ORIGINAL ARTICLE OPEN ACCESS

Trans-Specific Polymorphisms Between Cryptic *Daphnia* Species Affect Fitness and Behavior

Connor S. Murray[1,2] | Madison Karram[1] | David J. Bass[1] | Madison Doceti[1] | Dörthe Becker[1,3] | Joaquin C. B. Nunez[1,4] | Aakrosh Ratan[2] | Alan O. Bergland[1]

[1]Department of Biology, University of Virginia, Charlottesville, Virginia, USA | [2]Department of Genome Sciences, University of Virginia School of Medicine, Charlottesville, Virginia, USA | [3]School of Biosciences, Ecology and Evolutionary Biology, University of Sheffield, Sheffield, UK | [4]Department of Biology, University of Vermont, Burlington, Vermont, USA

Correspondence: Connor S. Murray (csm6hg@virginia.edu) | Alan O. Bergland (aob2x@virginia.edu)

ABSTRACT

Shared polymorphisms, loci with identical alleles across species, are of unique interest in evolutionary biology as they may represent cases of selection maintaining ancient genetic variation post-speciation, or contemporary selection promoting convergent evolution. In this study, we investigate the abundance of shared polymorphism between two members of the *Daphnia pulex* species complex. We test whether the presence of shared mutations is consistent with the action of balancing selection or alternative hypotheses such as hybridization, incomplete lineage sorting or convergent evolution. We analyzed over 2,000 genomes from six taxa in the *D. pulex* species group and examined the prevalence and distribution of shared alleles between the focal species pair, North American and European *D. pulex*. We show that North American and European *D. pulex* diverged over 10 million years ago, yet retained tens of thousands of shared polymorphisms. We suggest that the number of shared polymorphisms between North American and European *D. pulex* cannot be fully explained by hybridization or incomplete lineage sorting alone. We show that most shared polymorphisms could be the product of convergent evolution, that a limited number appear to be old trans-specific polymorphisms, and that balancing selection is affecting convergent and ancient mutations alike. Finally, we provide evidence that a blue wavelength opsin gene with trans-specific polymorphisms has functional effects on behavior and fitness in the wild.

1 | Introduction

Genetic diversity reflects a species' history and serves as the foundation for adaptation to ecological change. In nature, mutations arise and their persistence time is a function of their selective value and the effective population size of the focal species (Li and Nei 1977), as well as features of the environment (Cvijović et al. 2015). One distinct type of genetic variant is a shared polymorphism, in which mutations are identical by state across closely related species (Wang and Mitchell-Olds 2017). The abundance and frequency of shared polymorphisms between two species can provide insight into some of the most interesting processes in evolution. A shared polymorphism that arose prior to the split of two species is generally referred to as a trans-species polymorphism (Hedrick 2013; Wiuf et al. 2004; Wu et al. 2017). Trans-specific polymorphisms can be used to study the speciation process (Klein et al. 1998), helping refine estimates of the timing and

population sizes at divergence (Edwards and Beerli 2000). Unless divergence happens recently or there is ongoing gene-flow, it is unlikely that neutral polymorphisms will be retained in both species for long (Leffler et al. 2013). Therefore, the presence of trans-species polymorphism between two species with limited gene-flow can be a powerful way to identify balanced polymorphisms (Clark 1997). These polymorphisms are presumed to be maintained by temporal or spatial variation in the direction of natural selection or by genetic overdominance (e.g., Bergland et al. 2014; Cho et al. 2006; Cornetti et al. 2024; Wills 1975; Schield et al. 2022; Ségurel et al. 2012; Ségurel, Gao, and Przeworski 2013). Shared polymorphisms can also indicate convergent adaptive evolution (Castoe et al. 2009) or adaptive introgression (Hedrick 2013), and these polymorphisms themselves can also be the target of balancing selection (Wang and Mitchell-Olds 2017).

Identifying the forces that generate and maintain shared polymorphism is therefore an important problem in evolutionary genetics. However, identifying the contribution of demographic and adaptive evolutionary processes to the generation and maintenance of shared polymorphisms is challenging. Testing alternative hypotheses for the generation of shared polymorphism is a tractable problem for older species pairs because neutral trans-species polymorphisms are expected to be rare thereby eliminating incomplete lineage sorting as a possible explanation (Rosenberg 2003). If sufficient time has occurred for the fixation of species-specific alleles, patterns of adaptive introgression can be identified (e.g., Huerta-Sánchez et al. 2014; Setter et al. 2020), especially if it occurred recently. Evidence for long-term balancing selection of a trans-specific polymorphism becomes stronger if there are multiple sites at a locus that are shared, and these are tightly linked causing gene-trees to group by haplotype and not by species (Wang et al. 2020). Moreover, the presence of trans-specific haplotypes suggests that multiple functional sites at the locus are the target of balancing selection (Charlesworth 2006).

Daphnia are an excellent model to study the mechanisms that generate and maintain shared polymorphism. *Daphnia* are freshwater microcrustaceans that have been the focus of ecological and evolutionary research for over a century (Ebert 2022). Amongst the most widely studied taxa within this genus are *Daphnia magna* (Decaestecker et al. 2007), *Daphnia obtusa* (Spitze 1993), as well as *Daphnia pulex* (Lynch et al. 2017) and its close relatives (Colbourne et al. 2011). The *D. pulex* species group is currently undergoing adaptive radiation (Fryer 1991). Owing to their recent divergence, some members of the North American *D. pulex* species group are known to hybridize (Held, Koenemann, and Schubart 2016), resulting in introgression-induced obligate asexuality (Xu et al. 2015) and the loss of male production (Ye et al. 2019). Members of the *D. pulex* species group, including *D. obtusa, D. pulicaria* and *D. pulex*, are found across the Palearctic and Nearctic (Crease et al. 2012), and recently established populations can be found in other regions of the world as invasives (So et al. 2015). Although *D. obtusa*, *D. pulicaria* and *D. pulex* have been identified on multiple continents, each of these taxa represents polyphyletic groups (Černý and Hebert 1999). For instance, *D. pulex* found in North America are more closely related to North American *D. pulicaria* than

to European *D. pulex* (Crease et al. 2012). The confusion of species identification and naming in this genus is due to their similar morphology (Dodson 1981) and ecological niches (Chin and Cristescu 2021) plus their capacity to interbreed (Pantel, Juenger, and Leibold 2011), generally reflecting the taxonomic ambiguities within the species group (Hebert and Wilson 1994), and amongst zooplankton in general (Brooks 1957a, 1957b).

The evolutionary and ecological history of the *D. pulex* species group affords us an ideal opportunity to study the evolutionary forces that have shaped patterns of shared polymorphism. Here, we assessed alternative causes of shared polymorphisms between North American and European *D. pulex*. Using population genomic data, we first confirm that North American and European *D. pulex* are distinct species that diverged millions of years ago. Next, we show that North American and European *D. pulex* possess tens of thousands of shared polymorphisms, whose abundance cannot be explained by neutral, demographic models alone. Therefore, we conclude that many of these shared polymorphisms arose either via convergent evolution or have been maintained by selection since the split between these taxa. We identify a blue wavelength opsin with an excess of shared non-synonymous polymorphisms that are in linkage disequilibrium, consistent with long-term balancing selection operating on functionally distinct haplotypes. We show that European *D. pulex* clones harbouring alternate genotypes for this blue wavelength opsin have differences in movement and activity that are dependent on light conditions and provide evidence for overdominance in the wild. Taken together, our results highlight the abundance, selective history and function of shared polymorphisms in *Daphnia* and contribute to our understanding of the phylogeography for this model system.

2 | Materials and Methods

2.1 | Sampling and Sequencing of New European *Daphnia* Genomes

Daphnia were sampled from 16 ponds throughout England in 2018 (Table S1). Samples were transported to the University of Virginia and clonally derived isofemale lines were established. Samples were identified as either *D. pulex*, *D. pulicaria* or *D. obtusa* based on morphological characteristics using an online dichotomous key (http://cfb.unh.edu/cfbkey/html/anatomy/daphnia/daphnia.html). DNA extraction and library preparation followed methods outlined in Barnard-Kubow et al. (2022). Briefly, isofemale lines were exposed to antibiotics (streptomycin, tetracycline and ampicillin, 50 mg/L of each) and fed Sephadex G-25 beads to clear their gut of algae. Samples were homogenised using metal beads and a bead beater and DNA was extracted using the Agencourt DNAdvance kit (Beckman-Coulter). RNA was removed using RNase. DNA was quantified using the broad-range Quant-iT dsDNA kit (ThermoFisher Scientific) and normalized to 1–2 ng/μL before library construction. Full genome libraries were constructed using a scaled-down Nextera protocol (Baym et al. 2015). Libraries were size selected for fragments ranging from 450 to 550 bp using a Blue Pippin and quality

checked using a BioAnalyzer. Samples were sequenced on a HiSeq X platform, paired-end 150 bp.

2.2 | Publicly Available *Daphnia* Genomes

Genome sequences of North American and European *D. pulex*, *D. pulicaria* and *D. obtusa* were obtained from NCBI's Sequence Read Archive (Leinonen et al. 2011). Additionally, we obtained short-read sequence data for known hybrids between North American *D. pulex* and *D. pulicaria*. We incorporated wild-sequenced or isogenic female lineages (Alzbutas 2015; Barnard-Kubow et al. 2022; Jackson et al. 2021; Jiang et al. 2017; Lack 2017; Lynch et al. 2017; Maruki 2018; Maruki, Ye, and Lynch 2022; Tucker et al. 2013; Xu et al. 2015; Ye et al. 2022), and excluded samples that were from mutation accumulation studies. Species identity for these samples was based on annotations provided in each sequence read archive record. Information about each clone can be found in Table S1.

2.3 | Short-Read Mapping

Prior to mapping samples, adaptors were removed using *trimmomatic v0.39* (Bolger, Lohse, and Usadel 2014), and overlapping reads were merged using *pear v0.9.11* (Zhang et al. 2014). All samples were mapped to the European *D. pulex* genome (D84A; GenBank assembly: GCA_023526725.1; Barnard-Kubow et al. 2022) using *bwa mem v0.7.17* (Li 2013), and downstream data manipulation was performed using *samtools merge v1.12* (Li et al. 2009). Duplicate reads were marked and removed using *picard v2.23.4* (https://broadinstitute.github.io/picard/).

Additionally, we mapped North American *D. pulex* to the North American *D. pulex* reference genome (KAP4; GenBank assembly: GCF_021134715.1; Ye, Pfrender, and Lynch 2023) using the same mapping strategy outlined above. We created a liftOver file to translate features in KAP4 to D84A. We created the liftOver file by running pairwise alignments using *lastz v1.04.22* followed by the use of various UCSC tools to chain the alignments, sort them, filter them and convert them into UCSC nets and chains (Harris 2007). The chains exhibited good coverage, allowing us to translate 72.6% of the KAP4 genome to D84A (Lee et al. 2022). We used *LiftOverVCF* from *picard v2.23.4* to convert the KAP4-aligned VCF to the D84A genome coordinates.

2.4 | SNP Calling and Filtering

We performed SNP calling using *HaplotypeCaller* and *GenotypeGVCFs* from *gatk v4.1.6.0* (Poplin et al. 2017). We called SNPs by first using the short-reads for all taxa in the *D. pulex* species group mapped to the European *D. pulex* genome (D84A). Additionally, we called SNPs in North American *D. pulex* using the reads mapped back to the North American *D. pulex* genome (KAP4). In both strategies, we used *VariantFiltration* in *gatk* to remove low-quality SNPs recommended for organisms without reference panels: ('QD < 2.0', 'QUAL < 30.0', 'SOR > 3.0', 'FS > 60.0', 'MQ < 40.0', 'MQRankSum < −12.5' and 'ReadPosRankSum < −8.0'). We removed sites flanking indels

(±10 bp) using *bcftools filter* (Li et al. 2009) and removed indels. We annotated SNPs using *snpEff v4.3t* (Cingolani et al. 2012).

We filtered the dataset by first removing all samples with average genome-wide coverage <8×. Next, we filtered any genomic region (10 kbp window) with high (95% upper quantile tail; ≥35×) or low mean site read coverage (5% lower quantile; ≤8×) across all samples, along with chromosomal endpoints and regions ±500 bps surrounding unassembled gaps that likely represent transposable elements. Repetitive elements identified in both the European and North American *D. pulex* genomes were classified with *RepeatMasker v4.0.8* and removed (Tarailo-Graovac and Chen 2009). Most analyses removed SNPs that have a minor allele frequency (MAF) <0.01 within-species. After this initial filtering, 347,200 SNPs remained, representing the whole-genome SNP set. For the principal component, D/f4 and admixture analyses, we used the whole-genome SNP set.

We performed additional filtering of the SNP set for different analyses to produce a conservative dataset. For the phylogenetic analysis, we used the SNP set generated by mapping all genomes to the European *D. pulex* reference genome and restricted analysis to SNPs surrounding BUSCO genes (Mi et al. 2013; Seppey, Manni, and Zdobnov 2019; Simão et al. 2015). This multi-species BUSCO gene SNP set includes 138,024 SNPs. For the analyses of cophenetic distances, balancing selection and individual genes (e.g., see below BLOP gene), we restricted analyses to genic (3'UTR, 5'UTR, exon and intron) and non-genic (downstream and upstream up to 5 kbps) regions associated with the 6,544 single-copy ortholog genes between European *D. pulex* and North American *D. pulex* identified by *OrthoFinder v5* (Emms and Kelly 2019) for the SNPs that were retained from the lift-Over. Additionally, we only used the SNPs classified in the same manner (i.e., within-species polymorphisms, shared polymorphisms or fixed differences) when mapping to either European and North American *D. pulex* assembly (Figure S1A,B). We show that 88% of SNP classifications are unchanged between assemblies (Figure S1B,C; N = 28,983; Table S2).

A concern for aligning divergent sequences to the same assembly is for reference allele bias to decrease mapping efficiency and cause genotype errors (Günther and Nettelblad 2019). To assess this, we calculated the proportion of alternative and reference allele dosage for heterozygous BUSCO gene SNPs (N = 1,000; 100 bootstraps). On average, SNPs identified in North American or European *D. pulex, D. pulicaria* or *D. obtusa* had approximately the same alternative and reference allele dosage at heterozygous sites, revealing an absence of systematic reference allele bias (Figure S1D). Therefore, we conclude that the data are not systematically biased by mapping reads from non-European *D. pulex* to the European *D. pulex* assembly.

2.5 | Assigning Multi-Locus Genotypes

Every sample was assigned to a multi-locus genotype (MLG) using the *poppr v2.9.3* package (Kamvar, Tabima, and Grünwald 2014) in *R v4.0.3* using the genome-wide SNP set (N = 347,200). Unless otherwise noted, every analysis was subset based on picking a representative sample with the highest coverage for each MLG.

2.6 | Mitochondrial Tree

We annotated the D84A mitochondrial genome using *MITOs v1* (Bernt et al. 2013). We aligned and called SNPs using *bcftools mpileup v1.9* and *bcftools call*. We excluded reads that had low-quality scores ($Q < 20$) and high coverage ($DP > 100$) using *bcftools filter*, and generated consensus FASTA files using *bcftools consensus*. We mapped North American *D. pulex* and *D. pulicaria* to the North American *D. pulex* mitochondrial genome (GenBank accession: NC_000844.1) and mapped both North American and European *D. obtusa* samples to the North American *D. obtusa* mitochondrial genome (GenBank accession: CM028013.1). The mitochondrial sequence of European *D. magna* was used as an outgroup (GenBank accession: NC_026914.1). We assessed sequence similarity using *exonerate v2.4.0* (Slater and Birney 2005) for the 13 protein-coding genes and found high sequence similarity ($> 80\%$), except for *atp8*. Therefore, we assembled trees excluding *atp8*. We then used *mafft v7.475* (Katoh and Standley 2013) to assemble multiple sequence alignments (MSA). We concatenated these MSAs for each gene using *seqkit concat v2.2.0* (Shen et al. 2016) and ran *iqtree2 v2.1.2* with 1,000 bootstraps (Figure S2; Hoang et al. 2018).

2.7 | Estimating Divergence-Time

We used *Snapp v1.6.1* within *Beast2 v2.6.6* to estimate the split-time between-species (Bouckaert et al. 2014). We used two representative individuals with the highest coverage for each species. We used 3,000 randomly sampled BUSCO SNPs, one million iterations and a 10% burn-in. The output tree was time-constrained for the outgroup species, European *D. obtusa*, to 31 million years ago (mya with a confidence interval of 1 mya based upon a genus-wide tree; Chin and Cristescu 2021; Cornetti et al. 2019). We used *Tracer v1.7.1* to quantify convergence and set the generation time to one per year (Rambaut et al. 2018).

2.8 | Hybridization Statistics

We used *ADMIXTURE v1.3.0* (Alexander and Lange 2011), excluding any sites with MAF < 0.01 and thinned every 500 SNPs. We varied the number of *k*-clusters and calculated the cross-validation error for every value of *k*. We quantified the magnitude of introgression using *Dsuite v0.5* (Malinsky, Matschiner, and Svardal 2021) with European *D. obtusa* as the outgroup.

2.9 | Historic N_e, Neutral Models of Incomplete Lineage Sorting and Demographic Inference of Migration

To calculate historic N_e for European and North American *D. pulex*, we ran *MSMC2 v2.1.1* (Schiffels and Wang 2020) with a fixed recombination rate and the --timeSegmentPattern parameter set to $10 * 1 + 15 * 2$ following Nunez et al. (2021). We ran *eSMC v2.0.5* (Sellinger et al. 2020), a software designed for facultatively sexual organisms to estimate historic N_e and the hatching rate from resting eggs per generation (β). *eSMC* was run by chromosome and we scaled the recombination rate of 8×10^{-8}

per generation by 5 to account for five generations of parthenogenesis prior to a hatching event as described in Sellinger et al. (2020).

We tested whether the extent of shared polymorphism can be explained by incomplete lineage sorting (Novikova et al. 2016; Wiuf et al. 2004). The formula in Novikova et al. (2016) uses the expected coalescence times within and between-species to estimate the upper bound of the number of shared polymorphisms at synonymous sites given no migration. This method uses the level of polymorphism within-species to reflect the within-species coalescent time and assumes that the population size (and thus diversity level) of the ancestral species is the same as the larger of the two diversity estimates of the extant species. See Novikova et al. (2016) for a more in-depth derivation of the method. The upper bound of the number of shared polymorphisms given incomplete lineage sorting is

$$\exp\left(-2d_{\text{between}} + \frac{(p_{\text{NAm}} + p_{\text{Euro}})\max\{p_{\text{NAm}}, p_{\text{Euro}}\}}{p_{\text{NAm}}, p_{\text{Euro}}} \right)$$

where d_{between} is synonymous D_{xy} between North American and European *D. pulex* and p_{NAm} and p_{Euro} are the levels of synonymous polymorphism within-species. D_{xy} was calculated using *PopGenome v2.7.5* across all samples for each species (Revell 2019).

We performed demographic inference with *moments v1.1.0* (Jouganous et al. 2017) in *python3*. We tested two models: one with-migration ('*Split + Migration*') and one without-migration ('*Split*'). For the former model, we used *moments' split_mig* model. For the latter, we used *split_mig* with the migration rate fixed at zero. We ran inference on 20×20 SFS projections until model convergence and classified a shared polymorphism as an allele whose allele frequency is above $1/20$ in both species. Thus, the proportion of shared polymorphisms reported in this analysis is less than reported for the larger dataset (Table S2). We chose to use 20×20 SFS projections instead of the whole dataset because McCoy et al. (2014) showed that there is a marginal performance increase past even small projection sizes of 10×10. Also, we wanted to have equal projection sizes between-species and increase computational performance for bootstrapping. We used a fixed mutation rate of $\mu = 5.69 \times 10^{-9}$ (Lynch et al. 2017). We first performed *moments* inference and optimization across 1,000 bootstrap replicates for both the '*Split + Migration*' model and the '*Split*' model to derive optimized priors for parameters, including population sizes, split times and migration rates. Using these optimized priors, we generated 2D SFSs for North American and European *D. pulex* under each bootstrap model. We present the mean projections of the 2D SFS based on the results from 1,000 bootstrap replicates for each model. Mean standardized residuals were calculated from the allele counts for each element of the 2D SFS with the following formula: $\frac{\text{Empirical} - \text{Model}_x}{\sqrt{\text{Model}_x}}$, where Model_x is '*Split + Migration*' or '*Split*'.

2.10 | Balancing Selection Statistics

We used *BetaScan v1* (Siewert and Voight 2017) to calculate the β^1 statistic within-species using the folded SFS. We used the α_b

statistic to estimate the proportion of shared polymorphisms under balancing selection between-species-pairs from Soni, Vos, and Eyre-Walker (2022), where

$$\alpha_b = 1 - \frac{\text{Pol\_Syn} \times \text{SP\_NS}}{\text{Poly\_NS} \times \text{SP\_Syn}}$$

and *Poly_Syn* and *Poly_NS* are within-species synonymous and non-synonymous polymorphisms, respectively, and *SP_Syn* and *SP_NS* are shared polymorphisms between-species at synonymous and non-synonymous sites.

2.11 | Phylogenetic Tree Test Using Pairwise Cophenetic Distances

We tested the local sequence genealogy for trans-specificity versus convergent evolution (Koenig et al. 2019; Nunez et al. 2021). This test used trees built from 500 bps flanking high-frequency non-synonymous shared polymorphisms (MAF > 0.25) and calculated the median pairwise cophenetic distances (CPD) between samples (Cardona et al. 2013). We extracted haplotypes from a *WhatsHap v1.1* phased VCF (Martin, Ebert, and Marschall 2023) from 30 high-coverage individuals for each species. We chose 30 samples to keep the sample size consistent across species whilst decreasing convergence time. We aligned the phased haplotypes ($n = 60$ per species) using *mafft v7.475* and built trees using *iqtree2* (1,000 bootstraps).

2.12 | Light Exposure Experiments on *Daphnia* Activity

We developed a behavioural assay to collect activity data on 12 European *D. pulex* clones using a DAM Trikinetics monitor (TriKinetics Inc., Waltham, MA). In total, we measured activity for 216 individuals. Each well has an infrared light beam that when broken by a *Daphnia* will count as an activity event. We exposed individuals to white, blue and dark lighting conditions using blackout boxes mounted with LEDs (Erickson et al. 2020). Individuals were placed inside a plastic assay tube with media (ASTM; Standard 2007) whilst each monitor collected activity over a 12-h period, sampling every 5-seconds. We excluded measurements during the first hour to allow individuals to adjust to their new environment. For 95% of the 5-s intervals, 0 or 1 beam-breaks were recorded and 99.9% of intervals had four or fewer beam-breaks. Therefore, for each interval, we converted the number of beam-breaks recorded into a binary variable (≥ 1 beam-break vs. 0 beam-breaks) and calculated activity as the fraction of 5-s intervals with more than one beam-break per individual. We modelled activity with a generalized linear mixed effect model using *lme4 v1.1–27.1* (Bates et al. 2015) and performed likelihood ratio tests between the following models:

$$\text{Model 1}: y \sim \text{Light} + \text{Clone} + \text{Block} + \varepsilon$$

$$\text{Model 2}: y \sim \text{Light} + \text{Genotype} + \text{Clone} + \text{Block} + \varepsilon$$

$$\text{Model 3}: y \sim \text{Light} + \text{Genotype} + \text{Light: Genotype} + \text{Clone} + \text{Block} + \varepsilon$$

In these models, *y* is the fraction of intervals with activity, *Light* is the fixed effect of light treatment (white, blue and dark), *Genotype* is the fixed effect of genotype at the opsin (BLOP), *Light:Genotype* is the fixed interaction effect, (1|*Clone*) is the random effect of clone, (1|*Block*) is the random effect of one of the three experimental blocks run over successive weeks and ε is the binomial error with weights equal to the number of 5-s intervals (ca. 7,800). We conducted likelihood ratio tests between each model using the *anova()* function in *R* (Table S3). In addition, we performed an analysis that models elapsed time in the experiment as a fixed effect and includes the individual *Daphnia* as a random effect to account for repeated measures. The results of that analysis are in line with the simple model presented here and are shown in Table S4.

2.13 | BLOP Orthologs

We identified orthologs of the BLOP (*Daphnia11806* in the D84A annotation) by BLASTing the amino acid sequence against the NCBI database using *blastp v2.13.0* (Sayers et al. 2022).

2.14 | Statistics and Visualization

Most analyses were performed using *R v3.6.2–4.0.3* (R Core Development Team 2013). We used the following packages for analysis and visualization: *tidyverse v1.3.1* (Wickham et al. 2019), *ggplot2 v3.3.5* (Villanueva and Chen 2019), *ggtree v2.0.4* (Xu et al. 2022), *ape v5.4-1* (Paradis and Schliep 2019), *patchwork v1.0.1* (Thomas Lin Pedersen 2022), *data.table v1.12.8* (Dowle and Srinivasan 2023), *foreach v1.4.7*, *doMC v1.3.5* (Daniel et al. 2022) and *SeqArray v1.26.2* (Zheng et al. 2017). Principal component analysis (PCA) of SNPs was conducted in *SNPRelate v1.24.0* (Zheng et al. 2012). Linkage disequilibrium (r^2) between adjacent SNPs was calculated using *plink v1.9* (Purcell et al. 2007).

3 | Results

3.1 | Thousands of *Daphnia* Genomes

We first assembled short-read genomic data for 2,321 samples of *D. pulex*, *D. pulicaria* and *D. obtusa* collected from North American and European ponds (Figure 1A). This includes whole genomes published elsewhere, along with known hybrids of North American *D. pulex* and North American *D. pulicaria*, and 93 European samples of *D. pulex, D. obtusa* and *D. pulicaria* reported here for the first time (see Table S1). The SNPs that we used represent within-species SNPs, fixed differences, and shared polymorphisms between these taxa. The vast majority (97%) of the samples we used are either North American or European *D. pulex*, and therefore most of our analysis focuses on these taxa. Because lineages could be clonally derived from a recent common ancestor, each sample was assigned to a multi-locus genotype using the filtered genome-wide SNP set (MLG; Table S1). In all analyses, unless otherwise noted, we restricted to one sample per MLG ($N = 1,173$).
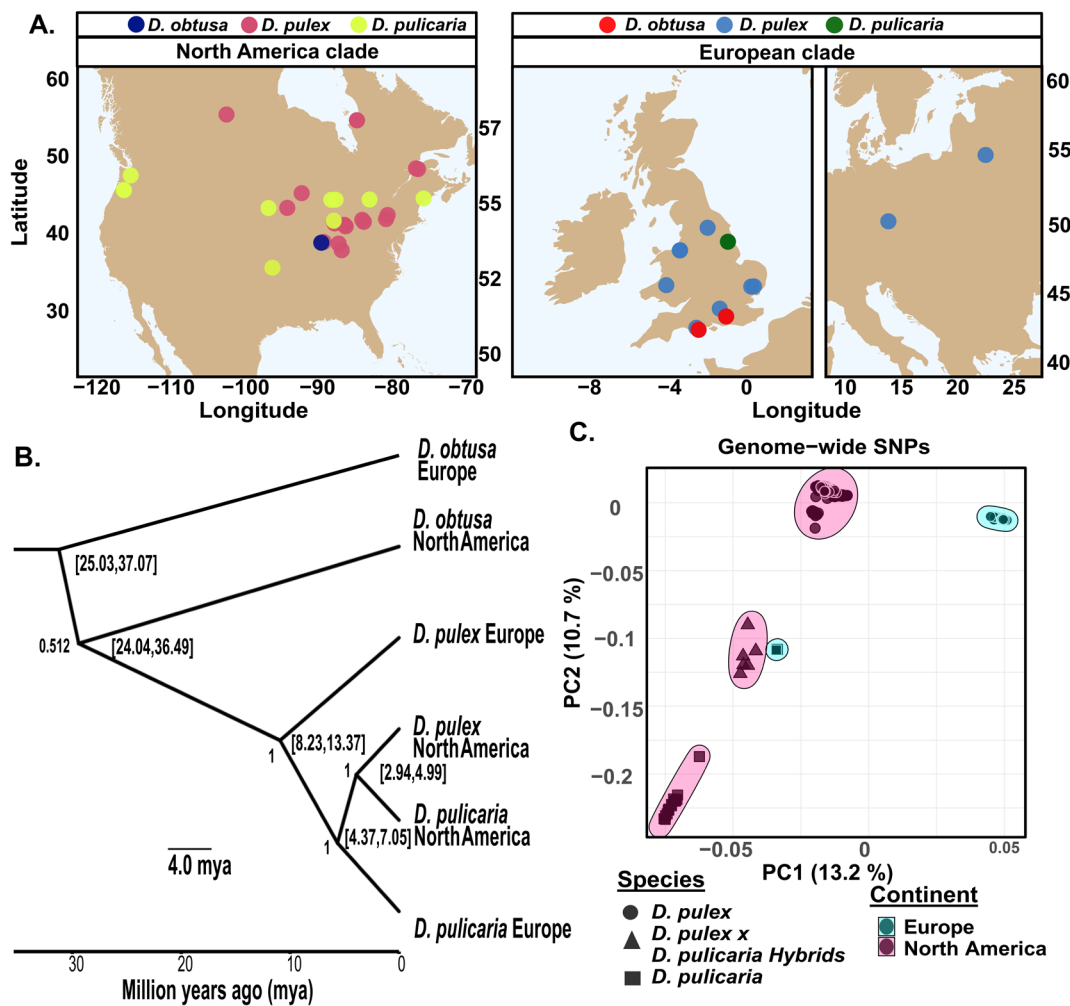
FIGURE 1 | Genetic divergence of the *Daphnia pulex* species complex. (A) Sample origin of the North American and European clades, each consisting of *D. pulex*, *D. pulicaria* and *D. obtusa*. Most of the European clade has samples in the United Kingdom but there is one sample in both the Czech Republic and Lithuania as shown in the rightmost subfigure. (B) Time-constrained phylogenetic tree restricted to two representative individuals within each species based on 3,000 BUSCO gene SNPs. This consensus tree is rooted with European *D. obtusa* to have 31 million years of divergence. Bracketed values are 95% confidence intervals in millions of years ago (mya). Node labels indicate the posterior probabilities estimated from one million bootstraps. (C) The principal component axes (PC1 and PC2) using filtered genome-wide SNPs. The proportion of variation explained by each PC is shown in parentheses. We restricted the PCA to the *D. pulex* and *D. pulicaria* taxa because the *D. obtusa* taxa are so distantly related.

3.2 | North American and European *D. pulex* are Distinct Species

To evaluate the nuclear phylogeny of the *D. pulex* species complex and to determine if North American *D. pulex* and European *D. pulex* are distinct taxa, we built a time-constrained phylogenetic tree using BUSCO gene SNPs (Figure 1B). The tree omitted hybrids between North American *D. pulex* and *D. pulicaria* because they prevented model convergence. The nodes that split the *D. pulex* species complex are generally well-supported. We estimate that the split-time between North American and European *D. pulex* is around 10 million years ago (95% CI 8.23, 13.37; Figure 1B). Previous studies based on mitochondrial and nuclear markers also showed reciprocally monophyletic relationships between North American and European *D. pulex* (Vergilino et al. 2011; Marková et al. 2013). Similarly, our mitochondrial phylogeny supports this reciprocally monophyletic relationship between North American and European *D. pulex*. In contrast, the mitochondrial tree shows that North American

D. pulex and North American *D. pulicaria* are not reciprocally monophyletic (Figure S2). The recent split-time between North American *D. pulex* and *D. pulicaria* (Ye et al. 2022), their propensity to hybridize (Pantel, Juenger, and Leibold 2011) and discordant phylogenies support the hypothesis that North American taxa are in the process of incipient speciation (Heier and Dudycha 2009).

We performed PCA of *D. pulex*, *D. pulicaria* and known hybrids using a genome-wide SNP dataset (Figure 1C). The first and second PC axes are significantly different between the North American and European *D. pulex*, *D. pulicaria* and hybrid species groups (ANOVA PC1: $F_{4,1154} = 70{,}617$, $p < 2 \times 10^{-16}$; ANOVA PC2: $F_{4,1154} = 27{,}940$, $p < 2 \times 10^{-16}$). Intriguingly, European *D. pulicaria* clusters near the known hybrids of North American *D. pulicaria* and *D. pulex* (Jackson et al. 2021; Tucker et al. 2013); below we test whether the samples identified as *D. pulicaria* collected in Europe are related hybrids between North American taxa or are themselves hybrids.

North American and European *D. pulex* possess marked differences in levels of diversity, consistent with long-term divergence. Principal component clusters are more dispersed amongst North American *D. pulex* than they are amongst European *D. pulex*, suggesting higher genetic variability within the North American clade (Figure 1C). Demographic history reconstruction further highlights these distinctions: the effective population size (N_e) for North American *D. pulex* is estimated at approximately 500,000 (95% CI: 487,787, 522,128) over a timescale of 1 kya to 1 mya, compared to ~200,000 (95% CI: 225,775, 235,111) for European *D. pulex* when using *MSMC2* (Figure S3). Estimates of historical population size from *MSMC2* do not incorporate the potential for contribution from the resting egg bank. To account for the contribution from the egg bank, and to estimate the egg bank deposition rate (the inverse of the per-generation germination rate, β) we ran *eSMC* with either 5 or 10 generations of parthenogenesis between every sexual generation (Figure S3). *eSMC* estimates of N_e are similar in scale to one other when assuming 5 generations per year to the *MSMC2* results: North American *D. pulex* (95% CI: 42,600, 43,211) and European *D. pulex* (95% CI: 26,574, 27,039). Under 10 generations per year, the estimates of N_e are increased: North American *D. pulex* (95% CI: 149,228, 152,899) and European *D. pulex* (95% CI: 123,245, 127,387). The *eSMC* analysis also suggests both species have substantial contributions to the egg bank because they have moderate rates of hatching each generation: $\beta_{Euro} = 0.336$ (95% CI: 0.332, 0.34) and $\beta_{NAm} = 0.322$ (95% CI: 0.321, 0.323) under an assumption of 5 generations per year. Doubling the generations to 10 per year increases β, suggesting a hatching rate of $\beta_{Euro} = 0.69$ (95% CI: 0.68,

0.71) and $\beta_{NAm} = 0.65$ (95% CI: 0.644, 0.65). Taken together, our analysis of the genomes supports the conclusion that *D. pulex* found in North America and Europe are different species that likely diverged millions of years ago.

3.3 | Hybridization in the *D. pulex* Species Group

Hybridization is common amongst North American *D. pulicaria* and *D. pulex* species (Pantel, Juenger, and Leibold 2011; Vergilino et al. 2011; Marková et al. 2013), however, signals of hybridization between North American and European *Daphnia* remain less well understood. Hybridization between European *D. pulicaria* and North American or closely related circumarctic species is not recent or is with other members of the complex North American *D. pulex-pulicaria* sub-group. For example, an *ADMIXTURE* analysis reveals that European *D. pulicaria* has distinct clusters from other species, whilst the recent North American *D. pulex-pulicaria* hybrid displays split ancestry between North American *D. pulex* and *D. pulicaria* (Figure 2A; Alexander and Lange 2011). We also examined heterozygosity at fixed differences (i.e., hybrid index) between North American *D. pulex* and North American *D. pulicaria* in European *D. pulicaria* and North American *D. pulex-pulicaria* hybrids. These fixed differences are heterozygotes 70% of the time in North American hybrids, but only 2% of the time in European *D. pulicaria*, suggesting a distinct evolutionary history of the European *D. pulicaria* clade (Figure 2B). In summary, our findings imply that European *D. pulicaria* is likely a member of the speciose North
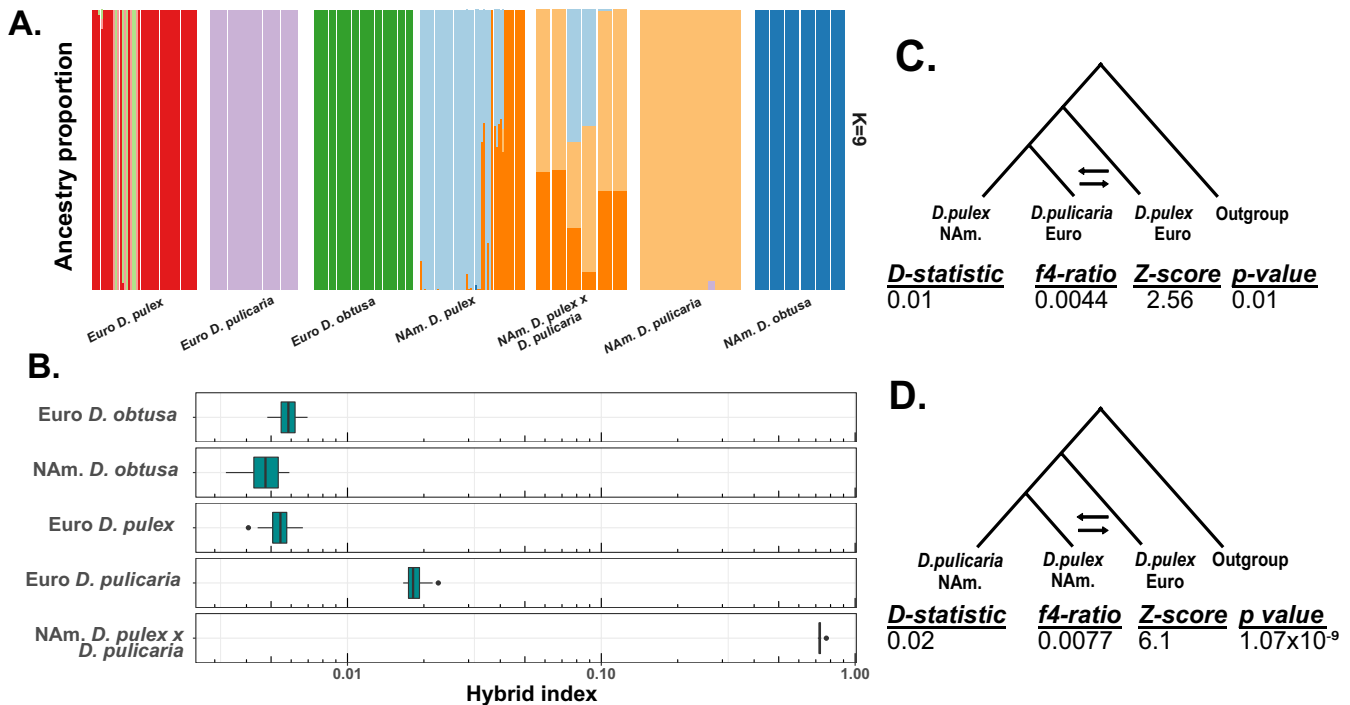

FIGURE 2 | Hybridization across the *D. pulex* species complex. (A) *ADMIXTURE* plot of the *D. pulex* species complex with $k = 9$ having the minimal cross-validation error. Each color represents a unique ancestry group for each sample. (B) We identified fixed differences between North American *D. pulex* and *D. pulicaria* and calculated the hybrid index (i.e., the proportion of fixed differences between North American *D. pulex* and *D. pulicaria* that are heterozygous in any individual of the tested species) in a randomly chosen individual from the remaining taxa. The boxplot shows the distribution of the hybrid index from randomly sampled clones (one per MLG). (C, D) Introgressions tests using various four-species trees. The outgroup is European *D. obtusa* in all tests. *D-statistic* and *f4-ratio* describe the extent of introgression between the 2nd and 3rd taxa on the tree being tested.

American *D. pulex* species sub-group, consistent with previous reports of a circumarctic *D. pulex* lineage predominant across Northern Eurasia (Colbourne et al. 1998).

Signals of hybridization are weak between European *D. pulex* and other taxa. For instance, signals of admixture between European *D. pulex* and European *D. pulicaria* are weak ($D = 0.01$, *f4-ratio* $= 0.0044$, $p = 0.01$; Figure 2C) as are signals of admixture between European *D. pulex* and North American *D. pulex* ($D = 0.02$, *f4-ratio* $= 0.0077$, $p = 1.07 \times 10^{-9}$; Figure 2D). *ADMIXTURE* analysis suggests that European *D. pulex* forms several distinct ancestry groups that do not appear within any species of North American *Daphnia* (Figure 2A). Only ~0.5% of fixed differences between North American *D. pulex* and North American *D. pulicaria* segregate as heterozygous sites in European *D. pulex* (Figure 2B). These results suggest that European *D. pulex* are distinct from the remaining taxa and likely do not have a recent history of hybridization with the other species studied.

3.4 | Extent of Shared Polymorphism Between North American and European *D. pulex* is Not Explained by Incomplete Lineage Sorting or Recent Migration

For species with deep split-times and low levels of migration or hybridization, we expect few shared polymorphisms to exist if such polymorphisms are neutral. For instance, based on a simple neutral model with no migration (Novikova et al. 2016; see Section 2) we expect to observe 336 shared polymorphisms at synonymous sites given the split-time between North American and European *D. pulex*. Yet, we observe at least 11,000 shared synonymous SNPs between these species (Table S2).

This prediction does not account for historic migration, so we performed demographic inference on the two-dimensional site-frequency spectrum (2D SFS) using *moments* (Figure 3A; Jouganous et al. 2017). First, we contrasted two models, one that allows constant migration (*Split + Migration*) and one where the migration rate was set to zero after population divergence (*Split*). The 'Split + Migration' model is the best model based on the mean Bayesian information criteria (BIC; 'Split + Migration' BIC = 20,637 and 'Split' BIC = 33,216). Notably, the 'Split' model severely underpredicts the number of shared polymorphisms, reflecting that incomplete lineage sorting alone is insufficient to explain the abundance of shared SNPs. The 'Split + Migration' model itself underpredicts the number of shared polymorphisms by 25% (Figure 3A), and the model prediction shows a notable deficit of common shared SNPs and an excess of shared SNPs that are at low frequencies compared to the empirical SFS (Figure 3B).

3.5 | Selective Forces Acting on Shared Polymorphisms

European and North American *D. pulex* share more polymorphisms than expected by neutral, demographic models, suggesting selection may maintain these polymorphisms. We aimed to differentiate old, trans-specific polymorphisms versus
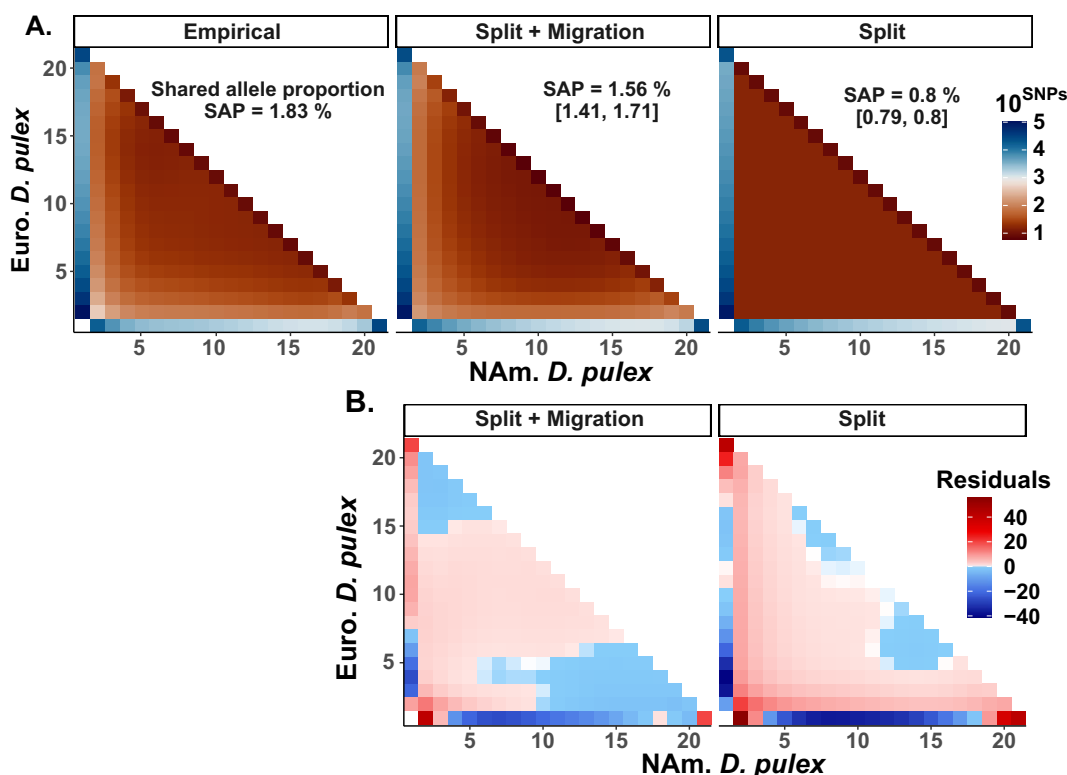
FIGURE 3 | An excess of shared polymorphisms between North American and European *D. pulex*. (A) Demographic model inference between North American and European *D. pulex* based on the folded site-frequency spectrum (SFS). The empirical SFS is constructed from the genome-wide SNP dataset. The split with migration ('Split + Migration') and split without migration ('Split') models were generated from *moments* and we are showing the mean projection based on 1,000 bootstraps. The x- and y-axis use a 20 × 20 SFS projection. (B) Average standardized residuals for both models tested against the empirical SFS.

convergently evolved polymorphisms by constructing phylogenetic trees for genes with shared polymorphisms. If convergent evolution occurred, gene-trees would match the species-tree, with haplotypes from the same species clustering together (Figure 4). However, in genes with old trans-specific polymorphisms, gene-trees might cluster by allele type, not species. However, for this to occur there must be multiple trans-specific polymorphisms in close linkage as a single trans-specific polymorphism would not be sufficient for the gene-tree to differ from the species tree (Gao, Przeworski, and Sella 2015). Our analysis therefore can highlight genes with multiple linked trans-specific polymorphisms, which may be targets of long-term balancing selection.

We summarized the topology of gene-trees by calculating the pairwise CPD (Cardona et al. 2013) within and between haplotypes of the same-species from gene-trees that contain high-frequency (MAF > 0.25), non-synonymous, shared polymorphisms. When gene-trees resemble the species-tree topology, the within-species distances will be lower than the between-species distances ($CPD_{Within} - CPD_{Between} = CPD_{Within-Between} < 0$; Figure 4A). However, if alleles from two species cluster together, and are discordant with the species-tree, the within-species distances will be larger than between-species distances ($CPD_{Within-Between} > 0$). A small number of gene-trees surrounding shared polymorphisms have a positive $CPD_{Within-Between}$ value, consistent with old trans-specific haplotypes (Figure 4A). However, most shared polymorphisms have negative $CPD_{Within-Between}$ values (Figure 4B), consistent either with convergent evolution or trans-specificity.

Regardless of whether shared polymorphisms arose via convergent evolution, or prior to the species-split, they could have been subject to balancing selection. To test this hypothesis, we first calculated α_b, a statistic to estimate the proportion of
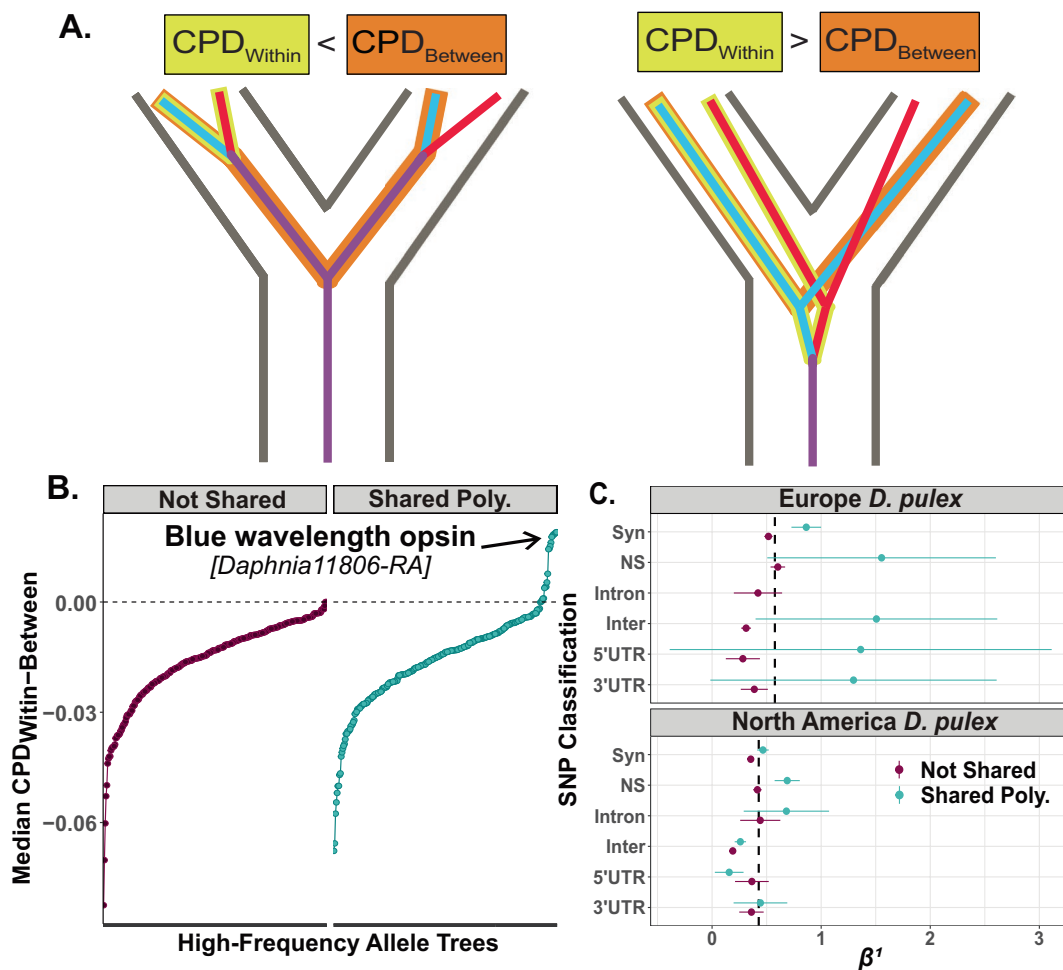

FIGURE 4 | Convergent evolution, trans-specificity and signatures of balancing selection. (A) Visualization of two adaptive hypotheses that produce shared polymorphisms, convergent evolution on the left and trans-specificity on the right. For each tree, we calculated the median pairwise cophenetic distance as the distance within-species (CPD_{Within}; yellow highlighted pair) – between-species ($CPD_{Between}$; orange highlight) for shared polymorphisms and non-shared polymorphisms. $CPD_{Within-Between} < 0$ describes the consensus species-tree topology (Left), whilst $CPD_{Within-Between} > 0$ describes an allele-specific tree topology consistent with multiple linked trans-specific polymorphisms (Right). The red and blue branches indicate examples of shared polymorphisms between-species. (B) $CPD_{Within-Between}$ for non-synonymous shared SNPs and non-shared SNPs above 0.25 minor allele frequency (MAF) in both species. Each allele-tree was made from 30 samples from North American and European *D. pulex*. At the focal SNP, we extracted 500 bps surrounding the focal SNP. (C) β^1 is a statistic that detects balancing selection. We show the mean with 95% standard errors for several SNP classifications (SYN = synonymous, NS = non-synonymous, Intron = intronic, Inter = intergenic, ±500 bps upstream and downstream variants, 5′ UTR = 5′ untranslated region, 3′ UTR = 3′ untranslated region). The dotted vertical line is the average β^1 within each species.

non-synonymous sites under balancing selection using counts of species-specific alleles and shared polymorphisms (Soni, Vos, and Eyre-Walker 2022). We found that α_b is significantly positive across the genome, indicating that balancing selection is influencing non-synonymous shared polymorphisms ($\alpha_b = 0.082$ [0.05, 0.114], $p = 1.5 \times 10^{-6}$). Next, we calculated β^1, an SFS-based statistic for detecting signals of balancing selection (Siewert and Voight 2017) at both shared and control SNPs. We found that β^1 at shared polymorphisms is significantly larger than zero in both species for non-synonymous SNPs (one sample t-test: Euro. $t = 7.8$, $df = 270$, $p = 1.75 \times 10^{-13}$; NAm. $t = 18.8$, $df = 1,563$, $p = 2.2 \times 10^{-16}$; Figure 4C). Shared synonymous sites also show significantly elevated β^1 in both species (NAm. $t = 25.12$, $df = 4,367$, $p = 2.2 \times 10^{-16}$; Euro. $t = 20.41$, $df = 1,481$, $p = 2.2 \times 10^{-16}$; Figure 4C).

3.6 | Trans-Specific Polymorphisms at an Opsin Affects Behavior and Shows Evidence of Genetic Overdominance

Of the common, non-synonymous shared polymorphisms, 14 (5%) have positive CPD$_{\text{Within-Between}}$ values (Figure 4B). Almost all of these shared polymorphisms (13/14) are within a rhabdomeric blue wavelength opsin (BLOP) gene (Brandon, Greenwold, and Dudycha 2017). BLOP is found as a single-copy in European and North American *D. pulex* (Figure S4A). The 13 non-synonymous shared SNPs reside across several exons (Figure 5A) and encompass a large linkage block within European *D. pulex* ($r^2 > 0.7$ across ~1.5 kbps; Figure 5B,C), thereby explaining the gene-tree species-tree discordance (Figures 4B and 5C) and suggest that these alleles are trans-specific polymorphisms that predate the split between North American and European *D. pulex*.

If these haplotypes at the BLOP have been maintained prior to the split between North American and European *D. pulex* 10 mya (Figure 1C), they may have a functional effect. To test this hypothesis, we measured the light-induced activity of European *D. pulex* clones that harbour distinct haplotypes bearing alternate shared alleles. We first assigned clonal haplotypes to one of two genetic clusters (Figure 5C) and tested the activity levels of all three genotypes (AA, AB and BB) in different light conditions. We found that genotype has a significant effect on activity that is dependent on light conditions ($\chi^2 = 5,849.71$, $df = 4$, $p < 2 \times 10^{-16}$; Table S3). In general, all genotypes had low activity in dark conditions. Heterozygotes have the highest activity levels when exposed to white light yet have the lowest activity when exposed to blue light consistent with shifts between genetic overdominance and underdominance affecting behavior (Figure 5D).

Overdominance affecting behavior could also translate into overdominance affecting fitness. If trans-specific polymorphisms at the BLOP cause overdominance in fitness, heterozygotes should be more common than expected under Hardy–Weinberg equilibrium (HWE). To test the hypothesis, we examined segregation patterns of trans-specific SNPs at the BLOP amongst F1 offspring derived from a cross between two clones that are both heterozygous for the trans-specific haplotypes we identified. These clones were previously referred to as 'super-clone A' and 'super-clone C' by Barnard-Kubow et al. (2022). Both clones had reached high frequency in the southern English (Dorset) pond

D8 by the end of the 2017 growing season. In 2018, most individuals in the D8 pond were the F1 offspring between super-clone A and C (74%) enabling us to directly test if there is an excess of heterozygotes relative to the expected Mendelian segregation patterns amongst the F1s. First, we calculated the frequency of AA, AB and BB genotypes at the BLOP locus compared to the rest of the genome, without downsampling to one clone per MLG. In this way, we are able to see how all samples collected were affected by selection at the BLOP locus. We find that there is an excess of heterozygotes in the wild-caught individuals compared to expectations from HWE and compared to random SNPs in the genome or other trans-specific polymorphisms (Figure 5E). We calculated the distribution of F_{IS}, a measure of the departure of HWE, at genes across the genome and found that the BLOP gene is amongst the most strongly negative F_{IS} compared to other genes ($F_{IS} = -0.54$; Figure 5E). The BLOP gene has amongst the smallest 2.6% of F_{IS} values. Even if we examine the genotype distribution by only sampling one individual per MLG, we still observe an excess of heterozygotes (Figure S5C–F). Finally, we examined genotype frequencies in lab-generated A×C and C×C F1s. In contrast to our field-sampled individuals, we do not observe an excess of heterozygotes from a lab-generated cross of the same clones (Figure S5A–E).

4 | Discussion

In this study, we examined the evolutionary forces that generate and maintain shared polymorphisms between two taxa in the *D. pulex* species complex. We used whole-genome sequences coupled with polymorphism data to resolve the nuclear phylogeny of members of this species group and to evaluate mechanisms that can generate shared polymorphisms between-species. We show that there is an excess of shared polymorphisms between North American and European *D. pulex* that cannot be explained by neutral or demographic processes, thereby implicating some form of natural selection as a force maintaining polymorphism. For one gene, a blue wavelength opsin, we show that shared polymorphisms are ancient, predating speciation and has functional consequences on behavior and fitness in the wild.

4.1 | Phylogenetics of the *D. pulex* Species Group

Members of the genus *Daphnia*, and the *D. pulex* species group in particular, have proven challenging from a taxonomic perspective since their early description. For instance, Leydig separated *D. pulex* from *D. magna* and *D. longispina* (Leydig 1860, p. 117), but did not further describe divisions. Richard (1896), identified *D. obtusa* as a distinct species from *D. pulex* (p. 260) but also described ten subspecies of *D. pulex* found across the Americas and Eurasia (p. 232–255). Scourfield (1942), reinforced the view that *D. obtusa* and *D. pulex* are distinct species and emphasized that this species group represents several lineages in various stages of speciation. Johnson (1951), in his description of British members of the *D. pulex* group, noted that American forms resembling species in the *D. pulex* group are not likely monophyletic with Eurasian species of the same name, although these taxonomic naming conventions have persisted (e.g., Brooks 1957b; Omilian and Lynch 2009; Ye, Pfrender, and Lynch 2023). The challenge of morphological classification in the *D. pulex* group
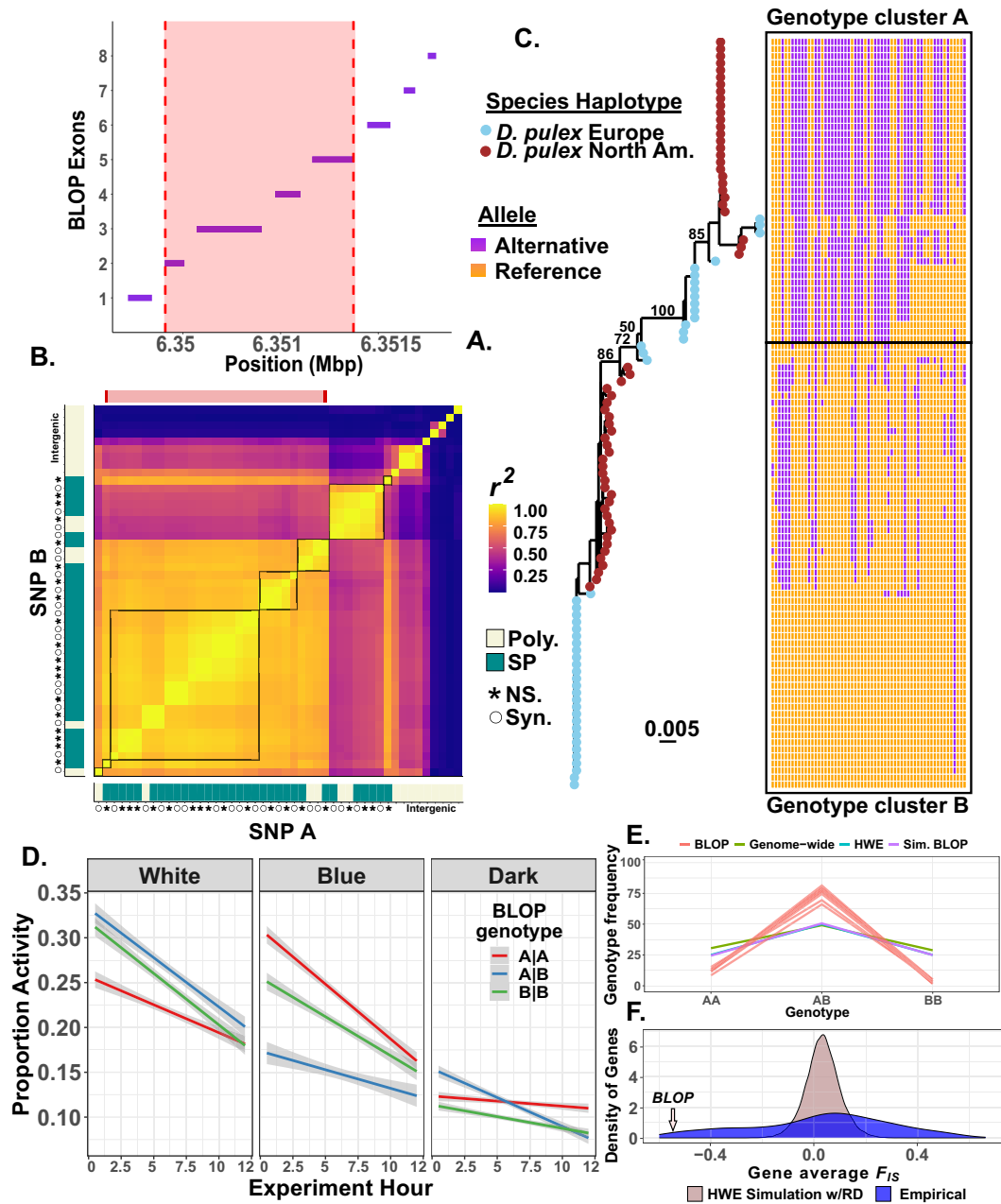
FIGURE 5 | Behavioral and fitness effects of trans-specific SNPs at a blue wavelength opsin. (A) Gene structure showing the length and position of exons within BLOP (*Daphnia11806*). The shaded red region indicates the location of a large high-linkage block identified in panel B. (B) Pairwise linkage disequilibrium (r^2) for every SNP within BLOP for European *D. pulex*, filtered for SNPs with a MAF > 0.01. The right and bottom tile objects indicate whether the SNP is polymorphic (Poly; khaki) or a shared polymorphism (SP; blue-green). NS refers to non-synonymous polymorphism and Syn refers to synonymous polymorphism represented by asterisks and open circles respectively. The grey boxes indicate within-exon SNP comparisons, exon 8 did not have any relevant SNPs. (C) Allele-tree made from the gene for a subset of phased samples of North American and European *D. pulex*. Tip symbols indicate whether the samples are North American or European *D. pulex*. Numbers indicate bootstrap support. The included haplotype plot and multiple-sequence alignment showcase the presence of each SNP within the gene, colored for whether the allele is derived (purple) or reference (gold). (D) The activity of individual European *D. pulex* was measured for 12 h for three genotypes in three different light conditions. Lines represent the best fit and 95% standard errors. (E) Average frequency of F1 genotypes expected based on a double heterozygous cross (i.e., AB×AB) using empirical read depth at each SNP. 'BLOP' is the empirical segregation of trans-specific polymorphisms within the BLOP gene - amongst F1 genotypes. 'Genome-wide' is the segregation for SNPs based on the read depth. 'HWE' is the segregation pattern expected for Hardy Weinberg equilibrium. 'Sim. BLOP' is the segregation pattern expected for the SNPs within BLOP based on empirical read depth. (F) Distribution of average gene F_{IS}. 'HWE Simulation w/RD' is the expected F_{IS} for each gene based on the empirical read depth for each SNP within every gene and 'Empirical' is the average F_{IS} across genes. The small arrow denotes where the gene average for the blue wavelength opsin falls along the empirical distribution.

stems from a limited number of diagnostic characteristics (Brooks 1957b; Dodson 1981), coupled with phenotypic plasticity (Colbourne et al. 1997), intraspecific mating type variation (Heier and Dudycha 2009; Jose and Dufresne 2010) and cytological variation (Gómez et al. 2016; Hosseinie 1966). However, recent phylogenetic analysis of mitochondrial markers has shown the *D. pulex* group consists of many distinct lineages and that the deepest splits within the *D. pulex* species group occur between Eurasian and North American taxa (Crease et al. 2012; Ye et al. 2022). Consistent with these results, allopatric speciation has been estimated to account for roughly 40% of cladogenetic events within *Daphnia* (Adamowicz et al. 2009), a process possibly enhanced by cycles of glaciation (Chin and Cristescu 2021). We show that substantial genetic division exists between North American and European taxa and that these taxa are separated by millions of years (Figure 1). Given the deep split time between members of the *D. pulex* species group, it is likely that they have distinct features ranging from their response to environmental stimuli to their impact on the ecosystem. Further study of the behavioral, physiological and ecological interactions that distinguish these taxa is warranted.

The complicated nature of the *D. pulex* species group is compounded by incomplete reproductive isolation. North American *D. pulex* and North American *D. pulicaria* are known to hybridize in the wild (Xu et al. 2015; Ye et al. 2019). Hybrids between these lineages are obligately asexual and fail to produce functional males (Tucker et al. 2013; Xu et al. 2015; Ye et al. 2019). These post-zygotic reproductive incompatibilities are a hallmark of taxa undergoing incipient speciation (Coughlan and Matute 2020). Consistent with this view, we show that the split-time between North American *D. pulex* and *D. pulicaria* based on the nuclear genome is recent, within 3 mya (Figure 1B). Our estimate is consistent with a study made from mitochondrial genomes (Colbourne et al. 1998; Marková et al. 2013), but older than another using a limited number of nuclear markers (Omilian and Lynch 2009). Nonetheless, genomic data clearly show that hybridization between these North American lineages occurs (Xu et al. 2015; Ye et al. 2019; Figure 2). Previous analysis of mitochondrial markers placed European *D. pulicaria* as a sister to the North American *D. pulex/pulicaria* clade (Marková et al. 2013), a result consistent with the nuclear phylogeny we constructed (Figure 1B). European *D. pulicaria* also shows evidence of hybridization with members of the North American *D. pulex/pulicaria* clade (Figure 2B), although such hybridization is not likely recent or could have occurred with other lineages in this complex that are not represented here. Although the North American taxa, along with European *D. pulicaria* show signals of hybridization with each other, European *D. pulex* appears to be a well-defined species. We show that European *D. pulex* split from the other *D. pulex/pulicaria* taxa approximately 10 mya (Figure 1B) and has little to no evidence of recent hybridization (Figure 2B–D). More extensive sampling of the *D. pulex* species group across the Northern hemisphere will be necessary to more exhaustively test complex patterns of hybridization and introgression.

4.2 | The Evolution of Shared Polymorphisms

Polymorphisms that are shared between-species represent a particularly interesting class of mutations because they can reflect a wide variety of evolutionary processes. On the one

hand, shared polymorphisms could reflect neutral processes when they occur between closely related species. For example, species that diverged relatively recently will share many neutral polymorphisms because there has been insufficient time for drift to cause fixation or loss (Hobolth et al. 2011), and because of ongoing gene-flow (Payseur and Rieseberg 2016). Whilst the presence of neutral shared polymorphisms due to incomplete lineage sorting or gene-flow is important for understanding features such as historical population size (Suh, Smeds, and Ellegren 2015) or barriers to migration (Kutschera et al. 2014), they can obscure selective forces such as convergent adaptive evolution or balancing selection that can also generate or maintain shared polymorphism. Therefore, to examine these selective forces, it is important to identify species that have diverged long enough ago that incomplete lineage sorting and ongoing gene-flow are limited. Our work identifies European and North American *D. pulex* as two such species because of their relatively deep split-time and limited evidence for hybridization.

We show that there are tens of thousands of polymorphisms that are shared between European and North American *D. pulex* (Figure 3A, Table S2) and suggest that natural selection is responsible for their presence. Natural selection has often been implicated as playing a key role in maintaining shared polymorphism. For instance, immune-related genes of invertebrates (Cornetti et al. 2024), vertebrates (Aguilar et al. 2004; Azevedo et al. 2015; Klein et al. 1993) and plants (Klein et al. 1998; Sutherland, Tobutt, and Robbins 2008; Novikova et al. 2016) are routinely found to possess trans-specific polymorphisms that are older than the species split and are thought to be maintained as polymorphic via mechanisms such as negative frequency dependence or genetic overdominance (Key et al. 2014). In other cases, shared polymorphisms in a variety of taxa have possibly arisen via convergent evolution to common selective pressures such as pathogens (Těšický and Vinkler 2015) and have been subject to balancing selection in both species (Solberg et al. 2008).

The shared non-synonymous polymorphisms that we identified have gene-trees that largely reflect the species-tree (Figure 4). Taken at face value, this result is consistent with convergent evolution. Others have suggested that widespread convergent evolution is an unlikely mechanism generating shared polymorphisms (Klein et al. 1993). Is this conclusion valid for *Daphnia*? The probability of a beneficial mutation arising in a population is a function of its census size (Pennings and Hermisson 2006) and its establishment in a population is a function of the selective value of the mutation (Haldane 1927). Whilst the long-term effective population size of both European and North American *D. pulex* is somewhat limited ($N_e < 1$ million; Figure S3), the census size at any single pond or lake can be quite large, possibly reaching into the millions of individuals (Dudycha 2004), whilst the global census size of either species can reach upwards of 10^{12} individuals (Buffalo 2021). Therefore, across the species range, these taxa are not likely mutation-limited. Recurrent *de novo* evolution of beneficial mutations has been hypothesized to occur rapidly and contribute to within-population variation in male production rates (Barnard-Kubow et al. 2022) and morphological responses to predators (Becker et al. 2022).

Positive and fluctuating selection have been inferred to drive adaptive differentiation amongst and within *Daphnia* populations (e.g., Barnard-Kubow et al. 2022; De Meester, Boersma, and Spaak 1999; Reger et al. 2018; Yampolsky, Schaer, and Ebert 2014) and balancing selection has been inferred to maintain heterozygosity within-populations, suggesting that selection on new beneficial mutations could be strong enough to prevent beneficial mutations from being lost (Chaturvedi et al. 2021; Flynn et al. 2017; Lynch 1987; Lynch et al. 2023). Taken together, it is conceivable that such shared polymorphisms between North American and European *D. pulex* arose independently. On the other hand, it is possible that many of these shared polymorphisms are old and predate speciation, even if the gene-trees surrounding the shared polymorphism are consistent with the species-tree. For instance, if a SNP is a trans-species polymorphism and is old, then the genetic footprint surrounding it will become eroded due to recombination (Gao, Przeworski, and Sella 2015) and the signature of trans-specificity will be lost. Regardless, the shared polymorphisms that we observe show signatures of balancing selection compared to non-shared polymorphisms in the genome (Figure 4C), suggesting that even if the shared mutations arose via convergent evolution many are relatively old. Moreover, theoretical work has shown that balanced polymorphisms in facultatively sexual organisms have especially deep coalescent times, compared to obligately sexual organisms (Agrawal and Hartfield 2016), consistent with the signatures of elevated polymorphism that we observe using the β^1 statistic. Furthermore, the ratio of non-synonymous to synonymous polymorphisms is higher for shared polymorphisms (0.58) than non-shared polymorphisms in both North American and European *D. pulex* (0.53 and 0.46, respectively; Table S2). Therefore, it is likely that many shared polymorphisms are functional and subject to balancing selection.

4.3 | Natural Selection Maintains Functional Trans-Specific Polymorphisms in a Blue Wavelength Opsin Gene

We show that one gene, a blue wavelength opsin harbours trans-specific mutations that predate the split between North American and European *D. pulex* (Figures 4B and 5C). At this locus, the gene-tree differs from the species-tree, a signal that is consistent with trans-specific polymorphism (Charlesworth 2006; Fijarczyk and Babik 2015). This BLOP gene has 15 non-synonymous trans-specific polymorphisms, extensive heterozygosity (Figure S4B) and elevated linkage disequilibrium making it a high-priority candidate for functional characterization.

Research into the North American *D. pulex* genome has shown ancient expansion of opsin genes that occurred over 145 mya (Brandon, Greenwold, and Dudycha 2017). Recent work showcases distinct selective pressures between North American *D. pulex* and *D. pulicaria* at opsins highlighting the complex patterns of selection acting upon them (Ye, Pfrender, and Lynch 2023). It could be that this blue wavelength opsin mediates behavioral responses like predator avoidance or vertical diel migration (Li et al. 2022). Our laboratory experimental work shows that alternate genotypes at the BLOP have variable behavioral activity patterns in response to different light conditions (Figure 5D). Indeed, it even appears that there are changes in dominance as a function of light treatment, a feature that is consistent with the long-term persistence of balanced polymorphisms (Wittmann et al. 2017).

Our observations of a natural population in Dorset, England identified a fitness advantage of heterozygotes in the wild (Figure 5E), consistent with previous work in *Daphnia* (Haag and Ebert 2007; Hebert, Ferrari, and Crease 1982). Our result relies on temporal sampling of a single wild population along with the reconstruction of the pedigree using genomic data of individuals (Barnard-Kubow et al. 2022). Barnard-Kubow et al. (2022), show that two clones became dominant in a pond by the end of 2017 and then crossed with each other, producing a population of F1 offspring the following year. The two dominant clones were heterozygous for the trans-specific polymorphisms at the BLOP and thus we expect their offspring to follow a simple Mendelian 1:2:1 ratio. In contrast, we observe an excess of heterozygous individuals at 81% (35/43 isofemale clones). This pattern is largely explained by heterozygous clones reaching higher frequency in the population by the time they were sampled, suggesting that heterozygotes had higher fitness and thus were more likely to survive. By contrasting genotype frequencies in F1s from the field and the lab (Figure S5), we conclude that the excess of heterozygotes in the field is not likely due to factors such as inbreeding depression or associative overdominance (Ohta 1971). Instead, the excess of heterozygotes in the field likely emerged due to the action of natural selection. Given the strong link between looming stimulus, movement and predator avoidance in *Daphnia* (Pijanowska and Kowalczewski 1997; Ringelberg 1999; Van Gool and Ringelberg 2003), we hypothesize that trans-specific polymorphisms at the BLOP locus may play a role in conferring a fitness advantage by reducing encounters with predators or by facilitating migration through the water column.

5 | Conclusion

Our study elucidates the evolutionary history and genetic structure of the *D. pulex* species complex and provides evidence that shared polymorphisms are common between North American and European *D. pulex*. We show that balancing selection broadly influences shared polymorphisms and that a small fraction predates the species-split. We demonstrate the functional significance of shared polymorphisms across specific ecological contexts and show that polymorphisms at a blue wavelength opsin gene are associated with fitness in the wild. Whilst we present four hypotheses related to the origin and maintenance of shared polymorphism (hybridization, incomplete lineage sorting, convergence and balancing selection), these hypotheses are not mutually exclusive. Additionally, the evolutionary mechanisms presented as hypotheses will all be affected by background levels of recombination, historic shifts in N_e and patterns of positive and purifying selection acting upon the genome (Charlesworth 2009, 2006). Despite this challenge, we laid the groundwork for understanding the mechanisms by which genetic diversity is maintained between members of the *D. pulex* species group.

Author Contributions

Connor S. Murray: conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, visualization, writing – original draft, writing – review and editing. **Madison Karram:** methodology, investigation, writing – review and editing. **David J. Bass:** formal analysis, software, visualization, writing – review and editing. **Madison Doceti:** investigation, writing – review and editing. **Dörthe Becker:** investigation, resources, writing – review and editing. **Joaquin C. B. Nunez:** formal analysis, software, methodology, writing – review and editing. **Aakrosh Ratan:** formal analysis, software, writing – review and editing. **Alan O. Bergland:** conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, software, supervision, validation, visualization, writing – review and editing.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The D84A mitochondrial genome was uploaded to NCBI under accession number JAHCQT000000000. The 93 new genomes described here are available on NCBI under the accession: PRJNA982532. The metadata for samples are in Table S1. The VCF and GDS are deposited on dryad: https://doi.org/10.5061/dryad.wwpzgmsv9. Scripts and data are deposited on GitHub: https://github.com/connor122721/SharedPolymorphismsDaphnia.

References

Adamowicz, S. J., A. Petrusek, J. K. Colbourne, P. D. N. Hebert, and J. D. S. Witt. 2009. "The Scale of Divergence: A Phylogenetic Appraisal of Intercontinental Allopatric Speciation in a Passively Dispersed Freshwater Zooplankton Genus." *Molecular Phylogenetics and Evolution* 50, no. 3: 423–436. https://doi.org/10.1016/j.ympev.2008.11.026.

Agrawal, A. F., and M. Hartfield. 2016. "Coalescence With Background and Balancing Selection in Systems With Bi-and Uniparental Reproduction: Contrasting Partial Asexuality and Selfing." *Genetics* 202, no. 1: 313–326. https://doi.org/10.1534/genetics.115.181024.

Aguilar, A., G. Roemer, S. Debenham, M. Binns, D. Garcelon, and R. K. Wayne. 2004. "High MHC Diversity Maintained by Balancing Selection in an Otherwise Genetically Monomorphic Mammal." *Proceedings of the National Academy of Sciences of the United States of America* 101, no. 10: 3490–3494. https://doi.org/10.1073/pnas.0306582101.

Alexander, D. H., and K. Lange. 2011. "Enhancements to the ADMIXTURE Algorithm for Individual Ancestry Estimation." *BMC Bioinformatics* 12: 246. https://doi.org/10.1186/1471-2105-12-246.

Alzbutas, G. 2015. "Genome Sequencing of European *Daphnia pulex*." NCBI Sequence Read Archive. PRJNA302704. https://www.ncbi.nlm.nih.gov/sra.

Azevedo, L., C. Serrano, A. Amorim, and D. N. Cooper. 2015. "Trans-Species Polymorphism in Humans and the Great Apes Is Generally Maintained by Balancing Selection That Modulates the Host Immune Response." *Human Genomics* 9, no. 1: 21. https://doi.org/10.1186/s40246-015-0043-1.

Barnard-Kubow, K. B., D. Becker, C. S. Murray, et al. 2022. "Genetic Variation in Reproductive Investment Across an Ephemerality Gradient in *Daphnia pulex*." *Molecular Biology and Evolution* 39, no. 6: msac121. https://doi.org/10.1093/molbev/msac121.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67, no. 1: 1–48. https://doi.org/10.18637/jss.v067.i01.

Baym, M., S. Kryazhimskiy, T. D. Lieberman, H. Chung, M. M. Desai, and R. Kishony. 2015. "Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes." *PLoS One* 10, no. 5: e0128036. https://doi.org/10.1371/journal.pone.0128036.

Becker, D., K. Barnard-Kubow, R. Porter, et al. 2022. "Adaptive Phenotypic Plasticity Is Under Stabilizing Selection in *Daphnia*." *Nature Ecology & Evolution* 6, no. 10: 1449–1457. https://doi.org/10.1038/s41559-022-01837-5.

Bergland, A. O., E. L. Behrman, K. R. O'Brien, P. S. Schmidt, and D. A. Petrov. 2014. "Genomic Evidence of Rapid and Stable Adaptive Oscillations Over Seasonal Time Scales in *Drosophila*." *PLoS Genetics* 10, no. 11: e1004775. https://doi.org/10.1371/journal.pgen.1004775.

Bernt, M., A. Donath, F. Jühling, et al. 2013. "MITOS: Improved de novo metazoan mitochondrial genome annotation." *Molecular Phylogenetics and Evolution* 69, no. 2: 313–319. https://doi.org/10.1016/j.ympev.2012.08.023.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30, no. 15: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Bouckaert, R., J. Heled, D. Kühnert, et al. 2014. "BEAST 2: A Software Platform for Bayesian Evolutionary Analysis." *PLoS Computational Biology* 10, no. 4: 1–6. https://doi.org/10.1371/journal.pcbi.1003537.

Brandon, C. S., M. J. Greenwold, and J. L. Dudycha. 2017. "Ancient and Recent Duplications Support Functional Diversity of *Daphnia* Opsins." *Journal of Molecular Evolution* 84, no. 1: 12–28. https://doi.org/10.1007/s00239-016-9777-1.

Brooks, J. L. 1957a. "The Species Problem in Freshwater Animals." In *The Species Problem*, 81–123. Washington DC: American Association for the. Advancement of Science.

Brooks, J. L. 1957b. "The Systematics of North American *Daphnia*: Memoirs of the Connecticut Academy of Arts & Science."

Buffalo, V. 2021. "Quantifying the Relationship Between Genetic Diversity and Population Size Suggests Natural Selection Cannot Explain Lewontin's Paradox." *eLife* 10: e67509. https://doi.org/10.7554/eLife.67509.

Cardona, G., A. Mir, F. Rosselló, L. Rotger, and D. Sánchez. 2013. "Cophenetic Metrics for Phylogenetic Trees, After Sokal and Rohlf." *BMC Bioinformatics* 14, no. 1: 3. https://doi.org/10.1186/1471-2105-14-3.

Castoe, T. A., A. P. J. De Koning, H.-M. Kim, et al. 2009. "Evidence for an Ancient Adaptive Episode of Convergent Molecular Evolution." *Proceedings of the National Academy of Sciences of the United States of America* 106, no. 22: 8986–8991. https://doi.org/10.1073/pnas.0900233106.

Černý, M., and P. D. N. Hebert. 1999. "Intercontinental Allozyme Differentiation Among Four Holarctic *Daphnia* Species." *Limnology and Oceanography* 44, no. 6: 1381–1387. https://doi.org/10.4319/lo.1999.44.6.1381.

Charlesworth, B. 2009. "Effective Population Size and Patterns of Molecular Evolution and Variation." *Nature Reviews Genetics* 10, no. 3: 195–205. https://doi.org/10.1038/nrg2526.

Charlesworth, D. 2006. "Balancing Selection and Its Effects on Sequences in Nearby Genome Regions." *PLoS Genetics* 2, no. 4: 379–384. https://doi.org/10.1371/journal.pgen.0020064.

Chaturvedi, A., J. Zhou, J. A. M. Raeymaekers, et al. 2021. "Extensive Standing Genetic Variation From a Small Number of Founders Enables Rapid Adaptation in *Daphnia*." *Nature Communications* 12, no. 1: 4306. https://doi.org/10.1038/s41467-021-24581-z.

Chin, T. A., and M. E. Cristescu. 2021. "Speciation in *Daphnia*." *Molecular Ecology* 30: 1398–1418. https://doi.org/10.1111/mec.15824.

Cho, S., Z. Y. Huang, D. R. Green, D. R. Smith, and J. Zhang. 2006. "Evolution of the Complementary Sex-Determination Gene of Honey Bees: Balancing Selection and Trans-Species Polymorphisms." *Genome Research* 16, no. 11: 1366–1375. https://doi.org/10.1101/gr.4695306.

Cingolani, P., A. Platts, L. L. Wang, et al. 2012. "A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of Drosophila Melanogaster Strain w1118; Iso-2; Iso-3." *Fly* 6, no. 2: 80–92. https://doi.org/10.4161/fly.19695.

Clark, A. G. 1997. "Neutral Behavior of Shared Polymorphism." *Proceedings of the National Academy of Sciences of the United States of America* 94, no. 15: 7730–7734. https://doi.org/10.1073/pnas.94.15.7730.

Colbourne, J. K., T. J. Crease, L. J. Weider, P. D. N. Hebert, F. Duferesne, and A. Hobaek. 1998. "Phylogenetics and Evolution of a Circumarctic Species Complex (Cladocera: *Daphnia pulex*)." *Biological Journal of the Linnean Society* 65, no. 3: 347–365. https://doi.org/10.1111/j.1095-8312.1998.tb01146.x.

Colbourne, J. K., M. E. Pfrender, D. Gilbert, et al. 2011. "The Ecoresponsive Genome of *Daphnia pulex*." *Science* 331, no. 6017: 555–561. https://doi.org/10.1126/science.1197761.

Colbourne, J. K. J., P. D. Hebert, D. J. Taylor, and T. J. Givnish. 1997. "Evolutionary Origins of Phenotypic Diversity in *Daphnia*." In *Molecular Evolution and Adaptive Radiation*, 163–188. Cambridge: Cambridge University Press.

Cornetti, L., P. D. Fields, L. Du Pasquier, and D. Ebert. 2024. "Long-Term Balancing Selection for Pathogen Resistance Maintains Trans-Species Polymorphisms in a Planktonic Crustacean." *Nature Communications* 15, no. 1: 5333. https://doi.org/10.1038/s41467-024-49726-8.

Cornetti, L., P. D. Fields, K. Van Damme, and D. Ebert. 2019. "A Fossil-Calibrated Phylogenomic Analysis of *Daphnia* and the Daphniidae." *Molecular Phylogenetics and Evolution* 137: 250–262. https://doi.org/10.1016/j.ympev.2019.05.018.

Coughlan, J. M., and D. R. Matute. 2020. "The Importance of Intrinsic Postzygotic Barriers Throughout the Speciation Process." *Philosophical Transactions of the Royal Society B: Biological Sciences* 375, no. 1806: 20190533. https://doi.org/10.1098/rstb.2019.0533.

Crease, T. J., A. R. Omilian, K. S. Costanzo, and D. J. Taylor. 2012. "Transcontinental Phylogeography of the *Daphnia pulex* Species Complex." *PLoS One* 7, no. 10: e46620. https://doi.org/10.1371/journal.pone.0046620.

Cvijović, I., B. H. Good, E. R. Jerison, and M. M. Desai. 2015. "Fate of a Mutation in a Fluctuating Environment." *Proceedings of the National Academy of Sciences of the United States of America* 112, no. 36: E5021–E5028. https://doi.org/10.1073/pnas.1505406112.

Daniel, F., Revolution Analytics, and S. Weston. 2022. "doMC: Foreach Parallel Adaptor for 'Parallel'."

De Meester, L., M. Boersma, and P. Spaak. 1999. "Environmental Stress and Local Adaptation in *Daphnia magna*." *Limnology and Oceanography* 44, no. 2: 393–402. https://doi.org/10.4319/lo.1999.44.2.0393.

Decaestecker, E., S. Gaba, J. A. M. Raeymaekers, et al. 2007. "Host–Parasite 'Red Queen' Dynamics Archived in Pond Sediment." *Nature* 450, no. 7171: 870–873. https://doi.org/10.1038/nature06291.

Dodson, S. I. 1981. "Morphological Variation of *Daphnia pulex* Leydig (Crustacea: Cladocera) and Related Species From North America." *Hydrobiologia* 83, no. 1: 101–114. https://doi.org/10.1007/BF02187155.

Dowle, M., and A. Srinivasan. 2023. data.table: Extension of "data.frame." https://R-Datatable.Com, https://Rdatatable.Gitlab.Io/Data.Table, https://Github.Com/Rdatatable/Data.Table.

Dudycha, J. L. 2004. "Mortality Dynamics of *Daphnia* in Contrasting Habitats and Their Role in Ecological Divergence." *Freshwater Biology* 49, no. 5: 505–514. https://doi.org/10.1111/j.1365-2427.2004.01201.x.

Ebert, D. 2022. "*Daphnia* as a Versatile Model System in Ecology and Evolution." *EvoDevo* 13, no. 1: 16. https://doi.org/10.1186/s13227-022-00199-0.

Edwards, S. V., and P. Beerli. 2000. "Perspective: Gene Divergence, Population Divergence, and the Variance in Coalescence Time in Phylogeographic Studies." *Evolution* 54, no. 6: 1839–1854. https://doi.org/10.1111/j.0014-3820.2000.tb01231.x.

Emms, D. M., and S. Kelly. 2019. "OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics." *Genome Biology* 20, no. 1: 238. https://doi.org/10.1186/s13059-019-1832-y.

Erickson, P. A., C. A. Weller, D. Y. Song, A. S. Bangerter, P. Schmidt, and A. O. Bergland. 2020. "Unique Genetic Signatures of Local Adaptation Over Space and Time for Diapause, an Ecologically Relevant Complex Trait, in *Drosophila melanogaster*." *PLOS Genetics* 16, no. 11: e1009110. https://doi.org/10.1371/journal.pgen.1009110.

Fijarczyk, A., and W. Babik. 2015. "Detecting Balancing Selection in Genomes: Limits and Prospects." *Molecular Ecology* 24, no. 14: 3529–3545. https://doi.org/10.1111/mec.13226.

Flynn, J. M., F. J. J. Chain, D. J. Schoen, and M. E. Cristescu. 2017. "Spontaneous Mutation Accumulation in *Daphnia pulex* in Selection-Free vs. Competitive Environments." *Molecular Biology and Evolution* 34, no. 1: 160–173. https://doi.org/10.1093/molbev/msw234.

Fryer, G. 1991. "Functional Morphology and the Adaptive Radiation of the Daphniidae (Branchiopoda: Anomopoda)." *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 331, no. 1259: 1–99. https://doi.org/10.1098/rstb.1991.0001.

Gao, Z., M. Przeworski, and G. Sella. 2015. "Footprints of Ancient-Balanced Polymorphisms in Genetic Variation Data From Closely Related Species." *Evolution* 69, no. 2: 431–446. https://doi.org/10.1111/evo.12567.

Gómez, R., K. Van Damme, J. Gosálvez, E. S. Morán, and J. K. Colbourne. 2016. "Male Meiosis in Crustacea: Synapsis, Recombination, Epigenetics and Fertility in *Daphnia magna*." *Chromosoma* 125, no. 4: 769–787. https://doi.org/10.1007/s00412-015-0558-1.

Günther, T., and C. Nettelblad. 2019. "The Presence and Impact of Reference Bias on Population Genomic Studies of Prehistoric Human Populations." *PLoS Genetics* 15, no. 7: e1008302. https://doi.org/10.1371/journal.pgen.1008302.

Haag, C. R., and D. Ebert. 2007. "Genotypic Selection in *Daphnia* Populations Consisting of Inbred Sibships." *Journal of Evolutionary Biology* 20, no. 3: 881–891. https://doi.org/10.1111/j.1420-9101.2007.01313.x.

Haldane, J. B. S. 1927. "A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation." *Mathematical Proceedings of the Cambridge Philosophical Society* 23, no. 7: 838–844. https://doi.org/10.1017/S0305004100015644.

Harris, R. S. 2007. *Improved Pairwise Alignment of Genomic DNA*. University Park, PA: Pennsylvania State University.

Hebert, P. D. N., D. C. Ferrari, and T. J. Crease. 1982. "Heterosis in *Daphnia*: A Reassessment." *American Naturalist* 119, no. 3: 427–434. https://doi.org/10.1086/283921.

Hebert, P. D. N., and C. C. Wilson. 1994. "Provincialism in Plankton: Endemism and Allopatric Speciation in Australian *Daphnia*." *Evolution* 48, no. 4: 1333–1349. https://doi.org/10.1111/j.1558-5646.1994.tb05317.x.

Hedrick, P. W. 2013. "Adaptive Introgression in Animals: Examples and Comparison to New Mutation and Standing Variation as Sources of Adaptive Variation." *Molecular Ecology* 22, no. 18: 4606–4618. https://doi.org/10.1111/mec.12415.

Heier, C. R., and J. L. Dudycha. 2009. "Ecological Speciation in a Cyclic Parthenogen: Sexual Capability of Experimental Hybrids Between *Daphnia Pulex* and *Daphnia pulicaria*." *Limnology and Oceanography* 54, no. 2: 492–502. https://doi.org/10.4319/lo.2009.54.2.0492.

Held, C., S. Koenemann, and C. D. Schubart, eds. 2016. *Phylogeography and Population Genetics in Crustacea*. Boca Raton, FL: CRC Press. https://doi.org/10.1201/b11113.

Hoang, D. T., O. Chernomor, A. von Haeseler, B. Q. Minh, and L. S. Vinh. 2018. "UFBoot2: Improving the Ultrafast Bootstrap Approximation." *Molecular Biology and Evolution* 35, no. 2: 518–522. https://doi.org/10.1093/molbev/msx281.

Hobolth, A., J. Y. Dutheil, J. Hawks, M. H. Schierup, and T. Mailund. 2011. "Incomplete Lineage Sorting Patterns Among Human, Chimpanzee, and Orangutan Suggest Recent Orangutan Speciation and Widespread Selection." *Genome Research* 21, no. 3: 349–356. https://doi.org/10.1101/gr.114751.110.

Hosseinie, F. 1966. *The Ecology and Reproductive Cytology of* Daphnia middendorffiana, *Fischer (Cladocera) From the Arctic*. Bloomington, IN: Indiana University ProQuest Dissertations Publishing.

Huerta-Sánchez, E., X. Jin, B. Z. Asan, et al. 2014. "Altitude Adaptation in Tibetans Caused by Introgression of Denisovan-Like DNA." *Nature* 512, no. 7513: 194–197. https://doi.org/10.1038/nature13408.

Jackson, C. E., S. Xu, Z. Ye, et al. 2021. "Chromosomal Rearrangements Preserve Adaptive Divergence in Ecological Speciation." *bioRxiv*. https://doi.org/10.1101/2021.08.20.457158.

Jiang, X., H. Tang, Z. Ye, and M. Lynch. 2017. "Insertion Polymorphisms of Mobile Genetic Elements in Sexual and Asexual Populations of *Daphnia pulex*." *Genome Biology and Evolution* 9, no. 2: 362–374. https://doi.org/10.1093/gbe/evw302.

Johnson, D. S. 1951. *A Study of the Physiology and Ecology of Certain Cladocera*. Bedford: University of London, Bedford College. ProQuest Dissertations Publishing.

Jose, C., and F. Dufresne. 2010. "Differential Survival Among Genotypes of *Daphnia pulex* Differing in Reproductive Mode, Ploidy Level, and Geographic Origin." *Evolutionary Ecology* 24, no. 2: 413–421. https://doi.org/10.1007/s10682-009-9314-4.

Jouganous, J., W. Long, A. P. Ragsdale, and S. Gravel. 2017. "Inferring the Joint Demographic History of Multiple Populations: Beyond the Diffusion Approximation." *Genetics* 206, no. 3: 1549–1567. https://doi.org/10.1534/genetics.117.200493.

Kamvar, Z. N., J. F. Tabima, and N. J. Grünwald. 2014. "*Poppr*: An R Package for Genetic Analysis of Populations With Clonal, Partially Clonal, and/or Sexual Reproduction." *PeerJ* 2: e281. https://doi.org/10.7717/peerj.281.

Katoh, K., and D. M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30, no. 4: 772–780. https://doi.org/10.1093/molbev/mst010.

Key, F. M., J. C. Teixeira, C. de Filippo, and A. M. Andrés. 2014. "Advantageous Diversity Maintained by Balancing Selection in Humans." *Current Opinion in Genetics and Development* 29: 45–51. https://doi.org/10.1016/j.gde.2014.08.001.

Klein, J., A. Sato, S. Nagl, and C. O'hUigín. 1998. "Molecular Trans-Species Polymorphism." *Annual Review of Ecology and Systematics* 29, no. 1: 1–21. https://doi.org/10.1146/annurev.ecolsys.29.1.1.

Klein, J., Y. Satta, N. Takahata, and C. O'hUigin. 1993. "Trans-Specific *Mhc* Polymorphism and the Origin of Species in Primates." *Journal of Medical Primatology* 22, no. 1: 57–64. https://doi.org/10.1111/j.1600-0684.1993.tb00637.x.

Koenig, D., J. Hagmann, R. Li, et al. 2019. "Long-Term Balancing Selection Drives Evolution of Immunity Genes in *Capsella*." *eLife* 8: e43606. https://doi.org/10.7554/eLife.43606.

Kutschera, V. E., T. Bidon, F. Hailer, J. L. Rodi, S. R. Fain, and A. Janke. 2014. "Bears in a Forest of Gene Trees: Phylogenetic Inference Is Complicated by Incomplete Lineage Sorting and Gene Flow." *Molecular Biology and Evolution* 31, no. 8: 2004–2017. https://doi.org/10.1093/molbev/msu186.

Lack, J. 2017. "*Daphnia pulicaria* Raw Sequence Reads." NCBI Sequence Read Archive. PRJNA386566. https://www.ncbi.nlm.nih.gov/sra.

Lee, B. T., G. P. Barber, A. Benet-Pagès, et al. 2022. "The UCSC Genome Browser Database: 2022 Update." *Nucleic Acids Research* 50, no. D1: D1115–D1122. https://doi.org/10.1093/nar/gkab959.

Leffler, E. M., Z. Gao, S. Pfeifer, et al. 2013. "Multiple Instances of Ancient Balancing Selection Shared Between Humans and Chimpanzees." *Science* 339, no. 6127: 1578–1582. https://doi.org/10.1126/science.1234070.

Leinonen, R., H. Sugawara, M. Shumway, and International Nucleotide Sequence Database Collaboration. 2011. "The Sequence Read Archive." *Nucleic Acids Research* 39: D19–D21. https://doi.org/10.1093/nar/gkq1019.

Leydig, F. 1860. *Naturgeschichte der Daphniden:(Crustacea cladocera)*. Tübingen: Laupp & Siebeck.

Li, D., J. Huang, Q. Zhou, et al. 2022. "Artificial Light Pollution With Different Wavelengths at Night Interferes With Development, Reproduction, and Antipredator Defenses of *Daphnia magna*." *Environmental Science and Technology* 56, no. 3: 1702–1712. https://doi.org/10.1021/acs.est.1c06286.

Li, H. 2013. "Aligning Sequence Reads, Clone Sequences and Assembly Contigs With BWA-MEM." *arXiv preprint arXiv*:1303.3997.

Li, H., B. Handsaker, A. Wysoker, et al. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25, no. 16: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Li, H., and M. Nei. 1977. "Persistence of Common Alleles in Two Related Populations or Species." *Genetics* 86, no. 4: 901–914. https://doi.org/10.1093/genetics/86.4.901.

Lynch, M. 1987. "The Consequences of Fluctuating Selection for Isozyme Polymorphisms in *Daphnia*." *Genetics* 115, no. 4: 657–669. https://doi.org/10.1093/genetics/115.4.657.

Lynch, M., R. Gutenkunst, M. Ackerman, et al. 2017. "Population Genomics of *Daphnia pulex*." *Genetics* 206, no. 1: 315–332. https://doi.org/10.1534/genetics.116.190611.

Lynch, M., W. Wei, Z. Ye, and M. Pfrender. 2023. "The Genome-Wide Signature of Short-Term Temporal Selection." *bioRxiv*. 2023.04.28.538790. https://doi.org/10.1101/2023.04.28.538790.

Malinsky, M., M. Matschiner, and H. Svardal. 2021. "Dsuite—Fast D-Statistics and Related Admixture Evidence From VCF Files." *Molecular Ecology Resources* 21, no. 2: 584–595. https://doi.org/10.1111/1755-0998.13265.

Marková, S., F. Dufresne, M. Manca, and P. Kotlík. 2013. "Mitochondrial Capture Misleads About Ecological Speciation in the *Daphnia pulex* Complex." *PLoS One* 8, no. 7: e69497. https://doi.org/10.1371/journal.pone.0069497.

Martin, M., P. Ebert, and T. Marschall. 2023. "Read-Based Phasing and Analysis of Phased Variants With WhatsHap." In *Haplotyping*, edited by B. A. Peters and R. Drmanac, vol. 2590, 127–138. New York, NY: Springer. https://doi.org/10.1007/978-1-0716-2819-5_8.

Maruki, T. 2018. "Sequencing Data of *Daphnia pulex* From Nine Populations." NCBI Sequence Read Archive PRJNA482684. https://www.ncbi.nlm.nih.gov/sra.

Maruki, T., Z. Ye, and M. Lynch. 2022. "Evolutionary Genomics of a Subdivided Species." *Molecular Biology and Evolution* 39, no. 8: msac152. https://doi.org/10.1093/molbev/msac152.

McCoy, R. C., N. R. Garud, J. L. Kelley, C. L. Boggs, and D. A. Petrov. 2014. "Genomic Inference Accurately Predicts the Timing and Severity of a Recent Bottleneck in a Nonmodel Insect Population." *Molecular Ecology* 23, no. 1: 136–150. https://doi.org/10.1111/mec.12591.

Mi, H., A. Muruganujan, J. T. Casagrande, and P. D. Thomas. 2013. "Large-Scale Gene Function Analysis With the PANTHER Classification System." *Nature Protocols* 8, no. 8: 1551–1566. https://doi.org/10.1038/nprot.2013.092.

Novikova, P. Y., N. Hohmann, V. Nizhynska, et al. 2016. "Sequencing of the Genus *Arabidopsis* Identifies a Complex History of Nonbifurcating Speciation and Abundant Trans-Specific Polymorphism." *Nature Genetics* 48, no. 9: 1077–1082. https://doi.org/10.1038/ng.3617.

Nunez, J. C. B., S. Rong, A. Damian-Serrano, et al. 2021. "Ecological Load and Balancing Selection in Circumboreal Barnacles." *Molecular Biology and Evolution* 38, no. 2: 676–685. https://doi.org/10.1093/molbev/msaa227.

Ohta, T. 1971. "Associative Overdominance Caused by Linked Detrimental Mutations." *Genetical Research* 18, no. 3: 277–286. https://doi.org/10.1017/S0016672300012684.

Omilian, A. R., and M. Lynch. 2009. "Patterns of Intraspecific DNA Variation in the *Daphnia* Nuclear Genome." *Genetics* 182, no. 1: 325–336. https://doi.org/10.1534/genetics.108.099549.

Pantel, J. H., T. E. Juenger, and M. A. Leibold. 2011. "Environmental Gradients Structure *Daphnia pulex × pulicaria* Clonal Distribution: Hybrid *Daphnia* Clonal Distribution." *Journal of Evolutionary Biology* 24, no. 4: 723–732. https://doi.org/10.1111/j.1420-9101.2010.02196.x.

Paradis, E., and K. Schliep. 2019. "Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R." *Bioinformatics* 35, no. 3: 526–528. https://doi.org/10.1093/bioinformatics/bty633.

Payseur, B. A., and L. H. Rieseberg. 2016. "A Genomic Perspective on Hybridization and Speciation." *Molecular Ecology* 25, no. 11: 2337–2360. https://doi.org/10.1111/mec.13557.

Pennings, P. S., and J. Hermisson. 2006. "Soft Sweeps II—Molecular Population Genetics of Adaptation From Recurrent Mutation or Migration." *Molecular Biology and Evolution* 23, no. 5: 1076–1084. https://doi.org/10.1093/molbev/msj117.

Pijanowska, J., and A. Kowalczewski. 1997. "Predators Can Induce Swarming Behaviour and Locomotory Responses in *Daphnia*." *Freshwater Biology* 37, no. 3: 649–656. https://doi.org/10.1046/j.1365-2427.1997.00192.x.

Poplin, R., V. Ruano-Rubio, M. A. DePristo, et al. 2017. "Scaling Accurate Genetic Variant Discovery to Tens of Thousands of Samples." *BioRxiv*. https://doi.org/10.1101/201178.

Purcell, S., B. Neale, K. Todd-Brown, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics* 81, no. 3: 559–575. https://doi.org/10.1086/519795.

R Core Development Team. 2013. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. "Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7." *Systematic Biology* 67, no. 5: 901–904. https://doi.org/10.1093/sysbio/syy032.

Reger, J., M. I. Lind, M. R. Robinson, and A. P. Beckerman. 2018. "Predation Drives Local Adaptation of Phenotypic Plasticity." *Nature Ecology and Evolution* 2, no. 1: 100–107. https://doi.org/10.1038/s41559-017-0373-6.

Revell, L. J. 2019. "*learnPopGen*: An R Package for Population Genetic Simulation and Numerical Analysis." *Ecology and Evolution* 9, no. 14: 7896–7902. https://doi.org/10.1002/ece3.5412.

Richard, J. 1896. "Révision des Cladocères. Deuxième Partie. Anomopoda. Famille III.—Daphnidae." *Annales des Sciences Naturelles, Zoologie, Serie 8* 8: 187–363.

Ringelberg, J. 1999. "The Photobehaviour of *Daphnia spp.* as a Model to Explain Diel Vertical Migration in Zooplankton." *Biological Reviews of the Cambridge Philosophical Society* 74, no. 4: 397–423. https://doi.org/10.1017/S0006323199005381.

Rosenberg, N. A. 2003. "The Shapes of Neutral Gene Genealogies in Two Species: Probabilities of Monophyly, Paraphyly, and Polyphyly in a Coalescent Model." *Evolution* 57, no. 7: 1465–1477. https://doi.org/10.1111/j.0014-3820.2003.tb00355.x.

Sayers, E. W., E. E. Bolton, J. R. Brister, et al. 2022. "Database Resources of the National Center for Biotechnology Information." *Nucleic Acids Research* 50, no. D1: D20–D26. https://doi.org/10.1093/nar/gkab1112.

Schield, D. R., B. W. Perry, R. H. Adams, et al. 2022. "The Roles of Balancing Selection and Recombination in the Evolution of Rattlesnake Venom." *Nature Ecology and Evolution* 6, no. 9: 1367–1380. https://doi.org/10.1038/s41559-022-01829-5.

Schiffels, S., and K. Wang. 2020. "MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent." In *Statistical Population Genomics*, edited by J. Y. Dutheil, vol. 2090, 147–166. New York, NY: Springer. https://doi.org/10.1007/978-1-0716-0199-0_7.

Scourfield, D. J. 1942. "XIX.—The "Pulex" Forms of *Daphnia* and Their Separation Into Two Distinct Series." *Annals and Magazine of Natural History* 9, no. 51: 202–219. https://doi.org/10.1080/03745481.1942.9755477.

Ségurel, L., Z. Gao, and M. Przeworski. 2013. "Ancestry Runs Deeper Than Blood: The Evolutionary History of *ABO* Points to Cryptic Variation of Functional Importance." *BioEssays* 35, no. 10: 862–867. https://doi.org/10.1002/bies.201300030.

Ségurel, L., E. E. Thompson, T. Flutre, et al. 2012. "The ABO Blood Group Is a Trans-Species Polymorphism in Primates." *Proceedings of the National Academy of Sciences of the United States of America* 109, no. 45: 18493–18498. https://doi.org/10.1073/pnas.1210603109.

Sellinger, T. P. P., D. Abu Awad, M. Moest, and A. Tellier. 2020. "Inference of Past Demography, Dormancy and Self-Fertilization Rates From Whole Genome Sequence Data." *PLoS Genetics* 16, no. 4: e1008698. https://doi.org/10.1371/journal.pgen.1008698.

Seppey, M., M. Manni, and E. M. Zdobnov. 2019. "BUSCO: Assessing Genome Assembly and Annotation Completeness." In *Gene Prediction*, edited by M. Kollmar, vol. 1962, 227–245. New York: Springer. https://doi.org/10.1007/978-1-4939-9173-0_14.

Setter, D., S. Mousset, X. Cheng, R. Nielsen, M. DeGiorgio, and J. Hermisson. 2020. "VolcanoFinder: Genomic Scans for Adaptive Introgression." *PLoS Genetics* 16, no. 6: e1008867. https://doi.org/10.1371/journal.pgen.1008867.

Shen, W., S. Le, Y. Li, and F. Hu. 2016. "SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation." *PLoS One* 11, no. 10: e0163962. https://doi.org/10.1371/journal.pone.0163962.

Siewert, K. M., and B. F. Voight. 2017. "Detecting Long-Term Balancing Selection Using Allele Frequency Correlation." *Molecular Biology and Evolution* 34, no. 11: 2996–3005. https://doi.org/10.1093/molbev/msx209.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. "BUSCO: Assessing Genome Assembly and Annotation Completeness With Single-Copy Orthologs." *Bioinformatics* 31, no. 19: 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Slater, G., and E. Birney. 2005. "Automated Generation of Heuristics for Biological Sequence Comparison." *BMC Bioinformatics* 6, no. 1: 31. https://doi.org/10.1186/1471-2105-6-31.

So, M., H. Ohtsuki, W. Makino, et al. 2015. "Invasion and Molecular Evolution of *Daphnia pulex* in Japan." *Limnology and Oceanography* 60, no. 4: 1129–1138. https://doi.org/10.1002/lno.10087.

Solberg, O. D., S. J. Mack, A. K. Lancaster, et al. 2008. "Balancing Selection and Heterogeneity Across the Classical Human Leukocyte Antigen Loci: A Meta-Analytic Review of 497 Population Studies." *Human Immunology* 69, no. 7: 443–464. https://doi.org/10.1016/j.humimm.2008.05.001.

Soni, V., M. Vos, and A. Eyre-Walker. 2022. "A New Test Suggests Hundreds of Amino Acid Polymorphisms in Humans Are Subject to Balancing Selection." *PLoS Biology* 20, no. 6: e3001645. https://doi.org/10.1371/journal.pbio.3001645.

Spitze, K. 1993. "Population Structure in *Daphnia obtusa*: Quantitative Genetic and Allozymic Variation." *Genetics* 135, no. 2: 367–374. https://doi.org/10.1093/genetics/135.2.367.

Standard, A. 2007. *Standard Guide for Conducting Acute Toxicity Tests on Test Materials With Fishes, Macroinvertebrates, and Amphibians.* West Conshohocken, PA: American Society for Testing and Materials. www.Atsm.Org.

Suh, A., L. Smeds, and H. Ellegren. 2015. "The Dynamics of Incomplete Lineage Sorting Across the Ancient Adaptive Radiation of Neoavian Birds." *PLoS Biology* 13, no. 8: e1002224. https://doi.org/10.1371/journal.pbio.1002224.

Sutherland, B. G., K. R. Tobutt, and T. P. Robbins. 2008. "Trans-Specific S-RNase and SFB Alleles in *Prunus* Self-Incompatibility Haplotypes." *Molecular Genetics and Genomics* 279: 95–106. https://doi.org/10.1007/s00438-007-0300-7.

Tarailo-Graovac, M., and N. Chen. 2009. "Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences." *Current Protocols in Bioinformatics* 25, no. 1: 4.10.1–4.10.14. https://doi.org/10.1002/0471250953.bi0410s25.

Těšický, M., and M. Vinkler. 2015. "Trans-Species Polymorphism in Immune Genes: General Pattern or MHC-Restricted Phenomenon?" *Journal of Immunology Research* 2015: 838035. https://doi.org/10.1155/2015/838035.

Thomas Lin Pedersen. 2022. "patchwork: The Composer of Plots." https://Patchwork.Data-Imaginist.Com, https://github.com/thomasp85/patchwork.

Tucker, A. E., M. S. Ackerman, B. D. Eads, S. Xu, and M. Lynch. 2013. "Population-Genomic Insights Into the Evolutionary Origin and Fate of Obligately Asexual *Daphnia pulex*." *Proceedings of the National Academy of Sciences of the United States of America* 110, no. 39: 15740–15745. https://doi.org/10.1073/pnas.1313388110.

Van Gool, E., and J. Ringelberg. 2003. "What Goes Down Must Come Up: Symmetry in Light-Induced Migration Behaviour of *Daphnia*." *Hydrobiologia* 491, no. 1–3: 301–307. https://doi.org/10.1023/A:1024406324317.

Vergilino, R., S. Markova, M. Ventura, M. Manca, and F. Dufresne. 2011. "Reticulate Evolution of the *Daphnia pulex* Complex as Revealed by Nuclear Markers." *Molecular Ecology* 20, no. 6: 1191–1207. https://doi.org/10.1111/j.1365-294X.2011.05004.x.

Villanueva, R. A. M., and Z. J. Chen. 2019. "ggplot2: Elegant Graphics for Data Analysis (2nd ed.)." *Measurement: Interdisciplinary Research and Perspectives* 17, no. 3: 160–167. https://doi.org/10.1080/15366367.2019.1565254.

Wang, B., and T. Mitchell-Olds. 2017. "Balancing Selection and Trans-Specific Polymorphisms." *Genome Biology* 18, no. 1: 231. https://doi.org/10.1186/s13059-017-1365-1.

Wang, M., L. Zhang, Z. Zhang, et al. 2020. "Phylogenomics of the Genus *Populus* Reveals Extensive Interspecific Gene Flow and Balancing Selection." *New Phytologist* 225, no. 3: 1370–1382. https://doi.org/10.1111/nph.16215.

Wickham, H., M. Averick, J. Bryan, et al. 2019. "Welcome to the Tidyverse." *Journal of Open Source Software* 4, no. 43: 1686. https://doi.org/10.21105/joss.01686.

Wills, C. 1975. "Marginal Overdominance in *Drosophila*." *Genetics* 81, no. 1: 177–189. https://doi.org/10.1093/genetics/81.1.177.

Wittmann, M. J., A. O. Bergland, M. W. Feldman, P. S. Schmidt, and D. A. Petrov. 2017. "Seasonally Fluctuating Selection Can Maintain Polymorphism at Many Loci via Segregation Lift." *Proceedings of the National Academy of Sciences of the United States of America* 114, no. 46: E9932–E9941. https://doi.org/10.1073/pnas.1702994114.

Wiuf, C., K. Zhao, H. Innan, and M. Nordborg. 2004. "The Probability and Chromosomal Extent of Trans-Specific Polymorphism." *Genetics* 168, no. 4: 2363–2372. https://doi.org/10.1534/genetics.104.029488.

Wu, Q., T.-S. Han, X. Chen, et al. 2017. "Long-Term Balancing Selection Contributes to Adaptation in *Arabidopsis* and Its Relatives." *Genome Biology* 18, no. 1: 217. https://doi.org/10.1186/s13059-017-1342-8.

Xu, S., L. Li, X. Luo, et al. 2022. "*Ggtree*: A Serialized Data Object for Visualization of a Phylogenetic Tree and Annotation Data." *iMeta* 1, no. 4: e56. https://doi.org/10.1002/imt2.56.

Xu, S., K. Spitze, M. S. Ackerman, et al. 2015. "Hybridization and the Origin of Contagious Asexuality in *Daphnia pulex*." *Molecular Biology and Evolution* 32: 3215–3225. https://doi.org/10.1093/molbev/msv190.

Yampolsky, L. Y., T. M. Schaer, and D. Ebert. 2014. "Adaptive Phenotypic Plasticity and Local Adaptation for Temperature Tolerance in Freshwater Zooplankton." *Proceedings of the Royal Society B: Biological Sciences* 281, no. 1776: 20132744. https://doi.org/10.1098/rspb.2013.2744.

Ye, Z., C. Molinier, C. Zhao, C. R. Haag, and M. Lynch. 2019. "Genetic Control of Male Production in *Daphnia pulex*." *Proceedings of the National Academy of Sciences of the United States of America* 116, no. 31: 15602–15609. https://doi.org/10.1073/pnas.1903553116.

Ye, Z., M. E. Pfrender, and M. Lynch. 2023. "Evolutionary Genomics of Sister Species Differing in Effective Population Sizes and Recombination Rates." *Genome Biology and Evolution* 15, no. 11: evad202. https://doi.org/10.1093/gbe/evad202.

Ye, Z., C. Zhao, R. T. Raborn, et al. 2022. "Genetic Diversity, Heteroplasmy, and Recombination in Mitochondrial Genomes of *Daphnia pulex*, *Daphnia pulicaria*, and *Daphnia obtusa*." *Molecular Biology and Evolution* 39, no. 4: msac059. https://doi.org/10.1093/molbev/msac059.

Zhang, J., K. Kobert, T. Flouri, and A. Stamatakis. 2014. "PEAR: A Fast and Accurate Illumina Paired-End reAd mergeR." *Bioinformatics* 30, no. 5: 614–620. https://doi.org/10.1093/bioinformatics/btt593.

Zheng, X., S. M. Gogarten, M. Lawrence, et al. 2017. "SeqArray—A Storage-Efficient High-Performance Data Format for WGS Variant Calls." *Bioinformatics* 33, no. 15: 2251–2257. https://doi.org/10.1093/bioinformatics/btx145.

Zheng, X., D. Levine, J. Shen, S. M. Gogarten, C. Laurie, and B. S. Weir. 2012. "A High-Performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data." *Bioinformatics* 28, no. 24: 3326–3328. https://doi.org/10.1093/bioinformatics/bts606.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.