scientific reports

Check for updates

OPEN

Three distinct genetic lineages of *Trichonephila clavata* based on mitochondrial *COI* and genome-wide SNPs on the Korean Peninsula

Gyeongmin Kim[1], Eun Hwa Choi[2] & Ui Wook Hwang[2,3,4,5,6]✉

Trichonephila clavata **L. Koch, 1878, an East Asian species renowned for its long-distance ballooning dispersal, has recently drawn attention as an invasive species following its introduction into the southeastern United States. However, few population genetic studies have been undertaken to investigate population genetic diversity and structure which will be helpful for its management. Here, we examined ten populations of *T. clavata* on the Korean Peninsula using mitochondrial *COI* and genome-wide SNP data to broadens the understanding of genetic and demographic processes of the species. Our results revealed: (1) high genetic diversity in *COI* but relatively low diversity in SNPs; (2) the presence of three genetic lineages detected by both markers; (4) population expansion in each *COI* genetic lineage and constant population size in each SNP genetic lineage; and (5) stepwise lineage divergence estimated based on *COI* and lineage divergence with admixture event based on SNPs. These findings suggest that multiple factors, which are ballooning dispersal, demographic dynamics and geological event, may play a pivotal role in shaping population genetic patterns with geographic co-occurrence among the lineages. Consequently, this study provides insights into the genetic architecture of natural populations and offers a valuable baseline for understanding the population genetic pools of invasive populations of this species.**

Keywords *Trichonephila clavata*, Invasive species, Ballooning dispersal, Population genetics, Mitochondrial *COI*, Genome-wide SNPs

The genus *Trichonephila* Dahl, 1911 (Araneae: Nephilidae) is globally distributed and functions as an important predator of diverse invertebrates in tropical and subtropical ecosystems[1,2]. Within this genus, *Trichonephila clavata* L. Koch, 1878 is endemic to East Asia, including South Korea, Japan, China, and Taiwan[3–7]. The species is particularly notable for its extensive dispersal capacity via ballooning, whereby juvenile spiders harness air currents to travel considerable distances—potentially over 300 km from land[5,8–10]. This dispersal strategy not only promotes range expansion but may also homogenize genetic variation across populations[11,12].

In recent years, *T. clavata* has emerged as an invasive species in the southeastern United States, where it was likely introduced from East Asia through human-mediated transport[6,13,14]. Subsequent studies suggest that the species now occupies at least 120,000 km^2, an expansion potentially enhanced by its long-distance dispersal capabilities[13,14]. Its recent establishment outside the native range underscores the importance of understanding the fundamental biology of *T. clavata*[14].

Molecular phylogenetics and population genetics have become indispensable tools in developing strategies for management, prevention, and ecological restoration for diverse taxa which is vulnerable to specific ecological issues such as environmental indicator species and invasive species[15–22]. In particular, population genetic studies—which assess genetic diversity, population structure, and demographic history—can guide proactive invasive species management by minimizing the risk of further introductions and evaluating the efficacy of control measures[16,23]. In particular, research on population dynamics, one of the major subjects of interest

[1]School of Life Sciences, Graduate School, Kyungpook National University, Daegu 41566, South Korea. [2]Department of Biology Education, Teachers College and Institute for Phylogenomics and Evolution, Kyungpook National University, Daegu 41566, South Korea. [3]Department of Advanced Bioconvergence, Kyungpook National University, Daegu 41566, South Korea. [4]Department of Biomedical Convergence Science and Technology, Kyungpook National University, Daegu 41566, South Korea. [5]Institute for Korean Herb-Bio Convergence Promotion, Kyungpook National University, Daegu 41566, South Korea. [6]Phylomics Inc., Daegu 41910, South Korea. ✉email: uwhwang1@gmail.com

within population genetics, provide insights into demographic processes such as rapid expansion of population size, population contraction through bottleneck, which are crucial for predicting a species' invasive potential under environmental changes for long timescales[16,23,24].

Although East Asia is considered the native range of *T. clavata*, few population genetic investigations have been performed in this region[6]. Jung et al. (2006) carried out the only such study, analyzing seven populations from the Korean Peninsula and one from Japan using amplified fragment length polymorphism (AFLP) markers[5]. Their results indicated (1) more pronounced genetic diversity within each population than among the populations and (2) an absence of significant correlation between genetic and geographic distance that among the Korean and Japanese *T. clavata* populations, likely attributable to ballooning. However, because the study utilized a single marker type (AFLP) and did not address population dynamics or historical divergence, its scope was limited. Jung et al. (2006) further underscored the need for broader genetic markers in subsequent investigations, given the constraints of AFLP data[5].

This study aims to provide a comprehensive understanding of the genetic structure of Korean *T. clavata* populations by utilizing both mitochondrial *COI* and genome-wide SNP data. The mitochondrial *COI* has been widely evaluated as suitable genetic markers to reconstruct the population-level phylogenies and investigate the evolutionary history of the particular species due to the rapid mutation rates, while genome-wide SNPs from biparentally-inherited nuclear DNA can reveal fine-scale patterns of population genetic structure, and detected signature of detection[25,26]. With these characteristics of both genetic markers, we investigate population genetic diversity, population structure, demographic history, and time-calibrated phylogenetic relationships. Our sampling includes 284 individuals for *COI* and 166 individuals for SNPs from 26 distinct localities within the Korean Peninsula, surpassing the sample size of Jung et al.[5]. By estimating genetic diversity for each population, examining population dynamics, and inferring divergence times on the Korean Peninsula, we shed new light on historical processes that have shaped population genetic structure and phylogeographic patterning within *T. clavata*.

Accordingly, this study addresses the following main questions:

1. What are the patterns of genetic diversity in *T. clavata* populations on the Korean Peninsula?
2. What is revealed about the population genetic structure from mitochondrial *COI* and SNP data?
3. Did the populations undergo historical demographic expansion, and if so, how did their effective population size fluctuate over historical time?
4. When did the observed divergence events among the genetic lineages occur?

By integrating mitochondrial *COI* and genome-wide SNP data, our work broadens the understanding of genetic and demographic processes in *T. clavata* and provides a basis for enhanced management strategies, both within its native range and in newly invaded habitats.

Results

Genetic diversity and differentiation analysis based on *COI*

A total of 284 T. *clavata* individuals were collected from 26 sites across the Korean Peninsula (Fig. 1 and Supplementary Table S1). Based on geographic proximity and mountain ranges, these individuals were grouped into ten populations for genetic analysis. All individuals were successfully amplified at the mitochondrial *COI* (993 bp) via PCR and sequenced using the Sanger method. In total, 113 haplotypes and 121 polymorphic sites were identified, comprising 57 singleton variable sites and 64 parsimoniously informative sites (Supplementary Table S2). The number of haplotypes per population ranged from 35 (GSD) to 5 (JNN) (Table 1).

To assess genetic diversity, segregation sites (S), haplotype diversity (h), and nucleotide diversity (π_{COI}) were calculated for each population (Table 1). The number of segregation sites ranged from 47 (JBE) to 13 (CNN). The highest haplotype diversity occurred in GSD ($h = 0.963$), whereas JNN had the lowest ($h = 0.731$). Nucleotide diversity was highest in JBE and GGD ($\pi_{COI} = 0.009$) and lowest in JCW ($\pi_{COI} = 0.004$).

Pairwise F_{ST} values based on nuclear divergence were computed to examine genetic differentiation among the ten populations (Table 2a). Significant F_{ST} values based on nuclear divergence ranged from 0.029 (JBE vs. DCB) to 0.347 (JJD vs. JCW). The significant F_{ST} values inferred from haplotype frequency ranged from 0.017 (JBE vs. GSD) to 0.189 (JNN vs. GWE), which were relatively lower than the values based on nuclear divergence (Supplementary Table S3). A Mantel test revealed a low and non-significant correlation between geographic and genetic distances ($r = 0.128$, $p > 0.05$) (Supplementary Figure S1).

Phylogenetic and population genetic analyses of *T. clavata* based on *COI*

Three previously reported *COI* sequences (KJ577713, MW369674, and NC008063) from NCBI GenBank were incorporated into the population genetic analysis, leading to the identification of a novel haplotype (NCCH114) distributed in Korean Peninsula. In total, 114 *COI* haplotypes were used for phylogenetic and population genetic analyses.

Phylogenetic analyses with unrooted ML and BI trees revealed three distinct genetic lineages, Lineages C-I, C-II, and C-III (BP = 98, BPP = 1.00) (Fig. 2a). Most populations contained both major genetic lineages, except JCW and CNN, which comprised only Lineage C-I, and two populations (JBE, DCB) that contained the Lineage C-III exclusively.

These lineages were also evident in the unrooted neighbor-net phylogenetic network and PCoA (Fig. 2b–d). Bootstrap support for the major splits defining the three genetic lineages in unrooted neighbor-net phylogenetic network was moderate to high (88.8% in the split separating Lineage C-II from the other lineages and 75.0% in the split distinguishing Lineage C-I from the others). In the PCoA, the first and second principal components explained 42.51% and 11.41% of the variance, respectively (Fig. 2c). In the TCS network, the Lineage C-III seems

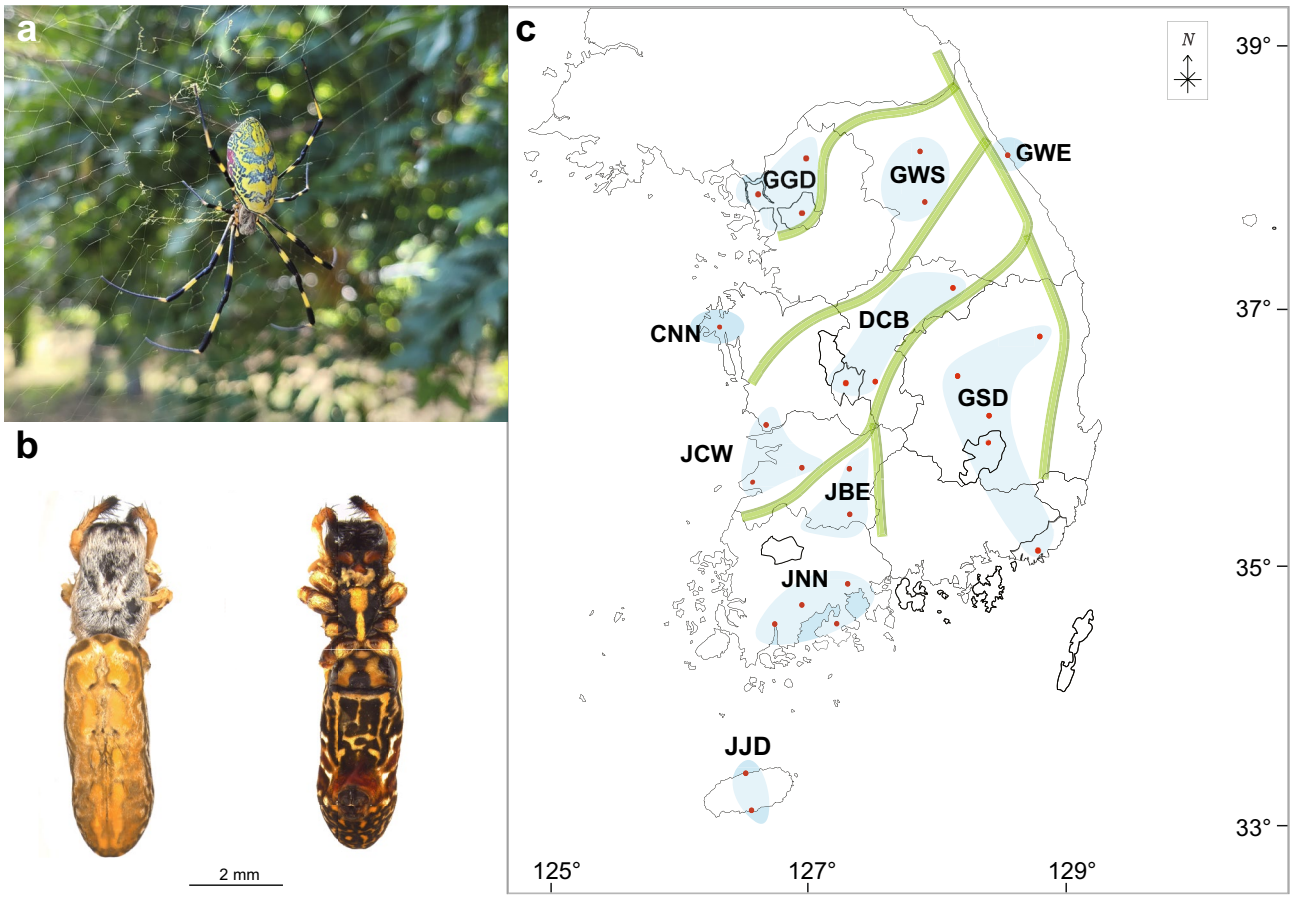

Fig. 1. Photographs of *Trichonephila clavata* and a map showing sampling localities, including population codes used in this study. (**a**) Wild specimen of *T. clavata* collected from Gangneung-si, Gangwon-do, South Korea, photographed by Gyeongmin Kim. (**b**) Distal (left) and ventral (right) views of *T. clavata*, photographed by Gyeongmin Kim. Images were edited using Adobe Illustrator v.22.2 (https://www.adobe.com). (**c**) 26 Collection sites of 284 T. *clavata* individuals, showing the distribution of ten population codes across South Korea. The maps were obtained from a free map-providing site (https://d-maps.com) and modified using Adobe Illustrator v.22.2 (https://www.adobe.com).

difficult to discern the difference with other genetic lineages, which only showed seven to eight mutation steps with Lineage C-I and two to five steps with Lineage C-II. Furthermore, a star-like pattern from several core haplotypes (NCCH01, NCCH10, NCCH25, NCCH27, and NCCH76) was observed in the network (Fig. 2d). AMOVA results supported the presence of the three major lineages: 63.65% of the total genetic variation was attributed to differences among the lineages, 2.85% to variation among populations within each lineage, and 33.50% to variation within populations (Supplementary Table S4).

Genetic distance analysis using the Kimura two-parameter (K2P) model showed that Lineage C-I and Lineage C-II had the largest mean genetic distance [0.015 (0012–0.018)] whereas Lineage C-II and Lineage C-III were most similar [0.011 (0.008–0.014)] (Supplementary Table S5). Within each lineage, Lineage C-I were the most genetically distinct [0.008 (0.007–0.009)], while Lineage C-III had the lowest [0.003 (0.002–0.004)].

Genome assembly, sequencing based on the PacBio platform, and SNP identification

A *T. clavata* reference genome was assembled using SMRT sequencing on the PacBio platform (GenBank accession number: JBMSTR000000000). The final assembly contained 22,392 contigs with an N50 of 121,913 bp, reaching a total size of 1.97 Gb (detailed assembly information will be reported in a future study).

For GBS-based SNP marker selection, 166 individuals (from the total of 284) with high DNA quality and quantity were used. Among 166 individuals, 15 individuals which selected for GBS analysis yielded insufficient DNA quality for SNP detection, and therefore alternative samples were analyzed, leading to a partially different sample composition between the *COI* and SNP datasets. A GBS library was constructed and sequenced to an average depth of 14.78×, producing 1.36 billion raw reads (8.21 million reads per individual). After trimming, 1.27 billion reads remained (7.67 million per individual) with an average mapping rate of 96.05% to the reference genome. Each individual retained a mean of 349,496 unique reads. Using the SEEDERS in-house pipeline and allowing up to 5% missing data, 207,749 SNPs were identified (Data S1). Neutral and positively selected SNPs detection was performed with Bayescan v2.1. Under an FDR threshold of 0.01, 1701 candidate SNPs

Pop	Mitochondrial *COI*					Whole SNPs						
	N_{seq}	N_{Hap}	*S*	*h*	π_{COI}	N	Polymorphic loci (%)	H_O	H_E	π_{SNP}	A_r	F_{IS}
JJD	41	19	38	0.844	0.007	30	29.81	0.273	0.388	0.362	1.350	0.296
JNN	13	5	17	0.731	0.006	13	24.26	0.262	0.382	0.340	1.306	0.314
JBE	42	24	47	0.954	0.009	21	34.50	0.226	0.324	0.299	1.315	0.302
JCW	26	10	16	0.889	0.004	14	13.82	0.234	0.390	0.331	1.273	0.400
DCB	39	23	39	0.935	0.007	16	27.15	0.301	0.341	0.304	1.283	0.117
CNN	8	6	13	0.929	0.005	5	27.53	0.192	0.450	0.291	1.237	0.573
GSD	53	35	43	0.963	0.007	30	23.04	0.205	0.341	0.327	1.349	0.399
GWE	14	8	22	0.890	0.007	8	20.20	0.211	0.365	0.290	1.279	0.422
GWS	16	11	23	0.908	0.008	8	21.01	0.197	0.379	0.210	1.239	0.480
GGD	32	17	43	0.889	0.009	21	33.32	0.208	0.326	0.295	1.321	0.362
Total	284	113	121	0.964	0.008	166	19.05	0.240	0.407	0.390	2.000	0.410

Table 1. Genetic diversity indices based on mitochondrial *COI* partial sequences and whole 6011 SNPs which included both neutral and positively selected SNPs from ten *Trichonephila clavata* populations on the Korean Peninsula. This table includes the abbreviation of the population (Pop), the number of sequences (N_{seq}), the number of haplotypes (N_{Hap}), the number of segregating sites (*S*), haplotype diversity (*h*), nucleotide diversity based on *COI* (π_{COI}), the number of samples (N), observed heterozygosity (H_O), expected heterozygosity (H_E), nucleotide diversity based on SNPs (π_{SNP}), allele richness (A_r), and inbreeding coefficients (F_{IS}).

(a) Mitochondrial *COI*										
	JJD	JNN	JBE	JCW	DCB	CNN	GSD	GWE	GWS	GGD
JJD										
JNN	**0.211*****									
JBE	**0.103*****	0.041								
JCW	**0.347*****	**0.065***	**0.104*****							
DCB	**0.218*****	0.039	**0.029***	**0.038***						
CNN	**0.311*****	0.034	0.078	-0.013	0.026					
GSD	**0.229*****	0.040	**0.047*****	**0.030****	0.007	0.029				
GWE	**0.219*****	0.080	**0.060***	**0.121***	0.015	0.052	**0.051***			
GWS	0.071	0.103	0.027	**0.216*****	**0.079***	**0.155***	**0.098*****	0.066		
GGD	**0.154*****	0.046	0.029	**0.087*****	0.014	0.064	**0.037***	0.012	0.004	
(b) Neutral SNPs										
	JJD	JNN	JBE	JCW	DCB	CNN	GSD	GWE	GWS	GGD
JJD										
JNN	**0.138*****									
JBE	**0.268*****	**0.212*****								
JCW	**0.181*****	0.029	**0.190*****							
DCB	**0.250*****	**0.275*****	0.034	**0.259*****						
CNN	**0.189*****	0.066	**0.193*****	0.034	**0.248*****					
GSD	**0.223*****	**0.117*****	0.032	0.086	0.088	0.083				
GWE	**0.270*****	**0.219*****	0.015	**0.180*****	0.058	0.184	0.030			
GWS	**0.349*****	**0.254*****	0.057	**0.210*****	**0.154*****	**0.223*****	0.051	0.067		
GGD	**0.264*****	**0.165*****	**0.049*****	**0.122*****	**0.108*****	0.112	0.021	0.037	0.043	

Table 2. Pairwise F_{ST} values estimated from (a) mitochondrial *COI* sequences based on nucleotide divergence and (b) 1701 neutral SNPs among ten *Trichonephila clavata* populations on the Korean Peninsula. Significant value: Bold (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$).

were identified as neutrally selected SNPs, and 4310 were identified as positively selected (diversified) SNPs (Supplementary Table S6 and Supplementary Figure S2).

Genetic diversity based on neutral and positively selected SNPs

Using the whole 6011 SNPs which include both neutral and positively selected SNPs, we estimated genetic diversity indices for each population (Table 1). The mean proportion of polymorphic loci was 19.05%, which ranged from 13.82% (JCW) to 34.50% (JBE). The mean observed heterozygosity (H_O) and expected heterozygosity (H_E) were 0.240 and 0.407, respectively. Among populations, DCB showed the highest H_O (0.301), and CNN had the lowest

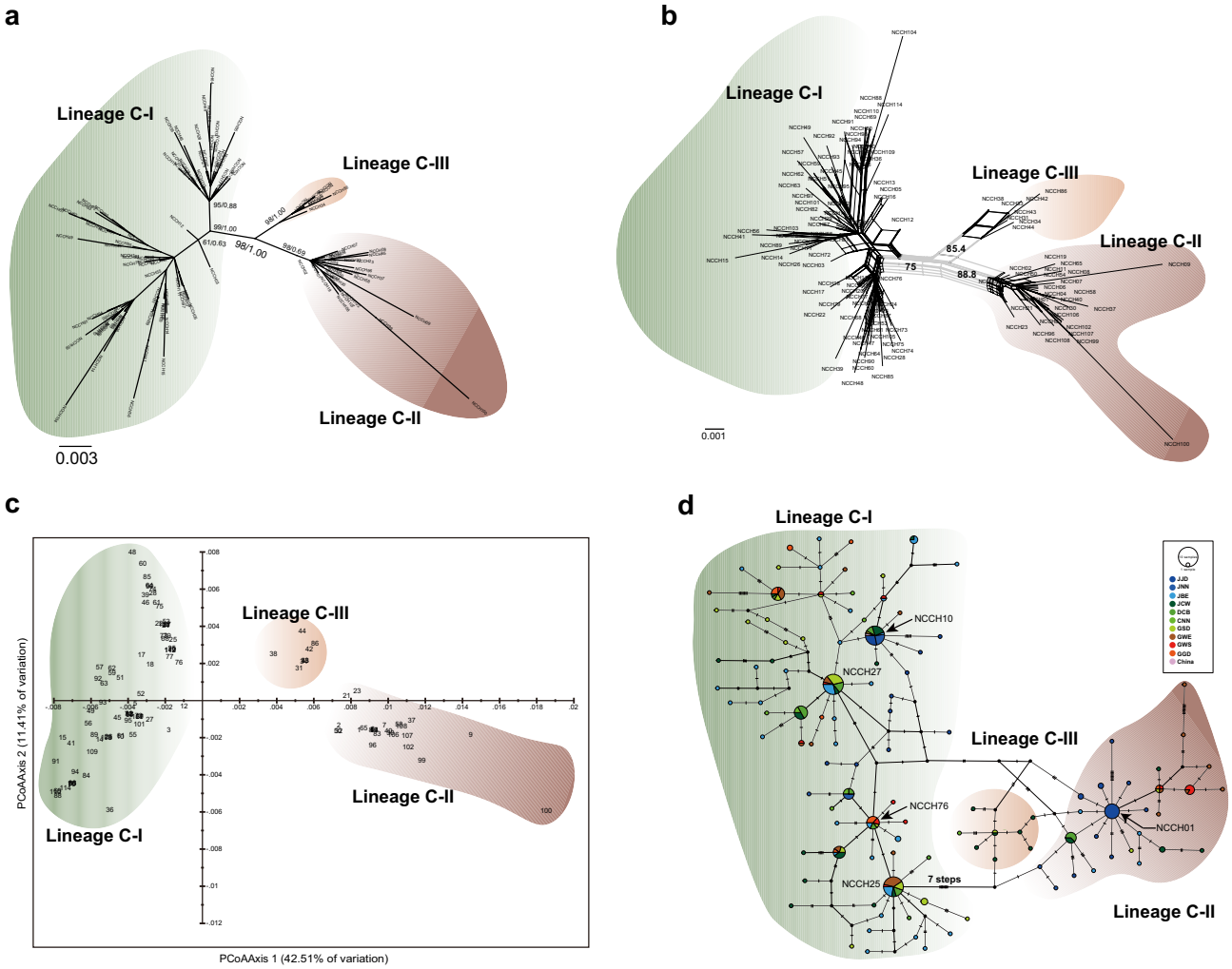

Fig. 2. Results of molecular phylogenetic and population analyses based on 114 *COI* haplotypes from 284 *Trichonephila clavata* individuals, suggesting the presence of three genetic lineages in South Korean populations. (**a**) Unrooted maximum likelihood (ML) phylogenetic tree, showing three distinct genetic lineages (Lineages C-I, C-II and C-III). Node support values are indicated as follows: bootstrap values (ML) and Bayesian posterior probabilities (BI), in order. (**b**) Unrooted Phylogenetic network reconstructed using the Neighbor-Net algorithm without an outgroup. Support values in the location that diverged the three genetic lineages are indicated. (**c**) Two-dimensional PCoA plot visualizing genetic differentiation. The variance explained by the first two axes is as follows: Axis 1 = 42.51%, Axis 2 = 11.41%. (**d**) Haplotype network reconstructed using the TCS algorithm. Circle size represents haplotype frequency.

(0.192). Nucleotide diversity (π_{SNP}) was highest in JJD (0.362) and lowest in GWS (0.210). Mean allelic richness (A_r) ranged from 1.237 (CNN) to 1.350 (JJD). Overall, the inbreeding coefficient (F_{IS}) was positive, indicating potential inbreeding within each population. Among populations, DCB had the lowest ($F_{IS} = 0.117$), and CNN had the highest ($F_{IS} = 0.573$). Hardy–Weinberg equilibrium (HWE) tests conducted within each population indicated that most loci conformed to HWE expectations, showing significant deviations was relatively low (ranging from 5 to 7% across populations).

F_{ST} analysis and Mantel test was conducted by 1701 neutral SNPs and 4310 positively selected SNPs, respectively. The significant pairwise F_{ST} values based on the neutral SNPs ranged from ranged from 0.049 (JBE vs. GGD) to 0.349 (JJD vs. GWS), with numerous comparisons showing significant population differentiation (Table 2b). The result of F_{ST} calculation based on positively selected SNPs ranged from 0.036 (JBE vs. GSD) to 0.467 (JJD vs. GWS) (Supplementary Table S7). A Mantel test revealed significant correlation between genetic and geographic distances ($r = 0.373$, $p < 0.001$ in neutral SNPs and $r = 0.379$, $p < 0.001$ in positively selected SNPs) (Supplementary Figure S3). Furthermore, analysis of molecular variance (one-level AMOVA) revealed a significant genetic differentiation both in neutral and positively selected SNPs ($\Phi_{ST} = 0.157$, $p < 0.001$ in neutral SNPs, and $\Phi_{ST} = 0.217$, $p < 0.001$ in positively selected SNPs) (Supplementary Table S8).

Population structure analysis based on neutral and positively selected SNPs

This study investigated the population genetic structure of neutral SNPs and positively selected SNPs separately and compared the respective results based on the STRUCTURE, unrooted phylogeny, DAPC, and Netview R plot. For STRUCTURE analysis, we removed the SNPs with the strong linkage disequilibrium ($r^2 > 0.2$) using the PLINK v 2.0. As a result, 1455 neutral SNPs and 2558 positively selected SNPs were used for the analysis, respectively. Based on the LD-pruned 1455 neutral SNPs, STRUCTURE analysis indicated that three clusters ($K = 3$) were the optimal number of genetic groups (Supplementary Figure S4a). These clusters were designated as Cluster I (Blue), Cluster II (Green), and Cluster III (Red) (Fig. 3a). Among the 166 individuals analyzed, 79 individuals were categorized as pure group and 87 as admixture group. Clusters I and II contained 73 and 72 individuals, respectively, with all populations represented except DCB in Cluster I and JCW and CNN in Cluster II. Cluster III comprised 21 individuals originating from JJD, JNN, JCW, DCB, CNN, and GSD. Among the admixture group, 68 individuals indicated skewed toward one cluster, However, balanced admixture was revealed from 19 individuals, including 15 from JJD, three from JNN, and one from DCB. All balance-admixed patterns were Clusters I + III or Clusters II + III, and Clusters I + II combination was not observed, suggesting that Cluster III may represent a genetic mosaic derived from the two Clusters.

Based on the no LD-pruned 1701 neutral SNPs, unrooted maximum likelihood (ML) phylogeny revealed three distinct genetic lineages (Lineages S-I, S-II, and S-III), nearly consistent with the clusters indicated in the result of STRUCTURE (Fig. 3b). The individual-level DAPC plot, unlike the phylogeny and STRUCTURE results, identified the five groups. (Fig. 3c and Supplementary Figure S5). Group 1 comprised all populations except JCW and CNN, whereas Group 2 was restricted to JJD, JNN, JCW, CNN, and GSD. Group 3 included all but DCB and GWS, and Group 4 all but JJD and DCB. Finally, Group 5 was composed of JJD, JBE, DCB, and GSD. In the Netview plot, which identified three modules (Modules I, II, and III). These three modules

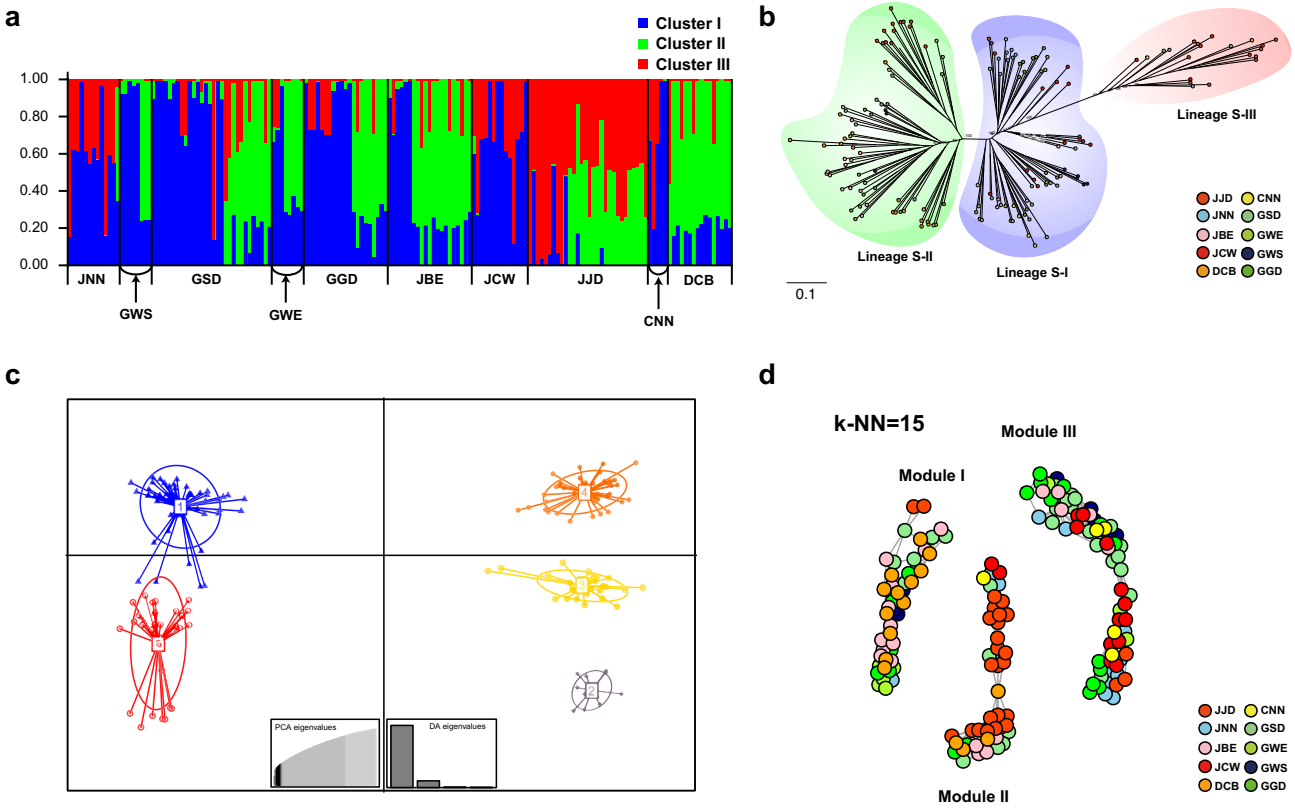

Fig. 3. Results of population genetic analyses based on neutral SNPs from 166 *Trichonephila clavata* individuals. (**a**) STRUCTURE output showing the optimal *K* value ($K = 3$) based on 1455 LD-pruned neutral SNPs using Δ*K* method by STRUCTURE HARVESTER (see Supplementary Figure S4a for details). Colors (blue, green and red) represent genotype composition within each population, and each genotype designate as Cluster I (blue), Cluster II (green), and Cluster III (red). Population labels correspond to those in Supplementary Table S1. (**b**) Unrooted phylogenetic tree reconstructed using maximum likelihood (ML) in the IQ-TREE web server (http://iqtree.cibiv.univie.ac.at/), based on concatenated sequences of 1701 no LD-pruned neutral SNPs, showing the three genetic lineages (Lineages S-I, S-II, and S-III). (**c**) Scatter plot from the discriminant analysis of principal components analysis (DAPC) based on 1701 no LD-pruned neutral SNPs using six clusters suggested by Bayesian information criterion (BIC) in the R package adegenet (see Supplementary Figure S5 for details), which shows five genetic groups (Groups I,II,III,IV, and V). (**d**) Netview R result at *k*-NN = 15 (see Supplementary Figure S6a for details) based on 1701 no LD-pruned neutral SNPs showing three genetic structures (Modules I, II, and III).

were further supported at *k*-NN = 15, which was included in the most appropriate range (10–20) by a *k*-NN selection plot (Fig. 3d and Supplementary Figure S6a). Module 1 consists of all populations except JCW and CNN. Module 2 included all except GWE and GWS, and Module 3 included all except DCB.

To evaluate alternative hypotheses of population structure inferred from these analyses, we performed AMOVA under four scenarios: three clusters from STRUCTURE, three lineages from unrooted phylogeny, five groups from DAPC, and three modules from the Netview plot (Supplementary Table S9). All results of the variation were significantly derived. Across all scenarios, the largest proportion of genetic variation was consistently explained by differences within individuals. The second largest component was variation among groups, with the highest value observed in DAPC (34.76%) and the lowest in STRUCTURE (28.75%). In contrast, variation among populations within groups was greatest under the unrooted phylogeny (5.69%) and lowest under DAPC (2.79%), whereas variation among individuals within populations was highest in STRUCTURE (10.34%) and lowest in DAPC (4.18%). Based on this comparison, we adopted the three lineages recovered from the unrooted phylogeny for demographic analyses, as they provide a more balanced resolution of genetic variation both among and within groups.

The STRUCTURE analysis based on 2558 LD-pruned positively selected SNPs identified two clusters (K = 2) as the optimal number of genetic groups, with all individuals assigned to pure groups (Supplementary Figures S4b and S7a). Cluster I (red) comprised 85 individuals, whereas Cluster II (green) included 81 individuals. All populations were represented in both clusters, except DCB (absent in Cluster I) and JCW and CNN (absent in Cluster II). Compared with the results from neutral SNPs, Cluster III in the neutral dataset was incorporated into Cluster II in the positively selected dataset. In contrast, the phylogeny inferred from 4310 no LD-pruned positively selected SNPs identified three genetic lineages (S-I, S-II, S-III), which corresponded to the lineages based on neutral SNPs (Supplementary Figure S7b). The DAPC plot identified six groups; relative to the neutral SNP results, group 1 and group 5 were further subdivided into groups 1, 5, and 6 in the positively selected dataset (Supplementary Figures S7c and S8). Finally, the Netview plot revealed three modules, consistent with the patterns observed for neutral SNPs (Supplementary Figures S6b and S7d).

AMOVA tests based on positively selected SNPs produced patterns consistent with those from neutral SNPs (Supplementary Table S10). Specifically, (1) within-individual variation accounted for the largest proportion of total variation across all analyses; (2) variation among groups was the second largest component, with the highest proportion in DAPC (41.01%); and (3) variation among populations within groups (3.78%) and among individuals within populations (0.62%) were both lowest in DAPC.

Except for 15 individuals for which both markers could not be obtained, genetic data from *COI* and SNPs were successfully generated for 151 individuals. Among these, 97 individuals showed concordant lineage assignments: Lineage C-I in *COI* corresponded to Lineage S-I in SNPs, and Lineage C-II in *COI* corresponded to Lineage S-II in SNPs. In contrast, 54 individuals exhibited discordant assignments. Of these, 32 involved mismatches between the two major lineages (e.g., Lineage C-I with Lineage S-II or Lineage C-II with Lineage S-I), while the remaining 22 involved either Lineage C-III with Lineage S-I or S-II or Lineage S-III with Lineage C-I or C-II.

Population dynamics and divergence times from mitochondrial *COI* for *T. clavata*

Neutrality tests (Tajima's *D*, Fu's F_S, and R_2) based on *COI* sequences yielded significant results for both the total population and each genetic lineage, rejecting the neutral hypothesis (Table 3). All Tajima's *D* and Fu's F_S values were negative, whereas R_2 values were positive.

Mismatch distribution (MMD) analysis using *COI* data revealed two moderate peaks for the overall *T. clavata* population, although the observed frequency patterns did not closely match the expected distribution under a population expansion model (Supplementary Figure S9a). All genetic lineages showed a unimodal distribution of observed frequency more similar to an expansion model than the result of total population (Fig. 4a). Furthermore, raggedness index (*r*: 0.006 to 0.051) and sums of squared deviations (SSD: 0.007 to 0.010) for each genetic lineage supported these distribution patterns which were not statistically significant ($p > 0.05$), except for the SSD of Lineage C-I.

Bayesian Skyline Plot (BSP) analysis based on *COI* indicated a strong population expansion between 6.5 and 2.0 Ma for the total population (Supplementary Figure S9b). Lineages C-I and C-II showed similar but more gradual patterns, whereas Lineage C-III lacked a clear expansion signal (Fig. 4b). The estimated expansion periods were 5.0–2.0 Ma for Lineage C-I, and 4.5–2.5 Ma for Lineage C-II (Fig. 4b). All BSP analyses had effective sample sizes (ESS) exceeding 10,000.

Divergence times were estimated via reconstruction of ultrametric tree in BEAST 2.7.6, which included *COI* sequences of three outgroups (*T. clavipes*, *T. inaurata*, and *Nephilengys malabarensis*) and two calibration points (Fig. 4c). As a result of Bayes factor (BF) comparison, relaxed lognormal clock model was selected as an better performance for the analysis, which were corresponded to our criteria (2 ln BF > 100 with strict clock model,

	Haplotype no	Tajima's *D*	Fu's *Fs*	R_2
Lineage C-I	79	**−1.895****	**−24.907*****	**0.097****
Lineage C-II	27	**−2.174*****	**−25.540*****	**0.123*****
Lineage C-III	8	**−1.421***	**−5.575***	**0.200***
Total	114	**−1.800****	**−24.375*****	**0.162*****

Table 3. Results of neutrality tests for mitochondrial *COI* sequences of *Trichonephila clavata*. Significant value: Bold ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$).

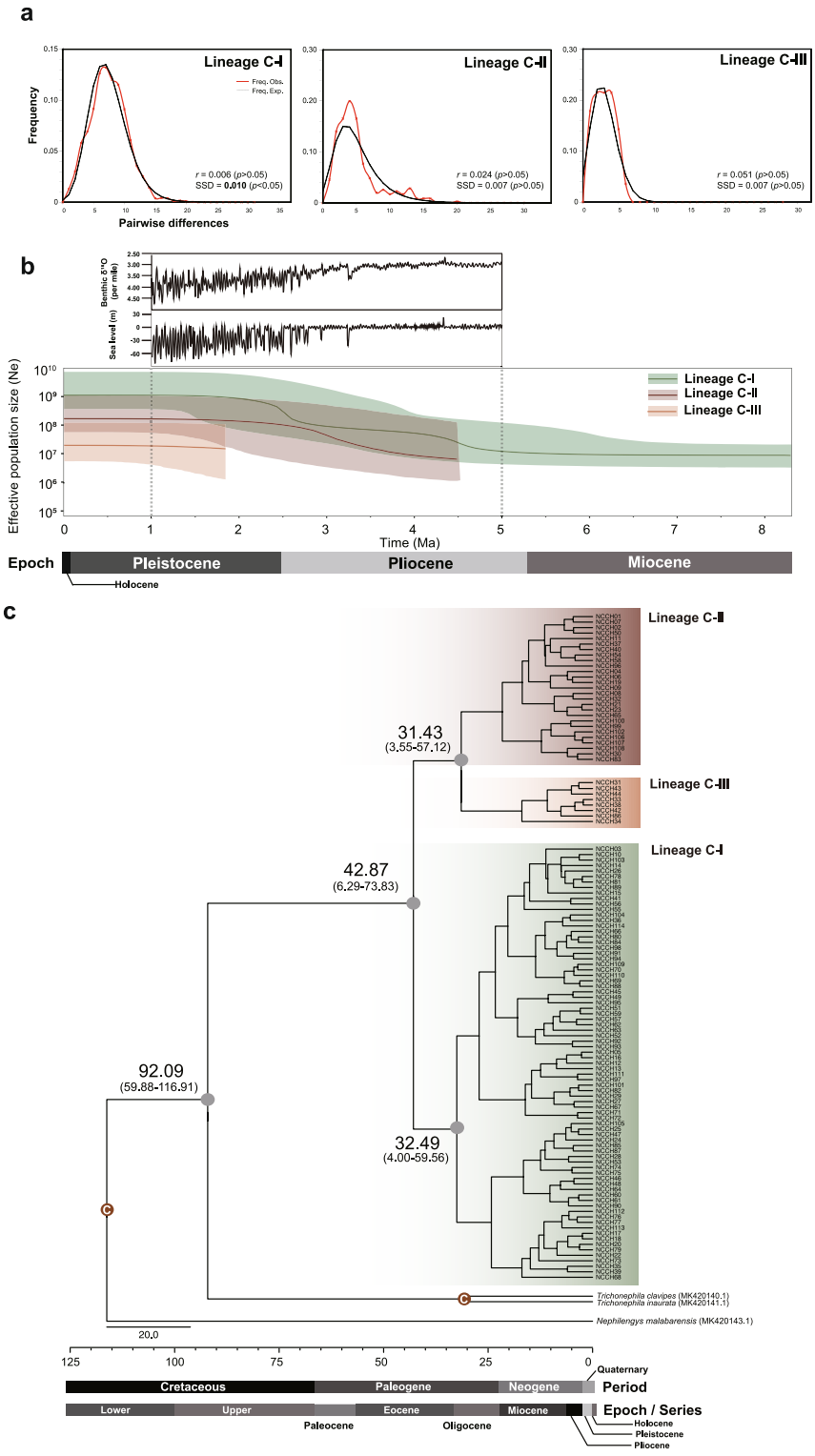

Fig. 4. Demographic history (**a**, **b**) and divergence time estimation (**c**) using mitochondrial *COI* sequences of *Trichonephila clavata* on the Korean Peninsula. (**a**) Mismatch distribution analysis (MMD) estimating pairwise difference frequency. The black dashed line represents the expected distribution under a population expansion model, while the red solid line with red-asterisk-shaped markers represents the observed distribution as the results of each dataset. Raggedness index (*r*) and sum of squared deviations (SSD) are shown at the bottom right of each plot, with statistically significant values in bold. (**b**) Bayesian skyline plot (BSP) analysis illustrating effective population size changes over time implemented in BEAST v.2.6.0. The above graph shows benthic Oxygen isotopic compositions ($\delta^{18}O$) and sea level changes from 5 to 1 Ma, based on Stap et al. (2016). (**c**) Time-calibrated phylogeny based on 114 *COI* haplotypes of *T. clavata* and three outgroups, inferred using a relaxed lognormal clock model with a birth–death process in BEAST v.2.7.6. Brown circles labeled "C" indicate calibration points.

and 2ln(BF) ≈ 7.4 witih relaxed exponential clock) (Supplementary Table S11). The resulting tree (ESS > 200) suggested that *T. clavata* diverged from other *Trichonephila* species in the Late Cretaceous (~ 92.09 Ma; 95% HPD: 59.88–116.91 Ma). Within *T. clavata*, Lineage C-I may have appeared to diverge from the genetic lineage of common ancestor of Lineages C-II and C-III emerged in the middle Eocene, around 42.87 Ma (95% HPD: 6.29–73.83 Ma). After the event, divergence event between the Lineages C-II and C-III may have occurred at approximately 31.43 Ma (95% HPD: 3.55–57.12 Ma). Further diversification within Lineage C-I likely began around 32.49 Ma (95% HPD: 4.00–59.56 Ma).

Inferring demographic history and divergence scenario from neutral SNPs using ABC analysis for *T. clavata*

An Approximate Bayesian Computation (ABC) approach in DIYABC v.2.1.0 using 1455 LD-pruned neutral SNPs to simulate population demographic history with the three scenarios and divergence history with the four scenarios for total lineages and each lineage based on unrooted phylogeny (Supplementary Tables S12–S17 and Supplementary Figures S10, S11). To establish biologically informed priors for ABC analyses, we estimated effective population sizes (Ne) for the whole population and each lineage using the linkage disequilibrium method implemented in NeEstimator v.2.0 (Supplementary Table S12). For the whole population, Ne was estimated at approximately 8.9 (95% CI 8.9–9.0; jackknife CI 7.6–10.3). Among lineages, S-I showed the largest Ne (38.8; 95% CI 38.4–39.2), followed by S-II (24.0; 95% CI 23.8–24.0) and S-III (9.5; 95% CI 9.4–9.7).

Among the three demographic scenarios, Scenario I (Constant population size) received the highest support, with both logistic ($0.00 \le p \le 1.00$) and direct ($0.75 < p \le 0.78$) methods favoring it over alternative models for total lineages (Supplementary Table S14a and Supplementary Figure S10). For each lineage, similar to the result of total population, the constant population size scenario was supported in both logistic ($0.72 < p < 0.99$ in Lineage S-I, $0.99 < p \le 1.00$ in Lineage S-II, and $0.96 < p < 1.00$ in Lineage S-III) and direct ($0.69 < p < 0.71$ in Lineage S-I, $0.61 < p \le 0.64$ in Lineage S-II, and $0.58 < p < 0.63$ in Lineage S-III) methods (Supplementary Table S15a and Supplementary Figure S11). Median value of assumed ancestral population size (N1) was estimate 67.9 in total population, 200 in Lineage S-I, 67.6 in Lineage S-II, and 29.7 in Lineage S-III (Supplementary Tables S14b and S15b).

Among the four divergence history scenarios, Scenario IV (admixture scenario) received the highest support, with both logistic ($0.99 < p \le 1.00$) and direct ($0.52 < p < 0.82$) methods favoring it over alternative models (Fig. 5 and Supplementary Table S17a). Median effective population sizes were estimated at 4,010 for Lineage S-I (N1), 4,960 for Lineage S-II (N2), 1 for Lineage S-III (N3), and 10,000 for ancestral population (Na) (Supplementary Table S17b). The divergence time between Lineages S-I and S-II was estimated as ~ 20,000 generations, and the time (ta) than has been an admixture event between the two parental populations, Lineage S-I and S-II was estimated as ~ 10 generations. Considering a univoltine life cycle of *T. clavata,* it suggests that divergence and admixture event between the two genetic lineages occurred around 20 Ka and 0.01 Ka respectively. The median value of admixture rate relative to Lineage S-I (ra) was estimated as 0.077, and based on the values, the rate relative to Lineage S-II (1-ra) was computed as 0.923.

Discussion

This study performed a comprehensive population genetic analysis of *T. clavata* using both mitochondrial *COI* and genome-wide SNPs, focusing on genetic diversity, differentiation, population structure, population dynamics, and divergence times among ten populations on the Korean Peninsula. Analyses of genetic diversity revealed contrasting patterns between *COI* and SNPs. Mitochondrial *COI* data showed generally high haplotype diversity ($h > 0.8$) in most populations except JNN, and nucleotide diversity (π_{COI}) exceeded 0.005 in all populations except JCW and CNN. In contrast, whole SNP datasets exhibited consistently high, positive inbreeding coefficients across all populations. Observed heterozygosity was notably lower than expected heterozygosity, indicating lower nuclear genetic diversity. Hardy–Weinberg disequilibrium analysis revealed that Korean *T. clavata* populations indicated a deficit of heterozygotes, which is remarkedly lower observed heterozygosity (H_o) than expected heterozygosity (H_E) at all SNP loci, which coincides with the high values of the inbreeding index.

These contrasting genetic patterns likely arise from the differing evolutionary and demographic factors that influence mitochondrial and nuclear markers. Nuclear SNPs are bi-parentally inherited with recombination, which results in larger effective population sizes and therefore a relatively long time to stabilize nucleotide polymorphisms. In contrast, mitochondrial genomes are haploid and maternally inherited, leading to smaller effective population sizes. Consequently, mtDNA is more prone to genetic drift and may accumulate nucleotide polymorphisms over a relatively short timescale[25,26]. Furthermore, infection of bacterial endosymbionts, such as *Wolbachia* strains, may influence the mitochondrial DNA-based genetic diversity[27,28].

The genetic differentiation analysis conducted in this study including F_{ST}, mantel test, and one-level AMOVA analyses revealed significant genetic differentiation among the Korean *N. clavata* populations. In particular, higher F_{ST} values based on nuclear divergence than based on haplotype frequency from *COI* suggested that the genetic differentiation may have occurred by historical processes such as quaternary glaciation events than gene flow. The result of Mantel test in this study were opposed to the result of Jung et al. (2006), which revealed no significant association between genetic and geographic distance ($p > 0.05$) based on AFLP. Among the results, those based on SNPs showed relatively more evident correlation between genetic distance and geographic distance than that of *COI*. Furthermore, the results of one-level AMOVA based on both neutral and positively selected SNPs intensified the significant evidence of genetic differentiation with moderate genetic distances.

Unlike our result, previous studies on *Nephila pilipes* and other ballooning spiders reported weak or no correlation between genetic and geographic distances, as extensive gene flow tends to overcome geographic barriers[11,12,29–31]. In contrast, non-ballooning spiders often exhibit stronger genetic structuring[32–36]. All these earlier findings considered that ballooning dispersal may be affected by these genetic structuring pattern,

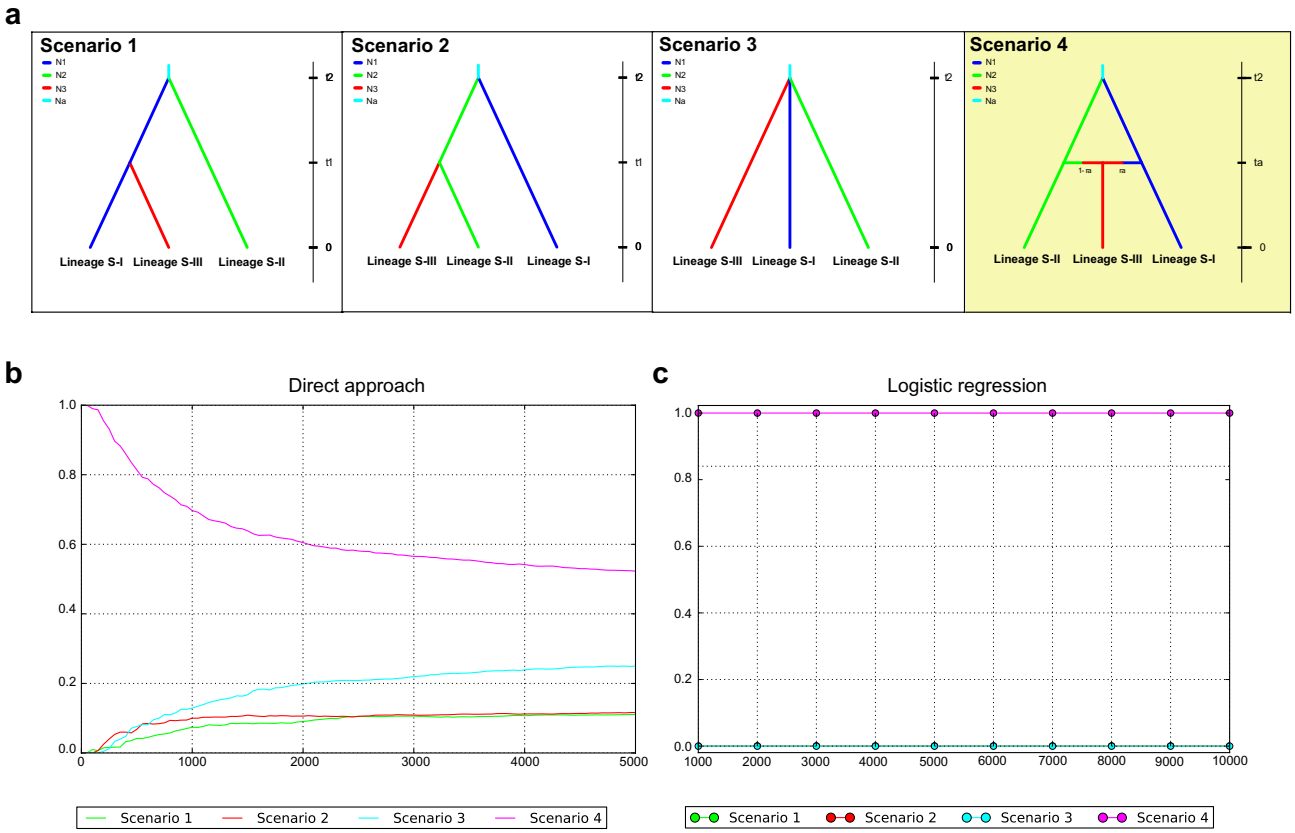

Fig. 5. Possible demographic history scenarios for the three genetic lineages based on 1455 LD-pruned neutral SNPs, analyzed using DIYABC v.2.1.0. (**a**) Four hypothesized population divergence scenarios (see "Divergence Time Estimation and ABC Analysis" in Methods for the detailed description of each scenario). (**b**) Logistic regression results and (**c**) model comparison using the direct approach for each scenario, suggesting Scenario 4 (colored yellowish) received the highest support.

which facilitate gene flow among populations[5]. Beyond innate ability, ballooning is also affected by favorable meteorological conditions such as wind speed, humidity, and temperature, as well as individual behaviors such as seeking suitable new habitats[37,38]. However, more detailed investigations of ballooning as a dispersal mechanism have so far been limited to a small number of spider taxa. Future research should therefore investigate ballooning mechanisms in greater detail across this group and assess their association with population genetic structure[37,38].

Population genetic analyses from both *COI* and SNP data consistently revealed three genetic structures (Lineages C-I, C-II, and C-III in *COI* and Lineages S-I, S-II, and S-III in SNP) with geographic mixing across the Korean Peninsula. Among them, the two main lineages (Lineages C-I and II in *COI*, and Lineages S-I and S-II) were broadly distributed, whereas the remaining lineages (Lineage C-III of *COI* and Lineage S-III of SNP) were genetically restricted, suggesting possible local adaptation.

Lineage C-III of *COI* was restricted to individuals from the JBE and DCB populations. These two populations are found in more geographically isolated and higher-elevation areas, suggesting potential local adaptation to cooler or mountainous environments. Recent physiology study for the species revealed that *T. clavata* is more suitable for cooler environment than its congener[13]. Furthermore, species distribution models (SDM) research highlights the tolerance of low temperature in this species[13,39]. These findings together raise the possibility that the persistence of Lineage C-III represents an adaptation to montane or cooler habitats, although the specific mechanisms by which elevation-related stress drives such adaptation remain to be clarified.

Interestingly, individuals showing discordance between mitochondrial (*COI*) and nuclear (SNP) lineage assignments were also concentrated in the JBE (10 individuals) and DCB (12 individuals) populations, where all cases involved Lineage C-I (*COI*) and Lineage S-II (SNP). This geographic concentration may reflect localized processes such as mitochondrial introgression, potentially driven by environmental conditions favoring the persistence of the Lineage C-I mitochondrial genotype over that of Lineage C-II in these regions.

In contrast to Lineages C-III of *COI*, Lineage S-III of SNPs was primarily found in JJD, JNN, JCW, CNN, and GSD. Its distribution was relatively southern compared to Lineages S-I and II and was absent from more northern populations such as GWE, GWS, and GGD. A southern-specific genetic group was also reported by Jung et al. (2006) (marked as Group 2 in that reference) based on AFLP data[5]. These findings may suggest that the southern region may have served as a historical refugium or center of lineage persistence during climatic fluctuations.

Overall population dynamic analyses based on *COI* consistently supported gradual population expansion in Lineages C-I, C-II considering the results of all neutrality tests, MMD, and BSP. Furthermore, in the TCS network, the presence of multiple star-like patterns from various core haplotypes suggest the expansion of population size with multiple mutational pathways and potentially linked to the interglacial colonization event[40,41].

When compared with sea-level and marine isotope data, population expansions coincided with glacial–interglacial cycles, paralleling patterns reported in other spiders[42–45]. The population expansions of Lineage C-I and Lineage C-II were mainly dated to the Pliocene (5.33–2.58 Ma), a period characterized by generally warmer conditions than today[46]. Toward the end of the epoch (~ 2.6 Ma), extensive ice-sheet growth in the Northern Hemisphere marked the onset of pronounced glacial stages[46]. Population expansion in both lineages appears to have begun during the warm phase, but the growth of effective population sizes gradually slowed with the progression of glacial conditions.

In contrast to the BSP results from *COI*, the ABC analysis based on LD-pruned neutral SNPs supported a scenario of constant population size across both the total population and individual lineages. Such discrepancies may reflect differences in inheritance mode, effective population size of each markers as well as the analytical methods employed[25,26,42].

A *COI*-based, time-calibrated phylogeny placed the divergence of total *T. clavata* lineages from other *Trichonephila* species in the Upper Cretaceous (92.09 Ma). Within the species, major splits among the lineages occurred in the Eocene. The divergence between Lineage C-I and the common ancestral lineage of Lineages C-II and C-III likely occurred after the Early Eocene Climatic Optimum (EECO), a period of elevated atmospheric pCO_2 which suggested globally warmer temperatures[47,48]. Subsequent divergence and diversification—between Lineages C-II and C-III and within Lineage C-I—was estimated for the early Oligocene (~ 31–32 Ma), following the Eocene–Oligocene Transition (EOT). This transition marked the onset of major continental glaciation and a dramatic shift from a warm climate to an icehouse world around 34 Ma[49–52]. Due to this glacial event, the period is recognized as the most significant biotic reorganization since the end-Cretaceous extinction[53]. The lineage divergence (between Lineages C-II and C-III) and diversification (within Lineage C-I) of *T. clavata* may have been driven in part by these events by causing small isolated refugia and droven genetic divergence of fragmentation[54,55].

Beyond climatic influences, subsequent geological changes may also have shaped diversification for each lineage. During the Oligocene to Miocene (ca. 30–15 Ma), the East Sea and the main Korean mountain ranges, including the Baekdudaegan, were sequentially formed through a tertiary tilted flexural process[56,57]. These mountainous areas of the Korean Peninsula may have played a critical role as refugia in buffering populations against subsequent extreme climate change by geological event. Previous studies have identified this mountain system as an important refugium for temperate plants and insects in northeastern Asia[58–60]. For widespread lineages (C-I and C-II from *COI*; S-I and S-II from SNPs), the areas may have provided multiple refugia across several regions. After severe glacial period, the warmer Miocene (20–5 Ma) with favorable climatic conditions may have enabled lineage expansion through ballooning as co-occurrence of multiple mitochondrial lineages[61].

Similar to the *COI* results, ABC analysis of LD-pruned neutral SNPs supported the initial divergence between Lineages S-I and S-II. In contrast to the strictly stepwise divergence pattern inferred from the *COI* time-calibrated tree, however, the SNP-based analysis favored a scenario in which Lineage S-III originated through admixture between parental lineages. Similar to the inconsistent pattern of genetic diversity and demographic estimation between the two markers, this discrepancy also likely arises from both marker-specific properties and methodological assumptions. In particular, mitochondrial *COI*-based phylogeny analysis captures a single genealogical history and enforces strictly bifurcating trees, whereas nuclear SNP-based ABC analysis explicitly evaluates alternative demographic scenarios—including divergence with gene flow and admixture—across genome-wide loci[62,63].

The divergence between Lineages S-I and S-II was further estimated at ~ 20,000 years ago, which is more recent than *COI*-based estimates. Although absolute values remain uncertain due to limited calibration of substitution rates in this species and methodological differences among markers, this discrepancy may reflect incomplete lineage sorting in nuclear genomes. Importantly, the estimated timing coincides with glacial–interglacial cycles of the Pleistocene, suggesting that climatic oscillations and associated geographic isolation likely contributed to the genetic divergence between Lineages S-I and S-II[64,65]. The repeated glacial–interglacial cycles likely drove alternating episodes of re-isolation and expansion as well as the two nuclear SNP-based lineages. During these climatic oscillations, ballooning dispersal may have continuously facilitated co-occurrence of the SNP-based lineages, maintaining relatively stable effective population sizes as inferred from SNP-based analyses. Finally, secondary contact among co-occurring lineages would have provided opportunities for interbreeding, potentially giving rise to the admixed lineage (S-III)[66,67].

Our study revealed two major genetic lineages within the native range (Lineages C-I and C-II from *COI*; Lineages S-I and S-II from SNPs) as well as an admixed lineage (Lineage S-III), providing a valuable baseline for future population genetic comparisons. The contrasting patterns between mitochondrial and nuclear markers highlight the importance of integrating both marker types when evaluating species-wide genetic structure. Moreover, linking genetic structure with past geological and climatic events provide insight into the evolutionary processes shaping *T. clavata* populations in East Asia. Such baseline knowledge will also facilitate future assessments of genetic diversity and range dynamics in non-native populations as the species continues to expand beyond its native range.

Methods

Sample collection and mitochondrial *COI* sequencing

All necessary permits for sample collection of this study were obtained from National Institute of Biological Resources, Ministry of Environment, South Korea. A total of 284 *Trichonephila clavata* specimens were collected

from 26 locations on the Korean Peninsula (Fig. 1 and Supplementary Table S1). Immediately after collection, each specimen was preserved in 95% ethanol and stored at − 20 °C until DNA extraction. Genomic DNA (gDNA) was extracted from the walking legs using the DNeasy Blood and Tissue Kit (Qiagen, United States). DNA quality and concentration were determined using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, United States) and further confirmed through 1% agarose gel electrophoresis.

Species identification was based on morphological traits (e.g., abdomen shape, web structure)[4]. Molecular data generated by amplifying and sequencing the mitochondrial *COI* gene via PCR (polymerase chain reaction) followed by Sanger sequencing. The *COI* region was amplified using the primers C1-J-1718 (5′-GGA GGA TTT GGA AAT TGA TTA GTT CC-3′) and C1-N-2776 (5′-GGA TAA TCA GAA TAT CGT CGA GG-3′), which amplified length was expected as 1059 bp[68,69]. Each 50 µL PCR reaction contained 50–100 ng of gDNA, 10 mM of each dNTP, 10 pM of each primer, and 0.25 U of Taq DNA polymerase (Solgent, South Korea). The thermal cycling protocol included an initial denaturation at 95 °C for 3 min, followed by 35 cycles of 95 °C for 20 s, 52–56 °C for 40 s, and 72 °C for 1 min, concluding with a final extension at 72 °C for 5 min. One microliter of each PCR product was electrophoresed on a 1% agarose gel containing eco-dye and visualized under UV light. Verified amplicons were purified using a QIAquick PCR Purification Kit (Qiagen, Germany) and sequenced on an ABI Prism 3730 DNA sequencer (PerkinElmer, United States) with a BigDye Terminator Sequencing Kit (PerkinElmer, United States).

Genetic diversity and differentiation analysis based on mitochondrial *COI* sequences

Raw *COI* sequences were aligned in Clustal X2 with default settings, and ambiguous 5′ and 3′ regions were trimmed in BioEdit v.7.2.5, yielding a final alignment of 993 bp[70,71]. The resulting sequences were compared using BLAST the NCBI GenBank database, which confirmed all samples as *T. clavata*. Haplotypes, singleton sites, and parsimoniously informative sites were identified in DnaSP v.6.11[72]. All newly identified haplotypes were deposited in NCBI GenBank: accession numbers PV413245–PV413357.

Based on geographic proximity and mountain ranges, these individuals were grouped into ten populations for genetic analysis (Supplementary Table S1). To evaluate genetic diversity within each population, three diversity indices—segregating sites (S), haplotype diversity (h), and nucleotide diversity (π)—were calculated in DnaSP v.6.11[72]. Genetic differentiation was assessed by pairwise fixation indices (F_{ST}) Genetic differentiation was assessed by pairwise fixation indices (F_{ST}) based on the genetic distance according to both nucleotide divergence and haplotype frequency, using Arlequin v.3.5[73]. The correlation between genetic and geographic distances was further examined using a Mantel test in Alleles in Space v.1.0[74].

Population genetic structure analysis based on mitochondrial *COI* sequences

Three *COI* sequences of *T. clavata*, KJ577713 from China and MW369674 and NC008063 from South Korea obtained from NCBI GenBank were included for comparative analyses. Maximum Likelihood (ML) and Bayesian Inference (BI) methods were employed to reconstruct phylogenetic relationships among *COI* haplotypes, with *T. clavata* haplotypes unrooted for tree topology.

The unrooted ML phylogeny was built using the IQ-TREE web server (http://iqtree.cibiv.univie.ac.at/), applying the HKY + F + I + G4 model, identified by IQ-TREE's ModelFinder based on the Bayesian Information Criterion (BIC)[75]. Branch support was evaluated via 1000 bootstrap replicates. BI analysis was performed with MrBayes v.3.2 under the same substitution model, running 10,000,000 MCMC generations with sampling every 1000 steps, discarding the first 15% of samples as burn-in[76].

Additionally, three analytical approaches were used to examine the population structure among Korean *T. clavata* populations:

1. An unrooted neighbor-net phylogenetic network in SplitsTree to present the phylogenetic information using the distance matrix that can guide more detailed analysis[77].
2. A haplotype network based on the TCS algorithm in PopART to examine the genealogical relationships among haplotypes[78].
3. A Principal Coordinate Analysis (PCoA) in Darwin v.6.0.9 to visualize the genetic structure among populations[79].

Genetic distances among and within *COI* lineages identified from these analyses were computed using MEGA v.11 based on the Kimura 2-Parameter (K2P) model which is widely used in DNA barcoding studies of mitochondrial *COI* sequences, and more realistic estimate of genetic divergence avoiding the overparameterization[80,81]. Based on the confirmed genetic lineage through the three analyses, Analysis of Molecular Variance (AMOVA) in Arlequin v.3.5 was conducted to partition the genetic variance within and among populations[73].

Reference genome construction for genome-wide SNP discovery

As of 2021, no complete *T. clavata* genome was available for population SNP analyses. Consequently, we assembled a whole-genome reference from a single individual collected in Daegu, Korea. Genomic DNA was extracted using a DNeasy Blood and Tissue Kit (Qiagen, Germany), and quantity/quality were assessed via NanoDrop 2000 (Thermo Fisher Scientific, United States). The extracted gDNA was submitted to DNA Link (Seoul, South Korea) for library preparation and sequencing. Approximately 12 µg of DNA was used to prepare a single-molecule real-time (SMRT) bell library with the SMRTbell® Express Template Preparation Kit v.2 (PacBio, USA). Library concentration and fragment size were measured using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, United States) and the DNA 12,000 Kit (Agilent, Germany). The BluePippin™ system (Sage Science, USA) removed small fragments, and DNA polymerase was bound to the complex using the Sequel II Binding Kit 2.0 (Pacific Biosciences, USA). Sequencing was performed on a Sequel II system (Pacific

Biosciences, USA), generating HiFi reads, which were assessed by LongQC[82]. Genome assembly was conducted using the FALCON-Unzip algorithm[83].

GBS sequencing pre-processing and SNP marker detection

Among the 284 individuals examined for *COI,* 166 were subjected to Genotyping-By-Sequencing (GBS) at SEEDERS (Daejeon, South Korea). Fifteen individuals which selected for GBS analysis yielded insufficient DNA quality for SNP detection, and therefore alternative samples were analyzed, leading to a partially different sample composition between the *COI* and SNP datasets. Sample quality and quantity were checked by NanoDrop 2000 and 1% agarose gels. Each sample was digested with ApeKI (New England Biolabs, United States), followed by adaptor ligation, pooling, purification, and multiplex PCR amplification. The resulting GBS library was sequenced on the Illumina HiSeq X10 platform (2 × 151 bp).

Raw reads were demultiplexed and trimmed with CutAdapt v.1.8.3 and SolexaQA v.1.13[84,85]. Filtered reads were aligned to the newly assembled reference genome using BWA v.0.7.17-r1188[86]. SAMtools v.0.1.16 was used to identify SNPs under the following parameters: minimum mapping quality (−Q) = 30, minimum mapping quality for gaps (−q = 15), minimum read depth (−d) = 3, maximum read depth (−D) = 160, indel proximity filters (–w = 15, –W = 15), and min indel score for nearby SNP filtering (−G = 30)[87]. Total 3,400,493 SNP matrix was yielded before the additional filtering processes. After filtering erroneous loci based on SEEDERS' in-house pipelines, 207,499 high-quality SNPs were retained for population analyses[88].

Bayescan v.2.1 was employed to identify SNPs potentially under selection or neutral mutation, using an input file generated via PGDSPIDER v.2.1.1.5[89,90]. The run parameters included: –n 1000, –thin 10, –nbp 20, –pilot 5000, –burn 50,000. Loci were classified by plotting F_{ST} coefficients against the log10 of q-values. Two indicators were used to select the SNP loci: (1) false discovery rate (FDR), which represents the expected proportion of false positives among SNP markers that is statistically inferred to be under selection despite being in reality a neutral loci, and (2) the alpha values, a locus-specific component shared across populations that may serve as an indicator of selection[89]. Finally, 1701 neutral and 4310 positively selected loci were retained for subsequent population genetic analyses.

Genetic diversity and population genetic structure analysis based on neutral and positively selected SNPs

Whole 6011 SNP dataset which included both 1701 neutral and 4310 positively selected SNPs was used to estimaete genetic diversity indices—percentage of polymorphic loci, observed heterozygosity (H_O), expected heterozygosity (H_E), nucleotide diversity (π)—were estimated with Arlequin v.3.5[73]. Allelic richness (A_r) was computed by the R package HIERFSTAT[91]. The inbreeding coefficient (F_{IS}) was calculated as $F_{IS} = 1 - H_O/H_E$[92]. Pairwise F_{ST} values were estimated in Arlequin v.3.5, and p-values for the indicator were adjusted for multiple comparisons using the Benjamini–Yekutieli FDR procedure ($q = 0.05$) following Narum[93]. The Mantel test in the R package VEGAN (9999 permutations) assessed the correlation between genetic and geographic distances[73,94]. Furthermore, one-level AMOVA analysis was accessed to examine the overall genetic differentiation, as implemented in Arlequin v.3.5[73].

To investigate the population genetic structures based on SNPs data, we employed STRUCTURE v.2.3.4.[95] Previously to conduct the analysis, each dataset of SNPs was pruned for Linkage disequilibrium (LD) in PLINK v. 2.0 using the command "-indep-pairwise". The parameters were as follows: a window size of 50 SNPs, 5 SNPs to shift the window at each step, and an r^2 threshold of 0.2. This procedure removed SNPs in strong LD ($r^2 > 0.2$), ensuring that the assumption of SNPs independence was sufficiently met.[96,97] Using the LD-pruned neutral or positively selected SNPs, the parameters of STRUCTURE v.2.3.4 was set for $K = 1$–8, each with 100 replicates[98]. The optimal K was determined using STRUCTURE HARVESTER based on the ΔK method[99]. To assign individuals to each genetic group based on the STRUCTURE results, we applied two categories: "pure group" and "admixture group." Individuals with an ancestry coefficient greater than 0.75 were classified as the pure group, while those with a coefficient below 0.75 were assigned to the admixture group. For individuals in the admixture group, a median threshold (≥ 0.50), which reflects more than half of an individual's inferred ancestry is derived from a single cluster, providing a straightforward and biologically reasonable criterion for assignment while minimizing misclassification, was additionally used to assign them to specific clusters following the majority assignment approach commonly used in population genetic studies[95,99]. To facilitate interpretation of admixture patterns, we categorized individuals into two groups, balanced admixture and skewed admixture. Balanced admixture was defined as cases where the two major clusters contributed approximately equally (0.40–0.60), consistent with recent hybrid ancestry. Skewed admixture referred to individuals in which one cluster predominated (0.60–0.75) but a secondary cluster still made a notable contribution (0.25–0.40).

In addition to the STRUCTURE, we conducted the unrooted phylogenetic analysis, Netview plot, and Discriminant Analysis of Principal Components (DAPC) to infer the population genetic structure. Unlike the STRUCTURE, these analyses are free of assumption about HWE or LD, and we conducted the analyses using the no LD-pruned 1701 neutral and 4310 positively selected SNPs respectively, which maximize the variant and avoid underestimation of genetic diversity due to SNP removal.

An unrooted ML phylogeny was reconstructed in IQ-TREE under the TPM2u + F + ASC + G4 model[75]. Input files for phylogeny were converted from VCF to FASTA format via PGDSPIDER v.2.1.1.5[90]. Genotypic relationships among individuals were also visualized with R package NetView[100]. The optimal range of k-nearest neighbor (k-NN) value for NetView analysis was selected by plotting each k-NN value against the number of communities based on the Fast-Greedy, Infomap, and Walktrap algorithm[100,101]. Lastly, Discriminant Analysis of Principal Components (DAPC) was performed using the R package ADEGENET to identify genetic clusters (K) based on the Bayesian Information Criterion (BIC)[102]. The best number of principal components was chosen based on a cross-validation test implemented in the *xvalDapc* function within ADEGENET[102].

Population dynamics and divergence time estimation based on mitochondrial *COI*

Three demographic methods—neutrality tests, mismatch distribution analysis (MMD), and Bayesian skyline plots (BSP)—were used to explore population size fluctuations within genetic lineages. Three neutrality tests, which are Tajima's *D*, Fu's *Fs* (by Arlequin v.3.5), and R_2 (by DnaSP v.6.11) examined deviations from neutral expectations[72,103–105]. MMD was conducted in DnaSP v.6.11 to estimate the frequency of pairwise differences and compare them with the expected patterns under a population expansion model[72,106]. The raggedness index (*r*) and sum of squared deviations (SSD), which evaluate the statistical significance of MMD, were assessed in Arlequin v.3.5[76,107].

BSP analyses were implemented in BEAST v.2.6.0 to infer historical population size changes[108]. A mutation rate 4.974×10^{-4} substitutions/site/Ma (95% HPD: 1.541×10^{-4} substitutions/site/Ma – 1.392×10^{-3} substitutions/site/Ma) was adopted for the BSP analyses, as estimated from the time-calibrated phylogeny based on the specified speciation events. (Please read the next paragraph). Substitution models were determined by jModelTest v.2.1.10, resulting in HKY + F + I + G4 for Lineage C-I, HKY + F + I for Lineage C-II, HKY + F for Lineage C-III, and GTR + F + I + G for the total dataset[109]. Markov Chain Monte Carlo runs continued for adequate sampling until reaching an effective sample size > 200, and results were visualized in Tracer v.1.7[110].

Divergence times among the identified lineages were first estimated via a time-calibrated *COI* phylogeny in BEAST v.2.7.6[108]. The HKY + F + I + G4 model was selected by jModelTest v.2.1.10[109]. Two calibration points derived from Scharff et al. (2020) were used with *Nephilengys malabarensis* (MK420140), *Trichonephila clavipes* (MK420141), and *Trichonephila inaurata* (MK420143) as outgroups: (1) 116 Ma for the *N. malabarensis*–*Trichonephila* split and (2) 30 Ma for the divergence among *T. clavipes* and *T. inaurata*[111,112]. We evaluated the appropriate molecular clock model among the three uncalibrated clock models (strict, lognormal relaxed, and exponential) by Bayes factor (BF) comparison, which the values of 2ln (BF) > 10 are regarded as firm evidence for a given model to be more likely than the other[113]. For the comparison, we computed the marginal likelihoods (In P) using Tracer v.1.6, and calculated the BF through the formula $(2\ln(BF) = 2 \times \Delta \ln P)$[114]. With the evaluation of the appropriate clock model, we employed Birth–Death model as a speciation model to better represent realistic speciation processes. A total of 20,000,000 MCMC generations were run, and ESS values were confirmed using Tracer v.1.7[110]. After discarding the first 25% of trees as burn-in, a maximum clade credibility tree was generated in TreeAnnotator v.2.6.0 and visualized in FigTree v.1.4.2[115,116].

Reconstructing demographic history and divergence patterns based on neutral SNPs

We implemented Approximate Bayesian Computation (ABC) analysis using 1455 LD-pruned neutral SNPs in DIYABC v.2.1.0 to compare alternative demographic and divergence scenarios[63]. For demographic history, we tested three scenarios for the total three lineages and each lineage (Lineages S-I, S-II, and S-III) of *T. clavata*, respectively: constant population size (Scenario 1), population expansion (Scenario 2), and population contraction through bottleneck (Scenario 3). To set biologically informed priors for effective population size (Ne), we estimated Ne for the total lineages and each lineage using NeEstimator v.2.0, based on the same set of SNPs. Input files were generated with PGDSpider v.2.1.1.5, and estimates were obtained using the linkage disequilibrium method under random mating, with a minimum allele frequency threshold of 0.05[90,117]. Point estimates and confidence intervals guided the prior ranges of Ne, ensuring parameters reflected empirical data. However, as expansion or bottleneck time (t1) was unknown, we applied a log-uniform distribution with a wide interval (see Supplementary Table S13).

For divergence history, we designed four scenarios to test alternative lineage relationships: (1) stepwise divergence with Lineages S-I and S-II splitting at t2, followed by divergence of S-III from S-I at t1; (2) stepwise divergence with S-I and S-II splitting at t2, followed by divergence of S-III from S-II at t1; (3) simultaneous divergence of all three lineages; and (4) an admixture scenario in which S-I and S-II served as parental populations and S-III as the admixed lineage, inferred from population structure results. Bottleneck events were not considered further, as they were unsupported by demographic ABC analyses. Ne priors from NeEstimator were also used in these divergence scenarios, while divergence times (t1, t2) and admixture rates were assigned log-uniform distributions with broad intervals (see Supplementary Table S16).

For all scenarios, summary statistics included mean genetic diversity, mean F_{ST}, and Nei's distance, with admixture proportions additionally considered in Scenario 4 of divergence history scenarios. Each scenario was simulated 1×10^7 times to generate reference tables. Prior to final inference, we evaluated the suitability of scenarios and prior distributions by conducting PCA on 1000 simulated datasets per scenario and projecting the observed dataset onto the same space[63]. Posterior probabilities and parameter estimates were then obtained using both direct and logistic regression approaches to compute the posterior probability of the best plausible scenario and parameter estimations in the scenarios. We finally selected the most plausible scenario as the one with the highest posterior probability in both the two approach.

We only employed SNPs data not *COI* data for the approximate Bayesian computation (ABC) analysis of alternative demographic diversification scenarios. This determination was due to the fact that ABC inference requires a large number of independent and unlinked loci to obtain sufficient statistical power and to minimize stochastic variance associated with single-locus genealogies. Because mitochondrial *COI* represents a single, non-recombining locus with a smaller effective population size and potentially different evolutionary dynamics (e.g., maternal inheritance, introgression, or selection), it is not suitable for robust ABC-based scenario testing.

Data availability

The sequence data have been deposited to the NCBI GenBank database (https://www.ncbi.nlm.nih.gov/genbank/) under the accession numbers PV413245-PV413357 of *COI* haplotypes obtained in this study, and reference genome of *T. clavata* under accession number JBMSTR000000000. The preprocessed dataset can be found in the

Supplementary Data as follows: Supplementary Data S1, 207,749 SNP genotype information in variant calling format for 166 T. *clavata* individuals.

Received: 24 June 2025; Accepted: 24 October 2025
Published online: 25 November 2025

References

1. Kuntner, M., Zhang, S., Gregorič, M. & Li, D. *Nephila* female gigantism attained through post-maturity molting. *J. Arachnol.* **40**, 345–347 (2012).
2. Kuntner, M. et al. Golden orbweavers ignore biological rules: phylogenomic and comparative analyses unravel a complex evolution of sexual size dimorphism. *Syst. Biol.* **68**, 555–572 (2019).
3. Kim, J. P., Kim, S. D. & Lee, Y. B. A revisional study of the Korean spiders, family Tetragnathidae Menge, 1866 (Arachnida: Araneae). *Korean Arach.* **15**, 41–100 (1999).
4. Namkung, J. *The Spiders of Korea* (Kyo-Hak, 2002).
5. Jung, J., Lee, J., Kim, J. & Kim, W. Genetic variations of the golden orb-web spider *Nephila clavata* (Araneae: Tetragnathidae) in Korea, using AFLP markers. *Korean J. Genet.* **28**, 325 (2006).
6. Hoebeke, E. R., Huffmaster, W. & Freeman, B. J. *Nephila clavata* L koch, the Joro spider of East Asia, newly recorded from North America (Araneae: Nephilidae). *PeerJ* **3**, e763 (2015).
7. Choi, E. H. & Hwang, U. W. Complete mitochondrial genome of a golden orb-web spider *Trichonephila clavata* (Chelicerata, Arachnida) from South Korea. *Mitochondrial DNA B: Resour.* **8**, 723–725 (2023).
8. Gertsch, W. J. *American Spiders* 2nd edn. (Van Nostrand Reinhold, 1979).
9. Decae, A. E. *Dispersal: Ballooning and Other Mechanisms in Ecophysiology of Spiders* (ed. Aitchison-Benell, C. W.) 348–356 (Springer, 1987).
10. Bonte, D., Vandenbroecke, N., Lens, L. & Maelfait, J. P. Low propensity for aerial dispersal in specialist spiders from fragmented landscapes. *Proc. R. Soc. Lond. B Biol. Sci.* **270**, 1601–1607 (2003).
11. Ramirez, M. G. & Haakonsen, K. E. Gene flow among habitat patches on a fragmented landscape in the spider *Argiope trifasciata* (Araneae: Araneidae). *Heredity* **83**, 580–585 (1999).
12. Dimassi, N., Khadra, Y. B., Othmen, A. B., Ezzine, I. K. & Said, K. High genetic diversity vs. low genetic and morphological differentiation of *Argiope trifasciata* (Araneae, Araneidae) in Tunisia. *Syst. Biodivers.* **15**, 1–15 (2017).
13. Davis, A. K. & Frick, B. L. Physiological evaluation of newly invasive jorō spiders (*Trichonephila clavata*) in the southeastern USA compared to their naturalized cousin, *Trichonephila clavipes*. *Physiol. Entomol.* **47**, 170–175 (2022).
14. Chuang, A., Deitsch, J. F., Nelsen, D. R., Sitvarin, M. I. & Coyle, D. R. The Jorō spider (*Trichonephila clavata*) in the southeastern US: An opportunity for research and a call for reasonable journalism. *Biol. Invasions* **25**, 17–26 (2023).
15. Bailly, X., Migeon, A. & Navajas, M. Analysis of microsatellite variation in the spider mite pest *Tetranychus turkestani* (Acari: Tetranychidae) reveals population genetic structure and raises questions about related ecological factors. *Biol. J. Linn. Soc.* **82**, 69–78 (2004).
16. Le Roux, J. & Wieczorek, A. M. Molecular systematics and population genetics of biological invasions: Towards a better understanding of invasive species management. *Ann. Appl. Biol.* **154**, 1–17 (2009).
17. Ryu, J. S. & Hwang, U. W. Complete mitochondrial genome of the longtail tadpole shrimp *Triops longicaudatus* (Crustacea, Branchiopoda, Notostraca). *Mitochondrial DNA* **21**, 170–172 (2010).
18. Baek, S. Y. et al. Complete mitochondrial genomes of *Carcinoscorpius rotundicauda* and *Tachypleus tridentatus* (Xiphosura, Arthropoda) and implications for chelicerate phylogenetic studies. *Int. J. Biol. Sci.* **10**, 479–489 (2014).
19. Choi, E. H. et al. Molecular phylogenetic, population genetic and demographic studies of *Nodularia douglasiae* and *Nodularia breviconcha* based on *CO1* and *16S rRNA*. *Sci. Rep.* **10**, 16572 (2020).
20. Shin, C. R. et al. Characterization of metapopulation of *Ellobium chinense* through Pleistocene expansions and four covariate *COI* guanine-hotspots linked to G-quadruplex conformation. *Sci. Rep.* **11**, 12239 (2021).
21. Hong, S., Park, B., Kim, G., Choi, E. H. & Hwang, U. W. Possible species discrimination of a blotched nerite *Nerita albicilla* with their distribution pattern and demographic history in the Indo-Pacific. *Sci. Rep.* **13**, 4545 (2023).
22. Kim, G. et al. Molecular population genetics and phylogeographic studies of *Ligia exotica* and *Ligia cinerascens* in East Asia. *Front. Mar. Sci.* **10**, 1260171 (2023).
23. Sakai, A. K. et al. The population biology of invasive species. *Annu. Rev. Ecol. Syst.* **32**, 305–332 (2001).
24. Faillace, C. A. & Morin, P. J. Evolution alters the consequences of invasions in experimental communities. *Nat. Ecol. Evol.* **1**(1), 0013 (2016).
25. Lansman, R. A., Avise, J. C. & Huettel, M. D. Critical experimental test of the possibility of" paternal leakage" of mitochondrial DNA. *Proc. Natl. Acad. Sci.* **80**, 1969–1971 (1983).
26. Allio, R., Donega, S., Galtier, N. & Nabholz, B. Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Mol. Biol. Evol.* **34**, 2762–2772 (2017).
27. Konecka, E. & Szymkowiak, P. Wolbachia supergroup A in *Enoplognatha latimana* (Araneae: Theridiidae) in Poland as an example of possible horizontal transfer of bacteria. *Sci. Rep.* **14**, 7486 (2024).
28. Russell, J. E. et al. Mitochondrial and *Wolbachia* phylogenetics of the introduced Jorō spider, *Trichonephila clavata* (Araneae: Araneidae) in North America. *PeerJ* **13**, e19952 (2025).
29. Ramirez, M. G. & Fandino, L. B. Genetic variability and gene flow in *Metepeira ventura* (Araneae, Araneidae). *J. Arachnol.* **24**, 1–8 (1996).
30. Schäfer, M. A., Hille, A. & Uhl, G. B. Geographical patterns of genetic subdivision in the cellar spider *Pholcus phalangioides* (Araneae). *Heredity* **86**, 94–102 (2001).
31. Lee, J. W., Jiang, L., Su, Y. C. & Tso, I. M. Is central mountain range a geographic barrier to the giant wood spider *Nephila pilipes* (Araneae: Tetragnathidae) in Taiwan? A population genetic approach. *Zool. Stud.* **43**, 112–122 (2004).
32. Johannesen, J., Baumann, T., Seitz, A. & Veith, M. The significance of relatedness and gene flow on population genetic structure in the subsocial spider *Eresus cinnaberinus* (Araneae: Eresidae). *Biol. J. Linn. Soc.* **63**, 81–98 (1998).
33. Johannesen, J. & Veith, M. Population history of *Eresus cinnaberinus* (Araneae: Eresidae) colour variants at putative species transition. *Heredity* **87**, 114–124 (2001).
34. Johannesen, J., Lubin, Y., Laufs, T. & Seitz, A. Dispersal history of a spider (*Stegodyphus lineatus*) across contiguous deserts: Vicariance and range expansion. *Biol. J. Linn. Soc.* **84**, 739–754 (2005).
35. Duncan, S. I., Riechert, S. E., Fitzpatrick, B. M. & Fordyce, J. A. Relatedness and genetic structure in a socially polymorphic population of the spider *Anelosimus studiosus*. *Mol. Ecol.* **19**, 810–818 (2005).
36. Tanikawa, A., Shinkai, A., Tatsuta, H. & Miyashita, T. Highly diversified population structure of the spider *Lycosa ishikariana* inhabiting sandy beach habitats. *Conserv. Genet.* **19**, 255–263 (2018).
37. Lee, V. M., Kuntner, M. & Li, D. Ballooning behavior in the golden orbweb spider *Nephila pilipes* (Araneae: Nephilidae). *Front. Ecol. Evol.* **3**, 2 (2015).

38. Wolz, M. et al. Dispersal and life-history traits in a spider with rapid range expansion. *Mov. Ecol.* **8**, 2 (2020).
39. Giulian, J., Jones, T. C. & Moore, D. Assessing the potential invasive range of *Trichonephila clavata* using species distribution models. *J. Asia-Pac. Biodivers.* **17**, 490–496 (2024).
40. Slatkin, M. & Hudson, R. R. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**, 555–562 (1991).
41. Jenkins, T. L., Castilho, R. & Stevens, J. R. Meta-analysis of northeast Atlantic marine taxa shows contrasting phylogeographic patterns following post-LGM expansions. *PeerJ* **6**, e5684 (2018).
42. Vawter, L. & Brown, W. M. Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. *Science* **234**, 194–196 (1986).
43. Ayoub, N. A. & Riechert, S. E. Molecular evidence for Pleistocene glacial cycles driving diversification of a North American desert spider. *Agelenopsis aperta. Mol. Ecol.* **13**, 3453–3465 (2004).
44. Gobbi, M., Fontaneto, D. & De Bernardi, F. Influence of climate changes on animal communities in space and time: The case of spider assemblages along an alpine glacier foreland. *Glob. Chang. Biol.* **12**, 1985–1992 (2006).
45. Krehenwinkel, H. & Tautz, D. Northern range expansion of European populations of the wasp spider *Argiope bruennichi* is associated with global warming–correlated genetic admixture and population-specific temperature adaptations. *Mol. Ecol.* **22**, 2232–2248 (2013).
46. McClymont, E. L. et al. Climate evolution through the onset and intensification of Northern Hemisphere glaciation. *Rev. Geophys.* **61**, e2022RG000793 (2023).
47. Bae, C. J., Hong, H. & Zhang, Z. Cenozoic climate change in eastern Asia: Part II. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **512**, 1–5 (2018).
48. Song, B. et al. Qaidam Basin paleosols reflect climate and weathering intensity on the northeastern Tibetan Plateau during the Early Eocene Climatic Optimum. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **512**, 6–22 (2018).
49. Zachos, J. C., Dickens, G. R. & Zeebe, R. E. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **451**, 279–283 (2008).
50. Houben, A. J., van Mourik, C. A., Montanari, A., Coccioni, R. & Brinkhuis, H. The Eocene-Oligocene transition: Changes in sea level, temperature or both?. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **335**, 75–83 (2012).
51. Sun, J. et al. Synchronous turnover of flora, fauna and climate at the Eocene-Oligocene Boundary in Asia. *Sci. Rep.* **4**, 7463 (2014).
52. Ye, C. et al. Chlorite chemical composition change in response to the Eocene-Oligocene climate transition on the northeastern Tibetan Plateau. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **512**, 23–32 (2018).
53. Kraatz, B. P. & Geisler, J. H. Eocene-Oligocene transition in Central Asia and its effects on mammalian evolution. *Geology* **38**, 111–114 (2010).
54. Trewick, S. A. & Wallis, G. P. Bridging the "beech-gap": New Zealand invertebrate phylogeography implicates Pleistocene glaciation and Pliocene isolation. *Evolution* **55**, 2170–2180 (2001).
55. Greaves, S. N., Chapple, D. G., Daugherty, C. H., Gleeson, D. M. & Ritchie, P. A. Genetic divergences pre-date Pleistocene glacial cycles in the New Zealand speckled skink, *Oligosoma infrapunctatum*. *J. Biogeogr.* **35**, 853–864 (2008).
56. Jin, P. S. & Ill, S. Discussions on the distribution and genesis of mountain ranges in the Korean Peninsula (I): The identification mountain ranges using a dEM and reconsideration of current issues on mountain range maps. *J. Korean Geogr. Soc.* **40**, 126–152 (2005).
57. Shin, J. & Hwang, S. A critical review on setting up the concept, timing and mechanism of tertiary tilted flexural mode of the Korean Peninsula: A new hypothesis derived from plate tectonics. *J. Korean Geogr. Soc.* **49**, 200–220 (2014).
58. Chung, M. Y., López-Pujol, J. & Chung, M. G. The role of the Baekudaegan (Korean Peninsula) as a major glacial refugium for plant species: A priority for conservation. *Biol. Conserv.* **206**, 236–248 (2017).
59. Chung, M. Y. et al. The Korean Baekdudaegan Mountains: A glacial refugium and a biodiversity hotspot that needs to be conserved. *Front. Genet.* **9**, 489 (2018).
60. Shin, S., Kim, J. H., Dang, J. H., Seo, I. S. & Lee, B. Y. Elevational distribution ranges of vascular plant species in the Baekdudaegan mountain range. *South Korea. J. Ecol. Environ.* **45**, 1–10 (2021).
61. Steinthorsdottir, M. et al. The Miocene: The future of the past. *Paleoceanogr. Paleoclimatol.* **36**, e20200PA04037 (2021).
62. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
63. Collin, F.-D. et al. Extending approximate Bayesian computation with Supervised Machine Learning to infer demographic history from genetic polymorphisms using DIYABC Random Forest. *Mol. Ecol. Resour.* **21**, 2598–2613. https://doi.org/10.1111/1755-0998.13413 (2021).
64. Waelbroeck, C. et al. Sea-level and deep-water temperature changes derived from benthic foraminifera isotopic records. *Quat. Sci. Rev.* **21**, 295–305 (2002).
65. Spratt, R. M. & Lisiecki, L. E. A late Pleistocene sea level stack. *Clim. Past Discuss.* **11**, 3699–3728 (2015).
66. Krosch, M. N., Baker, A. M., Mather, P. B. & Cranston, P. S. Comparison of intraspecific genetic structure among related chironomids (Diptera) from New Zealand and Patagonia: Disparity between potential and realized dispersal. *Freshw. Sci.* **31**, 1105–1120 (2012).
67. Krehenwinkel, H. et al. A phylogeographical survey of a highly dispersive spider reveals eastern Asia as a major glacial refugium for Palaearctic fauna. *J. Biogeogr.* **43**(8), 1583–1594 (2016).
68. Simon, C. et al. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* **87**, 651–701 (1994).
69. Hedin, M. C. & Maddison, W. P. A combined molecular approach to phylogeny of the jumping spider subfamily Dendryphantinae (Araneae: Salticidae). *Mol. Phylogenet. Evol.* **18**, 386–403 (2001).
70. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98 (1999).
71. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. 2007. *Bioinformatics* **23**, 2947–2948 (2007).
72. Rozas, J. et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017).
73. Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
74. Miller, M. P. Alleles In Space (AIS): Computer software for the joint analysis of interindividual spatial and genetic information. *J. Hered.* **96**, 722–724 (2005).
75. Trifinopoulos, J., Nguyen, L. T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**, W232–W235 (2016).
76. Ronquist, F. et al. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
77. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
78. Leigh, J. W. & Bryant, D. popart: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116 (2015).
79. Perrier, X. & Jacquemoud-Collet, J. P. DARwin software, available at: http://darwin.cirad.fr/Darwin (2006).
80. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120. https://doi.org/10.1007/BF01731581 (1980).

81. Tamura, K., Stecher, G. & Kumar, S. Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027. https://doi.org/10.1093/molbev/msab120 (2021).
82. Fukasawa, Y., Ermini, L., Wang, H., Carty, K. & Cheung, M. S. LongQC: A quality control tool for third generation sequencing long read data. *G3: Genes Genomes Genet.* **10**, 1193–1196 (2020).
83. Chin, C. S. et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050–1054 (2016).
84. Cox, M. P., Peterson, D. A. & Biggs, P. J. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinform.* **11**, 1–6 (2010).
85. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
86. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv:1303.3997, 1–3 (2013).
87. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
88. Kim, J. E., Oh, S. K., Lee, J. H., Lee, B. M. & Jo, S. H. Genome-wide SNP calling using next generation sequencing data in tomato. *Mol. Cells.* **37**, 36–42 (2014).
89. Foll, M. BayeScan v2.1 user manual. *Ecology* **20**, 1450–1462 (2012).
90. Lischer, H. E. & Excoffier, L. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**, 298–299 (2012).
91. Goudet, J. Hierfstat, a package for R to compute and test hierarchical *F*-statistics. *Mol. Ecol. Notes.* **5**, 184–186 (2005).
92. Li, C. C. & Horvitz, D. G. Some methods of estimating the inbreeding coefficient. *Am. J. Hum. Genet.* **5**, 107 (1953).
93. Narum, S. R. Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conserv. Genet.* **7**, 783–787 (2006).
94. Oksanen, J. et al. The vegan package. *Com. Ecol. Package.* **10**, 719 (2007).
95. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genet.* **155**, 945–959 (2000).
96. Purcell, S. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**(3), 559–575 (2007).
97. Chang, C. C. et al. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
98. Earl, D. A. & VonHoldt, B. M. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361 (2012).
99. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
100. Steinig, E. J., Neuditschko, M., Khatkar, M. S., Raadsma, H. W. & Zenger, K. R. netview p: A network visualization tool to unravel complex population structure using genome-wide SNPs. *Mol. Ecol. Resour.* **160**, 216–227 (2016).
101. Neuditschko, M., Khatkarm, M. & Raadsma, H. Fine scale population structure of global cattle breeds using dense haplotype data. *PLoS ONE* **7**, e48375 (2012).
102. Jombart, T. & Ahmed, I. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071. https://doi.org/10.1093/bioinformatics/btr521 (2011).
103. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
104. Fu, Y. X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925 (1997).
105. Ramos-Onsins, S. E. & Rozas, J. Statistical properties of new neutrality tests against population growth. *Mol. Biol. Evol.* **19**, 2092–2100 (2002).
106. Rogers, A. R. & Harpending, H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9**, 552–569 (1992).
107. Harpending, H. C. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum. Biol.* **66**, 591–600 (1994).
108. Bouckaert, R. et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, 1–28. https://doi.org/10.1371/journal.pcbi.1006650 (2019).
109. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods.* **9**, 772. https://doi.org/10.1038/nmeth.2109 (2012).
110. Rambaut, A. et al. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
111. Scharff, N. et al. Phylogeny of the orb-weaving spider family Araneidae (Araneae: Araneoidea). *Cladistics* **36**, 1–21 (2020).
112. Gatesy, J., Hayashi, C., Motriuk, D., Woods, J. & Lewis, R. Extreme diversity, conservation, and convergence of spider silk fibroin sequences. *Science* **291**, 2603–2605 (2001).
113. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
114. Rambaut, A., Drummond, A. J. & Suchard, M. Tracer v. 1. 6. http://tree.bio.ed.ac.uk/software/tracer/. (2013).
115. Rambaut, A. FigTree v.1.4. Available at: http://tree.bio.ed.ac.uk/sofware/figtree (2012).
116. Rambaut, A. & Drummond, A. J. TreeAnnotator v.2.3.0. http://beast2.org, (2014).
117. Do, C. et al. NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Mol. Ecol. Resour.* **14**, 209–214 (2014).

Acknowledgements

We thank the Bia Park and Cho Rong Shin for collecting samples. Furthermore, we used ChatGPT, a language model based on GPT-4-turbo developed by OpenAI (accessed March 2025) for editing the manuscript with clear and accurate English expression.

Author contributions

U.W.H. conceived the study. G.K, E.H.C, designed and set up the experiment, and performed the sample collection. G.K, E.H.C performed the experiment. G.K analyzed the data, checked the validation of the results, wrote the original draft of the manuscript, and visualize the result for the manuscript. All authors reviewed the manuscript.

Funding

We gratefully acknowledge support from the National Institute of Biological Resources (NIBR) under the Ministry of Environment (MOE), Republic of Korea (Grant Nos. NIBR202104202, NIBR202206203), and from the National Research Foundation of Korea (NRF) under the Ministry of Science and ICT (MSIT), Republic of Korea (Grant No. RS-2025–00561309).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-25890-9.

Correspondence and requests for materials should be addressed to U.W.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.