual ones independently in some species. Whether loss of sex will cause an increase or decrease in TE abundance remains undetermined. On the one hand, theoretical models predict that meiosis and recombination in sexual lineages facilitate the spread of TEs to new individuals in a population even if TEs are (mildly) deleterious (Hickey 1982). In contrast, in asexual lineages lacking recombination, within-lineage transmission of genes in asexual isolates is expected to limit the spread of TEs (Charlesworth and Langley 1986). On the other hand, in the absence of sex, TEs may become less virulent because the interests of TEs are consistent with the interests of their host. Regulatory mechanisms in asexuals may also play more important roles in reducing the rate of transposition than in sexuals to defend deleterious effects of TE load (Wright and Finnegan 2001). Moreover, asexual lineages typically arise from sexual ones and thus inherit TEs from sexuals. The reduced effective population size in asexual lineages may allow genetic drift to fix TEs by a Muller's ratchet-like process at a higher rate than in sexuals (Dolgin and Charlesworth 2006). Therefore, the abundance of TEs in asexual and sexual lineages is relevant to our understanding of the evolutionary mechanisms of driving the architecture of genomes.

An earlier study showed that when *Ty3* retrotransposons are introduced into the experimental sexual and asexual populations of the yeast *Saccharomyces cerevisiae*, the new TEs spread faster and are more likely to become established in sexual populations compared with asexual populations (Zeyl et al. 1996). Although the ancient bdelloid rotifers are not strictly asexual due to the presence of some gene exchanges between individuals (Debortoli et al. 2016), they have fewer DNA transposons and lack several classes of retrotransposons compared to their sexual relatives (Arkhipova and Meselson 2000, 2005; Flot et al. 2013). Moreover, the selfing plant *Arabidopsis thaliana* experienced a recent reduction in TE abundance compared to its outcrossing relative *Arabidopsis lyrata* (De La Chaux et al. 2012). Finally, a recent study compared the whole-genome TE loads of five asexual arthropod lineages and their sexual relatives, including crustaceans, insects, and mites, suggesting that there is no evidence of excess TE accumulation in genomes of asexuals as compared to sexual lineages (Bast et al. 2016).

The *Daphnia pulex* species complex, a group of freshwater microcrustaceans, where obligate parthenogenetic (asexual) lineages and cyclically parthenogenetic (sexual) lineages coexist, is one of the best-documented systems of contagious asexuality in animals (Lynch et al. 2008). Sexual *D. pulex* lineages typically reproduce offspring by females directly developing embryos through parthenogenesis; however, they still may produce males capable of meiosis in unfavorable conditions. In contrast, asexual *D. pulex* lineages lose the ability to engage in sexual reproduction and reproduce strictly by obligate parthenogenesis in any environment. Interestingly, the lack of sexual reproduction in asexual lineages is limited to females, and males produced by some asexual females are still capable of haploid gamete production, providing a way for transferring asexuality elements to sexual populations via backcrossing (Innes and Hebert 1988; Xu et al. 2015). Previous studies showed that the estimated age of obligately asexual *D. pulex* ranges from as young as ~1,000 years (Lynch et al. 2008) to as old as ~172,000 years (Paland et al. 2005). Recent genome-wide association studies also suggested that the spread of asexuality in *D. pulex* is perhaps as recent as 1,250 years, although the origin of the meiosis-suppressing elements could be substantially older (Tucker et al. 2013). The authors observed that all asexual clones share the same haplotypes in several genomic regions, including almost the entirety of chromosomes 8 and 9, both of which are thought to be associated with obligate asexuality. Recent studies suggested that the mutations that suppress meiosis and cause obligate asexuals originated via historical hybridization and introgression events between *D. pulex* and a closely related sister species, *D. pulicaria* (Xu et al. 2013). Given the widespread presence of sexually and asexually reproducing *D. pulex* populations and the independent origin of multiple asexual lineages (Innes and Hebert 1988), *D. pulex* complex is an ideal system for investigating how sexual or asexual reproduction shapes the dynamics of TEs.

TEs comprise ~10% of the *D. pulex* genome (*TCO*), among which LTR retrotransposons constitute ~8.0% of the *D. pulex* genome, while non-LTR retrotransposons and DNA transposons each constitute ~0.7% of the entire genome (Colbourne

et al. 2011). Schaack et al. (2010a, b) used transposon display to investigate the dynamics of DNA transposons in *D. pulex*, and showed that sexual lineages exhibit higher insertion-site polymorphism of DNA transposons than asexual lineages, suggesting that sex and recombination may facilitate the spread of DNA transposons. Eads et al. (2012) found a unique TE insertion site near a meiotic cohesin related gene *Rec8* in all tested obligate asexuals, while the TE insertion is completely absent in all sexuals. However, because empirical studies focused on only one or a few TE families in *D. pulex*, our understanding of TEs in *D. pulex* is still far from comprehensive due to the lack of pertinent large-scale comparison of TEs in asexual and sexual lineages.

In this study, we evaluated 20 asexual and 20 sexual isolates from North America by using whole-genome sequencing data to investigate whether the distribution pattern of TEs is different in asexual/sexual *D. pulex* isolates and whether TE insertions are associated with the origin of asexuality. With a recently available high-quality reference genome assembly of *D. pulex* (Ye et al. 2016), we investigated the insertion polymorphisms of TEs in a group of cyclically sexual and obligately asexual *D. pulex*. We first reconstructed an improved *in silico* TE library by searching for TE sequences in two available *D. pulex* genomes, *PA42* (Ye et al. 2016) and *TCO* (Colbourne et al. 2011), and then developed an algorithm based on paired-end reads to estimate the composition and the insertion polymorphisms of TEs in each isolate. This study constitutes the first comprehensive comparative analysis of TEs between sexual and asexual *D. pulex* populations, clarifying the different behavior of TEs in the two systems.

Materials and Methods

Data Source

Genomic next-generation sequencing data of *D. pulex* were partly retrieved from a previous study (Tucker et al. 2013). Among the available 22 *D. pulex* isolates (11 sexuals and 11 obligate asexuals), we abandoned 2 isolates with low sequencing quality and retained 10 sexual and 10 obligately asexual isolates each with coverage between 30X and 41X. Another 10 sets of genome sequencing data from asexual *D. pulex* isolates (with coverage between 21X and 29X) were retrieved from another recent study (Xu et al. 2015), named as "urban" isolates because they are often discovered in recently deforested areas and can inhabit in waters contaminated by heavy metals. The urban isolates also carry diagnostic *D. pulicaria* alleles in addition to chromosomes 8 and 9 (Hebert and Crease 1983; Xu et al. 2013). Although there are additional 11 asexual *D. pulex* isolates available (Xu et al. 2015), the sequencing data of these isolates are not suitable for this study due to short insert size (~150–200 bp) between paired-end reads and thus were not used here. Another 10 sexual *D. pulex* isolates (with coverage between 27X and 33X) were extracted from four different sequencing population data sets in Lynch lab. Among the 10 paired-end whole-genome sequencing data from *D. pulicaria*, 6 are from Xu et al. (2015), and the other 4 are sequenced by Joseph Shaw Lab at Indiana University used as the supplemental data to verify our conclusion. The unpublished raw fastq files of 10 sexual *D. pulex* and 4 *D. pulicaria* isolates have been deposited at NCBI Short Read Archive under the accession PRJNA355243, while the new *D. pulex* genomic assembly *PA42* has been deposited at NCBI under the accession PRJEB14656. The sources, reproduction modes, and coverage of all the used 20 asexual *D. pulex*, 20 sexual *D. pulex*, and 10 *D. pulicaria* isolates are summarized in supplementary table S1, Supplementary Material online.

De Novo Identification of TEs

Although there are available TE sequences from a previous *D. pulex* assembly *TCO* (Colbourne et al. 2011), we still searched for all the TEs in both *D. pulex* reference genome *de novo* (*PA42* and *TCO*) due to more available TE identification software now. The three types of full-length TEs, including LTR retrotransposons, non-LTR retrotransposons, and DNA transposons, were identified separately.

Identification of Intact LTR Retrotransposons

Two computer software were used here to search for all potential full-length LTR retrotransposons: LTRharvest (Ellinghaus et al. 2008) and MGEScan-LTR (Rho et al. 2007). A previous comparison of various *de novo* LTR-element identification programs has suggested that these two programs give the best results related to the number of true detected LTR retrotransposons (Lerat 2010). Both programs identify the structural features of LTRs and return candidate TE sequences. Because the outputs likely contain all kinds of repeats, several filtering rules based on multiple structural rules were applied to exclude non-TE sequences: detection of a pair of LTRs with length >80 bp and similarity $>80\%$ at the nuclear level, and the presence of putative *Pol* ORFs. To identify the four key protein domains (AP, RH, RT, and INT) of *Pol* ORFs included in LTR retrotransposons, we first translated each candidate TE into six frames with *transeq* from the EMBOSS package (Olson 2002), and then searched against Pfam HMM and CDD profiles by using RPS-Blast with stop codons (Marchler-Bauer et al. 2011; Finn et al. 2014). The profiles included: RVT_1 PF00078, RVT_2 PF07727, RT_LTR cd01647, RT_pepA17 cd01644, RT_DIRS1 cd03714, cd09272 RNase_HI_RT_Ty1, cd09273 RNase_HI_RT_Bel, cd09274 RNase_HI_RT_Ty3, cd09275 RNase_HI_RT_DIRS1, rve PF00665, cd00799 INT_Cre_C, Peptidase_A17 PF05380, Retrotran_gag PF03732, RVP PF00077. An LTR retrotransposon was retained for further analysis if at least one protein domain was found in the element with an *E*-value $<1e{-}10$. The elements identified by both programs were combined together. If overlapping or

nested elements were identified by different programs, we only kept the longer element by manually check.

To classify the different clades (*Ty1/Copia*, *Ty3/Gypsy*, *Bel/Pao*, and *Dirs*) of LTR retrotransposons, we first extracted the RT sequence of each element if existence of an RT protein domain, and then used RT sequences to classify elements. If absence of an RT sequence, the element was clustered with CD-HIT based on the sequence similarity (>80%) of LTRs (Fu et al. 2012). For a few elements unclassifiable by RT domain or LTRs, we aligned them to the already classified elements to obtain clade information. Reference RT sequences used to classify LTR retrotransposons were obtained from NCBI, listed in supplementary table S2, Supplementary Material online (Rho et al. 2010). Multiple alignments were carried out by using Mafft (Katoh et al. 2005) and the phylogenetic analysis was conducted by using FastTree (Price et al. 2009).

Identification of Non-LTR Retrotransposons

Autonomous non-LTR retrotransposons (LINE) were identified using program MGEScan-nonLTR, which uses probabilistic model to search for RT protein domain (Rho and Tang 2009). We only retained elements containing RT domain for further analysis because all autonomous non-LTR retrotransposons encode RT. Similar to the procedure of classifying LTR retrotransposons into different clades, putative non-LTR retrotransposon sequences were also translated into six frames and searched with RPS-Blast against RT domains of non-LTR retrotransposons in CDD profiles (RT_nLTR_like cd01650), while only the elements with an *E*-value <1e−10 were retained. RT sequence was extracted from each candidate element to classify non-LTR retrotransposon into different clades. Five clades of reference non-LTR RT sequences listed in supplementary table S3, Supplementary Material online were extracted from Repbase update library (Jurka et al. 2005).

Identification of DNA Transposons

Autonomous DNA transposons were identified in *D. pulex* genomes by using program MGEScan-DT (Ismail and Tang 2016, unpublished data) with the following procedure: first, six representative transposase sequences (*Harbinger*, *hAT*, *sil2eu*, *P element*, *Mariner, helicase*) are extracted from Repbase update library. Second, the hidden Markov model of each clade is constructed by using HMMer. MGEScan-DT uses HMMer to search for genes encoding transposases in the genome sequences against these clade-specific pHMMs. Once it identifies the hit region with acceptable *E*-value (<1e−10), the sequence is extended in both sides to locate the TIRs within the acceptable sequence length (3–20-kb) (Jiang 2013b). This program automatically classifies the DNA transposons into different clades according to their transposase sequences.

All three subclasses of TE sequences identified in both genomes were combined and redundant elements with the full sequence similarity >98% were removed by using CD-HIT clustering algorithm, constituting a non-redundant full-length TE sequence library.

Read Mapping

Raw paired-end reads from 20 asexual and 20 sexual *D. pulex* isolates were first trimmed with quality ($Q > 25$) and length (>60 bp) control using Trimmomatic (Bolger et al. 2014), and were subsequently mapped to the non-redundant TE library using the software BWA with the *mem* default parameters (Li 2013). Fractions of reads mapped to each element were extracted from SAM-formatted alignments.

Characterization of TE Insertions

The algorithm for detecting TE insertions has been described previously (Lee et al. 2014), attempting to identify paired-end reads in which one read is mapped to a TE sequence, while the other is mapped to a unique location in the reference genome. This algorithm takes three files as input: (1) the *D. pulex PA42* reference genome, (2) a pre-built TE sequence library, and (3) paired-end sequences from each of the isolates.

Our goal was to identify all possible TE insertions that are present in each isolate, thus, both full-length and fragmented TEs in *D. pulex* genomes were considered. In this step, we used non-redundant full-length TE sequences as the TE library in a homology search to detect all the fragmented TEs in *TCO* and *PA42* reference genomes using a commonly used tool RepeatMasker (Smit et al. 2004). For overlapping elements identified by the program, the element with the best score was kept. We only considered elements with a length of at least 500 bp in our further analysis. The TE insertion detection algorithm cannot distinguish redundant elements with nearly identical sequences in the end region, therefore, the full-length and fragmented elements were combined and clustered to further remove redundancy using sequencing similarity 95% at both end regions (250 bp for each end) as threshold. Moreover, to make sure the inserted element identified was present in the corresponding isolate and thus reduce the false-positive rate, we took the following steps: first mapped paired-end reads of each isolate to the entire TE library, and then calculated the fraction of each element mapped with reads using coverageBed in BEDTools package (Quinlan and Hall 2010). Only the element with more than 90% of its sequence covered by reads in the isolate was retained in the TE library for the TE insertion detection algorithm that applied to the sequencing data of the corresponding isolate.

The algorithm was then conducted in each isolate, and paired-end reads were simultaneously mapped to the input TE library as well as the *PA42* reference genome to detect TE

insertions. Three types of TE insertions were identified: paired-end reads supporting both forward and reverse directions around the insertion site, only paired-end reads supporting the forward directions, and only paired-end reads supporting the reverse directions. The final TE insertions used in the subsequent analysis include those supported by three or more pairs of reads at both the 5′ and 3′ ends of the TE insertion, or by five or more pairs of reads at either of the 5′ or 3′ ends. This empirical criteria used here may miss some of TE insertions, particularly when the paired-end read coverage is not very high.

Results

Identification and Annotation of TEs

We used an approach combining structure-based and homology-based methods to identify all TEs in two available draft *D. pulex* assembly genomes, *TCO* and *PA42*. In a first step, a structure-based approach identified 1,487 full-length TE sequences, including 1,139 LTR retrotransposons, 131 non-LTR retrotransposons, and 217 DNA transposons (table 1). The defined full-length TEs are limited to those containing a functional protein-coding domain, essential to the transposition structural components. As noted earlier, LTR retrotransposons are the most abundant subclass of TEs in the *D. pulex* genome. Here, 784 of the total 1,139 full-length LTR retrotransposons were derived from *TCO* genome, including all of the 314 intact elements identified in the previous study (Rho et al. 2010). Two probable reasons can explain the enlarged TE library: first, all four protein domains in the *Pol* region (RT, RH, INT, AP) were searched instead of only the RT domain in the previous study, and second, two software programs were used instead of only MGEScan-LTR. For non-LTR retrotransposons, the phylogenetic analysis of RT sequences revealed that *L2* is the most abundant clade among the four identified clades (*Jockey, L1, L2*, and *LOA*) in *D. pulex*. For DNA transposons, *isl2eu* and *helitron* are the most abundant of the six observed clades, whereas only a few *hat* and *P elements* were found. In a second step, we first clustered the 1,487 full-length TEs into a non-redundant TE library (similarity > 98%) containing 1,461 sequences, and then used these TE sequences in a homology search to annotate additional fragmented TEs with lengths longer than 500 bp. In total, the homology-based approach annotated 27,849 fragmented TEs, including 23,136 LTR retrotransposons, 2,066 non-LTR retrotransposons, and 2,647 DNA transposons.

The full-length and fragmented TEs identified in the *TCO* and *PA42* genome assemblies constitute an enlarged and comprehensive TE library, making up 12.9% of the *PA42* reference genome, with the majority being in the subclass of LTR retrotransposons (10.6%). The expanded full-length TE-sequence library significantly increases the estimated copies of TEs (17,303) in the *TCO* genome relative to that derived in the previous study (7,821) (Colbourne et al. 2011). We noted that there are almost twice the numbers of full-length (and fragmented) TE sequences in *TCO* as in *PA42*. It is not surprising that *TCO* contains more TE elements because *TCO* is a unique geographically isolated clade with a larger (with an assembly of ~195 Mbp) and more complex genome than *PA42* (with an assembly of ~156 Mbp). The *TCO* genome is derived from a population estimated to have a substantially reduced effective population size relative to the more broadly distributed Midwest U.S population from which *PA42* genome is derived, and effective population size is thought to strongly affect TE proliferation in the genome (Ye et al. 2016). Systematic comparison of other genomic-architectural differences between *TCO* and *PA42* is described in Ye et al. (2016).

TE-Derived Sequences in the *D. pulex* Genome

On average, ~90% of total sequencing reads from each *D. pulex* isolate can be aligned to the *PA42* reference genome, and ~18% of the whole-genome sequencing reads from each of the 40 *D. pulex* isolates can be aligned to the TE library consisting of 1,461 full-length TE sequences, of which on average ~16% were annotated as LTR retrotransposons, and ~1% each as non-LTR and DNA elements, respectively (supplementary table S4, Supplementary Material online). The fractions of reads mapped to each TE sequence in these genomes are listed in supplementary table S5, Supplementary Material online. The estimated TE fraction from read mapping (~18%) is higher than that from the annotation of the *PA42* reference assembly (~12.9%, table 1), indicating that the draft assembly may still be incomplete and missing some TE copies. Although the total fraction of TE-derived DNA sequences varies among the 40 *D. pulex* isolates, ranging from ~17 to 28% (supplementary table S4, Supplementary Material online), the estimated mean in asexual isolates (~21%) is not significantly different from that in sexual isolates (~22%, *t*-test, $P = 0.21$). Notably, of the 1,461 full-length TE sequences in the library, 6 LTR retrotransposons (including 1 *Bel*, 1 *Copia*, and 4 *Gypsy* TEs) identified from the *TCO* reference genome have almost no hits across all 40 Midwest *D. pulex* isolates (supplementary table S5, Supplementary Material online), implying that these elements may be restricted to *TCO*. In contrast, all of the TEs found in *PA42* have read hits, implying the close relationship between *PA42* and the evaluated isolates.

Polymorphisms of TE Insertion Sites in Asexual and Sexual *D. pulex* Isolates

We used a computational method to detect TE insertion sites based on the mapped paired-end reads (see Materials and Methods for details). This method takes as input a reference genome, a database of TE sequences, and a large set of paired-end sequencing data, and then reports a list of read-pair clusters. A TE insertion site is recognized if one of the

Table 1

Summary of TEs Identified in *TCO* and *PA42* Reference Genome

Class	Subclass	Clades	No. of TE[a]		Fraction (%)[b]	Ave. Length (bp)	
			TCO	*PA42*		TE[c]	LTRs
Retro-transposons	LTR	Bel/Pao	157(2,631)	61(1,358)	1.76	9,602	527
		Copia	267(5,267)	133(3,494)	4.15	8,687	565
		Dirs	23(887)	20(741)	0.93	11,802	540
		Gypsy	337(5,877)	141(2,881)	3.72	9,557	441
		subtotal	784(14,662)	355(8,474)	10.56	9,351	505
	Non-LTR	Loa	9(192)	15(175)	0.07	5,775	
		L1	9(115)	3(57)	0.46	4,615	
		L2	28(455)	27(343)	0.20	5,286	
		Jockey	22(404)	18(325)	0.46	5,703	
		subtotal	68(1,166)	63(900)	1.19	5,439	
Subtotal			852(15,828)	418(9,374)	11.75	8,989	
DNA-transposons	TIR	Isl2eu	48(312)	50(275)	0.30	3,484	
		Mariner	11(196)	19(150)	0.01	3,601	
		Harbinger	5(480)	13(506)	0.28	3,854	
		Hat	1(15)	1(7)	0.02	4,570	
		P element	3(11)	3(15)	0.19	3,897	
	Helitrons	Helitron	37(461)	26(219)	0.37	4,106	
Subtotal			105(1,475)	112(1,172)	1.18	3,738	
Total			957(17,303)	530(10,546)	12.93	8,244	

Note.—The full-length TEs are the elements with at least one protein coding domain for its transposition, while fragmental TEs are the elements obtained by RepeatMasker with >500 bp.

[a]The numbers of full-length TEs (fragmental TEs are listed within the bracket).

[b]The fraction of TE sequences in *PA42* reference genome.

[c]The average length of full-length TEs or LTRs.

paired-end reads is mapped to a unique location on the *PA42* reference assembly while the other one is mapped to either end of a TE sequence. The purpose here was to search for the insertion sites of all kinds of TEs in *D. pulex* isolates. We measured the insertion polymorphism of a TE insertion site (i.e., a specific location of a specific TE) in sexual/asexual isolates by the number of clones containing this TE insertion in each of the two groups, respectively: if a TE insertion is identified at a locus in single isolate, it is defined as singleton; if a TE insertion is observed in some (at least one) but not all isolates in the same (sexual or asexual) group, the insertion is considered to be polymorphic in this group; if a TE insertion is observed in all 20 isolates in the sample, the insertion is considered to be fixed. A total of 58,976 TE insertion sites were identified in 40 *D. pulex* isolates by applying this method (fig. 1). These elements inserted at these sites represent a broad taxonomic range, including 48,103 LTR-retrotransposons, 3,749 non-LTR retrotransposons, and 7,124 DNA transposons. The TE insertions in each group are mainly singletons (22,119 in asexuals vs. 20,418 in asexuals), which constitute more than half of the TE insertions in both populations (~51% in asexuals and ~54% in sexuals), implying that most TEs are in a dynamic process of expanding and diminishing. The distribution of TEs in asexual and sexual populations is significantly different (Wilcoxon signed rank test, $P = 0.0006$). Fewer TE insertions were found in the 20 asexual isolates (37,805) than in the 20 sexual isolates (42,920), with 21,749 common insertions in both samples, 16,056 and 21,171 unique insertions in asexual and sexual groups, respectively (fig. 2*A*). More fixed TE insertions were identified in the asexual group (2,405) than in the sexual group (1,387), although the fixed TE insertions are only a small portion of the total TE insertions (~6% in asexuals, and ~3% in sexuals).

Nevertheless, the numbers of TE insertions identified in each isolate by the TE identification algorithm is sensitive to the depth of read coverage and insert size (Lee et al. 2014), as higher coverage and longer insert size help to cover more TE insertions comprehensively (see supplementary table S6, Supplementary Material online, summarized all the TE insertions in each of the 40 *D. pulex* isolates). Here, we further focused on eight asexual isolates and eight sexual isolates (table 2) with almost the same average coverage (~33X) and insert size (~300 bp) in an attempt to avoid the bias introduced by the analytical pipeline in the comparison of TE insertion abundance between asexuals and sexuals. The number of TE insertions in each of the 16 focused isolates ranges from ~7,400 to ~8,500, with an average number of ~8,000 insertions. Although both groups contain almost the same number of TE insertions per isolate (8,174 in asexuals vs. 8,015 in sexuals on average; *t*-test, $P = 0.30$), the variation of the number of TE insertions among different isolates in the sexual group is significantly higher than that in the asexual

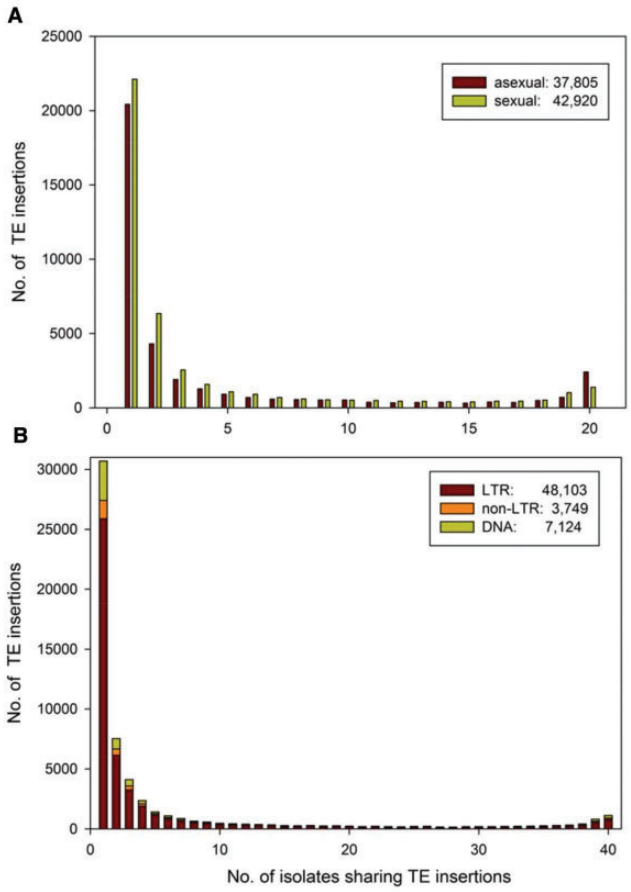

Fig. 1.—The distribution of TE insertions present in one or shared by different isolates. (*A*) 20 sexual and 20 asexual isolates. The numbers listed in the inserted box are the total numbers of different insertion sites occupied with TEs in the asexual/sexual groups. The distribution of TEs in asexual and sexual populations is significantly different (Wilcoxon signed rank test, $P = 0.0006$). (*B*) Summarized numbers of TE insertions of three different subclasses in all 40 *Daphnia pulex* isolates.

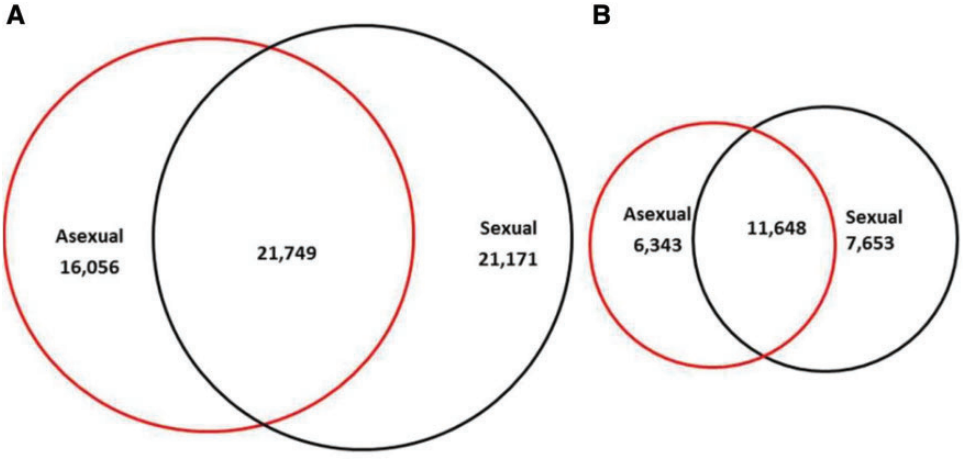

Fig. 2.—Number of TE insertions present in the asexual population (pink), the sexual population (blue), and shared by both population (overlap). (*A*) TE insertions observed in the survey of all 20 sexual and 20 asexual isolates. (*B*) TE insertions observed in the focused study of eight asexual and eight sexual isolates with the same coverage and insert size.

Table 2
Summary of TE Insertion Sites for Each of the Focused Eight Asexual and Eight Sexual *Daphnia pulex* Isolates

Isolates	Coverage	Insert Size (bp)	No. of TE Insertion Sites			
			LTR	Non-LTR	DNA	Total
Asexual						
5w-2	33	300	6,052	646	1,397	8,095
jol-2	33	300	6,411	702	1,334	8,447
lis-3	34	300	6,011	677	1,344	8,032
lyt-1	34	300	6,087	663	1,380	8,130
morg-5	34	300	6,059	675	1,344	8,078
sed-2	33	300	6,197	682	1,367	8,246
smi-1	33	300	6,161	707	1,368	8,236
tro-3	32	300	6,118	681	1,330	8,129
Average			6,137	679	1,358	8,174
Standard deviation			126	20	24	133
Sexual						
eb1	34	300	6,143	698	1,350	8,191
ksp10	32	290	5,840	684	1,258	7,782
lp8b6	32	300	5,994	663	1,250	7,907
nfl3	33	300	6,249	697	1,359	8,305
pa32	32	300	6,406	722	1,416	8,544
povi4	30	300	6,338	663	1,339	8,340
tex9	33	300	5,737	645	1,237	7,619
bw101	32	300	5,550	630	1,254	7,434
Ave			6,032	675	1,308	8,015
Standard deviation			305	31	66	389

NOTE.—Each insertion site is supported by at least three paired-end reads in each end, or five in either end. The focused 16 isolates with almost the same coverage and insert size.

group (table 2, *F*-test to compare two variances, $P = 0.01$). A comparison of the numbers of unique TE insertions in the focused 16 sexual/asexual isolates also shows the same tendency as in the 40 isolates (fig. 2*B*): fewer total TE insertions and more fixed TE insertions were found in the eight asexual isolates (17,991 and 3,721) than in the eight sexual isolates (19,301 and 3,191).

Asexual/Sexual-Specific Fixed TE Insertion Sites

When paired-end shotgun sequencing data have relatively low coverage or short insert size in isolates, some TE insertions may be missed by our TE insertion detection method as they are spanned by few or no read pairs. The average coverage of the 40 isolates used here is ~30X (ranging from 20X to 41X) and may not be sufficient to detect all TE insertions. Therefore, we relaxed the condition to define a TE insertion site as asexual/sexual-specific fixed if it was identified in at least 16 (instead of all 20) isolates in one group but absent in all 20 isolates from the other group. Using this criterion, a comparison of the distribution of polymorphic TE insertion sites in sexual and asexual isolates clearly shows that some TE insertions are fixed in asexual isolates while absent in sexual isolates (fig. 3 and supplementary table S7, Supplementary Material online). After manual inspection, we found nine potentially asexual-specific fixed TE insertion sites, contributed by eight LTR retrotransposons and one DNA transposons, but did not find any sexual-specific fixed TE insertions (table 3). Using genetic markers located on the scaffolds containing these TE insertion sites (Ye et al. 2016), we derived the chromosomal location of these asexual-specific fixed TE insertions: eight are located on chromosome 9, while the remaining one is located on chromosome 7.

We further evaluated whether the nine asexual-specific fixed TE insertion sites identified in our *in silico* analysis are associated with any genes related to the origin of asexuality by searching for genes within 10-kb flanking regions upstream and downstream from these TE insertion sites. In the total of 14 protein-coding genes located in flanking regions of the 9 TE insertion sites (supplementary table S8, Supplementary Material online), we found genes related to metabolic processes, components of membrane or extracellular, but no gene with obvious relevance to meiosis. Thus, more evidence is required to determine whether the participation of TE insertions have impacted the process of converting meiotically reproducing isolates into asexuals.

As noted earlier, previous studies showed that some genomic regions from *D. pulicaria* have been transferred into the obligately asexual isolates of *D. pulex* via historical

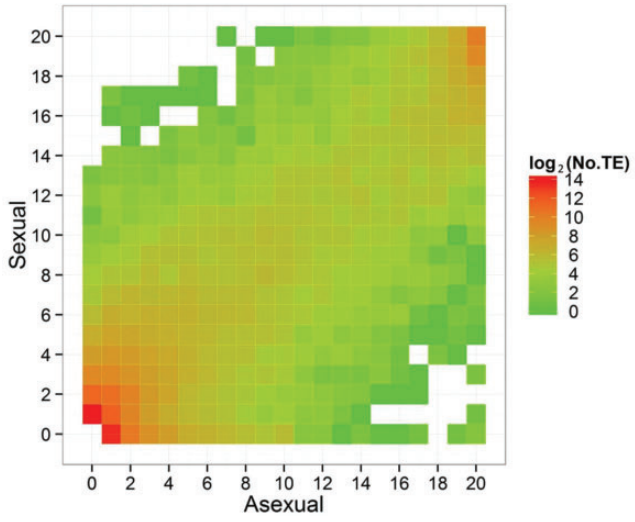

Fig. 3.—Occupancy distribution of polymorphic TE insertions at sites in 20 asexual and 20 sexual isolates. The numbers of polymorphic TE insertions are displayed in a heatmap illustration according to the numbers of asexual (*x*-axis) and sexual (*y*-axis) insertions in genotypes within which a TE insertion at a specific site is recognized. White in the figure means absence of corresponding numbers of TE insertion in both groups.

Table 3

Summary of Nine Asexual-Specific Fixed TE Insertion Sites

Name	Subclass/Clade	TE Insertion Sites Corresponding to the Location at *PA42* Assembly				
		Asex[a]	Scaffold	Start	End	Chrom[b]
TCO_107_283317_283825	LTR/Copia	20	103	100,226	101,237	9
TCO_219_144768_145894	LTR/Gypsy	16	109	363,334	363,798	9
TCO_13_1370849_1371446	DNA/harbinger	20	11	642,932	644,061	9
PA42_443_25059_26584	LTR/Copia	17	4	252,200	255,100	9
TCO_14_637384_638626	LTR/Copia	17	129	189,400	189,540	9
PA42_129_219941_220452	LTR/Copia	20	129	208,420	208,870	9
TCO_107_273836_274372	LTR/Bel	19	31	45,378	46,322	7
TCO_107_319799_320337	LTR/Copia	20	31	572,621	573,422	9
PA42_1454_1_1708	LTR/Gypsy	19	90	16,345	16,816	9

Note.—The left part includes the source information of the inserted TE, and the right lists the information on the inserted site in the *PA42* genome. Taking TCO_107_283317_283825 as example, it is a copia LTR retrotransposon, the TE sequence originally derived from the scaffold 107 of *TCO* assembly, with 5′ and 3′ boundaries 283,317 and 283,825 All eight asexual isolates have this insertion, corresponding to scaffold 103 of the *PA42* assembly, almost from ~100,226–101,237 (this is randomly picked up in one isolate, and the insertion location searched by our TE-insertion algorithm is slightly different in each isolate), while scaffold 103 is located on chromosome 9. Each of these TE insertions is also present in some *D. pulicaria* isolate.

[a]Number of clones in the asexual group containing the TE insertion which is also absent from all sexual clones.

[b]Chromosome the insertion site is located on.

introgression, with perhaps the entire chromosomes 8 and 9 in asexual *D. pulex* lineages originating from *D. pulicaria* (Tucker et al. 2013; Xu et al. 2013). To investigate whether these asexual-specific fixed TE insertions were inherited directly from *D. pulicaria* or occurred within the asexual isolates after introgression, we also searched for TE insertions with the available sequencing data for 10 *D. pulicaria* isolates (see supplementary table S1, Supplementary Material online, for details). It is possible that TE insertions inherited from a unique *D. pulicaria* ancestor that gave rise to asexuality was simply subsequently inherited by all descendant asexual isolates. Therefore, an asexual-specific TE insertion site was recognized as being potentially inherited from *D. pulicaria* if it was identified in at least one *D. pulicaria* isolate at an orthologous site. We observed that all nine fixed asexual-specific TE insertions, including a TE insertion site located on chromosome 7, were also identified in *D. pulicaria* isolates (table 3).

In this study, we used 20 asexual *D. pulex* isolates from two different sources: 10 normal isolates from Tucker et al (2013) and 10 "urban" isolates from Xu et al (2015). A principle component analysis using SNPs characterized in the latter study showed that the normal and "urban" asexual *D. pulex* isolates are clustered separately (see fig. 1 in Xu et al. 2015). The number of TE insertions identified in at

least 10 or more isolates in one group but absent in the other group are summarized in supplementary table S9, Supplementary Material online. With this somewhat relaxed criteria, 61 additional asexual-specific polymorphic TE insertion sites were found, among which 11 and 26 TE insertion sites were only identified in normal and urban isolates, respectively, indicating that asexual *D. pulex* isolates of normal and urban sources are extensively diverse in term of the TE distributions. We also identified 13 sexual-specific polymorphic TE insertion sites. Notably, 28 out of 61 asexual-specific polymorphic TE insertion sites can be identified in at least one *D. pulicaria* isolate, and furthermore, 17 out of the 28 insertion sites are located on asexuality-associated chromosomes 5, 8, 9, and 10. In contrast, only one (out of 13) sexual-specific polymorphic TE insertion site has an orthologous insertion in *D. pulicaria* isolates.

Discussion

We have presented the first systematic comparison of the fractions and distributions of all three types of TEs between asexual/sexual *D. pulex* isolates, providing a more comprehensive picture of polymorphisms of TE insertion sites in *D. pulex* than in previous reports. To achieve this, we first compiled a TE sequence library in two *D. pulex* reference genomes by using multiple *de novo* searching methods, expanding by almost threefold the TE library from previous studies of the *D. pulex* genome *TCO* (Colbourne et al. 2011). Using this TE library, we then estimated the fraction of TE-derived DNA sequences in the genomes of multiple *D. pulex*, and observed that there is no significant difference in the content of TE sequences between 20 asexual and 20 sexual *D. pulex* isolates. Moreover, results from the comparison of eight asexual and eight sexual *D. pulex* isolates with similar coverage and insert size indicated no difference in numbers of TE insertions between two groups. These results are consistent with a recent study in different asexual arthropods (Bast et al. 2016), which observed no evidence of increased TE load in genomes of asexual *D. pulex* isolates, although this previous study only used four of the isolates included in the current more comprehensive study.

TE insertions are quite polymorphic in different *D. pulex* isolates as in populations of many other species (Lisch 2013; Barrón et al. 2014; Joly-Lopez and Bureau 2014). For instance, a previous study conducted in a Portuguese population of *D. melanogaster* showed that most TE insertions occur only in a few individuals rather than in the whole population (Kofler et al. 2012). We also observed that TE insertions in *D. pulex* are often present as singletons (~52%) rather than being present in most isolates (fig. 1), consistent with the basic viewpoint from population genetics: persistence of a TE is possible without fixation if the insertion of TE is frequent enough to allow the successful generation of a new copy before being eliminated by selection (Charlesworth and Charlesworth 1983; Langley et al. 1983; Charlesworth and Barton 2004). In the future, we plan to characterize the locations of TE insertions in different *D. pulex* genomic regions (e.g., centromeric or gene rich), as soon as a more complete genome assembly including the chromosomal locations of contigs becomes available.

Although no significant difference in the average TE contents was observed between asexual/sexual groups, we observed more polymorphisms of TE insertions in the sexual than in the asexual group (fig. 1). Two earlier studies (Schaack, Choi, et al. 2010a; Schaack, Pritham, et al. 2010b) compared TE loads based on six different families of DNA transposons in asexual/sexual *D. pulex* populations, and showed that sexuals exhibit higher insertion polymorphisms among lineages, consistent with the conclusion of this study. This was explained by the fact that recombination provides a mechanism by which the rates of both TE gain and loss can be accelerated. In this study, we also found more fixed TE insertions in asexual isolates than in sexual isolates, with some asexual-specific TE insertions being derived from *D. pulicaria*, indicating that all observed asexuals are likely to have inherited, in an initial hybridization event, particular chromosomes from *D. pulicaria* carrying a specific load of TEs. However, previous studies also found that sexuals have more active DNA transposons than asexuals by using transposon display, whereas we observed no significant difference in total TE load (including DNA transposons) between asexual and sexual isolates. The discordance of this conclusion can be explained by the fact that different studies focused on different aspect of TEs: the previous studies focused on the dynamics of TEs and conducted experiments on only six families of potentially active DNA transposons, whereas here we used the whole-genome sequencing data to compare the difference of all potential TEs in the entire genome without distinguishing autonomous from non-autonomous elements. More RNA-seq data or experiments based on transposon display may help to further elucidate the difference of both autonomous from non-autonomous TEs across the entire genome between asexual/sexual populations.

Different complex phenomena related to TE abundance between asexual/sexual isolates have arisen in different species. For example, for evening primroses, almost twofold variation in genome size was observed in asexual/sexual isolates, but TEs seem not to play a major role in the genome size evolution (Ågren et al. 2015). For *D. pulex*, we observed no significant difference in the TE abundance between asexual and sexual lineages except for some asexual-specific fixed TE insertions. This can be explained by some unique features of the *D. pulex* system: first, the spread of asexual *D. pulex* lineages is very recent, perhaps only a few thousand years (Tucker et al. 2013), and such a short time scale may be insufficient to evolve substantially different TE contents between asexual and sexual *D. pulex* lineages. Second, males produced by asexuals can produce haploid gametes that permit backcrossing with sexual isolates, thereby facilitating transmission

of asexuality elements to sexuals (Xu et al. 2013), and further narrowing the difference in TE abundance between asexual/sexual *D. pulex* lineages. Third, recent evidence demonstrated the origination of asexual *D. pulex* lineages via hybridization between sexual *D. pulex* and *D. pulicaria* lineages, indicating asexual *D. pulex* lineages inherited TEs from both ancestors. Once asexual *D. pulex* lineages inherited different TEs from *D. pulicaria*, abandonment of sex presumably has made it hard to eliminate these TE insertions. This can partially explain why we observed no sexual-specific fixed TE insertions while the identified asexual-specific fixed TE insertions are located on chromosome 9 (except one located on chromosome 7) that was demonstrated to be associated with obligate-asexuality related markers and inherited from *D. pulicaria*. Inheritance of TEs from *D. pulicaria* chromosomes may be an important reason why asexual isolates contain some different TE insertions with sexual isolates.

We developed a pipeline for detecting polymorphic TE insertions between asexual and sexual populations that uses paired-end sequencing data. Although this method is particularly useful when the sequencing depth of each clone is not sufficiently high to confidently recover all TEs in their *de novo* assembly, it still suffers two limitations. First, it relies on a preassembled TE library, including only TEs identified in the reference genomes. Thus, some novel elements in natural *D. pulex* lineages might be missing, although we identified all potential TEs within two available reference genomes. For example, using the PCR method, Eads et al. (2012) identified a unique TE insertion located on chromosome 8 near a meiotic cohesion-related gene *Rec8* in all evaluated obligately asexual *D. pulex* but absent in all sexuals. This TE sequence was not included in our TE library because it is absent from the sexual *D. pulex* reference genomes (*PA42* and *TCO*). For similar reasons, we cannot recover all of the TE sequences present in the asexual *D. pulex* genomes with a sexual reference and may miss some asexual-specific TE insertions. However, the bias introduced by using a sexual reference genome should be small for two reasons: (1) asexual *D. pulex* lineages are very young, and (2) *Daphnia* are diploids, and only one of the two chromosomes 8 and 9 of the asexual *D. pulex* lineages are derived from *D. pulicaria*, which as a very close species to *D. pulex* will not have a greatly divergent genome from sexual *D. pulex*. Second, our method is also sensitive to the depth of read coverage and insert size, with increasing coverage of sequencing data enabling the identification of more TE insertion sites (Lee et al. 2014). However, to control for this, we used the 16 isolates with almost the same coverage depth and insertions size to conduct comparisons of TE abundance between asexual and sexual *D. pulex* genomes. Therefore, although having two limitations, our algorithm is still effective for conducting the comparison of TEs between asexual and sexual *D. pulex* groups.

To further confirm the effectiveness of our algorithm, we added the TE sequence found by Eads et al. (2012) to our TE library (downloaded from NCBI), and determined if the pipeline can identify the insertion of this TE in the lineages used in this study. We found that the TE insertion on chromosome 8 found by Eads et al. (2012) is confirmed in all the 10 isolates from Tucker et al. (2013) by this algorithm, demonstrating that our pipeline can efficiently identify TE insertion sites if the relevant TE sequences are included in the TE library. Interestingly, we also identified additional 5 insertion sites of this TE sequence in asexual isolates but absent in all 20 sexual isolates, among which 2 and 3 are identified in all the 10 isolates from Tucker et al (2013) and Xu et al (2015), respectively. However, there is no fixed asexual-specific insertion of this TE sequence. This suggested that our *in silico* searching method can help confirm the findings of previous experimental studies and predict more candidates of TE insertions not identified by experiments. Notably, the difference between normal and urban asexual *D. pulex* isolates is that the latter carry diagnostic *D. pulicaria* alleles in addition to those on chromosomes 8 and 9 (Hebert and Crease 1983; Xu et al. 2013). This is likely the reason why the TE insertion found by Eads et al. (2012) was not identified in the isolates from Xu et al. (2015) and why this TE sequence is distributed in normal and urban isolates at distinct locations.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We thank Heewook Lee and Wazim Mohammed Ismail for providing the TE-insertion detection software; Sen Xu, Hongan Long, and Tom Doak for providing helpful discussions; Kimberly Young for maintaining *D. pulex* clones; and Joseph Shaw and Craig Jackson for providing *D. pulicaria* sequencing data. We also thank two anonymous reviewers for their valuable comments. This project was supported by National Science Foundation (DBI-1262588 to H. Tang) and National Institutes of Health (GM101672 to M. Lynch).

Literature Cited

Ågren JA, Greiner S, Johnson MT, Wright SI. 2015. No evidence that sex and transposable elements drive genome size variation in evening primroses. Evolution 69:1053–1062.

Arkhipova I, Meselson M. 2005. Deleterious transposable elements and the extinction of asexuals. Bioessays 27:76–85.

Arkhipova I, Meselson M. 2000. Transposable elements in sexual and ancient asexual taxa. Proc Natl Acad Sci U S A. 97:14473–14477.

Barrón MG, Fiston-Lavier A-S, Petrov DA, González J. 2014. Population genomics of transposable elements in *Drosophila*. Annu Rev Genet. 48:561–581.

Bast J, et al. 2016. No accumulation of transposable elements in asexual arthropods. Mol Biol Evol. 33:697–706.

Biémont C. 2010. A brief history of the status of transposable elements: from junk DNA to major players in evolution. Genetics 186:1085–1093.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120..
Charlesworth B. 1994. Evolution in age-structured populations. Cambridge: Cambridge University Press.
Charlesworth B, Barton N. 2004. Genome size: does bigger mean worse? Curr Biol. 14:R233–R235.
Charlesworth B, Charlesworth D. 1983. The population dynamics of transposable elements. Genet. Res. 42:1–27.
Charlesworth B, Langley C. 1986. The evolution of self-regulated transposition of transposable elements. Genetics 112:359–383.
Colbourne JK, et al. 2011. The ecoresponsive genome of *Daphnia pulex*. Science 331:555–561.
De La Chaux N, Tsuchimatsu T, Shimizu KK, Wagner A. 2012. The predominantly selfing plant *Arabidopsis thaliana* experienced a recent reduction in transposable element abundance compared to its outcrossing relative *Arabidopsis lyrata*. Mob DNA 3:2.
de Souza FS, Franchini LF, Rubinstein M. 2013. Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong? Mol Biol Evol. 30:1239–1251.
Debortoli N, et al. 2016. Genetic exchange among bdelloid rotifers is more likely due to horizontal gene transfer than to meiotic sex. Curr Biol. 26:723–732.
Dolgin ES, Charlesworth B. 2006. The fate of transposable elements in asexual populations. Genetics 174:817–827.
Duret L, Marais G, Biémont C. 2000. Transposons but not retrotransposons are located preferentially in regions of high recombination rate in *Caenorhabditis elegans*. Genetics 156:1661–1669.
Eads BD, Tsuchiya D, Andrews J, Lynch M, Zolan ME. 2012. The spread of a transposon insertion in *Rec8* is associated with obligate asexuality in *Daphnia*. Proc Natl Acad Sci U S A. 109:858–863.
Eickbush T, Malik H. 2002. Origins and evolution of retrotransposons. Mobile DNA II 2:1111–1144.
Ellinghaus D, Kurtz S, Willhoeft U. 2008. *LTRharvest*, an efficient and flexible software for *de novo* detection of LTR retrotransposons. BMC Bioinformatics 9:18.
Finn RD, et al. 2014. Pfam: the protein families database. Nucleic Acids Res. 42:D222–D2230.
Finnegan DJ. 1992. Transposable elements. Curr Opin Genet Dev. 2:861–867.
Flot J-F, et al. 2013. Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. Nature 500:453–457.
Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152.
González J, Petrov DA. 2009. The adaptive role of transposable elements in the *Drosophila* genome. Gene 448:124–133.
Hebert PDN, Crease T. 1983. Clonal diversity in populations of *Daphnia pulex* reproducing by obligate parthenogenesis. Heredity 51:353–369.
Hickey DA. 1982. Selfish DNA: a sexually-transmitted nuclear parasite. Genetics 101:519–531.
Hurst GD, Werren JH. 2001. The role of selfish genetic elements in eukaryotic evolution. Nat Rev Genet. 2:597–606.
Innes DJ, Hebert PD. 1988. The origin and genetic basis of obligate parthenogenesis in *Daphnia pulex*. Evolution 1024–1035.
Jiang N. 2013a. Computational methods for identification of DNA transposons. Plant Transpos Elem: Methods Protoc 1057:289–304.
Jiang N. 2013b. Overview of repeat annotation and *de novo* repeat identification. Plant Transpo Elem. 1057:275–287.
Joly-Lopez Z, Bureau TE. 2014. Diversity and evolution of transposable elements in *Arabidopsis*. Chromos Res. 22:203–216.
Jurka J, et al. 2005. Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 110:462–467.
Kaminker JS, et al. 2002. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. Genome Biol. 3:RESEARCH0084.
Katoh K, Kuma K-I, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33:511–518.
Kazazian HH. 1998. Mobile elements and disease. Curr Opin Genet Dev. 8:343–350.
Kofler R, Betancourt AJ, Schlotterer C. 2012. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. PLoS Genet. 8:e1002487.
Lander ES, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.
Langley CH, Brookfield JF, Kaplan N. 1983. Transposable elements in Mendelian populations. I. A theory. Genetics 104:457–471.
Le Rouzic A, Boutin TS, Capy P. 2007. Long-term evolution of transposable elements. Proc Natl Acad Sci U S A. 104:19375–19380.
Le Rouzic A, Capy P. 2005. The first steps of transposable elements invasion. Genetics 169:1033–1043.
Lee H, Popodi E, Foster PL, Tang H. 2014. Detection of structural variants involving repetitive regions in the reference genome. J Comput Biol. 21:219–233.
Lerat E. 2010. Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. Heredity 104:520–533.
Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*
Lisch D. 2013. How important are transposons for plant evolution? Nat Rev Genet. 14:49–61.
Lynch M. 2007. The origins of genome architecture. Sunderland: Sinauer Associates.
Lynch M, Seyfert A, Eads B, Williams E. 2008. Localization of the genetic determinants of meiosis suppression in *Daphnia pulex*. Genetics 180:317–327.
Marchler-Bauer A, et al. 2011. CDD: a conserved domain database for the functional annotation of proteins. Nucleic Acids Res. 39:D225–D229.
Olson SA. 2002. Emboss opens up sequence analysis. Brief Bioinformatics 3:87–91.
Paland S, Colbourne JK, Lynch M. 2005. Evolutionary history of contagious asexuality in *Daphnia pulex*. Evolution 59:800–813.
Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 26:1641–1650.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.
Rho M, Choi JH, Kim S, Lynch M, Tang H. 2007. *De novo* identification of LTR retrotransposons in eukaryotic genomes. BMC Genomics 8:90.
Rho M, et al. 2010. LTR retroelements in the genome of *Daphnia pulex*. BMC Genomics 11:425.
Rho M, Tang H. 2009. MGEScan-non-LTR: computational identification and classification of autonomous non-LTR retrotransposons in eukaryotic genomes. Nucleic Acids Res. 37:e143.
SanMiguel P, et al. 1996. Nested retrotransposons in the intergenic regions of the *maize* genome. Science 274:765–768.
Schaack S, Choi E, Lynch M, Pritham EJ. 2010. DNA transposons and the role of recombination in mutation accumulation in *Daphnia pulex*. Genome Biol. 11:R46.
Schaack S, Pritham EJ, Wolf A, Lynch M. 2010. DNA transposon dynamics in populations of *Daphnia pulex* with and without sex. Proc Biol Sci. 277:2381–2387.
Smit A, Hubley R, Green P. 2004. RepeatMasker Open-3.0. 2004. Seattle: Institute for Systems Biology.
Tucker AE, Ackerman MS, Eads BD, Xu S, Lynch M. 2013. Population-genomic insights into the evolutionary origin and fate of obligately asexual *Daphnia pulex*. Proc Natl Acad Sci U S A. 110:15740–15745.

Wicker T, et al. 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 8:973–982.
Wright S, Finnegan D. 2001. Genome evolution: sex and the transposable element. Curr Biol. 11:R296–R299.
Wright SI, Agrawal N, Bureau TE. 2003. Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. Genome Res. 13:1897–1903.
Xu S, Innes DJ, Lynch M, Cristescu ME. 2013. The role of hybridization in the origin and spread of asexuality in *Daphnia*. Mol Ecol. 22:4549–4561.
Xu S, et al. 2015. Hybridization and the origin of contagious asexuality in *Daphnia pulex*. Mol Biol Evol. 32:3215–3225.
Ye Z, S, Xu K, Spitze J, Asselman X, Jiang ME, Pfrender M, et al. 2016. Comparative genomics 17 of the *Daphnia pulex* species complex. (Revised)
Zemojtel T, Kielbasa SM, Arndt PF, Chung HR, Vingron M. 2009. Methylation and deamination of CpGs generate p53-binding sites on a genomic scale. Trends Genet. 25:63–66.
Zeyl C, Bell G, Green DM. 1996. Sex and the spread of retrotransposon Ty3 in experimental populations of *Saccharomyces cerevisiae*. Genetics 143:1567–1577.

Associate editor: Judith Mank