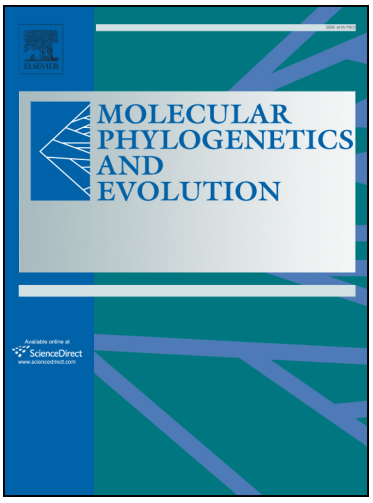
Mitogenome phylogeographic analysis of a planktonic crustacean



Mitogenome phylogeographic analysis of a planktonic crustacean

Peter D. Fields, Darren J. Obbard, Seanna J. McTaggart, Yan Galimov, Tom J. Little, Dieter Ebert

PII: S1055-7903(17)30918-1
DOI: https://doi.org/10.1016/j.ympev.2018.06.028
Reference: YMPEV 6211

To appear in: *Molecular Phylogenetics and Evolution*

Received Date: 18 December 2017
Revised Date: 14 June 2018
Accepted Date: 15 June 2018

Please cite this article as: Fields, P.D., Obbard, D.J., McTaggart, S.J., Galimov, Y., Little, T.J., Ebert, D., Mitogenome phylogeographic analysis of a planktonic crustacean, *Molecular Phylogenetics and Evolution* (2018), doi: https://doi.org/10.1016/j.ympev.2018.06.028

MITOGENOME PHYLOGEOGRAPHIC ANALYSIS OF A PLANKTONIC CRUSTACEAN

Running Head: Mitogenome analysis of Eurasian *D. magna*

Peter D. Fields[1,5], Darren J. Obbard[2,3], Seanna J. McTaggart[2,3], Yan Galimov[4], Tom J. Little[2,3], and Dieter Ebert [1]

[1] University of Basel, Zoological Institute, Vesalgasse 1, Basel, CH-4051, Switzerland

[2] Institute of Evolutionary Biology; School of Biological Sciences University of Edinburgh, Edinburgh, EH9 3JT, UK

[3] Centre for Infection, Immunity, and Evolution, University of Edinburgh, Edinburgh, United Kingdom

[4] Koltsov Institute of Developmental Biology RAS ul. Vavilova 26, 119334 Moscow, Russia

[5] Corresponding author: Email: peter.fields@unibas.ch

ACKNOWLEDGEMENTS

This work was supported by Natural Environment Research Council grant NE/J010790/1 to DJO and TJL, and work in DJOs laboratory was partly supported by a Wellcome Trust strategic award WT095831to the Centre for Immunity, Infection and Evolution. DE and PDF are supported by the Swiss National Science Foundation and the University of Basel.

ABSTRACT

Phylogeography places population genetics in an explicitly spatial context, and in doing so attempts to reconstruct the historical and contemporary evolutionary processes acting across a species range through space and time. Here we present the phylogeographical structure of *Daphnia magna* as determined for full mitochondrial genomes from samples of 60 populations throughout much of the species known range, including Europe, the Middle East, and Asia. Contrary to previous analyses, the present analysis of the mitochondrial genome reveals coarse-grained (continental scale) evidence for spatial structure, and in particular a deep split between Western Eurasia and East Asian *D. magna* lineages. In contrast to previous analyses with nuclear genetic markers, our mitogenomic analysis reveals much less structure within lineages. We quantify divergence between species using the full mitochondrial genome sequence of a closely related species, *D. similis*. The distribution of European and Middle Eastern genetic diversity is consistent with a rapid demographic expansion following the end of the most recent ice age about 10,000 years before present. By estimating species wide distributions of d_N/d_S in mtDNA, we provide evidence that the effectiveness of purifying selection on protein coding genes in the mitochondrial genome of coastal rock pool populations, which have pronounced extinction-colonization dynamics, is reduced compared to larger and more stable non-rock pool populations. The present study adds important insights into the evolutionary history of a widely used model organism in ecology, evolution and ecotoxicology, and highlights the utility of phylogeographic analysis of organellar genomes to understand evolutionary processes.

Keywords: Mitogenome; phylogeography; genetic drift; *Daphnia*; *Daphnia magna*; Palaearctic

INTRODUCTION

Some species have an essentially global distribution, while others are constrained to very limited 'islands' of appropriate habitat. Within these species ranges, population sizes can range from billions, as in the case of many bacterial species, to cases where just a few individuals represent the entire species, as in the case of the so-called 'Big Bird lineage' of Darwin's finch which recently arose as the result of homoploid hybrid speciation (Gaston, 1996; Lamichhaney et al., 2017). The causes of such variation in both range and census size of individual species is of great interest for basic and applied science (Runge et al., 2015). For example, effective population size is known to have a large effect on the efficacy of natural selection (Charlesworth, 2009). Additionally, native range size has been shown to be predictive of invasion potential in some species (Shah et al., 2012)

A species' present-day distribution and density may be very different from what it was in the past, and what it may be in the future. Several approaches can be used to determine past and future species distributions, though some of these are often quite resource and time intensive (Adler et al., 2010). Population genetics approaches use DNA based molecular information to reconstruct a species' demographic history, in so doing understanding the past, present, and possible future of a focal species (Cutter, 2013).

Construction of a species' specific phylogeography is a methodology allowing transition from range wide genetic diversity data to a reconstruction of the historical demography of individual populations, and concomitantly, the species as a whole (Knowles, 2009). Recent developments in model-based inference of phylogeography (Beaumont et al., 2010) have allowed more consistent statistical inference of the historical relationship of different populations. This includes estimates of historical

population sizes, population divergence times, and migration events between individual subpopulations (Maio et al., 2015). Phylogeographic methods have also been greatly advanced by the increase in genetic and genomic resources for a wide range of species resulting from the introduction of affordable next-generation sequencing (NGS) platforms (Bowen et al., 2014; Knowles, 2009).

The majority of phylogeographic and phylogenetic analysis in non-model systems have relied on the use of mitochondrial DNA (mtDNA) and on segments of organellar genomes (Morin et al., 2010; Moritz, 1994). The utility of mtDNA for phylogeographic analysis derives in part from their high rate of molecular evolution as compared to the nuclear genome (Avise, 1989; Brown et al., 1982; Moore, 1995) (though also see Galtier et al., 2009, wherein the authors describe particular limitations and assumptions for the application of mtDNA). Most uses of mtDNA for phylogeographic inference have relied on fragments of individual mitochondrial genes, and so only a limited inference is achieved (Galtier et al. 2009). A useful advance relies on low coverage (~1X) whole genome sequencing ('genome-skimming', (see Cronn et al., 2008; McPherson et al., 2013; Straub et al., 2012), allowing for construction of whole plastid and mitochondrial genomes deriving from high coverage of organelle genomes in the read pool even for non-model systems. Using a combination of iterative baiting and mapping assembly of reads, researchers can use whole genome references of distantly related species where a reference mitogenome is not available to construct whole mitogenomes of a target species, thereby allowing higher resolution phylogeographic analysis.

Daphnia magna is a widely used model in ecology, evolution and ecotoxicology, and has been the focus of a number of previous phylogenetic and phylogeographic studies to understand the species' evolutionary history. Using a 609-bp gene fragment

of the mitochondrial gene COX1, De Gelas and De Meester (2005) showed a general lack of geographic population genetic structure within Europe. However, subsequent work Fields et al. (2015), using a genome wide RADseq derived SNP dataset, suggested that there is a clear geographic population genetic structure in *D. magna* within the species' European range at the nuclear level, exhibiting a 'genes mirror geographic' isolation by distance signature.

Daphnia magna has a large, multi-continental range with very heterogeneous set of habitats. A particular distinction is seen between populations that inhabit persistent fresh-water ponds and lakes and populations that exist in ephemeral, coastal rock pools (Pajunen and Pajunen, 2003). Amongst the many differences between these two habitats is the increasing role of genetic drift via increased founder effects during colonization of pools that previously experienced local extinction (Haag et al., 2005; Pannell and Charlesworth, 2000). Importantly, while selection may occur equally in both persistent ponds and ephemeral rock-pool populations, genetic drift in rock pool metapopulations might be strong enough to diminish the effects of natural selection. Roulin et al. (2013) showed a clear signature of continent-scale local adaptation for sex/resting stage induction in *D. magna*, i.e. traits related to production of resting eggs and males exhibit a signature of local adaptation. However, in a follow up study, where *D. magna* populations across a Finnish rock pool metapopulation were sampled and assayed for sex induction, there was no evidence for local adaptation between populations which are known to occupy larger, more persistent pools vs. those that are prone to both within and among year desiccation (Roulin et al., 2015). A potential explanation for these results is that genetic drift prevents the evolution of locally adapted traits between the different rock pool types within the metapopulation (Walser and Haag 2012).

Here we present results of a large Eurasian sample of *D. magna* for which whole genome, paired-end Illumina sequencing was used to construct reference assisted *de novo* assemblies of 60 mitochondrial genomes. We also obtained a mitochondrial genome of the closely related *D. similis*, providing a useful outgroup for phylogenetic and phylogeographic analysis of *D. magna*. Our phylogeographic analysis provides insights in the deep phylogenetic split between *D. magna* clones deriving from Western Eurasia and Eastern Asia. We estimate the approximate timing of the last major population expansion of *D. magna* into Europe following the last glacial recession using Bayesian Skyline analysis. Finally, we show that, at the mitochondrial genome, significant differences in d_N/d_S ratios of the 13 protein-coding mitochondrial genes are suggestive of an increased role of genetic drift and thus reduced effectiveness of natural selection in some habitats.

MATERIALS AND METHODS

Samples and DNA sequencing

We analyzed complete mitogenomes of 60 *D. magna* clones (55 from Western Eurasia, five from Eastern Asia) and one clone of *D. similis*. The *D. magna* genotypes (clones; *D. magna* can be maintained as stable genotypes under lab conditions due to their cyclical-parthenogenetic life-cycle) used in this study originated either from field collected plankton samples, were hatched from field collected resting eggs, or resulted from inbred crosses in the laboratory (two clones). Field collected planktonic females were brought to the laboratory, and individual females were allowed to reproduce parthenogenetically. These lines were kept in the laboratory under conditions of continuous parthenogenetic reproduction. Field-collected resting eggs (ephippia) were collected on the surface of pond sediments and were washed and stimulated to hatch by

exposure to continuous light under room temperature in well-oxygenated medium. Hatchlings were isolated and iso-female lines were produced and kept under conditions of continuous parthenogenetic reproduction. Two clones were obtained by selfing of field collected females (clones produced by means of parthenogenesis lead to male offspring, which can fertilize sexual eggs of their clonal sisters). These two selfed clones are the parents of a standing *D. magna* QTL panel (Routtu et al. 2010; Roulin et al. 2013; Routtu & Ebert 2014). One clone (from Southern Germany; DE-Iinb1) is the result of one round of selfing, the other clone (from Finland, FI-Xinb3) resulted from three rounds of selfing. The Finnish clone has also been used for the *D. magna* reference genome (V2.4; Daphnia Genome Consortium). *D. similis* was collected in Israel close to the type locality of the species and the clone used in the present study derived from three generations of selfing in the laboratory.

To reduce non-focal DNA in our sequencing reactions (from microbiota and food items), individuals were treated for 72 h with three antibiotics (streptomycin, tetracycline, ampicillin) at a concentration of 50 mg/L each, and the microbiotic treatment was refreshed every 24-hours. Clones were fed with dextran beads (Sephadex 'Small' by Sigma Aldrich: 50 μm diameter) at a concentration of 0.5 g/100 mL to aid gut evacuation (Dukić et al., 2016). Animals were moved out of antibiotics and into 1.5-mL Eppendorf microcentrifuge tubes and excess fluids removed with a sterile pipette. Extraction buffer (Qiagen GenePure DNA Isolation Kit) was subsequently added to the tubes and tissue was disrupted using sterile and DNA-free plastic pestle. The resultant solution was incubated overnight with Proteinase K at 55 °C. RNA was degraded using RNAse treatment for one hour at 37 ° C. Protein removal and DNA precipitation, including the addition of glycogen (Qiagen) to aid DNA precipitation, were done using the Qiagen GenePure DNA Isolation Kit instructions. Resultant purified DNA

was suspended in 40 μL of Qiagen DNA hydration solution and subsequently tested for purity and concentration using a Nanodrop and Qubit 2.0, respectively. Libraries were either prepared using Kapa PCR-free kits and sequenced by the Quantitative Genomics Facility service platform at the Department of Biosystem Science and Engineering (D-BSSE, ETH), in Basel, Switzerland, on an Illumina HiSeq 2000, or were provided to Edinburgh Genomics (NCBI SRA Accessions X-Y) for library preparation using TruSeq DNA Nano gel free kits and paired-end 125nt sequencing using HiSeq v.4 .

mtDNA Assembly and Sequence Alignment

Whole mitochondrial genomes were assembled using MITObim, an approach which relies on a baiting and iterative mapping to assemble the focal genome (Hahn et al., 2013). Briefly, the utilized MITObim workflow begins with short-reads being mapped to conserved regions on a reference sequence, generating an initial reference for the focal genome using modules derived from the MIRA sequence assembler (Chevreux et al. 1999). Iterative fishing of reads with previously identified overlaps from the readpool then takes place. These baited reads are then used to extend the initially generated assembly from step one using MIRA. Steps two and three are then repeated until all gaps are closed and the number of reads fished from the readpool remains stationary (Hahn et al., 2013). The reference mitochondrial genome derived from XINB3 individual genome (V2.4; Daphnia Genome Consortium). Individual PE read datasets were subsampled down to two million reads using SEQTK (https://github.com/lh3/seqtk) four different times, each time using a different seed in order to generate four different input datasets. Each assembly was then interrogated for assembly error and consistency using the following quality control steps. The full read dataset was aligned to each reference using BWA MEM (Li, 2013; Li and Durbin, 2009),

the resulting sam alignment file being subsequently converted to a bam and coordinate sorted using SAMtools (Li et al., 2009). Alignments were then visualized using IGV (Robinson et al., 2011; Thorvaldsdóttir et al., 2013) in order to detect assembly and SNP errors by eye. All four individual mitochondrial assemblies were then aligned to one another using MAFFT v.7.305b (Katoh et al., 2002; Katoh and Standley, 2013). Any assembly specific inconsistencies amongst the four were subsequently interrogated by comparing the aforementioned alignments, resulting in a consensus assembly for each mitochondrial genome.

Individual mitochondrial genomes were independently annotated using the MITOS webserver, with the genetic code 'invertebrate' selected. Resultant annotations were then used to identify each of the thirteen coding genes, two structural rRNA genes, and 22 tRNA genes. Individual genes as well as whole mitochondrial sequences were aligned to one another using the multiple sequence aligner v.7.305b (Katoh et al., 2002; Katoh and Standley, 2013).

Sequence Variation and Pop- and Phylo-genetic Analyses

Measures of sequence variation (π) were computed using Arlequin (Excoffier and Lischer, 2010). We used the R (R core Team 2016) package *fields* (Nychka et al. 2015) to generate a spatial explicit summary of π over the sampled area of the species' range. We used the software BEAST2 (v.2.4.4) (Bouckaert et al., 2014; Drummond et al., 2005) to infer the timescale of *D. magna* mtDNA evolution in Europe and thereby explicitly link observed spatial patterns of π with demographic events using a strict clock Bayesian Skyline Plot (BSP) analysis. The data were divided into four partitions: tRNAs, other non-coding RNAs, four-fold degenerate codons (i.e. those four-fold degenerate in all sequences), and other codons. The two noncoding partitions were

given unlinked HKY substitution models with empirical base frequencies and gamma-distributed rate variation. The two coding partitions were also given unlinked HKY substitution models with empirical base frequencies and gamma-distributed rate variation, but were further sub-partitioned into codon positions 1+2 versus codon position 3. All partitions shared the same tree topology, but had unlinked clocks. To provide a time calibration for the tree, in generations, we applied a fully informative Gaussian prior to the third codon position rate of the four-fold degenerate codons (i.e. four-fold degenerate, and thus putatively neutral sites) of mean 1.37×10^{-7} mutations per site per generation, variance 4.0×10^{-8}. This was derived from the experimental estimate of Xu et al. (2012), which was based on parthenogenetic mutation-accumulation lines and are thus expected to approximately reflect the neutral rate. Note that the Gaussian prior distribution is an approximation of the confidence interval reported by Xu et al. (2012), and that we make no attempt to account for any differences in rate between parthenogenetic and sexual generations. In addition, to infer ancestral locations and population properties, we applied two discrete-trait models (Lemey et al., 2009) to infer the ancestral states of (internal) nodes. The first recorded the broad geographic region from which samples were obtained as traits (Northwest Europe, Nordic, Central Europe, Mediteranean, West Russia) and the rock-pool vs. non-rock pool status of the sample population. We ran five independent chains in each analysis with 50 million iterations (after discarding the first 20% of sampled iterations as burn-in) and samples drawn every 5,000 steps. Convergence was confirmed by visually inspecting the behavior of the sample chain, and ensuring that the effective sample size for posterior estimates (ESS) was greater than 200 using the program Tracer v.1.5 (Drummond and Rambaut 2007; available at http://beast.bio.ed.ac.uk/ Tracer). To provide date estimates for the more ancient divergence between European and Asian lineages of *D. magna*, and to

include *D. similis*, we also ran a constant population-size size coalescent model and expanded dataset, but otherwise using the same partitions, parameters and priors.

To test for isolation-by-distance (IBD) we measured the correlation between genetic relatedness and geographic distance among samples. We converted the multiple sequence alignment (MSA) into a variant call format (VCF) using the FastaToVCF.py script included in the PoMo package (De Maio et al., 2015). Pairwise relatedness between each of the sampled populations was estimated using the Yang et al. ((Yang et al., 2010) estimator as implemented in the VCFtools package (Danecek et al., 2011). We applied the recommendations of (Legendre et al., 2015) in order to determine the relationship between pairwise genetic relatedness and the pairwise distance between samples, wherein our matrix of pairwise relatedness was decomposed into principle components using the R base *stats* function *prcomp* and used as a response variable in a distance-based Moran's eigenvector map (dbMEM) analysis by redundancy analysis (RDA). Our pair-wise geographic distance matrix was used to compute a dbMEM using the R package *adespatial* function dbmem (Dray et al., 2017). RDA analysis was conducted using the R package *vegan* function *rda* (Oksanen et al., 2013), with significance assessed with 1000 permutations.

We estimated signatures of selection on protein-coding genes for specific sub-clades of the *D. magna* phylogeny using the branch models of ω ($\omega = d_N/d_S$; d_N, nonsynonymous substitutions; d_S, synonymous substitutions) implemented in the CODEML program of the package PAML v.4.7 (Yang 2007). To estimate different rates for 'rock pool' (RP) and 'non-rock-pool' (NRP) lineages, internal branches were labelled on the basis of the state inferred by the second BEAST discrete-trait analysis (above). We compared the log-likelihood values between models that had single ω (lnL0) and one that had branch type ω values for RP and NRP (lnL1) using a log-ratio test (LRT).

Significance was assessed using a χ^2 test, with the degrees of freedom (df) equal to one. We assessed the value of 2*(lnL1-lnL0) compared to the expected χ^2 value using the R base (R core Team 2016) function dchisq. To quantify potential uncertainty due to the inferred ancestral state of populations we performed the codon analysis on each of 1000 trees sampled randomly from the posterior tree set, and report bounds based on this distribution. In the course of this analysis, it became apparent that the distribution of terminal branch lengths of 'rock-pool' accessions was substantially different from that of non-rock pool accessions. To elucidate the relative impact of branch length *per se* versus rock-pool status, we ran codon analyses which considered 'long' (> than the approximate mean length of RP external branches, 20k generations) and 'short' (≤ 20k generations) terminal non-rock pool branches separately. We used a permutation test to determine when the ω values for RP were significantly greater than the NRP values. The permutation tests were conducted using the R package *perm* function *permTS* (Fay and Shaw, 2010), with significance assessed with 1000 Monte Carlo replications.

RESULTS

We used whole mitochondrial genome sequences of 60 *D. magna* from single individuals collected from populations throughout most of the species' known range and maintained parthogenetically within the laboratory, as well as a mitochondrial genome from the closely related species *D. similis*. The sequences could be unambiguously aligned, with an average pairwise nucleotide identity of 98.9% (range 97.52% – 100%) between European, Western Asia, and the Middle Eastern Range. The average pairwise nucleotide identity between Western Eurasian and East Asian lineages was lower at 95.9% (range 95.1 – 96.2%). The average pairwise nucleotide identity within East Asian lineages was 99.5% (range 99.14 – 99.6%). Between species, we

observed an average pairwise nucleotide identity of 78.3 % (range 77.6 – 78.4%) separating *D. magna* from *D. similis*, and an average nucleotide identity of 67 % to *D. pulex* (Crease, 1999).

We combined regional samples (Figure 1 A,B; Table 1.) to quantify range wide differences in genetic diversity. Mitochondrial-wide pairwise sequence diversity, or π, showed a clear cline, where π was higher in the Middle East and southern Europe as compared to Northern Europe, as previously reported by (De Gelas and De Meester, 2005) (Figure 2). Our limited sampling in Eastern Asia prevented a meaningful quantification of gradients in genetic diversity in this region.

A possible explanation for the observed cline in genetic diversity along the Middle East to Northern European latitudinal gradient is a rapid demographic expansion from a glacial refugium following the last glacial maximum (LGM), as has been observed in a large number of species (Hewitt, 1996; Holder et al., 1999; Stewart et al., 2010). Thus, we used Bayesian Skyline Analysis (BSA), as implemented in the software BEAST2 (Bouckaert et al., 2014) to detect historical changes in effective population size, or N_e that would accompany a post glacial expansion. We observe a rapid increase in N_e at approximately 10-25 (95% CI) thousand years before present (kyr. BP), consistent with well documented dating of the end of the last ice age (Figure 3A). A second analysis of only the protein coding genes in the mitochondrial genome yielded slightly lower estimates of N_e which is likely the consequence of purifying selection taking place on these genomic components (Figure 3B,C). Finally, BSA analysis of only the five East Asian samples suggests the sampled populations have not experienced recent shifts in N_e associated with range contraction and expansion taking place at the approximate timing of the end of the last ice age (Figure 3, D). The latter result is less surprising given the origin of the populations in areas wherein glacial encroachment is thought to

have not taken place (Ju et al., 2007). Larger samples are likely required to provide definitive estimates of shifts in N_e in the East Asian portion of the species' range.

Western Eurasian samples are divided into two distinct lineages which we refer to as Western Eurasia A and Western Eurasia B (Figure 4). Closer inspection of the membership of each of these groups seems to suggest a great deal of geographical overlap between these two lineages (as was also determined above with simpler, non-model based, distance methods (Suppl. Figure 1)). The observed posterior density of divergence time between Western and Eastern had a median of ~290 thousand generations before present (Gen. BP) [95% credibility interval, derived from the highest posterior density intervals CI = 180-450 Gen. BP]. Divergence time between *D. magna* and *D. similis* was estimated to be approximately 12 times longer (3.35 million GBP, 95% HPDI = 2.25 - 4.90 million Gen. BP).

A dbMEM analysis of the relationship between pairwise genetic relatedness and pairwise distance between sampling locations was significant for the full sample of mtDNA genomes ($p < 0.001$; $R^2 = 0.77$), consistent with a pattern of IBD. However, when the East-Asia samples were removed from the analysis the relationship between pairwise genetic relatedness and pairwise distance between sampling locations was no longer significant ($p = 0.125$; $R^2 = 0.26$), suggesting a much less strong pattern of IBD in the Western-Eurasia sample (Figure 5).

Variation in selection across different branches of *D. magna* was determined by estimating the d_N/d_S ratio, or ω. The ω value for Western Eurasia A and Asia was similar to estimates derived from a range of animal mitochondrial genomes, with a mean value 0.04 and 0.07, respectively, while the point-estimate of ω estimated for Europe group B was nearly three and half times larger with a value of 0.17. We reasoned that this observed increase in ω could be driven by the higher prevalence of rock pool samples in

the Western Eurasia B. To determine whether habitat type might affect ω we undertook additional post-hoc analysis of ω with the inferred phylogeny. We divided lineages into those inferred to be RP or or non-rock pool (NRP) (Figure 6). For RP and NRP genomes we estimated ω values of 0.1603 and 0.0798, respectively, with negligible uncertainty associated with tree topology, branch lengths or ancestral state inference (Figure 7A). A comparison of the two types of ω models with a LRT was significant ($p < 0.001$), suggesting the model which included a separate ω for RP and NRP populations was a better fit to the data. Comparison of RP to NRP values via permutation test suggests that the RP ω is significantly larger compared to NRP ω ($p < 0.001$). To test whether differences in omega between RP and NRP lineages might be solely attributable to differences in the distribution of branch lengths, we performed additional analyses, separating external and internal branches giving ω for external branches as RP 0.1547 and NRP 0.1084, respectively (Figure 7B). Comparison via permutation test again suggested the RP ω is significantly larger compared to NRP ω ($p < 0.001$). Finally, we compared NRP external branches shorter than 20k (the ~ mean length of RP external branches) generations to RP branches using a sample of 1000 posterior trees giving mean ω for RP as 0.159 [95% CI 0.1585, 0.1599] and 0.13 [95% CI 0.1365, 0.1376] for NRP populations (Figure 7C), and that the RP ω is still significantly larger compared to NRP ω ($p < 0.001$).

Great care is taken in order reduce or eliminate laboratory based selection taking place on individual lineages, but we cannot exclude the possibility that some loss of natural diversity has taken place as part of our cultivation procedures. If RP mitochondria are in fact metabolically defective it may be inferred that we are actually only able to observe a subset of the actual variation that exists in natural populations, and so the degree of difference in ω values between RP and NRP populations is actually

an underestimate. Low fitness of rock pool *D. magna* has been reported before, by using phenotypic fitness estimates (Lohr & Haag, 2015).

DISCUSSION

D. magna population sub-structure

Fields et al. (2015) showed that for the nuclear genome *D. magna* shows a clear isolation-by-distance (IBD) pattern of relatedness. Previous research by De Gelas and De Meester (2008), which used a 609-bp mitochondrial gene fragment, found instead a pattern of little to no geographic genetic structure. Here we found similar results based on the entire mitogenome, suggesting that the distribution of genetic diversity represented in the mitochondrial genomes shows little geographical signal, but retains the traces of a rapid, post-glacial expansion. Importantly, these differences in patterns and strength of IBD are scale dependent. There was a very strong signal of IBD across the full sample of mitochondrial genomes covering all of Western Eurasia, the Middle East, and Eastern Asia. Within the Europe and Western Asia lineages there is a much-diminished signal of IBD in the mitochondrial genome as compared the nuclear genome 'genes mirror geography' described in Fields et al. (2015), though the Western Eurasia A and Western Eurasia B lineages may have diverged as long ago as 100k GBP. One could speculate that prior to the LGM the Western Eurasia A and Western Eurasia B lineages may have had a more distinct spatial structuring. It is plausible with the discovery of ancient DNA sources that we will be able to investigate the demographic processes that may have taken place in Eurasia prior to the LGM, as has been possible in other biological systems (Marciniak and Perry, 2017; Orlando and Cooper, 2014).

Fields et al. (2015) identified a distinctly divergent lineage represented by a genotype from Mongolia, but as only a single sample from this geographic region was

available for analysis it was difficult to determine its phylogenetic placement or status as a separate species. However, this same Mongolian clone (MN-DM-1) from Fields et al. (2015) was included in the present study, as well as four additional clones sampled throughout East Asia (Table 1). They show that rather than MN-DM-1 being an anomalous sample, the entire geographic region contains a previously unexplored lineage of the species which diverged from the much better studied European lineages at least 100 thousand Gen. BP. This deep divergence between the *D. magna* lineages raises the possibility that we are in fact looking at different species. Recent work by Reisser *et al.* (2017) suggests that these East Asian lineages are the same species as *D. magna* from Europe, at least under the terms of the biological species concept (Mayr, 1940). Specifically, Reisser et al. (2017) generated crosses of a lineage having an Eastern Asia haplotype sampled in the same vicinity as our RU-YAK clones with a Western Russian clone having a haplotype consistent with other European and Middle Eastern clone. F1 clones were generated successfully and continued to generate viable F2 offspring, though direct assays of fitness of either these F1 or F2 individuals were not made. One might conclude at least at a preliminary level that while the observed divergence between these pairs of lineages is quite large, it is not sufficient to generate entirely different species. The present sampling and molecular marker type does not allow us to determine if admixture has taken place between these two distinct lineages.

De Gelas and De Meester (2005) analyzed a number of within population samples, and showed very high genetic structure among populations and limited haplotype diversity within populations for the sampled mitochondrial gene. This suggests that a single individual per population may be sufficient to capture most mitochondrial diversity at the population level. Importantly, including additional samples from more populations or multiple samples per population may increase

signals of population structure at a very local scale but a lack of IBD across Western Eurasian is unlikely to change.

Glacial refugium

Numerous phylogeographic analyses of species with extant ranges in the Europe and Middle East have suggested Southeastern Europe/Middle East as a likely glacial refugium during the last ice age (Hewitt 1996; Holder *et al.* 1999; Hewitt, 2000; Stewart *et al.* 2010). We show here the same pattern in *D. magna* for samples derived from Europe and Western Asia, though our present sample does not allow us to precisely pinpoint the locality of the refugium of the Western Eurasian clade of *D. magna*. Our BEAST2 Bayesian Skyline analysis suggests a rapid demographic expansion taking place approximately 10Ka, consistent with estimates of the end of the last ice age, when Central and Northern Europe became habitable for *Daphnia*. Importantly, with a significant caveat of limited sample size and the challenges inherent in phylogenetic dating (Graur and Martin, 2004), our analysis of samples from Eastern Asia do not suggest that these populations derived from the same glacial isolation and post-glacial expansion. The East Asian lineages may have escaped the range constricting effects of the LGM entirely, as these lineages existed in a geographic locality which may not have been subject to a major Quaternary glaciation (Sun and Chen, 1991; Zeng et al., 2015), or persisted in a different refugium that we have not sampled. With caution, our present data suggests the presence of only a single glacial refugium for the Europe and Western Asia lineages of *D. magna* as compared to two or more as has been observed in other species in the same geographical area (Hickerson *et al.* 2010; Eckert 2011) (Figure 2). Because the inferred phylogeographical patterns exist at such a large geographical scale we believe our sampling scheme at the population level is unlikely to bias our inference.

Metapopulation dynamics cause strong genetic drift

Most species are known to inhabit a wide range of habitats, often resulting in local adaptation. *Daphnia magna* is no exception, and occupies a huge geographical and ecological range with evidence of local adaptation for diverse traits (De Meester, 1993; Yampolsky et al., 2014). Roulin et al. (2013; 2015) showed clear signatures of local adaptation, both phenotypically and genomically, associated with male and ephippia production associated with different habitat persistence across Eurasia. However, when looking at a much smaller spatial area within a well-studied rock pool metapopulation of the species along the southern coast of Finland, Roulin et al. (2015) found no signature of local adaptation for the same traits. Lohr and Haag (2015) showed that *D. magna* from rock pools age faster than *Daphnia* from non-rock pool populations, suggesting a substantial genetic load. These results are consistent with research by Haag et al. (2005, 2006) and Ebert et al. (2013) revealing that genetic drift, through small population size and frequent founder effects during re-colonization of pools, has a strong effect on the rock-pool populations in southern Finland. With genetic drift dominating microevolutionary processes in rock pool metapopulations, local adaptation on the level of the rock pool and selection against deleterious mutations becomes inefficient. Rock-pool habitat as the one investigated in the above-mentioned studies is the dominant habitat for *D. magna* in the North, with well-known metapopulations described from the Finnish coast, the Swedish East and West coast, the Norwegian coast, the coast along the White Sea and also in North-America at the Northern East coast, Hudson Bay and Canadian Northwest coast.

We sought to determine whether purifying selection acting on mtDNA is generally homogeneous across the different branches of the *D. magna* phylogeny. In

determining the d_N/d_S ratio along different branches of the *D. magna* phylogeny, we identified clear rate heterogeneity, wherein the Europe A branches had a ω (0.04) consistent with strong purifying selection acting on coding mitochondrial genes, while Europe branches B had an ω more than three times as high (0.14). While there is a general lack of spatial structure in the overall phylogeny, the Europe B group does have a noticeably larger number of Nordic samples, which all exist as rock-pools populations. Following this observation, we estimated ω of rock-pool (RP) and not rock-pool (NRP) populations explicitly to determine if habitat type predicts the degree of purifying selection. Consistent with the interpretation that RP populations experience increased genetic drift, we found that the mean ω value was between 1.2 and 2 times as large as compared to NRP populations, depending upon if internal branches were included in the analysis or if branch length thresholds were imposed upon the NRP data subset. We speculate that this difference in the role of drift for RP and NRP populations may not be limited to the mitochondrial genome, but could also be present in genome wide data

CONCLUSIONS

Evolution of nuclear and mitochondrial genomes is characterized by fundamental differences in inheritance, recombination and effective population size (Barr et al., 2005), and thus in differences in selection and genetic drift affecting them. These differences are often reflected in different demographic dynamics for each respective genome. Therefore, reconstruction of a species' demographic history from one or the other genome may not correlate highly. In the current study we expanded the sampling range, sample size and genetic material included compared to previous studies of the focal species, aiming to achieve a comprehensive picture of the phylogeography of *D. magna*. We confirmed the suggestion of a limited spatial structure

at the within continent scale for the mitochondrial genome. We show that this is likely the result of a rapid, post-glacial expansion originating from a Southeast European/Middle-Eastern refugium. By comparing our results to those of Fields et al. (2015), wherein variation was assayed across the nuclear genome, we can discern that the nuclear genome of *D. magna*, and the species as a whole, likely experiences more gene flow so that the genetic structure, which may have arisen shortly after the ice age, has been superseded by frequent movements of alleles. By combining the paired approach of whole nuclear genome and mitogenome population genetics we can discern evolutionary processes both past and present.

The present analyses suggest a role of habitat in modulating the efficiency of natural selection to maintain a relatively consistent nucleotide sequence in the protein coding regions of the mitochondrial genome in rock-pool populations. Expanding our analysis to the nuclear genome will allow for a more rigorous test of this habitat hypothesis. In addition to our analysis, having a physiological assessment of mitochondrial function would be important to determine if the elevated d_N/d_S ratios and the apparent presence of deleterious substitutions reduce mitochondrial function in rock-pool lineages (see Lohr and Haag (2015)). By better characterizing variation in evolutionary processes happening among space, time, and even across different genomes within a single individual of the species, we can provide a more comprehensive understanding of the ecological and evolutionary processes acting in an important model system of basic and applied importance.

DATA ACCESSIBILITY STATEMENT

All analysis scripts as well as raw and processed data are available on Dryad (doi:XXXX), NCBI SRA database (BioProject ID PRJNA#####), and NCBI Genbank Accessions #X:Y. These data include the following:

1. 61 Illumina short-insert PE read files.
2. 61 FASTA formatted whole mitochondrial sequences.
3. BEAST(2) XML files for recreating the presented phylogentic and population genetic analyses.
4. Bash scripts for processing BEAST(2) and codeml outputs
5. R scripts describe range-wide genetic diversity maps and LRT tests.

LITERATURE CITED

Adler, P.B., Ellner, S.P., Levine, J.M., 2010. Coexistence of perennial plants: an embarrassment of niches. Ecol. Lett. 13, 1019-1029.

Avise, J.C., 1989. Gene trees and organismal histories - a phylogenetic approach to population biology. Evolution 43, 1192-1208.

Barr, C., Neiman, M., Taylor, D., 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. New Phytol 168, 39-50.

Beaumont, M.A., Nielsen, R., Robert, C., Hey, J., Gaggiotti, O., Knowles, L., Estoup, A., Panchal, M., Corander, J., Hickerson, M., Sisson, S.A., Fagundes, N., Chikhi, L., Beerli, P., Vitalis, R., Cornuet, J.-M., Huelsenbeck, J., Foll, M., Yang, Z., Rousset, F., Balding, D., Excoffier, L., 2010. In defence of model-based inference in phylogeography REPLY. Mol Ecol 19, 436-446.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comp. Biol. 10, e1003537.

Bowen, B.W., Shanker, K., Yasuda, N., Malay, M.C.D., von der Heyden, S., Paulay, G., Rocha, L.A., Selkoe, K.A., Barber, P.H., Williams, S.T., Lessios, H.A., Crandall, E.D., Bernardi, G., Meyer, C.P., Carpenter, K.E., Toonen, R.J., 2014. Phylogeography unplugged: comparative surveys in the genomic era. Bull. Mar. Sci. 90, 13-46.

Brown, W.M., Prager, E.M., Wang, A., Wilson, A.C., 1982. Mitochondrial DNA sequences of primates - tempo and mode of evolution. J. Mol. Evol. 18, 225-239.

Charlesworth, B., 2009. Effective population size and patterns of molecular evolution and variation. Nat Rev Genet 10, 195-205.

Crease, T.J., 1999. The complete sequence of the mitochondrial genome of Daphnia pulex (Cladocera: Crustacea). Gene 233, 89-99.

Cronn, R., Liston, A., Parks, M., Gernandt, D.S., Shen, R., Mockler, T., 2008. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. Nucleic Acids Res. 36, e122.

Cutter, A.D., 2013. Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory. Mol. Phylogen. Evol. 69, 1172-1185.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., Group, G.P.A., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156-2158.

De Gelas, K., De Meester, L., 2005. Phylogeography of Daphnia magna in Europe. Mol Ecol 14, 753-764.

De Maio, N., Schrempf, D., Kosiol, C., 2015. PoMo: An Allele Frequency-Based Approach for Species Tree Estimation. Syst. Biol. 64, 1018-1031.

De Meester, L., 1993. Genotype, Fish-Mediated Chemical, and Phototactic Behavior in Daphnia Magna. Ecology 74, 1467-1474.

Dray, S., Blanchet, F.G., Borcard, D., Clappe, S., Guenard, T.J., Larocque, G., Legendre, P., Madi, N., Wagner, H., 2017. adespatial: Multivariate Multiscale Spatial Analysis. R package version 0.1-0. https://CRAN.R-project.org/package=adespatial.

Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22.

Dukić, M., Berner, D., Roesti, M., Haag, C.R., Ebert, D., 2016. A high-density genetic map reveals variation in recombination rate across the genome of Daphnia magna. BMC Genet. 17, 137.

Ebert, D., Hottinger, J.W., Pajunen, V.I., 2013. Unsuitable habitat patches lead to severe underestimation of dynamics and gene flow in a zooplankton metapopulation. J. Anim. Ecol. 82, 759-769.

Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Res 10, 564-567.

Fay, M.P., Shaw, P.A., 2010. Exact and Asymptotic Weighted Logrank Tests for Interval Censored Data: The interval R Package. Journal of Statistical Software 36, 1-34.

Fields, P.D., Reisser, C., Dukić, M., Haag, C.R., Ebert, D., 2015. Genes mirror geography in Daphnia magna. Mol Ecol 24, 4521-4536.

Galtier, N., Nabholz, B., GlÉMin, S., Hurst, G.D.D., 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Mol Ecol 18, 4541-4550.

Gaston, K.J., 1996. Species-range-size distributions: patterns, mechanisms and implications. Trends Ecol. Evol. 11, 197-201.

Graur, D., Martin, W., 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. Trends Genet. 20, 80-86.

Haag, C.R., Riek, M., Hottinger, J.W., Pajunen, V.I., Ebert, D., 2005. Genetic diversity and genetic differentiation in Daphnia metapopulations with subpopulations of known age. Genetics 170, 1809-1820.

Haag, C.R., Riek, M., Hottinger, J.W., Pajunen, V.I., Ebert, D., 2006. Founder events as determinants of within-island and among-island genetic structure of Daphnia metapopulations. Heredity 96, 150-158.

Hahn, C., Bachmann, L., Chevreux, B., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Res. 41, e129.

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. Nature 405, 907.

Hewitt, G.M., 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. Biol. J. Linn. Soc. 58, 247-276.

Holder, K., Montgomerie, R., Friesen, V.L., 1999. A Test of the Glacial Refugium Hypothesis using Patterns of Mitochondrial and Nuclear DNA Sequence Variation in Rock Ptarmigan (Lagopus mutus). Evolution 53, 1936-1950.

Ju, L., Wang, H., Jiang, D., 2007. Simulation of the Last Glacial Maximum climate over East Asia with a regional climate model nested in a general circulation model. Palaeogeography, Palaeoclimatology, Palaeoecology 248, 376-390.

Katoh, K., Misawa, K., Kuma, K.i., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059-3066.

Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. 30, 772-780.

Knowles, L.L., 2009. Statistical Phylogeography. Annu Rev Ecol Evol Syst. Annual Reviews, Palo Alto, pp. 593-612.

Lamichhaney, S., Han, F., Webster, M.T., Andersson, L., Grant, B.R., Grant, P.R., 2017. Rapid hybrid speciation in Darwin's finches. Sci.

Legendre, P., Fortin, M.J., Borcard, D., Peres‐Neto, P., 2015. Should the Mantel test be used in spatial analysis? Methods in Ecology and Evolution 6, 1239-1247.

Lemey, P., Rambaut, A., Drummond, A.J., Suchard, M.A., 2009. Bayesian Phylogeography Finds Its Roots. PLoS Comp. Biol. 5, e1000520.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25.

Lohr, J.N., Haag, C.R., 2015. Genetic load, inbreeding depression, and hybrid vigor covary with population size: An empirical evaluation of theoretical predictions. Evolution 69, 3109-3122.

Maio, N.D., Wu, C.-H., O'Reilly, K.M., Wilson, D., 2015. New Routes to Phylogeography. arXiv:1503.08138.

Marciniak, S., Perry, G.H., 2017. Harnessing ancient genomes to study the history of human adaptation. Nat Rev Genet 18, 659.

Mayr, E., 1940. Speciation Phenomena in Birds. The American Naturalist 74, 249-278.

McPherson, H., van der Merwe, M., Delaney, S.K., Edwards, M.A., Henry, R.J., McIntosh, E., Rymer, P.D., Milner, M.L., Siow, J., Rossetto, M., 2013. Capturing chloroplast variation for molecular ecology studies: a simple next generation sequencing approach applied to a rainforest tree. BMC Ecol. 13.

Moore, W.S., 1995. Inferring phylogenies from mtDNA variation - mitochondrial-gene trees versus nuclear-gene trees. Evolution 49, 718-726.

Morin, P.A., Archer, F.I., Foote, A.D., Vilstrup, J., Allen, E.E., Wade, P., Durban, J., Parsons, K., Pitman, R., Li, L., Bouffard, P., Abel Nielsen, S.C., Rasmussen, M., Willerslev, E., Gilbert, M.T.P., Harkins, T., 2010. Complete mitochondrial genome phylogeographic analysis of killer whales (Orcinus orca) indicates multiple species. Genome Res. 20, 908-916.

Moritz, C., 1994. Defining evolutionarily significant units for conservation. Trends Ecol. Evol. 9, 373-375.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2013. vegan: Community Ecology Package.

Orlando, L., Cooper, A., 2014. Using Ancient DNA to Understand Evolutionary and Ecological Processes. Annual Review of Ecology, Evolution, and Systematics 45, 573-598.

Pajunen, V.I., Pajunen, I., 2003. Long-Term Dynamics in Rock pool Daphnia Metapopulations. Ecography 26, 731-738.

Pannell, J.R., Charlesworth, B., 2000. Effects of Metapopulation Processes on Measures of Genetic Diversity. Philosophical Transactions: Biological Sciences 355, 1851-1864.

Reisser, C.M.O., Fasel, D., Hürlimann, E., Dukič, M., Haag-Liautard, C., Thuillier, V., Galimov, Y., Haag, C.R., 2017. Transition from Environmental to Partial Genetic Sex Determination in Daphnia through the Evolution of a Female-Determining Incipient W Chromosome. Mol. Biol. Evol. 34, 575-588.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat Biotech 29, 24-26.

Roulin, A.C., Mariadassou, M., Hall, M.D., Walser, J.-C., Haag, C., Ebert, D., 2015. High genetic variation in resting-stage production in a metapopulation: Is there evidence for local adaptation? Evolution 69, 2747-2756.

Roulin, A.C., Routtu, J., Hall, M.D., Janicke, T., Colson, I., Haag, C.R., Ebert, D., 2013. Local adaptation of sex induction in a facultative sexual crustacean: insights from QTL mapping and natural populations of Daphnia magna. Mol Ecol 22, 3567-3579.

Runge, C.A., Tulloch, A., Hammill, E., Possingham, H.P., Fuller, R.A., 2015. Geographic range size and extinction risk assessment in nomadic species. Conserv. Biol. 29, 865-876.

Shah, M.A., Reshi, Z.A., Lavoie, C., 2012. Predicting plant invasiveness from native range size: clues from the Kashmir Himalaya. Journal of Plant Ecology 5, 167-173.

Stewart, J.R., Lister, A.M., Barnes, I., Dalén, L., 2010. Refugia revisited: individualistic responses of species in space and time. Proceedings of the Royal Society B: Biological Sciences 277, 661-671.

Straub, S.C.K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C., Liston, A., 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. Am. J. Bot. 99, 349-364.

Sun, X., Chen, Y., 1991. Palynological records of the last 11,000 years in China. Quat Sci Rev 10, 537-544.

Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics 14, 178-192.

Yampolsky, L.Y., Zeng, E., Lopez, J., Williams, P.J., Dick, K.B., Colbourne, J.K., Pfrender, M.E., 2014. Functional genomics of acclimation and adaptation in response to thermal stress in Daphnia. BMC Genomics 15, 859.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E., Visscher, P.M., 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42, 565-569.

Zeng, Y.F., Wang, W.T., Liao, W.J., Wang, H.F., Zhang, D.Y., 2015. Multiple glacial refugia for cool‐temperate deciduous trees in northern East Asia: the Mongolian oak as a case study. Mol Ecol 24, 5676-5691.

Table 1. Sample information. AS = East Asia; CE = Central Europe; MD = Mediterranean; ME = Middle East; NE = Northwest Europe; NO = Nordic; WR = Western Russia; RP = Rock pool population; NRP = Non-Rock pool population.

Clone ID	Country	Latitude	Longitude	Region	Rock-pool Status
BE-OHZ	Belgium	50.833333	4.650000	NE	NRP
BE-WH1	Belgium	51.336317	3.348728	NE	NRP
BY-G	Belarus	52.421464	31.013781	WR	NRP
CH-H	Switzerland	47.557769	8.862608	CE	NRP
CH-Z	Switzerland	47.51696	8.831272	CE	NRP
CN-W1	China	30.562365	114.390107	AS	NRP
CZ-N1	Czech Republic	48.775317	16.723528	CE	NRP
CZ-N2	Czech Republic	48.775317	16.723528	CE	NRP
DE-INB1	Germany	48.206375	11.709727	CE	NRP
DE-G1	Germany	54.281929	10.966506	CE	NRP
DE-K35	Germany	48.206647	11.709717	CE	NRP
DE-KA	Germany	50.935361	6.927944	CE	NRP
DE-KN1	Germany	54.176906	10.806683	CE	NRP
DE-S1	Germany	48.779872	9.182967	CE	NRP
DE-S3	Germany	48.806308	9.171758	CE	NRP
DK-RL	Denmark	55.964167	9.596389	CE	NRP
ES-RO	Spain	37.127934	-6.481599	MD	NRP
FI-FAV1	Finland	60.021612	19.902506	NO	RP
FI-FHS2	Finland	60.273783	27.21875	NO	RP
FI-FUT1	Finland	60.34705	27.478483	NO	RP
FI-FSE	Finland	60.257761	22.01515	NO	RP
FI-FSP1	Finland	60.1677	25.794617	NO	RP
FI-SK-58	Finland	59.833045	23.25743	NO	RP
FI-INB3	Finland	59.833183	23.260387	NO	RP
FR-C1	France	43.591583	4.591517	MD	NRP
FR-TR	France	43.738306	3.863302	MD	NRP
GB-C1	United Kingdom	51.734411	-1.336264	NE	NRP
GB-EK1-1	United Kingdom	55.702406	-2.340828	NE	NRP
GB-EK1-32	United Kingdom	55.702425	-2.340828	NE	NRP
GB-EL75	United Kingdom	51.527556	-0.158147	NE	NRP
GB-LK1	United Kingdom	55.707653	-2.190378	NE	NRP
GR-K	Greece	40.703997	23.144222	MD	NRP
HU-HO	Hungary	46.800000	19.133333	CE	NRP
IL-M1	Israel	31.714561	35.05099	ME	NRP
IL-PS	Israel	32.255706	34.853272	ME	NRP
IL-RAM	Israel	33.230073	35.765108	ME	NRP
IL-TY	Israel	32.243663	34.85448	ME	NRP

IR-GG1	Iran	37.918978	46.707003	ME	NRP
IT-ISR1	Italy	43.691944	10.288889	MD	NRP
MN-DM	Mongolia	45.032708	100.660481	AS	NRP
NO-AA	Norway	60.05095	5.074448	NO	RP
PL-KNP	Poland	52.322722	20.730514	CE	NRP
RU-AST1	Russia	45.903611	47.656389	WR	NRP
RU-BAL1	Russia	53.017778	106.885833	AS	NRP
RU-BN	Russia	50.155519	43.3894	WR	NRP
RU-BOL1	Russia	66.449692	33.856688	NO	RP
RU-KOR1	Russia	66.451887	33.799009	NO	RP
RU-R2	Russia	56.425003	37.602672	WR	NRP
RU-RM1	Russia	55.763472	37.581583	WR	NRP
RU-RT21	Russia	45.222097	36.684803	WR	NRP
RU-SPB	Russia	59.811111	30.133056	WR	NRP
RU-VOL	Russia	48.53	44.486944	WR	NRP
RU-YAK1	Russia	61.964047	129.630956	AS	NRP
RU-YAK3	Russia	61.937458	129.638192	AS	NRP
SE-BY	Sweden	55.675447	13.545069	NO	NRP
SE-G2	Sweden	60.25929	18.31342	NO	RP
SE-G3	Sweden	60.2539	18.33434	NO	RP
SE-GN1	Sweden	60.497167	18.431033	NO	RP
D. similis (IL-SIM-A20)	Israel	32.781095	35.407369	-	-

Figure 1. Geographic distribution of the samples used. The 60 *D. magna* samples were derived from Europe and Asia and the one *D. similis* genotype was derived from the Middle-East (Israel). A-B) Within the Europe we divided the samples into the regions Central Europe, Mediterranean, Nordic, Northwestern Europe, Western Russia; there was a single Middle-East region; and all samples from Eastern Asia were combined into a single region. C-D) *D. magna* occurs in two distinct freshwater environments, the first being rock-pools which are relatively shallow and prone to disturbance, and the second being more persistent and stable ponds which are not rock-pools.

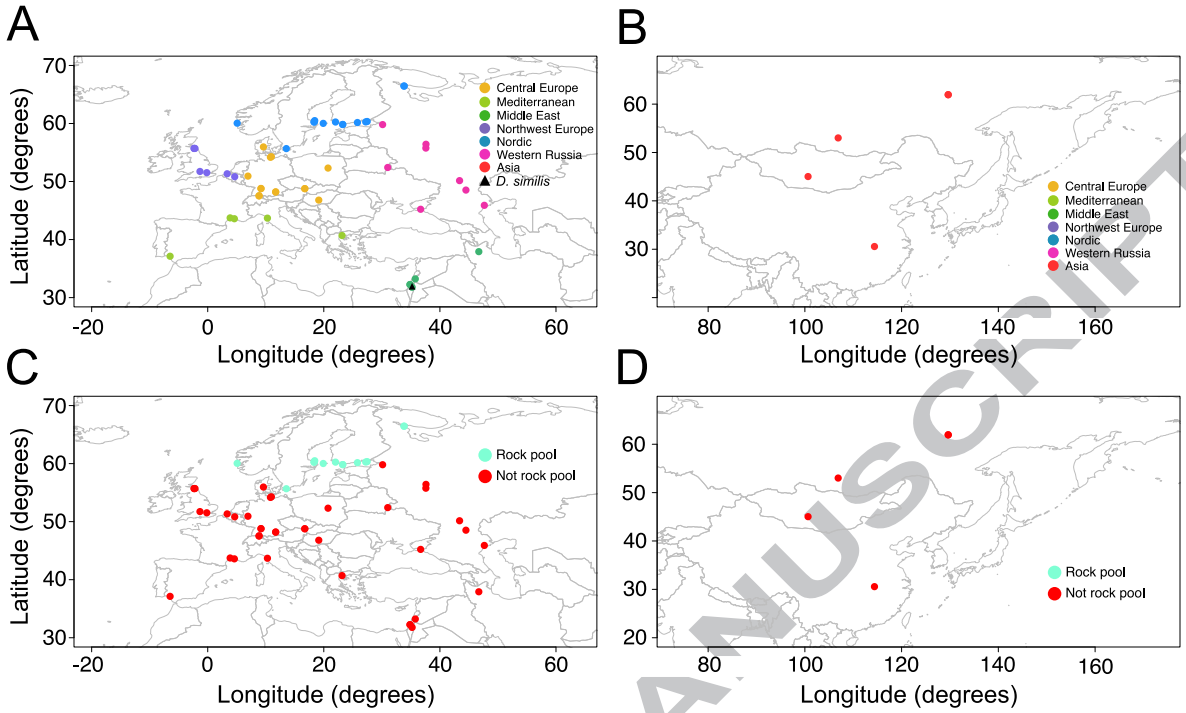


Figure 2. Genetic diversity across species range. Using samples combined by region, a measure of π was estimated using Arlequin. π had a mean value of 0.0799% and a range of 0.037%-0.114%. There is a clear pattern of decreasing genetic diversity in moving from the southeastern to the most northwestern portion of the species range.

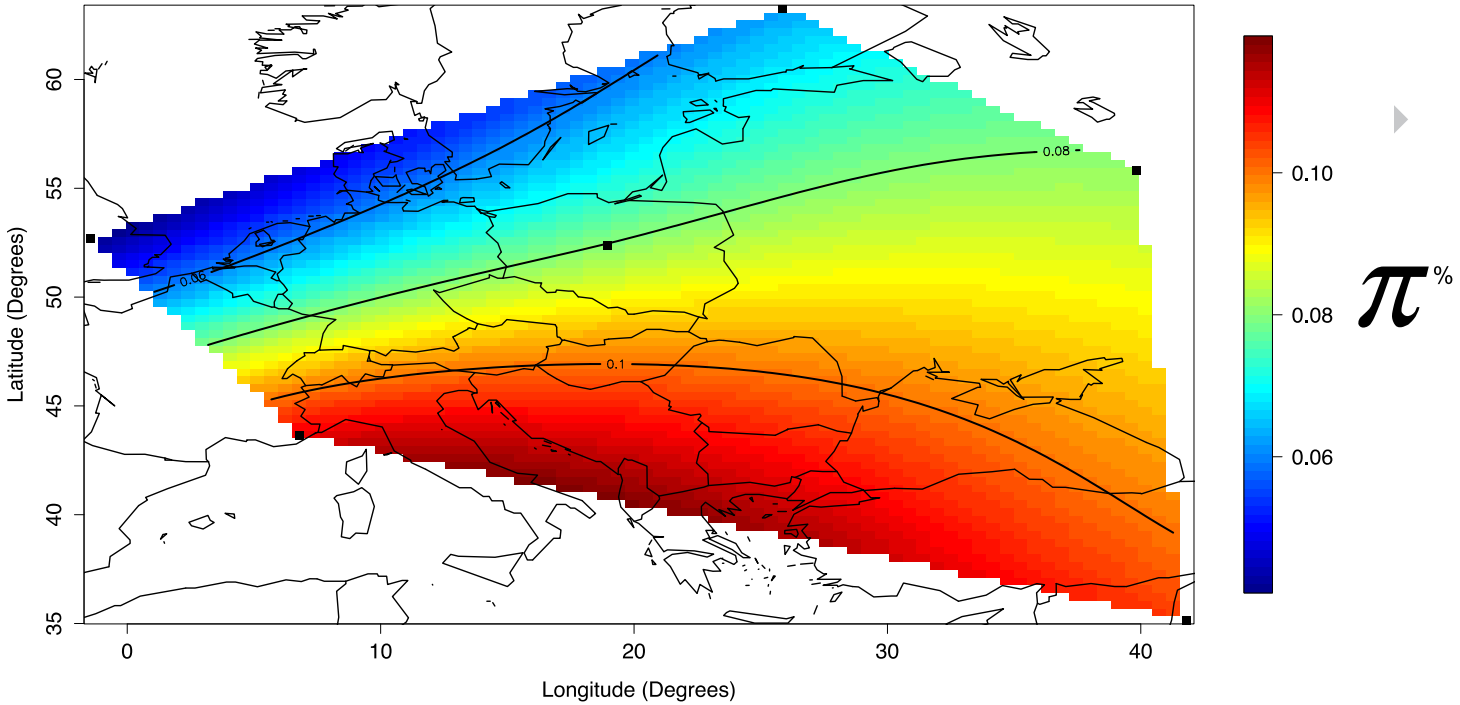


Figure 3. Bayesian Skyline analysis (BSA) of effective population size, N_E. A) When considering the entire mitochondrial genome, inclusive of coding and non-coding regions, we derive a median estimate of present N_E for the mitochondrial genome in the European range of the species to be ~ 350 thousand, and to have experienced a drastic increase approximately 10kyr. BP consistent with the end of large scale glaciation in Europe. B-C) The sample BSA applied to only the protein coding regions of the genome, and on Western Eurasian groups A and B, two genetically but not geographically distinct lineages, (see figure 5) showed a slightly lower estimate of N_E, likely the result of purifying selection acting on the protein coding portion of the mitochondrial genome. D) A BSA applied to East Asian samples did not show a significant shift in N_E though our comparatively limited sample size prevents robust conclusions regarding the demographic history of this region.

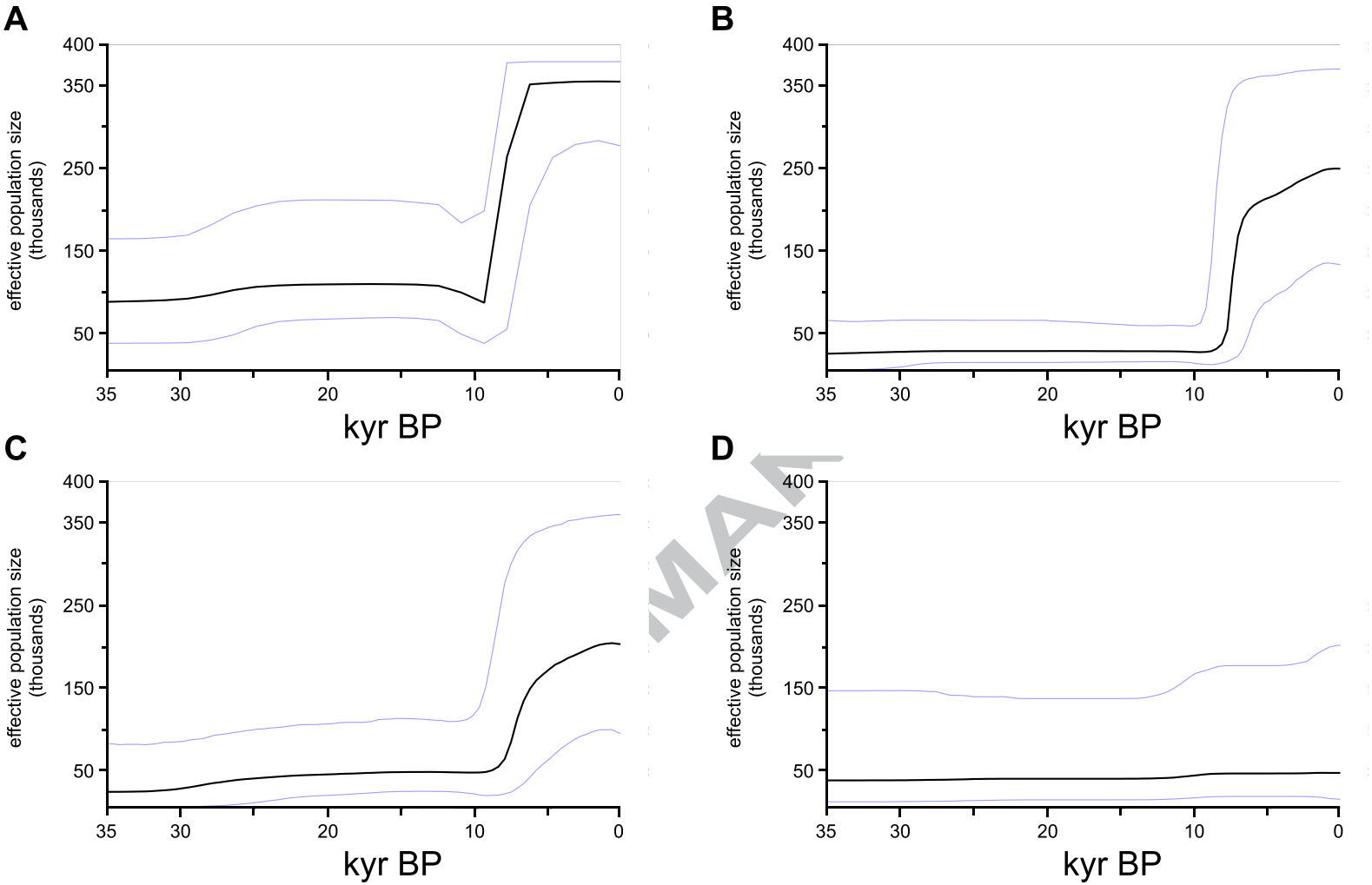


Figure 4. Phylogeny of *D. magna* and *D. similis* mitochondrial genomes. The topology of the tree comparing phylogenetic relationships amongst *D. magna* clones suggest a major split between Western Eurasian and Eastern Asian genotypes which is clear geographic signal which can be estimated to have taken place approximately 286,000 generations before present (Gen. BP) [95% HPDI = 178,249-454,561 Gen. BP], and a minor split between the two Western Eurasian lineages A and B happening much more recently but without the same clear geographical signal. The divergence time between *D. magna* and *D. similis* is much deeper, having taken place approximately 3,343,529 Gen. BP [95% HPDI = 2,252,655-4,908,550 Gen. BP]. The bars represent the 95% high probability density interval (HPDI) of median node height.

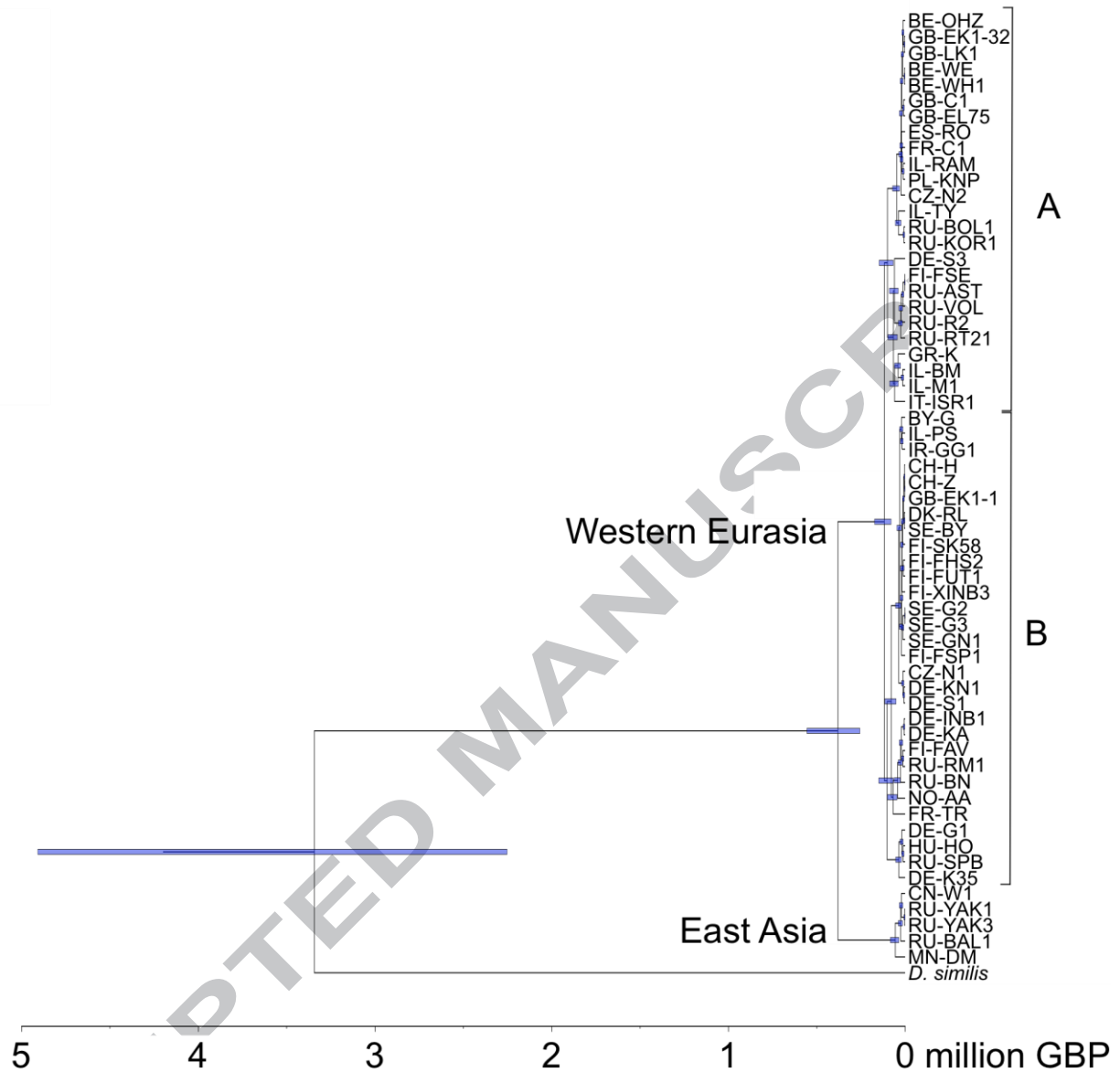


Figure 5. Geophylogeny of the mitochondrial genome of *D. magna.* Portions of the phylogeny representing Eurasian A (Red), Eurasian B (Blue), and East Asia (Cyan) are highlighted with distinct colors. Lines extend from individual phylogenetic tips to the geographical locality where each sample is derived. As with the phylogeny, lines and sampling localities are colored by which lineage the sample was inferred to belong to. A clear geographical signal is clear in the geophylogeny which distinguishes Western Eurasia and Eastern Asia samples. Within the Western Eurasia sample it can be discerned that samples associated with the Western Eurasia A and Western Eurasia B lineages have a great deal of geographical overlap, suggesting weaker geographical structuring of genetic diversity for these lineages.

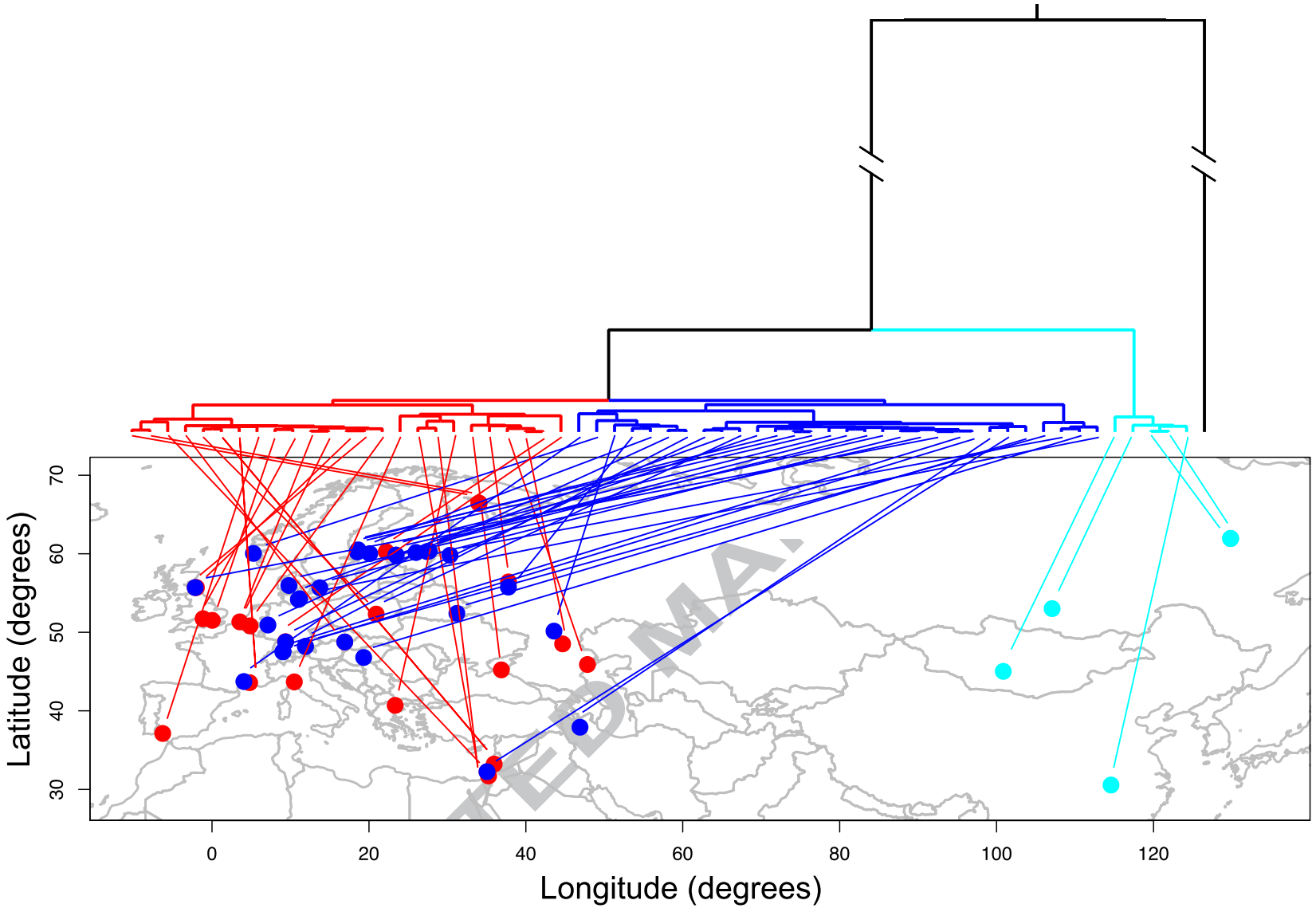


Figure 6. Phylogeny of sampled *D. magna* genotypes with region and habitat type included as traits. The observed branch colors show a relatively weak signal of geographical clustering within the species Eurasian range and a strong geographical signal delineating these groups from Eastern Asian clones. Rock-pool populations generally cluster into a major grouping in the Europe B group, though two rock-pool populations (RU-BOL1 and RU-KOR1) are present in Europe A. The bars represent the 95% high probability density interval (HPDI) of median node height. Coloring of node circles indicate the inferred habitat state.

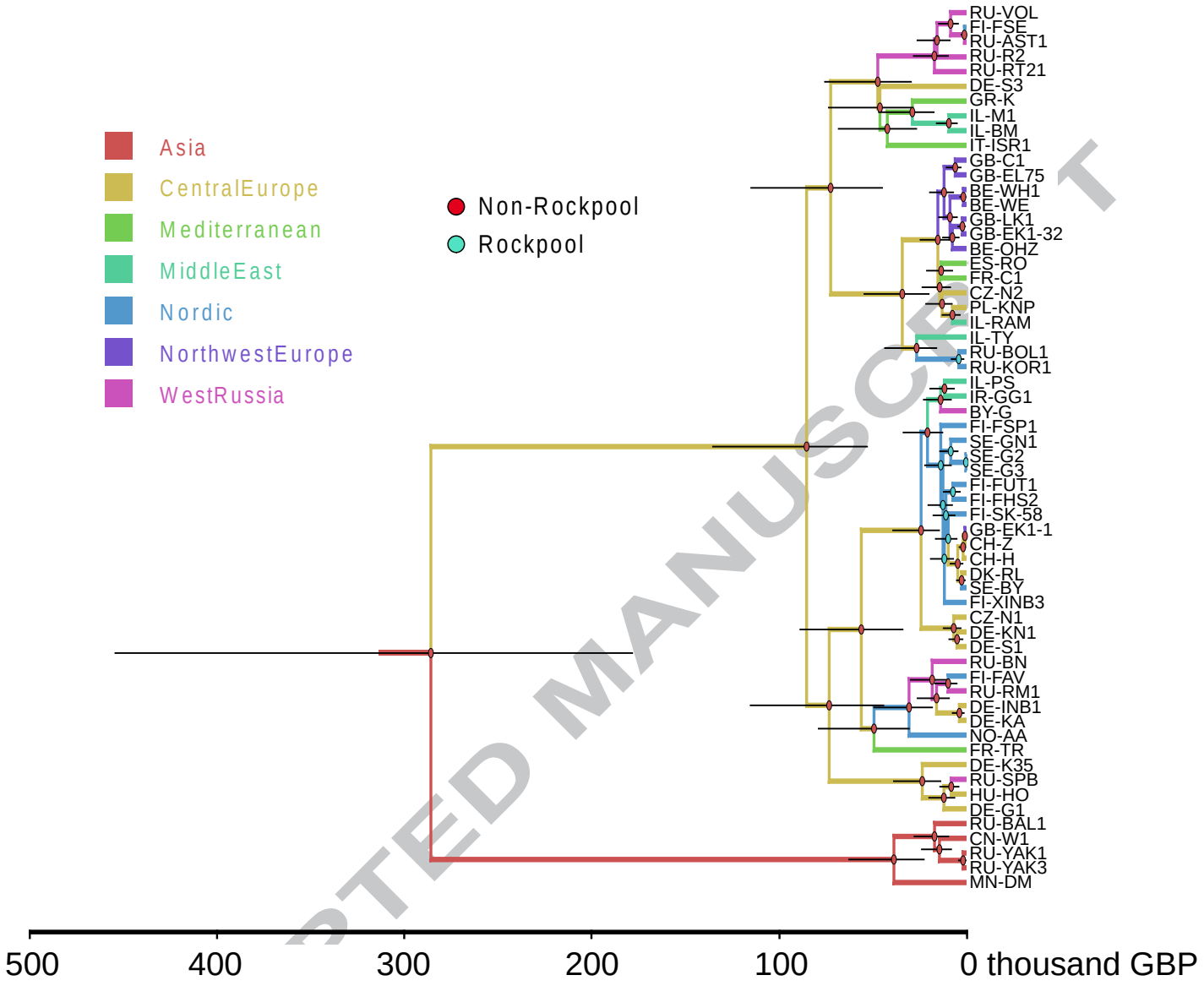
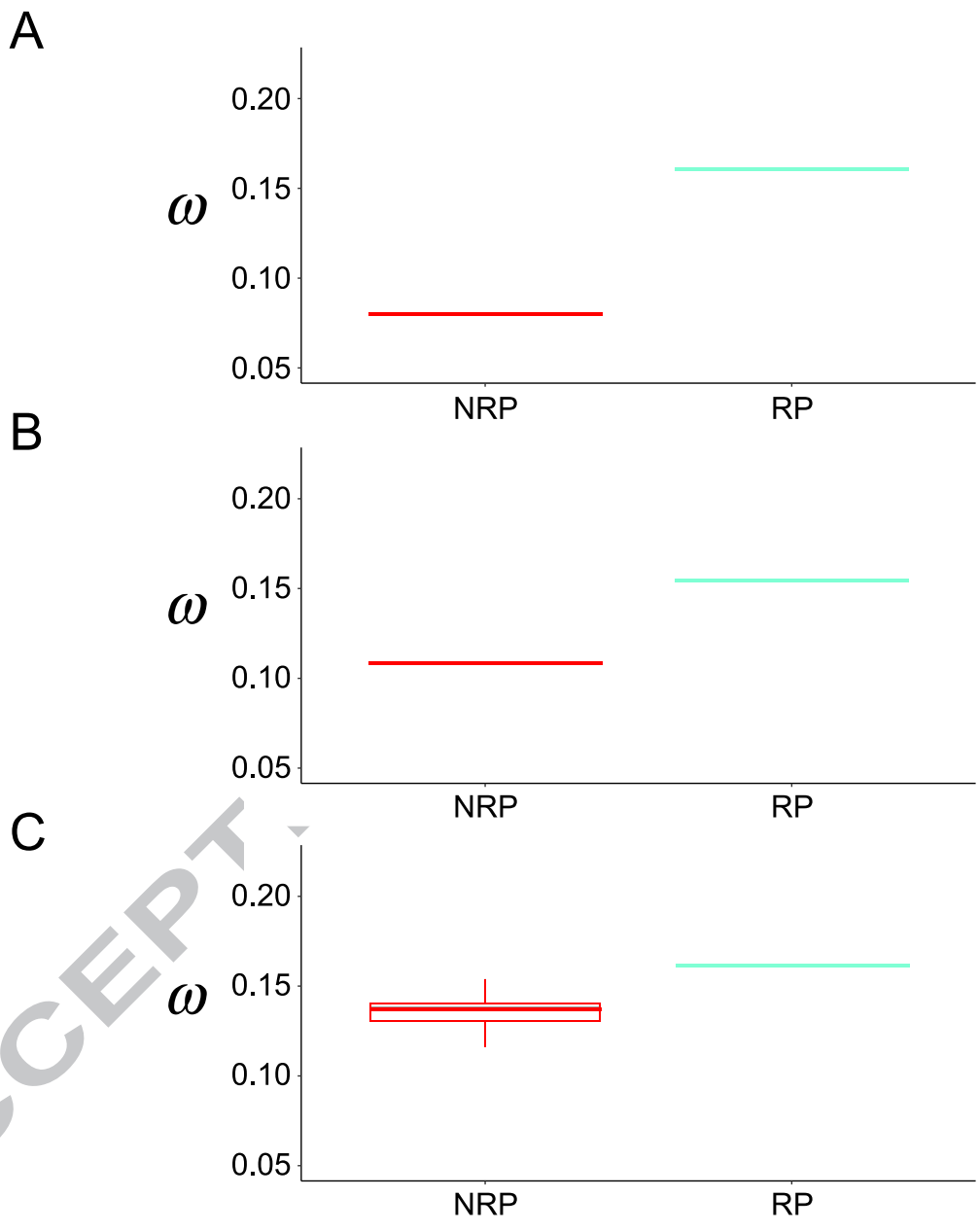
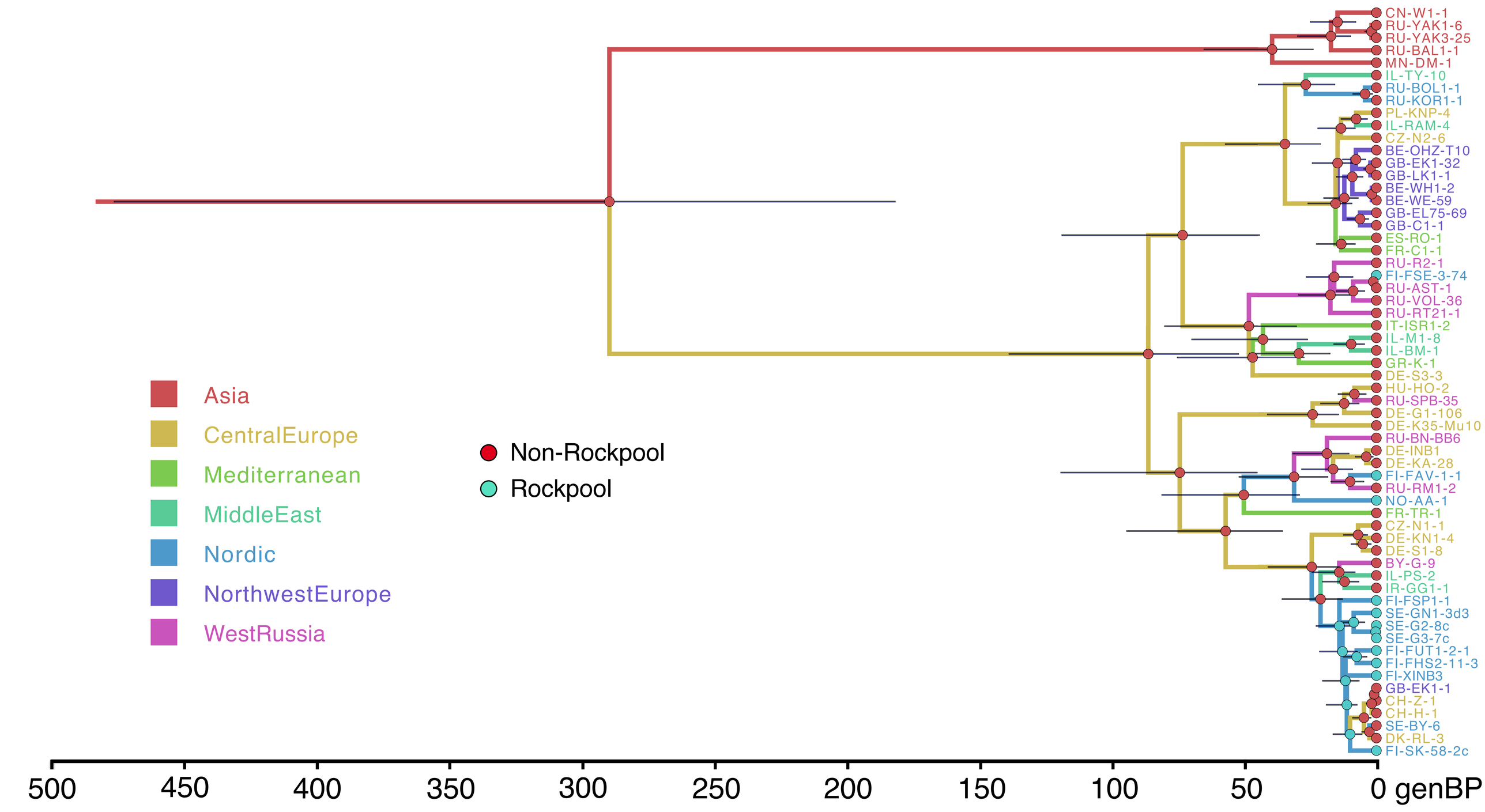


Figure 7. Boxplot summarizing the results of a PAML analysis of ω distinguishing two habitat types, genotypes derived from rock-pool (RP) or not rock-pool (NRP) habitats. A) When ω was estimated for both internal and external branches identified as one or the other habitat we found mean ω values of 0.1603 [95% CI 0.1601, 0.1605] and 0.0798 [95% CI 0.0798, 0.0798], respectively, with the RP value significantly greater than NRP based upon a permutation test ($p < 0.001$). B) When only external branches were used to estimate ω we found mean values of 0.1547 (95% CI 0.1546, 0.1547) and 0.1084 (95% CI 0.1084, 0.1084), respectively, with the RP value significantly greater than NRP based upon a permutation test ($p < 0.001$). C) When the NRP genotypes were subset in a manner to have external branches comparable to RP populations (equal or less than 20,000 generations), we estimated a mean ω of 0.159 [95% CI 0.1585, 0.1599] and 0.13 [95% CI 0.1365, 0.1376] for RP and NRP genotypes, respectively, with the RP value significantly greater than NRP based upon a permutation test ($p < 0.001$). A general pattern was observed of a much smaller variance in the estimate of ω for RP compared to NRP populations.



Highlights

- The average nucleotide identity within *D. magna* ranged between 98.9% and 95.9%. π showed a decreasing cline in moving from the Middle East to Northern Europe.
- BSA suggested a rapid demographic expansion 10-25 thousand years before present.
- The posterior density of divergence time between Western and Eastern populations had a median of ~290 thousand generations before present.
- The value of ω ranged between 0.04 and 0.17, whereas genomes sampled from non-rock pool populations was 0.0798 and rock pools was 0.1603.


CN-W1-1
RU-YAK1-6
RU-YAK3-25
RU-BAL1-1
MN-DM-1
IL-TY-10
RU-BOL1-1
RU-KOR1-1
PL-KNP-4
IL-RAM-4
CZ-N2-6
BE-OHZ-T10
GB-EK1-32
GB-LK1-1
BE-WH1-2
BE-WE-59
GB-EL75-69
GB-C1-1
ES-RO-1
FR-C1-1
RU-R2-1
FI-FSE-3-74
RU-AST-1
RU-VOL-36
RU-RT21-1
IT-ISR1-2
IL-M1-8
IL-BM-1
GR-K-1
DE-S3-3
HU-HO-2
RU-SPB-35
DE-G1-106
DE-K35-Mu10
RU-BN-BB6
DE-INB1
DE-KA-28
FI-FAV-1-1
RU-RM1-2
NO-AA-1
FR-TR-1
CZ-N1-1
DE-KN1-4
DE-S1-8
BY-G-9
IL-PS-2
IR-GG1-1
FI-FSP1-1
SE-GN1-3d3
SE-G2-8c
SE-G3-7c
FI-FUT1-2-1
FI-FHS2-11-3
FI-XINB3
GB-EK1-1
CH-Z-1
CH-H-1
SE-BY-6
DK-RL-3
FI-SK-58-2c
Asia
CentralEurope
Mediterranean
MiddleEast
Nordic
NorthwestEurope
WestRussia
Non-Rockpool
Rockpool
500
450
400
350
300
250
200
150
100
50
0 genBP