Check for updates

Molecular Ecology

WILEY

MOLECULAR ECOLOGY

ORIGINAL ARTICLE

How Did a Size-Divergent Water Strider Species Group Exhibiting an East Asia–Eastern North America Disjunction Originate and Diversify?

Yuqi Zai[1,2] | Matthew R. Pintar[3] | Mu Qiao[2] | Chen Liu[2] | Siying Fu[4] | Jakob Damgaard[5] | Jiayue Zhou[2] | Wenjun Bu[2] | Juanjuan Yuan[1] | Zhen Ye[2]

[1]College of Life Sciences, Zaozhuang University, Zaozhuang, Shandong, China | [2]Institute of Entomology, College of Life Sciences, Nankai University, Tianjin, China | [3]Institute of Environment, Florida International University, Miami, Florida, USA | [4]Research Institute of Applied Biology, Shanxi University, Taiyuan, Shanxi, China | [5]Natural History Museum of Denmark, Zoological Museum, Copenhagen Ø, Denmark

Correspondence: Juanjuan Yuan (qsdyuanjuanjuan@126.com) | Zhen Ye (yezhen1987331@nankai.edu.cn)

Received: 21 May 2026 | **Revised:** 13 August 2026 | **Accepted:** 18 August 2026

Keywords: aquatic insects | East Asia-eastern North America disjunction | intercontinental dispersal | phylogenomics | phylogeography

ABSTRACT

The classic East Asia–Eastern North America (EA-ENA) disjunction, well-documented in temperate plants, is typically attributed to the fragmentation of once-continuous high-latitude ranges by climatic cooling and regional extinction. Whether this model extends to insects, particularly aquatic groups, remains largely untested with genomic data. Here, we integrate whole-genome sequencing, ddRAD-seq and morphology with phylogenetic, phylogeographic, genotype–environment association (GEA) and niche modelling analyses to reconstruct the origin and diversification of the *Aquarius conformis* species group (Hemiptera: Gerridae). Phylogenomic and demographic analyses reject a simple widespread-ancestor vicariance scenario, instead supporting an East Asian origin followed by intercontinental dispersal to North America and subsequent divergence at approximately 460 ka, with no detectable post-divergence gene flow. The two continental lineages exhibit deep genetic divergence, marked body size differentiation and contrasting demographic histories, with precipitation seasonality (BIO15) emerging as the primary climatic axis separating their niches. GEA identifies candidate SNPs along this axis that map to conserved growth-regulating pathways (Hippo, mTOR, FoxO, Toll and Imd), suggesting a genomic basis for environment-associated divergence. Within East Asia, the widespread species *A. elongatus* comprises four phylogeographic lineages that diverged sequentially during the Pleistocene, accompanied by climate-associated allele-frequency variation and successful colonisation of highly heterogeneous environments. Collectively, our results reveal an alternative route to the EA-ENA disjunction, specifically source-area dispersal rather than relictual fragmentation, and demonstrate that post-isolation evolution in this system involves hydroclimatic differentiation, candidate pathways linked to growth and stress response, and lineage-specific demographic trajectories. This study provides a rare insect-based test of the EA-ENA disjunction and offers an integrative framework for linking biogeographic origin to post-isolation trait evolution in aquatic organisms.

1 | Introduction

Intercontinental disjunctions between East Asia and eastern North America (EA-ENA) represent a recurrent biogeographic pattern in the Northern Hemisphere. The prevailing model, developed largely from temperate plant lineages, posits that formerly widespread boreotropical or temperate taxa were fragmented by large-scale climatic deterioration during the Paleogene and Neogene, leading to extinction of intermediate populations and the survival of relict lineages on the

© 2026 John Wiley & Sons Ltd.

All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies.

two continents (Donoghue and Smith 2004; Tiffney 1985; Wen 1999). While many such disjunct plant lineages exhibit climatic niche conservatism, comparative phylogenetic and niche analyses increasingly reveal substantial heterogeneity in both assembly routes and post-isolation trajectories (Lin et al. 2023; Melton et al. 2022; Yin et al. 2021). Divergence times and reconstructed historical routes vary considerably among EA-ENA disjunct plant clades, with some lineages consistent with relictual fragmentation of formerly widespread Northern Hemisphere ancestors and others involving source-area origins followed by intercontinental dispersal through high-latitude routes (Lin et al. 2023; Rose et al. 2023; Wen et al. 2010). Post-disjunction evolution is likewise heterogeneous: many disjunct relatives retain substantial climatic niche conservatism (Melton et al. 2022; Yin et al. 2021), whereas others show niche divergence and environment-associated genetic differentiation, as illustrated by *Liriodendron* (Shen et al. 2022). These findings indicate that EA-ENA disjunctions are composite patterns whose formation routes and subsequent evolutionary trajectories are shaped by both shared climatic history and lineage-specific traits.

Documented cases of strict EA-ENA disjunction in insects remain rare (Ren et al. 2013, 2019; von Dohlen et al. 2002). In aquatic organisms, studies show that hydrological regimes potentially regulate habitat stability, spatial connectivity and the frequency of local extinction and recolonisation (Rolls et al. 2018; Sánchez-Montoya et al. 2023). In lentic habitats, enhanced dispersal ability and larger body size may facilitate repeated colonisation across ephemeral environments (Arribas et al. 2012; Ribera and Vogler 2000). By contrast, persistent flowing waters are more likely to favour local persistence, habitat specialisation and genetic structure shaped by dispersal limitation or freshwater-network configuration (Dijkstra et al. 2014; Phillipsen et al. 2015). These habitat characteristics provide an ecological context for assessing genomic, morphological and demographic differentiation following intercontinental isolation.

Species groups are clusters of closely related species that share a recent common ancestor and often retain similar morphological, genetic, or ecological characteristics (Mayr 1942). They provide an informative framework for investigating such processes. Because their members retain a shared evolutionary background while differing in geography, ecology, or phenotype, species groups allow the effects of geographic isolation, environmental differentiation and gene flow after divergence to be evaluated within a common comparative context (Coyne and Orr 2004; Nosil et al. 2005; Schluter 2009). This framework links population-level differentiation to species-level diversification (Harvey et al. 2019).

The *Aquarius conformis* species group provides such a system. This group comprises three species: *A. conformis* (Uhler, 1878) and *A. nebularis* (Drake & Hottes, 1925), both restricted to eastern North America and predominantly occupying lotic habitats (Andersen 1990; Damgaard and Cognato 2006), and *A. elongatus* (Uhler, 1896), which is endemic to East Asia and primarily inhabits lentic environments. These species also exhibit significant body size differentiation, with *A. elongatus* being substantially larger than its North American congeners. In addition, the two North American species have relatively restricted geographic distributions, whereas *A. elongatus* is widely distributed across East Asia (Figure 1). Together, these features make this species group an ideal case study for examining how the EA-ENA disjunction originated and how diversification proceeded following continental isolation. Accordingly, we tested hypotheses at two spatial scales. At the intercontinental scale, we contrasted fragmentation of a formerly widespread ancestor with source-area dispersal from either East Asia or eastern North America followed by intercontinental isolation. Within *A. elongatus*, we hypothesised that repeated regional colonisation and persistence across geographically and environmentally heterogeneous regions contributed to its broad East Asian distribution, while Pleistocene isolation, post-divergence gene flow and climate-associated genetic differentiation jointly shaped its present phylogeographic structure.

In this study, we integrate ddRAD-seq and whole-genome sequencing data with ecological and distributional information and combine phylogenetic, phylogeographic and morphological analyses to test these hypotheses. We conduct genotype–environment association analyses to identify candidate loci potentially associated with body-size regulation and bioclimatic differentiation. Focusing on *A. elongatus*, we further examine population structure, demographic history and local adaptation to explain its broad distribution across East Asia. Collectively, this study aims to place a semiaquatic insect lineage within the broader EA-ENA biogeographic framework and to evaluate how intercontinental isolation, regional diversification, phenotypic differentiation and environmental genomic signals are associated with the diversification of this species group.

2 | Materials and Methods

2.1 | Sample Collection and DNA Extraction

We collected 979 specimens of the *Aquarius conformis* species group from 104 sampling sites within its natural distribution range in East Asia and North America, including 906 specimens of *A. elongatus*, 20 specimens of *A. nebularis* and 53 specimens of *A. conformis*. The latitude and longitude of each collection site were recorded using a handheld GPS unit. All samples collected were preserved in 100% ethanol and stored at −20°C in freezers at the Institute of Entomology, College of Life Sciences, Nankai University, China (NKU). Genomic DNA was extracted from the thoracic muscle using a Universal Genomic DNA Kit (CWBIO, China). Detailed information for all specimens is provided in Tables S1 and S2.

2.2 | Phylogenetic and Genotype-Environment Association Analyses of the *A. conformis* Species Group

2.2.1 | High-Quality Reference Genome for Congeneric Species *A. elongatus*

To facilitate read mapping and variant calling across the *A. conformis* species group, we generated a chromosome-level reference genome for the East Asian congeneric species *A. elongatus*. High-molecular-weight genomic DNA was

13652294x, 2026, 16, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70533 by Zhen Ye - Nankai University , Wiley Online Library on [25/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

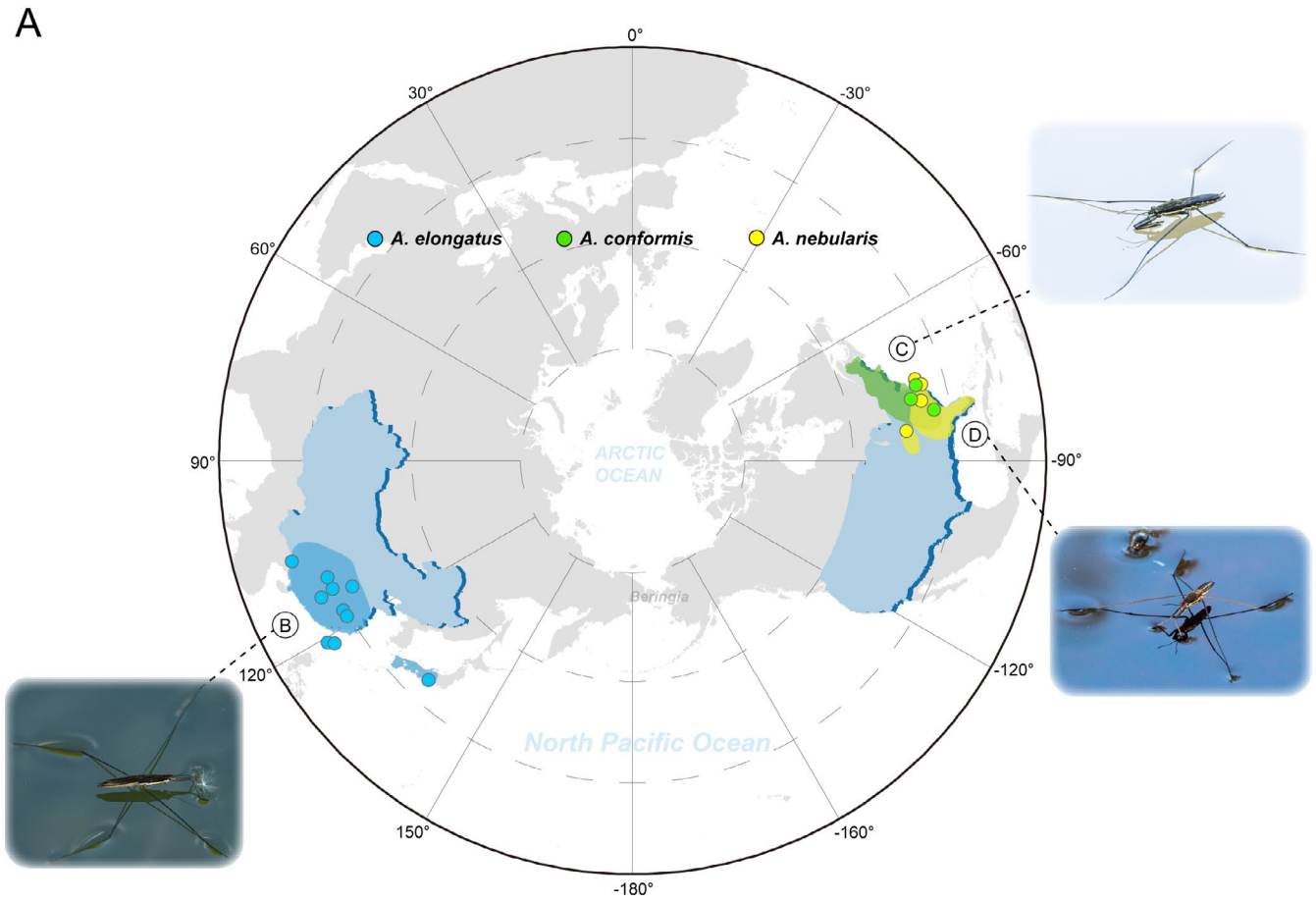

FIGURE 1 | Geographic distributions, sampling localities and representative habitat photographs of the *Aquarius conformis* species group. (A) Geographic ranges and sampling localities of *A. elongatus* in East Asia and *A. conformis* and *A. nebularis* in eastern North America. Shaded areas indicate species ranges, points indicate sampling localities used in this study, and colours correspond to species as shown in the legend. Circled letters and dashed lines link the map to the corresponding inset photographs: (B) *A. elongatus*, (C) *A. conformis* and (D) *A. nebularis* in their natural habitats. Photographs are not shown to the same scale. Photograph sources: *A. conformis* (https://www.inaturalist.org/photos/525938735); *A. nebularis* (https://www.inaturalist.org/photos/223521932).

extracted from a single adult specimen collected in Hubei, China. Long-read sequencing was performed on an Oxford Nanopore Technologies PromethION platform, supplemented with Illumina paired-end short reads (150 bp) for error correction and Hi-C libraries from the same individual for chromosome scaffolding.

De novo assembly was conducted using NextDenovo v2.3.1, followed by iterative polishing with Racon (ONT reads) and NextPolish (Illumina reads). Redundant contigs were removed based on self-alignment similarity. Chromosome-level scaffolding was performed using LACHESIS with Hi-C data. Detailed protocols, including library preparation, quality control and assembly parameters, are provided in Methods S1.

2.2.2 | Whole-Genome Sequencing and Variant Calling

We performed whole-genome sequencing on 16 individuals of three species selected from the *A. conformis* species group (Table S2). The clean reads were mapped to the reference genome of *A. elongatus* using BWA-MEM v.0.7.18 (Li and Durbin 2009). BAM files were sorted, and index files were created using SAMtools v.1.7 (Danecek et al. 2021). The average mapping rate of clean reads to the reference genome was 92.60% ± 0.51% (*A. elongatus*) and 56.10% ± 1.83% (*A. nebularis* and *A. conformis*). SNP calling was performed with GATK v4.1.9.0 (DePristo et al. 2011) after removing PCR duplicates using the 'rmdup' command in SAMtools v.1.7. To obtain reliable candidate SNPs, hard filters were applied to the clean SNPs using VCFtools v0.1.16 (Danecek et al. 2011) according to the following criteria: (1) minQ > 30, (2) min-DP > 10 and max-DP < 50, (3) max-missing counts = 1, (4) MAF > 0.05, (5) retain only biallelic SNPs.

2.2.3 | Genetic Structure and Phylogenetic Analyses

To infer evolutionary relationships and potential reticulation from nuclear markers, a Neighbour-Net network (Huson and Bryant 2006) was constructed in SplitsTree v4.14.5 using pairwise genetic distances calculated from the filtered resequencing-derived SNP dataset. Population genetic structure was further assessed by principal component analysis (PCA) on the same SNP dataset using PLINK v1.9 (Purcell et al. 2007). The variance-standardised relationship matrix

1365294x, 2026, 16, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70533 by Zhen Ye - Nankai University , Wiley Online Library on [25/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

was computed, and the top principal components were extracted; individuals were plotted along the first two principal components.

2.2.4 | Demographic Model Testing

Based on nuclear SNP data, the 16 resequenced individuals were grouped into three geographic lineages corresponding to the three species: East Asian *A. elongatus* and North American *A. nebularis* + *A. conformis*. To infer the intercontinental demographic history of the *A. conformis* species group, we tested eight isolation-with-migration (IM) models (Figure S1): recent bidirectional gene flow between adjacent lineages (M1–M3), between any two of the three lineages (M4–M6), across all three lineages (M7), or strict isolation with no post-divergence migration (M8). We used fastsimcoal2 v2.6 (Excoffier et al. 2013, 2021) to perform coalescent simulations and composite-likelihood parameter estimation. A nuclear mutation rate of 3.5×10^{-9} substitutions per site per generation was applied, based on estimates from *Drosophila melanogaster* (Keightley et al. 2009), with a generation time of 0.5 years (Andersen 1982). The joint site frequency spectrum (SFS) was generated using easySFS, projecting down to optimal sample sizes to minimise linkage effects while retaining information. For each model, 50 independent runs were conducted, each with 50,000 coalescent simulations and 40 ECM optimisation cycles; monomorphic sites were excluded to avoid overfitting. The best run per model was selected as the one with the smallest difference between maximum observed and maximum estimated likelihood (MaxObsLhood—MaxEstLhood). The overall best-fitting model was chosen based on the highest likelihood and lowest AIC. Confidence intervals (95% CIs) for parameter estimates were obtained by parametric bootstrapping: 100 SFS replicates were simulated from the best-fit parameter file, each subjected to 50 replicate runs as described above, and means with 95% CIs were calculated from the resulting 100 best-fit bootstraps.

2.2.5 | Biogeographic Scenario Testing

We performed approximate Bayesian computation with random forests (ABC-RF) in DIYABC-RF v1.0 (Collin et al. 2021; Pudlo et al. 2016) to compare three competing biogeographic scenarios for the origin of the EA-ENA disjunction. Individuals were grouped into East Asian (*A. elongatus*; EA, $n = 9$) and eastern North American (*A. conformis* + *A. nebularis*; ENA, $n = 7$) lineages. After LD filtering and minor allele frequency thresholding (0.05), 5000 SNP loci were subsampled from the 33,473 polymorphic sites for simulation to reduce computational burden. Three scenarios with equal prior probabilities were tested: (1) East Asian origin followed by dispersal to North America; (2) North American origin followed by dispersal to East Asia; and (3) widespread ancestral population with subsequent vicariance. Uniform priors were assigned to effective population sizes and divergence times. Sixteen summary statistics were calculated per simulated dataset, including heterozygosity, F_{st} and genetic distance measures. A reference table of 12,000 simulations (4000 per scenario) was generated. Model choice was conducted using random forest classification with 500 trees and two LDA axes. Model performance was assessed through prior error rates and LDA projection of the observed data. The posterior probability of the best-supported scenario was estimated via random forest regression.

2.2.6 | Historical Demographic Changes

To infer long-term changes in effective population size (N_e) of three species within the *A. conformis* species group, we applied the pairwise sequentially Markovian coalescent (PSMC) model (Li and Durbin 2011). Individual genomes were assembled de novo using MEGAHIT v1.2.8 (Li et al. 2015) with default parameters. Contigs were concatenated into pseudochromosomes using FaTools (He et al. 2013). Clean reads were mapped to these pseudochromosomes with BWA-MEM (Li and Durbin 2009), and alignments were processed into sorted BAM files using SAMtools v1.12 (Danecek et al. 2021). Variant calling was performed with bcftools mpileup (-C 50) and bcftools call, filtering sites with depth $< 10\times$ or $> 100\times$. Consensus sequences were generated with vcfutils.pl. vcf2fq and converted to PSMC input with fq2psmcfa. PSMC inferences used parameters -N 25 -t 15 -r 5 -p '4 + 25 × 2 + 4 + 6'. Trajectories were scaled using a nuclear mutation rate of 3.5E-9 substitutions per site per generation (consistent with intercontinental demographic modelling) and a generation time of 0.5 years (Andersen 1982). Robustness was assessed with 100 bootstrap replicates per individual.

2.2.7 | Morphological Cluster

We conducted morphological measurements on 163 individuals from the *A. conformis* species group collected across East Asia and North America. Measurements were performed using a stereomicroscope and recorded in millimetres. Fourteen diagnostic characteristics were quantified: body length, body width, head length, head width, pronotum length, fore femur length, fore tibia length, middle femur length, middle tibia length, hind femur length, hind tibia length, abdomen length, seventh abdominal sternum length and connexival spine length. PCA was carried out in SPSS (IBM Corp. 2010) based on these 14 morphological variables to visualise morphological relationships between individuals from East Asia and North America. Additionally, boxplots were generated using the ggplot2 package (Wickham 2009) in R to compare each of the 14 morphological variables between the two regions.

2.2.8 | Identification of Diagnostic Climate Axes

We obtained occurrence records for three *Aquarius* species (*A. elongatus, A. nebularis and A. conformis*) from global biodiversity databases (GBIF and iNaturalist) using R packages *spocc* (Chamberlain and Boettiger 2017). Following quality control to remove erroneous and imprecise records, we extracted 19 bioclimatic variables from WorldClim's near-current (1970–2000) climate layers at 2.5-arcminute resolution (Fick and Hijmans 2017). To identify the primary climatic axis differentiating *A. elongatus* from the two North American congeners (*A. nebularis* and *A. conformis*), we implemented a linear Support

1365294x, 2026, 16, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70533 by Zhen Ye - Nankai University , Wiley Online Library on [25/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Vector Machine (SVM) algorithm via the e1071 package (Cortes and Vapnik 1995; Meyer et al. 2024). Climate variables with pairwise correlation coefficients > 0.7 were excluded to mitigate multicollinearity effects, and retained variables were normalised to a 0–1 scale. As the objective focused on climate-space differentiation rather than predictive modelling, all quality-controlled records were utilised for training, with specimens categorised as binary classes (*A. elongatus* vs. non-*elongatus*). The SVM weight vector was transformed into feature importance metrics by calculating absolute weights and their proportional contributions. This yielded a composite differentiation index defined as: $\text{Weighted_Bio} = 0.338 \times \text{BIO15} + 0.285 \times \text{BIO2} + 0.126 \times \text{BIO3} + 0.121 \times \text{BIO14} + 0.072 \times \text{BIO8} + 0.056 \times \text{BIO9} + 0.002 \times \text{BIO5}$. To further examine hydroclimatic differences between East Asian and eastern North American records, Euclidean (straight-line) distances from occurrence localities to the nearest river were calculated. Distributions of these distances and of BIO15 were compared between the two continental groups, and the relationship between BIO15 and river distance was visualised with fitted linear trends.

2.2.9 | Genotype-Environment Association Analysis

To identify adaptive loci associated with bioclimatic differentiation, we performed genotype-environment association (GEA) analyses on the SNP_dataset_2 dataset using latent factor mixed models (LFMM) (Frichot et al. 2013) implemented in the R package LEA v3.3.2 (Frichot and François 2015). Specifically, we tested allelic associations with the Weighted_Bio index using two latent factors to account for population structure. Five independent Markov Chain Monte Carlo (MCMC) runs were executed, each with 5000 burn-in iterations followed by 10,000 sampling iterations. Across the five runs, the median *z*-score was calculated, recalibrated using lambda and the resulting *p* values were adjusted for multiple testing via false discovery rate (FDR) correction at a 5% significance threshold. Significantly associated loci (FDR < 0.05) ranking within the top 0.1% of association strength were retained. Genes overlapping these loci were functionally annotated and subjected to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis using KOBAS v3.0 (Bu et al. 2021).

2.3 | Phylogeographic Analyses of *A. elongatus*

2.3.1 | ddRAD-Seq Library Preparation, Sequencing and Variant Calling

Double-digest RAD-seq (ddRAD-seq) libraries were prepared for 134 individuals from 17 East Asian populations of *A. elongatus* (Table S1), following the protocol of Peterson et al. (2012) with EcoRI and MspI restriction enzymes. Libraries were size-selected (250–600 bp) and sequenced paired-end on an Illumina NovaSeq 6000 platform. Clean reads were processed de novo using ipyrad v0.9.45 (Eaton and Overcast 2020) with a clustering threshold of 0.85. The final dataset required loci present in ≥ 95% of individuals; one random SNP per locus was retained to reduce linkage disequilibrium, producing an unlinked ddRAD dataset for downstream analyses. Detailed protocols are provided in Methods S1.

2.3.2 | Population Genetic Structure

Population genetic structure was inferred primarily using Bayesian clustering in STRUCTURE v2.3.4 (Pritchard et al. 2000). For $K = 1$–5, we performed 10 independent runs with a burn-in of 20,000 steps followed by 1,200,000 MCMC iterations, assuming correlated allele frequencies and admixture. The optimal K was determined using the ΔK method in STRUCTURE HARVESTER (Earl and vonHoldt 2012). Results from replicate runs were combined using CLUMPP v1.1.2 (Jakobsson and Rosenberg 2007) and visualised with DISTRUCT v1.1 (Rosenberg 2004). Additional analyses of genetic diversity (observed and expected heterozygosity, nucleotide diversity), inbreeding coefficients and reticulate evolutionary relationships (Neighbour-Net network) were conducted as described in Methods S1. Discriminant analysis of principal components (DAPC) was also performed to complement the STRUCTURE results (Methods S1; Jombart 2008).

2.3.3 | Historical Demographic Changes

We inferred demographic history of *A. elongatus* using the stairway plot method (Liu and Fu 2015) based on a one-dimensional unfolded SFS. We generated lineage-specific SFS for each genetic group (EC, YN, TW, JPA) using easySFS. A nuclear mutation rate of 3.5E-9 per site per generation and a generation time of 0.5 years were applied, consistent with those used in demographic model testing. Temporal changes in effective population size (N_e) were then estimated for each lineage via stairway plot analyses.

2.3.4 | Demographic Model Testing

Demographic histories among the four East Asian lineages of *A. elongatus* (EC, YN, TW, JPA) were inferred using coalescent-based model testing in fastsimcoal2 v2.6 (Excoffier et al. 2021), following the same protocol and parameter settings described above for the intercontinental divergence in the *A. conformis* species group. An initial set of 16 models was evaluated (Figure S2): 15 dichotomous models (M1–M15) representing alternative sequential divergence topologies, and one trifurcating model (M16) assuming simultaneous splits. The best-supported model was further refined by adding bidirectional recent gene flow between neighbouring lineages (M17–M19), among two of three lineages (M20–M22), or across all four lineages (M23). Model selection was based on AIC scores, with 100 independent simulations per model.

2.3.5 | Partial Redundancy Analysis

To evaluate climate-associated allele-frequency variation while controlling for neutral population structure, we conducted partial redundancy analysis (pRDA) following standard genotype-environment association protocols (Capblancq et al. 2018; Forester et al. 2018). Analyses were performed in R v4.2.2 using the vegan package v2.6–1. The response matrix comprised population-level allele frequencies at biallelic SNPs from the filtered ddRAD VCF file. Allele counts were

1365294x, 2026, 16, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70533 by Zhen Ye - Nankai University , Wiley Online Library on [25/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

exported in CATG format using bcftools, and population frequencies were calculated per SNP with VCFtools. Results were merged and transposed into a 'population×SNP' matrix. Environmental predictors included the 19 bioclimatic variables from WorldClim v2.1 (Fick and Hijmans 2017) at 2.5 arc-minute resolution, extracted for each sampling locality using the raster package. Highly collinear variables (pairwise Pearson $r > 0.7$) were iteratively removed, and retained predictors were z-score standardised using scale() in R. Neutral population structure was controlled by conditioning on the admixture Q-matrix from STRUCTURE at $K = 4$. The pRDA model was: allele frequency matrix ~ climatic predictors + Condition(Q-matrix).

The analysis used the rda() function in vegan. Population scores on the first two constrained axes (RDA1 and RDA2) and biplot loadings of climatic variables were extracted for visualisation and interpretation of climate-associated genetic patterns after accounting for neutral structure.

3 | Results

3.1 | Phylogenetic and Genotype-Environment Association Analyses of the *A. conformis* Species Group

3.1.1 | High-Quality Reference Genome of *A. elongatus*

We assembled a high-quality chromosome-level reference genome for *A. elongatus* using Oxford Nanopore long reads (90× coverage), Illumina short reads (49× coverage) for polishing and Hi-C data (119× coverage) for scaffolding. The final assembly spanned 1.06 Gb with a GC content of 30.24%, a contig N50 of 13.51 Mb and a scaffold N50 of 94.65 Mb (Table S3; Figure S3A). Over 99.94% of sequences were anchored and oriented into 12 pseudochromosomes, consistent with known karyotypes in the Gerridae family (Table S4; Figure S3B). Pre-assembly k-mer analysis (17-mer) of Illumina reads estimated a genome size of 0.94 Gb with 1.60% heterozygosity, slightly smaller than the final assembly due to effective resolution of repetitive and heterozygous regions. Assembly completeness, assessed with BUSCO v4.0.5 (insecta_odb10 dataset), was 94.37% (93.64% single-copy, 0.73% duplicated, 3.22% fragmented, 2.41% missing), comparable to other high-quality insect genomes (Figure S3C). Illumina read mapping rates exceeded 98%, confirming high consensus accuracy. Repeats comprised 54.95% of the genome, and protein-coding gene prediction yielded 11,696 genes with similar BUSCO completeness in the annotated set.

3.1.2 | Genetic Structure and Phylogenetic Analyses

Nuclear SNP data revealed a clear genetic separation between the East Asian and North American lineages of the *A. conformis* species group. The Neighbour-Net network constructed from genome-wide SNPs showed a sharp split between a tightly clustered *A. elongatus* group from East Asia and a more reticulate North American clade containing both *A. conformis* and *A. nebularis* (Figure S4). Principal component analysis (PCA) of the same SNP dataset further confirmed distinct genetic clustering of the three species (Figure S5). PC1 sharply separated East Asian *A. elongatus* from the North American taxa, while PC2 distinguished *A. conformis* (positive scores) from *A. nebularis* (negative scores).

3.1.3 | Demographic Model Testing

For the whole-genome sequencing dataset, the 16 resequencing datasets were mapped to the reference genome of *A. elongatus*, yielding an average coverage depth of 10× and a breadth of coverage of 67.4%. We employed coalescent-based demographic simulations using fastsimcoal2 v2.6 (Excoffier et al. 2021) to infer the evolutionary history of the *A. conformis* species group, including *A. elongatus* in East Asia alongside *A. nebularis* and *A. conformis* in North America. Model 1 showed the highest likelihood and lowest Akaike Information Criterion (AIC), supporting a scenario with no gene flow between the East Asian and North American lineages (Table S5). Point estimates of the demographic parameters for model M1 were provided with 95% confidence intervals (CIs) (Table S6). Under this model, the East Asian lineage diverged from the North American lineage at approximately 460 ka (95% CI: 433–483 ka) (Figure 2A). Effective population sizes differed markedly among lineages, contracting to 5251 in *A. conformis* and 73,019 in *A. nebularis*, while remaining largest in East Asian *A. elongatus* (814,658). Within North America, the best model incorporated asymmetric recent gene flow following secondary contact, with substantially higher migration from *A. conformis* into *A. nebularis* (2.1E-5; 95% CI: 1.7–2.5E-5) than the reverse (4.5E-6 migrants per generation; 95% CI: 2.9–6.1E-6).

3.1.4 | Biogeographic Scenario Comparison

The DIYABC-RF model comparison favoured Scenario 1, corresponding to the East Asian-origin model, over the North American-origin and widespread-ancestor alternatives (Figure S6). Scenario 1 received the highest number of random forest votes (312 of 500), whereas Scenario 2 and Scenario 3 received 62 and 126 votes, respectively. The posterior probability of the selected East Asian-origin scenario was 0.595. Model-choice validation indicated relatively low global classification error across the simulated reference table. The prior error rate was 0.061, and class-specific prediction errors were 0.052 for Scenario 1, 0.058 for Scenario 2 and 0.074 for Scenario 3. Projection of the observed dataset together with simulated datasets showed that the observed summary statistics fell within the simulated reference space. Thus, DIYABC-RF favoured an East Asian source for the *A. conformis* species group.

3.1.5 | Historical Demographic Changes

Pairwise sequentially Markovian coalescent (PSMC) analyses revealed markedly different long-term demographic trajectories between East Asian and North American lineages (Figure 2B). The East Asian *A. elongatus* exhibited a relatively stable and gradual decline in effective population size (N_e) from 10^6 years ago, maintaining Ne above approximately 1.5×10^5 individuals through the late Pleistocene, with only

13652294x, 2026, 16, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70533 by Zhen Ye - Nankai University , Wiley Online Library on [25/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

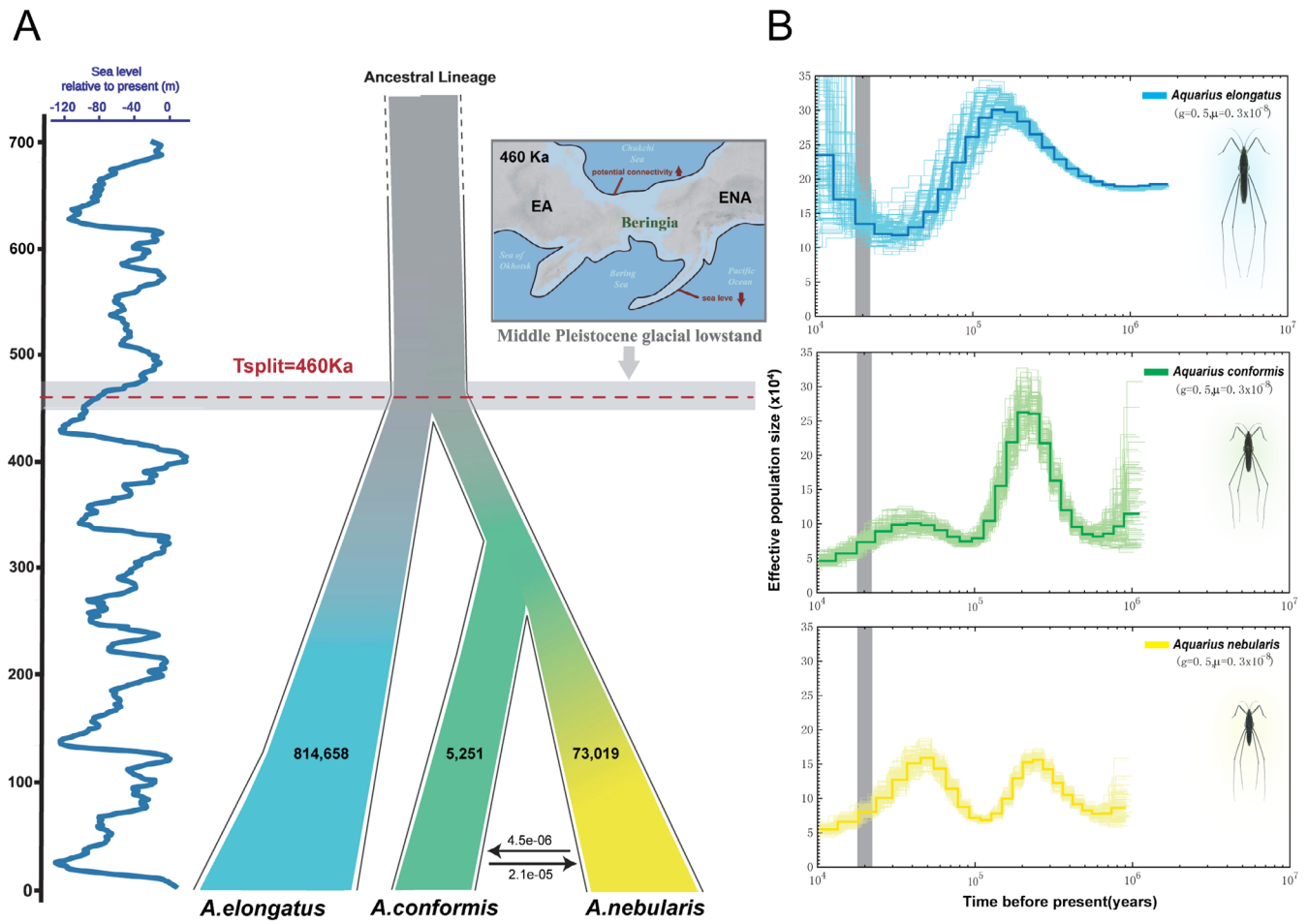

FIGURE 2 | Demographic model and historical population dynamics of the *Aquarius conformis* species group. (A) Best-fitting demographic model inferred with fastsimcoal2 from the joint site-frequency spectrum (SFS) of whole-genome resequencing SNPs. Boxes denote populations; arrows denote migration parameters; branch labels indicate model parameters, including divergence times and effective population sizes, as defined in the panel. The left inset shows the Spratt and Lisiecki (2016) global sea-level stack spanning 0–700 ka, with the model-based EA-ENA split time at approximately 0.46 Ma (460 ka) marked by a dashed vertical line. The right inset shows the chronological schematic aligned with the same split time. (B) PSMC curves for representative individuals of *A. elongatus*, *A. conformis* and *A. nebularis*. The *x*-axis shows time before present and the y-axis shows effective population size (N_e), scaled by the mutation rate and generation time specified in Methods. The grey shaded band denotes the LGM period. Lines represent individual trajectories; thin lines indicate 95% confidence intervals.

minor fluctuations and no evidence of severe bottlenecks. In contrast, both North American species showed pronounced fluctuations. *A. conformis* experienced a marked expansion peaking around 10^5–10^6 years ago, reaching an Ne of approximately 3.0×10^5 individuals, followed by a sharp contraction during the Last Glacial Maximum (LGM, 26.5–19 ka; Clark et al. 2009) and subsequent recovery. *A. nebularis* displayed similar dynamics but with more pronounced oscillations, including an earlier peak and deeper late-Pleistocene declines (Ne dropping below 10^5 in some intervals).

3.1.6 | Morphological Cluster

Representative dorsal habitus photographs of *A. nebularis*, *A. conformis* and *A. elongatus* are shown in Figure S7. PCA of the 14 morphological traits revealed a clear separation between East Asian *A. elongatus* and the two North American species along the first two principal components (Figure S8). Boxplot comparisons indicated that East Asian *A. elongatus* individuals exhibited significantly larger body-size-related traits, including body length and hind femur length, than *A. nebularis* and *A. conformis* from North America (Figure S9).

3.1.7 | Genotype-Environment Association Analysis

A. elongatus occupied a significantly broader climatic niche space than *A. nebularis* and *A. conformis* (Figure 3A), with linear SVM analysis identifying an optimal decision boundary separating *A. elongatus* from the combined *A. nebularis* and *A. conformis* species group, where the primary climatic separation vector was defined by the weighted bioclimatic index: Weighted_Bio = 33.83% BIO15 + 28.47% BIO2 + 12.65% BIO3 + 12.09% BIO14 + 7.19% BIO8 + 5.58% BIO9 + 0.18% BIO5 (Figure 3B). East Asian records also exhibited significantly greater distances to potential rivers and higher precipitation seasonality than eastern North American records (Figure 3C), and BIO15 was positively associated with distance to potential rivers within each region (Figure 3D). Subsequent LFMM

1365294x, 2026, 16, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70533 by Zhen Ye - Nankai University , Wiley Online Library on [25/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

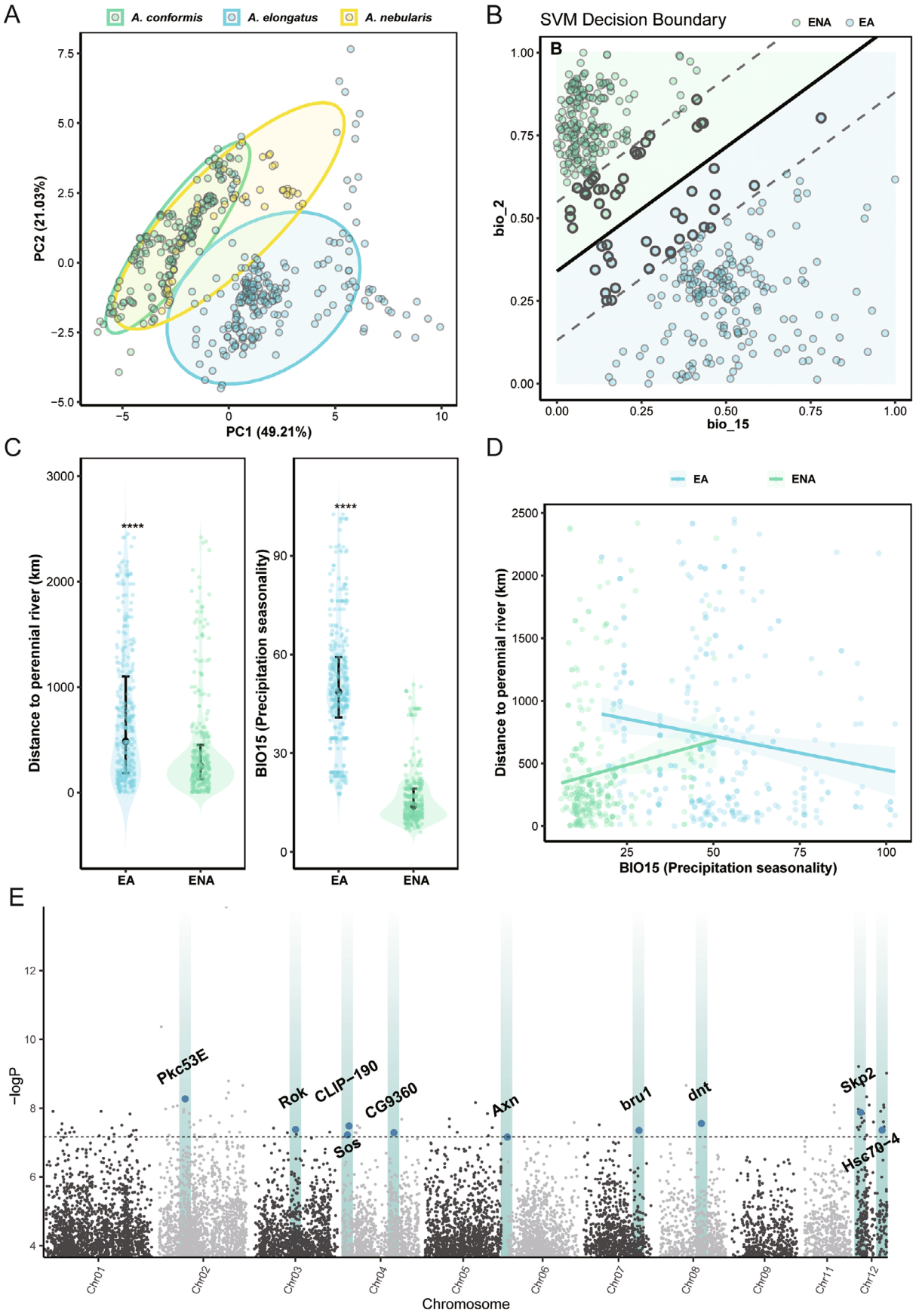

FIGURE 3 | Legend on next page.

1365294x, 2026, 16, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70533 by Zhen Ye - Nankai University , Wiley Online Library on [25/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

FIGURE 3 | Environmental differentiation and genotype-environment associations in the *Aquarius conformis* species group. (A) Principal component analysis of bioclimatic variables for occurrence records of *A. elongatus*, *A. conformis* and *A. nebularis*. Points represent occurrence records, and ellipses indicate species-level grouping. (B) Linear SVM decision boundary separating East Asian *A. elongatus* from eastern North American congeners (*A. conformis* + *A. nebularis*) using BIO15 and BIO2. (C) Distributions of distance to potential rivers and BIO15 for East Asian and eastern North American records. Asterisks indicate significance levels shown in the panel. (D) Relationship between BIO15 and distance to potential rivers for East Asian and eastern North American records. Lines indicate fitted trends. (E) Manhattan plot of LFMM associations between SNPs and the Weighted_Bio axis. The dashed line indicates the significance threshold, and labelled points indicate representative candidate genes. EA, East Asia; ENA, eastern North America; BIO15, precipitation seasonality; BIO2, mean diurnal temperature range.

analysis of the top 15,000 Weighted_Bio-associated SNPs revealed their localisation within 257 genes, among which 10 genes mapped to pathways with putative relationships to insect body size regulation including *jhamt* in insect hormone biosynthesis, *app* and *Act5C* in the Hippo signalling pathway (fly ortholog), *fz* in mTOR signalling, *Sirt1* in FoxO signalling and *Ank* in Toll and Imd signalling pathways, while additional pathways involving fatty acid and amino acid metabolism also showed potential connections to body size modulation (Figure 3E).

3.2 | Phylogeographic Analyses of *A. elongatus*

3.2.1 | Nuclear SNP Calling

For the ddRAD-seq dataset, a total of 134 *A. elongatus* individuals were subjected to sequencing, yielding 837,263,627 raw reads. The number of reads per individual ranged from 186,713 to 248,366 with an average of 6,067,127 reads per specimen. After filtering low-quality reads, 816,804,698 high-quality clean reads were retained. The final ddRAD_95 SNPs dataset comprised 3652 ddRAD-seq loci, 71,073 single nucleotide polymorphisms (SNPs) and 3.49% missing data. From this dataset, 3652 unlinked SNPs were randomly selected to construct the ddRAD_95 USNPs dataset (generated via ipyrad).

3.2.2 | Population Genetic Structure and Differentiation

Genetic clustering via STRUCTURE revealed significant population structure, with $K = 4$ identifying four genetically distinct groups: Eastern China lineage (EC), Taiwan lineage (TW), Yunnan lineage (YN) and Japan lineage (JPA) (Figure 4A). The EC lineage comprised populations sampled broadly across central, southern and eastern mainland China, whereas the YN lineage was geographically restricted to Yunnan. Neighbour-Net analysis of ddRAD-derived SNPs supported these divisions, revealing four primary lineages with similar geographic correspondence (Figure S10). DAPC further corroborated the four clusters (Figure S11). Genetic diversity metrics varied among lineages (Table S7). Nucleotide diversity (π) was highest in the eastern China lineage ($\pi = 0.0577$) and progressively lower in Taiwan ($\pi = 0.0490$), Yunnan ($\pi = 0.0417$) and Japan ($\pi = 0.0333$). Observed heterozygosity (H_O) ranged from 0.1959 to 0.3872 across all populations. Expected heterozygosity (H_E) varied between 0.2163 and 0.4067.

3.2.3 | Demographic Model Testing

Coalescent-based model testing in fastsimcoal2 identified a sequential divergence topology (M7) as the best-supported among 16 initial models for the four East Asian lineages of *A. elongatus* (lowest AIC; Table S8). Refinement with gene flow scenarios favoured M23, which included bidirectional migration among lineages. This model inferred initial divergence of the Japan (JPA) lineage, followed by separation of the Taiwan (TW) lineage, and finally the split between eastern China (EC) and Yunnan (YN) lineages (Figure 4C). Point estimates (with 95% CIs; Table S9) placed the JPA divergence at 258,230 years ago (235,338–281,122), TW at 37,858 years ago (36,645–39,072) and EC-YN at 35,101 years ago (33,793–36,408). Asymmetric bidirectional gene flow was detected, with the highest migration rate from JPA to EC (2.5E-5; 2.4E-5–2.7E-5) and the lowest from EC to TW (4.3E-9; 1.7E-9–6.9E-9).

3.2.4 | Historical Demographic Changes

Stairway plot analyses revealed distinct demographic trajectories among the four lineages of *A. elongatus*. The eastern China (EC) group maintained a relatively large and stable effective population size (N_e), whereas the Taiwan (TW) group underwent a pronounced contraction. The Japan (JPA) group expanded after the LGM and subsequently remained stable, while the Yunnan (YN) group showed moderate fluctuations around the LGM followed by relative stability (Figure 4D).

3.2.5 | Climate-Associated Allele-Frequency Variation

Partial redundancy analysis (pRDA), conditioning on neutral population structure (STRUCTURE Q-matrix at $K = 4$), revealed clear associations between allele frequency variation and climatic gradients in *Aquarius elongatus*. After removing highly collinear predictors, five bioclimatic variables were retained: annual precipitation (BIO12), precipitation seasonality (BIO15), maximum temperature of the warmest month (BIO5), minimum temperature of the coldest month (BIO6), and mean temperature of the wettest quarter (BIO8). Environmental loadings indicated that the first constrained axis (RDA1) primarily reflected a contrast gradient between precipitation- and temperature-related variables, with strong negative loadings for annual precipitation (BIO12 = −0.880) and minimum temperature of the coldest month (BIO6 = −0.787), and a strong positive loading for maximum temperature of the warmest month (BIO5 = 0.699). In contrast, the second axis (RDA2) was predominantly driven by mean temperature of the wettest quarter (BIO8 = −0.667), with a moderate contribution

1365294x, 2026, 16, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70533 by Zhen Ye - Nankai University , Wiley Online Library on [25/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

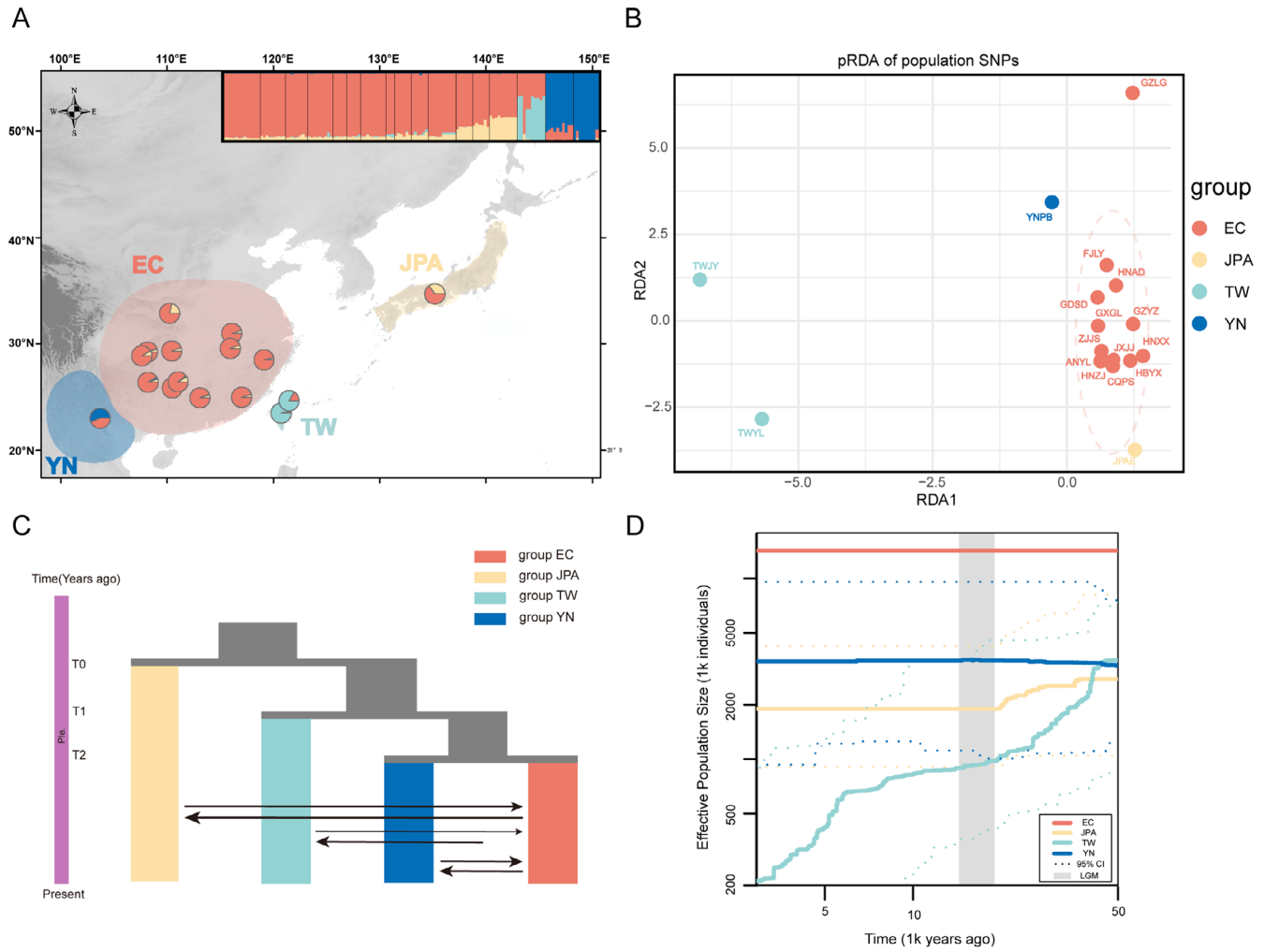

FIGURE 4 | Population structure, demographic history and climate-associated genetic variation within East Asian *A. elongatus*. (A) Sampling map of East Asian *A. elongatus* populations used for ddRAD-seq analyses, with the STRUCTURE clustering plot shown as an inset. Points indicate sampling sites; colours denote genetic lineages (EC, YN, TW, JPA) as defined in the legend. In the inset, each vertical bar represents an individual, colours represent ancestry proportions and individuals are grouped by population. (B) Partial redundancy analysis (pRDA) of allele-frequency variation against climatic variables, conditioning on neutral population structure using the STRUCTURE Q-matrix. Points represent populations and are coloured by lineage; axes correspond to constrained ordination components. (C) Demographic model inferred with fastsimcoal2 from the ddRAD joint SFS among East Asian lineages. Boxes represent lineages; arrows represent migration parameters; labels indicate model parameters, as defined in the panel. (D) Stairway plot reconstructions of historical effective population size (N_e) for East Asian lineages based on the SFS. The x-axis is time before present and the y-axis is Ne. The grey band denotes the Last Glacial Maximum (LGM; 26.5–19 ka; Clark et al. 2009). Solid lines represent point estimates and shaded areas represent confidence intervals.

from BIO5 (−0.472). In the constrained ordination space, most mainland Chinese populations formed a relatively tight central cluster. Populations from peripheral regions—Taiwan, Japan and Yunnan—were distinctly separated from this central cluster and positioned outside the main mainland group (Figure 4B).

4 | Discussion

4.1 | Biogeographic Origin of the EA-ENA Disjunction in the *Aquarius conformis* Species Group

Our nuclear genomic analyses indicated that the *A. conformis* species group is composed of an East Asian lineage represented by *A. elongatus* and a North American lineage comprising *A. conformis* and *A. nebularis*. The Neighbour-Net network and PCA based on genome-wide SNPs revealed a clear genetic separation between East Asia and North America, whereas the two North American species showed closer genetic affinity to each other. This pattern suggests that the present EA-ENA disjunction was shaped by an intercontinental split, followed by shallower diversification within North America. Model-based biogeographic inference further supported an East Asian origin of the species group. Among the alternative scenarios tested in DIYABC-RF, the East Asian-origin model was favoured over the North American-origin and widespread-ancestor models. This result was consistent with the broader phylogenetic and biogeographic pattern, in which the East Asian lineage occupied a basal position within the focal group and close relatives were inferred to have originated in East Asia. The agreement among the DIYABC-RF model comparison, phylogenetic placement and inferred source areas of close

1365294x, 2026, 16, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70533 by Zhen Ye - Nankai University , Wiley Online Library on [25/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

relatives supports East Asia as the most likely source region. The selected demographic model reinforced this scenario. The best-fitting fastsimcoal2 model favoured an initial separation between the East Asian and North American lineages without detectable post-divergence gene flow, whereas asymmetric migration was inferred only between *A. conformis* and *A. nebularis*. Under the assumed mutation rate and generation time, the estimated divergence between East Asia and North America occurred at approximately 458 ka, with a 95% confidence interval of 433–483 ka. This result places the origin of the EA-ENA split in the Middle Pleistocene, specifically around Marine Isotope Stage 12 (MIS 12), one of the most severe glacial stages of the Quaternary. We therefore suggest that the ancestor of the *A. conformis* species group most likely originated in East Asia and colonised North America during a major Middle Pleistocene glacial interval.

The timing of this intercontinental split provides an important palaeoclimatic context for the formation of the EA-ENA disjunction in this species group. MIS 12 was characterised by extensive Northern Hemisphere glaciation, enlarged continental ice sheets, cold and dry climatic conditions and substantially lowered global sea levels (Elderfield et al. 2012; Spratt and Lisiecki 2016). Although continental ice sheets expanded extensively during MIS 12, the Beringian lowlands remained largely unglaciated (Brigham-Grette 2001). Together with substantially lowered sea levels, this condition could have exposed a broad, largely ice-free land connection between northeastern Asia and northwestern North America, potentially facilitating high-latitude dispersal. For mobile semiaquatic insects, lowered sea levels and expanded high-latitude lowlands may have created temporary dispersal opportunities by increasing the availability of freshwater margins, floodplains, ponds and slow-flowing water bodies across northern routes. Under this interpretation, the ca. 460 ka divergence of the East Asian and North American lineages was likely associated with a glacial-stage dispersal window rather than with the fragmentation of a formerly widespread ancient Northern Hemisphere ancestor. Subsequent climatic amelioration after MIS 12 may have promoted long-term isolation between the two continental lineages. The transition from MIS 12 to MIS 11 involved major deglaciation, sea-level rise and reorganisation of high-latitude environments (Elderfield et al. 2012; Spratt and Lisiecki 2016; Tzedakis et al. 2022). Rising sea levels would have reduced or eliminated transient Beringian connections, while interglacial warming and the redistribution of freshwater habitats may have fragmented the ecological corridors that had facilitated earlier dispersal. This scenario is consistent with our demographic inference showing no detectable post-divergence gene flow between the East Asian and North American lineages. Thus, the EA-ENA disjunction in the *A. conformis* species group was probably shaped by a sequence of Middle Pleistocene events: East Asian origin, glacial-stage dispersal into North America around 460 ka, and subsequent intercontinental isolation during post-glacial environmental reorganisation. Combined with the above analyses, we favour an East Asian-origin and MIS 12-associated dispersal-mediated scenario for the formation of the EA-ENA disjunction in the *A. conformis* species group. In this scenario, the ancestor of the group originated in East Asia, expanded into North America during a major Middle Pleistocene glacial lowstand, and then became isolated from the East Asian source lineage as high-latitude connections and freshwater habitats were reorganised during deglaciation and interglacial warming.

4.2 | Hydroclimatic Differentiation and Phenotypic Divergence After EA-ENA Disjunction

After the establishment of the North American lineage and its isolation from the East Asian source lineage, the continental lineages appear to have followed distinct evolutionary trajectories. Morphological analyses based on fifteen diagnostic traits revealed pronounced phenotypic differentiation between the East Asian and North American species. In particular, *A. elongatus* from East Asia is substantially larger than the two North American congeners, *A. conformis* and *A. nebularis*, with body length and multiple appendage-related traits showing consistent intercontinental differences. This size divergence is one of the most conspicuous phenotypic features of the species group, and this phenotypic contrast is also associated with distinct habitat and climatic backgrounds. *A. elongatus* is widely distributed in East Asia and is mainly associated with lentic or slow-flowing habitats, whereas *A. conformis* and *A. nebularis* occur in eastern North America and are more frequently associated with lotic habitats (Andersen 1990; Damgaard and Cognato 2006). Our environmental analyses further suggested that the East Asian and North American lineages occupy different climatic spaces. The linear SVM analysis identified precipitation seasonality (BIO15) as the strongest climatic variable distinguishing *A. elongatus* from the two North American species, followed by mean diurnal temperature range (BIO2), isothermality (BIO3) and precipitation of the driest month (BIO14). The resulting composite environmental axis was therefore mainly determined by precipitation-related and temperature-variation variables, with BIO15 contributing the highest weight. These results suggest that the continental lineages differ not only in geography and morphology, but also in the hydroclimatic conditions under which they occur. For semiaquatic insects, such hydroclimatic differences may be biologically meaningful. Precipitation seasonality can affect the permanence, predictability and connectivity of freshwater habitats, especially the availability of temporary lentic habitats versus persistent flowing-water systems (Rolls et al. 2018; Sánchez-Montoya et al. 2023). In East Asia, stronger precipitation seasonality may generate more heterogeneous and temporally variable aquatic habitats, potentially favouring dispersal-oriented life-history strategies and repeated colonisation across fragmented water bodies (Arribas et al. 2012; Phillipsen et al. 2015). In contrast, more stable lotic habitats in eastern North America may favour local persistence in flowing-water environments. Under this interpretation, hydroclimatic heterogeneity provides a plausible ecological background for the independent evolution of the East Asian and North American lineages after intercontinental isolation.

The genomic results are consistent with this ecological interpretation. LFMM analyses identified candidate SNPs associated with the main environmental gradient separating the continental lineages. These loci were located within or near 257 genes, among which several were annotated to pathways with putative roles in growth regulation, nutrient sensing, hormone biosynthesis, metabolism, stress response and immune function. In

1365294x, 2026, 16, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70533 by Zhen Ye - Nankai University , Wiley Online Library on [25/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

particular, candidate genes were associated with pathways such as Hippo signalling, mTOR signalling, FoxO signalling, juvenile hormone biosynthesis and Toll and Imd immune-related pathways. Genes such as *jhamt*, *app*, *Act5C*, *fz*, *Sirt1* and *Ank* are of particular interest because homologous genes in insects or other model systems have been linked to developmental regulation, cell proliferation, nutrient sensing, hormonal control, stress response, or immune processes. Together, these annotations suggest that hydroclimatic differentiation may be associated with genome-wide allele-frequency variation in pathways relevant to growth, physiology and environmental response. However, functional annotation does not demonstrate that these loci directly regulate body size in the *A. conformis* species group and genotype–environment association analyses should be interpreted as hypothesis-generating rather than causal tests (Capblancq et al. 2018; Forester et al. 2018). PSMC analyses also indicate that the continental lineages responded differently to late Quaternary climatic oscillations. The East Asian *A. elongatus* maintained a relatively stable demographic trajectory, whereas the two North American species exhibited stronger fluctuations and contractions during late Pleistocene climatic changes, a period characterised by repeated glacial, sea-level and abrupt climate oscillations (Hemming 2004; Lambeck and Chappell 2001). This demographic contrast is consistent with the possibility that the East Asian and North American lineages experienced different habitat stability and climatic histories after their separation. However, the ancestral North American lineage may have been established by a small number of colonists from East Asia. This possibility is consistent with the much smaller effective population sizes estimated for *A. conformis* and *A. nebularis* than for *A. elongatus*. A founder effect and lower effective population sizes could have amplified the demographic effects of subsequent climatic and habitat changes. It reinforces the view that the two continental lineages evolved under different historical and ecological contexts after the formation of the EA-ENA disjunction. Combined with the above analyses, we suggest that hydroclimatic differentiation may have contributed to divergent post-disjunction evolution in the *A. conformis* species group.

4.3 | Pleistocene Divergence and Successful Colonisation of Heterogeneous Environments by East Asian *A. elongatus*

Population structure analyses based on ddRAD nuclear SNPs revealed that East Asian *A. elongatus* comprises four well-supported phylogeographic lineages corresponding to major geographic units: Eastern China (EC), Yunnan (YN), Taiwan (TW) and Japan (JPA). Coalescent-based demographic model selection supported a sequential divergence history, with the Japan (JPA) lineage splitting first (258 ka), followed by the Taiwan (TW) lineage (38 ka), and finally the split between eastern China (EC) and Yunnan (YN) lineages (35 ka). These divergence times align closely with late Pleistocene glacial–interglacial cycles, during which repeated exposure of continental shelves and land bridges could have facilitated colonisation of Taiwan and Japan, followed by renewed marine isolation during subsequent sea-level rises and abrupt climate events (Hemming 2004; Kimura 2000; Lambeck and Chappell 2001; Qi et al. 2014). Concurrently, the Yunnan divergence coincided with intensified climatic and topographic fragmentation in the Hengduan Mountains and adjacent regions, where steep elevational gradients and complex mountain building have promoted geographic isolation and lineage diversification (Spicer et al. 2021; Xing and Ree 2017). Despite these episodic barriers, divergence proceeded under incomplete isolation. Demographic models detected bidirectional but asymmetric post-divergence gene flow between the eastern China lineage (EC) and the YN, TW and JPA lineages, indicating recurrent connectivity after divergence rather than long-term complete separation. Stairway plot analyses further indicated lineage-specific demographic responses during the late Pleistocene, with EC and YN maintaining comparatively stable effective population sizes across the LGM, whereas JPA and, more notably, TW showed more pronounced demographic changes.

As a macropterous species, *A. elongatus* likely possesses greater dispersal potential than brachypterous water striders; more broadly, studies of aquatic insects show that dispersal ability and habitat requirements strongly influence whether gene flow or genetic drift dominates population structure across freshwater landscapes (Dijkstra et al. 2014; Phillipsen et al. 2015). This ability may have enabled recurrent colonisation of peripheral regions through transient connectivity via glacial land bridges, riverine corridors and low-elevation saddles during suitable climatic intervals (Kimura 2000; Lambeck and Chappell 2001; Qi et al. 2014). Partial redundancy analysis (pRDA) of ddRAD-seq SNPs, conditioning on neutral population structure, identified significant climate-associated allele-frequency variation, with strong associations to precipitation seasonality (BIO15), annual precipitation (BIO12) and temperature variables (BIO5, BIO6, BIO8). Because pRDA identifies climate-associated allele-frequency variation rather than directly demonstrating local adaptation, these results are best interpreted as genomic signals consistent with local adaptation hypotheses (Capblancq et al. 2018; Forester et al. 2018). Collectively, the sequential Pleistocene divergence, dynamic demographic histories and climate-associated genetic variation suggest that *A. elongatus* has been a successful regional coloniser, capable of establishing and persisting across a wide range of heterogeneous lentic habitats throughout East Asia despite repeated climatic fluctuations.

4.4 | Multiple Routes and Post-Disjunction Divergence in EA-ENA Lineages

The classical interpretation of EA-ENA disjunctions has largely been developed from temperate plant lineages. In this view, formerly widespread Northern Hemisphere taxa were fragmented by climatic deterioration, regional extinction and the loss of suitable intermediate habitats, leaving relict lineages in East Asia and eastern North America (Donoghue and Smith 2004; Tiffney 1985; Wen 1999). This model remains an important baseline because it captures a recurrent combination of high-latitude connectivity, climatic filtering, extinction in intervening regions and persistence in refugial areas (Figure 5A). However, comparative studies also show that EA-ENA disjunctions are not the product of a single historical route. Divergence times span multiple geological intervals, and ancestral-area reconstructions have supported East Asian origins, North American origins and

1365294x, 2026, 16, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70533 by Zhen Ye - Nankai University , Wiley Online Library on [25/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

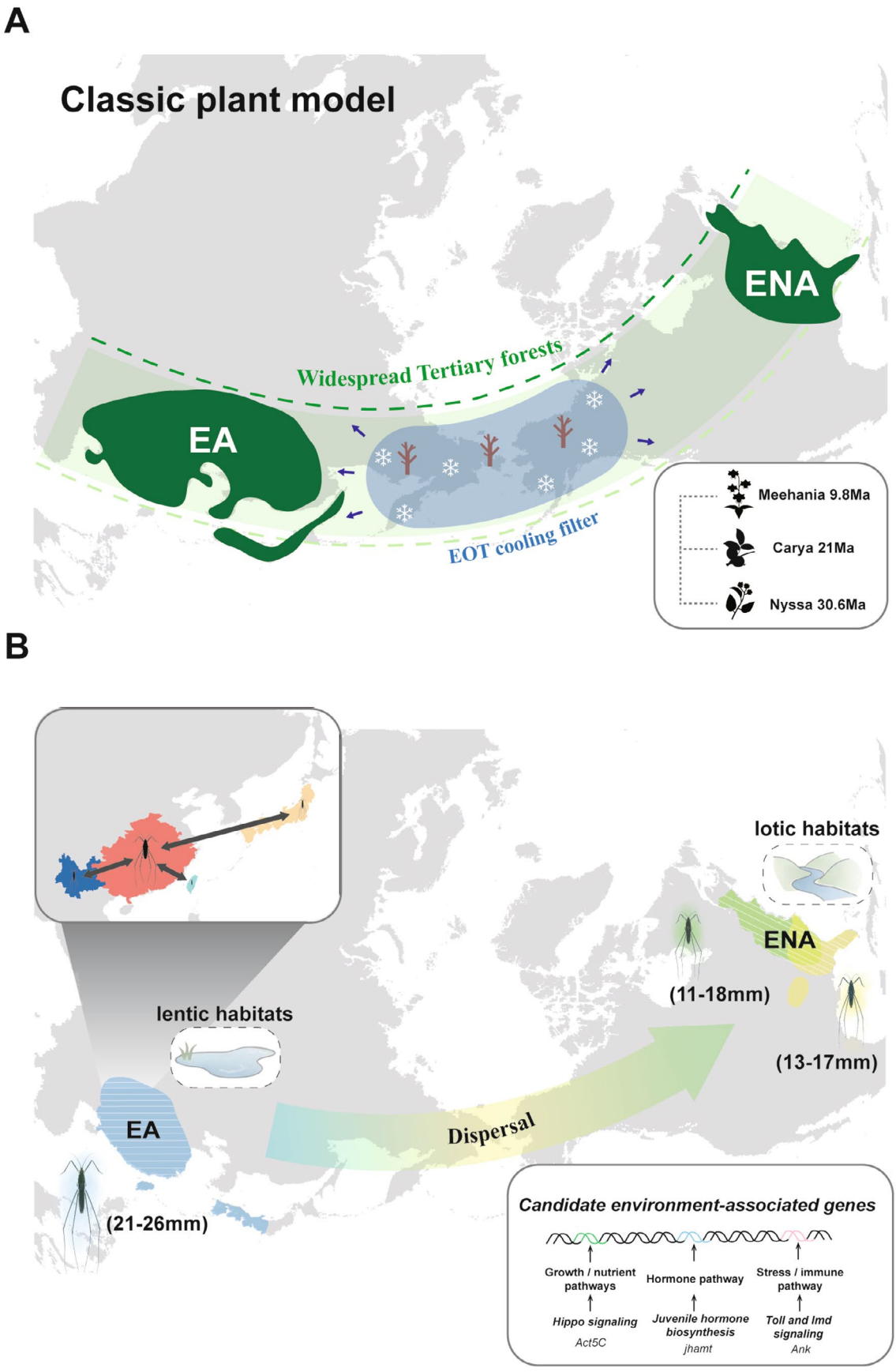

FIGURE 5 | Legend on next page.

1365294x, 2026, 16, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70533 by Zhen Ye - Nankai University , Wiley Online Library on [25/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

FIGURE 5 | Schematic diagrams of the widespread-ancestor vicariance model and the *Aquarius conformis* species group. (A) Schematic representation of the widespread-ancestor vicariance model for East Asia–Eastern North America disjunct lineages. Shaded areas indicate a formerly widespread Northern Hemisphere distribution and present relict regions in East Asia and eastern North America. Arrows and symbols denote range fragmentation and regional extinction as shown in the diagram. The lower-right inset lists representative plant cases and published divergence-time estimates included in the comparison. (B) Geographic schematic of the *A. conformis* species group. Shaded regions indicate the distributions of *A. elongatus*, *A. conformis* and *A. nebularis*. The inset shows East Asian lineages of *A. elongatus*. The large arrow indicates the dispersal arrow shown in the diagram. Habitat icons indicate lentic and lotic habitat categories. The candidate-gene box lists representative GEA candidate genes and annotated pathways shown in the diagram.

widespread ancestral ranges followed by fragmentation across different clades (Lin et al. 2023; Melton et al. 2022; Wen 1999). Our results place the *A. conformis* species group within a source-area dispersal route. Model-based biogeographic inference supports East Asia as the most likely source region, and demographic modelling indicates effective isolation between East Asia and North America, with asymmetric gene flow detected only between *A. conformis* and *A. nebularis*. This pattern is most consistent with East Asian origin, intercontinental colonisation, isolation from the source region and subsequent diversification within North America. Thus, the *A. conformis* species group is not a direct analogue of the classical widespread-ancestor relict model. Instead, it illustrates another way in which the same EA-ENA disjunct pattern can be assembled: a source lineage colonises a second continent and then diversifies regionally after establishment (Figure 5B).

Studies of EA-ENA plants further show that evolution after disjunction is not uniform. Comparative niche analyses have often emphasised climatic niche conservatism, with disjunct relatives retaining broadly comparable climatic spaces after continental isolation (Melton et al. 2022; Yin et al. 2021). Other cases indicate that ecological divergence and regional demographic history can also shape post-disjunction evolution. In *Liriodendron*, for example, intercontinental sister species exhibit niche divergence and environment-associated genetic differentiation, suggesting that heterogeneous environments may contribute to local adaptation after long-term separation (Shen et al. 2022). The *A. conformis* species group adds a complementary example from semiaquatic insects. Its inferred East Asian origin and North American diversification explain the formation of the disjunction, whereas body-size divergence, hydroclimatic differentiation, environment-associated candidate loci and contrasting demographic trajectories point to phenotypic, ecological and genomic divergence after isolation (Figure 5B). These GEA results should therefore be interpreted as candidate associations rather than causal evidence for body-size divergence (Capblancq et al. 2018; Forester et al. 2018). More broadly, our results suggest that the biological traits of a lineage may influence both how an EA-ENA disjunction forms and how the separated lineages continue to evolve. Plant disjunctions are often discussed in relation to forest continuity, macroclimatic niche space, physiographic heterogeneity, topography and lineage-specific environmental divergence (Donoghue and Smith 2004; Lin et al. 2023; Melton et al. 2022; Wen 1999). For freshwater insects, a different set of habitat features may also be important. Standing and running waters differ in persistence, connectivity and colonisation dynamics, and studies of aquatic beetles have shown that dispersal ability can explain range-size differences across the lentic–lotic divide better than ecological tolerance alone (Arribas et al. 2012; Ribera and Vogler 2000). Landscape-genetic studies of aquatic insects further show that dispersal ability and habitat requirements influence whether gene flow or genetic drift dominates population structure across freshwater networks (Phillipsen et al. 2015). Thus, freshwater connectivity, habitat permanence, lentic–lotic structure and precipitation seasonality may be especially relevant axes for semiaquatic insects (Dijkstra et al. 2014; Rolls et al. 2018; Sánchez-Montoya et al. 2023). Taken together, the *A. conformis* species group illustrates that an EA-ENA disjunction can arise through East Asian source-area dispersal followed by intercontinental isolation and regional diversification in North America, rather than through fragmentation of a widespread ancestor. The ensuing hydroclimatic, phenotypic and demographic divergence, together with genomic signals consistent with local adaptation, further suggests that lineage-specific dispersal capacity and habitat dependence may influence both how intercontinental disjunctions form and how the separated lineages subsequently diversify.

Author Contributions

Yuqi Zai: performed research; collected samples; analysed data; wrote the paper. **Matthew R. Pintar:** performed research; collected samples; reviewed and edited the paper. **Mu Qiao:** collected samples; analysed data. **Chen Liu:** collected samples; analysed data. **Siying Fu:** collected samples; analysed data. **Jakob Damgaard:** collected samples; reviewed and edited the paper. **Jiayue Zhou:** analysed data. **Wenjun Bu:** reviewed and edited the paper. **Juanjuan Yuan:** designed research; collected samples; acquired funding; reviewed and edited the paper. **Zhen Ye:** designed research; collected samples; acquired funding; wrote the paper; reviewed and edited the paper.

Acknowledgements

We thank our lab's members for help in collecting some specimens for this study.

Funding

This project is funded by the National Natural Science Foundation of China (No. 32200368, No. 32322012), Shandong Provincial University Youth Innovation and Technology Program, China (2024KJG077) and Shandong Provincial Natural Science Foundation (ZR2021QC229).

Ethics Statement

Samples were collected from non-protected areas with the permission of the landowners. We did not use artificial intelligence (AI) tools at any stage of this study.

13652946x, 2026, 16, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70533 by Zhen Ye - Nankai University , Wiley Online Library on [25/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All data generated or analysed during this study are included in this published article and its Supporting Information files S1. Assembled sequences newly generated in the present study are available in GenBank (accession number SRR38898735–SRR38898750, SRR39046210–SRR39046339, SRR39063676–SRR39063679). Original sequence reads of genomic data are available in GenBank (accession number JCBRLG000000000).

References

Andersen, N. M. 1982. "The Semiaquatic Bugs (Hemiptera, Gerromorpha): Phylogeny, Adaptations, Biogeography and Classification." *Entomon* 3: 1–455.

Andersen, N. M. 1990. "Phylogeny and Taxonomy of Water Striders, Genus Aquarius Schellenberg (Insecta, Hemiptera, Gerridae), with a New Species From Australia." *Steenstrupia* 16: 37–81.

Arribas, P., J. Velasco, P. Abellán, et al. 2012. "Dispersal Ability Rather Than Ecological Tolerance Drives Differences in Range Size Between Lentic and Lotic Water Beetles (Coleoptera: Hydrophilidae)." *Journal of Biogeography* 39, no. 5: 984–994. https://doi.org/10.1111/j.1365-2699.2011.02641.x.

Brigham-Grette, J. 2001. "New Perspectives on Beringian Quaternary Paleogeography, Stratigraphy, and Glacial History." *Quaternary Science Reviews* 20, no. 1–3: 15–24. https://doi.org/10.1016/S0277-3791(00)00134-7.

Bu, D., H. Luo, P. Huo, et al. 2021. "KOBAS-i: Intelligent Prioritization and Exploratory Visualization of Biological Functions for Gene Enrichment Analysis." *Nucleic Acids Research* 49, no. W1: W317–W325. https://doi.org/10.1093/nar/gkab447.

Capblancq, T., K. Luu, M. G. B. Blum, and E. Bazin. 2018. "Evaluation of Redundancy Analysis to Identify Signatures of Local Adaptation." *Molecular Ecology Resources* 18, no. 6: 1223–1233. https://doi.org/10.1111/1755-0998.12906.

Chamberlain, S. A., and C. Boettiger. 2017. "R, Python, and Ruby Clients for GBIF Species Occurrence Data." *PeerJ Preprints* 5: e3304v1. https://doi.org/10.7287/peerj.preprints.3304v1.

Clark, P. U., A. S. Dyke, J. D. Shakun, et al. 2009. "The Last Glacial Maximum." *Science* 325, no. 5941: 710–714. https://doi.org/10.1126/science.1172873.

Collin, F. D., G. Durif, L. Raynal, et al. 2021. "Extending Approximate Bayesian Computation With Supervised Machine Learning to Infer Demographic History From Genetic Polymorphisms Using DIYABC Random Forest." *Molecular Ecology Resources* 21, no. 8: 2598–2613. https://doi.org/10.1111/1755-0998.13413.

Cortes, C., and V. Vapnik. 1995. "Support-Vector Networks." *Machine Learning* 20, no. 3: 273–297. https://doi.org/10.1007/BF00994018.

Coyne, J. A., and H. A. Orr. 2004. *Speciation*. Sinauer Associates.

Damgaard, J., and A. I. Cognato. 2006. "Phylogeny and Reclassification of Species Groups in Aquarius Schellenberg, Limnoporus Stål and Gerris Fabricius (Insecta: Hemiptera-Heteroptera, Gerridae)." *Systematic Entomology* 31, no. 1: 93–112. https://doi.org/10.1111/j.1365-3113.2005.00302.x.

Danecek, P., A. Auton, G. Abecasis, et al. 2011. "The Variant Call Format and VCFtools." *Bioinformatics* 27, no. 15: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Danecek, P., J. K. Bonfield, J. Liddle, et al. 2021. "Twelve Years of SAMtools and BCFtools." *GigaScience* 10, no. 2: giab008. https://doi.org/10.1093/gigascience/giab008.

DePristo, M. A., E. Banks, R. Poplin, et al. 2011. "A Framework for Variation Discovery and Genotyping Using Next-Generation DNA Sequencing Data." *Nature Genetics* 43, no. 5: 491–498. https://doi.org/10.1038/ng.806.

Dijkstra, K.-D. B., M. T. Monaghan, and S. U. Pauls. 2014. "Freshwater Biodiversity and Aquatic Insect Diversification." *Annual Review of Entomology* 59: 143–163. https://doi.org/10.1146/annurev-ento-011613-161958.

Donoghue, M. J., and S. A. Smith. 2004. "Patterns in the Assembly of Temperate Forests Around the Northern Hemisphere." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 359, no. 1450: 1633–1644. https://doi.org/10.1098/rstb.2004.1538.

Earl, D. A., and B. M. vonHoldt. 2012. "STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method." *Conservation Genetics Resources* 4, no. 2: 359–361. https://doi.org/10.1007/s12686-011-9548-7.

Eaton, D. A. R., and I. Overcast. 2020. "Ipyrad: Interactive Assembly and Analysis of RADseq Datasets." *Bioinformatics* 36, no. 8: 2592–2594. https://doi.org/10.1093/bioinformatics/btz966.

Elderfield, H., P. Ferretti, M. Greaves, et al. 2012. "Evolution of Ocean Temperature and Ice Volume Through the Mid-Pleistocene Climate Transition." *Science* 337, no. 6095: 704–709. https://doi.org/10.1126/science.1221294.

Excoffier, L., I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, and M. Foll. 2013. "Robust Demographic Inference From Genomic and SNP Data." *PLoS Genetics* 9, no. 10: e1003905. https://doi.org/10.1371/journal.pgen.1003905.

Excoffier, L., N. Marchi, D. A. Marques, R. Matthey-Doret, A. Gouy, and V. C. Sousa. 2021. "fastsimcoal2: Demographic Inference Under Complex Evolutionary Scenarios." *Bioinformatics* 37, no. 24: 4882–4885. https://doi.org/10.1093/bioinformatics/btab468.

Fick, S. E., and R. J. Hijmans. 2017. "WorldClim 2: New 1-Km Spatial Resolution Climate Surfaces for Global Land Areas." *International Journal of Climatology* 37, no. 12: 4302–4315. https://doi.org/10.1002/joc.5086.

Forester, B. R., J. R. Lasky, H. H. Wagner, and D. L. Urban. 2018. "Comparing Methods for Detecting Multilocus Adaptation With Multivariate Genotype-Environment Associations." *Molecular Ecology* 27, no. 9: 2215–2233. https://doi.org/10.1111/mec.14584.

Frichot, E., and O. François. 2015. "LEA: An R Package for Landscape and Ecological Association Studies." *Methods in Ecology and Evolution* 6, no. 8: 925–929. https://doi.org/10.1111/2041-210X.12382.

Frichot, E., S. D. Schoville, G. Bouchard, and O. François. 2013. "Testing for Associations Between Loci and Environmental Gradients Using Latent Factor Mixed Models." *Molecular Biology and Evolution* 30, no. 7: 1687–1699. https://doi.org/10.1093/molbev/mst063.

Harvey, M. G., S. Singhal, and D. L. Rabosky. 2019. "Beyond Reproductive Isolation: Demographic Controls on the Speciation Process." *Annual Review of Ecology, Evolution, and Systematics* 50: 75–95. https://doi.org/10.1146/annurev-ecolsys-110218-024701.

He, W., S. Zhao, X. Liu, et al. 2013. "ReSeqTools: An Integrated Toolkit for Large-Scale Next-Generation Sequencing Based Resequencing Analysis." *Genetics and Molecular Research* 12, no. 4: 6275–6283. https://doi.org/10.4238/2013.December.4.15.

Hemming, S. R. 2004. "Heinrich Events: Massive Late Pleistocene Detritus Layers of the North Atlantic and Their Global Climate Imprint." *Reviews of Geophysics* 42, no. 1: RG1005. https://doi.org/10.1029/2003RG000128.

13652394x, 2026, 16, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70533 by Zhen Ye - Nankai University , Wiley Online Library on [25/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Huson, D. H., and D. Bryant. 2006. "Application of Phylogenetic Networks in Evolutionary Studies." *Molecular Biology and Evolution* 23, no. 2: 254–267. https://doi.org/10.1093/molbev/msj030.

IBM Corp. 2010. *IBM SPSS Statistics for Windows, Version 19.0*. IBM Corp.

Jakobsson, M., and N. A. Rosenberg. 2007. "CLUMPP: A Cluster Matching and Permutation Program for Dealing With Label Switching and Multimodality in Analysis of Population Structure." *Bioinformatics* 23, no. 14: 1801–1806. https://doi.org/10.1093/bioinformatics/btm233.

Jombart, T. 2008. "Adegenet: An R Package for the Multivariate Analysis of Genetic Markers." *Bioinformatics* 24, no. 11: 1403–1405. https://doi.org/10.1093/bioinformatics/btn129.

Keightley, P. D., U. Trivedi, M. Thomson, F. Oliver, S. Kumar, and M. L. Blaxter. 2009. "Analysis of the Genome Sequences of Three *Drosophila melanogaster* Spontaneous Mutation Accumulation Lines." *Genome Research* 19, no. 7: 1195–1201. https://doi.org/10.1101/gr.091231.109.

Kimura, M. 2000. "Paleogeography of the Ryukyu Islands." *Tropics* 10, no. 1: 5–24. https://doi.org/10.3759/tropics.10.5.

Lambeck, K., and J. Chappell. 2001. "Sea Level Change Through the Last Glacial Cycle." *Science* 292, no. 5517: 679–686. https://doi.org/10.1126/science.1059549.

Li, D., C. M. Liu, R. Luo, K. Sadakane, and T. W. Lam. 2015. "MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph." *Bioinformatics* 31, no. 10: 1674–1676. https://doi.org/10.1093/bioinformatics/btv033.

Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment With Burrows-Wheeler Transform." *Bioinformatics* 25, no. 14: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Li, H., and R. Durbin. 2011. "Inference of Human Population History From Individual Whole-Genome Sequences." *Nature* 475, no. 7357: 493–496. https://doi.org/10.1038/nature10231.

Lin, H. Y., M. Sun, Y. J. Hao, et al. 2023. "Phylogenetic Diversity of Eastern Asia-Eastern North America Disjunct Plants Is Mainly Associated With Divergence Time." *Plant Diversity* 45, no. 1: 27–35. https://doi.org/10.1016/j.pld.2022.09.008.

Liu, X., and Y. X. Fu. 2015. "Exploring Population Size Changes Using SNP Frequency Spectra." *Nature Genetics* 47, no. 5: 555–559. https://doi.org/10.1038/ng.3254.

Mayr, E. 1942. *Systematics and the Origin of Species*. Columbia University Press.

Melton, A. E., M. H. Clinton, D. N. Wasoff, et al. 2022. "Climatic Niche Comparisons of Eastern North American and Eastern Asian Disjunct Plant Genera." *Global Ecology and Biogeography* 31, no. 7: 1290–1302. https://doi.org/10.1111/geb.13504.

Meyer, D., E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch. 2024. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (formerly: E1071), TU Wien. R Package Version 1.7–16.

Nosil, P., T. H. Vines, and D. J. Funk. 2005. "Reproductive Isolation Caused by Natural Selection Against Immigrants From Divergent Habitats." *Evolution* 59, no. 4: 705–719. https://doi.org/10.1111/j.0014-3820.2005.tb01747.x.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. "Double Digest RADseq: An Inexpensive Method for de Novo SNP Discovery and Genotyping in Model and Non-Model Species." *PLoS One* 7, no. 5: e37135. https://doi.org/10.1371/journal.pone.0037135.

Phillipsen, I. C., E. H. Kirk, M. T. Bogan, M. C. Mims, J. D. Olden, and D. A. Lytle. 2015. "Dispersal Ability and Habitat Requirements Determine Landscape-Level Genetic Patterns in Desert Aquatic Insects." *Molecular Ecology* 24, no. 1: 54–69. https://doi.org/10.1111/mec.13003.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. "Inference of Population Structure Using Multilocus Genotype Data." *Genetics* 155, no. 2: 945–959. https://doi.org/10.1093/genetics/155.2.945.

Pudlo, P., J. M. Marin, A. Estoup, J. M. Cornuet, M. Gautier, and C. P. Robert. 2016. "Reliable ABC Model Choice via Random Forests." *Bioinformatics* 32, no. 6: 859–866. https://doi.org/10.1093/bioinformatics/btv684.

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

Qi, X. S., N. Yuan, H. P. Comes, S. Sakaguchi, and Y. X. Qiu. 2014. "A Strong "Filter" Effect of the East China Sea Land Bridge for East Asia's Temperate Plant Species: Inferences From Molecular Phylogeography and Ecological Niche Modelling of Platycrater Arguta (Hydrangeaceae)." *BMC Evolutionary Biology* 14: 41. https://doi.org/10.1186/1471-2148-14-41.

Ren, Z., C. D. von Dohlen, A. J. Harris, R. B. Dikow, X. Su, and J. Wen. 2019. "Congruent Phylogenetic Relationships of Melaphidina Aphids (Aphididae: Eriosomatinae: Fordini) According to Nuclear and Mitochondrial DNA Data With Taxonomic Implications on Generic Limits." *PLoS One* 14, no. 2: e0213181. https://doi.org/10.1371/journal.pone.0213181.

Ren, Z., Y. Zhong, U. Kurosu, et al. 2013. "Historical Biogeography of Eastern Asian-Eastern North American Disjunct Melaphidina Aphids (Hemiptera: Aphididae: Eriosomatinae) on Rhus Hosts (Anacardiaceae)." *Molecular Phylogenetics and Evolution* 69, no. 3: 1146–1158. https://doi.org/10.1016/j.ympev.2013.08.003.

Ribera, I., and A. P. Vogler. 2000. "Habitat Type as a Determinant of Species Range Sizes: The Example of Lotic-Lentic Differences in Aquatic Coleoptera." *Biological Journal of the Linnean Society* 71, no. 1: 33–52. https://doi.org/10.1006/bijl.1999.0412.

Rolls, R. J., J. Heino, D. S. Ryder, et al. 2018. "Scaling Biodiversity Responses to Hydrological Regimes." *Biological Reviews* 93, no. 2: 971–995. https://doi.org/10.1111/brv.12381.

Rose, J. P., J. Wiese, N. Pauley, et al. 2023. "East Asian-North American Disjunctions and Phylogenetic Relationships Within Subtribe Nepetinae (Lamiaceae)." *Molecular Phylogenetics and Evolution* 187: 107873. https://doi.org/10.1016/j.ympev.2023.107873.

Rosenberg, N. A. 2004. "DISTRUCT: A Program for the Graphical Display of Population Structure." *Molecular Ecology Notes* 4, no. 1: 137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x.

Sánchez-Montoya, M. M., T. Datry, A. Ruhi, S. M. Carlson, R. Corti, and K. Tockner. 2023. "Intermittent Rivers and Ephemeral Streams Are Pivotal Corridors for Aquatic and Terrestrial Animals." *Bioscience* 73, no. 4: 291–301. https://doi.org/10.1093/biosci/biad004.

Schluter, D. 2009. "Evidence for Ecological Speciation and Its Alternative." *Science* 323, no. 5915: 737–741. https://doi.org/10.1126/science.1160006.

Shen, Y., H. Xia, Z. Tu, Y. Zong, L. Yang, and H. Li. 2022. "Genetic Divergence and Local Adaptation of Liriodendron Driven by Heterogeneous Environments." *Molecular Ecology* 31, no. 3: 916–933. https://doi.org/10.1111/mec.16271.

Spicer, R. A., T. Su, P. J. Valdes, et al. 2021. "Why "the Uplift of the Tibetan Plateau" Is a Myth." *National Science Review* 8, no. 1: nwaa091. https://doi.org/10.1093/nsr/nwaa091.

Spratt, R. M., and L. E. Lisiecki. 2016. "A Late Pleistocene Sea Level Stack." *Climate of the Past* 12, no. 4: 1079–1092. https://doi.org/10.5194/cp-12-1079-2016.

Tiffney, B. H. 1985. "The Eocene North Atlantic Land Bridge: Its Importance in Tertiary and Modern Phytogeography of the Northern Hemisphere." *Journal of the Arnold Arboretum* 66, no. 2: 243–273. https://doi.org/10.5962/bhl.part.13183.

1365294x, 2026, 16, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70533 by Zhen Ye - Nankai University , Wiley Online Library on [25/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Tzedakis, P. C., D. A. Hodell, C. Nehrbass-Ahles, T. Mitsui, and E. W. Wolff. 2022. "Marine Isotope Stage 11c: An Unusual Interglacial." *Quaternary Science Reviews* 284: 107493. https://doi.org/10.1016/j.quascirev.2022.107493.

von Dohlen, C. D., U. Kurosu, and S. Aoki. 2002. "Phylogenetics and Evolution of the Eastern Asian-Eastern North American Disjunct Aphid Tribe Hormaphidini (Hemiptera: Aphididae)." *Molecular Phylogenetics and Evolution* 23, no. 2: 257–267. https://doi.org/10.1016/S1055-7903(02)00025-8.

Wen, J. 1999. "Evolution of Eastern Asian and Eastern North American Disjunct Distributions in Flowering Plants." *Annual Review of Ecology and Systematics* 30: 421–455. https://doi.org/10.1146/annurev.ecolsys.30.1.421.

Wen, J., S. M. Ickert-Bond, Z. L. Nie, and R. Li. 2010. "Timing and Modes of Evolution of Eastern Asian-North American Biogeographic Disjunctions in Seed Plants." In *Darwin's Heritage Today: Proceedings of the Darwin 200 Beijing International Conference*, edited by M. Long, H. Gu, and Z. Zhou, 252–269. Higher Education Press.

Wickham, H. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer. https://doi.org/10.1007/978-0-387-98141-3.

Xing, Y., and R. H. Ree. 2017. "Uplift-Driven Diversification in the Hengduan Mountains, a Temperate Biodiversity Hotspot." *Proceedings of the National Academy of Sciences of the United States of America* 114, no. 17: E3444–E3451. https://doi.org/10.1073/pnas.1616063114.

Yin, X., S. Jarvie, W. Y. Guo, et al. 2021. "Niche Overlap and Divergence Times Support Niche Conservatism in Eastern Asia-Eastern North America Disjunct Plants." *Global Ecology and Biogeography* 30, no. 10: 1990–2003. https://doi.org/10.1111/geb.13360.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1:** Schematic diagrams of the 8 demographic models tested with fastsimcoal2 based on whole-genome resequencing data to infer divergence history within the *A. conformis* species group. Each model represents a distinct hypothesis about the presence of post-divergence gene flow between lineage pairs. Arrows denote the presence and direction of gene flow between lineages. Model definitions and model-selection results are provided in Table S5. **Figure S2:** Schematics of the 23 demographic models compared in fastsimcoal2 using ddRAD-seq data to infer the evolutionary history of *A. elongatus*. In Step I, 16 alternative divergence topologies (M1-M16) were evaluated; in Step II, seven models (M17-M23) added gene flow to the best-supported topology from Step I. Arrows indicate the presence and direction of gene flow. Detailed model descriptions are provided in Table S8. **Figure S3:** Quality assessment of the de novo genome assembly and Hi-C scaffolding. (A) GC-depth distribution of the assembled genome. The *x*-axis shows GC content (%), and the y-axis shows mean sequencing depth, calculated in non-overlapping 10-kb windows across the assembly. The single prominent peak (average GC 30.24%) with minimal scattered points indicates low heterozygosity and no substantial contamination. (B) Genome-wide Hi-C interaction heatmap at chromosome scale. Chromosomes (Chr01–Chr12) are clearly delineated along the diagonal, with strong interaction signals within chromosomes (high intensity on the diagonal) and negligible signals between chromosomes (low off-diagonal intensity). This pattern confirms successful chromosome-level scaffolding and high assembly continuity. (C) BUSCO completeness assessment of the final genome assembly using the insecta_odb10 database. The assembly is highly complete, with 93.64% complete single-copy BUSCOs, 0.73% complete duplicated BUSCOs, 3.22% fragmented and 2.41% missing. **Figure S4:** Neighbour-Net phylogenetic network of the *A. conformis* species group inferred from nuclear SNP data. The network was constructed in SplitsTree based on pairwise genetic distances from the whole-genome resequencing SNP dataset. Branch colours indicate species and regions: blue, *A. elongatus* from East Asia; green, *A. conformis* from North America; yellow, *A. nebularis* from North America. **Figure S5:** Principal component analysis (PCA) of nuclear SNP data showing genetic clustering in the *A. conformis* group. Individuals are coloured by species: blue (*A. elongatus*, East Asia), green (*A. conformis*, North America), yellow (*A. nebularis*, North America). PC1 sharply separates East Asian *A. elongatus* from North American taxa, while PC2 distinguishes *A. conformis* (positive scores) from *A. nebularis* (negative scores). **Figure S6:** (A) Schematic diagrams of the three biogeographic scenarios compared using DIYABC-RF for the *A. conformis* species group. Scenario 1 represents an East Asian-source model, in which the East Asian lineage is ancestral to the eastern North American lineage. Scenario 2 represents an eastern North American-source model, in which the direction of source area is reversed. Scenario 3 represents a widespread-ancestor model, in which the East Asian and eastern North American lineages are derived from an older ancestral population. Branch colours indicate population-size parameters as shown in the legend. The vertical axis indicates time before present. EA, East Asia; ENA, eastern North America. (B) Random forest classification votes for the three scenarios. Bars indicate the number of votes assigned to each scenario among 500 random forest trees. DIYABC-RF was used for source-area scenario comparison only; demographic parameter estimates were not used for the main interpretation of divergence times or migration rates. **Figure S7:** Dorsal habitus of representative specimens of the *A. conformis* species group. Specimens are shown in dorsal view: *A. nebularis*, *A. conformis* and *A. elongatus* from left to right. Scale bar = 1 cm. **Figure S8:** Principal component analysis (PCA) of fourteen morphological variables associated with 163 specimens between East Asian *A. elongatus* and North American *A. conformis* and *A. nebularis*. **Figure S9:** Boxplots of fifteen morphological characters comparing East Asian *A. elongatus* with North American *A. conformis* and *A. nebularis*. Traits include abdomen length, body width, pronotum length, sternum 7 length, connexival spine length, head length, head width, fore femur length, fore tibia length, hind tibia length, middle femur length, middle tibia length, body length, hind femur length and first antennal segment/body length. Significance codes above brackets indicate pairwise comparisons: ****, $p < 0.001$; NS, not significant. **Figure S10:** Neighbour-Net phylogenetic network of *A. elongatus* inferred from ddRAD-seq data. The network was constructed in SplitsTree using the Neighbour-Net algorithm based on genetic distances from the ddRAD_95 USNP dataset (3652 unlinked SNPs) for 134 individuals. The reticulate structure highlights conflicting phylogenetic signals among populations, while delimiting four major genetic lineages corresponding to eastern China (EC, red), Japan (JPA, yellow), Taiwan (TW, light blue) and Yunnan (YN, dark blue). **Figure S11:** Discriminant analysis of principal components (DAPC) of *A. elongatus* based on the ddRAD_95 unlinked SNP (USNP) dataset. Scatterplot shows individuals projected onto the first two discriminant functions from the DAPC. Group dispersions are summarised by ellipses. Colours correspond to lineages as shown in the legend (EC, red; JPA, yellow; TW, light blue; YN, dark blue). Insets display eigenvalues of retained principal components and discriminant axes. **Table S1:** Sampling information for ddRAD-seq analyses of *A. elongatus*. The table summarises 134 individuals from 17 populations, reporting sample codes, sample sizes (N), sampling localities, collection dates and geographic coordinates (longitude and latitude in decimal degrees). **Table S2:** Sampling information for whole-genome sequencing (WGS) of the *A. conformis* group. The table lists the 16 individuals selected for WGS, representing *A. elongatus* (East Asia), *A. nebularis* and *A. conformis* (North America), with sample codes, sampling localities, collection dates and geographic coordinates (longitude and latitude in decimal degrees). **Table S3:** Results of reference genome assembly. **Table S4:** The statistics of contig clustering of 12 chromosomes by length. **Table S5:** fastsimcoal2 model-selection results for demographic models among three lineages of the *A. conformis* species group based on whole-genome resequencing data. Model numbers correspond to Figure S1. n indicates the number of free parameters. **Table S6:** Parameter estimates for the best-fitting intercontinental demographic model (M1) of the *A. conformis* species group inferred with fastsimcoal2 based on whole-genome resequencing data. Estimates were obtained assuming diploidy, a 0.5-year generation time and a nuclear mutation rate of 3.5E-9 substitutions per site per generation. Point estimates correspond to the best-fitting run among 50 independent replicate optimizations for the selected model. **Table S7:** Genetic diversity

1365294x, 2026, 16, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70533 by Zhen Ye - Nankai University , Wiley Online Library on [25/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

statistics for 17 populations of *A. elongatus* based on ddRAD-seq data, including observed heterozygosity (Obs.Het), expected heterozygosity (Exp.Het), nucleotide diversity (π) and inbreeding coefficient (F) for individual populations and grouped lineages. **Table S8:** Demographic model selection for the four East Asian lineages of *A. elongatus* inferred with fastsimcoal2 based on ddRAD-seq data (Step I: topology testing; Step II: gene flow testing based on best topology M7). Model numbers correspond to Figure S2. *n* indicates the number of free parameters. deltaL and AIC are reported for each model; models were ranked by AIC. **Table S9:** Parameter estimates for the best-fitting demographic model with gene flow (M23) among four East Asian lineages of *A. elongatus* inferred with fastsimcoal2. All estimates assume diploid cells, a 0.5-year generation time, and a nuclear mutation rate of 3.5E-9 per site per generation. Point estimates are provided with 95% confidence intervals in parentheses.

1365294x, 2026, 16, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70533 by Zhen Ye - Nankai University , Wiley Online Library on [25/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License