

RESEARCH ARTICLE



Unravelling plant diversification: Intraspecific genetic differentiation in hybridizing *Anacyclus* species in the western Mediterranean Basin

A. Bruno Agudo¹ | F. Xavier Picó² | Rubén G. Mateo^{3,4} | Arnald Marcer^{5,6} |
Rubén Torices⁷ | Inés Álvarez¹

¹Departamento de Biodiversidad y Conservación, Real Jardín Botánico (RJB), Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain

²Departamento de Ecología y Evolución, Estación Biológica de Doñana (EBD), Consejo Superior de Investigaciones Científicas (CSIC), Sevilla, Spain

³Departamento de Biología (Botánica), Universidad Autónoma de Madrid, Madrid, Spain

⁴Centro de Investigación en Biodiversidad y Cambio Global (CIBC-UAM), Universidad Autónoma de Madrid, Madrid, Spain

⁵CREAF, E 08193, Bellaterra (Cerdanyola del Vallès), Catalonia, Spain

⁶Universitat Autònoma de Barcelona, E 08193, Bellaterra (Cerdanyola del Vallès), Catalonia, Spain

⁷Área de Biodiversidad y Conservación, Departamento de Biología y Geología, Física y Química Inorgánica, Universidad Rey Juan Carlos, Móstoles, Madrid, Spain

Correspondence

Inés Álvarez, Departamento de Biodiversidad y Conservación, Real Jardín Botánico (RJB), Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain.
Email: ines@rjb.csic.es

Abstract

Premise: The interfertile species *Anacyclus clavatus*, *A. homogamos*, and *A. valentinus* represent a plant complex coexisting in large anthropic areas of the western Mediterranean Basin with phenotypically mixed populations exhibiting a great floral variation. The goal of this study was to estimate the genetic identity of each species, to infer the role of hybridization in the observed phenotypic diversity, and to explore the effect of climate on the geographic distribution of species and genetic clusters.

Methods: We used eight nuclear microsatellites to genotype 585 individuals from 31 populations of three *Anacyclus* species for population genetic analyses by using clustering algorithms based on Bayesian models and ordination methods. In addition, we used ecological niche models and niche overlap analyses for both the species and genetic clusters. We used an expanded data set, including 721 individuals from 129 populations for ecological niche models of the genetic clusters.

Results: We found a clear correspondence between species and genetic clusters, except for *A. clavatus* that included up to three genetic clusters. We detected individuals with admixed genetic ancestry in *A. clavatus* and in mixed populations. Ecological niche models predicted similar distributions for species and genetic clusters. For the two specific genetic clusters of *A. clavatus*, ecological niche models predicted remarkably different areas.

Conclusions: Gene flow between *Anacyclus* species likely explains phenotypic diversity in contact areas. In addition, we suggest that introgression could be involved in the origin of one of the two *A. clavatus* genetic clusters, which also showed ecological differentiation.

KEYWORDS

Anacyclus, Anthemideae, Asteraceae, ecological niche modeling, genetic structure, hybridization, intraspecific differentiation, Mediterranean Basin

During the diversification process, the ecological niche of a lineage (i.e., the population component of the niche corresponding to the distribution of individuals across environments within a region; Ricklefs, 2010) is expected to become increasingly occupied because of progressive differentiation and specialization to new environmental conditions. This complex phenomenon is molded by

niche-based processes normally acting in concert, such as environmental filtering, biotic interactions, and trade-offs (Chase and Myers, 2011), differentially affecting populations across varying spatial and temporal scales. During a lineage's diversification, gene flow may occur between locally adapted populations of either the same species or closely related ones with different degrees of incomplete reproductive isolation

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *American Journal of Botany* published by Wiley Periodicals LLC on behalf of Botanical Society of America.

(Rieseberg, 1997; Arnold, 1997, 2006; Abbott, 2017). A continuum of phenotypic variation often results, hindering the delimitation of species (Eckhart et al., 2004; Streisfeld and Kohn, 2005; Brennan et al., 2009; Wachowiak et al., 2015; Sung et al., 2018; Bresadola et al., 2019) and adding complexity to studies to disentangle the diversification process. In these cases, however, the study of phenotypic variation of reproductive characters in plants is particularly interesting because of the direct implications on their fitness (Endler, 1986; Holsinger, 2000; Fenster et al., 2004; Harder and Barrett, 2006; Armbruster et al., 2009; Broz et al., 2017). Besides affecting fitness, the suit of reproductive traits, such as flower shape, phenology, sexual and breeding system, or reward allocation will altogether shape mating patterns and thus determining the probability of gene flow between new evolving lineages.

The genus *Anacyclus* (Anthemideae, Asteraceae) encompasses eight species of herbs, mostly growing in open anthropic environments, with their center of diversification in western Mediterranean Basin (Humphries, 1979; Viales et al., 2018). From an evolutionary viewpoint, *Anacyclus* has been an object of study due to its diversity of floral phenotypes at the species (Humphries, 1981; Bello et al., 2013, 2017; Álvarez et al., 2020) and population levels (Agudo et al., 2019; Cerca et al., 2019), including various sexual systems and female flower shapes, colors, and sizes. Interspecific hybridization is likely responsible of such a high diversity of floral phenotypes because most of the *Anacyclus* species are interfertile (Humphries, 1981; Álvarez et al., 2020) and show some degree of spatial overlap in their distributions. Although the existence of current gene flow between these species has never been addressed, interspecific hybridization, which would be mediated by the active pollinator community visiting *Anacyclus* populations (Cerca et al., 2019) was supported by analyses on genome size (Agudo et al., 2019) and ribosomal sites variation (Rosato et al., 2017) of *A. clavatus* and *A. valentinus* populations. In our recent investigations including intra- and interspecific experimental crosses between *A. clavatus*, *A. homogamos*, and *A. valentinus*, Álvarez et al. (2020) found that parents, second-generation (F_2) hybrids, and backcrosses may show similar phenotypes. This highlights the relevance that interspecific hybridization may have on the phenotypic variation in this species complex. Hence, the identification of species in sympatry based solely on morphological characters may become troublesome and probably incorrect (Manzanilla et al., 2021).

In the present study, we characterized the climatic niche, estimated genetic structure, and inferred the impact of gene flow among populations of *A. clavatus*, *A. homogamos*, and *A. valentinus* across the western Mediterranean Basin. We expected to find well-differentiated climatic niches and genetic clusters for each species, except in sympatric populations, which might be represented by a combination of genetic backgrounds from the co-occurring species. We implemented a combination of two well-established methodological approaches to unravel the complexity of the lineage diversification in *Anacyclus*. On the one hand, we

estimated the genetic structure based on nuclear microsatellites previously developed for several species of the genus (Agudo et al., 2013) using Bayesian model-based clustering algorithms and complemented with ordination methods. This multispecies approach is rare in the literature because the difficulty of its implementation in cases of high genetic divergence among species (Barbará et al., 2007; Fu et al., 2016; He et al., 2017). On the other hand, based on a set of climatic factors summarizing species' niche and spatial properties, we applied ecological niche models (ENM; Guisan et al., 2017) and niche overlap analysis (Broennimann et al., 2012) to species and genetic clusters to estimate their environmental suitability and their degree of ecological niche similarity (Hespanhol et al., 2022). The climatic variables used in ENM represent putative selective pressures, which jointly with their demographic history would partially account for the geographic distribution of *Anacyclus* species in western Mediterranean Basin.

MATERIALS AND METHODS

Study system

The three study species—*A. clavatus*, *A. homogamos*, and *A. valentinus*—mainly differ in the type of peripheral florets of the capitulum (Humphries, 1979; Álvarez, 2019). *Anacyclus clavatus* presents heterogamous capitula (Figure 1), with 8–15 peripheral female flowers that display a showy white ligule (0.8–1.5 cm long). In contrast, the peripheral female flowers of *A. valentinus* may be white or yellow with a smaller ligule (0.3 cm long) usually hidden by the involucre bracts (Figure 1). Finally, in *A. homogamos*, all flowers are tubular and bisexual (Figure 1). Therefore, both *A. homogamos* and *A. valentinus* show discoid capitula, whereas *A. clavatus* displays rayed ones. These three species occur in the western Mediterranean Basin (Figure 1), inhabiting anthropogenic open places but in ecologically different areas. For example, *Anacyclus clavatus* occurs in both coastal and inland areas across the region, indicating that the species can cope with a gradient of Mediterranean subclimates mainly characterized by the severity of winter conditions from mild coastal areas to cold inland areas. In contrast, *A. valentinus* and *A. homogamos* are distributed in areas with particular environmental conditions: *A. valentinus* mainly occupies areas along the coast or with a strong coastal influence, whereas *A. homogamos* is mostly restricted to the Atlas region in Morocco (Álvarez et al., 2020).

Phenotypically mixed populations (hereinafter “mixed populations”) of *A. clavatus* and *A. valentinus* occur throughout the distribution of *A. valentinus*. These mixed populations may be composed by individuals with *clavatus*-like and *valentinus*-like phenotypes (Figure 1), and/or individuals exhibiting intermediate phenotypes between the two species (Figure 1). In mixed populations from the Middle Atlas in Morocco, *A. homogamos* may also be present (Humphries, 1979; Agudo et al., 2019).

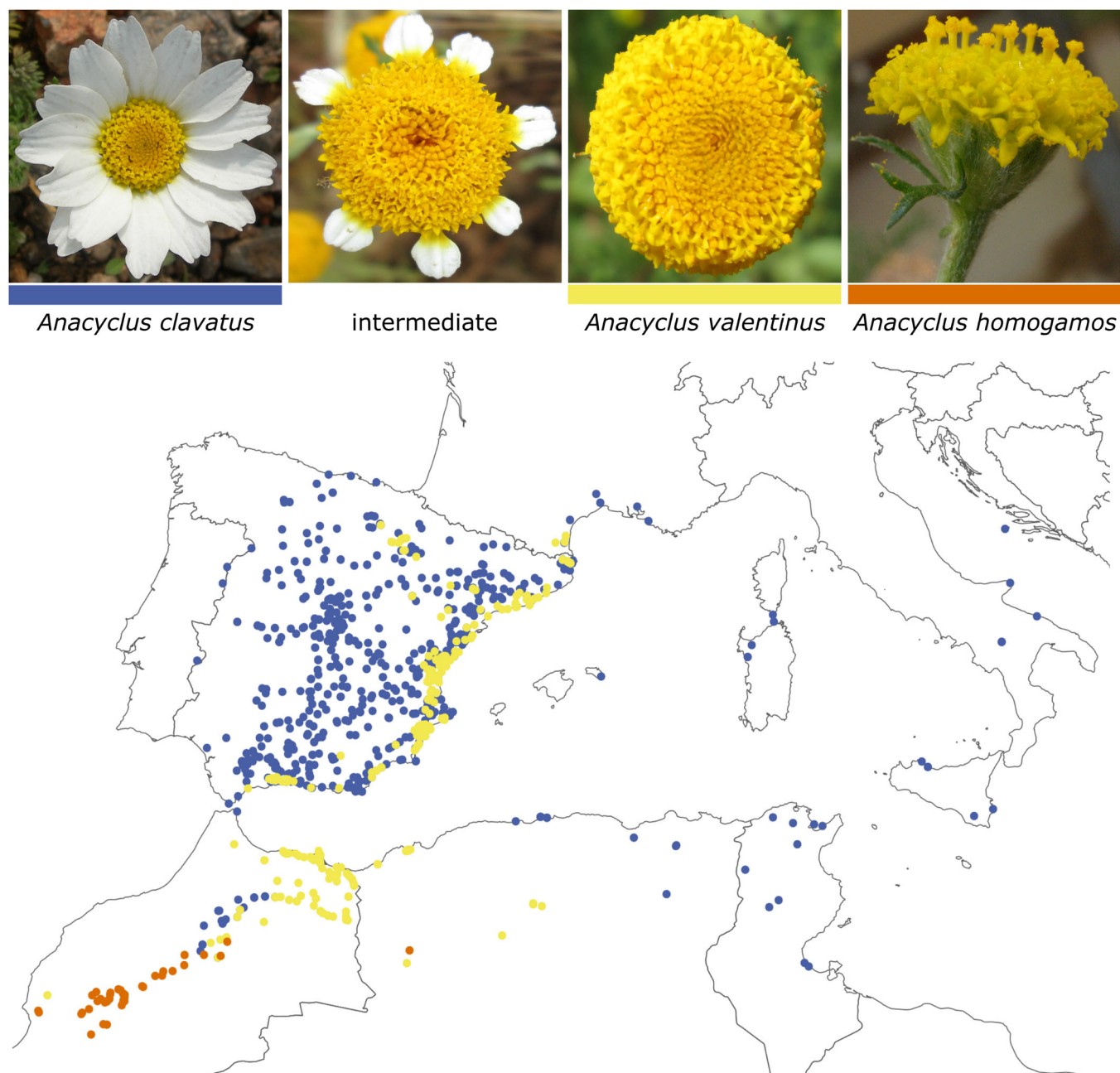


FIGURE 1 Distribution map of the studied species *Anacyclus clavatus* (blue), *A. valentinus* (yellow), and *A. homogamos* (red) based on herbarium specimens revised and on our field collections. Inflorescences (capitula) of floral phenotypes for each species are shown. “Intermediate” refers to an intermediate phenotype between *A. clavatus* and *A. valentinus*.

Study area and sampling

After reviewing 1562 *Anacyclus* specimens from B, BC, BCN, G, LISE, LISI, LISU, MA, MGC, SEV, and VAL herbaria (according to Thiers, 2016) and several field campaigns between 2010 and 2013, we eventually used 964 specimens to represent the geographic distribution of the *Anacyclus* species of study across western Mediterranean Basin (Figure 1) and to model their potential distributions (see below). To study the among- and within-population genetic structure and the extent of gene

flow, we used 12–36 individuals from each of 31 populations representing the study species, totaling 585 individuals (Appendix S1). Ten populations were of *A. clavatus*, nine of *A. valentinus*, seven of *A. homogamos*, and five represented mixed populations. In three of these five mixed populations, individual phenotypes were categorized as *clavatus*-like, *valentinus*-like, or as intermediate (Appendix S1). For the two other mixed populations, we were not able to categorize each individual, but we considered the ratios of different phenotypes observed in the field. Individuals from each population were haphazardly collected and separated by at

least 10 m from each other. Leaves from each individual were dried and stored in silica gel.

To cover the distribution range more densely and to increase ENM performance of the resulting genetic clusters (see below), we estimated genetic structure of a larger data set that included all individuals of the 31 populations used for the intrapopulation genetic analysis plus 1–3 individuals from other 98 populations (i.e., 69 populations of *A. clavatus*, 22 of *A. valentinus*, three of *A. homogamos*, and four of mixed populations), totaling 721 individuals (Appendix S2).

Microsatellite genotyping

We extracted total genomic DNA from silica-dried leaves using the DNeasy Plant Minikit (QIAGEN, Hilden, Germany). We used eight microsatellite markers (i.e., locus 9, 15, 17, 19, 20, 21, 24, and D3) previously developed for *A. clavatus* (Agudo et al., 2013). AllGenetics & Biology SL (A Coruña, Spain) amplified fragments using the protocols described by Agudo et al. (2013). We scored and manually checked the fragments with the software Geneious v.7.1.2 (Kearse et al., 2012). We tested for homozygote excess with MICRO-CHECKER v.2.2.3 (Van Oosterhout et al., 2004) to evaluate the presence of null alleles per locus at the species and population levels. The observed frequency of homozygote classes was analyzed using Monte Carlo simulations ($\times 1000$) and a Bonferroni (Dunn–Šidák correction)-adjusted 95% confidence interval. To avoid bias, we discarded individuals with amplification success below 50% across loci, ending up with a data set of 585 individuals from 31 populations (Appendix S1). The same criterion was applied to the extended data set (Appendix S2).

Effect of homoplasy and suitability of microsatellites

Scoring errors produced by the presence of insertions and deletions (indels) in microsatellite flanking regions may produce homoplasy, which tends to increase with phylogenetic distance between species (Germain-Aubrey et al., 2016). For determining the presence of such indels, all loci in 43 individuals (25 of *A. clavatus*, six of *A. homogamos*, seven of *A. valentinus* and five of phenotypically mixed individuals) were sequenced by AllGenetics & Biology SL (A Coruña, Spain) using a 3730XL DNA Analyzer (Applied Biosystems, Foster City, California, USA). Each reaction contained 12.5 μ L of Supreme NZYTaQ Green PCR Master Mix (NZYTech, Lisbon, Portugal), 0.5 μ M of each primer, 25 ng of template DNA, and PCR-grade water up to 25 μ L. The thermal cycling conditions were 95°C for 5 min; 35 cycles at 95°C for 1 min, 56°C for 1 min, and 72°C for 45 s; and a final extension step at 72°C for 5 min. PCR products were verified on 1% agarose gels stained with GreenSafe Premium (NZYTech, Lisbon, Portugal). We

aligned the sequences of the flanking regions and repeat motifs with Geneious using the known sequences of each locus as a reference (Agudo et al., 2013). Overall, we considered loci with a poor performance when they had low amplification rates and/or high frequencies of null alleles. To assess the effect of homoplasy and the presence of unsuitable loci, we compared the results of genetic structure analyses with and without loci with poor performance.

Genetic diversity and structure

We used the software GenAlex v.6.501 (Peakall and Smouse, 2012) to estimate percentage of polymorphic loci, number of alleles, observed (H_o) and expected heterozygosity (H_e), inbreeding coefficient (F_{IS}), and pairwise fixation index (F_{ST}) for each species and population. We conducted analyses of molecular variance (AMOVA) to decompose the genetic variance among species, among populations within species and among individuals within populations. We excluded the within-individual level of variation, interpolated missing data, and used 999 permutations to test the significance of F_{ST} statistics. We also conducted Mantel tests between pairwise Euclidian geographic and genetic distances (i.e., the proportion of pairwise allelic differences among individuals) with 999 permutations to test for significance of isolation-by-distance.

We estimated genetic structure among all *Anacyclus* individuals with two complementary methods. First, we performed a Bayesian model-based analysis to estimate the number of genetic clusters using STRUCTURE v.2.3.4 (Pritchard et al., 2000; Falush et al., 2003) under an admixture ancestry model with the default parameters. Burn-in period was set to 50,000 runs and MCMC repetitions after burn-in to 100,000. For each simulation, we ran 2–10 genetic populations (K) and 20 iterations per K . Analysis and visualization of the population structure was performed with R package POPHELPER v.1.1.9 (Francis, 2017). All R packages used here were run with R v4.1.1 (R Core Team 2021). We inferred the best K using the ΔK method (Earl and vonHoldt, 2012). The average matrix of multiple runs per K that was used for subsequent calculations, as well as the highest value of the symmetric similarity coefficient (H') across runs per K were obtained with CLUMPAK (Kopelman et al., 2015). A posterior probability of $Q \geq 0.90$ was used as a threshold to assign full ancestry to a single genetic cluster, and samples below this threshold were considered as individuals with admixed ancestry. Second, we used R package adegenet v.2.1.7 (Jombart, 2008) to perform a principal coordinate analysis (PCoA) based on synthetic variables summarizing the genetic distance matrix among individuals.

Ecological niche and spatial metrics

We estimated niche metrics with a recently described hypervolume-based approach and spatial suitability metrics

with ecological niche model (ENMs) projections for current timeframe (Hespanhol et al., 2022). We only used macroclimatic variables to estimate niche and spatial suitability metrics because all *Anacyclus* species studied grow in similar open habitats under various human disturbances (e.g., mostly road margins). We used the bioclimatic variables available in WorldClim v.2.1 at a 1-km resolution (Fick and Hijmans, 2017) as predictors in the ENMs. As background, we randomly selected 10,000 points over the entire study area. To avoid multicollinearity, we ran a correlation analysis on the background points and eliminated one of the variables in each pair with a Pearson correlation value >0.8 . The variables ultimately included in the models were annual mean temperature (BIO1), isothermality (BIO3), mean temperature of the wettest quarter (BIO8), mean temperature of the driest quarter (BIO9), precipitation seasonality (BIO15), and precipitation of the coldest quarter (BIO19).

We predicted the climatic suitability for *A. clavatus*, *A. valentinus*, and *A. homogamos* using presence data confirmed by our own observations during field campaigns and herbarium revisions. To avoid sampling bias (Syfert et al., 2013), we only retained points for each species that were separated by at least 1 km from each other and thus matching the resolution of the climatic data, resulting in 964 occurrences. For phenotypically mixed populations with individuals showing *clavatus*-like and *valentinus*-like phenotypes, we considered occurrences for both species.

The ENMs were based on an ensemble approach (Araújo and New, 2007) using three modeling techniques: generalized linear models (GLM; McCullagh and Nelder, 1989), gradient boosting machine (GBM; Friedman, 2001) and random forest (RF; Breiman, 2001). We performed models with R package BIOMOD v.2.0 (Thuiller et al., 2009) using the default parameters. We calibrated and evaluated all models with 70% and 30% of the data using the area under the ROC curve (AUC) and true skill statistics (TSS). For each modeling technique, we replicated the procedure 10 times with random training and evaluation data sets. To remove spurious models, we generated the ensembles using models with $AUC > 0.8$ and $TSS > 0.7$. The contribution of each model to the final ensemble model was proportional to their goodness-of-fit statistics. Climatic suitability was considered as a consensus across statistical techniques and their contribution to the ensemble was proportional to their AUC values. We converted the consensus model into a binary model (presence/absence) applying a threshold that allows a maximum 5% omission error (i.e., the percentage of the real presence predicted as absences in the model; false negative; Fielding and Bell, 1997).

We estimated the spatial overlap between species, between species and genetic clusters, and between genetic clusters for the selected final ensemble models with the Schoener's *D* (Schoener, 1970; Warren et al., 2008). We plotted and transformed observations of each species and genetic cluster into densities using a gridded environmental

space of 100×100 cells. We also plotted and calculated overlaps between species pairs and between genetic cluster pairs by using binary models.

RESULTS

Homoplasmy and marker performance

There were no indels in flanking regions of 51 sequences from loci D3, 17, 19, and 20. In contrast, loci 21 and 24 presented indels of 14 or 15 nucleotides in 18 and 21 sequences. We could not assess homoplasmy in six sequences from locus 9 and four from locus 15 because of low sequence quality. In addition, locus 15 was not amplified in any of the *A. homogamos* populations and in most of *A. valentinus*. To assess the effect of low performance of locus 15 and indels in loci 21 and 24, we ran two analyses of genetic structure, one including all loci and the other excluding loci 15, 21, and 24.

Our results showed that the most probable number of genetic clusters was four ($K = 4$) when considering all loci (see below), and five ($K = 5$) when excluding loci 15, 21, and 24. However, just a few samples (5.5%) scattered across different *A. clavatus* and *A. valentinus* populations represented the fifth genetic cluster, which did not provide useful information to interpret our results. Furthermore, in the $K = 5$ scenario, half of the *A. valentinus* samples showed *Q* values < 0.90 . In contrast, when all loci were included ($K = 4$), most of these *A. valentinus* samples were assigned to one of the four genetic clusters. Finally, the assignments to genetic clusters for all samples with $Q \geq 0.90$ were identical when including or excluding loci 15, 21, and 24. Therefore, we used all loci in all analyses to enhance the resolution of the genetic structure.

Genetic diversity

We identified 521 different genotypes across all 585 *Anacyclus* individuals sampled (89.1% of different genotypes), ranging from a low of 71.7% of different genotypes in *A. valentinus* populations to a high of 100% in *A. clavatus*. No genotype sharing was observed between different species, except in mixed populations in which 1–9 genotypes were shared among several other populations of *A. valentinus* and *A. clavatus*. Within species, the pattern of genotypes shared among populations was variable. For example, 1–3 *A. homogamos* genotypes were shared by 2–4 populations, 1–10 *A. valentinus* genotypes were shared by 2–9 populations, and no genotype sharing was observed among *A. clavatus* populations.

Anacyclus clavatus, *A. homogamos*, and mixed populations exhibited high genetic diversity parameters, whereas *A. valentinus* populations showed the lowest values (Table 1). In short, genetic diversity was higher in mixed populations and in *A. clavatus* ($H_e = 0.38$ – 0.40),

TABLE 1 Genetic diversity parameters for the *Anacyclus* species and for the phenotypically mixed populations studied. Parameters include number of populations (N_{POP}), number of individuals (N_{IND}), number of alleles (N_a), observed heterozygosity (H_o), expected heterozygosity (H_e), inbreeding coefficient (F_{IS}), pairwise fixation index among population pairs within each species (F_{ST}) and percentage of polymorphic loci (PL). For each species, means (\pm SD) among populations are given.

Species/group	N_{POP}	N_{IND}	N_a	H_o	H_e	F_{IS}	F_{ST}	PL
<i>A. clavatus</i>	10	225						
Mean		18.46	3.113	0.297	0.380	0.229	0.352	78.75
SD		0.85	0.192	0.032	0.029	0.147	0.074	5.61
<i>A. homogamos</i>	7	106						
Mean		9.79	2.625	0.212	0.309	0.273	0.378	67.86
SD		0.91	0.281	0.029	0.037	0.108	0.147	3.72
<i>A. valentinus</i>	9	159						
Mean		12.81	1.819	0.147	0.180	0.261	0.342	41.67
SD		0.89	0.192	0.026	0.030	0.139	0.150	5.10
Phenotypically mixed	5	95						
Mean		15.35	3.100	0.286	0.399	0.328	0.176	87.50
SD		0.84	0.270	0.042	0.041	0.152	0.036	6.85

intermediate in *A. homogamos* ($H_e = 0.31$) and lower in *A. valentinus* ($H_e = 0.18$). On average, the inbreeding coefficient (F_{IS}) varied from a low of 0.229 in *A. clavatus* to a high of 0.328 in mixed populations. The genetic differentiation among populations was similar in *A. homogamos* ($F_{ST} = 0.378$), *A. clavatus* ($F_{ST} = 0.352$) and *A. valentinus* ($F_{ST} = 0.342$), whereas mixed populations exhibited the lowest estimate ($F_{ST} = 0.176$).

The analysis of molecular variance (AMOVA) indicated that 79% of the genetic variance was found within populations, whereas the remaining 14% and 7% of the genetic variance was found among species and among populations within species, respectively. For each species separately, the highest genetic variance was also found within populations: 71% for *A. clavatus*, 89% for *A. homogamos*, 90% for *A. valentinus*, and 85% for mixed populations.

Mantel tests indicated a significant correlation between geographic and genetic distances when all samples were analyzed together ($r = 0.300$, $P = 0.001$). In contrast, isolation by distance slightly fluctuated around significance ($P = 0.05$) in *A. homogamos* populations ($r = 0.693$, $P = 0.040$) and *A. clavatus* populations ($r = 0.200$, $P = 0.068$), and it was not significant in *A. valentinus* populations ($r = 0.326$, $P = 0.129$) and mixed populations ($r = -0.352$, $P = 0.175$).

Genetic structure

The Bayesian analysis of genetic structure estimated four ($K = 4$) as the most probable number of genetic clusters (hereafter named GC1, 2, 3, and 4) for this data set (Figure 2A), with an average symmetric similarity

coefficient $H' = 0.99$. Principal coordinate analysis (PCoA) supported quite well this structure made of four genetic clusters (Figure 2B), including the assignment of populations to specific clusters of each species, and depicting the genetic complexity of mixed populations (see below).

Considering the maximum mean proportions of genetic cluster membership among individuals within populations (Figure 2A; Appendix S3), we found that populations of *A. valentinus* and *A. homogamos* were very homogenous falling within their respective genetic clusters (GC3 and GC4, respectively). *Anacyclus valentinus* and *A. homogamos* populations exhibited mean membership proportions of 0.95 (9 populations; range = 0.94–0.96) and 0.90 (7 populations; range = 0.70–0.90), respectively. In contrast, populations of *A. clavatus* fell within up to three different genetic clusters. Two clusters specific to *A. clavatus* (GC1 and GC2) with mean membership proportions of 0.87 (5 populations; range = 0.74–0.96) and 0.91 (3 populations; range = 0.85–0.95), respectively, and the GC3—the one specific to *A. valentinus*—with a mean membership proportion of 0.63 (2 populations; range = 0.60–0.66). Finally, the mixed populations were more erratic: one population clustered in GC2—one of the specific *A. clavatus* group—with a mean membership proportion of 0.60, three populations fell within the other *A. clavatus* cluster (GC1) with a mean membership proportion of 0.81 (range = 0.73–0.95), and another population clustered with the *A. valentinus* group (GC3) with a mean membership proportion of 0.63.

Individuals with admixed ancestry (i.e., with a low cluster assignment; $Q < 0.90$) were mostly present in *A. clavatus* (10 populations; mean proportion of individuals with $Q < 0.90$

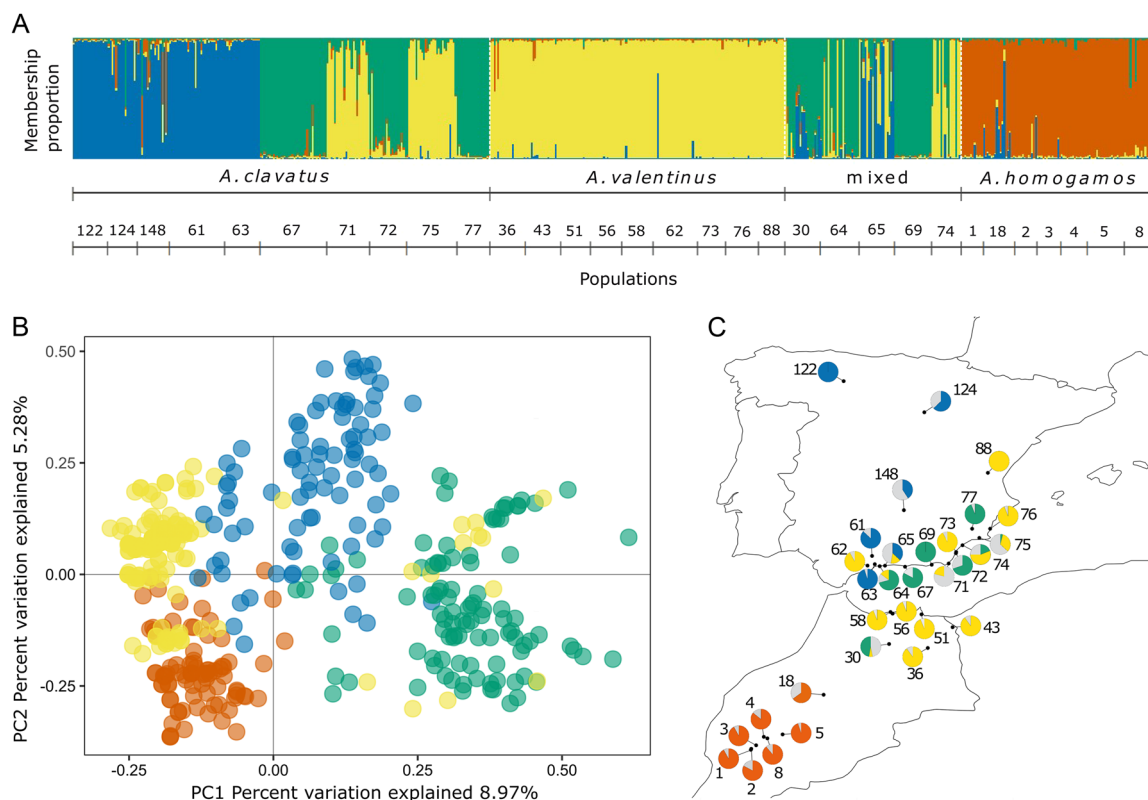


FIGURE 2 Analysis of the genetic structure of the 31 populations representing *Anacyclus clavatus*, *A. homogamos*, *A. valentinus*, and the phenotypically mixed populations. Colors represent the genetic clusters: GC1 (blue), GC2 (green), GC3 (yellow), and GC4 (red). Numbers indicate the population ID. (A) Bar plot of estimated membership proportions to the most probable number of genetic clusters ($K = 4$) of each individual inferred with *STRUCTURE* v.2.3.4. (B) Principal coordinate analysis (PCoA) of all individuals with a posterior probability ≥ 0.90 of genetic cluster membership. (C) Geographic distribution of the populations (black dots). Pie charts represent frequency of genetic cluster membership by population. Frequency of individuals with a posterior probability of genetic cluster membership < 0.90 are represented in grey.

per population = 0.35; range = 0.00–0.83) and in mixed populations (5 populations; mean = 0.27; range = 0.00–0.47). In contrast, few individuals with $Q < 0.90$ were present in *A. homogamos* (7 populations; mean = 0.18; 0.05–0.65) and *A. valentinus* (9 populations; mean = 0.07; 0.00–0.11). Interestingly, four populations of *A. clavatus* (populations 71, 72, 75, and 148) and one of *A. homogamos* (population 18) had high admixed ancestry (range $Q < 0.90 = 0.59$ –0.83). These populations were mostly located in contact zones between *A. clavatus* and *A. valentinus* (Figure 2C).

Ecological niche and spatial metrics

The climatic suitable distribution of species predicted by the ENMs mostly included all presence data and extended to larger areas where no presence was recorded (Figure 3A–C). The genetic cluster assignments of the samples used for the intrapopulation genetic analysis and the ENMs were mostly consistent (Appendices S3 and S4), indicating that the larger data set with populations represented by a lower number of individuals was not biasing the general picture of genetic structure in this system. The climatic suitable distribution of the genetic clusters mostly coincided with the suitable

distribution of the species, showing even a better fit than those of the species that they represent (Figure 4A–D). The final ensemble models across all species and genetic clusters were statistically reliable, showing AUC values > 0.9 .

The most important climatic variables accounting for the distribution of each species were precipitation of the coldest quarter (BIO19) and mean temperature of the driest quarter (BIO9) for *A. clavatus*, precipitation seasonality (BIO15) and annual mean temperature (BIO1) for *A. homogamos*, and precipitation of the coldest quarter (BIO19) and annual mean temperature (BIO1) for *A. valentinus* (Table 2). The suitable climatic distributions predicted by the ENMs for each of the species partially overlapped (Figure 3D–F). However, overlap was higher between *A. clavatus* and *A. valentinus* (Figure 3E; Schoener's $D = 0.73$, overlap with binary models = 0.42) than between *A. clavatus* and *A. homogamos* (Figure 3D; $D = 0.60$, binary overlap = 0.05) and between *A. valentinus* and *A. homogamos* (Figure 3F; $D = 0.64$, binary overlap = 0.06).

For genetic clusters, annual mean temperature (BIO1) and isothermality (BIO3) emerged as the most important climatic variables accounting for their suitable distribution (Table 2), except for GC4, in which precipitation seasonality (BIO15) was the most important predictor (Table 2). The highest spatial overlap in suitable distribution was observed

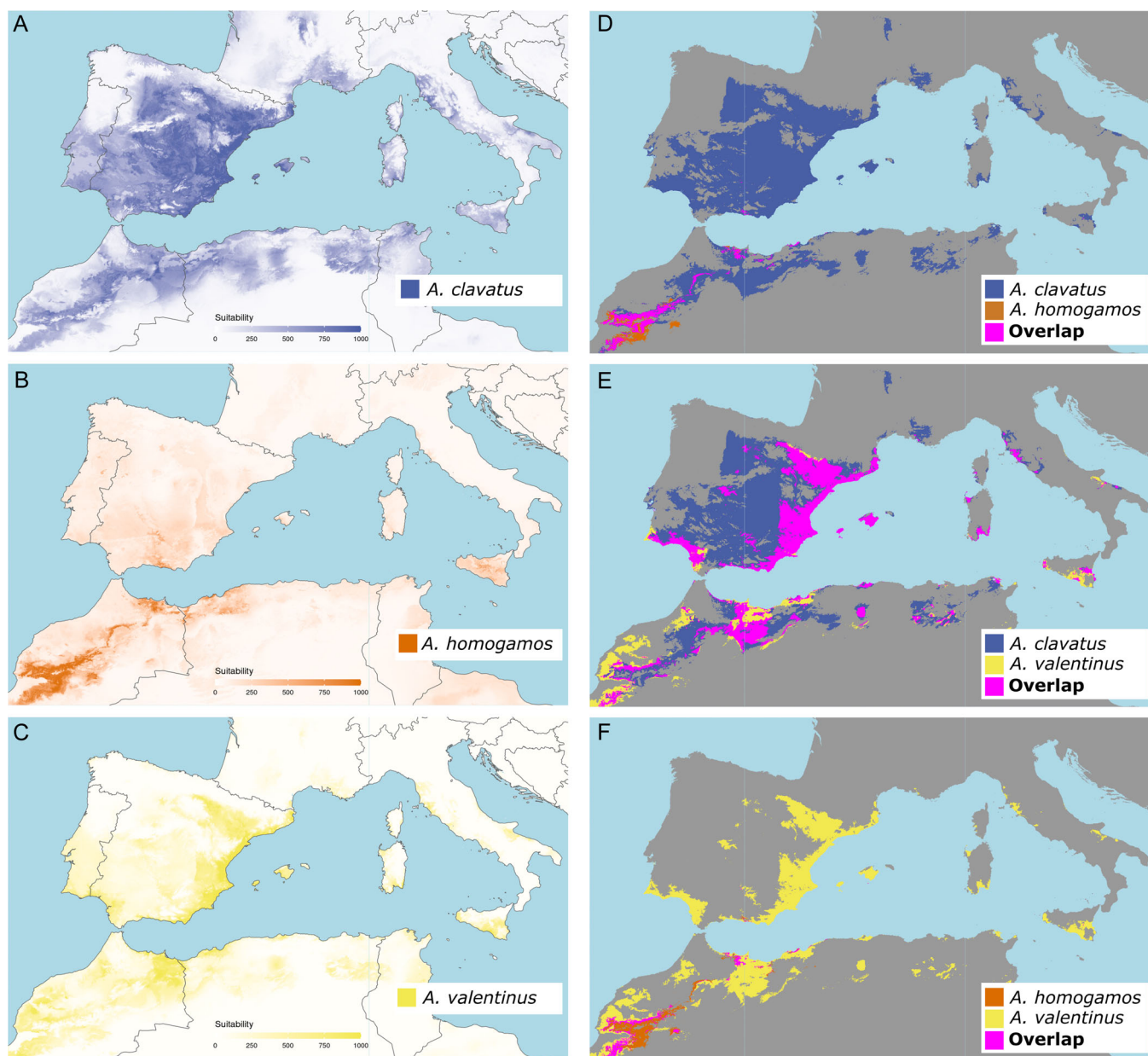


FIGURE 3 Ecological niche models for the studied species and the overlap of the distributions predicted between species pairs. Overlaps were plotted using binary models. (A) *Anacyclus clavatus*. (B) *A. homogamos*. (C) *A. valentinus*. (D) Overlap between *A. clavatus* and *A. homogamos*. (E) Overlap between *A. clavatus* and *A. valentinus*. (F) Overlap between *A. homogamos* and *A. valentinus*.

between GC2 and GC3 (Figure 4E; $D = 0.79$, binary overlap = 0.14), and the lowest between GC1 and GC2—the two specific clusters of *A. clavatus* (Figure 4E; $D = 0.60$, binary overlap = 0.006). The rest of the genetic cluster pairs exhibited intermediate overlaps in their suitable distributions (Figure 4E; range $D = 0.70$ – 0.76 , binary overlap = 0.002–0.09).

Finally, we also explored the overlaps between species and genetic clusters. The suitable area of GC1 had a high overlap with that of *A. clavatus* ($D = 0.77$) and to a lesser extent with that of *A. valentinus* and *A. homogamos* ($D = 0.68$ in both cases). The suitable distribution for GC2 had the highest overlap with that of *A. homogamos* ($D = 0.72$) and the lowest

with *A. clavatus* ($D = 0.55$). The GC3 had similar overlap with that of *A. valentinus* ($D = 0.73$) and with that of *A. homogamos* ($D = 0.75$). Finally, GC4 had the highest overlap with that of *A. homogamos* ($D = 0.85$) and much lower with those of the other two species ($D = 0.61$ for *A. clavatus* and 0.62 for *A. valentinus*).

DISCUSSION

Our study on the multispecies genetic structure for *Anacyclus* in the western Mediterranean Basin provided important insights into species identity and the likely role of

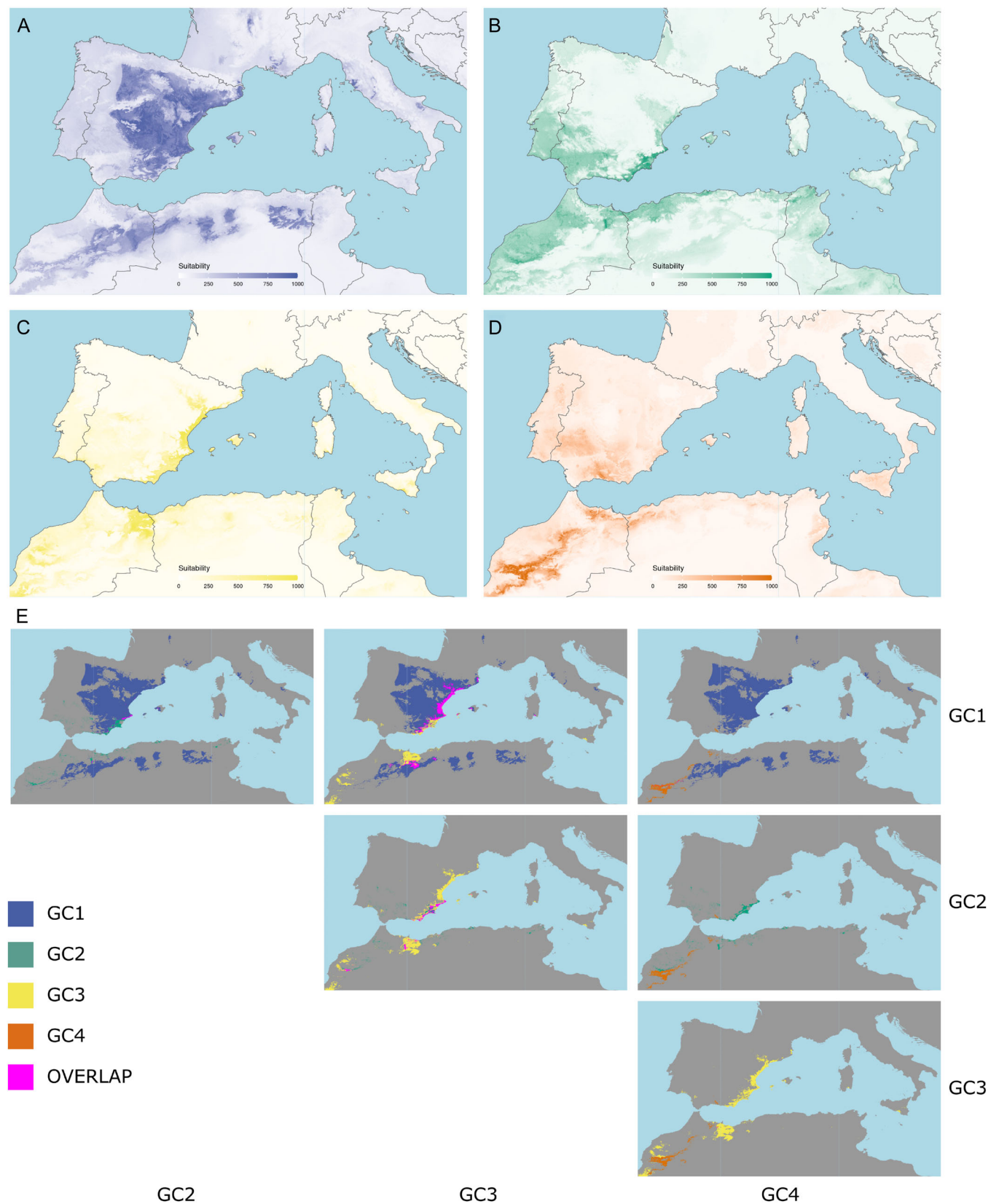


FIGURE 4 Ecological niche models for the genetic clusters estimated and the overlap of the distributions predicted between genetic clusters pairs. Overlaps were plotted using binary models. (A) Genetic cluster 1. (B) Genetic cluster 2. (C) Genetic cluster 3. (D) Genetic cluster 4. (E) Overlap between genetic clusters pairs.

TABLE 2 Percentage contribution of climatic variables to the fit of the models for each species and genetic cluster (GC). Climatic variables: BIO1, annual mean temperature; BIO3, isothermality; BIO8, mean temperature of the wettest quarter; BIO9, mean temperature of the driest quarter; BIO15, precipitation seasonality; BIO19, precipitation of the coldest quarter. The two largest contributions for each species and genetic cluster are given in boldface.

Variable	<i>A. clavatus</i>	<i>A. homogamos</i>	<i>A. valentinus</i>	GC 1	GC 2	GC 3	GC 4
BIO1	0.12 ± 0.02	0.58 ± 0.21	0.35 ± 0.05	0.38 ± 0.19	0.55 ± 0.21	0.53 ± 0.17	0.64 ± 0.20
BIO3	0.27 ± 0.03	0.29 ± 0.15	0.23 ± 0.02	0.34 ± 0.10	0.55 ± 0.20	0.48 ± 0.20	0.36 ± 0.17
BIO8	0.10 ± 0.01	0.22 ± 0.10	0.24 ± 0.03	0.13 ± 0.04	0.24 ± 0.06	0.22 ± 0.05	0.26 ± 0.13
BIO9	0.28 ± 0.03	0.37 ± 0.12	0.47 ± 0.05	0.34 ± 0.19	0.33 ± 0.05	0.42 ± 0.17	0.43 ± 0.16
BIO15	0.18 ± 0.01	0.78 ± 0.06	0.12 ± 0.03	0.25 ± 0.12	0.27 ± 0.10	0.17 ± 0.02	0.80 ± 0.14
BIO19	0.30 ± 0.03	0.39 ± 0.07	0.24 ± 0.04	0.26 ± 0.07	0.31 ± 0.15	0.29 ± 0.14	0.31 ± 0.07

gene flow and climatic variation for the genetic diversification in this species complex.

Species identity

In previous phylogenetic analyses of the genus *Anacyclus* based on nrDNA and cpDNA sequences (Oberprieler, 2004; Vitales et al., 2018), none of the three study *Anacyclus* species were sisters, although in both cases, *A. clavatus* and *A. valentinus* were more closely related to each other than to *A. homogamos*. Recent evidence also supported the divergence between these three species, such as the significant differences in their genome sizes (Agudo et al., 2019) and the partial post-zygotic reproductive barriers between them (Álvarez et al., 2020). This study provided additional support to the recognition of these three *Anacyclus* species, but also brought to light the complexity posed by gene flow among them and the necessity of a thorough review of their diagnostic characters. In addition, the climatic variables accounting for their suitability distributions differed among species (Table 2), suggesting an ecological differentiation.

However, there was not a correspondence between genetic clusters and phenotypes in contact areas between species, suggesting the existence of gene flow between them. Gene flow is supported by the geographic distribution of *A. clavatus* samples assigned to GC2 and GC3, which only occur in areas where *A. valentinus* is present (i.e., mainly in Mediterranean coastal areas; Figure 2C). We believe that samples with a *clavatus*-like phenotype assigned to GC3 could represent hybrids between *A. clavatus* and *A. valentinus*. In addition, samples assigned to GC2—one of the specific genetic clusters of *A. clavatus*—form mixed populations, showed *clavatus*-like, *valentinus*-like, and intermediate phenotypes, suggesting extensive gene flow in these populations. On the contrary, in areas where *A. valentinus* was absent, all *A. clavatus* populations fell within GC1.

Although there is no data on the geographic-morphologic relationship in *A. clavatus*, there exists a geographic pattern of genome size variation (Agudo et al., 2019) matching the genetic structure of *A. clavatus*

found in this study. In particular, *A. clavatus* populations from areas where *A. valentinus* is present had significantly smaller genomes compared with those in *A. clavatus* populations from other areas. The presence of two clearly differentiated genetic clusters in *A. clavatus* with a different genome size and occupying different geographic areas, suggests the existence of two distinct lineages within *A. clavatus*. The presence of clearly divergent intraspecific genetic clusters with a marked geographic pattern is a common result in regional-scale population genetic studies (Brennan et al., 2014; Mandák et al., 2016; Castilla et al., 2020), illustrating the evolutionary and demographic imprint on plants' genetic structure across their distributions.

In contrast, *A. valentinus* and *A. homogamos* had no intraspecific genetic structure, and each species had a clearer correspondence with its own genetic cluster. Although it is difficult to determine the causes of such a relationship among species and genetic clusters in the *Anacyclus* complex, the association among distribution, genetic cluster membership and genetic diversity are more complicated than previously anticipated. For example, *A. valentinus* had the lowest genetic diversity among all studied species, despite occurring across a large area along the Iberian and North African Mediterranean coast. In *A. valentinus*, inbreeding and isolation seem unlikely because it is self-incompatible (Álvarez et al., 2020) and mostly occurs along roadsides. Moreover, the inbreeding coefficient (F_{IS}) for *A. valentinus* did not indicate an excess of homozygosity, and the AMOVA showed a structure congruent with connected populations. In contrast, the genetic diversity of *A. homogamos* was similar to that of *A. clavatus* and to the group of phenotypically mixed populations, despite its more restricted distribution.

Gene flow, introgression, and phenotype variation

Our study provided the first genetic evidence of gene flow among *Anacyclus* species (Humphries, 1979, 1981; Rosato et al., 2017; Agudo et al., 2019; Álvarez, 2019; Vitales et al., 2020). As in other systems (e.g., Emanuelli et al., 2013;

Ortego et al., 2017; Castilla et al., 2020), the genetic admixture within populations, including individuals with varying admixed genetic compositions, indicates the existence of current gene flow among different genetic clusters and species. Genetic admixture was very clear in the group of phenotypically mixed populations, where hybrid individuals with admixed ancestry were frequent. These hybrids showed intermediate phenotypes, but they also presented *clavatus*-like phenotypes and, to a lesser extent, *valentinus*-like phenotypes, which might be the result of backcrossing or introgression between *A. clavatus* and *A. valentinus*.

Gene flow also seems to occur in populations from areas off contact zones. This was the case of population 148 of *A. clavatus* (Figure 2B, C), in which some individuals showed admixed genetic compositions. This population is located outside of what it can be considered as contact areas between species (Figures 1, 2C). Dispersal from those areas seems unlikely because seed dispersal in all annual *Anacyclus* species mostly occurs by means of rain (i.e., ombrohydrochory; Bastida et al., 2010; Torices et al., 2013). However, since *Anacyclus* species grow mainly as roadside weeds (Humphries, 1979; Álvarez, 2019), we cannot rule out occasional dispersal events facilitated by human activity. Alternatively, we might have underestimated the size of the contact areas between species. In any case, these results stress the need for further studies using phylogenomic approaches and more intense samplings, particularly within populations, for the delimitation of contact areas between species.

Species and genetic cluster distribution

Niche modeling for within-species genetic clusters is increasingly frequent in the literature (Piñeiro et al., 2007; Gotelli and Stanton-Geddes, 2015; Marcer et al., 2016; Milanesi et al., 2018; Martínez-Minaya et al., 2019). Here, ENMs largely predicted similar distributions for species and genetic clusters, although the climatic factors driving such distributions were different for species and their genetic clusters (Table 2). On top of methodological limitations to compare comprehensively distributions between species and genetic clusters (e.g., sample sizes for species and genetic clusters were very different), other factors may account for some discrepancies between observed and predicted distributions. For example, in the wide area across southwestern Iberia predicted for *A. clavatus*, where we know that the species is absent, the competition with other dominant congeneric species, such as *A. radiatus* (Álvarez, 2020), might explain such inconsistency.

As expected, *A. clavatus* had the largest geographic area predicted by the environmental niche models. The fact that the predicted areas for the two genetic clusters represented by this species (GC1 and GC2) were mutually exclusive (Figure 4A, B) partly explains the wide climatic

optimum for *A. clavatus*. The lack of geographic barriers between the predicted areas for the two *A. clavatus* genetic clusters suggests that climate might be a selective pressure explaining their exclusionary distributions. The analysis of niche overlap in the environmental space also supports this hypothesis, as the overlap between GC1 and GC2, was the lowest of all comparisons between genetic clusters. In contrast, the optimal area predicted for GC2 almost completely coincided with part of that of GC3—the genetic cluster corresponding to *A. valentinus*. Finally, the genetic cluster represented mainly by *A. homogamos* (GC4) showed the most exclusive predicted area (western Morocco and southwestern Iberia) whose climatic optimum exhibited low overlap with that of the rest of genetic clusters.

Recent studies on the genetic structure and genomic differentiation of the annual mustard *Arabidopsis thaliana* across the same region (Iberia and Morocco) showed some similarities with the results presented here for the *Anacyclus* complex that are worth mentioning. For example, *A. thaliana* in Morocco is made of a specific genetic cluster with a strong presence in the Atlas Mountains (Brennan et al., 2014; Durvasula et al., 2017). In the case of *Anacyclus*, this specific cluster corresponds to *A. homogamos*. Furthermore, there is a strong differentiation between southern and northern *A. thaliana* populations in Morocco (Brennan et al., 2014). In the case of *Anacyclus*, such a differentiation corresponds to the predominant occurrence of *A. homogamos* and *A. valentinus* in southern and northern Morocco, respectively. Finally, Iberian *A. thaliana* is made of four genetic clusters, one of them closely related to the Moroccan cluster (Brennan et al., 2014; Durvasula et al., 2017). In the case of *Anacyclus*, the predominance of *A. clavatus* in Iberia, followed by the presence of *A. valentinus* and traces of *A. homogamos*, also indicates a shared history between Moroccan and Iberian *Anacyclus*. Overall, we hypothesize that the shared geologic (e.g., the isolation of the two continents after the Messinian Crisis about 5.5 Ma) and climatic history of the region (e.g., repeated pluvial periods shifting from arid to moist conditions) might have shaped the intra- and interspecific genetic structure of multiple organisms in a similar manner.

We conclude by stressing the value of the multispecies approach and the simultaneous modeling of species and genetic clusters to provide a snapshot on the evolutionary dynamics of complex plant groups under a diversification process, such as the *Anacyclus* complex in western Mediterranean Basin. Although the origin of species and genetic clusters is beyond the scope of this study, our results suggest that introgression in contact zones might have played an important role in shaping genetic structure and possibly species delimitation. Further research should focus on the origin of these intraspecific genetic clusters with the incorporation of whole-genome sequence data for the genomic characterization of admixed lineages. In addition, we need to increase sampling efforts for each *Anacyclus* species of the complex with a particular emphasis on the

within-population level in contact zones harboring phenotypically mixed individuals.

AUTHOR CONTRIBUTIONS

Conceptualization: I.A. and R.T.; data curation: A.B.A., A.M., I.A., R.G.M., and F.X.P.; formal analysis: A.B.A., A.M., I.A., R.G.M., and F.X.P.; funding acquisition: A.B.A., A.M., I.A., R.G.M., R.T., and F.X.P.; investigation: A.B.A., I.A., R.T., and F.X.P.; methodology: A.M., I.A., R.G.M., R.T., and F.X.P.; project administration: I.A. and R.T.; resources: A.M., I.A., R.G.M., R.T., and F.X.P.; supervision: I.A., R.T., and F.X.P.; visualization: A.M., I.A., and R.G.M.; writing original draft: A.B.A. and I.A.; writing review and editing: A.M., I.A., R.G.M., R.T., and F.X.P. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We are grateful to Joel Calvo, Raúl Gonzalo, Leopoldo Medina, Alejandro Quintanar, María Santos, and Salvatore Tomasello for collection of some specimens. We thank Alberto Herrero for technical assistance. Thanks also to the two anonymous reviewers and to the Associate Editor for their suggestions and thorough work on improving this article. The European Regional Development Fund and the Spanish Ministry of Science and Innovation (CGL2010-18039) funded this study to I. Álvarez. A.B.A. and R.T. received support from grants of the Spanish Ministry of Economy and Competitiveness (BES-2011-048197 and BVA 2010-0375). A.M. and F.X.P. were supported by PID2019-104135GB-I00 from the Agencia Estatal de Investigación (AEI) of Spain and the European Regional Development Fund (FEDER, UE). The European Regional Development Fund and the Spanish Ministry of Science and Innovation supported R.G.M. through the NextDive project (PID2021-124187NB-I00).

DATA AVAILABILITY STATEMENT

Data sets for this study can be found in the Supporting Information section.

ORCID

F. Xavier Picó  <http://orcid.org/0000-0003-2849-4922>
 Rubén G. Mateo  <http://orcid.org/0000-0001-8577-001X>
 Arnald Marcer  <http://orcid.org/0000-0002-6532-7712>
 Rubén Torices  <http://orcid.org/0000-0001-9604-1111>
 Inés Álvarez  <http://orcid.org/0000-0002-5982-5424>

REFERENCES

- Abbott, R. J. 2017. Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. *Journal of Systematics and Evolution* 55: 238–258.
- Agudo, A., F. X., Picó, and I. Álvarez. 2013. Microsatellite primers in the weedy annual herb *Anacyclus clavatus* (Asteraceae) and four closely related species. *Applications in Plant Sciences* 1: 1300043.
- Agudo, A. B., R. Torices, J. Loureiro, S. Castro, M. Castro, and Álvarez, I. 2019. Genome size variation in a hybridizing diploid species complex in *Anacyclus* (Asteraceae: Anthemideae). *International Journal of Plant Science* 180: 374–385.
- Álvarez, I. 2019. *Anacyclus* L. In C. Benedí, A. Buira, E. Rico, M. B. Crespo, A. Quintanar, and C. Aedo [eds.], *Flora iberica*, 16 (3), 1777–1788. Consejo Superior de Investigaciones Científicas, Madrid, Spain.
- Álvarez, I., A. B. Agudo, A. Herrero, and R. Torices. 2020. The Mendelian inheritance of gynomonocy: insights from *Anacyclus* hybridizing species. *American Journal of Botany* 107: 116–125.
- Araújo, M. B., and M. New. 2007. Ensemble forecasting of species distributions. *Trends in Ecology and Evolution* 22: 42–47.
- Armbruster, W. S., T. F. Hansen, C. Pelabon, R. Pérez-Barrales, and J. Maad. 2009. The adaptive accuracy of flowers: measurement and microevolutionary patterns. *Annals of Botany* 103: 1529–1545.
- Arnold, M. L. 1997. Natural hybridization and evolution. Oxford University Press, Oxford, UK.
- Arnold, M. L. 2006. Evolution through genetic exchange. Oxford University Press, Oxford, UK.
- Barbará, T., C. Palma-Silva, G. M. Paggi, F. Bered, M. F. Fay, and C. Lexer. 2007. Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology* 16: 3759–3767.
- Bastida, F., J. L. González-Andújar, F. J. Monteagudo, and J. Menéndez. 2010. Aerial seed bank dynamics and seedling emergence patterns in two annual Mediterranean Asteraceae. *Journal of Vegetation Science* 21: 541–550.
- Bello, M. A., I. Álvarez, R. Torices, and J. Fuertes-Aguilar. 2013. Floral development and evolution of capitulum structure in *Anacyclus*. *Annals of Botany* 112: 1597–1612.
- Bello, M. A., P. Cubas, I. Álvarez, G. Sanjuanbenito, and J. Fuertes-Aguilar. 2017. Evolution and expression patterns of *CYC/TB1* genes in *Anacyclus*: phylogenetic insights for floral symmetry genes in Asteraceae. *Frontiers in Plant Science* 8: 589.
- Breiman, L. 2001. Random forests. *Machine Learning* 45: 5–32.
- Brennan, A. C., J. R. Bridle, A.-L. Wang, S. J. Hiscock, and R. J. Abbott. 2009. Adaptation and selection in the *Senecio* (Asteraceae) hybrid zone on Mount Etna, Sicily. *New Phytologist* 183: 702–717.
- Brennan, A. C., B. Méndez-Vigo, A. Haddioui, J. M. Martínez-Zapater, F. X. Picó, and C. Alonso-Blanco. 2014. The genetic structure of *Arabidopsis thaliana* in the south-western Mediterranean range reveals a shared history between North Africa and southern Europe. *BMC Plant Biology* 14: 17.
- Bresadola, L., C. Caseys, S. Castiglione, C. A. Buerkle, D. Wegmann, and C. Lexer. 2019. Admixture mapping in interspecific *Populus* hybrids identifies classes of genomic architectures for phytochemical, morphological and growth traits. *New Phytologist* 223: 2076–2089.
- Broennimann, O., M. C. Fitzpatrick, P. B. Pearman, B. Petitpierre, L. Pellissier, N. G. Yoccoz, W. Thuiller et al. 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography* 21: 481–497.
- Broz, A. K., A. M. Randle, S. A. Sianta, A. Tovar-Méndez, B. McClure, and P. A. Bedinger. 2017. Mating system transitions in *Solanum habrochaites* impact interactions between populations and species. *New Phytologist* 213: 440–454.
- Castilla, A. R., B. Méndez-Vigo, A. Marcer, J. Martínez-Minaya, D. Conesa, F. X. Picó, and C. Alonso-Blanco. 2020. Ecological, genetic and evolutionary drivers of regional genetic differentiation in *Arabidopsis thaliana*. *BMC Evolutionary Biology* 20: 71.
- Cerca, J., A. B. Agudo, S. Castro, A. Afonso, I. Álvarez, and R. Torices. 2019. Fitness benefits and costs of floral advertising traits: insights from rayed and rayless phenotypes of *Anacyclus* (Asteraceae). *American Journal of Botany* 106: 231–243.
- Chase, J. M., and J. A. Myers. 2011. Disentangling the importance of ecological niches from stochastic processes across scales. *Philosophical Transactions of the Royal Society, B, Biological Sciences* 366: 2351–2363.
- Durvasula, A., A. Fulgione, R. M. Gutaker, S. I. Alacakaptan, P. J. Flood, C. Neto, T. Tsuchimatsu et al. 2017. African genomes illuminate the early history and transition to selfing in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA* 114: 5213–5218.

- Earl, D. A. and B. M. vonHoldt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361.
- Eckhart, V. M., M. A. Geber, and C. M. McGuire. 2004. Experimental studies of adaptation in *Clarkia xantiana*. I. Sources of trait variation across a subspecies border. *Evolution* 58: 59–70.
- Emanuelli, F., S. Lorenzi, L. Grzeskowiak, V. Catalano, M. Stefanini, M. Troggio, S. Myles et al. 2013. Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. *BMC Plant Biology* 13: 39.
- Endler, J. A. 1986. Natural selection in the wild. Princeton University Press, Princeton, NJ, USA.
- Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.
- Fenster, C. B., W. S. Armbruster, P. Wilson, M. R. Dudash, and J. D. Thomson. 2004. Pollination syndromes and floral specialization. *Annual Review of Ecology, Evolution, and Systematics* 35: 375–403.
- Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37: 4302–4315.
- Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24: 38–49.
- Francis, R. M. 2017. POPHELPER: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources* 17: 27–32.
- Friedman, J. H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics* 29: 1189–1232.
- Fu, P.-Ch., Q.-B. Gao, F.-Q. Zhang, R. Xing, J.-L., Wang, H.-R., Liu, and S.-L. Chen. 2016. Gene flow results in high genetic similarity between *Sibiraea* (Rosaceae) species in the Qinghai-Tibetan plateau. *Frontiers in Plant Science* 7: 1596.
- Germain-Aubrey, C. C., C. Nelson, D. E. Soltis, P. S. Soltis, and M. A. Gitzendanner. 2016. Are microsatellite fragment lengths useful for population-level studies? The case of *Polygala lewtonii* (Polygalaceae). *Applications in Plant Science* 4: 1500115.
- Gotelli, N. J., and J. Stanton-Geddes. 2015. Climate change, genetic markers and species distribution modelling. *Journal of Biogeography* 42: 1577–1585.
- Guisan, A., W. Thuiller, and N. E. Zimmermann. 2017. Habitat suitability and distribution models: with applications in R. Cambridge University Press, Cambridge, UK.
- Harder, L. D., and S. C. H. Barrett. 2006. Ecology and evolution of flowers. Oxford University Press, NY, NY, USA.
- He, Y.-L., Y. He, L.-L. Gong, M.-F. Fang, and Z.-H. Li. 2017. Population genetic structure and interspecific differentiation between *Acer davidii* Franchi. and *A. morrisonense* Hayata (Aceraceae) based on SSR markers. *Biochemical Systematics and Ecology* 71: 42–49.
- Hespanhol, H., K. Cezón, J. Muñoz, R. G. Mateo, and J. Gonçalves. 2022. How vulnerable are bryophytes to climate change? Developing new species and community vulnerability indices. *Ecological Indicators* 136: 108643.
- Holsinger, K. E. 2000. Reproductive systems and evolution in vascular plants. *Proceedings of the National Academy of Sciences, USA* 97: 7037–7042.
- Humphries, C. J. 1979. A revision of the genus *Anacyclus* L. (Compositae: Anthemideae). *Bulletin of the British Museum (Natural History), Botany* 7: 83–142.
- Humphries, C. J. 1981. Cytogenetic and cladistic studies in *Anacyclus* (Compositae: Anthemideae). *Nordic Journal of Botany* 1: 83–96.
- Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403–1405.
- Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.
- Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose. 2015. CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* 15: 1179–1191.
- Mandák, B., P. Vít, K. Krak, P. Trávníček, A. Havrdová, V. Hadincová, P. Zákravský et al. 2016. Flow cytometry, microsatellites and niche models reveal the origins and geographical structure of *Alnus glutinosa* populations in Europe. *Annals of Botany* 117: 107–120.
- Marcer, A., B. Méndez-Vigo, C. Alonso-Blanco, and F. X. Picó. 2016. Tackling intraspecific genetic structure in distribution models better reflects species geographical range. *Ecology and Evolution* 6: 2084–2097.
- Martínez-Minaya, J., D. Conesa, M.-J. Fortin, C. Alonso-Blanco, F. X. Picó, and A. Marcer. 2019. A hierarchical Bayesian beta regression approach to study the effects of geographical genetic structure and spatial autocorrelation on species distribution range shifts. *Molecular Ecology Resources* 19: 929–943.
- McCullagh, P., and J. A. Nelder 1989. Generalized linear models. Chapman and Hall, London, UK.
- Milanesi, P., R. Caniglia, E. Fabbri, F. Puopolo, M. Galaverni, and R. Holderegger. 2018. Combining Bayesian genetic clustering and ecological niche modeling: insights into wolf intraspecific genetic structure. *Ecology and Evolution* 8: 11224–11234.
- Oberprieler, Ch. 2004. On the taxonomic status and the phylogenetic relationships of some unispecific Mediterranean genera of Compositae-Anthemideae I. *Brocchia, Endopappus and Heliocaula*. *Willdenowia* 34: 39–57.
- Ortego, J., P. F. Gugger, and V. L. Sork. 2017. Impacts of human-induced environmental disturbances on hybridization between two ecologically differentiated Californian oak species. *New Phytologist* 213: 942–955.
- Peakall, R., and P. E. Smouse. 2012. GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28: 2537–2539.
- Piñeiro, R., J. Fuertes-Aguilar, D. D. Munt, and G. Nieto-Feliner. 2007. Ecology matters: Atlantic-Mediterranean disjunction in the sand-dune shrub *Armeria pungens* (Plumbaginaceae). *Molecular Ecology* 16: 2155–2171.
- Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- R Core Team. 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website <https://www.R-project.org/>.
- Ricklefs, R. E. 2010. Evolutionary diversification, coevolution between populations and their antagonists, and the filling of niche space. *Proceedings of the National Academy of Sciences, USA* 107: 1265–1272.
- Rieseberg, L. H. 1997. Hybrid origins of plant species. *Annual Review of Ecology, Evolution, and Systematics* 28: 359–389.
- Rosato, M., I. Álvarez, G. Nieto-Feliner, and J. A. Rosselló. 2017. High and uneven levels of 45S rDNA site-number variation across wild populations of a diploid plant genus (*Anacyclus*, Asteraceae). *PLoS One* 12: e0187131.
- Schoener, T. W. 1970. Nonsynchronous spatial overlap of lizards in patchy habitats. *Ecology* 51: 408–418.
- Streisfeld, M. A., and Kohn, J. R. 2005. Contrasting patterns of floral and molecular variation across a cline in *Mimulus aurantiacus*. *Evolution* 59: 2548–2559.
- Sung, Ch.-J., K. L., C. C. Bell, C. C. Nice, and N. H. Martin. 2018. "Integrating Bayesian genomic cline analyses and association mapping of morphological and ecological traits to dissect reproductive isolation and introgression in a Louisiana *Iris* hybrid zone." *Molecular Ecology* 27: 959–978.
- Syfert, M. M., M. J. Smith, and D. A. Coomes. 2013. The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PLoS One* 8: e55158.
- Thiers, B. 2016. Index herbariorum: a global directory of public herbaria and associated staff. New York Botanical Garden's Virtual

- Herbarium. Website <http://sweetgum.nybg.org/science/ih/> [accessed 10 September 2021].
- Thuiller, W., B. Lafourcade, R. Engler, and M. B. Araújo. 2009. BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography* 32: 369–373.
- Torices, R., A. Agudo, and I. Álvarez. 2013. Not only size matters: achene morphology affects time of seedling emergence in three heterocarpic species of *Anacyclus* (Anthemideae, Asteraceae). *Anales del Jardín Botánico de Madrid* 70: 48–55.
- Van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4: 535–538.
- Vitales, D., G. Nieto-Feliner, J. Valles, T. Garnatje, M. Firat, and I. Álvarez. 2018. A new circumscription of the Mediterranean genus *Anacyclus* (Anthemideae, Asteraceae) based on plastid and nuclear DNA markers. *Phytotaxa* 349: 1–17.
- Vitales, D., I. Álvarez, S. García, O. Hidalgo, G. N. Feliner, J. Pellicer, J. Vallès et al. 2020. Genome size variation at constant chromosome number is not correlated with repetitive DNA dynamism in *Anacyclus* (Asteraceae). *Annals of Botany* 125: 611–623.
- Wachowiak, W., S. Cavers, and W. B. Żukowska. 2015. Interspecific gene flow and ecological selection in a pine (*Pinus* sp.) contact zone. *Plant Systematics and Evolution* 301: 1643–1652.
- Warren, D. L., R. E. Glor, and M. Turelli. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62: 2868–2883.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Voucher information and locus genotyping for the samples used for the genetic analysis. Colors indicate the population category: *Anacyclus clavatus* (blue), *A.*

homogamos (red), *A. valentinus* (yellow), and phenotypically mixed populations (grey). Vouchers for each accession are kept at MA herbarium.

Appendix S2. Voucher information and locus genotyping for samples used for modeling the potential distribution of the resulting genetic clusters. Colors indicate the population category: *Anacyclus clavatus* (blue), *A. homogamos* (red), *A. valentinus* (yellow), and phenotypically mixed populations (grey). Vouchers for each accession are kept at MA herbarium.

Appendix S3. Sample information and posterior probability (Q-value) of genetic cluster (GC) membership for each sample used in the intrapopulation genetic analyses.

Appendix S4. Posterior probability (Q-value) of genetic cluster (GC) membership for each sample used for the ecological niche models of the four genetic clusters.

How to cite this article: Agudo, A. B., F. X. Picó, R. G. Mateo, A. Marcer, R. Torices, and I. Álvarez. 2023. Unravelling plant diversification: intraspecific genetic differentiation in hybridizing *Anacyclus* species in the western Mediterranean Basin. *American Journal of Botany* 110(2): e16121. <https://doi.org/10.1002/ajb2.16121>