Genomic insights into the origin of trans-Mediterranean disjunct distributions

Víctor Noguerales1,2,3 | Pedro J. Cordero2,4 | L. Lacey Knowles3 | Joaquín Ortego5

1Department of Biological Sciences, University of Cyprus, Nicosia, Cyprus
2Genetic and Cultural Biodiversity Research Group, Instituto de Investigación en Recursos Cinegéticos (IREC, CSIC-UCLM-JCCM), Ciudad Real, Spain
3Department of Ecology and Evolutionary Biology, Museum of Zoology, University of Michigan, Ann Arbor, MI, USA
4Escuela Técnica Superior de Ingenieros Agrónomos (ETSIA), Universidad de Castilla-La Mancha (UCLM), Ciudad Real, Spain
5Department of Integrative Ecology, Estación Biológica de Doñana (EBD-CSIC), Sevilla, Spain

Correspondence
Víctor Noguerales, Department of Biological Sciences, University of Cyprus. PO Box 20537, Nicosia 1678, Cyprus.
Email: victor.noguerales@csic.es

Present address
Víctor Noguerales, Island Ecology and Evolution Research Group, Institute of Natural Products and Agrobiology (IPNA-CSIC), San Cristóbal de La Laguna, Tenerife, Spain

Funding information
Spanish Ministry of Economy and Competitiveness, Grant/Award Number: CGL2011-25053, CGL2016-80742-R and CGL2017-83433-P

Handling Editor: Rosemary Gillespie

Abstract

Aim: Two main biogeographical hypotheses have been proposed to explain the Mediterranean-Turanian disjunct distributions exhibited by numerous steppe-dwelling organisms, namely (a) dispersal during the Messinian salinity crisis (∼5.96–5.33 Ma) followed by range fragmentation and vicariance, and (b) Pleistocene colonization and recent processes of population subdivision (<2 Ma). Despite the two hypotheses postulate the role of climatic alterations and changes in landmass configuration on determining such disjunct distributions, estimates of the timing of lineage diversification have not been complemented so far with spatially-explicit tests providing independent evidence on the proximate processes underlying geographical patterns of population genetic connectivity/fragmentation.

Location: Mediterranean-Turanian region.

Taxon: Saltmarsh band-winged grasshopper (Mioscirtus wagneri).

Methods: We integrate different sources of genetic (mtDNA and ddRADseq) and spatial information (configuration of emerged lands and niche modelling) to evaluate competing hypotheses of lineage diversification in the saltmarsh band-winged grasshopper, a halophile species showing a classical Mediterranean-Turanian disjunct distribution.

Results: Phylogenomic analyses reveal the presence of two North African cryptic lineages and support that trans-Mediterranean populations of the species diverged in the Pleistocene, with evidence of post-Messinian permeability of the Strait of Gibraltar to gene flow likely associated with sea level drops during glacial periods. Accordingly, spatial patterns of genetic differentiation are best explained by a scenario of population connectivity defined by the configuration of emerged landmasses and environmentally suitable habitats during glacial periods, a time when effective population sizes of the species peaked as inferred by genomic-based demographic reconstructions.

Main conclusions: Our results support post-Messinian colonization and Pleistocene diversification as the biogeographical scenario best explaining the trans-Mediterranean disjunct distributions of halophilous organisms.
1 | INTRODUCTION

Inferring the processes structuring genetic variation, through their effects on genetic drift and gene flow, and understanding their evolutionary consequences remain a central focus of both population genetics and phylogeography (Avise, 2009; Habel et al., 2015). However, despite large overlap in their theoretical bodies, these disciplines have tended to evolve independently. In part, this was due to the contrasting spatiotemporal scales at which the main evolutionary processes that they deal with occur and whose resolution has traditionally required the usage of different molecular markers (e.g., nuclear microsatellite vs. mitochondrial sequence markers; Wang, 2010). This idiosyncrasy has in part hampered our capacity to understand how the microevolutionary processes operating at population level ultimately drive speciation and lineage divergence at broader geographical and temporal scales (Papadopoulou & Knowles, 2017). Currently, the advent of high-throughput sequencing technologies and the possibility of generating vast genomic datasets has opened the door to investigate patterns of genetic variation at a wide range of evolutionary scales, blurring the boundaries of population genetics, phylogeography and phylogenetics (Rissler, 2016).

Applications of genomic data, especially when viewed through a statistical phylogeographical framework (Knowles, 2009), are essential to studies of the divergence histories of organisms across geologically dynamic and environmentally heterogeneous landscapes, such as the Mediterranean. The modern Mediterranean basin is the legacy of complex climate and geological dynamics that have taken place since the Late Miocene (Blondel & Aronson, 1999). This includes the Messinian salinity crisis (MSC, ∼5.96–5.33 Ma), a geological episode during which the Mediterranean-Atlantic seaways progressively closed, leading to a dramatic sea-level drawdown (∼1,500 m) in the Mediterranean Sea and its partial desiccation (Krijgsman et al., 1999). The resulting land bridges connecting regions from southern Europe and northern Africa provided a favourable geographical setting for the expansion of terrestrial biota throughout the Mediterranean region, especially for dry-adapted and steppe-dwelling taxa that benefited by the increased availability of open landscapes and arid environments (García-Alix et al., 2016). The reopening of the Atlantic-Mediterranean marine connection (Strait of Gibraltar) and refilling of the Mediterranean basin (∼5.33 Ma; García-Castellanos et al., 2009) led to vicariant events in numerous organism groups that had previously expanded their distributions during the MSC (Chueca et al., 2015). The long-term persistence of populations in relict patches of optimal habitat since the Messinian has been often invoked as the most plausible explanation for the current disjunct distribution of many steppe or halophilous taxa with populations at the western- and easternmost portions of the Mediterranean basin (Ribera & Blasco-Zumeta, 1998). However, an alternative biogeographical hypothesis proposes that the origin of disjunct distributions in some trans-Mediterranean organisms reflect their ability to track existing suitable habitats in more recent times (i.e., post-Messinian; Allegrucci et al., 2011). According to this hypothesis, disjunctions arose by colonization events and subsequent isolation processes linked to Quaternary climate oscillations (Kadereit & Yaprak, 2008). Sea-level drawdown (∼125 m; Litcher et al., 2010) and expansions of steppe-like environments (Kajtoch et al., 2016; Kirschner et al., 2020) during Pleistocene glacial periods might have facilitated the dispersal of steppe species across seaways and favourable habitats and, ultimately, led to the colonization of geographically distant regions that have remained isolated since then (Ortiz et al., 2007). For example, the Strait of Gibraltar, where the distance between the Iberian Peninsula and Africa shortened to several kilometers during the Pleistocene coldest stages, which might have favoured the exchange of fauna and flora between the two continents (Graciá et al., 2013). However, the east-west vicariant distributions of several Mediterranean steppe-dwelling organisms have often been attributed to pre-Quaternary migration events linked to the MSC without proper evaluation of alternative phylogeographical hypotheses (Ortego et al., 2009; Ribera & Blasco-Zumeta, 1998).

Clearly important insights can be gained from integrating precise information on the timing of population- and lineage-level divergence, detailed reconstructions of the demographic fate of the populations, and formal tests of alternative spatially-explicit scenarios of gene flow representing expectations of population connectivity under contrasting biogeographical hypotheses (Papadopoulou & Knowles, 2015).

Here, we apply an integrative and hypothesis testing framework to distinguish between competing hypotheses about the Mediterranean-Turanian disjunct distribution of the saltmarsh band-winged grasshopper, Mioscirtus wagneri (Eversmann, 1859) (Orthoptera: Acrididae). This species is the only representative of the monotypic genus Mioscirtus Saussure, 1888 (Cigliano et al., 2020) and its vast range spans from the Iberian Peninsula to central Asia (Cordero et al., 2007). In the Mediterranean region it forms highly fragmented populations in Iberia, northwestern Africa (Morocco, Algeria and Tunisia) and the Middle East (Turkey, Israel, Palestine and Jordan), being absent from France, Balkan and Italian Peninsulas, northeastern Africa (Egypt and Libya), and all islands with the exception of Cyprus (Cigliano et al., 2020; Cordero et al., 2007; Figure 1). The saltmarsh band-winged grasshopper is a specialist species intimately associated to certain halophilic plants (Suaeda sp.) from the family Amaranthaceae on which it depends for feeding and shelter (Cordero et al., 2007). This restricted ecological requirement limits the distribution of the species to saline lowlands, including coastal marshes and inland endorheic lagoons and steppes. Three
subspecies of the saltmarsh band-winged grasshopper have been described so far based on subtle phenotypic variations and differences in body size (Cigliano et al., 2020): *M. w. wagneri* (Eversmann, 1859) (type locality: Asia-Temperate, Middle Asia, Turkmenistan), *M. w. rogenhoferi* (Saussure, 1888) (type locality: Asia-Temperate, Western Asia, Iraq) and *M. w. magrebi* Fernandes, 1968 (type locality: Africa, Northern Africa, Algeria). It has been hypothesized that the diversification of the saltmarsh band-winged grasshopper in the Mediterranean region could be linked to the MSC. This would suggest that its present-day disjunct distribution may constitute the relics of a wider range during the Late Miocene that has since experienced gradual fragmentation (Ortego et al., 2009). However, the
tempo and mode of divergence between the different populations and putative subspecies remain unresolved due to the lack of a comprehensive approach covering the entire Mediterranean distribution range of the species and formal testing of alternative biogeographical hypotheses.

In this study, we focus on the trans-Mediterranean distribution range of the saltmarsh band-winged grasshopper, including populations separated > 4,000 km that encompass the Mediterranean and Irano-Turanian regions and comprise the three putative subspecies (Cigilano et al., 2020) (Figure 1). Specifically, we integrate different sources of genetic (mtDNA and ddRADSeq) and spatial information (configuration of emerged lands and niche modelling) to evaluate competing phylogeographical hypotheses and determine whether lineage diversification is a consequence of dispersal during the MSC followed by range fragmentation and vicariance (~5.33 Ma) (Ribera & Blasco-Zumeta, 1998) or, alternatively, resulted from post-Messinian colonization and Pleistocene population subdivision (<2 Ma) (Graciá et al., 2013). First, we used coalescent-based methods to reconstruct the phylogenetic relationships among populations, estimate the timing of lineage divergence, and elucidate whether these are congruent with either Messinian or Pleistocene genetic fragmentation. Second, we integrated bathymetric information and environmental niche modelling (ENM) to determine the spatial configuration of emerged landmasses and climatically suitable areas at different time periods (Messinian, Pleistocene and present day) and generate alternative scenarios of population connectivity, the fit of which to our genetic data we evaluate based on the timing of population divergence estimated in the previous step. Finally, we examined the geographical distribution of genetic variation and reconstructed changes in effective population size through time in order to infer the demographic responses of the different populations to past climate oscillations, to identify pulses of population expansion/contraction and to link them with the historical processes underlying genetic fragmentation and lineage formation.

2 | MATERIALS AND METHODS

2.1 | Sample collection

We collected 180 individuals from 20 populations of the saltmarsh band-winged grasshopper distributed throughout its Mediterranean distribution range, including populations from Spain, Portugal, Morocco, Tunisia, Turkey and Jordan (Figure 1; Table S1).

2.2 | Genetic data

We extracted DNA and amplified and sequenced fragments of the 12S and 16S rRNA mitochondrial genes as detailed in González-Serna et al. (2018). After sequence alignment, trimming and editing, gene fragments had 395 bp for 12S and 460 bp for 16S. From the 180 specimens, we selected 7–9 individuals per population (144 samples in total; Table S1) to be processed into three genomic libraries following the double-digestion restriction-fragment-based procedure (ddRADseq) described in Peterson et al. (2012), with minor modifications detailed in Noguerales et al. (2018). Each library was sequenced in a single-read 101-bp lane on an Illumina HiSeq2500 platform at The Centre for Applied Genomics (Toronto, Canada). Raw sequences were demultiplexed and preprocessed using STACKS 1.35 (Catchen et al., 2013) and assembled using PYRAD 3.0.66 (Eaton, 2014). Appendix S1 provides all details on genomic data assembling and filtering.

2.3 | Phylogenetic analyses: mtDNA data

We used mitochondrial DNA data and BEAST 1.8.0 (Drummond et al., 2012) to reconstruct the phylogenetic relationships among the different populations and estimate the timing of lineage divergence. We applied a HKY + I model of sequence evolution for both gene fragments as determined in JMODELTEST2 (Darriba et al., 2012). We assumed a normal distributed substitution rate of 0.049 (SD = 0.0008) per site per million years for the 16S rRNA gene fragment (Papadopoulou et al., 2010) and applied a continuous-time Markov chain (CTMCs) model for the 12S rRNA, as no explicit priors for the substitution rate for this gene fragment are available (e.g., González-Serna et al., 2018). A strict clock and a constant demographic model was used for phylogenetic reconstructions as determined by model testing using a generalized stepping-stone sampling approach (Baele et al., 2016). We ran the analysis with two independent MCMC chains of 100 million generations each, sampling every 10,000 generations, and discarding the first 10% as burn-in.

2.4 | Phylogenetic analyses: Genome-wide nuclear data

First, we reconstructed the phylogenetic relationships among populations using genome-wide SNP data and the coalescent-based method implemented in SVDQUARTETS (Chifman & Kubatko, 2014). We ran SVDQUARTETS exhaustively evaluating all possible quartets and performing nonparametric bootstrapping with 100 replicates for quantifying uncertainty in relationships. Second, we used BPP 4.0 to estimate the timing of lineage and population divergence (Flouri et al., 2018). The phylogenetic tree inferred using SVDQUARTETS was fit as the fixed topology in BPP analyses (option A00). The loci file from PYRAD was edited, converted into a BPP input file and filtered using custom scripts (J-P. Huang, https://github.com/airbugs/; Huang et al., 2020). Due to high computational burden, branch length estimation was inferred using five datasets consisting of 100, 200, 300, 400 and 500 randomly chosen loci (of the 1,518 variable loci recovered) to confirm the consistency of the results (e.g., Huang et al., 2020). We applied an automatic adjustment of fine-tune parameters, allowing swapping rates to range between 0.30 and 0.70, and set the diploid option to indicate that the input sequences are unphased (for further details on the diploid option in BPP, see Flouri et al., 2018). Analyses were...
run for 500,000 generations, sampling every 10 generations, after a conservative burn in of 500,000 generations.

### 2.5 | Population genetic structure

We inferred population genetic structure using genome-wide SNP data and the model-based clustering algorithm implemented in FASTSTRUCTURE 1.0 (Raj et al., 2014). We performed 10 independent runs for each of the different possible K genetic clusters (from K = 1 to K = 10) using the simple prior and a convergence criterion of $1 \times 10^{-7}$. We assessed the number of genetic clusters that best describes our data by estimating the metrics $K^*$, the value of K that maximizes log-marginal likelihood lower bound (LLBO) of the data, and $K^\prime$, the smallest number of model components explaining at least 99% of cumulative ancestry contribution in our sample (Raj et al., 2014).

### 2.6 | Landscape genetic analyses

We implemented a spatially explicit approach based on circuit theory to model gene flow under alternative hypothetical scenarios of contemporary and historical population connectivity (McRae, 2006). Specifically, we used CIRCUITSCAPE 4.0.5 (McRae & Beier, 2007) to calculate resistance distance matrices between all pairs of populations based on isolation by resistance (IBR) scenarios defined according to (i) the spatial configuration of emerged lands during the present-day, last glacial maximum (LGM) and Messinian (IBR$_{\text{GEO}}$; see Appendix S2 for details) and (ii) the spatial configuration of climatically suitable habitats as inferred from projections of an environmental niche model (ENM) to present-day and LGM bioclimatic conditions (IBR$_{\text{CLI}}$; see Appendix S3 for details). In order to identify the resistance value for sea water that best explains observed estimates of genetic differentiation, we explored a range of resistance values (10–1,000,000) for this landscape feature plus a scenario that considered it as an impassable barrier to dispersal. For IBR$_{\text{GEO}}$ scenarios, pixels coded as “emerged land” were assigned a resistance value of 1. For IBR$_{\text{CLI}}$ scenarios, logistic environmental suitability scores ($x$) yielded by projections of the ENM were transformed as 1-($x$), so that smaller pixel values offer lower resistance to gene flow. Additionally, we tested a null model of isolation-by-distance (IBD$_{\text{NULL}}$) by generating a completely “flat landscape” scenario based on a map in which all pixels (including sea water and emerged lands) have a fixed resistance (=1) value. We used multiple matrix regressions with randomization (MMRR; Wang, 2013) to test all resistance matrices against matrices of population genetic differentiation ($F_{ST}$) calculated in ARLEQUIN 3.5 (Excoffier & Lischer, 2010) for mtDNA and genome-wide SNP data.

### 2.7 | Genetic diversity

We analysed spatial clines of genetic diversity ($\pi$, nucleotide diversity) estimated for both mtDNA and genome-wide nuclear data using DNASP 5.10/6.12 (Rozas et al., 2017). We tested the association between genetic diversity ($\pi$) and geography (latitude and longitude) using generalized linear models (GLMs) as implemented in the R package ‘lm4’. GLMs were constructed using a Gaussian error distribution, an identity link function and a weighted least square (WLS) method, where weight equals the sample size for each population (Bates et al., 2015).

### 2.8 | Demographic inference

We inferred the demographic history of each population using STAIRWAY PLOT 2.0, which implements a flexible multi-epoch demographic model to estimate changes in effective population size ($N_e$) over time using the site frequency spectrum (SFS) (Liu & Fu, 2015). We used the script easySFS.py (I. Overcast, https://github.com/isaacovercast/easySFS) to calculate a folded SFS for each population. To avoid the effects of linkage disequilibrium and remove all missing data for the calculation of the SFS, we considered a single SNP per locus and retained only loci present in ~66% of individuals. Populations with <7 individuals (PHU and MER) were excluded from STAIRWAY PLOT analyses. Demographic reconstructions in STAIRWAY PLOT were run fitting a mutation rate of $2.8 \times 10^{-9}$ per site per generation (Keightley et al., 2014), considering a one-year generation time (Cordero et al., 2007), and performing 200 bootstrap replicates to estimate 95% confidence intervals.

### 3 | RESULTS

#### 3.1 | Genetic data: mtDNA data

We found 16 and 21 unique haplotypes for 12S and 16S gene fragments, respectively (Table S1). In particular, 11 and 14 haplotypes for 12S and 16S gene fragments, respectively, were exclusively found in a single population (Table S1). The remaining haplotypes were shared among individuals belonging to nearby populations located within the same geographical region (Table S1). Remarkably, individuals from Iberian (TIN) and African (HOC) populations located at both sides of the Strait of Gibraltar shared the same 12S and 16S haplotypes (Table S1).

#### 3.2 | Genetic data: Genome-wide nuclear data

Four individuals (one individual from PHU and three individuals from MER) recovered a very low number of reads (<10,000) and were discarded from subsequent analyses. After filtering and assembly steps in PYRAD, each remaining individual ($n = 140$) retained on average 1.83 million sequence reads ($SD = 0.66M$), which represents 92.50% ($SD = 2.29\%$) of their initial number of reads (Figure S1). On average, we recovered 49,415 loci ($SD = 9,305$; range = 23,170–79,257) per individual.
3.3 | Phylogenetic analyses

Phylogenetic inference based on both mitochondrial (BEAST) and genome-wide nuclear (svdquartets) data consistently supported two highly divergent lineages only represented in the populations located in western Morocco (ZIM) and southern Tunisia (KEB) (Figure 2). The remaining populations were grouped into four clades corresponding to the Middle East, Africa along with southwestern Iberia, northeastern Iberia, and central-southeastern Iberia populations (Figure 2). All analyses showed consistently that Iberian populations were not reciprocally monophyletic, as populations from southwestern Iberia (CAS and TIN) were placed within the northern Africa clade. We found a discordance between mitochondrial and nuclear genealogies in the phylogenetic position of the northeastern Iberian clade (Ebro Valley): mtDNA data supported populations from northeastern Iberia as the most external lineage of the trans-Mediterranean clade, whereas nuclear data placed them as a sister lineage to the central-southeastern Iberian clade (Figure 2).

3.4 | Divergence time estimation

Estimates of divergence time based on BEAST analyses for mtDNA revealed that the ZIM + KEB clade and the remaining lineages diverged from a shared common ancestor during the Pliocene (~4.5 Ma; Figure 2). According to bpp analyses based on genome-wide nuclear data, the most ancient split was estimated around 0.01032 \( \tau \) units (Figure 3). Assuming a genomic mutation rate of 2.8 \( \times 10^{-9} \) per site per generation (Keightley et al., 2014) and a 1-year generation time (Cordero et al., 2007), the oldest diversification event may have taken place during the Early Pleistocene, around 1.84 Ma (\( \tau = 2\mu t \), being \( \mu \) the mutation rate and \( t \) the divergence time). The subsequent split between ZIM and KEB populations was dated to occur during the Early Pleistocene (mtDNA data: ~1.7 Ma; genome-wide nuclear data: ~1.08 Ma, 0.00608 \( \tau \) units). The diversification of the remaining clades triggered around the Middle Pleistocene (mtDNA data: ~1.0 Ma; genome-wide nuclear data: ~0.21 Ma, 0.00117 \( \tau \) units) and finished with the split of northern Morocco and southwestern Iberia populations in the Late Pleistocene (mtDNA data: ~100 ka;).

**FIGURE 2** (a, c) Phylogenetic relationships and (b) genetic structure of populations of the saltmarsh band-winged grasshopper (Mioscirtus wagneri). Panel (a) shows the maximum clade credibility tree for mitochondrial data (16S and 12S gene fragments) as inferred in BEAST. Estimates of divergence time for the main clades (median and lower and upper 95% highest posterior density, in brackets) and branch support values (*) are indicated. Panel (b) shows population genetic structure for the most likely number of clusters (K = 6) as inferred by FASTSTRUCTURE using genome-wide SNP data. Panel (c) shows the species-tree inferred by svdquartets using genome-wide SNP data. Bootstrapping support values are indicated on the nodes (*=100%). Putative membership of the populations to the different subspecies and population codes as in Table S1 [Colour figure can be viewed at wileyonlinelibrary.com]
genome-wide nuclear data: ∼47.5 ka, 0.00026 \( \tau \) units) (Figure 3). Inferences from \text{bpp} were consistent across analyses based on dataset considering a different number of loci (Figure S2).

3.5 | Population genetic structure

Results from \text{faststructure} analyses indicated \( K = 6 \) as the most likely number of genetic clusters (\( K_x^* = 6; K_y^* = 6 \)). The inferred genetic groups were consistent with phylogenetic inferences (Figure 2b). The highly divergent lineages from ZIM and KEB populations of North Africa were included in two distinct genetic clusters (Figure 2b). The other four genetic clusters corresponded to the rest of North African populations plus southwestern Iberian populations, and populations from northeastern Iberia, central-southeastern Iberia and the Middle East (Figure 2b).

3.6 | Landscape genetic analyses

Pairwise \( F_{ST} \) matrices calculated for mitochondrial and genome-wide nuclear data (Table S2) were significantly correlated (Mantel test, \( r = 0.581, p < .001 \)). Resistance distance matrices estimated under most scenarios were positively correlated with population genetic differentiation (\( F_{CT} \)) at both nuclear and mitochondrial levels (Table S3). However, multiple matrix regression with randomization (MMRR) analyses consistently supported that the scenario of population connectivity based on the spatial configuration of emerged lands and climatically suitable areas during the LGM (IBRCLI-LGM) was the best fit to our data and the only one retained into the final models (Table 1). This result was consistent across both mitochondrial and genome-wide nuclear data and when including and excluding from the analyses the highly divergent ZIM and KEB populations (Table 1).

3.7 | Genetic diversity

Mitochondrial and genome-wide nuclear genetic diversity were not significantly correlated (Pearson’s \( r = −0.127, p = .592 \)). The highest levels of mitochondrial genetic diversity were found in ZIM and KEB populations followed by the three populations from northeastern Iberia, whereas several populations from central, southeastern and southwestern Iberia, northern Morocco and Jordan exhibited no mitochondrial genetic variation (\( \pi = 0 \); Table S1). Accordingly, we did not find any significant relationship between mitochondrial genetic diversity and longitude and latitude either analysing all populations (longitude, \( t = −0.360, p = .723 \); latitude, \( t = 0.030, p = .977 \) or
excluding ZIM and KEB populations (longitude, t = 0.179, p = .860; latitude, t = 1.919, p = .074). With regards to genome-wide nuclear data, populations from the Middle East (TUZ and DEA) showed the highest levels of genetic diversity, whereas populations from central and northeastern Iberia presented the lowest estimates (Figure 1; Table S1). Nuclear genetic diversity was positively correlated with latitude (β = 1.16 × 10^{-4}, t = 8.981, p < .001; latitude: β = -4.74 × 10^{-4}, t = -8.408, p < .001).

### 3.8 | Demographic inference

**Stairway plot** analyses revealed strong declines in N_e for all populations after the end of the last glacial period (~10–20 ka; Figure 4). Phylogenetically closer populations shared more similar demographic histories in terms of the trends and magnitude of population size changes. In line with analyses of nuclear genetic diversity, populations from central, southeastern and northeastern Iberian Peninsula exhibited on average the lowest historical estimates of N_e. Middle East populations presented the highest estimates, and populations from northern Africa and southwestern Iberia had intermediate values (Figure 4).

### 4 | DISCUSSION

Our phylogenetic analyses and the estimated timing of lineage splits, complemented with evidence provided by spatially-explicit testing of alternative scenarios of gene flow, support a post-Messinian colonization and Pleistocene genetic fragmentation as the most plausible biogeographical hypothesis to explain the trans-Mediterranean disjunct distribution of the saltmarsh band-winged grasshopper. Intriguingly, our analyses revealed that the species exhibits a complex phylogeographical structure (Figure 2), with three main clade-genetic events and a deep genetic split between parapatric northern Africa lineages that indicate cryptic processes of allopatric divergence and call upon a taxonomic revision of this monotypic genus (Cigliano et al., 2020). Shared genetic lineages between southwestern Iberia and the Maghreb region support the recurrent permeability to gene flow of the Strait of Gibraltar during the Pleistocene and point to the important role of sea-level fluctuations linked to glacial cycles in promoting biotic exchanges between Europe and Africa.

### 4.1 | Ancient cryptic lineages in the Maghreb region

One of the major questions in biogeography research on the western Palearctic realm is elucidating the processes underlying the disjunct distributions of many Mediterranean organisms (Sammartin, 2003). We found that genetic variation in the saltmarsh band-winged grasshopper is organized into three main evolutionary lineages whose divergence most likely took place after the end of the Messinian age (Figures 2-3). The first two sister lineages are distributed in western Morocco (ZIM) and southern Tunisia (KEB), regions separated > 1,600 km apart from each other. These lineages present a parapatric distribution with respect to their geographically closest populations that belong to the third and more widespread trans-Mediterranean clade. This intriguing phylogeographical structure contrasts with the prevailing biogeographical pattern in the Maghreb,
where the foremost genetic discontinuities have been linked to the Moulouya River (northern Morocco) and the Kabylia region (northern Algeria) in numerous terrestrial organisms (e.g., Beddek et al., 2018). The genetic discontinuities and disjunct distributions of these lineages likely represent processes of allopatric isolation driven by inland landscape changes occurring along the Pliocene and Pleistocene throughout the Maghreb region. Since the Early Pliocene, the Mediterranean region experienced a progressive aridification-cooling climatic shift that led to the expansion of steppe formations (Jiménez-Moreno et al., 2010). This shift to a drier climate regime was particularly notable in North Africa (Griffin, 2002), which could have facilitated the establishment and expansion of halophilous vegetation along vast endorheic lacustrine areas (i.e., sabkhas and chotts) located across the Maghreb and whose development dates back to the Late Messinian associated to local tectonic dynamics (Capella et al., 2018). These geomorphological and climate alterations could have synergistically contributed to create refugial areas for the species in northern Africa, leading to long-term isolation processes that might have impeded gene flow (e.g., via the evolution of reproductive isolation or lack of secondary contact) with its parapatric sister lineage widely distributed across the Mediterranean.

4.2 | Pleistocene divergence throughout the Mediterranean region

Focusing on the trans-Mediterranean clade, our phylogenetic inferences support an east-to-west divergence that led to the formation

**FIGURE 4** Demographic history of the studied populations inferred using STAIRWAY PLOT. Lines represent the median estimate of the effective population size ($N_e$). Axes are logarithmically scaled and populations are grouped into the different panels according to the results of phylogenetic and clustering analyses. Upper panels represent temperature anomaly (ΔT °C) in the Late Quaternary as estimated from the EPICA (European Project for Ice Coring in Antarctica) Dome C ice core (Jouzel et al., 2007). Highlighted periods show the extent of the last glacial period (~110–11.7 ka), last glacial maximum (LGM: ~21–19 ka), and Holocene (~11.7 ka to present). Population codes as in Table S1 [Colour figure can be viewed at wileyonlinelibrary.com]
of three main lineages distributed in the Middle East, northern Africa and Iberian Peninsula with a distribution gap in Egypt and Libya. As consistently estimated on the basis of both mitochondrial and nuclear data (Figures 2-3), the split of these clades began around the Middle Pleistocene (~1.0–0.21 Ma) and continued until the Late Pleistocene (<100 ka). These findings are concordant with phylogeographical inferences (i.e., Quaternary westward expansion) for other trans-Mediterranean halophitic and steppe-dwelling taxa (Escudero et al., 2010; Kadereit & Yaprak, 2008; Pérez-Collazos et al., 2009; Weising & Freitag, 2007). Our analyses also revealed that the three main lineages previously described within the Iberian Peninsula on the basis of mitochondrial (Ortego et al., 2009) and nuclear microsatellite (Ortego et al., 2010) data are not reciprocally monophyletic (Figures 2-3), with populations from southwestern Iberia being nested within the northern Africa clade (see Husemann et al., 2014). In turn, we found a mitonuclear discordance in the phylogenetic position of the northeastern Iberian clade, which might have been caused by the interplay between historical secondary contact and different evolutionary and demographic processes (see Toews & Brärsford, 2012). These results point to two independent colonization events of the Iberian Peninsula during the Pleistocene followed by allopatric divergence from their respective African ancestors. Thus, our results strongly support that the Strait of Gibraltar was a permeable barrier to dispersal, probably in association with sea-level changes linked to Quaternary climatic oscillations (Graciá et al., 2013). The sea-level during Pleistocene glacial periods has been estimated to be approximately 125 meters lower than at present in the western Mediterranean region (Rohling et al., 2017), a fact that would have reduced the distance between European and African coasts to less than 5–7 km at the Camarín Sill area (Luján et al., 2011). This contributed to the emergence of small islands and shoals facilitating “stepping-stone” dispersal and the sporadic exchange of terrestrial faunas across the Strait of Gibraltar (Cosson et al., 2005; see also Ortiz et al., 2007). Accordingly, our landscape genetic analyses supported that genetic differentiation among populations is best explained by the configuration of shorelines and suitable habitats during Pleistocene glacial periods and when sea surface is modelled to offer much more resistance to movement than emerged lands but without acting as impassable barrier to gene flow (Table 1 and Table S3).

4.3 Genetic diversity and past demographic history

Demographic reconstructions revealed that all analysed populations have experienced population size declines after the LGM coinciding with the beginning of the current interglacial stage (Figure 4), a period during which climatic conditions likely resulted in the progressive shrink of suitable open habitats for the species (Weising & Freitag, 2007). Palynological studies have evidenced that extensive areas of the Mediterranean region were vegetated by steppe-like formations during glacial periods that became progressively replaced by temperate forests during interglacial stages (Carrirón et al., 2012; Sánchez-Góñi et al., 1999). Consequently, the confinement of the saltmarsh band-winged grasshopper in refuges of suitable habitat during unfavourable periods may have led to processes of allopatric divergence along the Quaternary as reported for steppe-like and halophytic species presenting similar environmental requirements (Kajtoch et al., 2016; Kirschner et al., 2020). Demographic inferences based on genomic data contrast with the predictions of our ENM, which suggests that the extent of climatically suitable habitats for the species have tended to increase since the LGM (Figure 1f). This might reflect the difficulty of climate-based niche modelling to identify refugial areas in highly specialized species tightly linked to scattered and small-size habitat patches (i.e., salt-marshes, sabkhas and chotts) for which spatially-explicit information about their past distribution and connectivity is not currently available (González-Serna et al., 2019). The very limited dispersal capacity of the saltmarsh band-winged grasshopper documented in previous studies (Ortego et al., 2015) points to postglacial contraction of suitable habitats and regional extinction in northeastern Africa, rather than long-distance dispersal, as the most likely cause of the species’ distribution gap in Egypt and Libya (Cigliano et al., 2020). Accordingly, our data showed a general pattern of genetic erosion, the total lack of genetic variation at mitochondrial level in many populations, and an east-to-west and south-to-north decline of nuclear genetic diversity (Figure 1b; see also Escudero et al., 2010). This pattern might reflect the strong fragmentation of most contemporary populations, historical declines revealed by demographic reconstructions, and founder events resulted from serial colonization from the putatively ancestral range located in the Middle East (Pérez-Collazos et al., 2009).

4.4 Taxonomic implications

Our study revealed cryptic diversity within the saltmarsh band-winged grasshopper, as well as incongruence between current subspecies designations and the inferred phylogenetic relationships (Cigliano et al., 2020). For example, the populations ZIM and KEB, putatively belonging to the subspecies *M. w. maghrebi*, accumulated more genetic divergence than the remaining populations assigned to the three subspecies (Figure 2). This finding highlights the need for evaluating the taxonomic status of these two North Africa populations as potentially distinct taxa, which would ideally require additional population sampling, the identification of potential contact zones with the other Maghrebian lineage to determine whether the two are reproductively isolated, and comprehensive morphological analyses that consider diagnostic phenotypic traits (e.g., Huang et al., 2020; Noguerales et al., 2018). Our phylogenetic analyses also shed light on the controversial taxonomic status of Iberian populations (Cordero et al., 2007), supporting the existence of two putative subspecies: *M. w. maghrebi* in southwestern Iberia (Fernandes, 1968) and
M. w. wagneri in central, southeastern and northeastern Iberia (Cordero et al., 2007). Finally, our study clarifies the phylogenetic position of Turkish populations, where subspecies M. w. wagneri and M. w. rogenhoferi have been indistinctly considered in the literature (Naskrecki & Únal, 1995; Únal, 2011). In this sense, our analyses supported the close phylogenetic relationship between Turkish and Jordanian populations (Figure 2), which is coherent from a geographical point of view and agrees with the assignment of Middle East populations to the subspecies M. w. rogenhoferi, as reported in most Orthoptera inventories from the region (Katbeh-Bader, 2001; Naskrecki & Únal, 1995).

5 | CONCLUSIONS

Our study demonstrates that the disjunct distribution of the most widespread lineage of the saltmarsh band-winged grasshopper across the Mediterranean basin was shaped by its capacity to track and colonize suitable habitats during the Pleistocene (Graciá across the Mediterranean basin was shaped by its capacity to widespread lineage of the saltmarsh band-winged grasshopper across the Mediterranean basin was shaped by its capacity to track and colonize suitable habitats during the Pleistocene (Graciá et al., 2013), rather than as a consequence of fragmentation and long-term persistence of relict populations after range expansions linked to the MSC (Ribera & Blasco-Zumeta, 1998). The cohesiveness of Moroccan and southern Iberia populations evidences the post-Messinian permeability of the Strait of Gibraltar, while the presence of highly divergent cryptic lineages in northern Africa highlights the pivotal role of the Maghreb as a diversification area during the Pliocene-Pleistocene (Husemann et al., 2014). Overall, our study emphasizes the power of combining phylogenetic inference with spatially-explicit testing of alternative biogeographical scenarios to unravel hidden diversification patterns and gain insights into the proximate processes underlying the origin of disjunct distributions.

ACKNOWLEDGEMENTS

We are grateful to the people at the Knowles Lab for their advice during the preparation of genomic libraries and valuable help in data analysis. We thank José Miguel Aparicio and Nabil Amor for their help during field sampling and two anonymous referees for their constructive and valuable comments on an earlier version of the manuscript. Centro de Supercomputación de Galicia (CESGA) and Doñana’s Singular Scientific-Technical Infrastructure (ICTS-RBD) provided access to computer resources. This work was funded by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (grants CGL2011-25053, CGL2016-80742-R, and CGL2017-83433-P).

DATA AVAILABILITY STATEMENT

Raw Illumina reads have been deposited at the NCBI Sequence Read Archive (SRA) under BioProject PRJNA663622. Mitochondrial DNA sequences for the 12S and 16S gene fragments are deposited in GenBank with accession numbers MW018492-MW018671 and MW018175-MW018354, respectively. Input files for all analyses are available for download from the Dryad Digital Repository (https://doi.org/10.5061/dryad.qfttdz0fk).

ORCID

Victor Noguerales https://orcid.org/0000-0003-3185-778X
Pedro J. Cordero https://orcid.org/0000-0002-1371-8009
L. Lacey Knowles https://orcid.org/0000-0002-6567-4853
Joaquin Ortego https://orcid.org/0000-0003-2709-429X

REFERENCES

Cordero, P. J., Llorente, V., & Aparicio, J. M. (2007). New data on morpho-


**BIOSKETCH**

Victor Noguerales is interested in understanding the ecological and evolutionary processes shaping spatial patterns of biological diversity. His research harnesses the power of genomic data to respond questions on how genetic variation arises and is maintained across different levels of biological organization.

Author contributions: All authors conceived and designed the study and analyses. VN, PJC and JO collected the samples. VN performed the lab work and analysed the data guided by JO, who performed the ENMs. VN wrote the manuscript with help of JO and inputs from PJC and LLK.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.