Genome-wide screening discerns incomplete lineage sorting from gene flow among wolf spiders from Madeira

Alba Enguídanos García[1], Luis C. Crespo[2], Miquel Arnedo[1], and Vanina Tonzo[3]

[1]Universitat de Barcelona Facultat de Biologia
[2]University of the Azores - Angra do Heroísmo Campus
[3]CEFE

December 25, 2024

Abstract

Genomic data offers valuable insights into population history and species divergence, but interpreting complex evolutionary processes remains challenging, particularly in cases of recent divergence and ancestral polymorphism. This study addresses the taxonomic and evolutionary complexities of two endemic Hogna spider species from Madeira, H. maderiana from Porto Santo and H. insularum from Madeira, Desertas, Bugio, and Porto Santo, which exhibit mitochondrial gene tree discordance and ambiguous morphological boundaries. Using ddRADseq genomic data, population genomic analyses, and coalescent-based demographic analyses, we aim to determine whether these two nominal species represent a case of early divergence with unsorted molecular markers due to incomplete lineage sorting (ILS), if they are hybridising within the contact zone on the island of Porto Santo, or if they represent a single species exhibiting extreme morphological polymorphism. Our genetic structure analyses suggested three potential genetic clusters, one for each nominal species and one compatible with hybridisation between the two species on Porto Santo Island. However, demographic modelling and D-statistics rejected gene flow, instead supporting the existence of a third independent lineage in Porto Santo. The lack of genetic separation between these lineages likely reflects short recent divergence and unsorted ancestral polymorphisms. Our findings highlight the challenges of inferring hybridisation solely based on population structure analyses, which may lead to an overestimation of gene flow. This study highlights the importance of integrating demographic modelling and genetic data to resolve complex evolutionary histories and emphasizes the need for careful interpretation of genomic data to avoid misattributing gene flow.

Genome-wide screening discerns incomplete lineage sorting from gene flow among wolf spiders from Madeira

Alba Enguídanos García[1,2], Luís C. Crespo[3], Miquel Arnedo[1,2], Vanina Tonzo[4]

1. Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Universitat de Barcelona, Avinguda Diagonal 643, 08028 Barcelona, Spain
2. Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona (UB), Barcelona, Spain.
3. cE3c – Centre for Ecology, Evolution and Environmental Changes/Azorean Biodiversity Group, CHANGE – Global Change and Sustainability Institute, University of the Azores, Rua Capitão João d´Ávila, Pico da Urze, 9700-042 Angra do Heroísmo, Azores, Portugal
4. CEFE, Université Montpellier, CNRS, EPHE, IRD, Montpellier, 34293, France.

corresponding author: albaengarcia@gmail.com

Running title: Genomic insights into Madeira wolf spiders

Abstract

Genomic data offers valuable insights into population history and species divergence, but interpreting complex evolutionary processes remains challenging, particularly in cases of recent divergence and ancestral polymorphism. This study addresses the taxonomic and evolutionary complexities of two endemic *Hogna* spider species from Madeira, *H. maderiana* from Porto Santo and*H. insularum* from Madeira, Desertas, Bugio, and Porto Santo, which exhibit mitochondrial gene tree discordance and ambiguous morphological boundaries. Using ddRADseq genomic data, population genomic analyses, and coalescent-based demographic analyses, we aim to determine whether these two nominal species represent a case of early divergence with unsorted molecular markers due to incomplete lineage sorting (ILS), if they are hybridising within the contact zone on the island of Porto Santo, or if they represent a single species exhibiting extreme morphological polymorphism. Our genetic structure analyses suggested three potential genetic clusters, one for each nominal species and one compatible with hybridisation between the two species on Porto Santo Island. However, demographic modelling and D-statistics rejected gene flow, instead supporting the existence of a third independent lineage in Porto Santo. The lack of genetic separation between these lineages likely reflects short recent divergence and unsorted ancestral polymorphisms. Our findings highlight the challenges of inferring hybridisation solely based on population structure analyses, which may lead to an overestimation of gene flow. This study highlights the importance of integrating demographic modelling and genetic data to resolve complex evolutionary histories and emphasizes the need for careful interpretation of genomic data to avoid misattributing gene flow.

Keywords

Genomic data, hybridisation, phylogenetic discordance, species boundaries, island evolution

Introduction

Understanding the mechanisms driving the diversity of life on Earth is a fundamental question in biology, with speciation being the critical driving process (Hernández-Hernández et al., 2021; Jia et al., 2021; Seehausen et al., 2014). Speciation is not a single event but a continuum, ranging from minimally differentiated to entirely distinct species. This continuum is often reflected in the genetic diversity within and between populations, which can be further modulated by processes such as incomplete lineage sorting (ILS) and hybridisation. Systematic studies integrating phenotypic and molecular data are essential for placing species within an evolutionary framework. However, the gradual divergence of species through time and processes such as ILS or gene flow can result in genetic and phenotypic variation gradients that challenge the reconstruction of accurate phylogenies and blur species boundaries.

Along with evolutionary processes, methodological and analytical decisions (e.g., data type and quality and probabilistic framework) may also contribute to phylogenetic uncertainty and fail to establish species limits (Steenwyk et al., 2023). When referring to genomic data, one of the main challenges in unravelling recently diverged evolutionary lineages is distinguishing similarity due to gene flow from that of shared ancestry. Detecting hybrid zones, especially in geographically constrained areas like islands, often leads to debates about whether observed genetic patterns reflect ongoing gene flow or result from historical separation followed by secondary contact (Gompert & Buerkle, 2016). Overestimating gene flow can lead to erroneous conclusions regarding the limits and evolutionary relationships among species, whereas underestimating it can conceal the role of hybridisation in species diversification. This issue is further complicated in cases of ILS, where ancestral genetic polymorphisms persist across multiple species, leading to phylogenetic discordance among loci (He et al., 2023; Hernández-Gutiérrez et al., 2022; Herrig et al., 2024). Detecting and correctly interpreting these signals is vital for reconstructing and understanding the Tree of Life, as well as for assessing biodiversity and conservation priorities.

Spiders are among the most diverse animals on Earth and play a fundamental role as generalist predators in shaping terrestrial ecosystems (Nyffeler & Birkhofer, 2017; Wheeler et al., 2017). As in many other organisms, hybridisation has been documented across multiple species, highlighting its potential role in driving diversification and evolutionary processes. For instance, hybrid zones have been identified among funnel weaver spiders as the *Eratigena atrica* (C.L. Koch, 1843) group in Europe (Croucher et al., 2007;

Oxford & Bolzern, 2018); in deeply divergent lineages of the red devil spider genus *Harpactocrates* in the Pyrenees (Bidegaray-Batista et al., 2016), or in Iberian representatives of the cobweb spider genus *Theridion (Domènech et al., 2020)* . Conversely, recent research has demonstrated that multiple evolutionary processes, such as ancestral hybridisation events and ILS, have contributed to the emergence of different ecomorphs in the Hawaiian spiny-leg *Tetragnatha* spiders (Cerca, Cotoras, Santander, et al., 2023).

The wolf spider genus *Hogna* in the Galapagos Islands exemplifies the significance of gene flow in evolutionary diversification, demonstrating that high levels of intra-island introgression have likely led to the parallel evolution of habitat specialisation across the archipelago (De Busschere et al., 2015). The worldwide distributed, species-rich genus *Hogna* Simon, 1885, includes 236 species (World Spider Catalog, 2024) of medium- to large-sized spiders. This richness is the result of the usage of the genus as a dump basket for Lycosinae species with conserved morphology, mostly because species identification has been plagued by the absence of distinctive diagnostic characters and low-quality illustration (Logunov, 2020), thus resulting in a polyphyletic genus (Crespo et al., 2022; Piacentini & Ramírez, 2019). The genus has managed to disperse and diversify in several archipelagos worldwide, including the Mediterranean, the Caribbean, the Atlantic, and the Pacific (Gloor et al., 2017). Seven endemic species of the wolf spider genus *Hogna* have been documented in the northeast Atlantic archipelago of Madeira, namely *Hogna maderiana* (Walckenaer, 1837), *H. ingens* (Blackwall, 1857), *H. blackwalli* (Johnson, 1863), *H. heeri* (Thorell, 1875), *H. insularum* (Kulczynski, 1899), *H. nonannulata* Wunderlich, 1995 and *H. isambertoi* Crespo, 2022. A recent study employing molecular and phenotypic data highlighted the importance of integrative methods in resolving species delimitation and understanding the endemic species' biogeographic connections from the Madeira archipelago (Crespo et al., 2022). However, the study failed to draw the boundaries and phylogenetic relationships of two endemic sibling species, namely *H. insularum* , which inhabits all the islands, and the large and colourful *H. maderiana* , circumscribed to Porto Santo and nearby islets, where the two species overlap. This pair of species defied classification due to the presence of intermediate phenotypes and the lack of mitochondrial and nuclear sorting, despite significant divergences in the DNA barcode gene (cytochrome c oxidase subunit I, *COI*) (Crespo et al., 2022).

In this study, we aim to resolve the taxonomic and evolutionary conundrum posed by the Madeiran endemic species pair *H. maderiana* and *H. insularum* , by implementing a genome-wide approach using double digest RADseq (ddRADseq). Specifically, we seek to determine whether these two nominal species represent a case of hybridisation due to gene flow on the Porto Santo Island or if they represent a single species exhibiting extreme morphological polymorphism. To achieve this goal, we used genome-wide data to analyse 61 specimens of *H. maderiana* and *H. insularum* sampled across their known distribution. Our results shed light on their taxonomic status, help to understand the historical events leading to the observed genetic pattern and evaluate the presence of genetic discordance across populations due to incomplete lineage sorting and/or ongoing gene flow.

Material and Methods

Sampling.

We sampled 61 specimens from the previous study of Crespo et al. (2022), representing the known range of the focal species in the Madeira Islands. These specimens were distributed as follows: 27 *H. insularum* , 32 *H. maderiana,* and two unidentified immatures. They included eight *Hogna insularum* from Deserta Grande, seven from Madeira, three from Bugio, nine from Porto Santo, and nearby islets (three from Ilhéu de Cima and one from Ilhéu de Ferro). Additionally, we collected 32 *H. maderiana* from Porto Santo (eight from Ilhéu de Ferro). Finally, we included two immatures from Porto Santo that could not be unambiguously assigned to any of the two species (Figure 1). We also included one *H. nonannulata* specimen from Madeira, seven *H. ingens* specimens from Deserta Grande, and one from Madeira as outgroups (Table S1).

Haplotype network.

To analyse and visualise the relationships among mitochondrial DNA sequences within *H. maderiana* and *H. insularum,* we used the 93 COI sequences of 676 base pairs (bp) available from the previous study (Crespo et al., 2022) (Table S1). We estimated the haplotype network for the COI matrix using the randomised

minimum spanning tree method (Paradis, 2018). We analyse and visualise the data with the help of the R packages ape (Paradis & Schliep, 2019) and pegas (Paradis, 2010).

ddRADseq library preparation and sequencing

For the double-digest restriction-site associated DNA (ddRADseq) library preparation, we used Speedtools Tissue DNA Extraction Kit (Biotools, Madrid, Spain) or DNeasy Blood & Tissue Kit, (Qiagen, Valencia, CA, USA) to extract and purify total genomic DNA, with minor modifications in the cell lysis time (overnight incubation). We estimated the amount of gDNA with Qubit dsDNA HS assay (Invitrogen, Carlsbad, CA, USA). Genomic DNA was individually barcoded and processed in-house using the ddRADseq procedure (Peterson et al., 2012). For ddRADseq library preparation: 300 ng of genomic DNA was digested in a reaction consisting of 0.5 µL high-fidelity restriction enzymes *MseI* (10,000 units ml^{-1}, New England Biolabs, UK*)* , 0.5 µL *EcoRI-HF* (20,000 units ml^{-1}, New England Biolabs) and 2 µl of CutSmart Buffer (New England Biolabs). It was then incubated for 4 hours at 37°C and enzymes were deactivated at 4ºC. We performed a purification step with AMPure XP 1.6X magnetic beads (Beckman Coulter, Inc., Brea, CA, USA) with a final elution in 40 µl. We measured the concentration of purified and digested DNA with Qubit dsDNA HS assay (Invitrogen) and used this value to normalise the DNA. We pooled them into three libraries of 24 samples each based on concentration measurements (Table S1). The resulting fragments were ligated to custom-made 24 P1 and one P2 adapters containing sample-specific 7-base-pair barcodes and primer annealing sites (Table S2). We added the listed buffers and enzymes in every sample for the ligation step: 4 µl T4 DNA Ligase Buffer (New England Biolabs), 1.5 µl T4 DNA Ligase (2,000,000 units ml^{-1}, New England Biolabs), 2 µl P1 adapter (1µM), 2 µl P2 adapter (4µM), 0.5 µl water and 30 µl of DNA digested-normalised. The ligation process was performed for 2 h at 23°C, enzymes were deactivated at 65°C for 10 min and the temperature decreased by 2degC per 90 seconds until reaching 23degC. Then, we pooled all the ligation products in three tubes and purified them with AMPure XP 1.5X magnetic beads (Beckman Coulter), eluting in 31 µl. We size-selected the pools or libraries at 300 bp (range 150–400 bp) with BluePippin (Sage Science, Beverly, MA, USA) using the cassette type "2% DF Marker V1" and the "tight" option. Finally, we amplified each library using the Phusion High-Fidelity PCR Kit (New England Biolabs,) and we conducted PCRs with an initial denaturation for 30 s at 98 °C, 10 cycles (10 s at 98 °C, 30 s at 65 °C, 30 s at 72 °C) and a final extension for 10 min at 72 °C. We cleaned and size-selected the resulting libraries to remove possible remaining adapters and primers using AMPure beads 0.8X (Beckman Coulter). We pooled the three libraries at equal concentrations and paired-end sequenced on an Illumina NovaSeq PE150 at Novogene Co.

ddRADseq locus assembly, filtering, and outlier detection

We demultiplexed and assembled paired-end reads into *de novo* loci using the ipyrad 0.9.83 (Eaton & Overcast, 2020). Raw reads were quality-trimmed to eliminate low-quality sequences, reads with uncalled bases, and reads without complete barcode or restriction cut site. We used the following parameters in ipyrad: "pairddrad" data type for double-digest RADseq; AATTC and TAA for the restriction overhang. We used default parameters for base quality (5), minimum (6), and maximum depth of reads in stacks (10,000). Reads retained were further quality-filtered to convert base calls with a Phred score below 20 into Ns and discard reads with more than two Ns. Within sample clustering, we used the 0.85 threshold. We set the maximum number of allowable barcode mismatches to 0 to ensure that each read was correctly assigned to its identifying barcode. Filtering for adapters/primers was set to 2, the strictest setting, to remove any barcodes that may have been left over from the demultiplexing and could interfere with downstream clustering. We set the minimum length of reads to 35 and the maximum alleles per site in the consensus sequences to two. We used the default settings for the maximum number of uncalled bases and heterozygotes in consensus sequences (0.05). Finally, we used default settings for the maximum number of single nucleotide polymorphisms (SNPs) per locus (0.2) and indels per locus (8). We assembled two different datasets, each differing in the number of species: dataset a) all specimens from the four*Hogna* species that were part of the study and dataset b) only specimens from *H. insularum* and *H. maderiana* .

Population genomic structure and genetic diversity

To examine the genetic structure among the specimens in dataset b), we performed a discriminant analysis of principal components (DAPC) (Table S1). The DAPC method combines discriminant and principal component analysis to maximise the genetic variability between groups and minimise the variability within each group (Jombart et al., 2010). The study was performed with the R package adegenet 2.1.10 (Jombart & Ahmed, 2011). We used the "xvalDapc" function to find the optimal number of principal components (PCs) to be retained and "find.clusters" to find the optimal number of clusters based on the Bayesian information criterion (BIC). The acquired values were utilised to do a DAPC on the data using the "dapc" function. We employed the "scatter" function to display the graphs.

To infer the population genetic structure and detect potential admixture among groups, we used structure 2.3 (Pritchard et al., 2000) under the admixture model based on one SNP per locus, for dataset b) (Table S1). We predefined the number of genetic clusters (K) from one to five. We conducted ten replicates for each K, executing 200,000 iterations and a burn-in of 10,000. Graphs from the 10 runs at each value of K were aggregated using clumpak (Kopelman et al., 2015) and the probable number of ancestral clusters was selected based on ΔK and Pritchard from Structure Harvester 0.6.94 (Earl & vonHoldt, 2012; Evanno et al., 2005).

In addition, we employed fineradstructure (Malinsky et al., 2018), a model-based method developed explicitly for evaluating extensive SNP datasets, to assess the genetic structure patterns while identifying the sources of ancestry in each population. This software provides high resolution for inferring recent shared ancestry by utilising haplotype linkage information and prioritising the most recent common ancestry among the studied individuals. It generates a co-ancestry matrix, summarising the dataset's closest neighbour haplotype relationships. More precisely, we employed the script finerad_input.py, which is part of the fineRADstructure-tools package available at https://github.com/edgardomortiz/fineRADstructure-tools, to transform the ".alleles" output from ipyrad into the appropriate input format for fineRADstructure. As part of the conversion process, we specifically filtered the dataset to include only loci that were not linked (using the default parameter) and had a minimum sample size of eight (–minsample 8). Individuals included in the analyses are summarized in Table S1. Subsequently, we created the co-ancestry matrix using the RADpainter module of fineradstructure. We employed the default parameters of 100,000 Markov chain Monte Carlo (MCMC) iterations, with a burn-in of 100,000 iterations and sampling taking place every 1000 iterations, to conduct the analysis. A tree was built using 10,000 iterations of the hill-climbing algorithm. We visualised the results using the scripts fineradstructureplot.r and finestructurelimrary.r, which may be accessed at http://cichlid.gurdon.cam.ac.uk/fineRADstructure.html.

Since we identified a genetic group with a signal of putative admixture, we investigated whether there is evidence of potential gene flow or incomplete lineage sorting (ILS) between the different genetic groups. To achieve this, we used Patterson's D-statistics, also known as the abba-baba test (Durand et al., 2011), implemented in dsuite 0.1 (Malinsky et al., 2021). The D-statistics analyses correlations in allele frequencies across closely related species since they share a significant amount of genetic variation due to common ancestry and ILS. This algorithm computes the D-statistic by considering a 4-taxon fixed phylogeny: (((P1, P2), P3), O) where O is an outgroup with the ancestral allele "A". The "A" allele and derived allele "B" should follow the pattern BBAA (P1 and P2 share the allele "B"); the pattern ABBA refers to P2 and P3 sharing the derived allele "B" and BABA refers to P1 and P3 sharing the derived allele. The allele patterns ABBA and BABA, which differ from the species tree pattern BBAA, can provide evidence of introgression when there is a disproportion between the two patterns, usually an excess of them. ILS is the most plausible explanation if these two patterns occur in equal frequency. We performed the four-taxon D-statistic test on two different comparisons with dataset a: a) taking into account the three genetic groups detected by the genetic structure analyses (*H. insularum* from Madeira, Desertas, and Bugio*, H. insularum* from Porto Santo and *H. maderiana)* and *H. ingens* as outgroup; b) considering the three genetic groups across the different islands, with *H. ingens* as outgroup (Table S1).

Demographic analyses

Based on the genetic structure analysis results, we conducted a demographic modelling analysis to further

elucidate between incomplete lineage sorting or introgression as a mechanism to explain the recovered evolutionary patterns within the *Hogna* species complex. We used two approaches based on the unfolded site frequency spectrum (SFS). On the one hand, using fastsimcoal2 (fsc28) (Marchi et al., 2024), we tested five demographic scenarios based on the multidimensional joint site frequency spectrum (SFS). We used easySFS (https://github.com/isaacovercast/easySFS) to generate a folded multidimensional SFS. We downsampled our dataset to a smaller number: four *H. maderiana* individuals, four *H. insularum* from Madeira, and three *H. insularum* from Porto Santo. Briefly, we simulated five demographic models (Figure S1). We first considered two scenarios without gene flow between lineages, and hence that admixture was due to ancestral polymorphism: model A) where*H. insularum* split into two lineages, one from Porto Santo and one from Madeira; and model B) where *H. maderiana* and *H. insularum* from Porto Santo shared a common ancestor. The following scenario (model C) considered recent gene flow between the two species from Porto Santo (*H. insularum* and *H. maderiana*). The fourth scenario (model D) considered that *H. insularum* from Porto Santo resulted from secondary contact between *H. maderiana* and*H. insularum* , with no subsequent gene flow. Finally, we tested the same secondary contact scenario but with subsequent gene flow between the three lineages (model E). We fixed the effective population size (*Ne*) for *H. maderiana* to enable the estimation of other parameters in fastsimcoal2. We calculated *Ne* fixed in the model from the level of nucleotide diversity (π) and estimates of mutation rate per site per generation (μ, because$N_e = (\pi/4\mu)$. We estimated π for *H. maderiana* from polymorphic and non-polymorphic loci using DnaSP ($\pi = 0.0048$). We assumed a mutation rate per site per generation of 6×10^{-9} as estimated for*Drosophila melanogaster* (Wang et al., 2023). All template files (*.tpl) that contain parameters and estimation files (*.est) that contain unknown parameters for estimation were provided in Supplementary Material. We ran 100 independent simulations of each model in fastsimcoal2. Each run comprised 100,000 coalescent simulations, 50 expectation–conditional maximisation (ECM) cycles, and a proportion of the initial search in the last cycle of 0.5 (-z0.5). We used Akaike's Information Criterion (AIC) to determine the probability of each model given the observed data and select the best model. Finally, we calculated each parameter estimate's confidence intervals (CI) from the best-supported model by simulating 20 replicates of 100 SFS and re-estimating the parameters each time. We obtained the point estimates of each demographic parameter for the best-supported model from the run with the highest composite maximum likelihood score.

On the other hand, we implemented a model-flexible approach using stairway plot2 (Liu & Fu, 2015) to infer the demographic history of each lineage separately and to understand the events that can explain the observed pattern in fastsimcoal2. We assumed a generation time of 1 year. Median estimates of *Ne* and confidence intervals were estimated with the built-in bootstrap function using 200 subsets of the input data, ignoring singletons. We graphically represented the results with ggplot2.

Species tree under the multispecies coalescent model

Finally, we used the Bayesian multispecies coalescent framework of snapp 2.7.3 (Bryant et al., 2012) as implemented in beast 2 2.7.3 (Bouckaert et al., 2014) to estimate a species tree and corroborate the presence of incomplete lineage sorting. Given that SNAPP does not consider gene flow or introgression in its coalescent process modelling, the species tree generated would reflect the most probable evolutionary relationships among species, assuming all discrepancies between gene trees and the species tree are attributable to ILS. The dataset consisted of 696 unlinked single nucleotide polymorphisms that were converted into a biallelic format. These SNPs were derived from the ingroup specimens, which comprised 50 specimens from *Hogna maderiana* and *Hogna insularum* . The individuals were categorised into three genetic groups: a) *H. maderiana* , b) *H. insularum* from Madeira, Desertas, and Bugio and c) *H. insularum* from Porto Santo (Table S1), based on the results of the genomic structure analyses. The .usnps file generated by ipyrad was modified and transformed into a snapp input file. The model was defined with specific parameter values: the mutation rates u and v were fixed at 1 and not subject to sampling, the coalescent rate was set to 10 and subject to sampling, non-polymorphic sites were excluded, and a log-likelihood correction was applied. snapp utilises a Yule prior, where the parameter lambda represents the rates at which new species originate. Prior probabilities: alpha = 1, beta = 250, gamma = 1, lambda = 0.01, no samples taken. We conducted four separate iterations, each consisting of a chain length of 100,000 generations. In each iteration, we collected samples at intervals of

100 generations. We assessed convergence by ensuring the effective sample size (ESS) was greater than 200. We determined that the burn-in period was 10% using tracer 1.7.2. Additionally, we analysed the posterior distribution of trees using densitree 3.0.2. We utilised LogCombiner to merge the four runs and generated a maximum clade credibility tree (MCC) using treeannotator 2.7.3. Finally, we visualised the tree using itol 6.8 (Letunic & Bork, 2024).

Results

Haplotype network

The 93 COI sequences yielded 51 haplotypes (Table S3, Figure S2). Among these haplotypes, 35 were singletons. A taxonomic segregation into the two species was not readily apparent in the haplotype network. In several instances, haplotypes of one species were more closely connected to the ones of the other species. Haplotypes of *H. insularum* from Madeira formed an exclusive cluster. They shared the same ancestral relative that most specimens sampled from Desertas and the closeby Bugio islet, except for two distantly related haplotypes more closely associated with Porto Santo ones. Haplotype IX was the most abundant (12 out of 93 individuals), present in Porto Santo and Ilhéu de Ferro, and restricted to *H. maderiana* . The second most abundant haplotype (XXXI) was exclusive to *H. insularum* from Madeira, Desertas, and Bugio (Table S3, Figure S2).

SNP ddRADseq data

Sequencing of the ddRADseq library resulted in 164,606,284 paired-end (PE) reads that contained a maximum of a single error per 1000 bases, with a mean individual sample coverage of 1,130,758.37 ± 474,932.71 (Table S4). The initial matrix with the four species included 70 individuals (4034 SNPs, 497 unlinked SNPs, and 67.25% of missing data). Still, after eliminating 11 individuals based on a low number of reads (below 750,000) or low read quality, our final ddRADseq datasets included 59 individuals, including two outgroups. After quality-trimming, we retained for the first dataset a total of 74,772,557 reads (1,267,331.47 ± 372,437.93 individual sample coverage) (Table S5), which resulted in 7158 SNPs, 803 unlinked SNPs and 67.2% missing data. For the *Hogna insularum and H. maderiana* dataset, we retained a total of 60,779,385 reads (1,215,587.70 ± 314,645.31) (Table S6), which resulted in 6456 SNPs, 865 unlinked SNPs, and 65,54% missing data for downstream analyse.

Population genomic structure and genomic diversity

Six PCs and two discriminant eigenvalues were retained during DAPC analysis to describe the relationship between the samples. Under the BIC criteria, three clusters were found as the optimal results. The first cluster contained *Hogna maderiana* individuals from Porto Santo, the second cluster gathered all *Hogna insularum* individuals from Madeira, Desertas and Bugio (hereafter referred as MDB), whereas the third comprised those individuals of *Hogna insularum* from Porto Santo (hereafter referred as PS). The different clusters were well-separated and did not overlap (Figure 2a).

structure analyses yielded an "optimal" clustering value for K=3 according to Evanno's ΔK criterion and K=2 according to Pritchard (Figure 2b, Figure S3). The structure results for K=2 revealed two genetically distinct groups of individuals, one corresponding to*H. maderiana* and one to *H. insularum* MDB; however, individuals of *H. insularum* PS shared approximately 50% genetic component of each genetic group. Considering K=3, there was no distinct genetic structure, with individuals exhibiting different levels of genetic similarity with a common dominant genetic component, but a distinct additional component in individuals of *H. insularum* MDB. Results for K=1 to K=5 are summarised in Figure S4.

The fineradstructure analysis was mainly in agreement with the structure and DAPC results. The dendogram and coancestry matrix resulting from fineradstructure resolved two major groupings in the data, corresponding to *H. insularum* and *H. maderiana* (Figure 2c). Within the *H. insularum* group, individuals clustered in two supported subclusters belonging to the individuals from Porto Santo on one side and the individuals from the remaining islands on the other. Overall, the co-ancestry level was high within the three groups, higher in *H. insularum* MDB. Slightly higher average values were observed between the *H.*

insularum groups compared to the coancestry levels between *H. insularum* PS and *H. maderiana* . However, the coancestry values remain relatively uniform and comparable across the species.

The abba-baba tests supported an incomplete lineage sorting (ILS) scenario across the three genetic groups. No significant evidence of gene flow was detected in any comparisons among *H. insularum* PS, *H. insularum* MDB, and *H. maderiana* , as indicated by the non-significant *p-values* and relatively low Z-scores (Table S7). On the other hand, although not directly relevant to the objective of our analysis, we observed signals of gene flow between populations of*H. insularum* from the Madeira and Bugio Islands. Detailed results of the abba-baba tests are presented in Table S7.

Coalescent-based demographic model

The fastsimcoal2 analysis identified the model of three independent lineages (*H. insularum* MDB, *H. insularum* PS, and *H. maderiana*) with neither past admixture nor contemporary gene flow as the one with the highest support (model A; see Table S8 and Figure 3a). According to this model, the lineage of *H. insularum* PS would constitute a distinct species closely related to *H. insularum* MDB. Assuming a 1-year generation time, fastsimcoal2 estimated that lineages diverged from a shared ancestor approximately divergence between the two *H. insularum* lineages occurred around estimated *Ne* for *H. insularum* from Porto Santo was*H. insularum* from Madeira, Desertas, and Bugio, it was presentation of all estimated demographic variables and their weighted averages can be found in Figure 3a and Table S9.

We repeated the demographic analyses using stairway plot, revealing a consistent pattern in the two lineages inhabiting Porto Santo Island. Both lineages exhibited a stepwise increase in effective population size, eventually stabilising at a constant level. On the other hand, in *H. insularum* MDB, we identified a rapid population increase followed by a more recent stabilisation (Figure 3b).

Species tree under the multispecies coalescent model

snapp inferred three incongruent species tree topologies from the posterior distribution. The species tree topology with the highest probability (45.93%) indicated a sister group relationship between the lineages from *H. maderiana* and *H. insularum* PS. The following topology (35.53%) recovered the *H. insularum* PS and MDB as sister groups. The third and least supported topology (21.54% of the gene trees) recovered *H. maderiana* as sister to *H. insularum* MDB (Figure 4). The consensus tree yielded a low support level (posterior probability of 0.47) for the clade of the two Porto Santo lineages (Figure S5).

Discussion

Genomic data has become an invaluable tool for unveiling complex evolutionary scenarios, providing insights into hidden aspects of population history and recent species divergence. However, current practice in population genomic studies fail to provide evidence on the full evolutionary history of a species and, therefore, achieve incorrect conclusions. Our study focused on two endemic *Hogna* species from the Madeira archipelago, exhibiting mitochondrial discordance and fuzzy morphological limits. We used ddRADseq data, integrating population genomic structure analyses and coalescent-based demographic modelling to understand the contribution of incomplete lineage sorting (ILS) and gene flow to the observed genetic patterns. Our genetic structure analyses suggested three potential genetic clusters, one for each nominal species and one compatible with hybridisation between the two species on Porto Santo Island. However, our demographic models and the D-statistic test explicitly rejected gene flow between the two lineages. Instead, they supported a third independent *Hogna* lineage on Porto Santo Island. The apparent absence of clear-cut limits between the lineages is likely explained by unsorted polymorphism, probably due to the recent divergence times, as some of our results suggest. The contrasting results between inferences of hybridisation based on population structure patterns and multi-population coalescent analyses highlight the evolutionary complexity of *Hogna* species from Madeira and underscore the need for careful interpretation of genomic data.

The perils of inferring hybridisation from population structure patterns

Our genome-wide data revealed a population structure closely matching the two nominal species, *Hogna*

insularum, and *H. maderiana* , but with a third component, formed by individuals from Porto Santo, first identified as *H. insularum,* of ambiguous interpretation. This pattern was consistently observed across both non-model-based (DAPC) and model-based clustering methods (structure and fineradstructure) (Figure 2). DAPC and fineradstructure analyses identified three genetic clusters, with slightly higher co-ancestry within each cluster than among them, suggesting a certain degree of genetic similarity across the three genetic clusters. While in general agreement, the structure analysis further supported this genetic continuity. At K=3 (as suggested by Evanno's criterion), we observed varying levels of genetic similarity among the groups with similar dominant components across them, but different secondary components for those individuals outside Porto Santo. At K=2, individuals of *H. insularum* from Porto Santo exhibited an equal ancestry proportion of 50% from each species, suggesting ongoing hybridisation between them. At this point, it is essential to note that determining the correct number of genetic clusters using Evanno and Pritchard methods in STRUCTURE is challenging, as these approaches can either underestimate or overestimate the number of clusters, particularly in populations with complex evolutionary histories (Janes et al., 2017). In the case of the *Hogna* species complex, the identification of K=2 suggests that the *H. insularum* lineage from Porto Santo may result from continued hybridisation between *H. insularum* and *H. maderiana* in Porto Santo. Surprisingly, none of the samples collected from Porto Santo belonged to the *H. insularum* Madeira, Desertas, and Bugio (MDB) (Figure 2), which challenges the putative scenario of a potential hybrid zone between the two nominal species co-occurring in Porto Santo.

Coalescence-based demographic analyses shed light on complex evolutionary scenarios.

Despite the different analyses that support the presence of three genetic clusters, one of them potentially resulting from the hybridisation of the other two, we could not rule out alternative explanations. To investigate further the underlying processes responsible for the observed pattern, we conducted a D-statistic test to differentiate between gene flow and ILS. The results found no significant support for gene flow between the three genetic clusters. Moreover, we observed symmetrical counts of the ABBA and BABA patterns, indicating no excess of shared alleles. Given the lack of evidence for gene flow, we explored alternative demographic models to understand the evolutionary history of the lineages better. By considering multiple evolutionary processes simultaneously, demographic analyses can disentangle complex scenarios that may produce ambiguous or conflicting signals, providing deeper insights into the species' evolutionary dynamics (Galià-Camps et al., 2024; Puckett & Munshi-South, 2019). We proposed five scenarios encompassing the observed patterns in genomic structuring and D-statistic analyses. These scenarios suggested: the existence of three recently diverged independent lineages, reflecting the potential effects of incomplete lineage sorting (Models A and B, Figure S1), ongoing gene flow as evidenced by the potential hybrids identified in Porto Santo (Model C, Figure S1), and, finally, we also considered the possibility that *H. insularum* PS arose from secondary contact between*H. insularum* MBD and *H. maderiana* , either episodic, without contemporary gene flow (Model D), or with contemporary gene flow (Model E, Figure S1). The scenario of three lineages showing high levels of co-ancestry was preferred over those involving hybridisation. Additionally, the effective population size estimation of the*Hogna* lineages, as inferred by the stairway plot 2, identified an ancestral split on Porto Santo, followed by a similar pattern of population growth between the two resulting lineages. Interestingly, the next split, which corresponds to the origin of the*H. insularum* MDB lineage, shows a much steeper, continuous population expansion compatible with a bottleneck following the colonisation of new islands (Madeira, Desertas) from the ancestral Porto Santo. This founder event inevitably depleted genetic variation (Cerca, Cotoras, Bieker, et al., 2023) and helps to explain the exclusive mtDNA haplotypes found in Desertas and, especially Madeira.

Results of snapp, however, failed to support the topology selected by the demographic analyses and revealed instead a high level of discordance among alternative tree topologies of *Hogna* lineages. High levels of gene-tree discordance suggest the occurrence of hybridisation and/or ILS (Mallet et al., 2016). Since our previous analyses discarded gene flow, we attribute that observed discordance to the presence of ILS. Short speciation times and large effective population sizes may result in the retention of ancestral polymorphisms across species (Maddison, 1997). Although they do not directly influence the description of the observed patterns, it is worth noting that our divergence time estimates differed across methods and from those based on

mitochondrial data by Crespo et al. (2022) (˜1.2 million years ago). These discrepancies could be attributed to the inherent characteristics of the methods: StairwayPlot is more informative for events spanning hundreds of generations, whereas fastsimcoal2 is better suited for detecting more recent demographic changes (Lapierre et al., 2017; Patton et al., 2019).

Regardless of the times, the inferred patterns of ancestral population sizes reconcile well with the higher levels of retained polymorphisms observed between *H. maderiana* and *H. insularum* PS, and, as stated above, the reduced polymorphisms in genome-wide data and mitochondrial haplotypes observed in *H. insularum* MDS.

Taxonomic implications

Our two starting hypotheses explained the reported lack of sorting in mitochondrial and nuclear genes among nominal species, either as a single species exhibiting extreme morphological polymorphism or as two nominal species with restricted hybridisation between them. Instead, our results support the existence of a third lineage, along with *H. maderiana* and *H. insularum* MDB, consisting of the individuals from Porto Santo initially assigned to *H. insularum* . Hence, the fuzzy boundaries between the species would be due to incomplete lineage sorting. Interestingly, *H. insularum* exhibits intraspecific variation in shape and size across the islands (Crespo et al., 2022), which led Wunderlich (1992) to describe the species *H. biscoitoi* Wunderlich, 1992, based on specimens from Porto Santo. Subsequently, Crespo et al. (2022) proposed *H. biscoitoi* as a junior synonym of*H. insularum* due to the lack of diagnostic morphological features and genetic distinctiveness. The lack of diagnostic features among*H. insularum* specimens from different island populations (Madeira, Desertas, Bugio and Porto Santo), with only minor variations, could also be attributed to homogeneous genetic variation caused by incomplete lineage sorting. This process can preserve symplesiomorphic morphological states, as suggested by (Futuyma, 2010). Crespo et al. (2022) further proposed the existence of a phenotypic continuity between the two extremes represented by the nominal species *H. maderiana* or *H. insularum* . Notably, most specimens with intermediate size and/or yellowish pilosity on the anterior legs, found in Porto Santo, correspond to *H. maderiana,* as also recovered in our results. Unfortunately, the limited sample of specimens from the different islands prevented us from conducting quantitative morphological analyses to further test the morphological distinctiveness of the island populations of *H. insularum.*

The rejection of gene flow across the different species in our Madeiran*Hogna* is surprising, given that recent studies have revealed the key role of gene flow between islands in promoting habitat adaptation in*Hogna* species on the Galapagos Islands (De Busschere et al., 2015). In this regard, wolf spiders are well-reputed for their great ability for long-distance dispersal (Bonte et al., 2006; Dries & Maelfait, 2001), as already illustrated by the presence of*H. insularum* in multiple islands. A possible explanation for the lack of hybridisation, or even a potential driver of speciation in our model system, could be found in the complex courtship behaviour exhibited by wolf spiders, which has frequently been used to diagnose species (Chiarle et al., 2013; Just et al., 2019). Although little is known about the courtship behaviour of Madeiran *Hogna* , the species pair *H. maderiana* and *H. insularum* differ in traits related to male signalling during courtship displays. Specifically, *H. maderiana* has pedipalps and front legs conspicuously marked in orange, which are absent in *H. insularum* . The absence of gene flow or hybridisation events in *Hogna* speciation may hint at the development of specific courtship behaviours during mating, which prevents genetic material from crossing between species and could promote speciation in Madeiran Hogna.

CONCLUSIONS AND FUTURE PERSPECTIVES

The combination of genome-wide data with the use of coalescent models of demographic history has enabled us to reject both extreme morphological polymorphism and hybridisation to explain mitochondrial discordance and lack of nuclear sorting in nominal species of *Hogna* spiders in the Madeiran Archipelago. Instead, our results identified a third independent evolutionary lineage, for which the name *Hogna biscoitoi* is available, on the small and highly disturbed island of Porto Santo. With this species, the number of endemic *Hogna* species in Madeira rises to eight.

Our study constitutes a further example of the perils of inferring evolutionary processes based on the evidence provided by the results of population structure analyses. With the introduction of genomics and genome-wide

markers, many studies have claimed the existence of hybrid populations or introgressive hybridisation, without further evaluating alternative scenarios compatible with the recovered patterns. We encourage future studies to consider alternative evolutionary scenarios to distinguish between ILS and gene flow and resolve challenging taxonomic situations. In this study, we implemented a straightforward protocol to fully disentangle between genetic signals in recently diverged species, which will contribute to improving the correct assessment of biodiversity in similar situations across biogeographic regions. Although our data have greatly helped to elucidate one of the major challenges in the current taxonomy of an island lineage of spiders, some of which are of conservation concern, we encourage further research to increase both the number of markers and, mainly, the individuals and population sample size to test more specific evolutionary events, such as bottlenecks associated with the colonisation of new islands. Equally important, reconciling the genetic evidence with a quantitative assessment of morphological data will allow for the future identification of the drivers of speciation in this exceptional island model.

Acknowledgements

We are in debt to Alejandro Sánchez-Gracia for his advice on running demographic models and for a critical review of an early draft. Also to Carles Galià-Camps for his help and advice on working with the supercomputer and implementing the softwares.

References

Bidegaray-Batista, L., Sánchez-Gracia, A., Santulli, G., Maiorano, L., Guisan, A., Vogler, A. P., & Arnedo, M. A. (2016). Imprints of multiple glacial refugia in the Pyrenees revealed by phylogeography and palaeodistribution modelling of an endemic spider. *Molecular Ecology*, *25*(9), 2046–2064. https://doi.org/10.1111/mec.13585

Bonte, D., Borre, J. V., Lens, L., & Maelfait, J.-P. (2006). Geographical variation in wolf spider dispersal behaviour is related to landscape structure. *Animal Behaviour*, *72*(3), 655–662. https://doi.org/10.1016/j.anbehav.2005.11.026

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *10*(4), e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, *29*(8), 1917–1932. https://doi.org/10.1093/molbev/mss086

Cerca, J., Cotoras, D. D., Bieker, V. C., De-Kayne, R., Vargas, P., Fernández-Mazuecos, M., López-Delgado, J., White, O., Stervander, M., Geneva, A. J., Guevara Andino, J. E., Meier, J. I., Roeble, L., Brée, B., Patiño, J., Guayasamin, J. M., Torres, M. de L., Valdebenito, H., Castañeda, M. D. R., … Martin, M. D. (2023). Evolutionary genomics of oceanic island radiations. *Trends in Ecology & Evolution*, *38*(7), 631–642. https://doi.org/10.1016/j.tree.2023.02.003

Cerca, J., Cotoras, D. D., Santander, C. G., Bieker, V. C., Hutchins, L., Morin-Lagos, J., Prada, C. F., Kennedy, S., Krehenwinkel, H., Rominger, A. J., Meier, J., Dimitrov, D., Struck, T. H., & Gillespie, R. G. (2023). Multiple paths toward repeated phenotypic evolution in the spiny-leg adaptive radiation (*Tetragnatha*; Hawai'i). *Molecular Ecology*, *32*(18), 4971–4985. https://doi.org/10.1111/mec.17082

Chiarle, A., Kronestedt, T., & Isaia, M. (2013). Courtship behavior in European species of the genus*Pardosa* (Araneae, Lycosidae). *The Journal of Arachnology*,*41*(2), 108–125. https://doi.org/10.1636/hi12-09.1

Crespo, L. C., Silva, I., Enguídanos, A., Cardoso, P., & Arnedo, M. (2022). Island hoppers: Integrative taxonomic revision of wolf spiders (Araneae, Lycosidae) endemic to the Madeira islands with description of a new species.*ZooKeys*, *1086*, 84–135. https://doi.org/10.3897/zookeys.1086.68015

Croucher, P. J. P., Jones, R. M., Searle, J. B., & Oxford, G. S. (2007). Contrasting patterns of hybridization

in large house spiders (*Tegenaria atrica* group, Agelenidae). *Evolution; International Journal of Organic Evolution*, *61*(7), 1622–1640. https://doi.org/10.1111/j.1558-5646.2007.00146.x

De Busschere, C., Van Belleghem, S. M., & Hendrickx, F. (2015). Inter and intra island introgression in a wolf spider radiation from the Galápagos, and its implications for parallel evolution. *Molecular Phylogenetics and Evolution*, *84*, 73–84. https://doi.org/10.1016/j.ympev.2014.11.004

Domènech, M., Crespo, L. C., Enguídanos, A., & Arnedo, M. A. (2020). Mitochondrial discordance in closely related *Theridion* spiders (Araneae, Theridiidae), with description of a new species of the *T. melanurum* group.*Zoosystematics and Evolution*, *96*(1), 159–173. https://doi.org/10.3897/zse.96.49946

Dries, B., & Maelfait, J.-P. (2001). Life history, habitat use and dispersal of a dune wolf spider (Pardosa monticola (Clerck, 1757) Lycosidae, Araneae) in the Flemish coastal dunes (Belgium). *Belgian Journal of Zoology*.

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*,*28*(8), 2239–2252. https://doi.org/10.1093/molbev/msr048

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Eaton, D. A. R., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets.*Bioinformatics* , *36*(8), 2592–2594. https://doi.org/10.1093/bioinformatics/btz966

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*,*14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Futuyma, D. J. (2010). Evolutionary constraint and ecological consequences. *Evolution; International Journal of Organic Evolution*, *64*(7), 1865–1884. https://doi.org/10.1111/j.1558-5646.2010.00960.x

Galià-Camps, C., Enguídanos, A., Turon, X., Pascual, M., & Carreras, C. (2024). The past, the recent, and the ongoing evolutionary processes of the worldwide invasive ascidian *Styela plicata*. *Molecular Ecology*, e17502. https://doi.org/10.1111/mec.17502

Gloor, D., Nentwig, W., Blick, T., & Kropf, C. (2017). *World Spider Catalog*. Natural History Museum Bern. https://doi.org/10.24436/2

Gompert, Z., & Buerkle, C. A. (2016). What, if anything, are hybrids: enduring truths and challenges associated with population structure and gene flow. *Evolutionary Applications*, *9*(7), 909–923. https://doi.org/10.1111/eva.12380

He, B., Zhao, Y., Su, C., Lin, G., Wang, Y., Li, L., Ma, J., Yang, Q., & Hao, J. (2023). Phylogenomics reveal extensive phylogenetic discordance due to incomplete lineage sorting following the rapid radiation of alpine butterflies (Papilionidae: *Parnassius*). *Systematic Entomology*. https://doi.org/10.1111/syen.12592

Hernández-Gutiérrez, R., van den Berg, C., Granados Mendoza, C., Peñafiel Cevallos, M., Freire M, E., Lemmon, E. M., Lemmon, A. R., & Magallón, S. (2022). Localized phylogenetic discordance among nuclear loci due to incomplete lineage sorting and introgression in the family of cotton and cacao (*Malvaceae)*. *Frontiers in Plant Science*, *13*, 850521. https://doi.org/10.3389/fpls.2022.850521

Hernández-Hernández, T., Miller, E. C., Román-Palacios, C., & Wiens, J. J. (2021). Speciation across the Tree of Life. *Biological Reviews of the Cambridge Philosophical Society*, *96*(4), 1205–1242. https://doi.org/10.1111/brv.12698

Herrig, D. K., Ridenbaugh, R. D., Vertacnik, K. L., Everson, K. M., Sim, S. B., Geib, S. M., Weisrock, D. W., & Linnen, C. R. (2024). Whole genomes reveal evolutionary relationships and mechanisms underlying

gene-tree discordance in*Neodiprion* sawflies. *Systematic Biology*. https://doi.org/10.1093/sysbio/syae036

Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., & Andrew, R. L. (2017). The K = 2 conundrum. *Molecular Ecology*,*26*(14), 3594–3602. https://doi.org/10.1111/mec.14187

Jia, T. Z., Caudan, M., & Mamajanov, I. (2021). Origin of Species before Origin of Life: The Role of Speciation in Chemical Evolution. *Life* , *11*(2). https://doi.org/10.3390/life11020154

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* , *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*, 94. https://doi.org/10.1186/1471-2156-11-94

Just, P., Opatova, V., & Dolejš, P. (2019). Does reproductive behaviour reflect phylogenetic relationships? An example from Central European *Alopecosa* wolf spiders (Araneae: Lycosidae). *Zoological Journal of the Linnean Society*, *185*(4), 1039–1056. https://doi.org/10.1093/zoolinnean/zly060

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*,*15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Lapierre, M., Lambert, A., & Achaz, G. (2017). Accuracy of Demographic Inferences from the Site Frequency Spectrum: The Case of the Yoruba Population. *Genetics*,*206*(1), 439–449. https://doi.org/10.1534/genetics.116.192708

Letunic, I., & Bork, P. (2024). Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research*, *52*(W1), W78–W82. https://doi.org/10.1093/nar/gkae268

Liu, X., & Fu, Y.-X. (2015). Corrigendum: Exploring population size changes using SNP frequency spectra. *Nature Genetics*, *47*(9), 1099. https://doi.org/10.1038/ng0915-1099a

Logunov, D. V. (2020). On three species of *Hogna* Simon, 1885 (Aranei: Lycosidae) from the Near East and Central Asia. In *Arthropoda Selecta* (Vol. 29, Issue 3, pp. 349–360). https://doi.org/10.15298/arthsel.29.3.08

Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology*, *46*(3), 523–536. https://doi.org/10.1093/sysbio/46.3.523

Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*,*21*(2), 584–595. https://doi.org/10.1111/1755-0998.13265

Malinsky, M., Trucchi, E., Lawson, D. J., & Falush, D. (2018). RADpainter and fineRADstructure: Population Inference from RADseq Data. *Molecular Biology and Evolution*, *35*(5), 1284–1290. https://doi.org/10.1093/molbev/msy023

Mallet, J., Besansky, N., & Hahn, M. W. (2016). How reticulated are species? *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*,*38*(2), 140–149. https://doi.org/10.1002/bies.201500149

Marchi, N., Kapopoulou, A., & Excoffier, L. (2024). Demogenomic inference from spatially and temporally heterogeneous samples. *Molecular Ecology Resources*,*24*(1), e13877. https://doi.org/10.1111/1755-0998.13877

Nyffeler, M., & Birkhofer, K. (2017). An estimated 400-800 million tons of prey are annually killed by the global spider community. *Die Naturwissenschaften*,*104*(3-4), 30. https://doi.org/10.1007/s00114-017-1440-1

Oxford, G. S., & Bolzern, A. (2018). Molecules v. Morphology—is *Eratigena atrica* (Araneae: Agelenidae) one species or three? *Arachnology*, *17*(7), 337–357. https://doi.org/10.13156/arac.2017.17.7.337

Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach.*Bioinformatics* , *26*(3), 419–420. https://doi.org/10.1093/bioinformatics/btp696

Paradis, E. (2018). Analysis of haplotype networks: the randomized minimum spanning tree method.*Methods in Ecology and Evolution*, *9*, 1308–1317. https://doi.org/10.1111/2041-210X.12969

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* , *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Patton, A. H., Margres, M. J., Stahlke, A. R., Hendricks, S., Lewallen, K., Hamede, R. K., Ruiz-Aravena, M., Ryder, O., McCallum, H. I., Jones, M. E., Hohenlohe, P. A., & Storfer, A. (2019). Contemporary Demographic Reconstruction Methods Are Robust to Genome Assembly Quality: A Case Study in Tasmanian Devils. *Molecular Biology and Evolution*, *36*(12), 2906–2921. https://doi.org/10.1093/molbev/msz191

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS One*, *7*(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Piacentini, L. N., & Ramírez, M. J. (2019). Hunting the wolf: A molecular phylogeny of the wolf spiders (Araneae, Lycosidae). In *Molecular Phylogenetics and Evolution* (Vol. 136, pp. 227–240). https://doi.org/10.1016/j.ympev.2019.04.004

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Puckett, E. E., & Munshi-South, J. (2019). Brown rat demography reveals pre-commensal structure in eastern Asia before expansion into Southeast Asia.*Genome Research*, *29*(5), 762–770. https://doi.org/10.1101/gr.235754.118

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, A., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., ... Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews. Genetics*,*15*(3), 176–192. https://doi.org/10.1038/nrg3644

Steenwyk, J. L., Li, Y., Zhou, X., Shen, X.-X., & Rokas, A. (2023). Incongruence in the phylogenomics era. *Nature Reviews. Genetics*, *24*(12), 834–850. https://doi.org/10.1038/s41576-023-00620-x

Wang, Y., McNeil, P., Abdulazeez, R., Pascual, M., Johnston, S. E., Keightley, P. D., & Obbard, D. J. (2023). Variation in mutation, recombination, and transposition rates in and. *Genome Research*, *33*(4), 587–598. https://doi.org/10.1101/gr.277383.122

Wheeler, W. C., Coddington, J. A., Crowley, L. M., Dimitrov, D., Goloboff, P. A., Griswold, C. E., Hormiga, G., Prendini, L., Ramírez, M. J., Sierwald, P., Almeida-Silva, L., Alvarez-Padilla, F., Arnedo, M. A., Benavides Silva, L. R., Benjamin, S. P., Bond, J. E., Grismado, C. J., Hasan, E., Hedin, M., ... Zhang, J. (2017). The spider tree of life: phylogeny of Araneae based on target-gene analyses from an extensive taxon sampling.*Cladistics : The International Journal of the Willi Hennig Society*, *33*(6), 574–616. https://doi.org/10.1111/cla.12182

Data Accessibility

Raw sequence reads are deposited in the SRA (BioProject SUB14946694)

Conflict of interest statement

The authors have no conflicts of interest to disclose.

Funding

VT was funded by an individual MSC fellowship (P.N.:101025947). This work was supported by project CGL2016-80651-P from the Spanish Ministry of Economy and Competitivity and PID2019-105794GB-I00 from the Spanish Ministry of Science and innovation (both to MA). Additional funds were provided by the project 2021SGR689 from the Catalan Government (MA).

Author Contributions

All authors contributed to conceptualising this study. **MA** acquired funding. **LC** conducted the fieldwork. **AE** conducted molecular work and curated the data. **AE** and**VT** performed genomic analyses. **AE, VT** and **MA** interpreted the results. **AE** created figures and graphical content and prepared the first draft of the manuscript. All authors contributed to reviewing & editing and approving the final manuscript.

Figures

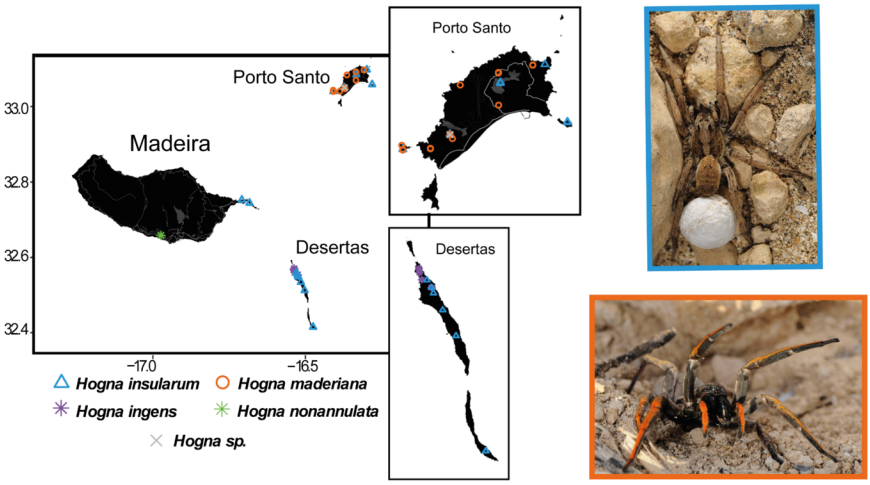

Figure 1. Distribution of the sampled localities in the Madeira archipelago (Madeira, Porto Santo and Desertas islands). Blue empty triangle: *Hogna insularum* ; orange empty circle: *Hogna maderiana* ; purple asterisk: *Hogna ingens* ; green asterisk:*Hogna nonannulata* ; grey cross: *Hogna sp* . Photograph of*H. insularum* in blue. A female specimen with ootheca in the field. Photograph credit Pedro Cardoso. Photograph of *H. maderiana* in orange. A female specimen in the field. Photograph credit Pedro Cardoso.

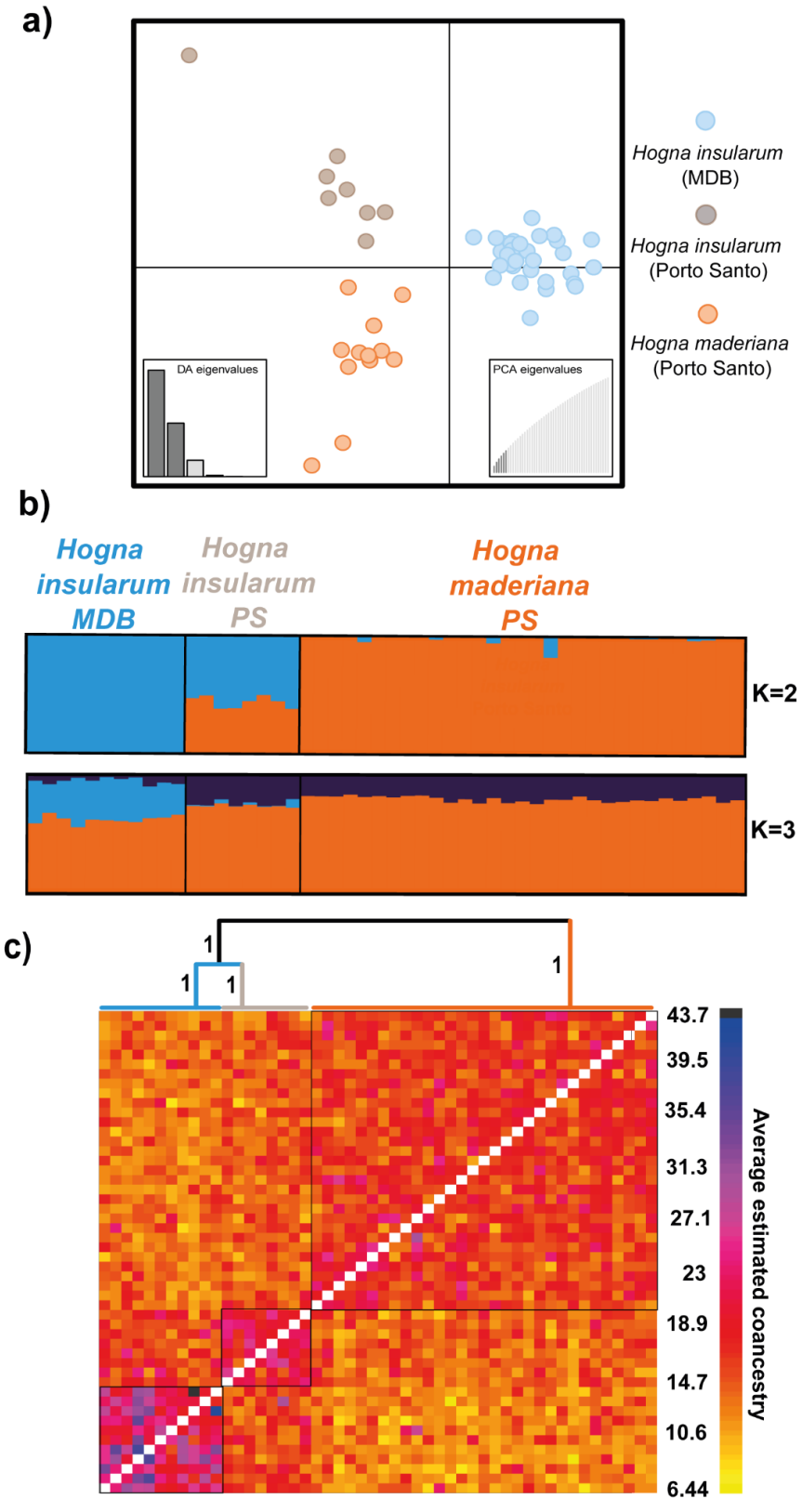

Figure 2. Population structure of *Hogna maderiana* and*H. insularum* species. a) Discriminant analysis of principal components (DAPC) for 50 specimens using 865 SNPs set. The axes represent the first two Linear Discriminants (LD). Dots represent individuals, and the clusters are presented in different colours.**b)** Population structure analysis of 50 specimens inferred using STRUCTURE software based on 865 SNPs for K = 2 and K=3. Each individual is represented by a horizontal bar proportionally coloured according to the probability of assignment to each cluster. **c)** Co-ancestry matrix based on fineRADstructure analysis of ddRAD-seq loci. The dendogram depicts clustering of individual samples based on the pairwise matrix of co-ancestry coefficients. Pairwise coefficients of co-ancestry are colour-coded from low (yellow) to high (blue) according to the scale shown on the right. Square black box demarcations identify individuals showing higher co-ancestry.

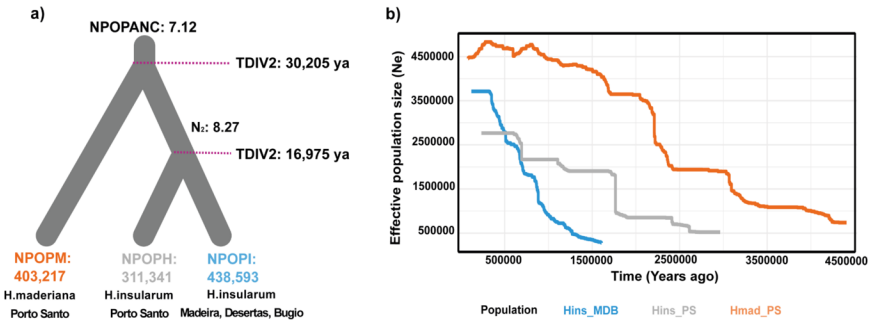

Figure 3. Demographic inference using fastsimcoal2 and Stairway plot. a) Parameters inferred from coalescent simulations with fastsimcoal2 under the best-supported demographic model (MODEL A). For each parameter, we show its point estimate. Lower and upper 95% confidence intervals are found in Table S11. Model parameters include effective population sizes (NPOPM: *H. maderiana* from Porto Santo; NPOPH: *H. insularum* from Porto Santo; NPOPI: *H. insularum* from Madeira, Desertas and Bugio), the timing of divergence (TDIV1, TDIV2) in years, and new deme sizes (NPOPANC, N2). **b)** Historical Ne of the studied populations of *H.insularum* from Madeira, Desertas and Bugio, *H.insularum* from Porto Santo and*H.maderiana* from Porto Santo, inferred using Stairway plot 2. Panels show the median of effective population size (Ne) over time, estimated assuming a mutation rate of 6 × 10-9 and 1-year generation time.

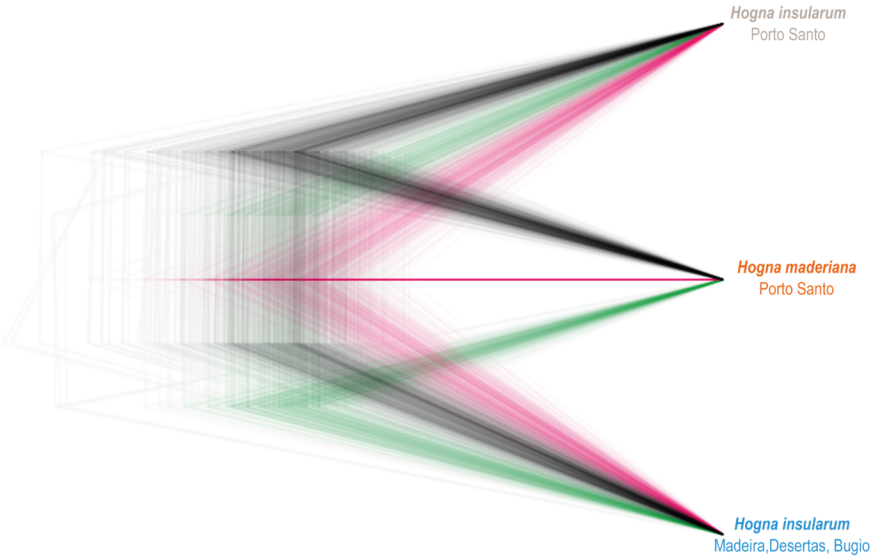

Figure 4. SNP species tree of 50 individuals assigned to three groups: *H. insularum* from Madeira, Desertas and Bugio, *H. insularum* from Porto Santo and *H. maderiana* from Porto Santo based on 696 unlinked SNPs . The maximum clade credibility tree and congruent trees are in black. The cloudogram produced by SNAPP represents the range of different topologies: the black tree displays the first topology, representing 45.93% of the accumulative trees. Pink (covering 35.53% of trees) and green (21.54%) trees display the second and third topologies, respectively. Higher-density areas of the tree indicate greater topology agreement.