Evolutionary Genomics of Sister Species Differing in Effective Population Sizes and Recombination Rates

Zhiqiang Ye [1,*], Michael E. Pfrender[2], and Michael Lynch[3]

[1]Hubei Key Laboratory of Genetic Regulation and Integrative Biology, School of Life Sciences, Central China Normal University, Wuhan, China

[2]Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana, USA

[3]Biodesign Center for Mechanisms of Evolution, Arizona State University, Tempe, Arizona, USA

*Corresponding author: E-mail: zhiqiang.yes@gmail.com.

Accepted: November 02, 2023

Abstract

Studies of closely related species with known ecological differences provide exceptional opportunities for understanding the genetic mechanisms of evolution. In this study, we compared population-genomics data between *Daphnia pulex* and *Daphnia pulicaria*, two reproductively compatible sister species experiencing ecological speciation, the first largely confined to intermittent ponds and the second to permanent lakes in the same geographic region. *Daphnia pulicaria* has lower genome-wide nucleotide diversity, a smaller effective population size, a higher incidence of private alleles, and a substantially more linkage disequilibrium than *D. pulex*. Positively selected genes in *D. pulicaria* are enriched in potentially aging-related categories such as cellular homeostasis, which may explain the extended life span in *D. pulicaria*. We also found that opsin-related genes, which may mediate photoperiodic responses, are under different selection pressures in these two species. Genes involved in mitochondrial functions, ribosomes, and responses to environmental stimuli are found to be under positive selection in both species. Additionally, we found that the two species have similar average evolutionary rates at the DNA-sequence level, although approximately 160 genes have significantly different rates in the two lineages. Our results provide insights into the physiological traits that differ within this regionally sympatric sister-species pair that occupies unique microhabitats.

Key words: *Daphnia*, hybridization, linkage disequilibrium, local adaptation, population genomics, recombination, sister species.

Significance

Comparisons of closely related species with habitat differentiation provide exceptional opportunities for understanding ecological speciation. This comparative population-genomics analysis for the lake-dwelling *Daphnia pulicaria* and its sister species *Daphnia pulex*, a temporary pond specialist, reveals that positively selected genes in *D. pulicaria* are enriched in aging-related categories, which may be related to the 2-fold increase in life span in *D. pulicaria.* In addition, opsin genes, which likely mediate photoperiodic responses, are under different selection pressures in these two species, which occupy very different light environments. The two species have similar average evolutionary rates at the DNA-sequence level, although approximately 160 genes have significantly different rates in the 2 lineages. These discoveries set the stage for future functional work on the molecular, cellular, and developmental determinants of differentiation in this globally distributed species pair.

Introduction

Comparative population-genomics analyses of related species with known population-genetic and/or ecological differences provide unique opportunities for obtaining insights into the genetic mechanisms of evolution. Animal species subject to such work include fish (Peichel and Marques 2017; Crawford et al. 2020), mice (Bedford and Hoekstra 2015; Wang and Payseur 2017), and dipterans (Bergland et al. 2016; Bradshaw et al. 2018; Rodrigues et al. 2021). Of special interest is the North American species pair of microcrustaceans, *Daphnia pulex* and *Daphnia pulicaria*, each of which is the other's closest relative. The two species are nearly identical morphologically (Brandlova et al. 1972), but whereas *D. pulex* mainly inhabits ephemeral ponds, *D. pulicaria* are generally restricted to permanent lakes. Often, the bodies of water involved are separated by only a few meters. Given the lack of compelling morphological distinctions, species ascertainment was historically based on allozyme markers, with *D. pulex* clones typically being homozygous for the slow allele (SS) and *D. pulicaria* homozygous for the fast allele (FF) at the lactate dehydrogenase (*Ldh*) locus (Hebert et al. 1989; Crease et al. 2011). More recently, >20,000 species-specific single-nucleotide markers have been found to distinguish the 2 species across the genome (Ye et al. 2022).

The cyclically parthenogenetic lifestyle of *Daphnia* results in a strong correlation between habitat specialization and the degree of recombination in the species pair. Throughout the bulk of the growing season, *Daphnia* populations reproduce by ameiotic parthenogenesis, but can switch to sexual reproduction and resting-egg production following the environmental induction of male production. (Although obligately asexual forms of the two species exist, here we focus entirely on populations capable of sexual reproduction.) As occupants of temporary ponds, the *D. pulex* populations referred to in this study typically reproduce asexually every 3–5 generations, and consequently experience as much if not more recombination per generation than obligately sexual species such as *Drosophila* with smaller numbers of chromosomes and the lack of recombination in males (Lynch et al. 2017). Genotype frequencies in these populations typically adhere very closely to Hardy–Weinberg (HW) expectations (Maruki et al. 2022).

On the other hand, as residents of permanent lakes, *D. pulicaria* are capable of foregoing sexual reproduction for multiple years, which in some cases could extend over tens to hundreds of generations. This often leads to substantial deviations from HW expectations within loci and elevated linkage disequilibrium (LD) among loci, including extreme cases of transient near-fixation of single clones within individual populations (Lynch 1987; Allen and Lynch 2012). *Daphnia pulicaria* typically live in stratified lakes containing zooplanktivorous fish, where they engage in a defensive behavior known as diel vertical migration, allowing them to migrate to deeper, darker waters during the day and returning to surface waters at night to feed (Gliwicz 1986; Glaholt et al. 2016). Although ages at first reproduction are similar in the two species, life spans in *D. pulicaria* are double those in *D. pulex* (Dudycha 2003), likely associated with the greater opportunities for longer life in lake environments.

Whereas the average level of sequence divergence between the two species is just approximately 1% (in excess of within-species variation; Tucker et al. 2013; Ye et al. 2019; Ye, Jiang, et al. 2021), the opportunities for gene flow may be limited by prezygotic isolation in nature, as some evidence suggests that sexual reproduction is induced by long-day photoperiods in *D. pulex* but by short-day photoperiods in *D. pulicaria* (Deng 1997). Nonetheless, the two species are physically capable of hybridization in the laboratory (Heier and Dudycha 2009). F_1 hybrids in nature are typically obligate asexuals (Crease et al. 1990; Hebert et al. 1993; Xu et al. 2013, 2015), implying a breakdown in meiotic processes in hybrid backgrounds and minimal opportunities for backcrossing (Ye, Jiang, et al. 2021; Ye, Williams, et al. 2021). In addition, an approximately 1.2-Mb region in *D. pulex* is thought be segregating for an introgressed fragment of the *D. pulicaria* genome that confers the inability to produce male offspring (Ye et al. 2019). A recent large-scale population-genomics study suggests that a typical *D. pulex* genome contains a small (<1%) fraction of discernible *D. pulicaria*–derived DNA (Maruki et al. 2022). On the other hand, analyses involving the mitochondrial genome suggest a more reticulate evolutionary pattern, with multiple introductions of *D. pulex* mtDNA into *D. pulicaria* (Cristescu et al. 2012; Ye et al. 2022).

In summary, the existing data suggest that the closely related sister taxa *D. pulex* and *D. pulicaria* experience limited gene flow despite their effectively sympatric distributions, presumably reinforced by ecological differentiation. Here, we carry out population-genomics analyses for multiple populations of the lake-dwelling *D. pulicaria*, taking advantage of prior work on >800 sequenced genomes of North American *D. pulex* distributed over 10 populations (Lynch et al. 2017, 2022; Maruki et al. 2022). Comparisons of the two enable us to quantify the magnitude and patterns of nucleotide diversity and recombination rate differences. In addition, data at the whole-genome level provide further insight into the degree to which genomic regions are impacted by hybridization as well as an evaluation of the patterns of differences in the rates of molecular evolution between the two species.

Results

Here, we primarily report on new observations for populations of *D. pulicaria*, integrating the results with those from

Table 1
Summary of Within-population Genetic Variation

Population	π_T	No. of Sites	F_{IS}	SE (F_{IS})	No. of SNPs	N_e
BRA	0.0029	117,123,707	−0.1850	0.0003	1,656,187	304,000
CLO	0.0017	120,113,119	0.0137	0.0003	1,034,526	165,000
TF	0.0031	117,149,728	−0.0340	0.0002	1,877,097	304,000
D. pulicaria	0.0026		−0.1259			258,000
SE	0.0004		0.0317			46,000
D. pulex	0.0078		−0.0168			640,000
SE	0.0004		0.0100			31,000

NOTE.—π_T is the mean nucleotide diversity over all sites; the standard errors of the population estimates (not shown) are all <0.00002. The average inbreeding coefficients, F_{IS}, are based on polymorphisms with ML estimates of MAFs significantly >0.02 at the 5% level. Effective population size estimates, N_e, estimated as $\pi_s/4\mu$, were obtained from diversity at silent sites and restricted intron positions (internally, 8–34 from both ends; Lynch et al. 2017), using the mutation rate estimate of $\mu = 5.69\times10^{-9}$ per site per generation from Keith et al. (2016).

broad surveys recently reported on the population-genomic features of *D. pulex* (Lynch et al. 2017, 2022; Maruki et al. 2022; Ye et al. 2023) throughout the narrative.

Within-Population Genetic Variation

In total, this study encompassed the genomic sequences of 233 clones from 3 *D. pulicaria* populations from permanent lakes in the midwestern United States and southern Canada, 70–84 clones per population (supplementary table S1, Supplementary Material online). On average, 1.2×10^8 nucleotide sites were assayed per population (table 1), 98.7% of which were monomorphic for the same allele across all populations (supplementary table S2, Supplementary Material online). Among polymorphic sites, >96% contain just 2 nucleotides per site, that is, tri- and tetra-allelic sites are rare, as in the case of *D. pulex* (Maruki et al. 2022). The mean genome-wide nucleotide diversity (π) is similar across these *D. pulicaria* populations, averaging 0.0026 (SE = 0.0004), which is about one-third of that within individual *D. pulex* populations (table 1). Although the mean inbreeding coefficient in the CLO sample is positive, it is negative for the other two populations, substantially so for population BRA. This yields an overall average F_{IS} (inbreeding coefficient) estimate that is significantly negative (excess average heterozygosity), in contrast to the situation in *D. pulex* where the average F_{IS} is close to zero, although slightly negative (Ye et al. 2023). For population CLO, 17% of the polymorphic sites exhibit deviations from HW genotype-frequency expectations, 34% of which are in the direction of heterozygote excess. In contrast, for population BRA, 50% of sites deviate from HW expectations, 84% in the direction of heterozygote excess, with these numbers becoming 16% and 65%, respectively, for population TF.

Consistent differences in nucleotide diversity exist among functionally different genomic sites (fig. 1). Nucleotide diversity is highest at silent sites, slightly lower at restricted intron sites (internal positions 8–34 bp from both ends; Lynch et al. 2017), and lowest at amino-acid replacement sites, with intergenic sites having intermediate levels. These observations are consistent with a predominant role for purifying selection in decreasing heterozygosity estimates at functionally more important sites, as found in essentially all species, including *D. pulex* (Maruki et al. 2022; Ye et al. 2023).

Using the average estimates of nucleotide diversity at silent sites within codons and internal intron sites, which we have inferred to be evolving in a nearly neutral fashion (Lynch et al. 2017), and the estimated mutation rate per nucleotide site of $u = 5.69\times10^9$ per generation for *D. pulex* (Keith et al. 2016), we infer the mean effective population size of *D. pulicaria* to be approximately 40% of that for *D. pulex* (table 1). Similarly, Wersebe et al. (2022) estimated that *D. pulicaria* has an effective population size about a third of that of *D. pulex*.

Divergence Among Subpopulations

Owing to the downward bias of the F_{ST} measure of population subdivision with low minor-allele frequencies (MAFs; Alcala and Rosenberg 2017), estimates of this index increase with increasing MAF, leveling off at the point at which the latter exceeds approximately 0.3, with minimal differences among different functional site types (fig. 1). The overall mean F_{ST} for *D. pulicaria* is 0.227 (SE = 0.0001), whereas by focusing just on silent sites with MAF > 0.15, the mean increases to 0.411 (0.001), and for such sites with MAF > 0.30, it further increases to 0.467 (0.0003).

Given the similarity of N_e estimates across populations (table 1), and the similar behavior of all functional categories of sites, an estimate of the average migration rate using equilibrium theory appears to be justified as a first-order approximation. To this end, we used Wright's (1951) island model wherein a metapopulation consists of a large number of demes of size N_e, with each deme experiencing an equivalent migration rate of m per individual (randomly distributed among demes), which yields an expected equilibrium value of F_{ST} of $(1 + 4N_em)^{-1}$. Setting $F_{ST} = 0.467$, the mean effective number of migrants per deme per generation is estimated to be $N_em = 0.29$ for *D. pulicaria*, which with the average N_e in table 1 implies a migration rate of

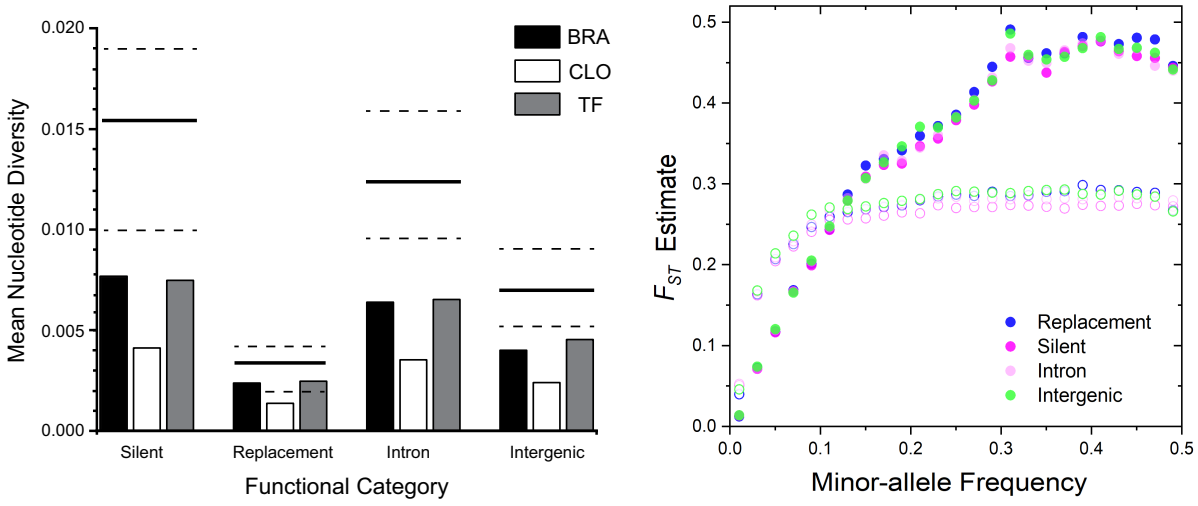

Fig. 1.—Measures of within- and among-population nucleotide variation. (Left) Genome-wide within-population nucleotide diversity estimates at sites in different functional categories. Silent, replacement, and intron sites are, respectively, 4-fold redundant, 0-fold redundant, and restricted intron (internal position 8–34 from both ends; Lynch et al. 2017) sites; intergenic sites are those >500 bp from translation initiation and termination codons. Tri- and tetra-allelic sites are excluded from the calculations but make trivial contributions. The standard errors of the estimates are too small to visualize on the graph ($<10^{-5}$). The solid and dashed lines denote the average, upper, and lower limits of estimates from *Daphnia pulex* populations (from Maruki et al. 2022). (Right) A comparison of the F_{ST} profiles of *Daphnia pulicaria* (closed circles) and *D. pulex* (open circles) as a function of MAFs at the metapopulation level.

$m \sim 1.1 \times 10^{-6}$. In contrast, the unbiased estimated mean F_{ST} for *D. pulex* populations is 0.269 (0.002), with a mean number of migrants per generation of $N_e m \sim 0.59$, and a migration rate of $m = 9.2 \times 10^{-7}$ (Maruki et al. 2022). Thus, both species have per-capita migration rates $\asymp 10^{-6}$, whereas *D. pulex* has essentially double the absolute migration rate of *D. pulicaria*, owing to the elevated N_e. The net effect of the latter is a substantial overall reduction in the level of population subdivision in *D. pulex*.

The elevated level of population subdivision in *D. pulicaria* is not a consequence of increased geographic distance between the sampled populations. The pairwise average F_{ST} (MAF > 0.3) for BRA versus CLO is 0.521, for BRA versus TF is 0.262, and for CLO versus TF is 0.535 (all with SE $\asymp$ 0.005). *Daphnia pulex* exhibits a weak pattern of isolation by distance (Maruki et al. 2022), and for the distances separating the above pairs (551, 787, and 524 km, respectively), the expected F_{ST} for *D. pulex* falls in the narrow range of 0.246–0.270, implying reduced migration per physical distance in *D. pulicaria*.

Finally, some idea of the degree to which the results from these 3 populations reflect the overall features of the entire North American metapopulation of *D. pulicaria* can be acquired by considering an evaluation of the genomic sequences of 14 clones distributed over the North American continent, from east-to-west coasts (Jackson et al. 2021). Because only a single clone was sampled per population in the latter study, the average within-population diversity cannot be estimated. However, the total variation across populations should provide a good estimate of the species-wide diversity, as the average within-population variance will be contained within the overall sample. The total nucleotide diversities are 0.0134, 0.0109, 0.0077, and 0.0045 for synonymous, intron, intergenic, and nonsynonymous sites, respectively, whereas the metapopulation-wide diversities for the 3 populations reported herein are 0.0109, 0.0098, 0.0063, and 0.0033, respectively (in all cases, the SEs of these measures are <0.0001). Thus, the entire species appears to harbor approximately 23%, 11%, 22%, and 36% more diversity for these categorical sites than revealed in our 3 study populations.

Private Alleles Within Populations

An alternative way to consider population subdivision is to evaluate the incidence of significant private alleles (SPAs), defined as polymorphisms that are restricted to single populations, and have an observed frequency that is significantly greater than expected under the null hypothesis of random sampling from a single panmictic population (Maruki et al. 2022). The maximum frequency of a private allele in a set of n populations is equal to p/n, where p is the metapopulation MAF, and rare alleles are unlikely to be designated as SPAs owing to the high probability of emerging by chance sampling of frequency 0.0 in all but one population. In *D. pulex* populations, the fraction of SPAs among all polymorphisms exhibits an exponential decline with increasing metapopulation MAF (Maruki et al. 2022). In striking contrast, in these *D. pulicaria* populations, the incidence of SPAs is quite nonresponsive to the

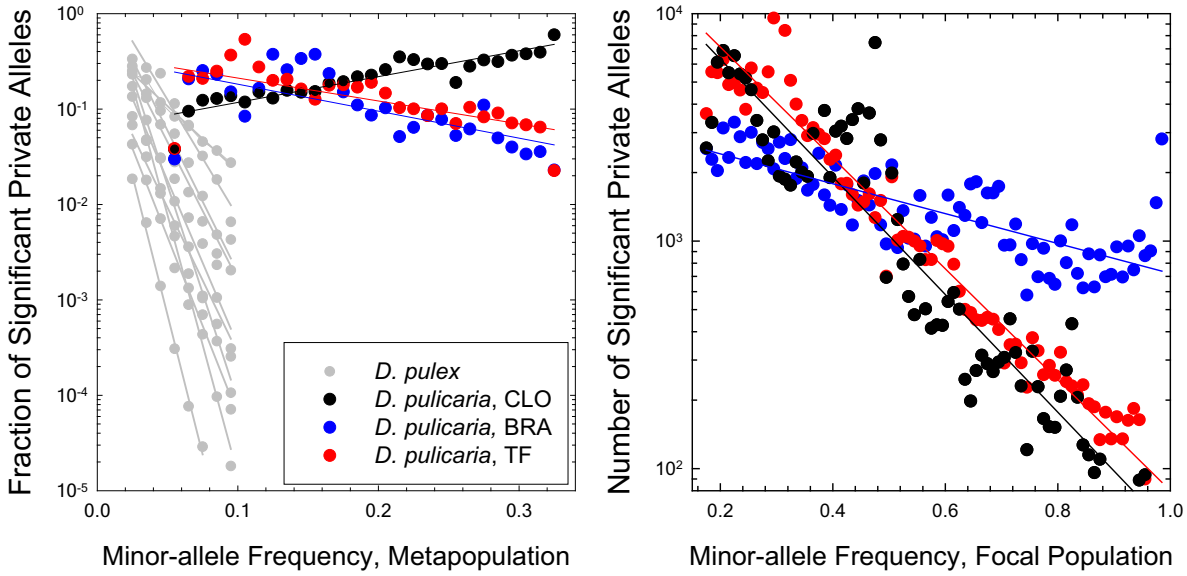

Fig. 2.—Incidence of private alleles in the three *D. pulicaria* populations. (Left) The fraction of private SNPs with the designated metapopulation-wide MAF. Gray data points and regression lines for *D. pulex* are obtained from Maruki et al. (2022). (Right) Counts of SPAs in each focal population as a function of the local MAF.

metapopulation MAF (fig. 2). Between 19,000 and 40,000 SPAs were found in each population, with their incidence declining with increasing frequency in the focal population (fig. 2). Owing to the fact that ten populations of *D. pulex*, but only three of *D. pulicaria* enter these analyses, the maximum frequency of a SPA in the former is about one-third of that in the latter. Nonetheless, it remains clear that *D. pulicaria* populations have exceptionally high incidences of high-frequency SPAs (fig. 2). This condition is almost certainly a consequence of prolonged clonal selection in the absence of recombination, resulting in the periodic emergence of a few clones in each population, which drags all SNPs within such clones to high frequency.

Using the genome-wide estimate of the frequency of private alleles per site as the null hypothesis, for each protein-coding gene in each population, assuming a Poisson distribution for the number of private alleles over entire genic regions (from the start of the 5′ UTR to the termination of the 3′ UTR), we determined the subset of genes containing significantly more private alleles than expected by chance in each population (as in Maruki et al. 2022). After Bonferroni correction, this analysis revealed that 134, 112, and 80 genes in the CLO, TF, and BRA populations, respectively, were enriched for private alleles beyond the background expectation, none of which overlap among populations. On average, these enriched genes had 13.0 (SE = 0.4) private SNPs, with a mean within-population private-allele frequency of 0.345 (0.0003). Thus, although restricted to individual *D. pulicaria* populations, these divergent haplotypes are quite abundant locally.

As noted above, the incidence of private alleles in a metapopulation sample will depend on the number of distinct population samples, declining as the latter increases. Nonetheless, several features of the *D. pulicaria* private-enriched genes are strikingly apparent. First, they are strongly concentrated on chromosome 11 (29%) in BRA and chromosome 4 (30%) in TF (supplementary fig. S1, Supplementary Material online), and such enrichment is not due to increased chromosome length or gene number on these two chromosomes (supplementary table S3, Supplementary Material online). Second, the private-enriched genes are somewhat concentrated in the "islands of divergence" identified in Jackson et al. (2021) for this species, most of which are associated with chromosomal rearrangements with respect to *D. pulex*; 59.8% of the *D. pulicaria* private-enriched genes are located in such regions, which constitute only 36.4% of the *D. pulicaria* genome.

To evaluate whether the *D. pulicaria* private SNPs might result from introgression, we determined the fraction of such SNPs that match known nucleotide variants in the most closely related species, *D. pulex*, as the two are known to hybridize in the field and in the laboratory (Heier and Dudycha 2009; Vergilino et al. 2011; Xu et al. 2015; Moy et al. 2021). Notably, although 8.6% of *D. pulicaria* private SNPs are also found in *D. pulex*, this incidence is twice as enriched (17.3%) in coding regions. Nearly half (46.9%) of private-enriched *D. pulicaria* genes contain private SNPs matching *D. pulex* nucleotides (Maruki et al. 2022), with an average fraction 0.162 (SE = 0.006) of the private SNPs in such genes matching unique variants in *D. pulex*. Thus, there is little doubt that some SNP sharing occurs between

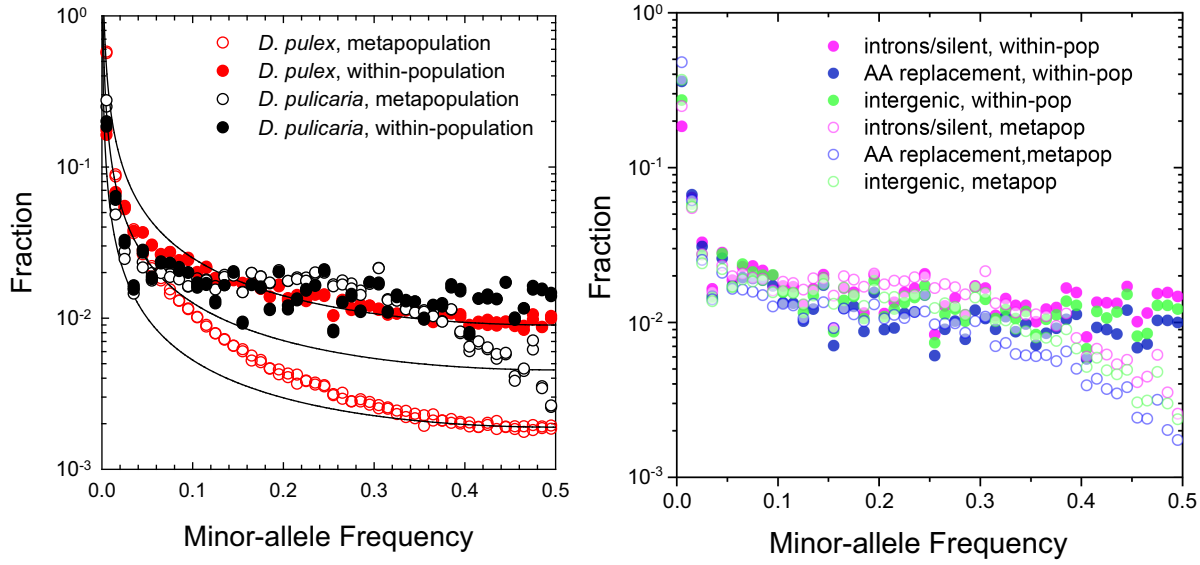

Fig. 3.—SFSs for polymorphisms in *D. pulicaria* and *D. pulex*. (Left) SFSs for putatively neutral sites given for both the average within-population pattern and for the entire metapopulation. For each MAF bin (of width 0.01), there are two points for each species, one for 4-fold redundant (silent) sites in protein-coding genes and the other for restricted intron sites (8–34 from both ends; Lynch et al. 2017). The plotted SFSs are exclusive of the metapopulation-wide monomorphic classes. The curved lines are benchmarks of arbitrary height, showing the scaling $[p(1-p)]^{-1}$, where p is the MAF, expected in the ideal case of neutral sites in a single panmictic, random-mating population at drift-mutation equilibrium. Data for *D. pulex* are obtained from Maruki et al. (2022). (Right) SFSs for functional categorical sites in *D. pulicaria* at the within-population (averaged over all three populations) and metapopulation levels.

these two species, and that such overlap is particularly enriched in protein-coding genes, although the number of genes involved comprises <1% of the genomic total within each *D. pulicaria* population. About 18% of the private *D. pulicaria* SNPs are fixed in *D. pulex* metapopulations, a strong indication of gene flow from *D. pulex*. However, the majority (82%) of the private SNPs are still polymorphic in the *D. pulex* metapopulation, and therefore, we are unable to determine whether they represent shared ancestral polymorphism or result from ongoing gene flow.

Site-Frequency Spectra

A site-frequency spectrum (SFS) describes the genome-wide distribution of SNP frequencies, here described by the folded SFS based on MAFs (with frequencies between 0.0 and 0.5), as we are not yet able to fully assign ancestral states across the genome. Under neutral evolution in a single panmictic population of constant size, the folded SFS is expected to scale as $1/[p(1-p)]$, where p is the MAF, with the elevation of the SFS being defined by the composite parameter $4N_eu$ (Wright 1938; Messer 2009). As was previously found for *D. pulex* (Maruki et al. 2022), the SFSs for 4-fold redundant sites in protein-coding genes and for restricted internal intron sites are essentially indistinguishable, and these are averaged in figure 3.

At the average within-population level, the SFSs for neutral sites are fairly similar in both species (fig. 3, left), and in both cases, there tends to be a deficit of rare alleles at the within-population level, with the SFS otherwise corresponding fairly closely to the neutral expectation. There are, however, three substantial differences in other aspects of SFS behavior between the two species. First, whereas the metapopulation SFS for neutral sites in *D. pulex* declines in a fairly continuous manner, that for *D. pulicaria* exhibits a more discontinuous behavior. In both cases, there is an excess of rare neutral alleles, as expected when migration is limited, as young alleles require time to move and equilibrate in frequency across the metapopulation (De and Durrett 2007; Städler et al. 2009). However, there is a huge bulge in the neutral metapopulation SFS in the MAF range of 0.1–0.4 for *D. pulicaria*, whereas the *D. pulex* metapopulation SFS declines in a smooth fashion. Second, the within-population SFSs for intergenic and amino-acid replacement sites are quite similar to the neutral-site SFS in *D. pulicaria* (fig. 3, right). This contrasts with the substantial depression in the frequencies of all but the lowest MAFs for these categorical sites in the *D. pulex* within-population SFS, by nearly an order of magnitude in the highest frequency sites (Lynch et al. 2017; Maruki et al. 2022). Third, the distinctions between the metapopulation SFSs for the three functional classes of sites in *D. pulicaria* remain far less extensive than observed in *D. pulex*, where all types of sites become increasingly rare with high MAFs (Maruki et al. 2022).

Linkage Disequilibrium

We measured the decline of LD with physical distance along chromosomes with two different methods. First, the

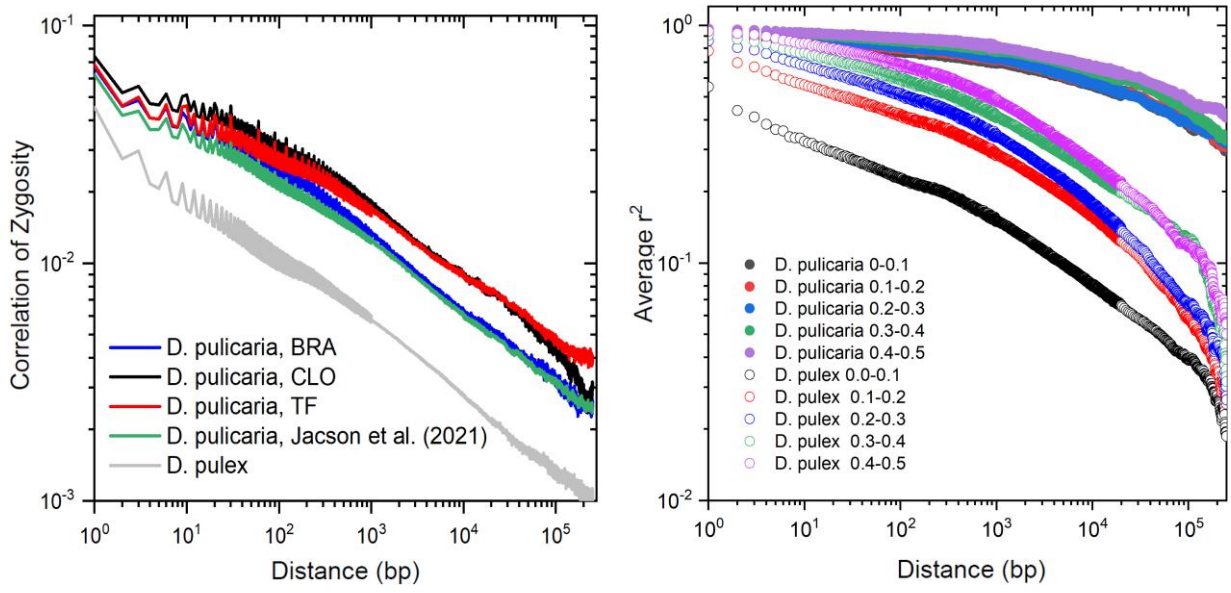

Fig. 4.—LD as a function of physical distance along chromosomes. (Left) Individual-based correlation of heterozygosity. Jackon et al. (2021) denote points derived from a collection of single clones from 14 different populations (Jackson et al. 2021). Data for *D. pulex* are averaged across eight *D. pulex* populations in Lynch et al. (2022) (SE < 0.001 for all distances). (Right) Population-level measures of LD given as averages over the three *D. pulicaria* (pc) populations (closed points), compared with averages from eight *D. pulex* (px) populations (open points; from Lynch et al. 2022), and given using five allele-frequency bins.

Table 2

Comparison of Sequence Diversity and Selection Intensities Between *D. pulex* and *D. pulicaria*

		π_N	π_S	π_N/π_S	d_N	d_S	d_N/d_S	NI_B
All	px	0.0029 (0.0000)	0.0160 (0.0001)	0.2504 (0.0030)	0.0172 (0.0002)	0.1232 (0.0005)	0.1511 (0.0038)	1.4779 (0.0124)
	pc	0.0015 (0.0000)	0.0081 (0.0001)	0.2425 (0.0034)	0.0183 (0.0002)	0.1303 (0.0005)	0.1531 (0.0044)	1.6464 (0.0015)
With orthologs	px	0.0026 (0.0000)	0.0169 (0.0001)	0.1961 (0.0022)	0.0170 (0.0001)	0.1230 (0.0004)	0.1478 (0.0034)	1.3796 (0.0125)
	pc	0.0013 (0.0000)	0.0081 (0.0001)	0.2086 (0.0026)	0.0172 (0.0002)	0.1301 (0.0005)	0.1401 (0.0056)	1.6147 (0.0171)
Without	px	0.0038 (0.0000)	0.0141 (0.0002)	0.4352 (0.0109)	0.0242 (0.0004)	0.1221 (0.0013)	0.2591 (0.0122)	1.4779 (0.0124)
orthologs	pc	0.0026 (0.0001)	0.0079 (0.0003)	0.5359 (0.0066)	0.0297 (0.0009)	0.1359 (0.0006)	0.2685 (0.0035)	1.7826 (0.0169)

Note.—"All" indicates all genes with π_S estimates >0.0005 (to avoid spurious upward bias in ratios). "With orthologs" indicates genes that are one-to-one orthologs between *D. pulex* and *D. pulicaria*, and "without orthologs" indicates genes lacking orthologous data in one or the other species. NI estimates are restricted to genes with $NI_B < 10.0$. Numbers in the parentheses are standard errors. px, *D. pulex*; pc, *D. pulicaria*.

correlation of zygosity (Δ) is an individual-based method, based on the degree to which pairs of sites are jointly heterozygous (Lynch 2008; Lynch et al. 2014). In random-mating populations, when scaled by the average heterozygosity, Δ is expected to decline with physical distance among sites in the same manner as population-level LD. The more general advantages of the approach are that it makes no assumptions about random mating and requires no haplotype phasing. Using this approach, it is clear that there is substantially more LD at all length scales in all *D. pulicaria* populations than in *D. pulex* (fig. 4, left). The magnitude of difference for LD gradually increases from approximately 1.5× at nearly adjacent sites to approximately 5× at distances in excess of 10^5 bp. These elevated levels of Δ are also seen in all of the clones (all from different populations) evaluated in Jackson et al. (2021), showing that this is a species-wide pattern. Second, population-level LD (r^2) based on genotype-frequency distributions (Rogers and Huff 2009) in *D. pulicaria* is also substantially greater than that in *D. pulex* at all length scales, exceeding 0.3–0.4 on average even at distances as large as 250 kb, even for low-frequency SNPs that tend to lead to downwardly biased estimates of r^2 (fig. 4, right).

Local adaptation is associated with protein-coding genes. Given the substantial excess heterozygosity (negative F_{IS}) in *D. pulicaria* populations, it is of interest to know whether this is reflected in unusually high levels of within-population diversity at nonsynonymous sites in protein-coding genes. However, restricting analyses to the 10,275 genes with $\pi_S > 0.0005$ (to avoid spurious upward bias in ratios), average π_N/π_S is 0.2425 (0.0034), nearly identical to the estimate for *D. pulex* obtained in the same way, 0.2504 (0.0030; 12,843 genes; table 2). For the subset of 9,204 protein-coding genes that could be compared across species

(with confident orthology and enough data coverage), the similarity between the two taxa remains: $\pi_N/\pi_S = 0.1961$ (0.0022) for *D. pulex*, and 0.2086 (0.0026) for *D. pulicaria* (Table 2).

To identify genes potentially under diversifying selection at the amino-acid level, we tested the difference between π_N and π_S in units of sampling standard errors. Of the 11,205 *D. pulicaria* genes that could be evaluated, 733 have π_N/π_S values exceeding 1.0, but with a critical *Z*-score value of 4.46 after correcting for multiple comparisons, only 51 (0.45%) are significant. This is more than double the incidence of such significant outlier genes in *D. pulex* (0.22%; Maruki et al. 2022). However, 15 of these 51 genes are not found in the annotated *D. pulex* genome (including 6 out of the 10 genes with the highest π_N/π_S values, with values ranging from 14.8 to 53.7), and of the remaining 36, only 5 have $\pi_N/\pi_S > 1.0$ in *D. pulex*. Notably, only 2 of these genes have π_N greater than the average π_S for all *D. pulicaria* genes, 0.0081 (0.0001), the average π_S for the 51 being just 0.0011 (0.0003). Thus, it is plausible that the majority of these high π_N/π_S genes are stochastic results of low π_S rather than indicators of elevated π_N, although the latter has an average of 0.0033 (0.0007), which is approximately 3× higher than the genome-wide mean of 0.0015 (0.00002). We tentatively conclude that the incidence of balancing selection at amino-acid replacement sites in these populations is very low. Interestingly, for the set of species-specific genes, that is, those exist in one species but missing in the other and the outgroup species, π_N exhibited significant elevations in both species, although π_S remains similar to genes with orthologs (table 2). This result suggests that, at the amino-acid sequence level, species-specific genes in each lineage may evolve faster than orthologous genes.

To determine whether there are any genes under strong diversifying selection among the three populations, we considered the ratio ϕ_N/ϕ_S, where ϕ_x is the among-deme measure of nucleotide diversity. This ratio is expected to exceed 1.0 in cases of strong local adaptation at the amino-acid sequence level. Of the 10,406 genes that could be evaluated, only in 48 cases (<0.5%) was ϕ_N more than two standard errors greater than ϕ_S, and after accounting for multiple comparisons, none of these retained significance. With only three populations evaluated, this analysis does not have great power, but it is notable that two prior studies, involving *D. pulex* (Maruki et al. 2022) and *Daphnia magna* (Fields et al. 2015), also failed to find evidence of local adaptation with this approach.

As tests for $\pi_N/\pi_S > 1$ and $\phi_N/\phi_S > 1$ are quite stringent, owing to the likelihood that a substantial fraction of mutations is deleterious even on a background of positive selection, in a further search for genes exhibiting a signature of local adaptation, we considered the metapopulation neutrality index, NI_W, which is the ratio of π_N/π_S to ϕ_N/ϕ_S (Maruki et al. 2022). Protein-coding genes under diversifying selection among demes are expected to have excess amino-acid sequence divergence (scaled to the silent-site diversities), that is, $NI_W < 1$. For the 6,376 genes that could be compared across both species, the average index is 2.27 (0.09) in *D. pulex* and 3.66 (0.34) in *D. pulicaria.* Removing outliers with $NI_W > 10$, which constitute 2.6% of genes in *D. pulex* and 5.1% of genes in *D. pulicaria*, these means become nearly identical, 1.56 (0.01) and 1.55 (0.02), respectively. The medians and modes of the distributions are also quite similar between species, 1.28 and 1.10 for *D. pulex*, and 1.15 and 1.12 for *D. pulicaria*. Combined with prior results for the sister taxon *D. pulex* and together comprising nearly 1,000 sequenced genomes, these 3 sets of analyses for individual protein-coding genes provide little evidence for diversifying/balancing selection within or among populations.

Between-Species Divergence

To obtain insight into longer term aspects of the evolution of protein-coding genes, we considered levels of divergence with respect to the outgroup *Daphnia obtusa.* The full set of 10,439 genes with adequate data for at least two *D. pulicaria* populations and putative orthologs in *D. obtusa* yielded average values of $d_N = 0.0183$ (0.0002), $d_S = 0.1303$ (0.0005), and $d_N/d_S = 0.1531$ (0.0044; table 2). Comparable average values for the full set of 9,599 *D. pulex/D. obtusa* orthologs are: $d_N = 0.0172$ (0.0002), $d_S = 0.1232$ (0.0005), and $d_N/d_S = 0.1511$ (0.0038; table 2). Considering just the 9,467 protein-coding genes for which both species could be compared with the outgroup, mean d_N is essentially identical for both species: 0.0172 (0.0002) for *D. pulicaria*, and 0.0170 (0.0001) for *D. pulex* (Table 2). On the other hand, mean d_S is slightly higher for *D. pulicaria* than for *D. pulex*, 0.1301 (0.0005) and 0.1230 (0.0004), respectively (table 2). This then renders mean d_N/d_S slightly lower for *D. pulicaria* than for *D. pulex* orthologs, 0.1401 (0.0056) and 0.1478 (0.0034), respectively (table 2). For genes without orthologs in either species, d_N exhibited significant elevations compared with genes with *D. pulex/pulicaria* orthologs, leading to higher d_N/d_S values (table 2).

Using the three two-species divergence estimates, the overall levels of divergence on the *D. pulex* and *D. pulicaria* branches to their common ancestor (rooted by *D. obtusa*) can be obtained by the solution of simultaneous equations (Sarich and Wilson 1967; Wu and Li 1985). This reveals a 29% increase in the amount of divergence at silent sites on the *D. pulicaria* versus the *D. pulex* branch. This need not imply a difference in mutation rates between taxa, however, as there are almost certainly differences in the annual numbers of generations between the two. Were there an average of 29% more generations/year in *D. pulicaria* than in *D. pulex* populations, for example, the above results would imply equal per-generation mutation rates in the

two taxa. Given that *D. pulicaria* occupies permanent lakes, whereas *D. pulex* is dormant most of the year, it is plausible that *D. pulicaria* has twice the number of annual generations, in which case the difference in silent-site divergence would imply a *D. pulicaria* mutation rate approximately 64% of that in *D. pulex*.

To gain further insight into whether genes on the *D. pulicaria* versus *D. pulex* branches experience different levels of purifying and/or positive selection relative to the outgroup, we evaluated the between-species NI_B, which is the ratio of Π_N/Π_S to d_N/d_S, where $\Pi_x = \pi_x + \phi_x$ is the total nucleotide diversity in the metapopulation of a species. The expected value of NI_B is equal to 1.0 under neutrality, declines toward 0.0 as positive selection promotes amino-acid sequence divergence between species, and exceeds 1.0 when the majority of amino-acid substitution sites are under purifying selection.

Regardless of the measure of statistical centrality, these analyses suggest higher NI_B for *D. pulicaria* than for *D. pulex*. Excluding genes with extreme values >100.0, mean $NI_B = 3.42$ (SE = 0.09; $n = 10{,}197$) in *D. pulicaria* and 2.10 (SE = 0.05; $n = 9{,}506$) in *D. pulex*. Further restricting the analyses to genes with $NI_B < 10.0$, the respective values become 1.65 (SE = 0.02; $n = 9{,}550$) and 1.48 (SE = 0.01; $n = 9{,}244$). Using the full set of genes for each species, the respective medians are 1.26 versus 1.12, and the respective modes are 0.94 versus 0.86. Whereas the among-gene variance in NI_B estimates is greater for *D. pulicaria* than for *D. pulex*, this may largely be a consequence of the former being based on three populations and the latter based on ten. These results also remain virtually identical if the analyses are restricted to *D. pulicaria*/*pulex* orthologs. Although the tendency for most of these centrality measures to exceed 1.0 implies an overall predominance of purifying selection acting on amino-acid-altering mutations, mode values <1 imply some operation of positive selection with respect to divergence, more so in *D. pulex*.

For the 9,858 *D. pulicaria* genes that could be tested (with confident orthology with *D. obtusa* and enough data coverage for at least 2 populations), 1,578 had NI_B estimates significantly smaller than 1.0 (using critical *Z*-scores at the 5% level) before Bonferroni correction, and with a critical *Z*-score of 4.44 after correction, 808 (8.2%) remain significant. By comparison, 5.3% of the genes in *D. pulex* were deemed to be under significant positive selection (Maruki et al. 2022). Of these 808 *D. pulicaria* genes, 741 had parallel data for *D. pulex*, 192 (26%) of which were also flagged as being under significant positive selection in this species. Functional-enrichment analysis reveals that positively selected genes in *D. pulicaria* are concentrated in aging and metabolic-related categories such as cellular homeostasis, single-organism catabolic process, carboxylic-acid metabolic process, and small-molecule metabolic process (fig. 5), none of which are highlighted in *D. pulex* (supplementary fig. S2, Supplementary Material online). In contrast, GO categories such as visual perception, one of the most significant GO categories in *D. pulex* (supplementary fig. S2, Supplementary Material online), are not enriched in *D. pulicaria*. However, genes in GO categories involving ribosomes, mitochondria, and responses to environmental stimuli were found to be under positive selection in both species (fig. 5 and supplementary fig. S2, Supplementary Material online).

With the goal of discovering the nature of the genes involved in the differentiation of *D. pulicaria* and *D. pulex*, we evaluated the degree to which specific genes have experienced accelerated rates of amino-acid substitutions in one lineage versus the other. The average raw sequence divergence between the two species at silent sites (d_S) is 0.0506 (SE = 0.0006, $n = 10{,}116$), but after subtracting the average within-species variation at such sites, this declines to an average net divergence $d'_S = 0.0331$ (SE = 0.0006). The average net divergence at replacement sites, d'_N, is 0.0179 (SE = 0.0004). To infer candidate genes potentially harboring mutations responsible for the differentiation between the two species, we computed for each gene the difference in *D. pulex/pulicaria* d'_N from the genomic average of 0.0179 in units of standard errors for the gene. Using the *Z*-scores as a rough guide, after correcting for multiple comparisons for the 9,418 genes compared at the 0.05% level, with a critical one-tailed *Z*-score of 4.43, 539 genes were identified as candidates for excess divergence in one or both species.

We then queried as to whether the rates of evolution were particularly accelerated in one lineage versus the other. The difference in d_N between each of the species contrasted with *D. obtusa* measures the difference in the divergence occurring on *D. pulex* versus the *D. pulicaria* branches emanating from the common ancestor of the two species, with positive versus negative values indicating excess divergence on the *D. pulex* versus *D. pulicaria* branches, respectively. The average deviation in d_N across the full set of *D. pulex/pulicaria* orthologs is just 0.00023 (SE = 0.00008, $n = 9{,}467$), indicating only a very minor difference in overall divergence at the amino-acid sequence level. Of the 539 candidate genes with accelerated evolution within the *D. pulex/pulicaria* lineage, 453 could be confidently contrasted with *D. obtusa*, 243 of which had higher rates of evolution on the *D. pulicaria* branch, while 210 had higher rates on the *D. pulex* branch. With a critical 2-tailed *Z*-score of 3.88, 77 of the accelerated *D. pulicaria* genes were deemed significantly so, whereas 82 were significantly accelerated on the *D. pulex* branch. As in Maruki et al. (2022) and Ye et al. (2023), rapidly evolving genes in *D. pulex* are associated with visual perception, ribosomes, mitochondrial functions, and life span determination, while those in *D. pulicaria* are related to metabolism, ribosomes, and mitochondrial functions. It is worth noting that

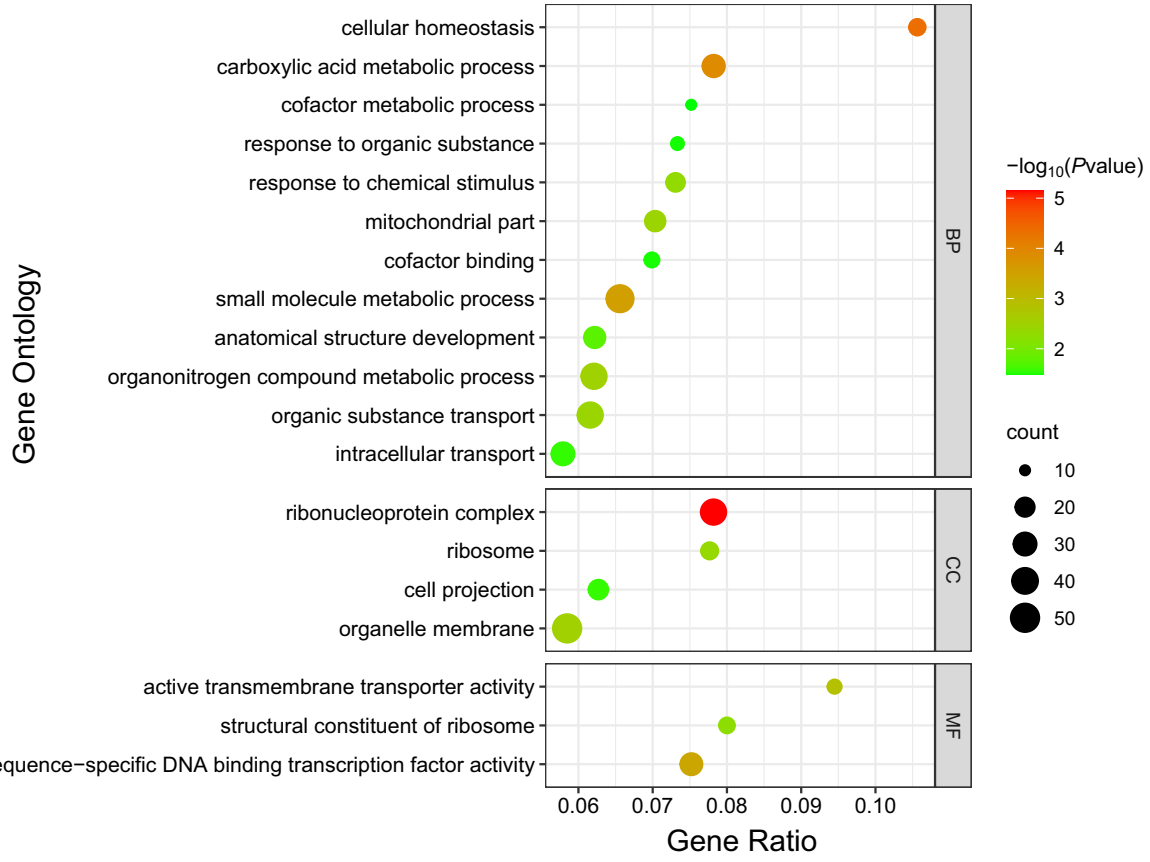

Fig. 5.—GO categories enriched with positively selected genes in *D. pulicaria*. *P*-values are derived from χ^2 tests corrected for multiple testing using the Benjamini and Hochberg (1995) method. The gene ratio is the number of positively selected genes divided by the total number of genes tested under each GO term. BP, biological process; CC, cell component; MF, molecular function.

ribosomes and mitochondrial-function-related categories are under positive selection in both species, although the specific genes involved are different. This suggests that there is different wiring of metabolic networks in the two species, which responds to the needs of their respective microhabitats.

Discussion

Contrasting sister species experiencing ecological differentiation is of particular interest, as a central goal of evolutionary biology is to understand how species adapt to distinct habitats. In this study, we compared population-genomics data between the pond-dwelling *D. pulex* and lake-dwelling *D. pulicaria*, showing that the population-genetic structure differs substantially between the two species. Genome-wide nucleotide diversity for *D. pulicaria* is much lower than that in *D. pulex*, likely reflecting a smaller N_e than in *D. pulex* (assuming equal mutation rates in the two species), despite the fact that individual populations of the former occupy much larger bodies of water. Our estimate for N_e in *D. pulicaria* populations is very similar to that in Wersebe et al. (2022) (258,000 vs. 221,377), although the latter were estimated from multiple *D. pulicaria* populations from a much wider geographic distribution (Wersebe et al. 2022). Therefore, the reduced N_e for *D. pulicaria* populations compared with that of *D. pulex* is unlikely caused by sampling bias. Unlike *D. pulex*, *D. pulicaria* commonly exhibits heterozygosity at nucleotide sites in excess of HW expectations, as previously observed in other permanent-lake populations, including *D. pulicaria* (Hebert 1974; Morgan et al. 2001). The per-capita migration rates of the two species are fairly similar, but owing to the larger N_e, *D. pulex* has a higher absolute migration rate, and hence a lower level of population subdivision.

At the average within-population level, the SFSs for neutral sites are similar between the two species, with a deficit of rare alleles and a close correspondence to the neutral expectation. However, there are three substantial differences in SFS behavior between the two species. First, the metapopulation SFS for neutral sites in *D. pulicaria* exhibits a more discontinuous behavior compared with *D. pulex*. Second, the within-population SFSs for intergenic and amino-acid replacement sites are quite similar to the neutral-site SFS in *D. pulicaria*, in contrast to the substantial depression

observed in *D. pulex*. Third, the distinctions between the metapopulation SFSs for the three functional classes (intergenic, intron/silent, replacement) of sites in *D. pulicaria* remain far less pronounced than in *D. pulex*. These results suggest that there are significant differences in the evolutionary forces shaping the genetic diversity of these two species, with *D. pulicaria* exhibiting a more complex and diverse SFS compared with *D. pulex*. This might be related to the hitch-hiking and Hill–Robertson (selective interference) effects enhanced by elevated LD in *D. pulicaria*, which in turn is a result of a reduction in the recombination rate associated with longer periods of uninterrupted clonal propagation in this species. Moreover, Wersebe et al. (2022) found that *D. pulex* and *D. pulicaria* have experienced different levels of population expansion; hence, the diverse SFS could be a joint effect of demographic changes and selection.

To determine how differences in evolutionary forces/ecological settings might lead to divergent phenotypes (e.g., longer life span, vertical migration in *D. pulicaria*; different responses to short-day photoperiods) in the two species, we searched for genes under positive selection in *D. pulicaria*. Eight hundred and eight genes were deemed to be under significant positive selection, 192 (26%) of which overlapped with those behaving similarly in *D. pulex*. Functional-enrichment analysis revealed that positively selected genes in *D. pulicaria* are enriched in aging-related categories such as cellular homeostasis and carboxylic-acid metabolic processes (fig. 5). Cellular homeostasis is known to be related to aging (Zhang and Cuervo 2008; Morimoto and Cuervo 2009; Hartl 2016; Pomatto and Davies 2017). The expression of genes in carboxylic-acid metabolic processes is affected by food availability (i.e., reduced calories; Zhang et al. 2011), which might indirectly link to aging, as caloric restriction affects aging in numerous organisms, including *Daphnia* (Lynch 1989; Cao et al. 2001; Piper and Partridge 2007; Latta et al. 2011). Thus, it is plausible that a number of positively selected genes associated with these processes might be playing important roles in the extended life span of *D. pulicaria* (Dudycha and Tessier 1999; Dudycha 2003). Both this study and Wersebe et al. (2022) found that metabolism-related genes are under different selection pressures in the species pair of *D. pulex* and *D. pulicaria*, which is likely related to various differences in aspects of life-history traits such as life span, migration behavior, predator response, etc.

Another interesting discovery is that genes related to visual perception and cellular response to light stimuli are exclusively under positive selection in *D. pulex* (Maruki et al. 2022; Ye et al. 2023). Among the genes under strong positive selection are opsin, rhodopsin, and melanopsin, all of which have been shown to mediate photoperiodic responses (Foster et al. 1985; Collantes-Alegre et al. 2018; Zhao et al. 2018; Senthilan et al. 2019; Frau et al. 2022). Consistently, opsin genes were also found to be under positive selection in a previous study (Jackson et al. 2021), as revealed by high F_{ST} between *D. pulex* and *D. pulicaria*. *Daphnia* living in ponds and lakes have significantly different responses to photoperiod (Deng 1997), likely as a consequence of differences in ambient light duration and intensity. Thus, our results suggest that the different light responses in the two species may be mediated through evolutionary changes in opsin, rhodopsin, and/or melanopsin-related genes. Our results also revealed that lactate dehydrogenase A (*ldh*A) experienced different evolutionary rates on the *D. pulex* and *D. pulicaria* branches. *ldh*A is a gene used to distinguish *D. pulex* and *D. pulicaria* at the molecular level, and it has been suggested to be under adaptive evolution (Crease et al. 2011). Our results confirmed its role in separating the species pair.

Previous studies have suggested that *D. pulex* may evolve faster than *D. pulicaria* due to higher growth rate and stronger selection pressures from the environments (Dudycha and Tessier 1999; Dudycha 2004). However, our study revealed that the two species have similar average evolutionary rates at the DNA-sequence level, with <1% of genes experiencing significantly different rates in the two lineages. This result suggests that only a small set of genes may be responsible for ecological differentiation in the species pair. These small sets of genes are tightly linked with the diverse traits in these two species, for example, extended life span in *D. pulicaria* and different visual responses to photoperiod. The list of genes with different evolutionary rates could serve as a resource for future studies on the speciation of the species pair.

Materials and Methods

Sample Preparation and Sequencing

We collected *D. pulicaria* clones from three randomly chosen permanent lakes in the midwestern United States and southern Canada in spring of 2015 and 2018 (supplementary table S1, Supplementary Material online). As in the *D. pulex* populations (Lynch et al. 2017; Maruki et al. 2022), adult individuals were collected early in the growing season, although in this case, there is no guarantee that clones were derived from the recent hatching of resting eggs as opposed to having been propagating by overwintering individuals. As in prior work (Maruki et al. 2022), after clonal expansion of individual clones in the laboratory, DNA was extracted for 75–95 clones per population for library preparation using the Nextera kit. The DNA from each clone was tagged by unique oligomer barcodes, with paired-end short (150 bp) reads obtained with Illumina technology.

Data Preparation

As in Maruki et al. (2022), we filtered the raw reads using Trimmomatic (version 0.36; Bolger et al. 2014), and the filtered reads were mapped to the *D. pulicaria* reference

genome (LK16; Jackson et al. 2021, accession number GCA_017493165.1) using NovoAlign (version 3.02.11). We further removed duplicated reads using Picard (version 2.9.2) and generated mpileup files using SAMtools (version 1.8). We applied proview command from MAPGD (version 0.4.26; Ackerman et al. 2017) to the mpileup files and obtained nucleotide-read quartets for all individuals, called a pro file.

Several procedures have been applied to filter sites in the pro file. First, repeat regions identified by RepeatMasker (version 4.0.5; www.repeatmasker.org) were excluded from the analysis as they may cause mapping issues; second, we eliminated polymorphic sites with nonbinomial deviations in ≥4 deviant individuals; third, we removed clones with mean coverage over sites <3×, which could easily lead to false genotype calls; fourth, we removed clones with a sum of the goodness-of-fit values across the genome <−0.4, which can happen when contaminated with another clone; fifth, we set population-coverage cut-offs to avoid analyzing problematic sites (supplementary table S1, Supplementary Material online). After the above filtering, allele and genotype frequencies in each population were estimated with the maximum-likelihood (ML) method of Maruki and Lynch (2015).

Analyses of Genetic Differentiation Among Populations

To quantify the degree of genetic differentiation among populations, we estimated Wright's F_{ST} from genotype-frequency estimates using Weir and Cockerham's method (1984), taking a procedure described in Maruki et al. (2022). As in Maruki et al. (2022), we first examined the relationship between F_{ST} and metapopulation-wide MAF estimates to consider the confounding effect of the maximum possible value of F_{ST} limited by MAF (Alcala and Rosenberg 2017; Maruki et al. 2022) on the analyses of the F_{ST} estimates.

Linkage-Disequilibrium Analysis

To examine how LD decays with distance between sites, we estimated the correlation of zygosity Δ (Lynch 2008) from the nucleotide-read quartets of single clones using mlRho (version 2.9; Haubold et al. 2010). We applied mlRho to three clones with the highest coverage and calculated the mean of the Δ estimates over the clones in each population. Population-level LD (r^2) in *D. pulicaria* is calculated using the method in Rogers and Huff (2009), restricting analysis to significant polymorphic sites and sites with no <50 genotypes.

Analyses of Private Alleles

To analyze alleles uniquely found in each of the populations (private alleles; Slatkin 1985), we identified private alleles taking the procedure described in Maruki et al. (2022). Using the mean frequencies of the private alleles over sites, the migration rate was inferred in each population using Equation (3) in Slatkin (1985) and adjusting the sample size. To infer the contribution of introgression from closely related species to the observed private alleles in genes significantly ($P < 0.05$) enriched with private alleles, we aligned reference sequences of *D. pulex* (KAP4) and *D. obtusa* to LK16 genome assembly using LAST (version 1066) and examined the fractions of private alleles matching reference nucleotides of *D. pulex* (KAP4), *D. obtusa*, and both species in each of the genes.

Analyses of Selection Signatures in Protein-Coding Sequences

We compared per-site differences in protein-coding sequences between amino-acid replacement and silent sites at three levels (between species, among populations, and within populations), as in Maruki et al. (2022). To estimate divergence between species, we mapped sequence reads of the *D. obtusa* draft reference sequence to LK16 and called genotypes using high-coverage genotype caller (Maruki and Lynch 2017). Then, we converted homozygous genotype calls into nucleotides. Using the estimates of the divergence and heterozygosity within populations, we also estimated analogs of the NI (Rand and Kann 1996). The analysis of among-population heterozygosity ϕ estimates was carried out using ϕ calculated as $\Pi - \pi$, where Π and π are the heterozygosity in the total population and mean heterozygosity within populations, respectively, and estimated in the framework of Weir and Cockerham (1984).

Analysis of Selection on Coding-Region DNA

As in Maruki et al. (2022), we estimated diversity at synonymous and nonsynonymous sites, π_S and π_N, for each gene within each population. The population level π_N/π_S for each gene is calculated as mean π_N across populations divided by the mean π_S, and the sampling variance was obtained using the Delta method equation (A1.19b, in Lynch and Walsh 1998). Similar measures have been applied to the estimation of between-species divergence, d_N and d_S, by drawing comparisons with orthologous genes contained within the closely related outgroup species *D. obtusa*. d_N and d_S estimations were restricted to genes with >300 coding sites. For each gene, d_N and d_S were estimated as the mean divergence estimates across all populations, and multiple hits were corrected using Jukes and Cantor's (1969) method. Then, d_N/d_S for each gene was obtained as the mean d_N divided by d_S, and the variance of the ratio was again calculated using the delta method.

Enrichment Analysis of Gene Ontology Terms in the Outlier Genes

To determine the enriched gene ontology (GO) terms for the outlier genes, we performed enrichment analysis by using a χ^2 test (fulfilled using the Perl module Statistics::ChisqIndep). To reduce the redundancy of overlapping GO terms, we removed GO terms when >80% of the genes overlapped with another higher-level GO term using web tool http://revigo.irb.hr/. The Benjamini–Hochberg method is used to adjust the resulted *P*-values for multiple testing correction.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This work was financially supported by NIH grant R35-GM122566-01 to M.L. and NSF Enabling Discovery through GEnomics (EDGE) grants IOS-1922914 to M.L. and Andrew Zelhof (Indiana University) and IOS-2220696 to M.E.P. The authors thank Bret Coggins for assistance with the sample collections, Rebecca Shelley and Nhu-Y Nguyen for assistance in the laboratory, and Joe Shaw for sharing population BRA.

Data Availability

The FASTQ files of the raw sequencing data for the *D. pulicaria* populations are available at the NCBI Sequence Read Archive (accession number: SRP439008). The FASTQ files of the *D. obtusa* sequence reads are available at the NCBI SRA (accession number SAMN12816670), and the *D. obtusa* genome assembly is available at the NCBI GenBank (accession number JAACYE000000000). The *D. pulicaria* genome assembly LK16 is available at GenBank under accession GCA_017493165.1, and the annotation is available at https://osf.io/nr7t6/. The *D. pulex* KAP4 assembly and annotation can be accessed at NCBI under accession GCF_021134715.1. C++ programs for analyzing population structure (PSA) and finding private alleles (FPAs) are available at https://github.com/Takahiro-Maruki/PSA and https://github.com/Takahiro-Maruki/FPA, respectively.

Literature Cited

Ackerman MS, et al. 2017. Estimating seven coefficients of pairwise relatedness using population-genomic data. Genetics. 206(1): 105–118.

Alcala N, Rosenberg NA. 2017. Mathematical constraints on F_{ST}: biallelic markers in arbitrarily many populations. Genetics 206(3): 1581–1600.

Allen DE, Lynch M. 2012. The effect of variable frequency of sexual reproduction on the genetic structure of natural populations of a cyclical parthenogen. Evolution 66(3):919–926.

Bedford NL, Hoekstra HE. 2015. *Peromyscus* mice as a model for studying natural variation. eLife 4:e06813.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Stat Methodol. 57(1):289–300.

Bergland AO, Tobler R, González J, Schmidt P, Petrov D. 2016. Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in *Drosophila melanogaster*. Mol Ecol. 25(5):1157–1174.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30(15):2114–2120.

Bradshaw WE, et al. 2018. Evolutionary transition from blood feeding to obligate nonbiting in a mosquito. Proc Natl Acad Sci U S A. 115(5):1009–1014.

Brandlova J, Brandl Z, Fernando CH. 1972. The Cladocera of Ontario with remarks on some species and distribution. Can J Zool. 50(11):1373–1403.

Cao SX, Dhahbi JM, Mote PL, Spindler SR. 2001. Genomic profiling of short-and long-term caloric restriction effects in the liver of aging mice. Proc Natl Acad Sci U S A. 98(19):10630–10635.

Collantes-Alegre JM, Mattenberger F, Barberà M, Martínez-Torres D. 2018. Characterisation, analysis of expression and localisation of the opsin gene repertoire from the perspective of photoperiodism in the aphid *Acyrthosiphon pisum*. J Insect Physiol. 104:48–59.

Crawford DL, Schulte PM, Whitehead A, Oleksiak MF. 2020. Evolutionary physiology and genomics in the highly adaptable killifish (*Fundulus heteroclitus*). Compr Physiol. 10(2):637–671.

Crease TJ, Floyd R, Cristescu ME, Innes D. 2011. Evolutionary factors affecting lactate dehydrogenase A and B variation in the *Daphnia pulex* species complex. BMC Evol Biol. 11:212.

Crease TJ, Lynch M, Spitze K. 1990. Hierarchical analysis of population genetic variation in mitochondrial and nuclear genes of *Daphnia pulex*. Mol Biol Evol. 7(5):444–458.

Cristescu ME, Constantin A, Bock DG, Caceres CE, Crease TJ. 2012. Speciation with gene flow and the genetics of habitat transitions. Mol Ecol. 21(6):1411–1422.

De A, Durrett R. 2007. Stepping-stone spatial structure causes slow decay of linkage disequilibrium and shifts the site frequency spectrum. Genetics 176(2):969–981.

Deng HW. 1997. Photoperiodic response of sexual reproduction in the *Daphnia pulex* group is reversed in two distinct habitats. Limnol Oceanogr. 42(3):609–611.

Dudycha JL. 2003. A multi-environment comparison of senescence between sister species of *Daphnia*. Oecologia 135:555–563.

Dudycha JL. 2004. Mortality dynamics of *Daphnia* in contrasting habitats and their role in ecological divergence. Freshw Biol. 49(5): 505–514.

Dudycha JL, Tessier AJ. 1999. Natural genetic variation of life span, reproduction, and juvenile growth in *Daphnia*. Evolution 53(6): 1744–1756.

Fields PD, Reisser C, Dukić M, Haag CR, Ebert D. 2015. Genes mirror geography in *Daphnia magna*. Mol Ecol. 24:4521–4536.

Foster RG, Follett BK, Lythgoe JN. 1985. Rhodopsin-like sensitivity of extra-retinal photoreceptors mediating the photoperiodic response in quail. Nature 313(5997):50–52.

Frau S, et al. 2022. From embryo to adult life: differential expression of visual opsins in the flatfish *Solea senegalensis* under different light spectra and photoperiods. Front Mar Sci. 9:85.

Glaholt SP, Kennedy ML, Turner E, Colbourne JK, Shaw JR. 2016. Thermal variation and factors influencing vertical migration behavior in *Daphnia* populations. J Therm Biol. 60:70–78.

Gliwicz MZ. 1986. Predation and the evolution of vertical migration in zooplankton. Nature 320(6064):746–748.
Hartl FU. 2016. Cellular homeostasis and aging. Annu Rev Biochem. 85:1–4.
Haubold B, Pfaffelhuber P, Lynch M. 2010. mlRho–a program for estimating the population mutation and recombination rates from shotgun-sequenced diploid genomes. Mol Ecol. 19: 277–284.
Hebert PD. 1974. Enzyme variability in natural populations of *Daphnia magna*. II. Genotypic frequencies in permanent populations. Genetics 77(2):323–334.
Hebert PD, Beaton MJ, Schwartz SS, Stanton DJ. 1989. Polyphyletic origins of asexuality in *Daphnia pulex*. I. Breeding-system variation and levels of clonal diversity. Evolution 43(5):1004–1015.
Hebert PDN, Schwartz SS, Ward RD, Finston TL. 1993. Macrogeographic patterns of breeding system diversity in the *Daphnia pulex* group. I. Breeding systems of Canadian populations. Heredity 70(2): 148–161.
Heier CR, Dudycha JL. 2009. Ecological speciation in a cyclic parthenogen: sexual capability of experimental hybrids between *Daphnia pulex* and *Daphnia pulicaria*. Limnol Oceanogr. 54(2):492–502.
Jackson CE, et al. 2021. Chromosomal rearrangements preserve adaptive divergence in ecological speciation. bioRxiv doi:10.1101/2021.08.20.457158.
Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In: Munro HN, editor. Mammalian protein metabolism. New York: Academic Press. p. 21–132.
Keith N, et al. 2016. High mutational rates of large-scale duplication and deletion in *Daphnia pulex*. Genome Res. 26(1):60–69.
Latta LC, Frederick S, Pfrender ME. 2011. Diet restriction and life history trade-offs in short- and long-lived species of *Daphnia*. J Exp Zool. 315(10):610–617.
Lynch M. 1987. The consequences of fluctuating selection for isozyme polymorphisms in *Daphnia*. Genetics 115(4):657–669.
Lynch M. 1989. The life history consequences of resource depression in *Daphnia pulex*. Ecology 70(1):246–256.
Lynch M. 2008. Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. Mol Biol Evol. 25(11):2421–2431.
Lynch M, et al. 2014. Genome-wide linkage-disequilibrium profiles from single individuals. Genetics 198(1):269–281.
Lynch M, Gutenkunst R, Ackerman M, Spitze K, Ye Z, Maruki T, Jia Z. 2017. Population genomics of *Daphnia pulex*. Genetics 206(1): 315–332.
Lynch M, Walsh B. 1998. Genetics and analysis of quantitative traits. Sunderland (MA): Sinauer.
Lynch M, Ye Z, Urban L, Maruki T, Wei W. 2022. The linkage-disequilibrium and recombinational landscape in *Daphnia pulex*. Genome Biol Evol. 14(11):evac145.
Maruki T, Lynch M. 2015. Genotype-frequency estimation from high-throughput sequencing data. Genetics 201(2):473–486.
Maruki T, Lynch M. 2017. Genotype calling from population-genomic sequencing data. G3: Genes Genom Genet. 7(5): 1393–1404.
Maruki T, Ye Z, Lynch M. 2022. Evolutionary genomics of a subdivided species. Mol Biol Evol. 39(8):msac152.
Messer PW. 2009. Measuring the rates of spontaneous mutation from deep and large-scale polymorphism data. Genetics 182(4):1219–1232.
Morgan KK, et al. 2001. Patterns of genetic architecture for life-history traits and molecular markers in a subdivided species. Evolution 55(9):1753–1761.
Morimoto RI, Cuervo AM. 2009. Protein homeostasis and aging: taking care of proteins from the cradle to the grave. J Med Biomed Sci. 64(2):167–170.
Moy I, Green M, Pham TP, Luu DH, Xu S. 2021. The life history fitness of F1 hybrids of the microcrustaceans *Daphnia pulex* and *Daphnia pulicaria* (Crustacea, Anomopoda). Invert Biol. 140:e12333.
Peichel CL, Marques DA. 2017. The genetic and molecular architecture of phenotypic diversity in sticklebacks. Philos Trans R Soc Lond B Biol Sci. 372(1713):20150486.
Piper MDW, Partridge L. 2007. Dietary restriction in *Drosophila*: delayed aging or experimental artefact? PLoS Genet. 3(4):e57.
Pomatto LCD, Davies KJA. 2017. The role of declining adaptive homeostasis in ageing. Physiol J. 595(24):7275–7309.
Rand DM, Kann LM. 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Mol Biol Evol. 13(6):735–748.
Rodrigues MF, Vibranovski MD, Cogni R. 2021. Clinal and seasonal changes are correlated in *Drosophila melanogaster* natural populations. Evolution 75(8):2042–2054.
Rogers AR, Huff C. 2009. Linkage disequilibrium between loci with unknown phase. Genetics 182(3):839–844.
Sarich VM, Wilson AC. 1967. Rates of albumin evolution in primates. Proc Natl Acad Sci. 58(1):142–148.
Senthilan PR, Grebler R, Reinhard N, Rieger D, Helfrich-Förster C. 2019. Role of rhodopsins as circadian photoreceptors in the *Drosophila melanogaster*. Biology (Basel) 8(1):6.
Slatkin M. 1985. Rare alleles as indicators of gene flow. Evolution. 39(1):53–65.
Städler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P. 2009. The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. Genetics 182(1):205–216.
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(39):15740–15745.
Vergilino R, Markova S, Ventura M, Manca M, Dufresne F. 2011. Reticulate evolution of the *Daphnia pulex* complex as revealed by nuclear markers. Mol Ecol. 20(6):1191–1207.
Wang RJ, Payseur BA. 2017. Genetics of genome-wide recombination rate evolution in mice from an isolated island. Genetics 206(4): 1841–1852.
Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 1358–1370.
Wersebe MJ, Sherman RE, Jeyasingh PD, Weider LJ. 2022. The roles of recombination and selection in shaping genomic divergence in an incipient ecological species complex. Mol Ecol. 32(6):1478–1496.
Wright S. 1938. The distribution of gene frequencies under irreversible mutation. Proc Natl Acad Sci U S A. 24(7):253–259.
Wright S. 1951. The genetical structure of populations. Ann Eugen. 15(1):323–354.
Wu CI, Li WH. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proc Natl Acad Sci. 82(6):1741–1745.
Xu S, et al. 2015. Hybridization and the origin of contagious asexuality in *Daphnia pulex*. Mol Biol Evol. 32(12):3215–3225.
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(17): 4549–4561.
Ye Z, Jiang X, Pfrender ME, Lynch M. 2021. Genome-wide allele-specific expression in obligately asexual *Daphnia pulex* and the implications for the genetic basis of asexuality. Genome Biol Evol. 13(11):evab243.
Ye Z, Molinier C, Zhao C, Haag CR, Lynch M. 2019. Genetic control of male production in *Daphnia pulex*. Proc Natl Acad Sci U S A. 116(31):15602–15609.
Ye Z, Wei W, Pfrender ME, Lynch M. 2023. Evolutionary insights from a large-scale survey of population-genomic variation. Mol Biol Evol. 40(11):msad233.

Ye Z, Williams E, Zhao C, Burns CW, Lynch M. 2021. The rapid, mass invasion of New Zealand by North American "*Daphnia pulex*". Limnol Oceanogr. 66(7):2672–2683.
Ye Z, Zhao C, Raborn RT, Lin M, Wei W, Hao Y, Lynch M. 2022. Genetic diversity, heteroplasmy, and recombination in mitochondrial genomes of *Daphnia pulex*, *Daphnia pulicaria*, and *Daphnia obtusa*. Mol Biol Evol. 39(4):msac059.
Zhang C, Cuervo AM. 2008. Restoration of chaperone-mediated autophagy in aging liver improves cellular maintenance and hepatic function. Nat Med. 14(9):959–965.
Zhang F, Xu X, Zhou B, He Z, Zhai Q. 2011. Gene expression profile change and associated physiological and pathological effects in mouse liver induced by fasting and refeeding. PLoS One 6(11):e27553.
Zhao H, Jiang J, Wang G, Le C, Wingfield JC. 2018. Daily, circadian and seasonal changes of rhodopsin-like encephalic photoreceptor and its involvement in mediating photoperiodic responses of Gambel's white-crowned sparrow, *Zonotrichia leucophrys gambelii*. Brain Res. 1687:104–116.

Associate editor: Josefa Gonzalez