Genome-Wide Allele Frequency Changes Reveal That Dynamic Metapopulations Evolve Differently

Pascal Angst ⓘ,[1,*] Christoph R. Haag ⓘ,[2,3] Frida Ben-Ami ⓘ,[3,4] Peter D. Fields ⓘ,[1,3,†]
and Dieter Ebert ⓘ [1,3,†]

[1]Department of Environmental Sciences, Zoology, University of Basel, Basel 4051, Switzerland

[2]CEFE, Université de Montpellier, CNRS, EPHE, IRD, Montpellier 34293, France

[3]Tvärminne Zoological Station, University of Helsinki, Hanko 10900, Finland

[4]School of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 6997801, Israel

[†]These authors share last authorship.

*Corresponding author: E-mail: pascal.angst@unibas.ch.

Associate editor: Yuseob Kim

Abstract

Two important characteristics of metapopulations are extinction–(re)colonization dynamics and gene flow between subpopulations. These processes can cause strong shifts in genome-wide allele frequencies that are generally not observed in "classical" (large, stable, and panmictic) populations. Subpopulations founded by one or a few individuals, the so-called propagule model, are initially expected to show intermediate allele frequencies at polymorphic sites until natural selection and genetic drift drive allele frequencies toward a mutation–selection–drift equilibrium characterized by a negative exponential-like distribution of the site frequency spectrum. We followed changes in site frequency spectrum distribution in a natural metapopulation of the cyclically parthenogenetic pond-dwelling microcrustacean *Daphnia magna* using biannual pool-seq samples collected over a 5-yr period from 118 ponds occupied by subpopulations of known age. As expected under the propagule model, site frequency spectra in newly founded subpopulations trended toward intermediate allele frequencies and shifted toward right-skewed distributions as the populations aged. Immigration and subsequent hybrid vigor altered this dynamic. We show that the analysis of site frequency spectrum dynamics is a powerful approach to understand evolution in metapopulations. It allowed us to disentangle evolutionary processes occurring in a natural metapopulation, where many subpopulations evolve in parallel. Thereby, stochastic processes like founder and immigration events lead to a pattern of subpopulation divergence, while genetic drift leads to converging site frequency spectrum distributions in the persisting subpopulations. The observed processes are well explained by the propagule model and highlight that metapopulations evolve differently from classical populations.

Key words: metapopulation, colonization, founder bottleneck, genomics, turnover dynamics, time series, *Daphnia*, *Cladocera*, site frequency spectrum, hybrid vigor.

Introduction

In heterogeneous landscapes, populations are often fragmented, forming a patchwork of subpopulations connected by gene flow. Subpopulations may display strong dynamics like local extinction and habitat patch (re)colonization (Hanski 1999; Colombo and Anteneodo 2015) that may create strong evolutionary changes in genome-wide allele frequencies, i.e. in the site frequency spectrum (SFS). The evolution of such metapopulations is, therefore, expected to differ from evolution in classical populations, here defined as large, stable, and panmictic populations, which are predominantly considered in population genetic studies (Charlesworth 2009). The SFS of classical populations is shaped by a combination of deterministic

(e.g. natural selection) and stochastic (e.g. genetic drift) evolutionary processes (Buffalo and Coop 2020). These processes also shape the SFSs of metapopulations, but their relative impact is expected to differ (Whitlock 2004). In dynamic metapopulations, founder events, i.e. strong bottlenecks when a single or a few individuals colonize vacant habitat patches, are frequent and are expected to result in low genetic diversity and a higher abundance of alleles with intermediate frequencies (Montero-Pau et al. 2018; Saastamoinen et al. 2018). In the extreme case, genetic bottlenecks are caused by a single selfing colonizer, founding a subpopulation in which all polymorphic alleles are initially at frequencies close to 50%. We, therefore, expect the SFSs of subpopulations in

Open Access

dynamic metapopulations to be skewed more toward intermediate frequencies than the right-skewed, negative exponential-like distributions observed in classical populations (Crawford and Lazzaro 2012). However, this theoretical predication has not been assessed. Here, we quantify the distribution of genome-wide allele frequencies in a natural metapopulation and examine the processes that shape it over time.

The propagule model of colonization is often used to explain population genetic processes in metapopulations. It assumes that subpopulations are colonized by one or a few individuals, often originating from a single source population (Slatkin 1977). After the founder bottleneck, evolutionary processes act on the intermediate frequency skewed distribution of the SFS (Whitlock 2004). In the subpopulations, which are initially small and prone to inbreeding (high genetic load) (Whitlock et al. 2000; Lohr and Haag 2015), genetic drift dominates the evolutionary process so that the efficacy of deterministic processes like purifying or spatially varying selection is low (Kutschera et al. 2020; Szép et al. 2021). In just a few (sexual) generations, genetic drift may reduce the skew of the SFS from intermediate frequency alleles to a successively flattening SFS. Pervasive genetic drift and the occurrence of new mutations shift the SFS further toward mutation–drift equilibrium with a negative exponential-like distribution similar to those known from classical populations (Crawford and Lazzaro 2012; Hahn 2018). However, this long-term effect is unlikely to contribute much to the observed SFS dynamics in dynamic metapopulations because subpopulations may not persist long enough. Gene flow is another important evolutionary process that shapes SFSs in metapopulations. It introduces new alleles into subpopulations, thus accelerating the evolution of SFSs toward a right-skewed distribution (Pannell and Charlesworth 1999). Gene flow may also lead to hybrid vigor, which is expected when residents and immigrant individuals mate to produce outbred offspring, especially if one or both parents suffered from high genetic load (Whitlock et al. 2000; Ebert et al. 2002). Under hybrid vigor, the hybrid offspring has high fitness, i.e. a selective advantage over the resident population and the immigrant parent, because deleterious mutations, which are homozygous in the parents, are masked in a recessive state. Therefore, immigrant alleles can quickly increase in frequency, up to 50% in the most extreme cases (Whitlock et al. 2000; Ebert et al. 2002), leading SFSs to evolve toward alleles with intermediate frequencies and higher heterozygosity. Selection for heterozygotes may also be observed when recombination creates variation in the level of genetic load, for example, among offspring after selfing of a single founder (Wang et al. 1999). Taken together, the propagule model predicts that founder events, genetic drift, inbreeding, gene flow, and natural selection will jointly drive the evolution of SFSs in a metapopulation. However, how SFSs of natural subpopulations evolve in the light of these evolutionary processes is not clear, in particular because these processes are influenced by ecology-related chance events

(e.g. immigration and bottleneck size), generating distinct evolutionary trajectories for individual subpopulations (Zhou et al. 2010). At any given time, genetic bottlenecks, hybrid vigor, and extinction events can occur in some, but not other, subpopulations. Despite these local differences, the overall trend of increasing genetic diversity and flattening of the SFS distribution with increasing age of a subpopulation is expected to be shared by all subpopulations but may only be visible in a large enough sample. To investigate the entire range of evolutionary relevant processes, thus, one should study many subpopulations of a metapopulation in parallel and over an extended time period.

The genetic consequences of metapopulation dynamics might be most obvious in small, unstable habitat patches where local extinctions and (re)colonization occur frequently. Here, we studied 118 occupied habitat patches of a dynamic metapopulation of the pond-dwelling crustacean *Daphnia magna* on the Skerry Islands of Southwestern Finland, where subpopulations of the cyclical parthenogen occur in distinct, small water bodies (Fig. 1). In this metapopulation, *D. magna* undergoes about six to ten asexual generations annually (planktonic or active phase) and one or two sexual generations annually. Sexual reproduction results in the production of resting stages, allowing *D. magna* to survive the inhospitable winters. An ongoing study with biannual censuses (Pajunen 1986; Pajunen and Pajunen 2003; Ebert et al. 2013; Dubart et al. 2020) has identified high extinction–(re)colonization dynamics in this metapopulation, with about 20% of all ponds in the study area containing *D. magna* subpopulations, about one-fifth of which going extinct every year. This pattern is balanced by a yearly colonization rate of about 5% for empty ponds. Colonization involves one or a few individuals, most likely originating from a single source population (Haag et al. 2005). Thus, subpopulation founding is common and associated with severe genomic bottlenecks that might lead to low genomic diversity in newly founded subpopulations compared with older subpopulations with increased genomic diversity (Haag et al. 2005). Colonization under the propagule model assumes that stochastic processes like genetic drift drive the evolutionary process of the metapopulation (Angst, Ameline, et al. 2022), although deterministic processes like selection for heterozygotes (through hybrid vigor [Ebert et al. 2002] and because of variation in the level of genetic load [Haag and Ebert 2007]) and local adaptation (Roulin et al. 2015; Cabalzar et al. 2019; Halter et al. 2022) are also expected to shape allele frequency dynamics. The dynamic nature and well-studied ecology of the *D. magna* metapopulation make it a particularly suitable model to investigate the dynamics of genome-wide allele frequency changes in dynamic metapopulations.

This population genomic study aimed to investigate the evolutionary process of a dynamic metapopulation by biannually sampling subpopulations of different ages and ecology for five consecutive years. Previous studies on this metapopulation were based on single cross-sections and did not follow the local dynamic across time (Haag
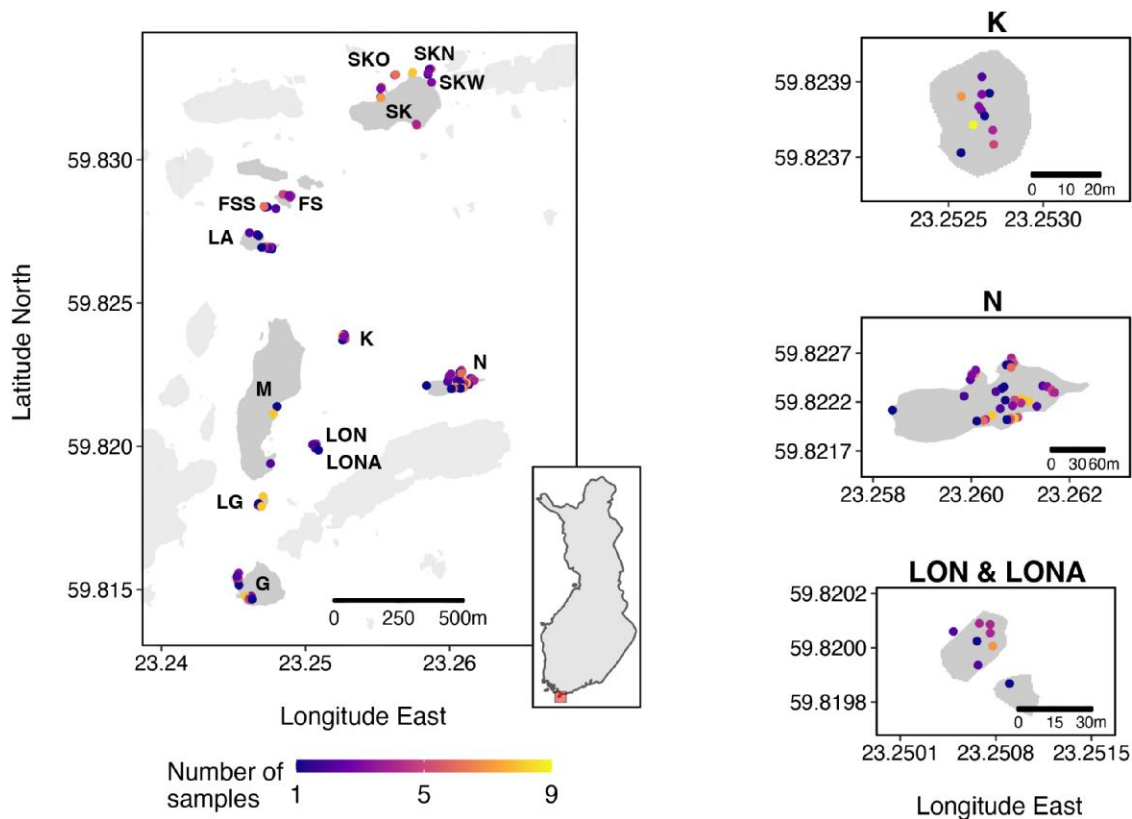
Fig. 1. Geographic locations of sampled ponds. The map on the left gives an overview of the study area. The inset map in the middle shows Finland with the study area indicated (small square in Southern Finland). Islands K, N, and LON/LONA are enlarged as insets on the right to better illustrate the spatial arrangement of individual ponds. Island names and abbreviations were adopted from the long-term study (see e.g. Angst, Ameline, et al. 2022). Ponds are colored according to the number of samples taken during the 5-yr study period. The study area contains about five times more habitat patches (ponds), but *D. magna* did not occur there during the study period. Lightened islands are not part of the biennial census and were, therefore, not sampled.

et al. 2005; Cabalzar et al. 2019; Angst, Ameline, et al. 2022). Instead, temporal dynamics were studied indirectly by regressing patterns against the recorded population age. This approach is limited, however, as older populations occur in more stable habitat patches and are, thus, a biased subsample of all subpopulations (Dubart et al. 2020). Our new approach—to quantify the distributions of SFSs in individual subpopulations and follow them over time—disentangles the roles of the distinct evolutionary forces that drive genomic changes in dynamic metapopulations. With a 5-yr sampling period of all subpopulations in a metapopulation with over 550 habitat patches, we are able to witness subpopulation founding and gene flow events with subsequent hybrid vigor in subpopulations of diverse ages. To supplement the analysis of our natural samples, we also conducted simulations of the population genetic processes across time. Together with previous findings from empirical and experimental studies on the same metapopulation, these approaches seek to understand the effects of founder events, natural selection, dispersal, hybrid vigor, and the impact of ecological factors on metapopulation evolution. While these scenarios have been described previously as separate events, here, we quantify their frequency and individual contribution to the overall

dynamics in order to answer questions about how metapopulations evolve and how this differs from expectations for classical populations.

Results

Samples and Sequencing

We sequenced 418 pooled *D. magna* samples from 118 ponds (Fig. 1, mean per pond: 4) across 14 islands (range per island: 1 to 142 samples, mean: 30) biannually from 2014 to 2018 (range per year: 53 to 108, mean: 84). One hundred ninety-six of these samples were collected in spring and 222 in summer. Sequencing resulted in average whole-genome coverages for the *D. magna* reference genome mostly well above 20× (supplementary table S1, Supplementary Material online). After variant filtering, 3,084,861 single nucleotide polymorphisms (SNPs) were retained for the analysis.

In individual subpopulations, alleles and their frequencies are not expected to change quickly unless specific events like hybrid vigor after gene flow are present. It has been suggested that hybrid vigor is important in shaping SFSs in focal metapopulations, and its effects on allele
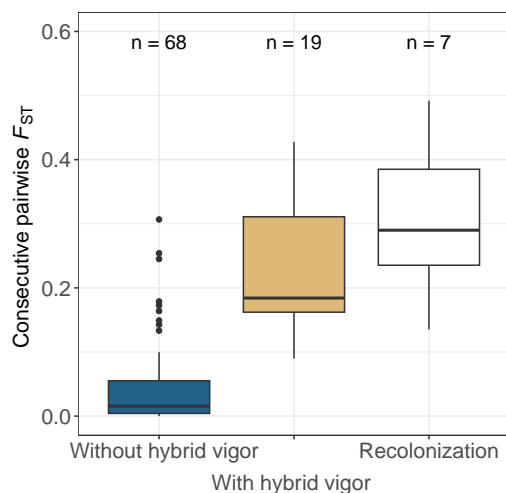
Fig. 2. Classification of subpopulations into groups undergoing different processes. The figure shows pairwise F_{ST} for consecutive samples from subpopulations with (yellow) and without (blue) hybrid vigor events. Hybrid vigor will lead to high F_{ST} between consecutive samples. Without hybrid vigor, F_{ST} is generally low but includes a few high values in subpopulations where we observed no increase in heterozygous sites. High F_{ST} without a gain in heterozygous sites might stem from population bottlenecks or from extinction–recolonization that happened too quickly to be detected by our sampling (cryptic extinction–recolonization). The right box (white) shows pairwise F_{ST} between subpopulations of the same pond before and after a recorded true positive extinction–recolonization event. This and subsequent analyses focused on the temporal dynamics only use subpopulations with at least 2 yr of sampling success.

frequencies in subpopulations—namely a high pairwise F_{ST} between two consecutive sampling time points with a simultaneous increase in the number of heterozygous sites—are well understood. By comparing all pairs of consecutive samples, we found 19 subpopulations that matched indications of gene flow plus hybrid vigor (Fig. 2; Table 1A). Therefore, in addition to the overall data, we analyzed these subpopulations separately from those presumably without hybrid vigor events. Further, we observed ten extinction–recolonization events in the field, where a subpopulation went extinct and a new subpopulation recolonized the same pond. Subpopulations of the pond before and after these extinction–recolonization events were treated as separate subpopulations. We calculated pairwise F_{ST} between them to detect potential false positives (little change in allele frequencies unnoticeable in the SFS distribution; $n = 3$; supplementary table S2, Supplementary Material online) and to visualize true positives (strong change in allele frequencies affecting the shape of the SFS; $n = 7$; Fig. 2, Table 1B, and supplementary table S2, Supplementary Material online). Thus, our data set included 125 subpopulations from 118 ponds, with 94 ponds having multiple samples (Fig. 2).

Shifts in Genome-Wide Allele Frequencies

We expect the folded SFS distribution of a subpopulation to generally evolve from left to right skewed with increasing age. However, this prediction is based on neutral

evolutionary processes. Deterministic processes may push the SFS distribution (back) in the opposite direction to become more left skewed. As a simple summary statistic for describing the distribution of the SFS, we use the median frequency of minor alleles (median MAF), because it is relatively unsensitive to the exact shape of a SFS. As a first approach to investigate whole-genome allele frequency distributions in subpopulations, we tested whether median MAF was associated with the age of the subpopulation by comparing subpopulations of different ages. We found that SFSs of young subpopulations showed higher median MAFs (more SNPs with intermediate frequencies, i.e. left skewed) than older subpopulations (see SFSs in Fig. 3a, Table 1C and D, and statistics in Fig. 3b and supplementary fig. S1, Supplementary Material online; young versus intermediate: $t(87.131) = 3.5602$, $P = 6.029 \times 10^{-4}$, young versus old: $t(76.424) = 5.8076$, $P = 1.372 \times 10^{-7}$, intermediate versus old: $t(63.114) = 1.3772$, $P = 0.173$, and mixed-effect model: $F = 26.73$, $P = 6.16 \times 10^{-7}$).

In following the genome-wide allele frequencies of individual subpopulations for up to 5 yr, we found an overall decrease in median MAF over time, i.e. the SNPs' MAFs decreased ($F = 4.54$, $P = 0.034$; Table 1D). This decrease was mostly due to allele frequency changes over winter (in the resting stage phase), not during the planktonic phase in summer, where we observed a slight increase in median MAF (paired Wilcoxon: $V = 570$, $P = 0.011$; Fig. 4c and supplementary fig. S2, Supplementary Material online, and Table 1E and F). The decrease of median MAF was stronger in the subset of young subpopulations ($F = 6.89$, $P = 0.011$) and in subpopulations without hybrid vigor ($F = 5.70$, $P = 0.018$; Fig. 3c) than overall. This effect was not observed in subpopulations with a genomic signature of gene flow plus hybrid vigor, where we expected and observed an increase in median MAF (Fig. 3d; Table 1A).

Even though whole-genome allele frequency distributions of individual subpopulations diverge from those expected in a classical population (some strongly), the overall metapopulation might diverge less. This pattern arises because—analogous to individuals in a classical population—individual subpopulations likely have private alleles and only a few alleles are shared among most subpopulations. To examine this, we estimated the allele frequencies on the metapopulation level. We averaged the allele frequencies of all subpopulations, obtaining a distribution similar to what would be expected for a classical population (Fig. 3e). To compare the evolution of our composite subpopulation's genome-wide allele frequency distribution with a single large population of the same species, we used temporal data (three samples across about 18 yr) from the Oud-Heverlee *D. magna* population in Belgium (Chaturvedi et al. 2021). Based on the analysis of 2,092,135 SNPs, we estimated that the median MAF of this large population has been stable at 0.2 for about 18 yr (Fig. 3f). This estimate is lower than the median MAF of the old subpopulations of the metapopulation (about 0.25; Fig. 3b).

Table 1 Summary of biological dynamics inferred through the analysis of genomic summary statistics in the metapopulation

Observation	Population genomic signature	Interpretation
A: Hybrid vigor	Genomic variation in allele frequencies (F_{ST}) strongly increases in consecutive samples from subpopulations without observed extinction–recolonization event and is accompanied by an increase in the number of variable sites (Fig. 2).	Outbred offspring from matings between resident and immigrant genotypes are rapidly spreading as a result of hybrid vigor.
B: Extinction–recolonization event	In habitat patches vacant for over a year, genomic samples from before and after this period were genetically strongly differentiated (Fig. 2).	The samples represent two different subpopulations—before and after an extinction–recolonization event.
C: Founder event	Newly established subpopulations are enriched in SNPs with intermediate allele frequencies (= high median MAF) (Fig. 3a and b).	Genetic bottlenecks happen during colonization of a vacant habitat patch.
D: Evolution toward mutation–selection–drift equilibrium	Median MAF is lower in older subpopulations than in newly established subpopulations. It declines with time in individual subpopulations, mostly due to neutral processes (Fig. 3b and c).	The SFS evolves from the initial excess of intermediate frequency variants through diverse processes, including genetic drift and recombination. Gene flow and mutation introduce initially rare alleles.
E: Strong genetic drift associated with the overwintering phase	Median MAF between samples from summer and the following spring (over winter) decreases (Fig. 4c). They show higher pairwise F_{ST} than consecutive samples from spring and the following summer (Fig. 4b).	Gene pools experience more change during the overwintering than the planktonic phase because sexual reproduction (allele reshuffling) and overwintering bottlenecks associated with resting stage production, survival, and hatching introduce variance in allele frequencies.
F: Heterozygote advantage during planktonic phase	The SFS distribution moves slightly toward frequencies of 0.5 (increasing median MAF) during the planktonic phase (Fig. 4c).	During the planktonic phase, clonal selection favors heterozygote genotypes.
G: Variation in genetic diversity	Subpopulations show stable genomic diversity (θ), but with variation among subpopulations (supplementary fig. S3, Supplementary Material online).	Different processes act together or counteract each other in their effect on θ. For example, hybrid vigor events might increase genetic diversity while bottlenecks decrease it. Subpopulations might persist longer because of higher diversity.
H: Temporal F_{ST} increases over time within subpopulations	F_{ST} increases with time between samples from the same subpopulation, most notably in less isolated subpopulations (Fig. 4a).	Genetic drift and gene flow create a pattern of population differentiation over time. Less isolated subpopulations are a more likely target of gene flow, leading to higher differentiation between consecutive samples.
I: Isolation by distance	Geographically close subpopulations (from the same or nearby islands) are more similar genetically (Fig. 6).	Gene flow happens at short distances, which leads to a pattern of isolation by distance.
J: Rare, long-distance gene flow	Some subpopulations separated by relatively large geographical distances (i.e. from different islands) are genetically similar (Fig. 6).	There is rare, long-distance gene flow.

Genomic Diversity and Differentiation Over Time

Genomic diversity measured as Watterson's theta (θ) of individual subpopulations did not generally increase over time (supplementary fig. S3, Supplementary Material online). The only subpopulations that showed an increase were those with gene flow plus hybrid vigor, which was part of how we detected this subset (Table 1G). When modeling θ, the interactive effect between the explanatory factors time and isolation was weak (NN2) ($F = 3.39$, $P = 0.067$), suggesting that temporal changes in θ could only marginally be caused by the subpopulation's degree of isolation. In the pairwise comparisons of consecutive samples, the difference in θ between spring samples and those from the following summer was indistinguishable from the difference in θ between summer and next year's spring samples (supplementary fig. S4, Supplementary Material online).

Population differentiation, estimated as pairwise F_{ST} between samples of the same subpopulation, increased with time between samples, ∂t, overall, and in both isolated and less isolated samples, with isolation measured as the mean distance to the two nearest neighboring subpopulations (NN2) (Fig. 4a; Table 1H). A nonlinear model described by $F_{ST} = ab^{1/\partial t}$, with $a = 0.13$ and $b = 0.51$, fit the observed data better than a linear model, i.e. $F_{ST} = a + b\partial t$ (Akaike information criterion [AIC] nonlinear: $-1,436.778$; AIC linear: $-1,422.533$). Samples of the same pond were less differentiated if taken from isolated ponds than taken from less isolated ponds, likely because a main source of variation, gene flow, is higher in the less isolated ponds (Fig. 4a). Pairwise F_{ST} estimates between consecutive samples of the same subpopulation were three times higher between summer and the following year's spring samples, i.e. across the winter (on average about 0.012), than between spring and the following summer samples (0.001; paired Wilcoxon: $V = 191$, $P = 0.005$; Fig. 4b and supplementary fig. S2, Supplementary Material online; Table 1E), indicating stronger evolutionary dynamics across the winter than the summer.

Regarding spatial population differentiation, pairwise F_{ST} between all 418 samples revealed clusters of low differentiation encompassing samples from the
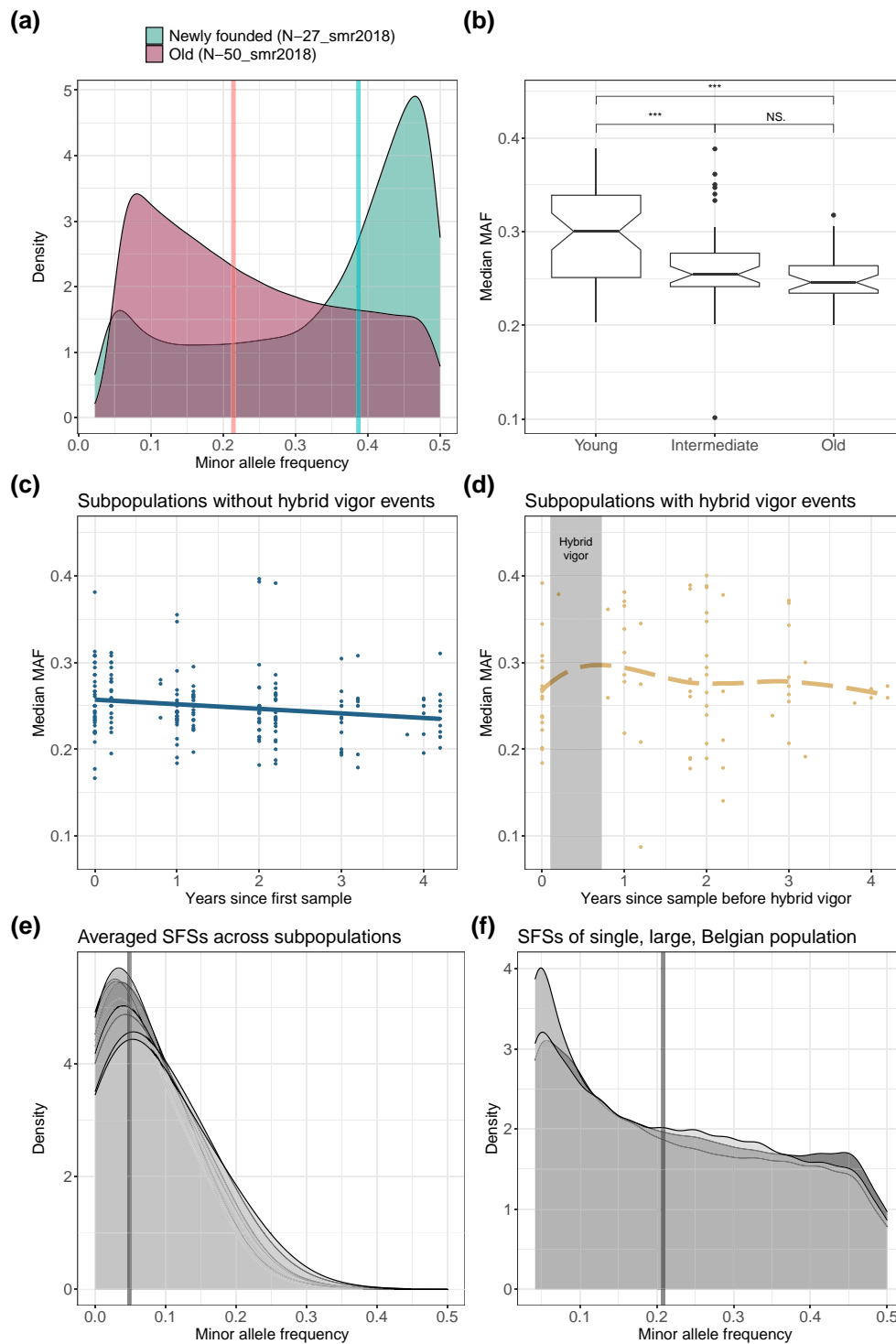
Fig. 3. MAF distributions and their median in subpopulations of different age and status. a) Example of a newly founded subpopulation with a left-skewed SFS distribution versus. one of the oldest subpopulations (age = 36 yr) with a right-skewed SFS distribution. b) On average, median MAF is higher in younger subpopulations than in older ones. Age class definitions are from Angst, Ameline, et al. (2022): young: ≤ 2 yr (≃ newly founded, $n = 50$), intermediate: >2 and ≤15 ($n = 41$), and old: >15 (= most stable, $n = 33$). c) Median MAF decreases in individual subpopulations during the study period. The regression line is from a simple linear model. d) Subpopulations that showed a hybrid vigor event (indicated with a shaded area) deviate from those without such an event by an initial increase in allele frequencies, as revealed by local regressions with LOESS (dashed line). e) In the overall metapopulation, the median MAF is stable, and the SFS distribution resembles a classical population. Each of the ten curves represents allele frequencies averaged across all subpopulations at a given sampling point. Alleles close to a frequency of 0.5 are rare, indicating that only a few alleles are common in all subpopulations. f) In a single, large, stable population from Oud-Heverlee, Belgium, median MAF is stable and similar to that of old subpopulations from the metapopulation. Each of the three curves is from a different population sample ($n = 12$), separated by about 9 yr. The smallest possible allele frequency for these samples is $1/(2 \times 12)$. In a), e), and f), bold vertical lines represent the median MAF. Colors in c) and d) are used as in Fig. 2.
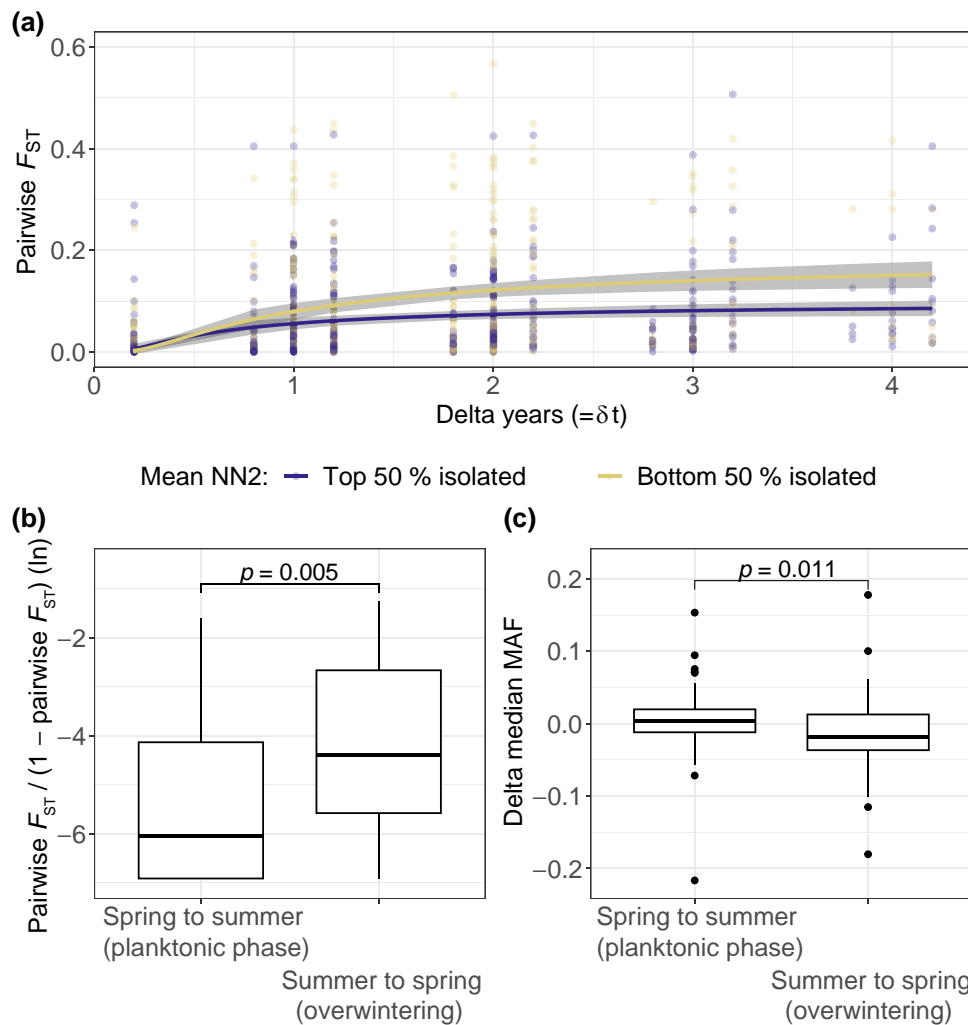
Fig. 4. Variation in genome-wide allele frequencies among temporally separated samples measured as pairwise F_{ST}. a) Two categorical subsets of subpopulations based on their degree of isolation (NN2). Lines represent predicted F_{ST} values based on nonlinear regression ($F_{ST} = ab^{1/\partial t}$) as used in Bergland et al. (2014) with 95% confidence intervals in gray. The line for the more isolated ponds (top 50% isolated) is lower than for the less isolated ponds (bottom 50% isolated), indicating less pronounced changes in observed allele frequencies, possibly due to lower local gene flow. b) Pairwise F_{ST} between summer and the following spring samples is larger than between samples from the same year. Untransformed median pairwise F_{ST}s of the two groups are 0.012 and 0.001, respectively. c) Median MAF decreases on average between the summer and the following spring samples (delta median MAF = −0.013), unlike between samples from the same year where median MAF increases slightly (delta median MAF = 0.005). In b) and c), individual groups were compared with paired Wilcoxon tests.

same subpopulation and island (supplementary fig. S5, Supplementary Material online). This result coincides with previously described patterns of isolation by distance (IBD) (Haag et al. 2006; Roulin et al. 2016; Angst, Ameline, et al. 2022), indicating that geographically proximate samples are typically less differentiated than more distant samples (Table 1I).

Simulations

Our simulations focused on isolating the effects of individual processes to help understanding how they shape the SFS. We simulated newly founded populations, tracking changes in their allele frequency under different scenarios for 36 yr, the age of the oldest subpopulations in our empirical data set. For median MAF, we observed the same patterns as in natural subpopulations: in the base

simulation, i.e. without immigration or bottlenecks, median MAF decreased with population age (−0.004/yr; Fig. 5a). This decrease was stronger in the early years (−0.033/yr), which corresponds to young subpopulations in the empirical data, and during the winter (with sexual reproduction) (mean = −0.013) than during the planktonic phase (only asexual reproduction) (mean = 0.006; Fig. 5b). The winter decrease in median MAF can be attributed to increased allele frequency dynamics caused by sexual reproduction and genetic drift. These dynamics are much weaker without recombination (supplementary fig. S6, Supplementary Material online). In contrast, the increase in median MAF is caused by clonal expansion (for which there is greater potential after overwintering bottlenecks) and clonal selection. Investigating the effects of genetic drift and newly arising mutations, we found that median MAF in larger populations was initially more
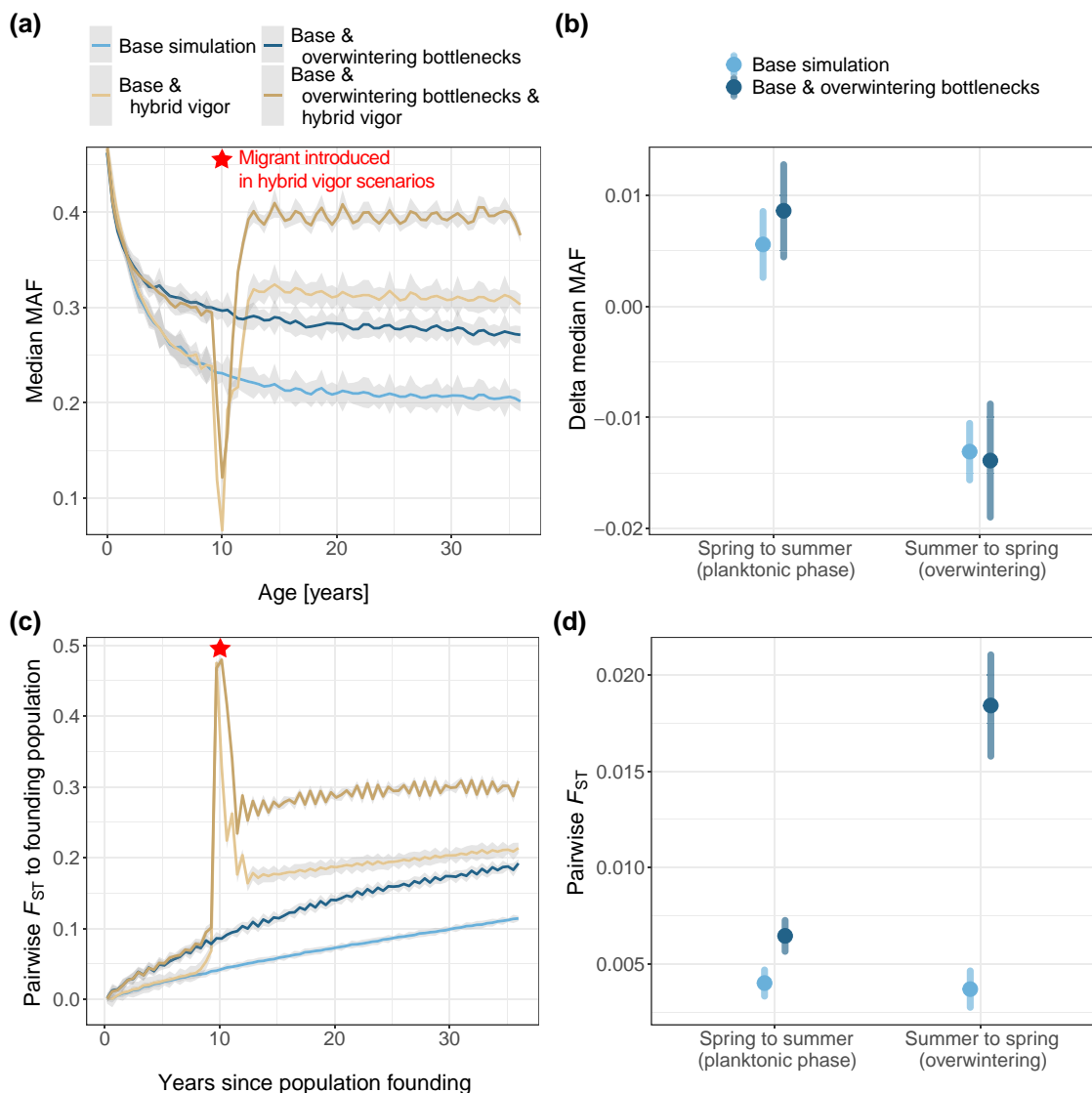
Fig. 5. Simulations of processes relevant to the propagule model adapted to the *D. magna* metapopulation. a) Median MAF decreases with population age, especially in the early years (base simulation). Higher median MAF values are found if rare alleles are removed by overwintering bottlenecks (base and overwintering bottlenecks) and/or if the alleles of immigrants spread through hybrid vigor (base and hybrid vigor). Migrants are introduced in the hybrid vigor scenarios after 10 yr, indicated with a star. b) The decrease in median MAF is mainly associated with overwintering. In contrast, clonal selection likely increases median MAF during the planktonic phase. c) The evolution of pairwise F_{ST} of a population, relative to its founding population, under the different scenarios outlined in the legend of a) suggests a temporal change in genome-wide allele frequencies, especially in the hybrid vigor scenario. d) Populations are more different before and after overwintering bottlenecks (if they occur) than at the beginning and end of the planktonic phase. a) and c) depict cubic spline regression lines along with 95% confidence intervals; the error bars in b) and d) represent 95% confidence intervals around the mean.

stable, close to frequencies of 0.5, suggesting less impact from genetic drift. Subsequently, however, median MAF decreased more in larger populations, because of the presence of more novel mutations, compared with smaller populations (supplementary fig. S7, Supplementary Material online).

In the simulated scenario focused on hybrid vigor (with migration), successful immigration was observed in about half of all replicates, leading to an average increase in median MAF of 0.065 within 2 yr after migrant introduction. Upon the immigrant's arrival, we observed a steep trough in median MAF because even though mutations of the unrelated immigrant are rare in the population, they make up

a significant portion of all mutations (Fig. 5a). Observing such a trough in a natural subpopulation sample would be unlikely because immigrant alleles are too rare to be picked up by empirical methodology, and immigrants in nature are likely more genetically similar to residents (they mostly arrive from nearby ponds) introducing fewer initially rare alleles than in the simulated scenarios. Immigrant alleles in the empirical samples would likely be detected only after selection for hybrids raised their frequencies. After being rare initially, immigrant alleles increased in frequency and remained at intermediate frequency in our simulations. This is because the population becomes dominated by hybrids, with deleterious
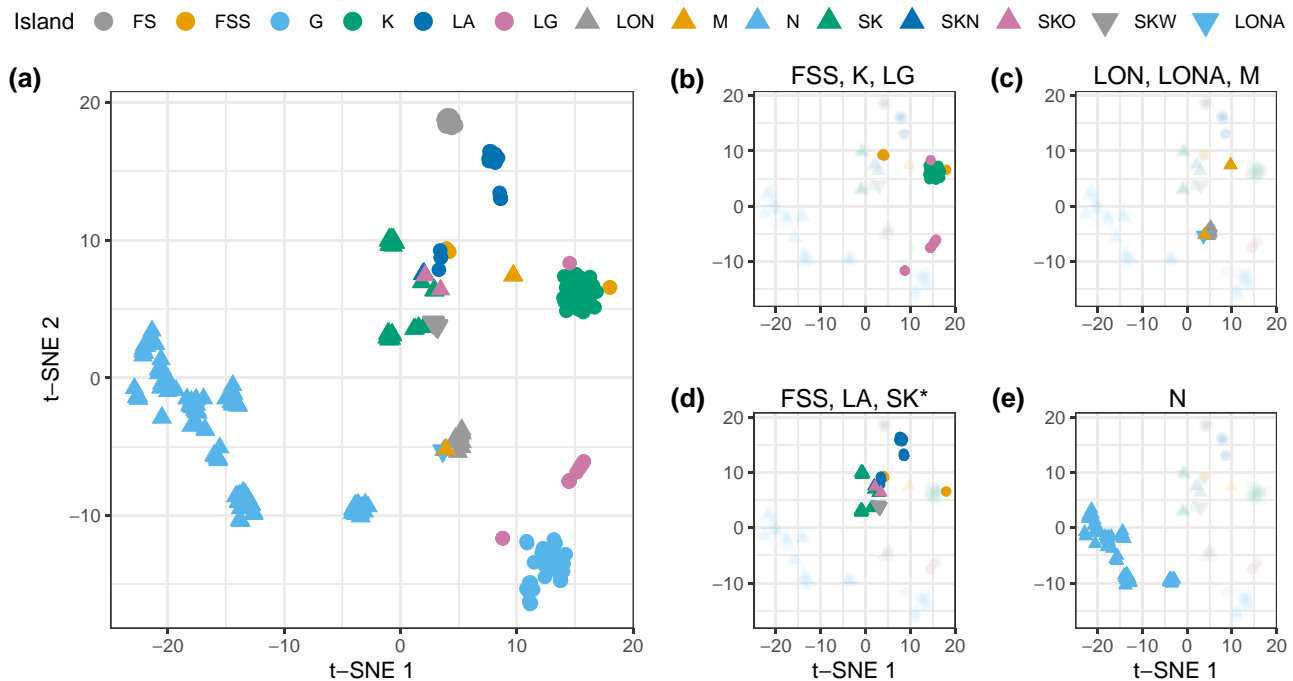
Fig. 6. Spatial population structure of the metapopulation. Dimensionality reduction of whole-genome allele frequencies using t-SNE reveals clustering of samples from the same island in most cases a). b) to e) highlight samples from different islands clustered together and samples from the same island separated into multiple clusters, indicating long-distance dispersal and distinct colonization events. Shape and color of symbols indicate the island of origin (see legend) and SK* stands for the island group SK, SKN, SKO, and SKW.

mutations from both parents being masked in a recessive state.

Regarding allelic variation, the simulated data support our empirical finding that there is a nonlinear temporal increase in pairwise F_{ST} (AIC nonlinear: $-387,258.7$; AIC linear: $-383,872.5$; Fig. 5c). In the hybrid vigor scenario, we observed a sharp peak in pairwise F_{ST} when comparing the population at the time of the migrant's introduction to the time of its founding. This pattern corresponded to the concurrently observed trough in median MAF. While pairwise F_{ST} in the founding population increased upon immigrant introduction, a decrease then follows because some immigrant alleles are lost, and the resident population's allele frequencies are pushed back to frequencies close to 0.5 because of selection for the hybrid offspring (hybrid vigor), making them slightly more similar to the founding population. In populations without successful gene flow, pairwise F_{ST} decreases because the immigrant alleles are disappearing. Our simulations further suggest that genetic bottlenecks associated with overwintering, e.g. when only a portion of animals hatch from resting stages or when members of the parent population contribute unequally to the resting egg pool, enhance genetic differentiation over winter: simulated pairwise F_{ST} between consecutive summer and spring samples (mean = 0.018) was higher than simulated pairwise F_{ST} between consecutive spring and summer samples (mean = 0.006) only in a scenario with bottlenecks over winter, not in scenarios without bottlenecks (both means = 0.004; Fig. 5d). In the simulated populations, θ was stable except

for the hybrid vigor scenario, where we observed an almost 3-fold average increase.

Population Structure

As observed earlier (Angst, Ameline, et al. 2022), the overall structure of our larger genomic data set reflected the spatial structure of the focal metapopulation: most geographically close samples (from the same pond or island) formed clusters (Fig. 6). Also, samples from close neighboring islands, such as LON and LONA, as well as SK, SKN, SKO, and SKW, formed single clusters. This clustering supports the observation of IBD. However, despite strong clustering by island, we see evidence for interisland dispersal: some samples from different islands cluster together (Table 1J). For example, island K's cluster includes a subpopulation each from FSS and LG (Fig. 6b), a subpopulation from M clusters with samples from LON and LONA (Fig. 6c), and the SK island group clusters with a subpopulation from each FSS and LA (Fig. 6d). Furthermore, subpopulations form some islands separate into multiple clusters (e.g. FSS, LA, LG, M, and N) (Fig. 6), whereby the ancestors of the subpopulations of these subclusters presumably originate from distinct colonization events (supplementary fig. S8, Supplementary Material online).

Discussion

We used intensive spatiotemporal genomic sampling to investigate how general and metapopulation-specific

processes shape the evolution of genome-wide allele frequency patterns in an ecologically well-studied dynamic metapopulation (Pajunen 1986; Pajunen and Pajunen 2003; Ebert et al. 2013; Dubart et al. 2020). Previous studies have suggested that this metapopulation follows the propagule model (Zumbrunn 2011; Ebert et al. 2013; Fields et al. 2018; Dubart et al. 2020; Angst, Ameline, et al. 2022), which this study corroborates. We found that genomic bottlenecks during subpopulation founding skew the distribution of genome-wide allele frequencies to intermediate frequencies. We hypothesized that microevolutionary processes, including gene flow, genetic drift, and newly arising mutations, gradually shift SFS distributions away from intermediate frequencies. Our analysis confirmed this hypothesis and further revealed shifts in the opposite direction in some subpopulations, likely due to the rapid spread of hybrid offspring from inbred resident and immigrant genotypes, i.e. hybrid vigor.

Deepening Our Understanding of the Propagule Model

By unraveling the dynamics of SFS distributions in the *D. magna* metapopulation and by applying numerical simulation, we advance the understanding of how metapopulations evolve and explore the population genomic consequences of the propagule model. Our analyses of both empirical and simulated data quantify the contribution of previously observed and newly discovered factors in metapopulation evolution and support the hypothesis of evolution under the propagule model by showing that newly founded subpopulations have high median MAF, consistent with the idea that they are founded by a single or a small number of animals (Haag et al. 2006). This finding was corroborated by simulations and is also supported by previous studies on turnover dynamics, the number of colonizers, and mitochondrial as well as nuclear genomic data, which have suggested evolution under the propagule model for this metapopulation (Zumbrunn 2011; Ebert et al. 2013; Fields et al. 2018; Dubart et al. 2020; Angst, Ameline, et al. 2022). A previously unknown phenomenon is that changes in the local allele frequencies were considerably stronger from summer to the following spring (the overwintering period) than from spring to the following summer. This seasonal pattern highlights the previously unknown role that processes related to sexual reproduction (resulting in resting stages), resting, and hatching play in the evolution of SFSs in subpopulations and, more generally, highlights the advantage of our holistic approach. Sexual reproduction in cyclic parthenogens involves various stochastic aspects, which may lead to unequal contribution of members of the planktonic population to the resting egg pool and to the following planktonic population. For example, genotypes have an unequal propensity to reproduce sexually (Yampolsky 1992), the production and survival of resting stages can vary, and hatching can be temporally asynchronous. Thus, overwintering may be associated with reductions

in the effective population size, genetic bottlenecks, and the previously described increased risks of extinction (Dubart et al. 2020). Furthermore, we found clear evidence for long-distance dispersal. Previously thought to be rare in the focal metapopulation (Pajunen 1986; Pajunen and Pajunen 2003; Dubart et al. 2020), we observed at least five cases where long-distance dispersal between islands up to 500 m apart is the most plausible explanation (Fig. 6). Dispersal also led to hybrid vigor, which can act on hybrids of resident and immigrant genotypes at the earliest in the planktonic season of the year after a dispersal event, i.e. after sexual reproduction. We observed in about 22% of all subpopulations for which multiple samples were available (Fig. 2). The observed temporal increase in median MAF in subpopulations with hybrid vigor has been suggested by previous studies (Ebert et al. 2002; Lohr and Haag 2015). Our data suggest a hybrid vigor event can occur in a given subpopulation about once every 21 yr or at a chance rate of about 5%/yr. The remarkably similar values of hybrid vigor rate and colonization rate suggest that about 5% of occupied and unoccupied ponds in this metapopulation are likely to experience successful immigration each year.

Prior to this study, the genetic model of the focal metapopulation was inferred from empirical and experimental studies that provided valuable insights and also had distinct inferential weaknesses. First, the empirical studies of natural subpopulations were based on single cross-sections, and the effects of time were approximated using subpopulations of different ages (Haag et al. 2005; Cabalzar et al. 2019; Angst, Ameline, et al. 2022). Subpopulation age was thereby confounded with other potentially age-related factors like habitat stability and quality, infection status, and isolation. Second, although experiments demonstrated strong effects such as hybrid vigor, the relative importance of these effects for the overall evolutionary process could not be assessed (Ebert et al. 2002; Lohr and Haag 2015). Third, potential confounding factors were controlled for in the experiments and thus did not contribute to the overall variance (Ebert et al. 2002). Fourth, experiments tended to test contrasting conditions, thus potentially overestimating the strength of a factor. For example, Ebert et al. (2002) tested the effects of inbreeding and gene flow by contrasting inbred versus outbred individuals and distant immigrants versus residents. Although necessary for the chosen experimental design, this approach did not consider that variation is not bimodal but continuous. Overall, the picture of the evolutionary processes in this metapopulation had been simplified in previous studies, creating the impression of rather strong individual effects. Here, we provide a larger context for those studies—including findings of high turnover dynamics, small numbers of subpopulation founders, low effective population sizes, weak purifying selection, and low rates of adaptive evolution—and present a quantitative picture of metapopulation evolution, as well as compelling evidence for the propagule model, for the strength and frequency of hybrid vigor events, and for

the effects of subpopulation aging. Furthermore, by showing the genomic consequences of high extinction–(re)colonization dynamics, this study demonstrates the effectiveness of using genome-wide allele frequencies to study evolution in metapopulations. We conclude that the SFSs of dynamic metapopulations do not match the norms of classical populations, upon which many widely used population genetics models are based.

Genomic Consequences of Founder Effects and Gene Flow

In line with the propagule model of population founding, we found SFSs skewed to intermediate frequencies in newly founded subpopulations. The median MAF was higher in younger subpopulations than in older ones, suggesting that the SFS distribution shifts to the left over time, converging to a right-skewed SFS distribution. We observed this change in real time as we followed the genome-wide allele frequencies of subpopulations for up to 5 yr. This is supported by in silico simulations. The use of the median to quantify the distribution of genome-wide MAFs is robust in that it depends little on the exact shape of the SFS. Given the variety of concomitant biological processes in the focal metapopulation, this simple measure turned out to be informative.

In contrast to the overall pattern of a decline of median MAF over time, this median increased in some of the subpopulations over time, alongside with rapid and strong genomic differentiation and an increase in the number of heterozygous sites. We interpret these observations as gene flow plus hybrid vigor events because such fast changes are strongly suggestive of a deterministic process, particularly if they are visible on a genome-wide scale. Without a selective advantage, alleles introduced by immigration events or mutation would mostly disappear, because of their rareness. Only a few alleles would stochastically rise in frequency over time. Our study could not directly test for hybrid vigor due to the lack of genotype information in our data; however, Ebert et al. (2002) speculated hybrid vigor to be common in this metapopulation of *D. magna* and showed hybrid offspring resulting from the mating of residents and immigrants outcompeting parental lineages usually within one season. This strong effect may have been a consequence of their experimental design, however, which had hybrid offspring at a relatively high frequency at the beginning of the season. This scenario is unlikely in natural populations, where hybrid offspring will initially be rare and where, therefore, their spread will take longer. Furthermore, in Ebert et al. (2002), the introduced migrants had to differ in marker genes from the residents, so they may have been more differentiated than an average immigrant, which is more likely to originate from a nearby, genetically similar subpopulation. This large differentiation may also have amplified the hybrid vigor effect. Here, we may have missed some hybrid vigor events, as their effects were not visible as clearly. Furthermore, we cannot fully rule out the possibility that

we misclassified some subpopulations as having experienced gene flow plus hybrid vigor, that do, in fact, represent false negative (very fast and undetected) extinction–recolonization events. Nonetheless, our method using biannual censuses for identifying extinction–recolonization events has been shown to be robust (Dubart et al. 2020), and the observed substantial increase in the number of heterozygote sites is not expected after a recolonization event.

Cross-sectional samples of this metapopulation, where older subpopulations were more diverse, have suggested increasing genomic diversity (Haag et al. 2005; Angst, Ameline, et al. 2022), while our temporal sampling scheme showed no general increase in the genomic diversity of individual subpopulations. Diversity is expected to increase as a consequence of gene flow (plus hybrid vigor), selection for heterozygote genotypes (clonal selection for less inbred genotypes) (Haag and Ebert 2007), and de novo mutations. It may decrease as a consequence of genetic drift and population bottlenecks. Selection for heterozygotes is observed here and includes hybrid vigor after gene flow, which was observed only in a few subpopulations, and selection for heterozygotes resulting from variation in local inbreeding. The latter has been used to explain small increases in heterozygosity in rock pool populations during the planktonic phase (Haag and Ebert 2007) and is a likely explanation for the here observed increase in median MAF during the summer, which has a greater intensity in the simulations than what can be detected from the noisier empirical data. The effect of de novo mutations is unlikely visible given the low mutation rates and the short time scale of our study. Genetic drift is a strong force, in particular in small populations and in our *D. magna* metapopulation (Angst, Ameline, et al. 2022), and may lead to the random fixation of alleles and therefore lower local diversity (Pannell and Charlesworth 2000). Genetic drift is even stronger if genetic bottlenecks occur, as seems to be the case during the overwintering of subpopulations. Taken together, the absence of an overall change in genomic diversity over time in our current study reflects the conflicting trends of selection for heterozygotes versus genetic drift and bottlenecks. As these processes differ among habitat patches, they will result in variable evolutionary trajectories across subpopulations. These considerations highlight the difficulties in predicting a general trend for the change of genomic diversity over time. The absence of an increase in genomic diversity over time in the current study, while older subpopulations have been reported to be more diverse than younger subpopulations based on a single cross-section of our data and an independent data set (Haag et al. 2005; Angst, Ameline, et al. 2022), has, however, a different explanation. Older subpopulations are a nonrandom, small subset of surviving subpopulations. Most subpopulations have a low life expectancy, with about 80% of subpopulations not becoming older than 3 yr, while some subpopulations survived for as long as the biennial census was done (>36 yr) (D. Ebert, pers. observation) (Pajunen and Pajunen 2003). The few

old subpopulations are typically associated with larger pond (and thus population) size and higher habitat patch stability, both of which contributes to their higher survival chances (Pajunen 2003, 2007). As a consequence, they may have experienced more hybrid vigor events and less genetic drift and, thus, have higher genetic diversity than a random sample of young subpopulations. Therefore, old subpopulations are a nonrandom sample of all subpopulations, so conclusions drawn from them regarding the evolutionary process are biased.

Our longitudinal sampling revealed that pairwise F_{ST} increased with time between samples of the same subpopulation (Fig. 4a), although samples from the same subpopulation and island remain more similar to each other than to others (Fig. 6; supplementary figs. S5 and S8, Supplementary Material online). Bergland et al. (2014) likewise found an increase in F_{ST} with time between samples in a natural *Drosophila melanogaster* population. They attributed this change to rapid response to selection driven by seasonal changes. For our *D. magna* metapopulation, genetic drift likely contributes to this increase because of its dominant role in this metapopulation and because adaptation to the local environment is thought to be weak (Roulin et al. 2015; Cabalzar et al. 2019; Angst, Ameline, et al. 2022). Gene flow also likely contributes to the gradual rise in population differentiation over time between samples from the same subpopulation, as suggested by the steeper rise in less isolated samples with more potential for gene flow (Fig. 4a). Our analysis of F_{ST} further shows that changes in allele frequencies are more profound after hatching from sexually produced resting stages in spring than after a period of repeated asexual (clonal) reproduction. After hatching, fewer alleles are at intermediate frequencies. In contrast, intermediate frequency alleles are slightly enriched during the planktonic phase, potentially due to clonal selection associated with heterozygote advantage (Haag and Ebert 2007). Taken together, these results suggest that the period of sexual reproduction, overwintering, and hatching plays an important role in the evolution of metapopulation SFSs. Seasonal differences have recently been highlighted as key drivers of metapopulation dynamics in a demographic study of the same metapopulation (Dubart et al. 2020).

Our results add knowledge to the *D. magna* system and are generalizable to many aquatic organisms in patchy habitats with a similar life cycle, such as ostracods, copepods, rotifers, and other *Cladocera* (Ranta 1982; Mort 1991; Boileau et al. 1992; Brendonck et al. 1998; Okamura and Freeland 2002), many of which also occur in the focal rock pool habitat. We expect many of these systems also evolve under the propagule module and show similar genomic patterns. We further think our findings can be, in part, generalized to many other species with metapopulation ecologies driven by extinction–(re)colonization dynamics, as other major model systems in metapopulation research, for example, the Glanville fritillary butterfly *Melitaea cinxia* and the herb *Silene dioica*, partially follow the propagule model and show patterns of genetic

variation consistent with our findings (Giles et al. 1998; Austin et al. 2011). Even species without a natural metapopulation structure are affected by extinction and (re)colonization dynamics due to the ongoing extinction crisis (Kurland et al. 2023). This study of the highly dynamic *D. magna* metapopulation thus adds empirical data to the scarce genomic resources of dynamic metapopulations, empirical knowledge to the study of how they evolve, and hopefully the kinds of insights that will aid in their conservation.

Potential Limitations due to Metadata and Sampling Procedure

A potential bias in our analysis might arise from our method for estimating subpopulation age. If we found no animals on three or more consecutive habitat patch inspections, we assumed subpopulation extinction. If we only had one or two visits with no animals, we did not assume extinction to account for the potential presence of undetected animals or of animals present only as resting stages. Although this method has been shown to work well for the focal metapopulation overall (Pajunen 1986; Pajunen and Pajunen 2007; Dubart et al. 2020; Angst, Ameline, et al. 2022), some subpopulations may remain undetected for longer periods and, thus, were classified as extinct even though they were not. We tried to correct for such false-positive extinction–recolonization events by comparing summary statistics from the genomic data before and after potential recolonization. In so doing, we confirmed seven true, but detected three false extinction events. Still, even after this correction, some young subpopulations show low median MAF, which seems to contradict the propagule model. These could result from subpopulation founding by multiple individuals, either simultaneously or in short sequence (potentially from multiple source subpopulations or the migrant pool model (Slatkin 1977). Sporadic founding by multiple individuals from the same source subpopulation has previously been suggested based on field observations (Zumbrunn 2011).

The other form of age estimation error occurs when true extinction and recolonization events happen within less than a year. In such cases, we would have estimated subpopulations to be older than they were, which would explain why some intermediate-aged subpopulations show an enrichment of intermediate frequency alleles. Another explanation, apart from hybrid vigor, for intermediate-aged subpopulations with an enrichment of intermediate frequency alleles is the occurrence of strong population bottlenecks, for example, when only very few animals hatch after the resting phase or when one hatchling hatches much earlier than others and has a head start in asexual propagation. Both strong genetic bottlenecks and false-negative extinction–recolonization events likely result in high pairwise F_{ST} between consecutive samples and a decrease in the number of heterozygous

sites, which we observed in about 8% of the subpopulations (Fig. 2).

Further difficulties arose from the field sampling itself. We were not always able to collect the target number of 50 individuals per sample. Some ponds contained low numbers of individuals at the time of sampling, so we had to exclude extant but insufficiently large populations from sampling to avoid disturbing the natural dynamics. We also had to exclude samples with exceptionally low sequencing coverage. Although our measure for summarizing the samples' allele frequencies, the median MAF, is robust regarding the shape of their distribution, low sequencing coverage in pool-seq can lead to biased allele frequency estimates of mainly rare alleles, which are uncommon in this system due to extinction–(re)colonization, clonal reproduction, and other dynamics. For all these reasons, and mainly because of the frequent subpopulation extinction, some ponds' time series were not complete. To avoid an imbalanced study design, we focused analyses on temporal dynamics only on subpopulations with at least 2 yr of sampling success.

Conclusions

Because evolutionary processes in classical populations are, unlike metapopulations, well studied and understood, the assumptions about the population genetics of natural populations are based largely on these stable populations. However, our study highlights that the evolutionary process in dynamic metapopulations with unstable subpopulations deviates strongly from classical populations. This difference is mainly evident in three factors. First, local extinction limits the lifespan of subpopulations and, thus, the degree to which subpopulations can approximate a classical population. Second, (re)colonization of empty patches leads to population bottlenecks and subpopulations with excess alleles at intermediate frequencies, including deleterious alleles. Third, because of genetic bottlenecks, subpopulations attain high levels of homozygosity and genetic load through inbreeding, opening the door for hybrid vigor events after gene flow that mask deleterious mutations in a recessive state. Together, these factors reduce the efficiency of natural selection, while genetic drift plays a major role in the overall evolutionary dynamics. Metapopulation theory provides a framework—the propagule model—for accurately understanding how these processes affect genome-wide allele frequency distributions (Slatkin 1977; Pannell and Charlesworth 1999; Whitlock 2004). Here, we corroborate the theoretical hypotheses of the propagule model and show empirically how subpopulations in a large metapopulation deviate strongly from our understanding of classical populations. We reaffirm concerns that assumptions like panmixia and Hardy–Weinberg equilibrium for natural populations should be tested before applying common population genetic models to study their evolution (De and Durrett 2007; Städler et al. 2009; Sachdeva et al. 2022). Dynamic metapopulations evolve differently.

Materials and Methods

Study System

In this study, we used samples from natural *D. magna* populations inhabiting shallow rock pools (depressions in the bare rock) on the Tvärminne archipelago in Southwestern Finland (59°50′ N, 23°15′ E). These rock pools have a mean volume of about 200 to 300 L and will be referred to as ponds in this paper, so as not to confuse them with the term "pool" from a genomic pool-seq sample. *Daphnia magna* is a planktonic freshwater crustacean whose mentioned populations form a metapopulation with extinction–(re)colonization dynamics and gene flow (Pajunen and Pajunen 2003; Ebert et al. 2013; Dubart et al. 2020). More than 550 ponds, most of which freeze solid in the winter and many of which dry up during summer, have been surveyed biannually since 1982 for the presence of *D. magna* (Pajunen and Pajunen 2003; Ebert et al. 2013). Sexual reproduction in the cyclical parthenogenetic *D. magna* results in resting eggs that enable local populations to survive the inhospitable winter freezes and summer droughts, with strong consequences for the metapopulation dynamics (Altermatt et al. 2008; Altermatt and Ebert 2010). Furthermore, the resting eggs, which are dispersed by wind, water, and birds, are crucial for migration and colonization of vacant habitat patches.

Samples and Sequencing

We collected *D. magna* from all occupied ponds in the core sampling area of the long-term study area during regular biannual visits in late May/early June (= spring samples) and late July/August (= summer samples) from 2014 to 2018 (Fig. 1; supplementary table S1, Supplementary Material online). Collecting individuals, preparing samples, and extracting DNA followed the procedures described in Angst, Ameline, et al. (2022). Briefly, we randomly sampled 50 individuals per pond, which were pooled for DNA extraction and whole-genome sequencing (pool-seq). We reduced nontarget DNA (gut content and microbiota) by applying an antibiotic treatment and inducing gut evacuation. DNA integrity during transport was preserved by adding RNAlater (Ambion); for DNA extraction, we used the Qiagen GenePure DNA Isolation Kit supplementing glycogen (Sigma-Aldrich). Library preparation and sequencing were done by the Quantitative Genomics Facility service platform Basel (D-BSSE, ETH), Switzerland using Kapa PCR-free kits and an Illumina HiSeq 2500 and Illumina NovaSeq 6000 sequencer, respectively. Genomic samples from the spring 2014 (NCBI database; Bioproject ID: PRJNA862292) were used in a previous study (Angst, Ameline, et al. 2022).

Mapping Genomic Reads and Variant Calling

We assessed the quality of the raw paired-end sequencing reads using FastQC v.0.11.8 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc) and trimmed them to

remove low-quality sequences and adapter contamination using Trimmomatic v.0.39 (Bolger et al. 2014). All bioinformatic software were used with default settings unless stated otherwise. We ran FastQC again to confirm successful trimming and interleaved trimmed reads using seqtk v.1.2 mergepe (https://github.com/lh3/seqtk). We used a snakemake (Mölder et al. 2021) workflow described in Angst, Ameline, et al. (2022) (https://github.com/pascalangst/Angst_etal_2022_MBE) to process interleaved reads to a VCF file and used the XINB3 individual genome (version 3.0; BioProject ID: PRJNA624896; Fields et al., in prep.) as the reference genome. This genome originates from a genotype collected from our study area and is, therefore, closely related to the samples used in our study. The snakemake workflow uses the following software: bwa-mem2 v.2.2.1 (Vasimuddin et al. 2019) for mapping reads, SAMtools v.1.7 (Li et al. 2009) for file manipulations, Picard Toolkit v.2.23.9 (Broad Institute 2020) for adding read groups and identifying duplicates, and GATK v.3.8 (McKenna et al. 2010; Van der Auwera et al. 2013) for realigning INDELs and calling SNPs using the UnifiedGenotyper. Furthermore, we estimated the average read depth of each sample using SAMtools function depth. These calculations were performed at sciCORE (https://scicore.unibas.ch/) scientific computing center at the University of Basel. We filtered the resulting VCF file to include high-quality sites (QUAL > 30, MQ > 40, QD > 2.0, and FS < 60) with biallelic SNP variants (i.e. excluding INDEL variants) using vcffilter from the C++ library vcflib v.1.0.0_rc2 (Garrison et al. 2022) and VCFtools v.0.1.16 (Danecek et al. 2011). Because GATK includes uninformative reads in the depth estimate but not in the allelic depth estimates, we recalculated depth estimates based on the allelic depths using VcfFilterJdk v.1f97a34 (Lindenbaum and Redon 2018). Afterward, we masked entries with DP less than ten and larger than twice the mode, calculated for each sample separately using VCFtools and BCFtools v.1.9 (Danecek et al. 2021) and masked such with AD of the minor allele equal to one using VcfFilterJdk. This minor allele read count filter is, in pool-seq studies, a more conservative, coverage-aware filter to avoid sequencing errors than the more commonly used MAF filter (Gautier et al. 2013).

When examining the sequencing data, we noticed that the mapping percentages of genomic sequencing reads for some samples were comparatively low (supplementary table S1, Supplementary Material online). This pattern was not unexpected, given that many samples contained a natural parasite of *D. magna*, the microsporidian parasite *Hamiltosporidium tvaerminnensis* (Angst, Ebert, et al. 2022). Unfortunately, this resulted in low coverage of the host's genome in four samples, which we therefore had to exclude from the analysis (i.e. N-71_spr2014, LON-1_smr2014, N-61_smr2014, and N-89_smr2014; supplementary table S1, Supplementary Material online). Two samples failed the sequencing entirely (i.e. N-71_spr2015 and SKW-1_spr2015; supplementary table S1, Supplementary Material online).

Population Genetic Analyses

We estimated genomic diversity, Watterson's theta (θ), by averaging estimates of 100 kbp windows from grenedalf v.0.2.0-beta (Czech et al. 2023), a recent reimplementation of PoPoolation (Kofler et al. 2011). We chose θ because its calculation is less affected by the SNPs' allele frequencies than π. Genome-wide allele frequencies were analyzed separately. To investigate spatiotemporal population differentiation, we estimated pairwise genomic differentiation (pairwise F_{ST}) between subpopulations using the R package *poolfstat* v.2.0.0 (Gautier et al. 2022) in R v.4.2.2 (R Core Team 2022). From *poolfstat*'s pooldata object, we also estimated allele frequencies and retained positions with low missingness for conducting t-SNE using the *Rtsne* v.0.15 (Krijthe 2015) R package with nine dimensions kept from the initial PCA step and 5,000 iterations.

During our sampling, we noted ten extinction–recolonization events where a previously occupied pond contained animals after being vacant for at least three consecutive visits. The likelihood of not detecting a subpopulation for three consecutive visits has been estimated at below 2%, as the detection probability of *D. magna* is 0.74 in this survey (Dubart et al. 2020). Seven of these ten recolonizations were true positives based on our interpretation of the genomic data; i.e. they showed high pairwise F_{ST} and SNP turnover before versus after recolonization (supplementary table S2, Supplementary Material online). Thus, we treated samples from before and after recolonization events in these seven ponds as separate subpopulations in our analyses. For the three (apparently) false-positive recolonization events, the four samples collected after the event showed little change in allele constitution compared with samples from the original subpopulations, so it was not clear if they stem from a separate subpopulation and were excluded from analyses requiring a (relative) age estimate. By following F_{ST} and the gain/loss ratio of heterozygous sites over time, we further classified subpopulations into those with gene flow plus hybrid vigor and those without. Gene flow in combination with hybrid vigor is expected to cause comparatively strong differentiation between two consecutive samples as well as an increase in heterozygous sites. For analyses focusing on understanding temporal changes of whole-genome allele frequencies within subpopulations, we only used subpopulations whose sampling period spanned at least two samples from 2 yr. Subpopulations where we could only obtain samples for a shorter time span, an unavoidable consequence of frequent extinctions, were excluded from these analyses because they would be uninformative.

SFSs in Subpopulations of Different Ages

To test our assumption that intermediate frequency alleles are enriched after subpopulation founding, we estimated the median frequency of all minor alleles (median MAF) separately for each subpopulation sample. Then, we tested if median MAF was associated with the $\log_{10}(\text{age} + 1)$ of the subpopulations using a mixed-effect model and the

lme4 v.1.1-21 (Bates et al. 2015) R package with island group of origin and subpopulation of origin (nested within island group) as random effects. For this analysis, closely neighboring islands FS and FSS, LON and LONA, and SK, SKN, SKO, and SKW were combined into island groups, as their shorelines are separated by less than 10 m (Fig. 1). Statistical significance was assessed using the lmerTest v.3.1-3 (Kuznetsova et al. 2017) R package. Subpopulation age was calculated based on the biannual sampling data, which started in 1982 (Pajunen and Pajunen 2003; Ebert et al. 2013). For subpopulations that have been extant since the survey began and whose age is thus not exactly known, age was set to the duration of the survey. We tested if age groups (Young: ≤2 yr, intermediate: >2 and ≤15, and old: >15; following Angst, Ameline, et al. 2022) differed in median MAF using Welch's t-tests. For multiple samples per subpopulation, we used the average of all median MAF estimates from the given subpopulation.

Genome-Wide Allele Frequency Changes

For consecutive samples from the same subpopulations, we tested if median MAF changed over the course of the study period using mixed-effect models with island group of origin and subpopulation of origin (nested within island group) as random effects. We also tested the same model excluding subpopulations with potential gene flow plus hybrid vigor and two subsets of the different age groups using the same model (all the young subpopulations; all intermediate and old subpopulations that did not differ in median MAF in the Welch's t-tests). We used paired Wilcoxon tests in R to test if genome-wide allele frequencies, measured as difference in median MAF, change differently during the planktonic phase (spring to summer) compared with the resting phase (summer to spring). In cases where several consecutive observations of a single subpopulation were available, we randomly chose one to avoid pseudo-replication.

Temporal Changes in Genomic Diversity and Genomic Differentiation

Along with temporal changes in whole-genome allele frequencies, we also investigated changes in genomic diversity and differentiation. For genomic diversity, we tested if a subpopulation's θ was stable over time using consecutive samples from the same subpopulation. Specifically, we used linear mixed-effect models to test if θ was associated with time since the first sampling and included the subpopulation of origin nested within the island group of origin as a random effect in the model. As for our analysis of median MAF, we also tested subsets of the data with the same model. To test whether the association between θ and time depends on the degree of isolation, i.e. if a potential increase in θ over time strengthens when isolation is less, we included the interaction of time and the mean $\log_{10}$ distance to the subpopulation's two nearest neighbors (NN2) using additional models with otherwise identical

parameters. Isolation has been suggested as an important predictor of genomic diversity in this metapopulation and NN2 a useful estimator of it (Angst, Ameline, et al. 2022). Regarding population differentiation, we used pairwise F_{ST} between all samples of the same subpopulation to test if the nonlinear increase described by $F_{ST} = ab^{1/\partial t}$ (Bergland et al. 2014), where ∂t is the time between samples, is a better fit than a simple linear model, i.e. $F_{ST} = a + b\partial t$, using AIC estimation. Furthermore, to test if there is more genomic differentiation or a greater change in θ during the planktonic phase than the resting phase, we used paired Wilcoxon tests in R as described above for median MAF. For plotting, we transformed pairwise F_{ST} using $\ln\{[F_{ST} + 0.001]/[1 - (F_{ST} + 0.001)]\}$.

Empirical Data of Single, Large, Stable Population

To compare subpopulations in our metapopulation with a temporal sample from a single, large, more stable population, we estimated genome-wide allele frequencies for the D. magna population from Oud-Heverlee pond near Leuven, Belgium (50°50′22.16″ N, 4°39′18.16″ E). We downloaded genomic sequences of 36 individual D. magna from GenBank (NCBI database; Bioproject ID: PRJNA344883) that stem from sediment core samples of three times 12 genotypes about 9 yr apart (Chaturvedi et al. 2021) and used the above described bioinformatic pipeline to map sequencing reads and variant calling, except that we used GATK v.4.1.9.0 HaplotypeCaller (Van der Auwera and O'Connor 2020), a more powerful approach for calling variants from single genotypes. In the resulting VCF file, we used the same parameters for filtering SNPs by quality and calculated median MAF for each of the three samples using VCFtools and R.

Simulations

We conducted numerical simulations using the forward simulation software SLiM v.4.0.1 (Haller and Messer 2023) to investigate the isolated effects of evolutionary processes on genome-wide allele frequencies. These simulations were guided by a previous simulation framework that had been calibrated with parameters that reflect the biology of the focal subpopulations (Angst, Ameline, et al. 2022) from a parameter range suggested by the authors of the simulation software (Haller and Messer 2023). Specifically, we adopted the reproduction model (sexual reproduction only every eighth generation, else asexually reproducing), recombination rate (1×10^{-8}), carrying capacity (300), and mutation rate (1×10^{-8}) generating synonymous as well as nonsynonymous mutations (the distribution of fitness effects for nonsynonymous mutations followed a gamma distribution with mean = −0.03 and shape = 0.2; dominance coefficient = 0.1). For each setting (= scenario), we conducted 1,000 replicate simulations of 10-Mbp sequence. A first modeled scenario, the base simulation, was subpopulation founding. For that, an average of 1.5 founders (range of founders: 1 to 7), the suggested number of colonizers in this system (Haag

et al. 2005; Zumbrunn 2011), was taken from a population at mutation–drift equilibrium, i.e. after a burn-in of 10,000 generations. Median MAF and θ of the newly founded subpopulation were calculated before (summer) and after (spring) sexual recombination; pairwise F_{ST} was calculated between all consecutive spring and summer, as well as summer and spring time points. In another scenario built on the base simulation, we included bottlenecks of size $n = 10$ directly after sexual recombination to examine the effects of overwintering bottlenecks. To isolate the effect of recombination from the effect of genetic drift and overwintering bottlenecks, we simulated this and the base scenario with a recombination rate of 0. A further scenario, designed to simulate a hybrid vigor event, introduced an immigrant from an independent population at mutation–drift equilibrium into the base simulation population 10 yr after the population's founding. A final simulation focused on investigating the effects of genetic drift and newly arising mutations on the temporal dynamics of median MAF. For that, we used the parameters of the base simulation and 100 replicates for each of seven different carrying capacities between 100 and 10,000. The simulated data were analyzed in the same manner as the empirical data.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Authors' Contributions

P.A., P.D.F., and D.E. designed the study. P.A. analyzed the data. D.E., C.R.H., and F.B.-A. collected the long-term metapopulation data. D.E. supervised the long-term metapopulation data collection and curated the data. P.D.F., D.E., and F.B.-A. collected and P.A. and P.D.F. prepared the metapopulation samples. P.A. wrote the manuscript. All authors reviewed and approved the final manuscript.

Conflict of Interests

The authors declare that they have no competing interests.

Data Availability

Analysis scripts are available at https://github.com/pascalangst/Angst_etal_2024_MBE, raw data are deposited at the NCBI SRA database (BioProject IDs PRJNA862292), and the filtered VCF file is deposited at https://doi.org/10.6084/m9.figshare.24324100.

References

Altermatt F, Ebert D. Populations in small, ephemeral habitat patches may drive dynamics in a *Daphnia magna* metapopulation. *Ecology*. 2010:**91**(10):2975–2982. https://doi.org/10.1890/09-2016.1.

Altermatt F, Pajunen VI, Ebert D. Climate change affects colonization dynamics in a metacommunity of three *Daphnia* species. *Global Change Biol*. 2008:**14**(6):1209–1220. https://doi.org/10.1111/j.1365-2486.2008.01588.x.

Angst P, Ameline C, Haag CR, Ben-Ami F, Ebert D, Fields PD. Genetic drift shapes the evolution of a highly dynamic metapopulation. *Mol Biol Evol*. 2022:**39**(12):msac264. https://doi.org/10.1093/molbev/msac264.

Angst P, Ebert D, Fields PD. Demographic history shapes genomic variation in an intracellular parasite with a wide geographical distribution. *Mol Ecol*. 2022:**31**(9):2528–2544. https://doi.org/10.1111/mec.16419.

Austin A, Ovaskainen O, Hanski I. Size and genetic composition of the colonizing propagules in a butterfly metapopulation. *Oikos*. 2011:**120**(9):1357–1365. https://doi.org/10.1111/j.1600-0706.2010.18992.x.

Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Statis Soft*. 2015:**67**(1):1–48. https://doi.org/10.18637/jss.v067.i01.

Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. *PLoS Genet*. 2014:**10**(11):11. https://doi.org/10.1371/journal.pgen.1004775.

Boileau MG, Hebert PDN, Schwartz SS. Non-equilibrium gene frequency divergence: persistent founder effects in natural populations. *J Evol Biol*. 1992:**5**(1):25–39. https://doi.org/10.1046/j.1420-9101.1992.5010025.x.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014:**30**(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Brendonck L, de Meester L, Hairston N. Evolutionary and ecological aspects of crustacean diapause. *Arch Hydrobiol Beih Ergebn Limnology*. 1998:**52**:1–561.

Broad Institute. (2020). "Picard toolkit." [Computer software]. Broad Institute, GitHub repository. https://broadinstitute.github.io/picard/.

Buffalo V, Coop G. Estimating the genome-wide contribution of selection to temporal allele frequency change. *Proc Natl Acad Sci U S A*. 2020:**117**(34):20672–20680. https://doi.org/10.1073/pnas.1919039117.

Cabalzar AP, Fields PD, Kato Y, Watanabe H, Ebert D. Parasite-mediated selection in a natural metapopulation of *Daphnia magna*. *Mol Ecol*. 2019:**28**(21):4770–4785. https://doi.org/10.1111/mec.15260.

Charlesworth B. Effective population size and patterns of molecular evolution and variation. *Nat Rev Genet*. 2009:**10**(3):195–205. https://doi.org/10.1038/nrg2526.

Chaturvedi A, Zhou J, Raeymaekers JAM, Czypionka T, Orsini L, Jackson CE, Spanier KI, Shaw JR, Colbourne JK, De Meester L. Extensive standing genetic variation from a small number of founders enables rapid adaptation in *Daphnia*. *Nat Commun*. 2021:**12**:4306. https://doi.org/10.1038/s41467-021-24581-z.

Colombo EH, Anteneodo C. Metapopulation dynamics in a complex ecological landscape. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2015:**92**(2):022714. https://doi.org/10.1103/PhysRevE.92.022714.

Crawford JE, Lazzaro BP. Assessing the accuracy and power of population genetic inference from low-pass next-generation sequencing data. *Front Genet.* 2012:**3**:66. https://doi.org/10.3389/fgene.2012.00066.

Czech L, Spence JP, Expsito-Alonso M. grenedalf: population genetic statistics for the next generation of pool sequencing. arXiv:2306. https://doi.org/10.48550/arXiv.2306.11622, 2023, preprint: not peer reviewed.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. *Bioinformatics.* 2011:**27**(15):2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. Twelve years of SAMtools and BCFtools. *GigaScience.* 2021:**10**(2):giab008. https://doi.org/10.1093/gigascience/giab008.

De A, Durrett R. Stepping-stone spatial structure causes slow decay of linkage disequilibrium and shifts the site frequency Spectrum. *Genetics.* 2007:**176**(2):969–981. https://doi.org/10.1534/genetics.107.071464.

Dubart M, David P, Ben-Ami F, Haag CR, Pajunen VI, Ebert D. Species coexistence and overlapping distributions in a metacommunity are compatible with niche differences and competition at a local scale. bioRxiv 196758. https://doi.org/10.1101/2020.07.10.196758, 10 July 2020, preprint: not peer reviewed.

Ebert D, Haag C, Kirkpatrick M, Riek M, Hottinger JW, Pajunen VI. A selective advantage to immigrant genes in a *Daphnia* metapopulation. *Science.* 2002:**295**(5554):485–488. https://doi.org/10.1126/science.1067485.

Ebert D, Pajunen VI, Hottinger JW. Unsuitable habitat patches lead to severe underestimation of dynamics and gene flow in a zooplankton metapopulation. *J Animal Ecol.* 2013:**82**(4):759–769. https://doi.org/10.1111/1365-2656.12044.

Fields PD, Obbard DJ, McTaggart SJ, Galimov Y, Little TJ, Ebert D. Mitogenome phylogeographic analysis of a planktonic crustacean. *Mol Phylogenet Evol.* 2018:**129**:138–148. https://doi.org/10.1016/j.ympev.2018.06.028.

Garrison E, Kronenberg ZN, Dawson ET, Pedersen BS, Prins P. A spectrum of free software tools for processing the VCF variant call format: vcflib, bio-vcf, cyvcf2, hts-nim and slivar. *PLoS Comput Biol.* 2022:**18**(5):e1009123. https://doi.org/10.1371/journal.pcbi.1009123.

Gautier M, Foucaud J, Gharbi K, Cézard T, Galan M, Loiseau A, Thomson M, Pudlo P, Kerdelhué C, Estoup A. Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Mol Ecol.* 2013:**22**(14):3766–3779. https://doi.org/10.1111/mec.12360.

Gautier M, Vitalis R, Flori L, Estoup A. *F*-statistics estimation and admixture graph construction with pool-seq or allele count data using the R package *poolfstat. Mol Ecol Res.* 2022:**22**(4):1394–1416. https://doi.org/10.1111/1755-0998.13557.

Giles BE, Lundqvist E, Goudet J. Restricted gene flow and subpopulation differentiation in *Silene dioica. Heredity (Edinb).* 1998:**80**(6):715–723. https://doi.org/10.1046/j.1365-2540.1998.00348.x.

Haag CR, Ebert D. Genotypic selection in *Daphnia* populations consisting of inbred sibships. *J Evol Biol.* 2007:**20**(3):881–891. https://doi.org/10.1111/j.1420-9101.2007.01313.x.

Haag CR, Riek M, Hottinger JW, Pajunen VI, Ebert D. Genetic diversity and genetic differentiation in *Daphnia* metapopulations with subpopulations of known age. *Genetics.* 2005:**170**(4):1809–1820. https://doi.org/10.1534/genetics.104.036814.

Haag CR, Riek M, Hottinger JW, Pajunen VI, Ebert D. Founder events as determinants of within-island and among-island genetic structure of *Daphnia* metapopulations. *Heredity (Edinb).* 2006:**96**(2):150–158. https://doi.org/10.1038/sj.hdy.6800774.

Hahn Matthew W. *Molecular population genetics.* Sunderland (MA): Sinauer Associates/Oxford University Press; 2018.

Haller BC, Messer PW. SLim 4: multispecies eco-evolutionary modeling. *Am Nat.* 2023:**201**(5):E127–E139. https://doi.org/10.1086/723601.

Halter M, Du Pasquier L, Ebert D, Fields PD. A genomic analysis of parasite-mediated population differentiation in a metapopulation. bioRxiv 483675. https://doi.org/10.1101/2022.03.10.483675, 2022, preprint: not peer reviewed.2022.

Hanski I. *Metapopulation ecology.* Oxford (UK): OUP; 1999.

Kofler R, Orozco-terWengel P, Maio ND, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlötterer C. PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One.* 2011:**6**(1):e15925. https://doi.org/10.1371/journal.pone.0015925.

Krijthe JH. 2015. Rtsne: T-distributed stochastic neighbor embedding using a Barnes–Hut implementation [Computer software]. https://github.com/jkrijthe/Rtsne.

Kurland S, Ryman N, Hössjer O, Laikre L. Effects of subpopulation extinction on effective size (N_e) of metapopulations. *Conservation Genetics.* 2023:**24**(4):417–433. https://doi.org/10.1007/s10592-023-01510-9.

Kutschera VE, Poelstra JW, Botero-Castro F, Dussex N, Gemmell NJ, Hunt GR, Ritchie MG, Rutz C, Wiberg RAW, Wolf JBW. Purifying selection in corvids is less efficient on islands. *Mol Biol Evol.* 2020:**37**(2):469–474. https://doi.org/10.1093/molbev/msz233.

Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. *J Statist Sof.* 2017:**82**(13):1–26. https://doi.org/10.18637/jss.v082.i13.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009:**25**(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Lindenbaum P, Redon R. Bioalcidae, samjs and vcffilterjs: object-oriented formatters and filters for bioinformatics files. *Bioinformatics.* 2018:**34**(7):1224–1225. https://doi.org/10.1093/bioinformatics/btx734.

Lohr JN, Haag CR. Genetic load, inbreeding depression, and hybrid vigor covary with population size: an empirical evaluation of theoretical predictions. *Evolution.* 2015:**69**(12):3109–3122. https://doi.org/10.1111/evo.12802.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010:**20**(9):1297–1303. https://doi.org/10.1101/gr.107524.110.

Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, Forster J, Lee S, Twardziok SO, Kanitz A, et al. Sustainable data analysis with snakemake. *F1000Res.* 2021:**10**:33. https://doi.org/10.12688/f1000research.29032.2.

Montero-Pau J, Gómez A, Serra M. Founder effects drive the genetic structure of passively dispersed aquatic invertebrates. *Peer J.* 2018:**6**:e6094. https://doi.org/10.7717/peerj.6094.

Mort M. Bridging the gap between ecology and genetics: the case of freshwater zooplankton. *Trends Ecol Evol.* 1991:**6**(2):41–45. https://doi.org/10.1016/0169-5347(91)90120-M.

Okamura B, Freeland JR. *Gene flow and the evolutionary ecology of passively dispersing aquatic invertebrates.* London (UK): Blackwell Scientific Publications; 2002.

Pajunen VI. Distributional dynamics of *Daphnia* species in a rock–pool environment. *Ann Zool Fenn.* 1986:**23**(2):131–140.

Pajunen VI, Pajunen I. Long-term dynamics in rock pool *Daphnia* metapopulations. *Ecography.* 2003:**26**(6):731–738. https://doi.org/10.1111/j.0906-7590.2003.03542.x.

Pajunen VI, Pajunen I. Habitat characteristics contributing to local occupancy and habitat use in rock pool *Daphnia* metapopulations. *Hydrobiologia.* 2007:**592**(1):291–302. https://doi.org/10.1007/s10750-007-0769-4.

Pannell JR, Charlesworth B. Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution*. 1999:**53**(3):664–676. https://doi.org/10.1111/j.1558-5646.1999.tb05362.x.

Pannell JR, Charlesworth B. Effects of metapopulation processes on measures of genetic diversity. *Philos Trans R Soc Lond B Biol Sci*. 2000:**355**(1404):1851–1864. https://doi.org/10.1098/rstb.2000.0740.

Ranta E. Animal communities in rock pools. *Ann Zool Fenn*. 1982:**19**(4):337–347.

R Core Team. 2022. R: the R project for statistical computing [Computer software]. [R Foundation for Statistical Computing. https://www.r-project.org/.

Roulin AC, Bourgeois Y, Stiefel U, Walser J-C, Ebert D. A photoreceptor contributes to the natural variation of diapause induction in *Daphnia magna*. *Mol Biol Evol*. 2016:**33**(12):3194–3204. https://doi.org/10.1093/molbev/msw200.

Roulin AC, Mariadassou M, Hall MD, Walser J-C, Haag C, Ebert D. High genetic variation in resting-stage production in a metapopulation: is there evidence for local adaptation? *Evolution*. 2015:**69**(10):2747–2756. https://doi.org/10.1111/evo.12770.

Saastamoinen M, Bocedi G, Cote J, Legrand D, Guillaume F, Wheat CW, Fronhofer EA, Garcia C, Henry R, Husby A, *et al*. Genetics of dispersal. *Biol Rev*. 2018:**93**(1):574–599. https://doi.org/10.1111/brv.12356.

Sachdeva H, Olusanya O, Barton N. Genetic load and extinction in peripheral populations: the roles of migration, drift and demographic stochasticity. *Philos Trans R Soc B Biol Sci*. 2022:**377**(1846):20210010. https://doi.org/10.1098/rstb.2021.0010.

Slatkin M. Gene flow and genetic drift in a species subject to frequent local extinctions. *Theor Popul Biol*. 1977:**12**(3):253–262. https://doi.org/10.1016/0040-5809(77)90045-4.

Städler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P. The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. *Genetics*. 2009:**182**(1):205–216. https://doi.org/10.1534/genetics.108.094904.

Szép E, Sachdeva H, Barton NH. Polygenic local adaptation in metapopulations: a stochastic eco-evolutionary model. *Evolution*. 2021:**75**(5):1030–1045. https://doi.org/10.1111/evo.14210.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, *et al*. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013:**11**(1110):11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43.

Van der Auwera GA, O'Connor BD. *Genomics in the cloud: using Docker, GATK, and WDL in Terra*. 1st ed. Sebastopol (CA): O'Reilly Media; 2020.

Vasimuddin M, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS); Rio de Janeiro, Brazil. p. 314–324. https://doi.org/10.1109/IPDPS.2019.00041.

Wang J, Hill WG, Charlesworth D, Charlesworth B. Dynamics of inbreeding depression due to deleterious mutations in small populations: mutation parameters and inbreeding rate. *Genet Res*. 1999:**74**(2):165–178. https://doi.org/10.1017/S0016672399003900.

Whitlock MC. 7—selection and drift in metapopulations. In: Hanski I, Gaggiotti OE, editors. *Ecology, genetics and evolution of metapopulations*. Cambridge (MA): Academic Press; 2004. p. 153–173. https://doi.org/10.1016/B978-012323448-3/50009-X.

Whitlock MC, Ingvarsson PK, Hatfield T. Local drift load and the heterosis of interconnected populations. *Heredity (Edinb)*. 2000:**84**(4):452–457. https://doi.org/10.1046/j.1365-2540.2000.00693.x.

Yampolsky LY. Genetic variation in the sexual reproduction rate within a population of a cyclic parthenogen, *Daphnia magna*. *Evolution*. 1992:**46**(3):833–837. https://doi.org/10.1111/j.1558-5646.1992.tb02089.x.

Zhou S, Zhou C, Pannell JR. Genetic load, inbreeding depression and heterosis in an age-structured metapopulation. *J Evol Biol*. 2010:**23**(11):2324–2332. https://doi.org/10.1111/j.1420-9101.2010.02091.x.

Zumbrunn T. (2011). *Daphnia* metacommunity dynamics: the roles of inbreeding, parasitism, competition, and dispersal [Doctoral dissertation, University of Basel]. edoc.unibas.ch. https://doi.org/10.5451/unibas-005651148.