MOLECULAR ECOLOGY

Ecological divergence combined with ancient allopatry in lizard populations from a small volcanic island

Journal:	*Molecular Ecology*
Manuscript ID:	MEC-14-0440.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Suarez, Nicolas; Universidad de Las Palmas de Gran Canaria, Departamento de Genética Pestano, Jose; Universidad de Las Palmas de Gran Canaria, Departamento de Genética Brown, Richard; Liverpool John Moores University, School of Natural Sciences & Psycholog
Keywords:	Ecological Genetics, Niche Modelling, Phylogeography, Reptiles, Speciation

Ecological divergence combined with ancient allopatry in lizard populations from a small volcanic island

N. M. Suárez[1], J. Pestano[1] and R. P. Brown[2†]

[1]Departamento de Genética, Facultad de Medicina, Universidad de Las Palmas de Gran Canaria, Las Palmas 35080, Gran Canaria, Spain.

[2]School of Natural Sciences & Psychology, Liverpool John Moores University, Liverpool L3 3AF, UK.

†Correspondence

Keywords: mtDNA; microsatellites; gene flow; ecological speciation; allopatric speciation; island

Correspondence author information:

Richard P. Brown: School of Natural Sciences & Psychology, Liverpool John Moores University, Liverpool L3 3AF, UK; Tel: (+44) 151 231-2159: r.p.brown@ljmu.ac.uk

Running Title: Ecological and allopatric divergence within an island

Abstract

Population divergence and speciation are often explained by geographical isolation, but may also be possible under high gene flow due to strong ecology-related differences in selection pressures. This study combines coalescent analyses of genetic data (11 microsatellite loci and 1 Kbp of mtDNA) and ecological modelling to examine the relative contributions of isolation and ecology to incipient speciation in the scincid lizard *Chalcides sexlineatus* within the volcanic island of Gran Canaria. Bayesian multispecies coalescent dating of within-island genetic divergence of northern and southern populations showed correspondence with the timing of volcanic activity in the north of the island 1.5-3.0 Ma ago. Coalescent estimates of demographic changes reveal historical increases in the size of northern populations, consistent with expansions from a volcanic refuge. Nevertheless, ecological divergence is also supported. First, species distribution modelling shows that the northern morph is associated with mesic habitat types and the southern morph with xeric habitat types. It seems likely that the colour morphs are associated with different anti-predator strategies in the different habitats. Second, coalescent estimation of gene copy migration (based on microsatellites and mtDNA) suggest high rates from northern to southern morphs demonstrating the strength of ecology-mediated selection pressures that maintain the divergent southern morph. Together, these findings underline the complexity of the speciation process by providing evidence for the combined effects of ecological divergence and ancient divergence in allopatry.

Introduction

Geographical isolation has traditionally been considered the main driving force behind population divergence (Mayr 1963). However, there has been considerable recent interest in ecological speciation, the process by which selection pressures promote speciation despite high gene flow (Rundle et al., 2000; Rundle and Nosil 2005; Egan et al. 2008). Under this model, divergent regions within the genome can arise due to reproductive incompatibilities or strong selection. Greatly reduced gene flow is expected in these regions, compared with higher gene flow in neutral regions (Hey 2006). Evidence for these patterns is starting to emerge (e.g., Nosil et al. 2012). Nevertheless, it may often be over-simplistic to assume that ranges and habitats have remained the same for long periods of time and that populations have diverged *in situ*. Historic geographical interruptions to gene flow may have played a role in shaping current patterns, although it is often difficult to demonstrate the combined effects of isolation and ecology (but see Thorpe et al. 1996; Thorpe and Richard 2001; Wang et al., 2013). Identification of good ecological models will help reveal new insights into the complex interplay of past and present gene flow and selection on the speciation process (Cowie and Holland 2006; Heaney 2007; Schilthuizen et al. 2011; Strasburg and Rieseberg 2011).

The Canary Island archipelago is of volcanic origin and located in the eastern Atlantic Ocean, off NW Africa. The scincid lizard *Chalcides sexlineatus* is endemic to the central island of Gran Canaria which it appears to have colonized at the beginning of the Pliocene or earlier (Brown and Pestano 1998). The island is only 1532 km^2 but reaches an altitude of 1949 m and shows strong zonation of habitat (Figure 1). Trade winds blow onto the north-facing slopes of Gran Canaria throughout the year causing relief rainfall. The north slopes are consequently more densely vegetated and were once home to laurisilva forest. In contrast, the southern slopes experience warmer, more arid conditions with low cloud and sparse vegetation. Two different

morphs of *C. sexlineatus* have been described and largely correspond to these two areas (Brown and Thorpe 1991a,b; Brown et al. 1991). The northern (N) morph has a relatively uniform brown dorsum and orange ventrum, while the southern (S) morph tends to have a black dorsum with light stripes and bright blue tail (Brown and Thorpe 1991b; Brown et al. 1991). The morphs also differ substantially in body dimensions and scalation (Brown and Thorpe 1991a). The transition between the two morphs is quite sharp but populations with intermediate morphologies are present in these regions. The correlation between morphology and habitat type, and the finding of a similar habitat-morphology association on a neighbouring island initially led Brown et al. (1991) to propose that different selection regimes had led to morphological divergence.

Geographical structuring of mtDNA is strongly N-S within Gran Canaria and concordant with the morphological variation (Pestano and Brown 1999). Timing of mtDNA divergence appears to coincide with the last major eruptive cycle on Gran Canaria, which began about 3 Ma, and covered large parts of the NE of the island (Carracedo 2011). This lends greater support to the hypothesis that the two morphological forms had originated in isolated volcanic refugia (although it should be pointed out that relatively short slowly-evolving mtDNA sequences were analysed and relationships were not fully resolved) (Pestano and Brown 1999).

While these studies have pointed to the effects of both historical isolation and natural selection, more detailed investigation is required. To achieve this we analyse: i) more informative mtDNA sequence from a larger number of individuals than analysed previously and ii) previously identified microsatellite markers (Suarez et al. 2008) to investigate divergence across the nuclear genome. Coalescent-based methods are employed to estimate levels of gene flow between populations and more rigorously date the timing of population divergence with the aim of understanding how such large morphological differences could arise within such a small

island. We also use species distribution modelling to investigate whether distributions predicted from biotic and abiotic features of the environment are distinct for the different morphotypes, allowing further evaluation of the ecological speciation hypothesis.

Materials and methods

Samples

A total of 650 *C. sexlineatus* were captured by hand from 26 evenly distributed sample sites covering the entire distribution of the species between October, 2001 and July, 2002 (Figure 1 and Supplementary Table S1). Tail-tips were removed, placed in 100% ethanol, and individuals released at the site of capture. Genomic DNA was extracted using the PureGene DNA Purification Kit (Gentra) following the manufacturer's instructions. DNA from other Canary Island *Chalcides* was also available from previous projects: 16 *C. viridanus* from Tenerife, representing the 3 main lineages within that island (Brown et al. 2000), 6 *C. coeruleopunctatus* from the islands of El Hierro (3 individuals) and La Gomera (3 individuals) and 2 *C. simonyi* from Lanzarote.

Ethics statement: Field permits were granted by the Consejeria de Medio Ambiente, Cabildo Insular de Gran Canaria.

Mitochondrial DNA amplification, sequencing and characterization

MtDNA sequences were obtained from a subsample of 137 *C. sexlineatus*, representing all sample sites within Gran Canaria (Figure 1), and also from all other Canary Island *Chalcides* from which DNA was available. A single 997-999 bp mtDNA fragment was amplified. It contained partial sequences from the NADH dehydrogenase gene subunits 1 (ND1) and 2 (ND2)

and three intervening tRNAs ($tRNA^{Ile}$, $tRNA^{Gln}$ and $tRNA^{Met}$). Polymerase chain reactions (PCRs) were carried out in 25 µl volumes using: 20–50 ng DNA, 1X buffer (Bioline; 16 mM $(NH_4)_2SO_4$, 67 mM Tris–HCl (pH 8.8) and 0.01% Tween 20), 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.2 U of Taq DNA polymerase (Bioline) and 1 µM of each primer. Primers used by previous studies were applied to all specimens, except for those from sites 19 and 38 in Gran Canaria for which new primers were designed (see Supplementary Table S2; Macey et al. 1997, Macey et al. 1998). PCR conditions were as follows: 94ºC for 3 min, followed by 30 cycles at 94ºC for 1 min, 55ºC for 1 min and 72ºC for 2 min with a final extension at 72ºC for 10 min, performed in a GeneAmp© PCR System 2700 (Applied Biosystems). PCR products were purified using MicroSpin™ S-400 HR Columns (GE Healthcare) and sequenced on an ABI PRISM 3130XL automatic sequencer (Applied Biosystems). Chromatograms were checked by eye for ambiguities and sequences were edited and aligned using ClustalW within BioEdit ver. 7.1.3.0 (Hall 1999). Estimation of genetic diversity (i.e., nucleotide and haplotype diversity) and tests of neutrality (Tajima's D [(Tajima 1989)] and Fu's F_S [(Fu 1997)]) were carried out using DnaSP ver. 5.10.1 (Librado and Rozas 2009).

Intraspecific mtDNA tree

The *C. sexlineatus* mtDNA tree was estimated using the Bayesian inference approach implemented within MRBAYES v3.1.2 (Ronquist and Huelsenbeck 2003). Sequences were partitioned into the following functional sets: 1) 1st codon positions (ND1+ND2), 2) 2nd codon positions (ND1+ND2), 3) 3rd codon positions (ND1+ND2), and 4) tRNAs. MRMODELTEST ver. 2.3 (Nylander 2004) was used to test models of molecular evolution for each partition by analyses of their log-likelihoods using the Akaike Information Criterion (AIC). Two independent

MRBAYES analyses were run from different starting points for $2x10^6$ steps and the results compared. Each run comprised four chains, with genealogies being sampled every 100 steps. MCMC performance was assessed by examination of convergence of posteriors using TRACER ver. 1.5 (Rambaut and Drummond 2007). $4x10^5$ steps were discarded as burnin. A 50% majority rule consensus tree was constructed from the post-burnin posterior tree sample from one of the runs.

Population divergence times and demographic changes

The multispecies coalescent method implemented in *BEAST ver. 1.7.4 (Heled and Drummond 2010, Drummond et al. 2012) was applied to the mtDNA with the aim of estimating time of divergence between groups within Gran Canaria using a well-established external calibration. This approach takes ancestral polymorphism into account. Species units within the analysis were generally represented by divergent lineages rather than by formally-recognized species and so will be referred to as "population groups". All described Canary Island specimens were used in the analysis. Ten population groups were defined, with two or more sequences available for each group (recognized species in parentheses): i) N Gran Canaria (*C. sexlineatus*), ii) SE Gran Canaria (*C. sexlineatus*), iii) W Gran Canaria (*C. sexlineatus*), iv) S Gran Canaria (*C. sexlineatus*), v) NE Tenerife (*C. viridanus*), vi) NW Tenerife (*C. viridanus*), vii) Central Tenerife (*C. viridanus*), viii) La Gomera (*C. coeruleopunctatus*), ix) El Hierro (*C. coeruleopunctatus*), x) Lanzarote (*C. simonyi*).

Following previous findings, monophyly constraints were applied to: i) all Gran Canaria population groups, ii) N, SE, and W Gran Canaria groups, iii) all Tenerife groups, iv) La Gomera and El Hierro groups, v) all Gran Canaria, Tenerife, La Gomera, and El Hierro population groups

(see Brown and Pestano 1998; Carranza et al. 2008). A Yule prior was used to specify divergence times across the tree. The prior on the divergence time for the (El Hierro, La Gomera) node was specified from the Gamma distribution G(12.5, 2.0), where the respective values are the shape and scale parameters, but with hard minimum and maximum limits of (0, 1.12). This provided increasing density between 0 and 1.12, reflecting prior knowledge that El Hierro was colonized from La Gomera soon after its emergence 1.12 Ma (this is supported by the degree of sequence divergence described by Brown and Pestano 1998). A prior hard maximum bound of 11.6 Ma was placed on the node that was most basal to all population groups from *C. sexlineatus*, *C. viridanus*, and *C. coeruleopunctatus*. This corresponded to the age of the second oldest of the islands on which they are found (Tenerife). The rationale for this prior is that at least two emerged islands must have been present to allow dispersal-mediated speciation. Finally, a maximum bound of 20.6 Ma was placed on the root node. This represents the time of appearance of the first (eastern) Canary Island, and appears to considerably predate the divergence time of the *Chalcides* group containing *C. simonyi* from the other Canary Island *Chalcides* (which has been previously estimated at around 7 Ma; Carranza et al. 2008).

Sequences were partitioned into: i) codons 1 and 2, ii) codon 3, and iii) tRNAs. The HKY+G model of substitution was applied to each partition and a lognormal uncorrelated rates relaxed clock model used. The *Chalcides* tree was quite shallow which can have the effect of making some priors (such as the prior on times) quite influential, in the absence of suitable prior knowledge, particularly under a relaxed clock (Brown and Yang 2010; Brown and Yang 2011). Hence, the results were compared with those from a strict clock analysis. MCMC chains were run for $4x10^7$ cycles sampled at intervals of 2000, providing 20000 samples from the posterior, of which the first 2000 were discarded as burnin, leaving 18000 samples for analysis.

Historical demographic changes in the four Gran Canarian population groups were analysed using Bayesian skyline plots (BSPs) under the piecewise-constant model (Drummond et al. 2005). Priors on rates of the four partitions were specified using normal distributions, the means and variances of which were derived from the posterior distributions of rates from the dating analyses. The BSP approach requires user-specification of the number of groups of coalescent intervals (this reduces potential noise associated with a large number of short intervals: Drummond et al. 2005). We specified 4 groups, but results were similar when larger numbers of groups (up to 10) were tested.

Nuclear DNA amplification, genotyping and characterization

Eleven autosomal microsatellite loci were analysed for all 650 individuals. All loci contained tetranucleotide (AAAG) repeats. We use the same locus names and multiplex PCR protocol described previously (Suarez et al. 2008). Genotyping was performed on an ABI PRISM 3130XL genetic analyser (Applied Biosystems) with G5 matrix and GeneScan-500 (LIZ) as size standard. Alleles were scored using GeneMapper v4.0 software (Applied Biosystems). Measures of genetic diversity and other statistics were obtained using ARLEQUIN version 3.11 (Excoffier et al. 2005) and FSTAT 2.9.3 (Goudet 1995).

Genetic structure (microsatellite DNA)

Genetic structuring of nuclear DNA was inferred by application of the model-based clustering method implemented in the program STRUCTURE ver.2.3.4 (Pritchard et al. 2000) to all 650 specimens. An admixture model with correlated allele frequencies among populations was applied. Twenty STRUCTURE runs (chain length = 10^6 steps, burn-in = 10^5) were performed for

different numbers of genetic clusters (*K*) between 1 and 10 (see Gilbert et al. 2012). We used STRUCTURE HARVESTER web version 0.6.92 (Earl and vonHoldt 2012) to analyse the output using the Δ*K* metric approach proposed by Evanno et al. (2005). This provides an objective and therefore preferable alternative to simply selecting *K* according to the magnitude of its log-likelihood (which may lead to overestimation of the number of genetic clusters; Evanno et al. 2005). Prior information on the origin of each sampled individual was not used in the analysis. CLUMPP (Jakobsson and Rosenberg 2007) was used to concatenate the data from the multiple runs for each *K* and assign individuals to clusters using their membership coefficient (Q). A threshold value of Q = 0.2 was used because it is efficient and accurate at differentiating between purebreds and hybrids (Vaha and Primmer 2006).

Analysis of migration and isolation

Estimation of timing of divergence and migration between the two main morphotypes within Gran Canaria was carried out using the coalescent method implemented within IMa2 (Hey and Nielsen 2007; Hey 2010). Sampled locations were assigned to either N or S morphs according to geographical position relative to the midpoint of the morphological variation that has been described previously (see Figure 1, Brown and Thorpe 1991b and Brown et al. 1991 for more details). All 137 mtDNA sequences (60 N and 77 S morphs) and 50 microsatellite genotypes (25 N and 25 S morphs with representatives from all 26 sites, for all 11 loci), were analysed. The microsatellite data had to be subsampled in this way because several months were required to run the MCMC chains for the complete data set.

The HKY model of DNA substitution (Hasegawa et al. 1985) was used for the mtDNA fragment, and the stepwise mutation model (SMM; Kimura and Ohta 1978) was used for the

microsatellites. Following preliminary runs using diffuse priors, tighter uniform priors were specified: U(0,6) on divergence time, U(0,300) on population sizes and U(0,1) on migration rates. Consistency of results of results was compared between three replicate runs starting from different positions. A final definitive MCMC chain was run for 1.01×10^8 steps, with parameters sampled every 100 steps, and the first 1×10^6 steps discarded as burnin.

In order to convert the estimates into more interpretable demographic units, a generation time of 2 years was used (derived from personal observations and evidence that similar species from cooler regions in northern Spain reach sexual maturity within 2-3 years: Galan 2003). A mutation rate of 1.8351×10^{-5} mutations/locus/year for the entire mtDNA sequence was determined from the *BEAST analysis of divergence times. IMa2 provides migration rates looking backwards in time, but here we present the results in the more intuitive forward direction of time.

Species distribution modelling and spatial analyses

Species distribution models (SDMs) were constructed separately for the N and S morphs using the maximum entropy algorithm implemented in MAXENT ver. 3.3.3 (Phillips et al. 2006). We used the coordinates of the 46 sample localities (25 N sites and 21 S sites) in Brown and Thorpe 1991a,b as evidence of presence (Supplementary Table S3). Note that the sample sites used here are a subset of these 46 sites.

The environment was modelled from a subset of 56 climatic layers obtained from the WorldClim global climate database (http://www.worldclim.org). The climatic layers had a spatial resolution of 30 arc-seconds (ca. $1 Km^2$). A categorical variable representing potential vegetation was obtained from land characterisation maps published by the Canary Island

Government (http://visor.grafcan.es/visorweb/). Seventeen vegetation categories were used with presence of each vegetation type being recorded for each sample square using the viewer tool (30 arc-seconds grid) provided on the database (Supplementary Figure 1).

There was no prior biological evidence to support objective determination of suitable climate predictors in the MAXENT analyses. In addition, most climate predictors were correlated. We therefore used two approaches to select climatic variables: 1) after preliminary runs using all 56 climate variables, a subset of 6 variables was determined according to permutation importance which is an indirect estimator of the dependence of the model on the selected variable (see MAXENT documentation and Supplementary Table S3), 2) just two uncorrelated climate variables were selected: precipitation seasonality (which also had high permutation importance) and temperature seasonality. In both analyses, the selected climatic variables were combined with potential vegetation and the results compared.

The N and S morph SDMs were tested for statistical significance by comparison of the observed area under the curve (AUC) for the receiver operating characteristic (ROC) plot with the same AUCs obtained by random sampling of the same number of sample squares (Raes and ter Steege 2007). This null model approach prevents interpretation of model quality using an arbitrary AUC threshold and removes the need to set aside samples for model testing. Randomized point data were created with ENMTools ver. 1.3 (Warren et al. 2010). A total of 500 AUCs were generated (including the observed AUC). Statistical significance was established when the magnitude of the observed AUC was equal or greater than the value of the 475th rank-ordered AUC (corresponding to $P \leq 0.05$).

We examined niche overlap using a principal components analysis (PCA) on the two subsets of climatic variables. Schoener's *D* index was used to test for niche overlap between N

and S morphs, with its significance being tested using a randomization test (see Warren et al., 2008). *D* can take values from 0 (no overlap) to 1 (complete overlap). Significance of *D* was achieved by comparison with 100 datasets containing random partitions of N and S occurrences.

Results

MtDNA diversity and phylogeography

Lengths of genes/partial genes that were sequenced were as follows: ND1, 307 bp; tRNAs, 215-217 bp ($tRNA^{Ile}$, 77-79 bp; $tRNA^{Gln}$, 71 bp; $tRNA^{Met}$, 67 bp); ND2, 475 bp (GeneBank accession numbers: KJ463905-KJ464030). One hundred and seven haplotypes were detected within *C. sexlineatus*, with polymorphisms at 234 sites (Supplementary Table S4).

The four main mitochondrial lineages were designated as N, S, SE and W according to their distributions within Gran Canaria (Figure 2). The basal node representing divergence between the (N, SE, W) and S lineages was strongly supported. Most sample sites provided individuals from a single mtDNA lineage although two or three lineages were identified at three sites (site 14: S and E lineages detected, site 27: N and SE lineages, and site 10: N, S and SE lineages).

Nucleotide diversity was lowest in the SE mtDNA lineage and highest in the W lineage (Table 1). In the N lineage, there was a significant deviation from neutral expectation for Tajima's D and Fu's F_S, a small Rozas' R^2 and a significant signal in sequence mismatch distributions, consistent with recent expansion/dispersal (Table 1). BSPs provided evidence of two substantial increases in population size in this lineage, one of which was during the last 50 ka. Evidence of less-pronounced increases in population sizes of the S and SE lineages was also detected by the BSPs (Figure 3).

Population divergence times

A likelihood ratio test was used to compare the likelihoods of mtDNA trees with and without a constant rate assumption (HKY+G model) and revealed significant violation of the clock ($2\Delta l$=460.19, P<0.0001). However, the *BEAST posterior median divergence time for the basal Gran Canaria node was similar under the strict (2.00 Ma [95% HPD: 0.88-3.59 Ma]) and the relaxed clock analyses (1.94 Ma [0.86-3.46]) (Figure 4). This was the case for all other nodes, such as the root (strict clock: 7.97 Ma [3.93-14.02], relaxed clock: 7.77 Ma [3.76-13.67]). Hence, only relaxed clock estimates will be discussed from here onwards.

Population genetic analysis of microsatellite loci

Microsatellite polymorphism was high: the number of alleles per locus ranged from 23 (locus *Csex*11) to 38 (locus *Csex*01), with a mean of 28.6 (site summary statistics are in Supplementary Table S5). There was significant deviation from HWE for some loci within populations, but there was no clear pattern across localities. There was significant LD between some pairs of loci (after Bonferroni correction), which appeared slightly more prevalent in populations with intermediate N/S morphologies (Supplementary Table S6). Allelic differentiation among the 26 samples was significant (Fisher's method; $P< 0.01$), allowing rejection of the null hypothesis that alleles are drawn from the same distribution in all samples. All between-site pairwise F_{ST}'s were also significant ($P<0.01$ in all cases) which implied major genetic differentiation (results not shown).

Two genetically distinct clusters were detected by STRUCTURE/STRUCTURE HARVESTER (highest value of $\Delta K = 52.67$) (Figure 5A,B). Although the approach used cannot reject one genetic cluster (K=1), the significant genetic differentiation and clear geographical

structuring of the two genetic clusters rule this out: clusters were closely associated with the N/S variation in morphology. Also, sites containing individuals with Q values around 0.2-0.8 (indicative of hybridization between individuals from different clusters) were most prevalent in areas of greatest morphological transition (Figure 5C). For example, the highest proportions of hybrid individuals (>40%) were found at sites 6 and 19.

Analysis of N-S migration

Replicated IMa2 analyses that started from different positions converged on the same posterior. The value of t corresponding to the highest posterior density (HiPt) scaled in years, was 258.5 ka (95% HPD: 168.3-630.5 ka). Population migration (2NM) is estimated as effective number of migrating gene copies per generation and was found to be high from the N to the S morph(HiPt: 3.54, 95% HPD: 1.03-8.79) and differed significantly from zero (LRT: $2\Delta l$=7.487, P< 0.01) (Figure 6). The posterior on 2NM for migration of gene copies from the S to N morph was lower (HiPt: 0.022, 95% HPD: 0.00-6.00) and did not significantly differ from zero (LRT: $2\Delta l$=0.012, P>0.05).

Species distribution modelling

The contribution of the available predictor variables varied considerably (Supplementary Table S3), but potential vegetation (20-23%) was most influential, followed by precipitation seasonality (6-8%). We selected the 7 variables that had a permutation importance of >5 for either the northern and/or the southern morph for use in SDM modelling of both N and S morphs. Generally high correlations were found between all climate variables (r>0.95) although temperature and precipitation seasonality showed generally low correlations (r<0.62) and so we

used these to provide alternative analyses of uncorrelated variables. Using the six climatic variables of high permutation importance plus potential vegetation, the species distribution models showed a high discriminatory power between presences and background. The AUCs for the calibration data sets were 0.823 for the northern morph, and 0.855 for the southern one (i.e., 82.3 and 85.5% of the records were correctly predicted, respectively). Randomization tests revealed that the AUCs were significant for both the northern (P=0.008) and the southern (P=0.024) morphs. Results were similar when we used just the two uncorrelated climatic variables (plus potential vegetation) instead of the six climatic variables with highest permutation importance. The SDMs were spatially non-overlapping for the N and S morphs indicating distinct environmental requirements (Figure 7).

Comparison of climate niche overlap between N and S morphs revealed significant deviation from the null distribution, indicating non-equivalence of niches between N and S morphs. This finding appears to be robust as it was supported by the analysis of 6 climatic variables with highest permutation importance (D=0.522, P= 0.0198) as well as the alternative analysis of just two uncorrelated climatic variables (D=0.439, P= 0.0198).

Discussion

We find evidence that within-island incipient speciation in *C. sexlineatus* is associated with both current ecological conditions and historical divergence in allopatry. We corroborate initial findings that the latter may have been mediated by volcanic activity through the creation of two or more disjunct habitat refuges within the island (Pestano and Brown 1999). This was achieved using more informative mtDNA sequences which allowed the detection of four well-supported mtDNA lineages (as opposed to three weakly supported lineages in Pestano and Brown [1999]).

We also analysed nuclear markers for the first time. The within-island morphological variation appears to correspond more closely to geographical structuring of nuclear microsatellite polymorphisms than to mtDNA phylogeography. This is not too surprising given that morphological differences should originate from divergence in the nuclear genome. Greatest levels of admixture are found in areas of intermediate morphology, as would be expected in a hybrid zone.

Environment-based distribution models of N and S morphs indicate close correspondence to the respective ecologically distinct N and S regions. This strengthens previous inferences that the morphs represent ecological forms that are adapted to the xeric and mesic habitat types (Brown et al. 1991). The isolation-with-migration analysis indicates that nuclear/mitochondrial gene flow is asymmetric with relatively high migration of gene copies from the N to the S morph but lower migration in the opposite direction. The N to S morph population migration estimate (3.5 migrant gene copies per generation) is much higher than the 0-1 range at which divergence is impeded (Slatkin 1995). Hence, strong habitat-related selection in the face of relatively high gene flow could explain the origin and maintenance of the southern morph, as expected under ecological divergence. The hypothesis that the southern morph has been subject to strong directional selection is further supported by the observation that it has quite a divergent morphology compared with the other Canary Island skinks of the same clade, at least in terms of colour which is indicative of different chromatophores in the skin (Kuriyama et al. 2006; Carranza et al. 2008). In contrast, the influx of southern gene copies into the northern morph appears negligible, possibly explaining why it remains morphologically distinct from the southern morph. An inability to survive and reproduce in foreign habitats has been described in several taxa (see Nosil et al. 2005) and could potentially explain this restricted introgression. The

present data do not reveal why gene flow appears to be asymmetric.

How the two morphs may have evolved under different selection pressures has been discussed previously (Brown and Thorpe 1991b). It was postulated that uniform brown northern skinks were suited to the more mesic N areas because they allow a more cryptic anti-predation strategy against birds such as the kestrel. The bright blue-tailed skinks appear suited to the more open xeric S areas where crypsis may be less successful. Escape from anti-predator attacks in these open habitats would be achieved by attracting predatory attacks towards the tail, which can autotomize increasing the chances of escape. It has been observed that lizards with distinctive tail colorations tend to be associated with more open habitats (Arnold 1984) which fits well with the pattern on Gran Canaria. The finding of a parallel, albeit weaker, pattern of tail colour variation in *Chalcides* from the neighbouring island of Tenerife, also supports this hypothesis (Brown et al. 1991).

Despite support for ecological divergence there is also clear evidence of additional historical vicariance. Population divergence began during the early Pleistocene or the late Pliocene and is one element that supports the role of the last major eruptive cycle in Gran Canaria, 1.5-3 Ma ago. During this period, eruptions covered most of the NE of the island with lava to approximately 500m depth, except for an isolated area in the extreme NE close to the current location of the city of Las Palmas (Carracedo 2011, and references therein) providing an isolated northern refuge. The south of the island was largely unaffected by these eruptions. Hence the spatial correspondence between these eruptions and the observed N/S genetic pattern adds support to the volcanism-mediated geographical isolation hypothesis. Finally, the finding of a strong signal of population increase found in the N mtDNA is concordant with range expansion from a northern volcanic refuge. The earliest split between the W, SE and N mtDNA lineages

(1.5 Ma ago) also fits in with the timing of northern volcanic activity and may have been part of this process, although the most recent split between these clades (0.4 Ma ago) is clearly too early to be associated with this period.

It is worth considering why the phylogenetic multispecies coalescent analysis provided a very different estimate of N-S divergence time (2 Ma) to the analysis of isolation-with-migration (260 ka) with non-overlapping posterior intervals. The isolation-with-migration analysis depends on an estimate of generation time (and of mutation rate but this was derived from the *BEAST analysis) but even an error as large as 50% in this estimate would not explain the difference. Instead, it is more likely to be because our *BEAST analyses simply date the divergence between distinct mtDNA lineages, equivalent to an analysis on two completely sorted populations. This was a suitable approach given that multispecies coalescent analyses do not take gene flow into account. In contrast, the IMa2 analysis examines both splitting time and microsatellite/mtDNA gene flow between the two morphological groups, which share microsatellite alleles and mtDNA haplotypes. Thus, the *BEAST analyses should provide a better estimate of divergence time of mtDNA lineages, while IMa2 may confound divergence time with levels of gene flow. Nevertheless, if IMa2 incorrectly attributed greater similarity between populations to more recent divergence rather than high gene flow, then this would lead to migration of gene copies being underestimated which would not affect our inferences.

One final cautionary point about the IMa2 analyses is that we cannot establish the relative influences of mtDNA and microsatellites on the results. Test analyses on microsatellite loci alone did not provide reliable posterior distributions and therefore are not helpful. Clearly, migration of nuclear alleles should be more relevant to morphological divergence than mtDNA migration, but we cannot decisively show that a significant component of the observed migration is accounted

for by microsatellite alleles.

In summary, our analyses support the ecological origins of the two primary skink morphs because their current distributions can be largely predicted from bioclimatic modelling. The finding of high rates of migration of gene copies from N to S suggest that these differences are maintained by strong selection pressures, at least within the arid southern habitats. These effects seem to be additional to ancient population vicariance mediated by Pleistocene volcanic activity in NE Gran Canaria. Studies of population divergence frequently focus on one particular causal mechanism in isolation, but here we show how different processes can combine to shape genetic and morphological diversity within a very small geographic area.

Acknowledgements

NMS acknowledges the financial support provided by the Ministerio de Educación y Cultura (AP98-1999-02-05 N). RPB was funded by an EU Marie Curie senior fellowship (HPMF-CT-2000-00886).We wish to thank the Consejeria de Medio Ambiente of the Cabildo Insular de Gran Canaria for fieldwork permits. The authors declare no conflict of interest (relationship, financial or otherwise). We thank four anonymous referees for their comments on an earlier draft of this MS.

References

Arnold EN (1984) Evolutionary Aspects of Tail Shedding in Lizards and Their Relatives. *Journal of Natural History* **18**, 127-169.

Brown R, Thorpe R, Báez M (1991) Parallel within-island microevolution of lizards on neighbouring islands. *Nature* **352**, 60-62.

462 Brown RP, Campos-Delgado R, Pestano J (2000) Mitochondrial DNA evolution and population
463 history of the Tenerife skink *Chalcides viridanus*. *Mol Ecol* **9**, 1061-1067.
464 Brown RP, Pestano J (1998) Phylogeography of skinks (*Chalcides*) in the Canary Islands
465 inferred from mitochondrial DNA sequences. *Mol Ecol* **7**, 1183-1191.
466 Brown RP, Thorpe RS (1991a) Within-island microgeographic variation in body dimensions and
467 scalation of the skink *Chalcides sexlineatus*, with testing of causal hypotheses. *Biological*
468 *Journal of the Linnean Society* **44**, 47-64.
469 Brown RP, Thorpe RS (1991b) Within-island microgeographic variation in the colour pattern of
470 the skink, *Chalcides sexlineatus*: Pattern and cause. *J Evol Biol* **4**, 557-574.
471 Brown RP, Yang Z (2010) Bayesian dating of shallow phylogenies with a relaxed clock. *Syst*
472 *Biol* **59**, 119-131.
473 Brown RP, Yang ZH (2011) Rate variation and estimation of divergence times using strict and
474 relaxed clocks. *BMC Evol Biol* **11**.
475 Carracedo JC (2011) *Geología de Canarias: Origen, evolución, edad y volcanismo. I* Editorial
476 Rueda.
477 Carranza S, Arnold EN, Geniez P, Roca J, Mateo JA (2008) Radiation, multiple dispersal and
478 parallelism in the skinks, *Chalcides* and *Sphenops* (Squamata: Scincidae), with comments
479 on Scincus and Scincopus and the age of the Sahara Desert. *Mol Phylogenet Evol* **46**,
480 1071-1094.
481 Cowie RH, Holland BS (2006) Dispersal is fundamental to biogeography and the evolution of
482 biodiversity on oceanic islands. *Journal of Biogeography* **33**, 193-198.
483 Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past
484 population dynamics from molecular sequences. *Mol Biol Evol* **22**, 1185-1192.

485 Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUti
486 and the BEAST 1.7. *Mol Biol Evol* **29**, 1969-1973.
487 Earl D, vonHoldt B (2012) STRUCTURE HARVESTER: a website and program for visualizing
488 STRUCTURE output and implementing the Evanno method. *Conservation Genetics*
489 *Resources* **4**, 359-361.
490 Egan SP, Nosil P, Funk DJ (2008) Selection and genomic differentiation during ecological
491 speciation: isolating the contributions of host association via a comparative genome scan
492 of *Neochlamisus bebbianae* leaf beetles. *Evolution* **62**, 1162-1181.
493 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the
494 software STRUCTURE: a simulation study. *Mol Ecol* **14**, 2611-2620.
495 Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package
496 for population genetics data analysis. *Evol Bioinform Online* **1**, 47-50.
497 Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking
498 and background selection. *Genetics* **147**, 915-925.
499 Galan P (2003) Female reproductive characteristics of the viviparous skink *Chalcides bedriagai*
500 pistaciae (Reptilia, Squamata, Scincidae) from an Atlantic beach in north-west Spain.
501 *Amphibia-Reptilia* **24**, 79-85.
502 Gilbert KJ, Andrew RL, Bock DG*, et al.* (2012) Recommendations for utilizing and reporting
503 population genetic analyses: the reproducibility of genetic clustering using the program
504 STRUCTURE. *Mol Ecol* **21**, 4925-4930.
505 Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. *J Heredity*
506 **86**, 485-486.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT **41**, 95-98.

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* **22**, 160-174.

Heaney LR (2007) Is a new paradigm emerging for oceanic island biogeography? *Journal of Biogeography* **34**, 753-757.

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Mol Biol Evol* **27**, 570-580.

Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. *Curr Opin Genet Dev* **16**, 592-596.

Hey J (2010) Isolation with migration models for more than two populations. *Mol Biol Evol* **27**, 905-920.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc Natl Acad Sci USA* **104**, 2785-2790.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-1806.

Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proc Natl Acad Sci USA* **75**, 2868-2872.

Kuriyama T, Miyaji K, Sugimoto M, Hasegawa M (2006) Ultrastructure of the dermal chromatophores in a lizard (Scincidae: *Plestiodon latiscutatus*) with conspicuous body and tail coloration. *Zoological Science* **23**, 793-799.

529 Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA
530 polymorphism data. *Bioinformatics* **25**, 1451-1452.
531 Macey JR, Larson A, Ananjeva NB, Fang Z, Papenfuss TJ (1997) Two novel gene orders and the
532 role of light-strand replication in rearrangement of the vertebrate mitochondrial genome.
533 *Mol Biol Evol* **14**, 91-104.
534 Macey JR, Schulte JA, 2nd, Larson A*, et al.* (1998) Phylogenetic relationships of toads in the
535 *Bufo bufo* species group from the eastern escarpment of the Tibetan Plateau: a case of
536 vicariance and dispersal. *Mol Phylogenet Evol* **9**, 80-87.
537 Mayr E (1963) *Animal species and evolution* Belknap Press of Harvard University Press,
538 Cambridge, MA.
539 Nosil P, Parchman TL, Feder JL, Gompert Z (2012) Do highly divergent loci reside in genomic
540 regions affecting reproductive isolation? A test using next-generation sequence data in
541 Timema stick insects. *BMC Evol Biol* **12**, 164.
542 Nosil P, Vines TH, Funk DJ (2005) Perspective: Reproductive isolation caused by natural
543 selection against immigrants from divergent habitats. *Evolution* **59**, 705-719.
544 Nylander J (2004) MrModeltest v2. Program distributed by the author. *Evolutionary Biology*
545 *Centre, Uppsala University*.
546 Pestano J, Brown RP (1999) Geographical structuring of mitochondrial DNA in *Chalcides*
547 *sexlineatus* within the island of Gran Canaria. *Proc Roy Soc B-Biol Sci* **266**, 805-812.
548 Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species
549 geographic distributions. *Ecological modelling* **190**, 231-259.
550 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus
551 genotype data. *Genetics* **155**, 945-959.

552 Raes N, ter Steege H (2007) A null-model for significance testing of presence-only species
553 distribution models. *Ecography* **30**, 727-736.
554 Rambaut A, Drummond A (2007) Tracer v. 1.5. Computer program and documentation
555 distributed by the authors.
556 Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population
557 growth. *Mol Biol Evol* **19**, 2092-2100.
558 Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed
559 models. *Bioinformatics* **19**, 1572-1574.
560 Rundle HD, Nosil P (2005) Ecological speciation. *Ecol Lett* **8**, 336-352.
561 Rundle, HD, Nagel L., Boughman JW, Schluter D (2000) Natural selection and parallel
562 speciation in sympatric sticklebacks. *Science*, **287**(5451), 306-308.
563 Schilthuizen M, Giesbers MC, Beukeboom LW (2011) Haldane's rule in the 21st century.
564 *Heredity (Edinb)* **107**, 95-102.
565 Strasburg JL, Rieseberg LH (2011) Interpreting the estimated timing of migration events
566 between hybridizing species. *Mol Ecol* **20**, 2353-2366.
567 Suarez NM, Bloor P, Brown RP, Pestano J (2008) Highly polymorphic microsatellite loci for the
568 Gran Canarian skink (*Chalcides sexlineatus*) and their applicability in other Canarian
569 Chalcides. *Mol Ecol Resour* **8**, 666-668.
570 Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA
571 polymorphism. *Genetics* **123**, 585-595.
572 Thorpe RS, Black H, Malhotra A (1996) Matrix correspondence tests on the DNA phylogeny of
573 the Tenerife lacertid elucidate both historical causes and morphological adaptation.
574 *Systematic Biology* **45**, 335-343.

Thorpe RS, Richard M (2001) Evidence that ultraviolet markings are associated with patterns of molecular gene flow. *Proc Natl Acad Sci U S A* **98**, 3929-3934.Vaha JP, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Mol Ecol* **15**, 63-72.

Wang, IJ, Glor, RE, Losos, JB (2013). Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters* **16**, 175-182.

Warren DL, Glor RE, Turelli M (2008) Environmentalniche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* **62**, 2868–2883.

Warren DL, Glor RE, Turelli M (2010) ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* **33**, 607-611.

Wright S (1931). Evolution in Mendelian populations. *Genetics* **16**, 96-159.

Data Accessibility

- DNA sequences: GeneBank accession numbers: KJ463905-KJ464030
- Microsatellite genotypes: Dryad doi: http://dx.doi.org/10.5061/dryad.db451/1
- MtDNA alignment and partition data: Dryad doi: http://dx.doi.org/10.5061/dryad.db451/2
- Occurrence data: Dryad doi: http://dx.doi.org/10.5061/dryad.db451/3
- Species Distribution Modelling data: Dryad doi: http://dx.doi.org/10.5061/dryad.db451/4

Author Contributions

596 This work originated from NMS's PhD that he carried out in JPs laboratory at the University of
597 Las Palmas. The study was originally formulated by RPB during an EU research fellowship held
598 at the University of Las Palmas. NMS and RPB recently reanalysed the data and wrote the paper.
599

600 **Supporting Information**

601 Figures S1

602 Tables S1-6

603 **Figure Legends**

604 Figure 1. Geographical locations of *C. sexlineatus* sample sites. The line across the island
605 represents the midpoint of the N/S morphological variation (Brown& Thorpe 1991b).

606

607 Figure 2. The 50% majority rule consensus of the posterior mtDNA trees obtained from the
608 Bayesian analysis. Bayesian posterior probabilities are shown at each node. The geographical
609 distributions of the four main lineages are shown on the map, as well as the areas affected by
610 volcanism (dark shading: Holocene volcanism, medium shading: rift volcanism1.5-3 Ma, light
611 shading: inferred rift volcanism with the rift axis shown as a dotted line (adapted from
612 (Carracedo 2011).

613

614 Figure 3. Bayesian skyline plots showing estimated demographic changes over time in the four
615 mtDNA lineages. Lines represent posterior medians (continuous), upper and lower 95% HPDs
616 (dotted).

617

618 Figure 4. *BEAST population tree chronogram. Median posterior ages of nodes are provided,
619 together with bars representing 95% HPDs. Scale bar provides times in millions of years.

620

621 Figure 5. Genetic structure inferred from microsatellites using STRUCTURE. A) Individual
622 assignment to clusters (K=2) based on B) ΔK (Evanno *et al.* 2005). C) Site compositions.

623

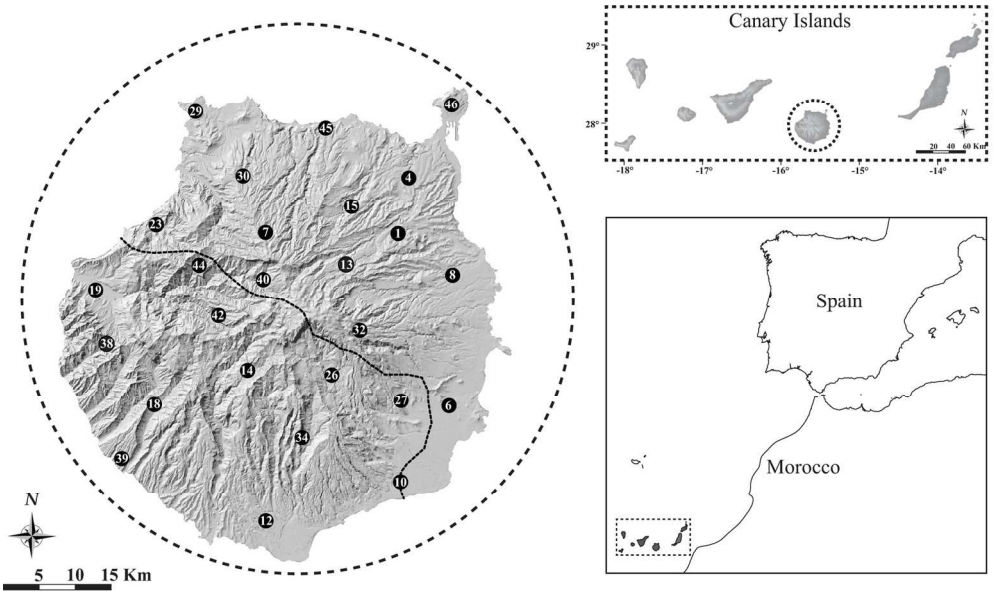
624 Figure 6. Posterior densities for population migration (2NM) estimated using IMa2.

625

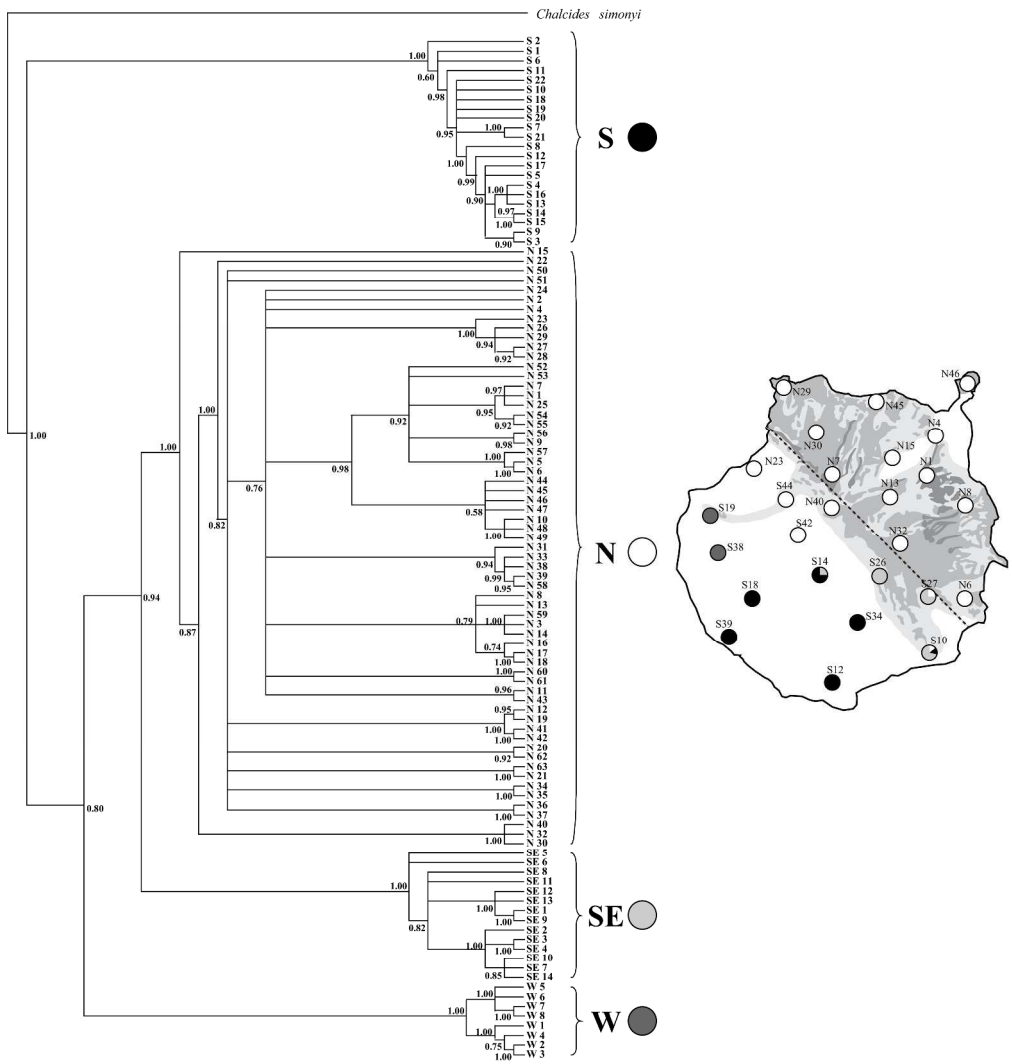
626 Figure 7. Species distribution models for the northern and southern morphs. Higher values
627 indicate higher predicted environmental suitability.
628
629

Table 1. Summary statistics for the four main mtDNA lineages identified in *C. sexlineatus*: *n*, number of individuals; PS, number of polymorphic sites; NH, number of haplotypes; R^2, Ramos-Onsins and Rozas statistic (Ramos-Onsins& Rozas 2002). *P<0.1, **P<0.05, ***P<0.001.

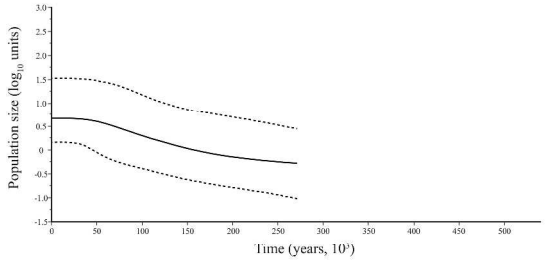
Lineage	***n***	**PS**	**Parsimony informative sites**	**NH**	**Haplotype diversity**	**Nucleotide diversity**	**R^2**	**Fu´s F_S(1997)**	**Fu & Li´s D (1993)**	**Fu & Li´s F (1993)**	**Tajima´s D (1989)**
North	78	139	89	63	0.994	0.012	0.0401^{**}	-48.11^{***}	-1.820^{ns}	-2.280^{ns}	-2.016^{**}
South	23	70	28	22	0.996	0.012	0.0627^{**}	-10.90^{***}	-2.112^{ns}	-2.282^{ns}	-1.580^{ns}
South-East	19	35	19	14	0.953	0.007	0.0906^{*}	-3.38^{*}	-0.858^{ns}	-1.043^{ns}	-0.992^{ns}
West	17	51	36	8	0.838	0.017	0.1585^{ns}	5.16^{ns}	0.035^{ns}	0.218^{ns}	0.566^{ns}
All	137	234	183	107	0.995	0.036	0.0743^{ns}	-46.57^{***}	-0.990^{ns}	-1.088^{ns}	-0.813^{ns}



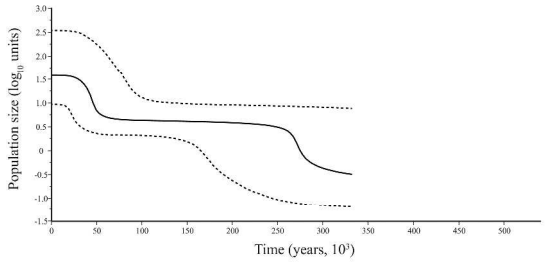
Geographical locations of C. sexlineatus sample sites. The line across the island represents the midpoint of the N/S morphological variation (Brown& Thorpe 1991b).
156x91mm (300 x 300 DPI)



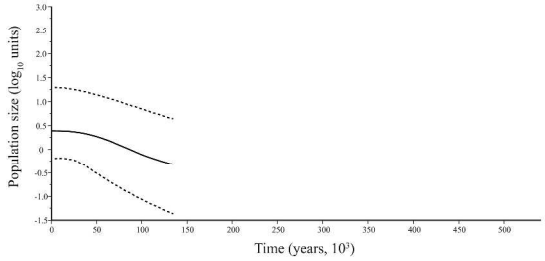
The 50% majority rule consensus of the posterior mtDNA trees obtained from the Bayesian analysis. Bayesian posterior probabilities are shown at each node. The geographical distributions of the four main lineages are shown on the map, as well as the areas affected by volcanism (dark shading: Holocene volcanism, medium shading: rift volcanism1.5-3 Ma, light shading: inferred rift volcanism with the rift axis shown as a dotted line (adapted from (Carracedo 2011).
346x365mm (300 x 300 DPI)



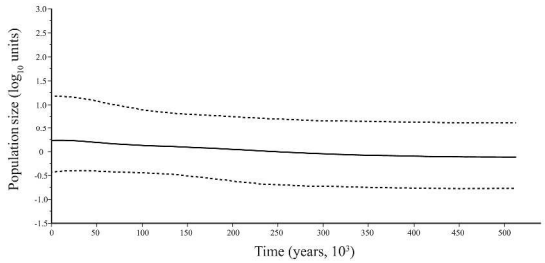
S



N

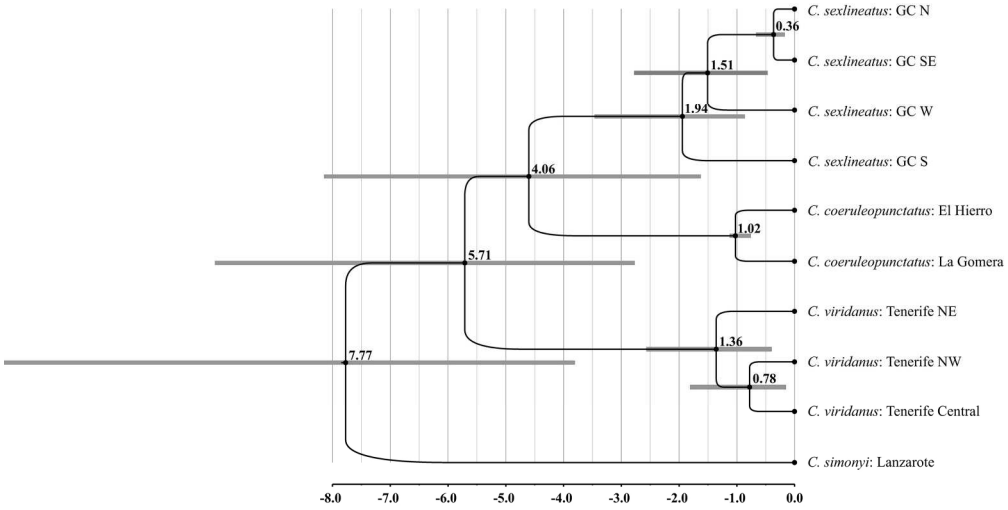


SE

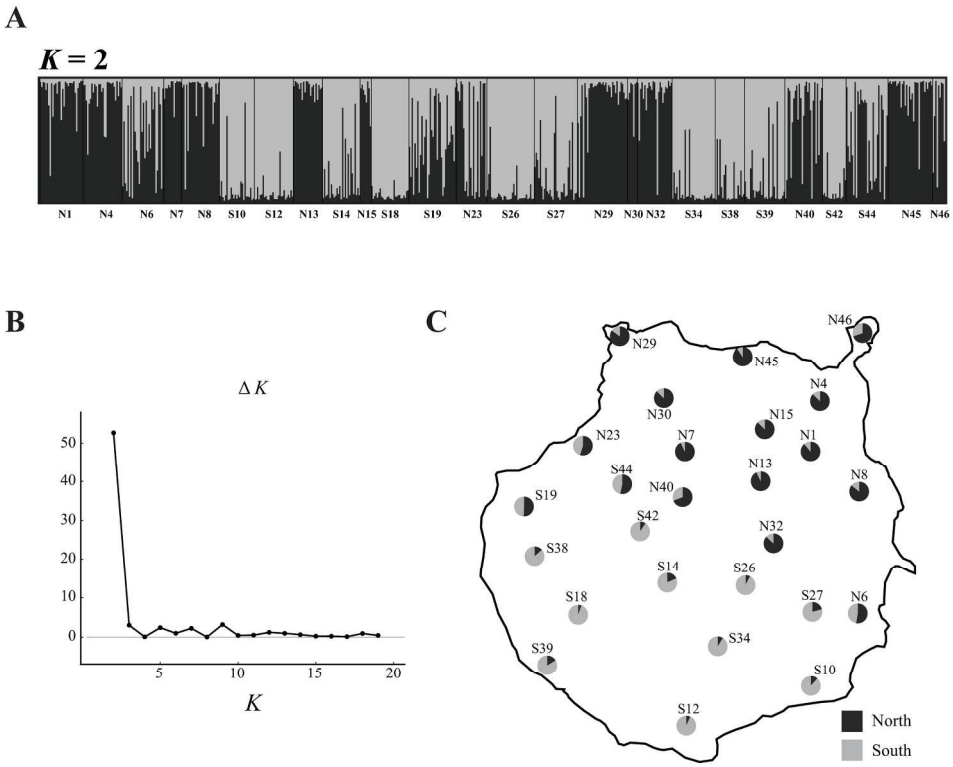


W

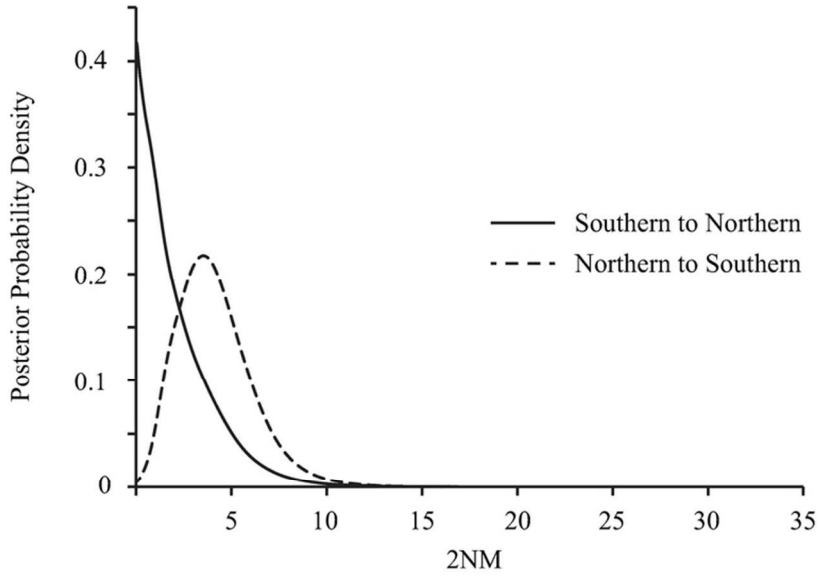
Bayesian skyline plots showing estimated demographic changes over time in the four mtDNA lineages. Lines represent posterior medians (continuous), upper and lower 95% HPDs (dotted).
449x695mm (600 x 600 DPI)



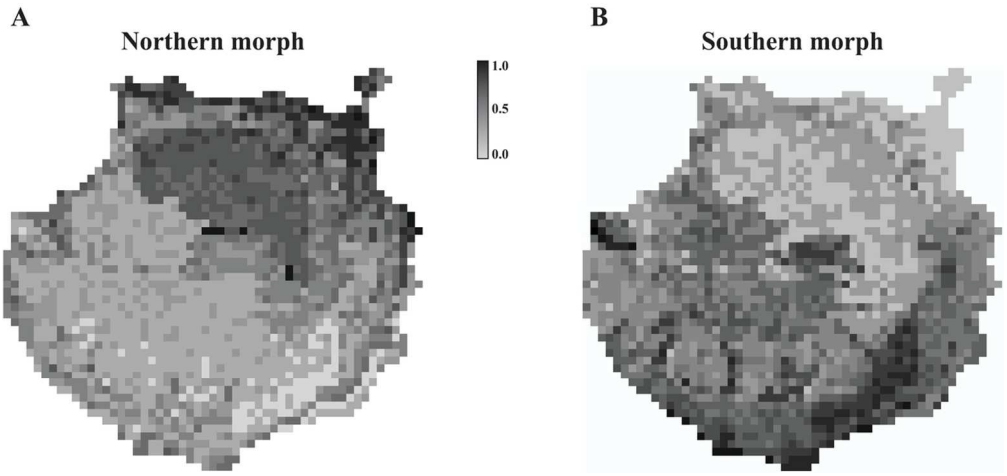
*BEAST population tree chronogram. Median posterior ages of nodes are provided, together with bars representing 95% HPDs. Scale bar provides times in millions of years.
184x92mm (300 x 300 DPI)



Genetic structure inferred from microsatellites using STRUCTURE. A) Individual assignment to clusters (K=2) based on B) ΔK (Evanno et al. 2005). C) Site compositions.
206x170mm (300 x 300 DPI)



Posterior densities for population migration (2NM) estimated using IMa2.
79x48mm (300 x 300 DPI)



Species distribution models for the northern and southern morphs. Higher values indicate higher predicted environmental suitability.
122x57mm (300 x 300 DPI)