

# Independent Homoploid Hybrid Speciation events in the Macaronesian endemic genus *Argyranthemum*

Running title: Homoploid Hybrid Speciation in *Argyranthemum*

Oliver W. White<sup>1,2</sup>, Alfredo Reyes-Betancort<sup>3</sup>, Mark A. Chapman<sup>2</sup> and Mark A. Carine<sup>1</sup>

<sup>1</sup> Algae, Fungi and Plants Division, Department of Life Sciences, The Natural History Museum, Cromwell Road, London, SW7 5BD, United Kingdom

<sup>2</sup> Biological Sciences, University of Southampton, Southampton, SO17 1BJ, United Kingdom.

<sup>3</sup> Jardín de Aclimatación de La Oratava (ICIA), C/ Retama 2 38400 Puerto de la Cruz, Tenerife, Spain

Mark A. Chapman (m.chapman@soton.ac.uk) and Mark A. Carine (m.carine@nhm.ac.uk) should be considered joint senior authors.

Corresponding author is Mark Chapman

m.chapman@soton.ac.uk

The authors have no conflict of interest.

## Abstract

Well-characterised examples of homoploid hybrid speciation (HHS) are rare in nature yet they offer the potential to study a number of evolutionary processes. In this study we investigate putative homoploid hybrid species in the genus *Argyranthemum* (Asteraceae), a group of plants endemic to the Macaronesian archipelagos of the North Atlantic Ocean. We specifically address a number of knowledge gaps surrounding the origin(s) of *A. sundingii* and *A. lemsii*, which are thought to be derived from the same parental cross. Comparisons of leaf morphology suggest that *A. sundingii* and *A. lemsii* are distinct from their parental progenitors and distinguishable from each other based on leaf area. Ecological niche modelling (ENM) demonstrated that the homoploid hybrid species occupy novel habitats that are intermediate relative to the parental species. Nuclear SSRs and SNP data indicate that the homoploid hybrid species are distinct from the parental taxa, whilst population level sampling of chloroplast SSRs and Approximate Bayesian Computation show that *A. sundingii* and *A. lemsii* are independently derived from the same parental cross. As such, *Argyranthemum* represents an example of independent homoploid hybrid speciation events

with evidence of divergence in leaf morphology and adaptation to novel intermediate habitats. On oceanic islands, which are often typified by steep ecological gradients and inhabited by recently derived species with weak reproductive barriers, multiple HHS events from the same parental cross are not only possible but are likely to have played a more important role in oceanic island radiations than we currently think.

## **Keywords**

Hybridisation, speciation, *Argyranthemum*, Macaronesia

## **Introduction**

Homoploid hybrid speciation (HHS), the origin of a new species by hybridisation without a change in chromosome number, is generally considered to be exceptionally rare in angiosperm evolution (Hegarty & Hiscock 2005; Soltis & Soltis 2009; Schumer *et al.* 2014; Goulet *et al.* 2017) with a recent review identifying 28 putative examples of HHS in flowering plants (Kadereit 2015). This is in sharp contrast to polyploid hybrid speciation, involving a doubling of chromosome number, which is thought to account for approximately 15 % of angiosperm speciation events (Wood *et al.* 2009). The relative rarity of HHS has largely been attributed to two main factors. Firstly, early generation hybrids frequently have reduced fitness caused by a loss of fertility and or viability, characteristics generally referred to as hybrid incompatibilities (Rieseberg & Willis 2007). Secondly, since early generation homoploid hybrids share the same chromosome number as their parental progenitors they are particularly susceptible to introgression and assimilation (Rieseberg & Willis 2007; Schumer *et al.* 2014). As such, the lack of habitats not occupied by one or both of the parental species may also limit the existence of homoploid hybrid species.

The accurate identification of a homoploid hybrid species is experimentally challenging, often requiring a combination of morphological and molecular analyses (Mallet *et al.* 2007). In addition, there is a lack of agreement on the defining criteria of a homoploid hybrid species (Schumer *et al.* 2014, 2018; Feliner Nieto *et al.* 2017). Nevertheless, given that occasional hybridisation between closely related species is frequent in plants (Mallet 2005) and that homoploid hybrids are difficult to identify, opinion has recently shifted to speculate that HHS might play a more prominent role in plant evolution than previously thought (Mallet 2005; Mavárez & Linares 2008; Nolte & Tautz 2010).

Examples of HHS provide natural experiments for the study of fundamental evolutionary processes such as reproductive isolation, hybridisation, adaptation, ecological speciation and genome evolution. An excellent example of this can be found in North American sunflowers *Helianthus* L., where hybridisation between the same parental combination of *H. annuus* L. and *H. petiolaris* Nutt. has resulted in the formation of three distinct homoploid hybrid species (*H. anomalus* S.F.Blake, *H. paradoxus* Heiser, and *H. deserticola* Heiser; Rieseberg *et al.* 2003). The fact that three species in *Helianthus* are all derived from the same parental cross has allowed for comparative investigations of genomic composition (Rieseberg *et al.* 2003), genome size (Baack *et al.* 2005) and ecological selection (Donovan *et al.* 2010).

*Argyranthemum* Webb, a genus of 24 species endemic to the Macaronesian archipelagos of Madeira, the Salvage Islands and the Canary Islands provides another of the few documented examples of HHS in plants. *Argyranthemum* is thought to have diverged relatively recently, with estimates ranging from 2.5-3.0 mya based on isozyme differentiation (Francisco-Ortega *et al.* 1995) to 0.26-2.1 mya based on ITS sequences (Francisco-ortega *et al.* 1997). All members of the genus are diploid ( $2n = 2x = 18$ ; Gonzalez *et al.* 1997) and outcrossing because of limited self-compatibility (Francisco-Ortega *et al.* 1997a) and wind dispersed although not specialised for this purpose. There is little intrinsic reproductive isolation between taxa, with isolation mainly due to occupancy of different islands or ecological habitats (Francisco-Ortega *et al.* 1997a). The breakdown of these external barriers often results in the formation of hybrid swarms (Borgen 1976; Brochmann 1984) and crosses can easily be made under common garden conditions (Humphries 1973; Brochmann *et al.* 2000). The presence of polyphyletic taxa in a chloroplast restriction site phylogeny has been interpreted as evidence that hybridisation was prevalent in the diversification of *Argyranthemum* (Francisco-Ortega *et al.* 1996).

Within *Argyranthemum*, *A. sundingii* Borgen and *A. lemsii* Humphries are thought to have originated by homoploid hybridisation (Brochmann *et al.* 2000; Fjellheim *et al.* 2009; Supplementary Plate 1). Both species are found in the Anaga peninsula of Tenerife and were discovered and described independently in southern and north-eastern valleys respectively (Humphries 1976; Borgen 1980; Figure 1). Morphological comparisons and artificial crossing experiments have demonstrated that these two species are both likely the result of hybridisation between *A. frutescens* (L.) Sch.Bip. and *A. broussonetii*(Pers.) Humphries (Brochmann *et al.* 2000; Supplementary Plate 1). Different *A.*

*frutescens* subspecies have been implicated in the parentage of the putative homoploid hybrid species based on their geographical distributions: subsp. *frutescens* as a parent of *A. sundingii* and subsp. *succulentum* Humphries as a parent of *A. lemsii*. While *A. sundingii* and *A. lemsii* provide a potential case study of HHS, a number of significant knowledge gaps concerning the delimitation of species and their origin(s) remain to be addressed.

The first concerns the morphological distinction of the homoploid hybrid species. The parental species differ conspicuously in their morphology. *Argyranthemum broussonetii* forms a large shrub up to 1.2 m in height with large bipinnatifid leaves, large capitula and dark brown ray cypselae (dry single-seeded fruits). In contrast, *A. frutescens* is a smaller branching shrub 20 – 80 cm in height, with small pinnatisect to bipinnatisect leaves, small capitula and cypselae that are pale brown in colour. *Argyranthemum frutescens* subsp. *succulentum* is a cushion-like shrub with pale green succulent leaves found along the North coast of the Anaga peninsula under the influence of the cooling trade winds whereas subsp. *frutescens* is more upright and branched with less succulent and darker green leaves distributed on the South coast of the Anaga peninsula. *Argyranthemum sundingii* and *A. lemsii* are morphologically intermediate between *A. broussonetii* and *A. frutescens* supporting their hybrid origin (Brochmann *et al.* 2000). They were found also to be morphologically distinct from their parental progenitors and exhibit high fertility, characteristics which are inconsistent with a typical hybrid swarm (Brochmann *et al.* 2000). Despite their broadly similar appearance, *A. sundingii* and *A. lemsii* are recognised as distinct in recent floras (Bramwell & Bramwell 2001; Schönfelder & Schönfelder 2012) and are both of conservation concern (Moreno 2008). However, an analysis of leaf morphology by Brochmann *et al.* (2000) suggested that *A. sundingii* and *A. lemsii* should be treated as a single species. This study was based on material collected only from the type localities of *A. sundingii* and *A. lemsii*; populations found more recently were not examined and the distinctiveness of taxa considering their full geographical extent remains to be robustly tested.

Although ecological selection has been implicated as an important factor in the origin of the homoploid hybrid species in *Argyranthemum* (Brochmann *et al.* 2000; Fjellheim *et al.* 2009), the evidence is largely anecdotal and based solely on species distributions. The parental species occupy the altitudinal extremes of the Anaga peninsula with *A. broussonetii* restricted to the higher altitudes between 500 – 1000 m in laurel forest clearings whereas *A. frutescens* typically occurs in coastal

xerophytic habitats below 100 m. Both *A. sundingii* and *A. lemsii* are found at intermediate altitudes of the strong humidity gradient that exists in the peninsula in habitats created by deforestation of lower parts of the laurel forest (Brochmann *et al.* 2000; Fjellheim *et al.* 2009). Models of HHS suggest that a hybrid species can arise from chromosomal rearrangements, ecological and/or geographical isolation (Rieseberg 1997; Buerkle *et al.* 2000; James & Abbott 2005). Although these models are not mutually exclusive, ecological isolation is increasingly seen as fundamental because simulations indicate HHS is unlikely without ecological divergence (Buerkle *et al.* 2000). Of the 28 instances of HHS reported by Kadereit (2015), 21 were associated with habitat or geographical displacement. However, as of yet there has been no explicit ecological analysis of the putative homoploid hybrid species in *Argyranthemum*.

The evidence to support the hypothesis that *A. sundingii* and *A. lemsii* originated independently through hybridisation events between *A. broussonetii* and *A. frutescens* is based on an earlier chloroplast (cp) restriction site phylogeny (Francisco-Ortega *et al.* 1996) which showed that the single individual sampled for each of *A. sundingii* and *A. lemsii* had distinct cp types: the *A. sundingii* sample shared a haplotype with *A. broussonetii* whereas the *A. lemsii* sample shared the *A. frutescens* haplotype, implying that *A. broussonetii* and *A. frutescens* are the maternal parents of *A. sundingii* and *A. lemsii* respectively. However, given that sampling was limited to one accession for each putative homoploid hybrid species, these results need to be interpreted with extreme caution in a genus with weak intrinsic reproductive barriers. Differences in Amplified Fragment Length Polymorphisms (AFLPs) (Fjellheim *et al.* 2009) have been identified between *A. sundingii* and *A. lemsii*, but were unable to differentiate between a single HHS event and divergence versus independent HHS events. Furthermore, the hypothesis that two subspecies of *A. frutescens* are involved in the parentage of the homoploid hybrid species was based on the distributions of the *A. frutescens* subspecies relative to *A. sundingii* and *A. lemsii* (Figure 1). More rigorous sampling is necessary to test the independent origins hypothesis and establish the contribution of the two *A. frutescens* subspecies to the homoploid hybrid species in the Anaga peninsula.

Fjellheim *et al.* (2009) identified a population in the Anaga peninsula of unknown identity that they referred to as '*A. cf. lemsii*'. This population was located in Barranco de Igueste almost equidistant between the then known populations of *A. sundingii* and *A. lemsii*. Analysis of AFLPs found that this population was genetically intermediate between *A. broussonetii* and *A. frutescens* suggesting a

hybrid origin. However, based on their AFLP analysis this population also appeared to be influenced by introgression suggesting it could be a hybrid swarm (Fjellheim *et al.* 2009). Further populations in this area have subsequently been discovered (see Methods; Sampling) and whether plants in this area represent a hybrid swarm, a contact zone between *A. sundingii* and *A. lemsii* or an independently derived homoploid hybrid species is unclear.

This study addresses a number of outstanding questions concerning the origin of the homoploid hybrid species in *Argyranthemum*. We first test if *A. sundingii* and *A. lemsii* are morphologically distinct using leaf characters from plants grown under controlled glasshouse conditions. We use material collected from previously recorded as well as recently discovered populations that are yet to be assessed. The hypothesis that *A. sundingii* and *A. lemsii* are adapted to novel ecological conditions along a strong humidity gradient that exists in the Anaga peninsula (Fjellheim *et al.* 2009) is also investigated using ecological niche modelling (ENM). Population level sampling of nuclear Simple Sequence Repeat markers (nSSRs) is employed to assess the genetic distinctiveness of the two homoploid hybrid species and chloroplast SSR markers (cpSSRs) are used to test the hypothesis of Brochmann *et al.* (2000) that *A. sundingii* and *A. lemsii* exhibit distinct maternal parentages. The origin(s) of the populations of uncertain status near Barranco de Iguete are clarified using a combination of evidence from morphology, ecology and population genetics. Finally, this study leverages a dataset of nuclear single nucleotide polymorphisms (nSNPs) obtained through Genotyping-By-Sequencing (GBS) and Approximate Bayesian Computation (ABC) to explicitly test whether the homoploid hybrid species originated by independent hybridisation events.

## Materials and Methods

### *Sampling*

From July to August 2015 multiple populations of *A. broussonetii*, *A. frutescens*, *A. lemsii* and *A. sundingii* in the Anaga peninsula of Tenerife were sampled for leaf, seed and voucher material (Table 1; Figure 1). Plant material was collected under a permit from the Cabildo de Tenerife, number 18297 and Gobierno de Canarias permit number 2015/939. We sampled populations investigated in earlier studies (Brochmann *et al.* 2000; Fjellheim *et al.* 2009) as well as populations recently discovered that had not been sampled previously. Population K that was sampled by Fjellheim *et al.* (2009) from Barranco de Iguete and population L from Lomo de las Casillas appeared to be morphologically intermediate between *A. broussonetii* and *A. frutescens* and were not clearly

referable to either *A. sundingii* or *A. lemsii*. These populations are therefore referred to hereafter as *A. frutescens* × *A. broussonetii*.

Wild collected cypselae from 70 plants representing 20 populations were used to grow 123 replicates (up to two replicates per parent) under common glasshouse conditions at the University of Southampton for leaf morphological analysis (Table 1; Supplementary Table 1). Georeferenced samples from a total of 594 collections were used as distribution data for each taxon in our ENM analysis. Representative voucher specimens are deposited at the Natural History Museum, London (BM; Supplementary Table 2). Silica dried leaf material from a total of 198 field-collected individuals across 20 populations were sampled for nuclear and chloroplast Simple Sequence Repeat (SSR) markers (Table 1). Nuclear SNP data were obtained for a total of 28 individuals using Genotyping-By-Sequencing (GBS), eight of each parental species and six of each homoploid hybrid species (Table 1). Populations referred to as *A. frutescens* × *A. broussonetii* (i.e. populations K and L) were not sampled for GBS as they were admixed and not clearly identifiable as either *A. sundingii* or *A. lemsii* (see results).

#### *Morphological analysis*

Plants sown from wild collected seed (Table 1) were grown for six months under common glasshouse conditions before sampling a leaf approximately halfway up the stem. No plants had started to flower, therefore minimising differences due to life stage of the plants. Whole leaves were imaged with a Canon EOS 600D digital SLR camera on a tripod with a scale bar next to each leaf. ImageJ v.1.5 (Schneider *et al.* 2012) was used to measure leaf area, perimeter, length, and width. Leaf length and width proved effective in earlier morphological comparisons (Brochmann 1987; Brochmann *et al.* 2000) while leaf area and perimeter have not been used in previous studies of the homoploid hybrid species in *Argyranthemum*. Measurements related to primary lobes have been used previously (Brochmann 1987) but were excluded to avoid the inclusion of over-correlated variables. To examine morphological differences between taxa, we used linear models and generalised linear models for normal and non-normally distributed data respectively, where each morphological character was treated as a response variable and the taxa as fixed factors. Normality was inferred using Q-Q norm plots, histograms and a Shapiro-Wilk test. Significant differences between taxa were identified using a *post hoc* Tukey's honest significant difference (HSD) test with false discovery rate (FDR) correction (Benjamini & Hochberg 1995) for multiple comparisons. To

visualise the differences between taxa, boxplots and a Principal Components Analysis (PCA) were generated. All analyses were performed using R (R Core Team 2018).

#### *Ecological niche modelling*

Climate data with a spatial resolution of 50 m were downloaded from the Climate Impact Project for Tenerife <http://climainpacto.eu/efectos/catalogos-climaticos/5-tenerife/> based on scenario B (current 1981-2010; 15 [PubMed](#) temperature and 6 precipitation variables). Altitude data for Tenerife were downloaded from <http://www.eorc.jaxa.jp/ALOS/en/aw3d30/> at a resolution of 30 m. ArcGIS was used to project the altitude data at the same resolution as the climate data and create an aspect and slope model. As aspect is a circular variable, it was treated as a categorical variable with nine values (flat, North, Northeast, East, Southeast, South, Southwest, West and Northwest). Georeferenced samples were filtered such that only a single accession per species occupied each 50 m<sup>2</sup> pixel, before extracting values from the climate, altitude, aspect and slope data sets. To avoid the inclusion of over-correlated variables, we selected mean annual temperature and mean annual precipitation and then selected further variables that were not over-correlated with these or each other (Pearson correlation < 0.7).

ENMs were generated for each taxon using the maximum entropy method of MAXENT version 3.4.1 (Phillips *et al.* 2006, 2018). We employed MAXENT with a random test percentage of 25 %, regularization multiplier of one, 10,000 background points, 10 replicates, a maximum of 5000 iterations and a convergence threshold of  $10^{-5}$ . ENMTools version 1.4.4 (Warren *et al.* 2010) was then used to calculate niche overlap between each species based on the Schoener's *D* (Schoener 1968) and Warren's *I* statistics (Warren *et al.* 2008), where a value of 0 denotes no overlap and 1 completely overlapping. To test whether the ENMs of two species are identical as expected under the null hypothesis we used the niche equivalency test initially proposed by Warren *et al.* (2008) in ENMtools. This test compares the observed scores of niche overlap statistics *D* and *I* with their null distribution generated with 100 pseudoreplicates (see Warren *et al.* (2008) for details). The null hypothesis is rejected when the observed value for the niche overlap statistics are significantly lower than the values expected from the pseudoreplicated data sets.

In order to assess the degree of niche overlap, we also employed PCA-env, initially implemented by Broennimann *et al.* (2012). In this approach, the multidimensional environmental space of the selected variables is first translated into two dimensions by means of principal components analysis



(PCA). The PCA is divided into a grid with a resolution of 100, where each cell corresponds to a unique set of environmental conditions. Species occurrences are projected onto this grid and a smoothed density of occurrence for each species was estimated using a kernel density function. Only continuous variables could be used for this analysis so aspect was excluded. This method is advantageous in that it accounts for spatial resolution biases, makes optimal use of both geographical and environmental spaces and corrects observed occurrence densities in light of the availability of environmental space (Broennimann *et al.* 2012). As above, niche overlap was quantified using the *D* and *I* statistics and the niche equivalency test was employed to test whether the environmental niche space of two species are identical using 100 pseudoreplicates. All analyses for PCA-env were performed in R using the package *ecospat* (Di Cola *et al.* 2017; R Core Team 2018).

#### *DNA extraction*

DNA was extracted from silica-dried leaf material using a modified CTAB based method (Doyle & Doyle 1987). Briefly, silica dried leaf material was homogenised and extracted with a CTAB-based buffer. Lipids and other debris were removed by mixing with 24:1 chloroform: isoamyl alcohol and centrifugation. DNA was precipitated using isopropanol, pelleted by centrifugation and washed with ethanol. DNA was re-suspended in TE buffer and treated with RNase at 37°C for 30 minutes.

#### *Polymerase Chain Reaction (PCR) of SSRs*

A combination of eight nuclear and four chloroplast SSR markers were used for population genetic analyses (Supplementary Table 3). The nuclear SSRs (nSSRs) were developed by White *et al.* (2016) and are known to be variable between the parental species *A. broussonetii* and *A. frutescens*. The chloroplast SSR (cpSSR) markers were developed by Bryan *et al.* (1999) using *Nicotiana tabacum* L. but are universal for flowering plants. The forward primer for each locus was designed with the sequence M13(-29) sequence (CACGACGTTGTAAAACGAC) appended to the 5' end such that a third fluorescently-labelled M13(-29) primer (either FAM, NED or TET) could be incorporated in the PCR (Schuelke 2000). PCR reactions were carried out as described in White *et al.* (2016).

#### *Population genetic analyses of nuclear SSRs*

Samples were excluded from nSSR analyses if there were missing data in more than three of the eight markers scored. Summary statistics were calculated in GenAlEx version 6.502 (Peakall & Smouse 2012) and Principal Component Analysis (PCA) was performed using the R package

adeget (Jombart & Ahmed 2011; R Core Team 2018). STRUCTURE (Pritchard *et al.* 2000) was used to identify clustering, where  $K$  (the number of populations) was tested from 1 to 10. For each value of  $K$ , 10 runs of 2,000,000 replicates after a 500,000 burn-in were implemented. The most likely  $K$  was then determined using the delta  $K$  method of Evanno *et al.* (2005) with STRUCTURE HARVESTER (Earl & vonHoldt 2012). For each accession, the proportion of membership to each of the clusters was determined using CLUMPP (Jakobsson & Rosenberg 2007) using the full search method and the structure plots were drawn using the R package ggplot2 (Wickham 2009; R Core Team 2018).

#### *Haplotype analysis of chloroplast SSRs*

Samples were excluded from cpSSR analyses if there were any missing data and summary statistics were calculated in GenAEx. Allele scores for cpSSRs were used to identify haplotypes and construct a median-joining haplotype network in Network 5.0 (Bandelt *et al.* 1999). For the purposes of identifying the parentage of the homoploid hybrid species, haplotypes were grouped according to their presence in the parental species. These groups included haplotypes that were only found in *A. brossonnetii*, only found in *A. frutescens*, identified in both parental species and those found in neither (i.e. restricted to one or other of the homoploid hybrid species or *A. brossonnetii* × *A. frutescens*).

#### *Processing of GBS SNP data*

DNA samples were sent to the Genomic Diversity Facility at Cornell University for GBS where they were digested using *EcoT22I* and single-end 100 bp reads were sequenced using an Illumina HiSeq. Ipyrad (version 0.7.15) was used to process the raw sequences, cluster loci and call SNPs (Eaton & Ree 2013). The raw fastq data were de-multiplexed based on barcode sequences with no barcode mismatches allowed. Bases with low quality scores (<33) were converted to N and reads with more than five Ns were discarded. The strict setting (2) in ipyrad was used to filter adapter sequences as recommended for GBS data. Prior to clustering, samples with less than 0.5 M filtered reads were removed to avoid the inclusion of samples with large amounts of missing data. Filtered reads were assembled with three different levels of clustering threshold (80 %, 85 % and 90 %) using the *de novo*-reference assembly method. In this approach, reads that map to a reference are removed from downstream processes. Therefore we were able to use this to identify and remove clusters that mapped to the chloroplast genomes of *Arabidopsis thaliana* (L.) Heynh. (Genbank accession number

NC\_000932.1), *Helianthus annuus* (NC\_007977.1) and *Chrysanthemum indicum* L. (NC\_020320.1) and the mitochondrial genomes of *A. thaliana* (NC\_001284.2) and *H. annuus* (KF815390.1).

Consensus allele sequences within individuals were estimated with a minimum depth of six reads required for base calling. Consensus sequences were then clustered across samples resulting in assembled loci. For the three levels of clustering employed, loci were filtered at two levels of missing data (minimum samples per locus of 10 and 13 individuals) resulting in a total of six assemblies. Loci with a shared heterozygous site in 20 % or more of samples were also removed from the final assembly as these may result from the clustering of paralogs.

#### *Genome-wide SNP analysis*

Assembled loci were further filtered using vcfTools (Danecek *et al.* 2011) for a minimum minor allele frequency of 0.05 while selecting the first SNP from each locus to reduce the likelihood that SNPs were linked. PCA and STRUCTURE were used to analyse the datasets generated using each assembly method. PCA was employed using PLINK v.1.9 (Purcell *et al.* 2007). STRUCTURE (Pritchard *et al.* 2000) was implemented as above for the SSR markers except for a shorter run length of 50,000 replicates after a 20,000 burn-in over five iterations and using the Greedy search algorithm in CLUMPP (Jakobsson & Rosenberg 2007). Plots were drawn using the R package ggplot2 (Wickham 2009; R Core Team 2018).

#### *Testing independent hybrid origins with ABC*

We used Approximate Bayesian Computation (ABC) to compare different evolutionary scenarios for the origin of *A. sundingii* and *A. lemsii* (Brochmann *et al.* 2000; Fjellheim *et al.* 2009) based on our population genetic results using the software package DIYABC (Cornuet *et al.* 2008). A total of nine scenarios were tested (Figure 6). For scenarios one and two, *A. sundingii* and *A. lemsii* originate independently from crosses between *A. broussonetii* and different subspecies of *A. frutescens*. For scenarios three and four *A. sundingii* and *A. lemsii* originate independently but from the same parentage. In scenario five, *A. sundingii* and *A. lemsii* originate independently from hybridisation events between *A. broussonetii* and the common ancestor of the two *A. frutescens* subspecies. Scenarios six, seven and eight involve a single origin based on a hybridisation event between *A. broussonetii* and either *A. frutescens* subsp. *frutescens*, subsp. *succulentum* or the common ancestor of the two subspecies. Finally, in scenario nine *A. sundingii* and *A. lemsii* diverge by cladogenesis from *A. broussonetii*.

A vcf file of unlinked SNPs was converted to DIYABC format using the python script vcf2DIYABC.py available from <https://github.com/loire/vcf2DIYABC.py>. To meet the requirements of DIYABC SNPs were removed if they were not present in at least one individual from each of the five taxa (*A. broussonetii*, *A. sundingii*, *A. lemsii*, *A. frutescens* subsp. *frutescens* and subsp. *succulentum*).

For each prior a uniform distribution with a large interval was chosen due to lack of knowledge concerning population sizes, divergence times and admixture rate. Specifically, the interval for population sizes and divergence times were set to  $10^2 - 10^7$  and the interval for admixture rates was 0.001 – 0.999. A total of 29 summary statistics were used to compare the observed and simulated data comparisons, including the means of genic diversity, pairwise sample  $F_{st}$  and Nei's distance as well as admixture summary statistics (Choisy *et al.* 2004). A total of  $10^6$  simulations were performed for each scenario and the posterior probability of each scenario was computed by performing a logistic regression on the 1% of simulated data closest to the observed dataset (Cornuet *et al.* 2008, 2010).

Confidence in scenario choice was assessed by simulating 1000 test datasets (pseudo-observed datasets) of the selected scenario, drawing parameter values from the prior distribution. The posterior probabilities of each scenario were estimated for each simulated dataset by performing a logistic regression as above. Type I and type II errors were then evaluated by measuring the fraction of datasets simulated under the best scenario that were assigned to other scenarios and the fraction of datasets simulated under other scenarios that were assigned to the best scenario respectively (Cornuet *et al.* 2010).

Goodness-of-fit of the selected scenario was assessed using the model checking function of DIYABC, which evaluates the ability of a given scenario to produce datasets similar to the observed dataset (Cornuet *et al.* 2010; Barrès *et al.* 2012). For the selected scenario, 1000 datasets simulated using posterior distribution of parameter values were compared with the observed dataset using different summary statistics than those used for model selection and parameter estimation to avoid overweighting the fit of the scenario (Cornuet *et al.* 2010; Capblancq *et al.* 2015). PCA was implemented to visually assess the position of the observed dataset with regards to the simulated datasets. A *P*-value was estimated for each summary statistic by ranking the observed value among the values obtained with simulated datasets.

The posterior distribution of the parameters for the selected model were also estimated using a local linear regression on the 1 % of simulations closest to the observed dataset and applying a logit transformation to the parameter values (Cornuet *et al.* 2010). These distributions estimate the most probable value and the width of the distribution for each historical parameter (Capblancq *et al.* 2015).

## Results

### *Morphological analyses*

Of the four morphological characters used only leaf area showed a non-normal distribution (Shapiro test:  $W = 0.9001$ ,  $P < 0.001$ ). A GLM of leaf area identified significant differences between taxa (GLM with a quasi link log function:  $F_{5,117} = 77.416$ ,  $P < 0.001$ ). Linear models identified significant differences between taxa for perimeter ( $F_{5,117} = 19.756$ ,  $P < 0.001$ ), leaf length ( $F_{5,117} = 39.934$ ,  $P < 0.001$ ) and leaf width ( $F_{5,117} = 22.794$ ,  $P < 0.001$ ). Summary plots for each model are in Supplementary Figure 1A-D.

For each character *A. broussonetii* and *A. frutescens* were significantly different, while the two subspecies of *A. frutescens* were largely indistinguishable from each other (Figure 2A-D). Leaf area and leaf perimeter were effective in delimiting *A. sundingii* and *A. lemsii* from both parental species. For leaf length, *A. sundingii* and *A. lemsii* were not significantly different from *A.*

*broussonetii*. *Argyranthemum lemsii* was significantly different to both parents based on leaf width but *A. sundingii* showed no significant difference from *A. broussonetii* based on this character. Only leaf area identified a significant difference the homoploid hybrid species with the remaining characters showing no significant differences. The putative hybrid populations of *A. broussonetii* × *A. frutescens* were significantly different from both of the homoploid hybrid species and *A. broussonetii* for all characters but were not significantly different from either subspecies of *A. frutescens*.

*Argyranthemum sundingii* and *A. lemsii* are intermediate relative to their parental progenitors based on the PCA whereas the putative hybrid populations of *A. broussonetii* × *A. frutescens* showed greater similarity to *A. frutescens* than to either of the homoploid hybrid species (Figure 2E).

### *Ecological niche modelling*

A total of 185 georeferenced localities remained after filtering samples to include only one record per species per 50 m<sup>2</sup> pixel (Supplementary Table 4). A total of six variables were not correlated (Pearson correlation < 0.7) and were used in our analysis: mean annual temperature, mean annual precipitation, isothermality (diurnal thermal range divided by annual thermal range), daytime thermal range (difference between maximum annual average temperature and minimum annual average temperature), annual thermal range (difference between maximum temperature of the warmest month and minimum temperature of the coldest month) and slope.

Niche distributions for each species predicted using Maxent had high predictive ability based on mean area under the receiver operator curve (AUC) values for *A. broussonetii* (0.987), *A. frutescens* subsp. *frutescens* (0.984), *A. frutescens* subsp. *succulentum* (0.969), *A. sundingii* (0.950), *A. lemsii* (0.974) and *A. frutescens* × *A. frutescens* (0.949). Predicted distributions were also congruent with the observed distributions of the taxa (Figure 3). Mean annual temperature contributed most to the model predictions for *A. broussonetii* (37.5 %) and *A. frutescens* subsp. *frutescens* (55.4 %), and mean annual precipitation provided a greater contribution to the predictions of *A. lemsii* (31.6 %) and *A. broussonetii* × *A. frutescens* (36.1 %). Daytime thermal range contributed most to the model predictions of *A. frutescens* subsp. *succulentum* (80.2 %) and isothermality contributed most to the prediction of *A. sundingii* (33.5 %).

In our PCA-env analysis, the proportion of variation explained by the first and second axis was 43.91 % and 26.8 % respectively. The first axis was primarily associated with mean annual temperature and mean annual precipitation whereas the second axis was largely explained by the remaining four variables (Supplementary Figure 2).

Niche overlap statistics demonstrated that each taxon occupied a distinct niche (Table 2). The null hypothesis that the ENM of two species are identical was rejected for all comparisons ( $P < 0.05$ ) except the  $D$  statistic of our PCA-env comparison for *A. frutescens* subsp. *frutescens* vs. subsp. *succulentum* ( $P = 0.0990$ ). Niche equivalency test plots are shown in Supplementary Figure 2A-D.

#### *Population genetic analyses of nuclear SSRs*

A total of 186 individuals had sufficient data to be included in the nuclear SSR analysis. Consistent with their hypothesised hybrid origin, expected heterozygosity ( $H_e$ ) and unbiased expected heterozygosity ( $uH_e$ ) are higher in *A. sundingii* (0.72 and 0.73 respectively), *A. lemsii* (0.69 and 0.70)

and *A. broussonetii* × *A. frutescens* (0.70 and 0.73) compared with *A. broussonetii* (0.57 and 0.57), *A. frutescens* subsp. *frutescens* (0.61 and 0.62) and subsp. *succulentum* (0.36 and 0.37) (Supplementary Table 5A).

Principal Component Analysis (PCA) of the nSSRs demonstrated a genetic distinction between *A. broussonetii* and *A. frutescens* and between the two subspecies of *A. frutescens* (Figure 4A). The homoploid hybrid species and *A. broussonetii* × *A. frutescens* are intermediate to the parents and appear separated to some extent, but do not form distinct genetic clusters. Examination of more principal components did not reveal any further distinction between taxa (Supplementary Figure 4).

STRUCTURE analysis of the nuclear SSR markers identified the most likely number of *K* as 2, 3 and 5 in descending order based on the Evanno delta *K* method (Figure 4B-C). For *K* = 2, the parent species are separated into different clusters. The average membership of *A. broussonetii* to cluster one was 94.52 % and the average membership of *A. frutescens* subsp. *frutescens* and subsp. *succulentum* to cluster two was 88.84 % and 96.99 % respectively. *Argyranthemum sundingii* and *A. lemsii* had an average membership of 81.67 % and 89.30 % to cluster one, suggesting the hybrid species have a greater affinity to *A. broussonetii*. Populations of *A. broussonetii* × *A. frutescens* had an average membership of 41.62 % and 58.38 % to clusters one and two respectively, although this was clearly variable, ranging from 1.60 % to 95.40 % for cluster one and from 4.60 % to 98.40 % for cluster two.

*K* values of 3 and 5 had similar levels of support and are both biologically informative. For *K* = 3, the clusters subdivide *A. broussonetii*, *A. frutescens* and the homoploid hybrid species. The average membership of *A. broussonetii* to cluster one was 91.58 % and the average membership of *A. frutescens* subsp. *frutescens* and subsp. *succulentum* to cluster two was 84.24 % and 94.27 % respectively. The average memberships of *A. sundingii* and *A. lemsii* to cluster three were 83.95 % and 81.90 %, respectively. Populations of *A. broussonetii* × *A. frutescens* share 66.15 % membership with cluster three, with 8.48 % and 25.39 % belonging to clusters one and two respectively.

For *K* = 5, the five named taxa occupy different clusters. A distinction between the two homoploid hybrid species is identified with *A. sundingii* having an average membership of 75.76 % and *A. lemsii* an average membership of 80.10 % to separate clusters. The two subspecies of *A. frutescens* are also delimited, with subsp. *frutescens* forming a cluster with an average membership of 84.11 % and subsp. *succulentum* an average of 90.43 %. Populations of *A. broussonetii* × *A.*

*frutescens* are admixed, showing greatest affinity to *A. sundingii* (40.52 %), *A. frutescens* subsp. *frutescens* (average = 27.36 %) and *A. lemsii* (26.01 %).

#### *Haplotype analysis of chloroplast SSRs*

After removing samples with missing data, 164 individuals remained and 16 haplotypes were identified (Table 3; Supplementary Figure 4). Ten of these haplotypes were specific to one of parental species with four found in *A. broussonetii* (40 individuals) and six specific for *A. frutescens* (27 individuals). Of the remaining six haplotypes two were shared between the parental species (12 individuals) and four were not found in either of the parental species (five individuals). Of the 37 *A. sundingii* individuals sampled, 24 shared a haplotype with *A. broussonetii*, four with *A. frutescens* and the remaining nine possessed haplotypes that were either shared between the parental species or not present in either. For *A. lemsii*, 29 of the 37 individuals sampled had a haplotype shared with *A. frutescens*, two shared a haplotype with *A. broussonetii* and six had haplotypes that were shared between each of the parental species or found in neither. Of the eleven individuals sampled from *A. broussonetii* × *A. frutescens*, 9 possessed a haplotype shared with *A. frutescens*, one individual had a haplotype shared between both parents and one individual had a haplotype not found in the plants of either parent sampled.

#### *Processing of GBS SNP data*

Genotyping-By-Sequencing resulted in an average of 1.91 M reads across all samples (range 0.23 to 5.24 M). Filtering out of low quality reads removed 7.59 - 8.52 % of the total reads resulting in an average of 1.77 M across all samples (range 0.22 to 4.79 M) (Supplementary Figure 5). Two samples (one each of *A. frutescens* subsp. *succulentum* [frs242] and *A. sundingii* [sun288]) with less than 0.5 M reads were removed. An average of 1705 reads per sample mapped to either chloroplast or mitochondrial references and were removed. Increasing the clustering threshold through 80 %, 85 % and 90 % increased the resulting number of clusters with an average of 59,494, 67,313 and 87,312 clusters per individual for each threshold respectively (Supplementary Table 6). This was reduced to 18,735, 20,818 and 25,621 clusters, respectively, that passed the minimum depth requirement. Increasing the clustering threshold reduced heterozygosity estimates from 0.0394 to 0.0284 and error estimates from 0.0164 to 0.0130. The average number of loci consensus reads per sample increased to 16,194, 18,586 and 23,988 for each clustering threshold. Two levels were used for the minimum number of samples required to process a locus. Increasing the minimum number of



samples required to process a locus from 10 to 13 reduced the average number of loci across samples included in the assembly as well as the number of SNPs (Supplementary Table 6).

#### *Nuclear SNP cluster analysis*

For each assembly generated in this study, PCA and STRUCTURE showed a similar grouping of individuals and the Evanno delta  $K$  plot as part of the STRUCTURE analysis always provided strongest support for  $K = 3$ , providing confidence that the results were robust to the read clustering approach used. Therefore, we selected the assembly based on a clustering threshold of 90 % with loci present in a minimum of 10 samples since it had the greatest number of SNPs (Supplementary Table 6). PCA and STRUCTURE plots for this dataset are presented in Figure 5 but results for all other assemblies are also presented in Supplementary Figures 6 and 7.

PCA identified clear differences between the parental species, as well as the two subspecies of *A. frutescens* (Figure 4A). The homoploid hybrid species form a cluster intermediate to the parents but closely allied to *A. broussonetii*. No clear distinction between *A. sundingii* and *A. lemsii* is apparent. The STRUCTURE analysis for the most strongly supported  $K = 3$  (Figure 4C) separated the parent species into distinct clusters with the two homoploid hybrid species sharing the third cluster (Figure 4B). Increasing the value of  $K$  does not increase the separation between the hybrid species. For  $K = 3$ , *A. broussonetii* has 100 % membership to cluster one while *A. frutescens* subsp. *frutescens* and subsp. *succulentum* have 99.96 % and 100 % membership to cluster three respectively. *Argyranthemum sundingii* and *A. lemsii* have 94.32 % and 82.82 % membership to cluster two respectively. Where  $K = 2$ , there is a clear distinction between the parental species with *A. broussonetii* having 100 % membership to cluster one while *A. frutescens* subsp. *frutescens* and subsp. *succulentum* have 99.97 % and 99.97 % membership to cluster two respectively. In keeping with the results of  $K = 2$  for the nSSRs, the homoploid hybrid species show greater similarity to *A. broussonetii*, with *A. sundingii* sharing an average membership of 78.00 % and *A. lemsii* a membership of 86.91 % with the *A. broussonetii* cluster.

#### *Approximate Bayesian Computation*

The ipyrad assembly used for our ABC analysis contained a total of 4736 putatively unlinked SNPs (Supplementary Table 6). After the removal of SNPs that were not found in at least one individual per species (a requirement of DIYABC), the total number of SNPs included in the ABC analysis was 2152. Scenario 1 (Figure 6) involving two independent hybridisation events was estimated to be the

most likely with a posterior probability of 0.5999 and a confidence interval of 0.4045 - 0.7953. In this scenario, *A. sundingii* and *A. lemsii* arose independently, and from different subspecies of *A. frutescens*. Scenario 3 which also involved two hybridisation events had some support with a posterior probability of 0.4001 and a confidence interval of 0.2047 - 0.5955. In this scenario the homoploid hybrid species again originated from independent hybridisation events but they shared the same parentage of *A. broussonetii* and *A. frutescens* subsp. *frutescens*.

Estimation of error rates provided confidence in scenario choice. Type I error rate (probability that datasets simulated under the selected scenario were assigned to other scenarios) was estimated to be 11.57 %, whereas, the type II error rate (the probability that datasets simulated under other scenarios were assigned to the best scenario) was estimated to be 1.41 %.

The goodness-of-fit of scenario 1 was demonstrated using simulations from posterior distributions and different summary statistics to those employed for scenario choice. PCA showed that the observed dataset was similar to these simulated datasets (Supplementary Figure 8). A *P*-value was also estimated for each summary statistic by ranking the observed value among the values obtained with simulated datasets. Of the 29 summary statistics used for model checking, 12 differed significantly from their simulation distribution ( $P < 0.05$ ; Supplementary Table 7). Despite being the most strongly supported model this suggests some discordance between the scenario-posterior combinations and the observed dataset.

Parameter estimates for scenario one (Figure 6) suggest that hybridisation events that gave rise to *A. sundingii* and *A. lemsii* occurred 2.22 (95% confidence interval 1.24 – 3.29) and 4.21 (CI 2.73 – 5.70) Mya respectively (based on one generation per year). The relative admixture rates suggests that the initial parental contributions were unequal for *A. sundingii*, with *A.*

*frutescens* subsp. *frutescens* contributing 74 % (parameter  $r_a$ ; Figure 6) and *A.*

*broussonetii* contributing 26 % ( $1 - r_a$ ). In contrast, the parental admixture for *A. lemsii* is more equal, with *A. frutescens* subsp. *succulentum* contributing 47 % ( $r_b$ ) and *A. broussonetii* 53 % ( $1 - r_b$ ). The mean divergence time for the two subspecies of *A. frutescens* is estimated to be 8.07 (CI  $6.01 \times 10^6 - 9.51 \times 10^6$ ) Mya and for *A. broussonetii* and an ancestral *A. frutescens* 5.88 (CI  $2.46 \times 10^6 - 9.42 \times 10^6$ ) Mya. Whilst the mean values are biologically implausible, the confidence intervals are large and overlapping suggesting that these events are indistinguishable.

## Discussion

In this study we employed a range of approaches to address a number of questions surrounding the status of putative homoploid hybrid species in *Argyranthemum*. Using collections from recently discovered populations we investigated if the homoploid hybrid species are morphologically distinct based on leaf characters. Fine resolution climatic data were used to test if the homoploid hybrid species occupy a novel ecological niche with respect to the parents. Using a combination of nuclear and chloroplast SSR markers and sampling at the population level we investigated the distinctiveness of the two homoploid hybrid species at a population genetic level while testing the hypothesis of Brochmann *et al.* (2000) that *A. sundingii* and *A. lemsii* exhibit distinct maternal parentages. We further sought to clarify the origin(s) of the populations of uncertain status near Barranco de Iguste using these data. Finally, we leveraged a nSNP dataset to test whether the homoploid hybrid species originated by independent hybridisation events using Approximate Bayesian Computation (ABC).

*The hybrid species are morphologically distinct from the parent species*

There is considerable variation in leaf morphology within *Argyranthemum* and leaf characters are important for species identification (Humphries 1973, 1976; Francisco-Ortega *et al.* 1997b). Each of the four leaf characters used in this study distinguished the parental species *A. broussonetii* and *A. frutescens* (Figure 2A-D). The two subspecies of *A. frutescens* were indistinguishable but this is not surprising since the main differences between these taxa are growth form and leaf succulence (Humphries 1976) which would not have been captured by our analysis. The homoploid hybrid species *A. sundingii* and *A. lemsii* were intermediate between and distinct from the parental species based on leaf area and perimeter in particular. *Argyranthemum sundingii* and *A. lemsii* did not exhibit the high degree of variability between the parental species that one would associate with ongoing hybridisation and backcrossing (Tovar-Sánchez & Oyama 2004; Worth *et al.* 2016) conforming to the hypothesis that *A. sundingii* and *A. lemsii* are stabilised homoploid hybrid species as opposed to hybrid swarms. This finding, with broader sampling, corroborates the earlier work of Brochmann *et al.* (2000) who found the extent of morphological variation in the homoploid hybrid species was comparable to the parental taxa, whereas hybrid swarms and synthetic F2 hybrids exhibited a much wider range in variation.

With the characters used in this analysis, only leaf area was able to distinguish *A. sundingii* and *A. lemsii*. This disagrees with their treatment as a single entity as proposed by Brochmann *et al.* (2000) and rather supports the current taxonomic treatment. While the homoploid hybrid species

are largely overlapping for most morphological traits examined here we were unable to include characters of the ligule (floral) and cypselae (seed) which are known to be informative for species identification in this genus (Humphries 1976; Brochmann 1987), since *A. broussonetii* did not flower under glasshouse conditions. A comprehensive study of morphology would be needed before making firm taxonomic conclusions regarding the status of the homoploid hybrid species in *Argyranthemum*. Our results do indicate that whilst the two homoploid hybrid species share broadly similar leaf traits, they are distinguishable and their distinctiveness might be greater when floral traits are considered.

#### *The hybrid species are ecologically intermediate*

Ecological isolation has been considered a contributing factor in most cases of HHS (Kadereit 2015). Few studies have explicitly investigated this hypothesis although Liu *et al.* (2014) demonstrated that the homoploid hybrid species *Ostryopsis intermedia* B.Tian & J.Q.Liu occupies a largely distinct niche from the parental progenitors and Capblancq *et al.* (2015) demonstrated significant overlap between the homoploid hybrid species *Coenonympha darwiniana* Staudinger and one of the progenitors.

The parental species are almost entirely non-overlapping in their ecological niches and it is clear from the Maxent and PCA-env analyses that the homoploid hybrid species occupy an altitude and ecological niche space intermediate between the parents. Despite being significantly different there is still some overlap in the niche space of the homoploid hybrid species and their parental progenitors, in particular between *A. broussonetii* and *A. lemsii* (Maxent:  $D = 0.369$ ,  $I = 0.632$ ; PCA-env:  $D = 0.343$ ,  $I = 0.417$ ), which is not surprising considering their geographical proximity (e.g. populations D and N; Figure 1). That such populations of a homoploid hybrid species can persist despite proximity to parental populations without assimilation, suggests either some degree of intrinsic reproductive isolation between the homoploid hybrid species and their parental progenitors or ecological isolation over very short distances. Brochmann *et al.* (2000) found a moderate reduction in the fertility of F1 hybrids between *A. broussonetii* and *A. frutescens* based on pollen staining. However, to our knowledge no such crosses have yet been made between the homoploid hybrid species and the parents. The fact that *A. sundingii* and *A. lemsii* are both adapted to intermediate habitats with respect to their parental progenitors might explain why each of the homoploid hybrid species is broadly similar leaf morphology. However, our analysis also identified a

significant difference in the ecological niches occupied by *A. sundingii* and *A. lemsii*, which might explain the differences in leaf area identified.

*A. sundingii* and *A. lemsii* are genetically distinct from the putative parents

Nuclear SSRs have been used successfully in studies investigating the origin of homoploid hybrid species (Sherman & Burke 2009; Liu *et al.* 2014) and the nSSRs identified by White *et al.* (2016) were effective in identifying genetic clusters for this study. The nSNP dataset (n = 4736 SNPs) obtained by GBS was also informative for investigating the genetic distinctiveness of the hybrid species. For a non-model system such as *Argyranthemum* where no reference genome is available, GBS followed by *de novo* assembly of reads using ipyrad was an effective protocol for the identification of nSNPs. A similar approach proved effective in a study of HHS in the butterfly genus *Coenonympha* which employed a reduced representation approach (Capblancq *et al.* 2015).

The PCA based on the nSSR data revealed a clear distinction between the parental progenitors and demonstrated the relative intermediacy of the homoploid hybrid species (Figure 4A). PCA of the nSNP dataset shows that the homoploid hybrid species are intermediate relative to their putative parental progenitors (Figure 4A) and the distinction is clearer compared with the PCA based on nSSRs.

In the STRUCTURE analysis of both nSSRs and nSNPs, the HHS formed a distinct cluster with respect to the parents where  $K = 3$  (Figure 4B; Figure 5B), supporting the distinctiveness of the homoploid hybrid species from the parental taxa. In the case of the nSNP data, these results are consistent across different assemblies suggesting that the results are robust (Supplementary Figure 6; Supplementary Figure 7).

Where  $K = 2$ , the STRUCTURE analyses of both datasets demonstrated that *A. sundingii* and *A. lemsii*, whilst being intermediate to the parents, showed greater similarity to *A. broussonetii*. This agrees with the expectation that a homoploid hybrid species is unlikely to inherit an equal proportion from each parent as backcrossing is expected to occur (Mallet *et al.* 2007).

In the nSSR analysis,  $K = 5$  showed evidence of differentiation between the two homoploid hybrid species which conforms to previous analyses based on AFLPs (Fjellheim *et al.* 2009) and evidence of chromosomal rearrangements between *A. sundingii* and *A. lemsii* based on genomic *in situ* hybridization (GISH; Borgen *et al.* 2003). This was not evident in the nSNPs analysis. This could

potentially be caused by missing data as a result of low sequence coverage commonly associated with reduced representation approaches such as GBS (Davey *et al.* 2011; Poland & Rife 2012).

Thus, both nSSRs and nSNPs support the genetic differentiation of the HHS from the putative parents with a greater contribution from *A. broussonetii* than *A. frutescens* to their genetic composition. Genetic differentiation between HHS is only supported by nSSR data, although this may be an artefact of the processing of nSNP data.

#### *Chloroplast haplotype patterns indicate independent origins*

Previous studies of HHS have also leveraged maternally inherited chloroplast markers as a method of inferring parentage and number of origins. For example, *Helianthus anomalus* (Schwarzbach & Rieseberg 2002), *Hippophae gyantsensis* Y.S.Lian (Wang *et al.* 2001) and *Pinus densata* Masters (Wang *et al.* 2001) are homoploid hybrid species that are thought to be derived from multiple hybridisation events based on the haplotype patterns identified using cp markers.

In *Argyranthemum*, 10 of the 16 haplotypes recovered were informative as to the parentage of the HHS because they occurred in one parental species and not the other. Of the remaining six haplotypes, two were shared between the parents and four were found in neither. Excluding uninformative haplotypes, 86 % (24/28) of *A. sundingii* samples have an *A. broussonetii* specific haplotype and 94 % (29/31) of *A. lemsii* samples have an *A. frutescens* haplotype. This pattern of haplotype distribution agrees with the restriction site phylogeny of Francisco-Ortega *et al.* (1996). A small number of samples of *A. sundingii* (four individuals) and *A. lemsii* (two individuals) have haplotypes that suggest the reverse parentage and this could be attributed to more recent gene flow. The most parsimonious explanation of the pattern of haplotypes identified agrees with the hypothesis that the two homoploid hybrid species currently recognised arose independently (Brochmann *et al.* 2000) and through crossing in different directions, although with some subsequent backcrossing.

#### *What are the plants from Igueste referred to as *A. broussonetii* × *A. frutescens*?*

Based on previous studies the origin(s) of populations near Barranco de Igueste (populations K and L in this study) remains unclear. In the field, plants at this site appeared morphologically intermediate between *A. broussonetii* and *A. frutescens* and considering the widespread distribution of the parental species in the Anaga peninsula, it is certainly plausible for hybridisation events to occur multiple times in different valleys (Brochmann *et al.* 2000). The comparison of leaf characters in this

study showed that populations of *A. broussonetii* × *A. frutescens* were similar to *A. frutescens*. Our ecological analysis found that *A. broussonetii* × *A. frutescens* occupied intermediate habitats similar to the homoploid hybrid species, which is consistent with their distribution at intermediate altitudes. The hybrid origin of *A. broussonetii* × *A. frutescens* was supported in our PCA of nSSRs which found these populations to be intermediate between the parental species. STRUCTURE analysis of nSSRs also showed that a number of individuals in these populations were admixed, with some individuals sharing a greater proportion of their genetic composition with *A. frutescens*. All but two of the individuals of *A. broussonetii* × *A. frutescens* possessed a haplotype shared with *A. frutescens*, with one having a haplotype shared between the parents and another a haplotype found in neither, suggesting that *A. frutescens* is the maternal parent of these crosses. Taken together, the data suggest that individuals sampled from population K and L are likely of hybrid origin and subsequently introgressed with *A. frutescens*.

*Approximate Bayesian Computation supports independent origins of A. sundingii and A. lemsii*

Approximate Bayesian Computation offers a flexible framework to investigate different evolutionary scenarios and is particularly useful for inferring historical hybridisation events (Bertorelle *et al.* 2010). This approach has been used effectively in previous studies of HHS, confirming a hybrid origin of *Ostryopsis intermedia* (Liu *et al.* 2014) and demonstrating that two subspecies of *Coenonympha darwiniana* originated by a single hybridisation event and subsequent divergence (Capblancq *et al.* 2015).

Our study supports two independent hybridisation events in the origin of the homoploid hybrid species with two different subspecies of *A. frutescens* involved. Specifically, a hybridisation event between *A. broussonetii* and *A. frutescens* subsp. *frutescens* gave rise to *A. sundingii* and a hybridisation event between *A. broussonetii* and *A. frutescens* subsp. *succulentum* resulted in *A. lemsii*. It is also noteworthy that the only other scenario which had had any level of support (scenario 3) also involved two independent hybridisation events, providing confidence in the hypothesis that the two homoploid hybrid species currently recognised arose independently.

While there is support for scenario 1 being the most likely, the model does not fit the observed data particularly well. This is apparent in the posterior values of parameters for scenario 1, which suggest that *A. frutescens* subsp. *frutescens* contributed a greater proportion of the genome to *A. sundingii* relative to *A. broussonetii*: the relative genomic contribution of *A. broussonetii* to the

hybrid species would be expected to be greater based on our STRUCTURE analyses (Figure 4B; Figure 5B). In addition, we would expect the split between *A. broussonetii* and *A. frutescens* to be ancestral to the subspecies of *A. frutescens*. Despite our tree topology showing this relationship the DIYABC analysis still provides the alternate order of events. The poor fit of the model could be attributed to the simplicity of our model design, which involved two discrete hybridisation events. In reality, it is likely that there would have been some degree of backcrossing particularly during the early stages of divergence. Indeed the identification of hybrid populations near Barannco de Igueste (populations K and L in the present study), which appear to be backcrossing with *A. frutescens*, and the unequal genomic contributions of the parental species to the homoploid hybrid species (Figure 4B; Figure 5B) indicates that occasional gene flow is likely. It is not possible to model gene flow using DIYABC and while there is software available that permits the simulation of gene flow over time (Excoffier & Foll 2011), it was not possible to use this due to the amount of missing data, an observation often associated with reduced representation datasets. Despite the simplicity and the limitations of the analysis, two independent hybridisation events appear to have resulted in the origin of the homoploid hybrid species in *Argyranthemum* consistent with other data presented. Future work should attempt to incorporate more complex models that include gene flow as opposed to discrete admixture events, which might allow for more accurate identification of model parameters.

### *Conclusions*

From a morphological perspective, *A. sundingii* and *A. lemsii* are distinct from their putative parents. Although both are similar in appearance, the homoploid hybrid species can be distinguished from each other based on leaf area. Each of the homoploid hybrid species occupy novel ecological niches with respect to their parents and each other. However, given the proximity of populations and climatic niche overlap, populations of the homoploid hybrid species are unlikely to be completely reproductively isolated from the parents based on ecological isolation alone. Our genetic analyses also demonstrate that *A. sundingii* and *A. lemsii* originated by independent hybridisation events. As such, *Argyranthemum* represents an example of independent homoploid hybrid speciation events, with evidence of divergence in morphology and parallel adaptation to novel intermediate ecological niches.

The molecular data provided support for the independent hybrid origin of *A. sundingii* and *A. lemsii*. Whether or not they represent a hybrid species depends largely on the definition employed.



Schumer *et al.* (2014) provided stringent criteria for a homoploid hybrid species which included evidence of reproductive isolation of hybrid lineages from the parental species, hybridisation in the genome, and of reproductive isolation as a consequence of hybridisation. These criteria have been debated in subsequent papers (Feliner Nieto *et al.* 2017; Schumer *et al.* 2018) as only *Helianthus* and *Heliconius* butterflies satisfy all three criteria, leading to a potential underestimate of the importance of HHS in evolution. Evidence of reproductive isolation and reproductive isolation derived from hybridisation are lacking from most putative examples of HHS (Schumer *et al.* 2014). In this study we have demonstrated the hybrid origin of *A. sundingii* and *A. lemsii* as well as evidence of adaptation to intermediate altitudes, which likely confers some, albeit incomplete, reproductive isolation from their parental progenitors. However, as of yet it is unclear whether or not hybridisation was directly responsible for the occupation of a novel habitat. Reciprocal transplant experiments to investigate if the hybrid species are more fit than parental species in intermediate habitats would allow this to be investigated.

Identifying the genetic mechanisms responsible for HHS is an important consideration for any future work. Jiggins *et al.* (2008) and Salazar *et al.* (2010) distinguished two different models of HHS: hybrid trait speciation and mosaic genome hybrid speciation. The former involves the introgression of a small number of adaptive loci (so-called magic traits) which confer reproductive isolation by a novel adaptation whereas the latter involves the stabilisation of the hybrid genome and reproductive isolation by intrinsic incompatibilities. It is not clear which mechanism may be acting in *Argyranthemum* and each is not necessarily mutually exclusive. The homoploid hybrid species in *Argyranthemum* appear to have inherited a larger proportion of the genome from *A. broussonetii* than *A. frutescens*, so it is plausible that the homoploid hybrid species have inherited a limited number of adaptive loci from *A. frutescens* although this would need to be demonstrated experimentally. Indeed, future studies in *Argyranthemum* or other examples of HHS should attempt to identify the genetic changes responsible for the morphological differences and/or ecological adaptation in the homoploid hybrid species relative to the parents. One approach would be to use transcriptomes to identify differentially expressed transcripts in the homoploid species relative to the parents.

Interspecific hybridisation is frequent in island floras (Howarth & Baum 2005; Carine *et al.* 2007; Friar *et al.* 2008; Lindhardt *et al.* 2009; Jones *et al.* 2014). This is likely the result of endemic lineages

possessing little intrinsic reproductive isolation and the dynamic ecological landscape of oceanic islands with natural and anthropogenic disturbances bringing previously isolated species into contact (Francisco-Ortega *et al.* 2000; van Hengstum *et al.* 2012; Crawford & Stuessy 2016; Kerbs *et al.* 2017). In addition, the variety of distinctive habitats in close proximity on oceanic islands also offers the opportunity for the establishment of hybrids free from competition with their parental progenitors. Although the evolutionary consequences of hybridisation are varied, there is increasing evidence that hybridisation within island lineages has promoted diversification. (Howarth & Baum 2005; Friar *et al.* 2008). Given the propensity for hybridisation in island lineages as well as the difficulty associated with homoploid hybrid species identification, it is plausible that HHS is more common in island lineages than we currently think and the processes that we document here may have played a much greater role in generating novel diversity in island plant radiations than we currently appreciate.

## Acknowledgements

The authors would like to acknowledge Cabildo de Tenerife and Gobierno de Canarias for providing permits, Arnaldo Santos Santos-Guerra for his knowledge of collection sites included in this study, Mike Cotton for the maintenance of glasshouse plants and members of Chapman lab for their comments on earlier drafts of this study. Suggestions from Roberto Novella Fernandez (University of Southampton) were very helpful in performing our ecological niche modelling analysis and comments from Yuval Sapir (Tel Aviv University) and two anonymous reviewers greatly improved the quality of the manuscript. We would also like to acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work. This work was funded by a NHM-UoS PhD studentship.

## References

- Baack EJ, Whitney KD, Rieseberg LH (2005) Hybridization and genome size evolution: Timing and magnitude of nuclear DNA content increases in *Helianthus* homoploid hybrid species. *New Phytologist*, **167**, 623–630.
- Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.
- Barrès B, Carlier J, Seguin M *et al.* (2012) Understanding the recent colonization history of a plant pathogenic fungus using population genetic tools and Approximate Bayesian Computation. *Heredity*, **109**, 269–279.
- Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*, **57**, 289–300.

- Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: Some cons, many pros. *Molecular Ecology*, **19**, 2609–2625.
- Borgen L (1976) Analysis of a hybrid swarm between *Argyranthemum aductum* and *A. filifolium* in the Canary Islands. *Nordic Journal of Botany*, **23**, 121–137.
- Borgen L (1980) A new species of *Argyranthemum* (Compositae) from the Canary Islands. *Nordic Journal of Botany*, **27**, 163–165.
- Borgen L, Leitch I, Santos-Guerra A (2003) Genome organization in diploid hybrid species of *Argyranthemum* (Asteraceae) in the Canary Islands. *Botanical Journal of the Linnean Society*, **141**, 491–501.
- Bramwell D, Bramwell Z (2001) *Wild Flowers of the Canary Islands*. Editorial Rueda, Spain.
- Brochmann C (1984) Hybridization and distribution of *Argyranthemum coronopifolium* (Asteraceae, Anthemideae) in the Canary Islands. *Nordic Journal of Botany*, **4**, 729–736.
- Brochmann C (1987) Evaluation of some methods for hybrid analysis, exemplified by hybridisation in *Argyranthemum*. *Nordic Journal of Botany*, **7**, 609–630.
- Brochmann C, Borgen L, Stabbetorp OE (2000) Multiple diploid hybrid speciation of the Canary Island endemic *Argyranthemum sundingii* (Asteraceae). *Plant Systematics and Evolution*, **220**, 77–92.
- Broennimann O, Fitzpatrick MC, Pearman PB *et al.* (2012) Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography*, **21**, 481–497.
- Bryan GJ, McNicoll J, Ramsay G, Meyer RC, De Jong WS (1999) Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. *Theoretical and Applied Genetics*, **99**, 859–867.
- Buerkle CA, Morris RJ, Asmussen MA, Rieseberg LH (2000) The likelihood of homoploid hybrid speciation. *Heredity*, **84**, 441–451.
- Capblancq T, D L, Rioux D, Mavarez J (2015) Hybridization promotes speciation in *Coenonympha* butterflies. *Molecular Ecology*, **24**, 6209–6222.
- Carine MA, Robba L, Little R, Russell S, Guerra AS (2007) Molecular and morphological evidence for hybridization between endemic Canary Island *Convolvulus*. *Botanical Journal of the Linnean Society*, **154**, 187–204.
- Choisy M, Franck P, Cornuet JM (2004) Estimating admixture proportions with microsatellites : comparison of methods based on simulated data. *Molecular Ecology*, **13**, 955–968.
- Di Cola V, Broennimann O, Petitpierre B *et al.* (2017) ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography*, **40**, 774–787.
- Cornuet J-M, Ravigné V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC bioinformatics*, **11**, 401.
- Cornuet JM, Santos F, Beaumont MA *et al.* (2008) Inferring population history with DIY ABC: A user-friendly approach to approximate Bayesian computation. *Bioinformatics*, **24**, 2713–2719.
- Crawford DJ, Stuessy TF (2016) Cryptic variation, molecular data, and the challenge of conserving plant diversity in oceanic archipelagos: the critical role of plant systematics. *Taxon*, **46**, 129–148.
- Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
- Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.
- Donovan LA, Rosenthal DR, Sanchez-Velenosi M, Rieseberg LH, Ludwig F (2010) Are hybrid species more fit than ancestral parent species in the current hybrid species habitats? *Journal of Evolutionary Biology*, **23**, 805–816.
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.
- Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

- Eaton D, Ree RH (2013) Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (Pedicularis: Orobanchaceae). *Systematic Biology*, **62**, 689–706.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, **14**, 2611–2620.
- Excoffier L, Foll M (2011) fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, **27**, 1332–1334.
- Feliner Nieto G, Álvarez I, Fuertes-Aguilar J *et al.* (2017) Is homoploid hybrid speciation that rare? An empiricist's view. *Heredity*, 513–516.
- Fjellheim S, Jørgensen MH, Kjos M, Borgen L (2009) A molecular study of hybridization and homoploid hybrid speciation in *Argyranthemum* (Asteraceae) on Tenerife, the Canary Islands. *Botanical Journal of the Linnean Society*, **159**, 19–31.
- Francisco-Ortega J, Crawford DJ, Santos-Guerra A, Jansen RK (1997a) Origin and Evolution of *Argyranthemum*. In: *In Molecular Evolution and Adaptive Radiation* (eds Givnish TJ, Sytsma KJ), pp. 407–432.
- Francisco-Ortega J, Crawford DJ, Santos-Guerra A, Sa-Fontinha S (1995) Genetic divergence among Mediterranean and Macronesian genera of the subtribe Chrysantheminae (Asteraceae). *American Journal of Botany*, **82**, 1321–1328.
- Francisco-Ortega J, Jansen RK, Santos-Guerra A (1996) Chloroplast DNA evidence of colonization, adaptive radiation, and hybridization in the evolution of the Macaronesian flora. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 4085–4090.
- Francisco-ortega AJ, Santos-guerra A, Hines A, Robert K (1997) Molecular Evidence for a Mediterranean Origin of the Macaronesian Endemic Genus *Argyranthemum* (Asteraceae). , **84**, 1595–1613.
- Francisco-Ortega J, Santos-Guerra A, Jansen RK (1997b) Origin and Evolution of *Argyranthemum*. In *Molecular Evolution and Adaptive Radiation*. In: *Molecular Evolution and Adaptive Radiation*, pp. 407–432. Cambridge University Press, Cambridge.
- Francisco-Ortega J, Santos-Guerra A, Kim SC, Crawford DJ (2000) Plant genetic diversity in the Canary Islands: A conservation perspective. *American Journal of Botany*, **87**, 909–919.
- Friar E a., Prince LM, Cruse-Sanders JM *et al.* (2008) Hybrid Origin and Genomic Mosaicism of *Dubautia scabra* (Hawaiian Silversword Alliance; Asteraceae, Madiinae). *Systematic Botany*, **33**, 589–597.
- Gonzalez AG, Estevez-Reyes R, Estevez-Braun A *et al.* (1997) Biological activities of some *Argyranthemum* species. *Phytochemistry*, **45**, 963–967.
- Goulet BE, Roda F, Hopkins R (2017) Hybridization in Plants: Old Ideas, New Techniques. *Plant Physiology*, **173**, 65–78.
- Hegarty MJ, Hiscock SJ (2005) Hybrid speciation in plants: new insights from molecular studies. *The New phytologist*, **165**, 411–423.
- van Hengstum T, Lachmuth S, Oostermeijer JGB *et al.* (2012) Human-induced hybridization among congeneric endemic plants on Tenerife, Canary Islands. *Plant Systematics and Evolution*, **298**, 1119–1131.
- Howarth DG, Baum DA (2005) Genealogical Evidence of Homoploid Hybrid Speciation in an Adaptive Radiation of *Scaevola* (Goodeniaceae) in the Hawaiian Islands. *Evolution*, **59**, 948–961.
- Humphries CJ (1973) A Taxonomic study of the genus *Argyranthemum* Webb Ex Sch. BIP. University of Reading.
- Humphries CJ (1976) A Revision of the Macaronesian Genus *Argyranthemum* Webb Ex Schultz BIP. (Compositae-Anthemideae). *Bulletin of the British Museum Natural History*, **5**, 145–240.
- 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.

- James JK, Abbott RJ (2005) Recent, allopatric, homoploid hybrid speciation: the origin of *Senecio squalidus* (Asteraceae) in the British Isles from a hybrid zone on Mount Etna, Sicily. *Evolution; international journal of organic evolution*, **59**, 2533–2547.
- Jiggins CD, Salazar C, Linares M, Mavarez J, B PTRS (2008) Review. Hybrid trait speciation and Heliconius butterflies. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **363**, 3047–3054.
- Jombart T, Ahmed I (2011) adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.
- Jones KE, Reyes-Betancort JA, Hiscock SJ, Carine MA (2014) Allopatric diversification, multiple habitat shifts, and hybridization in the evolution of *Pericallis* (Asteraceae), a macaronesian endemic genus. *American Journal of Botany*, **101**, 637–651.
- Kadereit JW (2015) The geography of hybrid speciation in plants. *Taxon*, **64**, 673–687.
- Kerbs B, Ressler J, Kelly JK *et al.* (2017) The potential role of hybridization in diversification and speciation in an insular plant lineage: insights from synthetic interspecific hybrids. *AoB PLANTS*, **9**, 1–12.
- Lindhardt MS, Philipp M, Tye A, Nielsen LR (2009) Molecular, morphological, and experimental evidence for hybridization between threatened species of the galapagos endemic genus *scalesia* (asteraceae). *International Journal of Plant Sciences*, **170**, 1019–1030.
- Liu B, Abbott RJ, Lu Z, Tian B, Liu J (2014) Diploid hybrid origin of *Ostryopsis intermedia* (Betulaceae) in the Qinghai-Tibet Plateau triggered by Quaternary climate change. *Molecular Ecology*, **23**, 3013–3027.
- Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, **20**, 229–237.
- Mallet J, James M, Mallet J (2007) Hybrid speciation. *Nature*, **446**, 279–283.
- Mavárez J, Linares M (2008) Homoploid hybrid speciation in animals. *Molecular Ecology*, **17**, 4181–4185.
- Moreno JC (2008) *Lista Roja 2008 de la flora vascular española*. Dirección General de Medio Natural y Política Forestal (Ministerio de Medio Ambiente y Medio Rural y Marino), y Sociedad Española de Biología de la Conservación de Plantas, Madrid, Spain.
- Nolte AW, Tautz D (2010) Understanding the onset of hybrid speciation. *Trends in genetics*, **26**, 54–8.
- Peakall R, Smouse PE (2012) GenALEX 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, **28**, 2537–2539.
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.
- Phillips SJ, Dudík M, Schapire RE (2018) Maxent software for modeling species niches and distributions (Version 3.4.1).
- Poland JA, Rife TW (2012) Genotyping-by-Sequencing for Plant Breeding and Genetics. *The Plant Genome Journal*, **5**, 92.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Purcell S, Neale B, Todd-brown K *et al.* (2007) PLINK : A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. , **81**, 559–575.
- R Core Team (2018) R: A Language and Environment for Statistical Computing.
- Rieseberg LH (1997) Hybrid Origins of Plant Species. *Annual Review of Ecology, Evolution and Systematics*, 359–389.
- Rieseberg LH, Raymond O, Rosenthal DM *et al.* (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, **301**, 1211–6.
- Rieseberg LH, Willis JH (2007) Plant speciation. *Science (New York, N.Y.)*, **317**, 910–4.
- Salazar C, Baxter SW, Pardo-Díaz C *et al.* (2010) Genetic evidence for hybrid trait speciation in *Heliconius* butterflies. *PLoS Genetics*, **6**.

- Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, **9**, 671–675.
- Schoener TW (1968) The Anolis Lizards of Bimini : Resource Partitioning in a Complex Fauna. *Ecology*, **49**, 704–726.
- Schönfelder I, Schönfelder P (2012) *Die Kosmos-Kanarenflora: Über 1000 Arten und 60 tropische Ziergehölze*.
- Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature biotechnology*, **18**, 233–234.
- Schumer M, Rosenthal GG, Andolfatto P (2014) How common is homoploid hybrid speciation? *Evolution*, **68**, 1553–1560.
- Schumer M, Rosenthal GG, Andolfatto P (2018) What do we mean when we talk about hybrid speciation? *Heredity*.
- Schwarzbach AE, Rieseberg LH (2002) Likely multiple origins of a diploid hybrid sunflower species. *Molecular Ecology*, **11**, 1703–1715.
- Sherman NA, Burke JM (2009) Population genetic analysis reveals a homoploid hybrid origin of *Stephanomeria diegensis* (Asteraceae). *Molecular Ecology*, **18**, 4049–4060.
- Soltis PS, Soltis DE (2009) The role of hybridization in plant speciation. *Annual Review of Plant Biology*, **60**, 561–88.
- Tovar-Sánchez E, Oyama K (2004) Natural hybridization and hybrid zones between *Quercus crassifolia* and *Quercus crassipes* (Fagaceae) in Mexico: Morphological and molecular evidence. *American Journal of Botany*, **91**, 1352–1363.
- Wang XR, Szmidt AE, Savolainen O (2001) Genetic composition and diploid hybrid speciation of a high mountain pine, *Pinus densata*, native to the Tibetan plateau. *Genetics*, **159**, 337–346.
- Warren DL, Glor RE, Turelli M (2010) ENMTools: A toolbox for comparative studies of environmental niche models. *Ecography*, **33**, 607–611.
- Warren DL, Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*, **62**, 2868–2883.
- White OW, Doo B, Carine MA, Chapman MA (2016) Transcriptome Sequencing and Simple Sequence Repeat Marker Development for Three Macaronesian Endemic Plant Species. *Applications in Plant Sciences*, **4**, 1600050.
- Wickham H (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wood TE, Takebayashi N, Barker MS *et al.* (2009) The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 13875–9.
- Worth JRP, Larcombe MJ, Sakaguchi S *et al.* (2016) Transient hybridization, not homoploid hybrid speciation, between ancient and deeply divergent conifers. *American Journal of Botany*, **103**, 246–259.

## Data Accessibility Statement

Leaf morphological data and ecological data extracted from the filtered points are available in Supplementary Tables 1 and 2 respectively. Demultiplexed Genotyping-By-Sequencing (GBS) reads for each sample have been submitted to Dryad (<https://doi.org/10.5061/dryad.q4d2v6h>) and other data files are available from the authors upon request.

## Tables

Table 1 – Number of individuals sampled for population genetics of Simple Sequence Repeat (SSR) markers, leaf morphological analysis and Genotyping-By-Sequencing (GBS). Taxa abbreviations used throughout are shown in brackets.

Pop.	Taxon	Locality	Latitude	Longitude
A	<i>A. broussonetii</i> subsp. <i>broussonetii</i> (bro)	Barranco de Valle Crispin	28.5306	-16.2
B		Las Casas de la Cumbre	28.5354	-16.2
C		Path to Mesa del Sabinal	28.5585	-16.1
D		La Cumbrilla	28.5663	-16.1
E		Chamorga	28.5722	-16.1
F		Roques del Fraile	28.5535	-16.2
G	<i>A. sundingii</i> (sun)	Valle Crispin	28.5148	-16.2
H		Valle Brosque	28.5196	-16.2
I		Roque Cubo	28.5239	-16.2
J		Barranco del Cercado de Andrés	28.5298	-16.2
K	<i>A. broussonetii</i> × <i>A. frutescens</i> (bro × fru)	Barranco de Igueste	28.5427	-16.1
L		Lomo de las Casillas	28.5529	-16.1
M	<i>A. lemsii</i> (lem)	Path to Mesa del Sabinal	28.5582	-16.1
N		La Cumbrilla	28.5667	-16.1
O		Barranco de Roque Bermejo	28.5741	-16.1
P	<i>A. frutescens</i> subsp. <i>frutescens</i> (frf)	Maria Jiménez	28.5053	-16.2
Q		San Andrés	28.5149	-16.1
R		Igueste de San Andreas	28.5281	-16.1
S	<i>A. frutescens</i> subsp. <i>succulentum</i> (frs)	Roque Bermejo	28.5782	-16.1
T		Between Almáciga and Roque Bermejo	28.5819	-16.1

Table 2 – Schoener's *D* (Schoener 1968) and Warren's *I* (Warren *et al.* 2008) niche overlap statistics based on our Maxent and PCA-env niche predictions. *P* values were generated using the\_niche equivalency test. Taxa are abbreviated to bro (*A. broussonetii*), sun (*A. sundingii*), bro × fru (*A. broussonetii* × *A. frutescens*), frf (*A. frutescens* subsp. *frutescens*) and frs (*A. frutescens* subsp. *succulentum*).

	Maxent		PCA-env	
	D	I	D	I
bro vs. frf	0.016 **	0.059 **	0.000 **	0.000 **
bro vs. frs	0.042 **	0.140 **	0.000 **	0.001 **
bro vs. sun	0.230 **	0.448 **	0.163 **	0.216 **
bro vs. lem	0.369 **	0.632 **	0.343 **	0.417 **
bro vs. bro × fru	0.267 **	0.526 **	0.175 **	0.253 **
frf vs. frs	0.264 **	0.552 **	0.3130	0.410 *
frf vs. sun	0.231 **	0.477 **	0.102 **	0.175 **
frf vs. lem	0.135 **	0.325 **	0.065 **	0.116 **

frf vs. bro × fru	0.207 **	0.433 **	0.120 **	0.179 **
frs vs. sun	0.155 **	0.374 **	0.064 **	0.178 **
frs vs. lem	0.279 **	0.561 **	0.046 **	0.125 **
frs vs. bro × fru	0.183 **	0.431 **	0.067 **	0.130 **
sun vs. lem	0.376 **	0.649 **	0.620 *	0.760 *
sun vs. bro × fru	0.513 **	0.790 **	0.387 **	0.536 **
lem vs. bro × fru	0.468 **	0.765 **	0.360 **	0.487 **

Niche overlap test significance denoted as  $P < 0.05$  (\*),  $P < 0.01$  (\*\*),  $P < 0.001$  (\*\*\*)

Table 3 - Distribution of haplotypes across species and groups as defined by presence in parental species. Taxa are abbreviated to bro (*A. broussonetii*), sun (*A. sundingii*), bro × fru (*A. broussonetii* × *A. frutescens*), frf (*A. frutescens* subsp. *frutescens*) and frs (*A. frutescens* subsp. *succulentum*).

Taxon	Parental groups and haplotypes															
	bro				fru						shared		neither			
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI
bro	1	2	9	28	0	0	0	0	0	0	5	1	0	0	0	0
frf	0	0	0	0	1	6	3	1	1	3	2	4	0	0	0	0
frs	0	0	0	0	0	0	0	0	4	8	0	0	0	0	0	0
sun	0	2	0	22	0	4	0	0	0	0	8	0	0	0	1	0
bro × fru	0	0	0	0	0	6	0	0	3	0	1	0	0	0	0	1
lem	0	0	0	2	3	7	11	0	2	6	3	0	1	2	0	0

## Figures

Figure 1 - Populations sampled across (A) Tenerife in the (B) Anaga Peninsula. Taxa abbreviations used throughout are shown in brackets. Contour lines represent a 200m change in altitude.



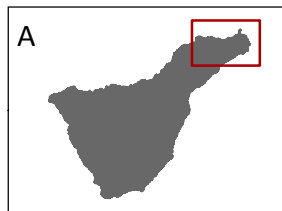
Figure 2 – Boxplots (A-D) and PCA (E) based on morphological characters leaf area, perimeter, length and width. Letters above each boxplot plot represent the groupings identified by a *post hoc* Tukey test with false discovery rate (FDR) correction (Benjamini & Hochberg 1995) for multiple comparisons. Taxa are abbreviated to bro (*A. broussonetii*), sun (*A. sundingii*), bro × fru (*A. broussonetii* × *A. frutescens*), frf (*A. frutescens* subsp. *frutescens*) and frs (*A. frutescens* subsp. *succulentum*).

Figure 3 – niche space for each species based on (A) Maxent and (B) ecospat. The Maxent maps were created using the average predictions across 10 replicates and average cloglog thresholds under the maximum training sensitivity plus specificity criterion. Ordinations generated in ecospat show the niche space for each species across the first two principal components. The density of occurrences for each species is represented by grey shading and the solid and dashed contour lines show 100 % and 50 % of the available background environment respectively. Taxa are abbreviated to bro (*A. broussonetii*), sun (*A. sundingii*), bro × fru (*A. broussonetii* × *A. frutescens*), lem (*A. lemsii*), frf (*A. frutescens* subsp. *frutescens*) and frs (*A. frutescens* subsp. *succulentum*).

Figure 4 – Evanno delta *K* (A), STRUCTURE results for *K* = 2 to 5 (B) and PCA (C) based on nuclear SSR clustering analyses. Taxa included in the PCA and STRUCTURE plots are abbreviated to bro (*A. broussonetii*), sun (*A. sundingii*), bro × fru (*A. broussonetii* × *A. frutescens*), lem (*A. lemsii*), frf (*A. frutescens* subsp. *frutescens*) and frs (*A. frutescens* subsp. *succulentum*).

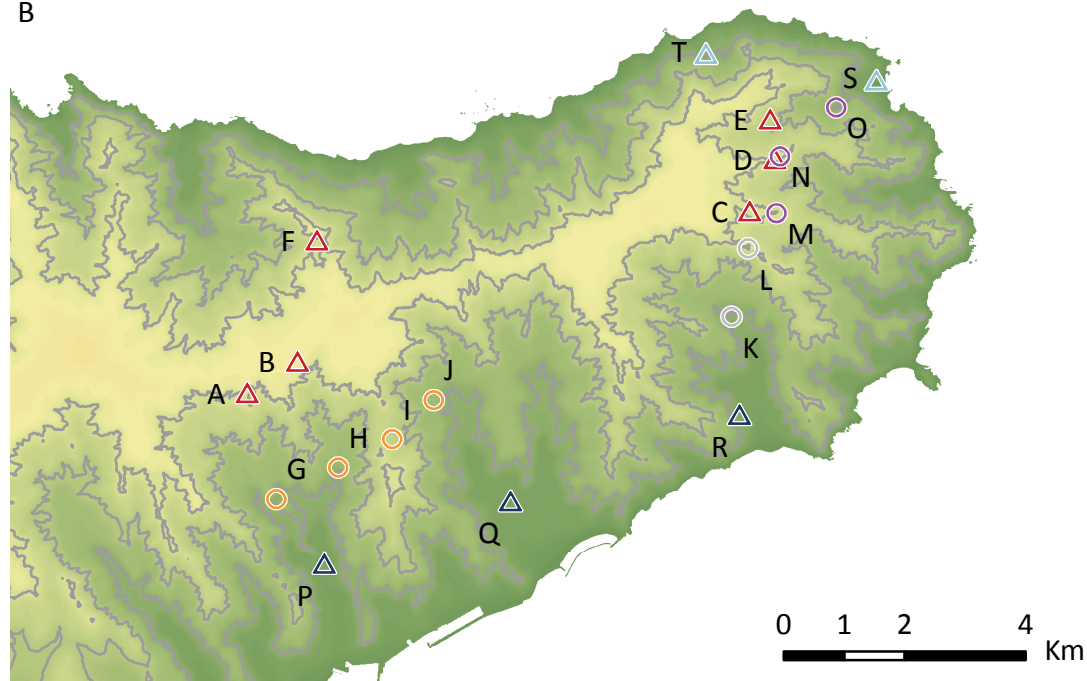
Figure 5 - Evanno delta *K* (A), STRUCTURE results for *K* = 2 to 5 (B) and PCA (C) based on an ipyrad assembly of GBS reads, using a 90 % clustering threshold and a minimum of 10 samples required to process a locus. For the STRUCTURE and PCA plots taxa are abbreviated to bro (*A. broussonetii*), sun (*A. sundingii*), bro × fru (*A. broussonetii* × *A. frutescens*), lem (*A. lemsii*), frf (*A. frutescens* subsp. *frutescens*) and frs (*A. frutescens* subsp. *succulentum*). Figure 6 – scenarios 1 to 9 included in the DIYABC analysis based on nuclear SNPs comparing two hybridisation events (scenarios 1-5), a single hybridisation event (scenarios 6-8) and cladogenesis (scenario 9). Scenario 1 was selected as being most likely in describing the origin of the homoploid hybrid species

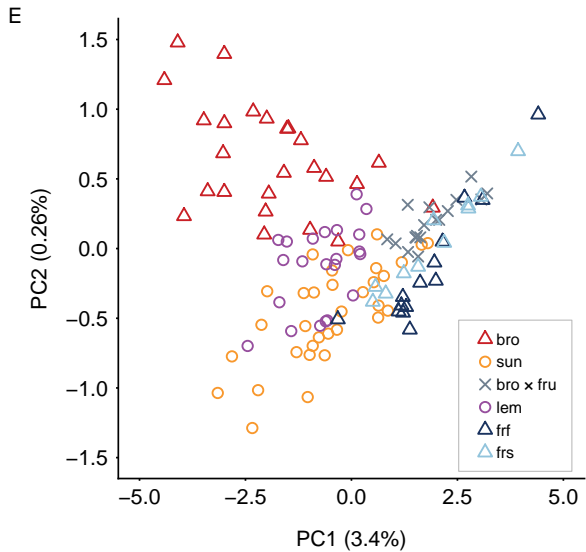
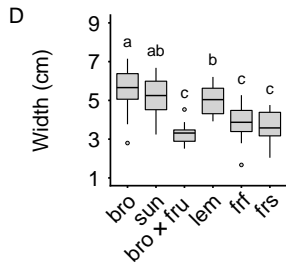
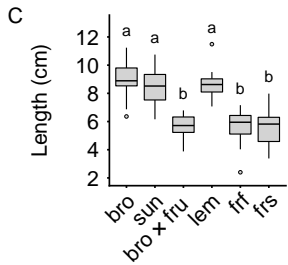
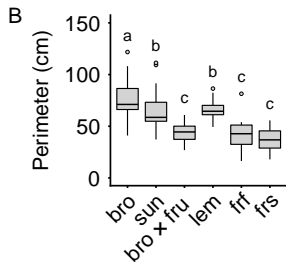
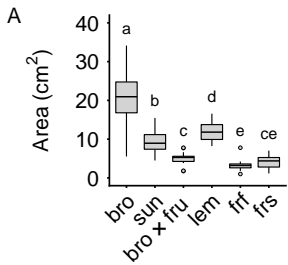
in *Argyranthemum*. Taxa are abbreviated to bro (*A. broussonetii*), sun (*A. sundingii*), lem (*A. lemsii*), frf (*A. frutescens* subsp. *frutescens*) and frs (*A. frutescens* subsp. *succulentum*). Generations are shown on the y axis (t0 to td) and admixture proportions from each parent are shown on scenario one.



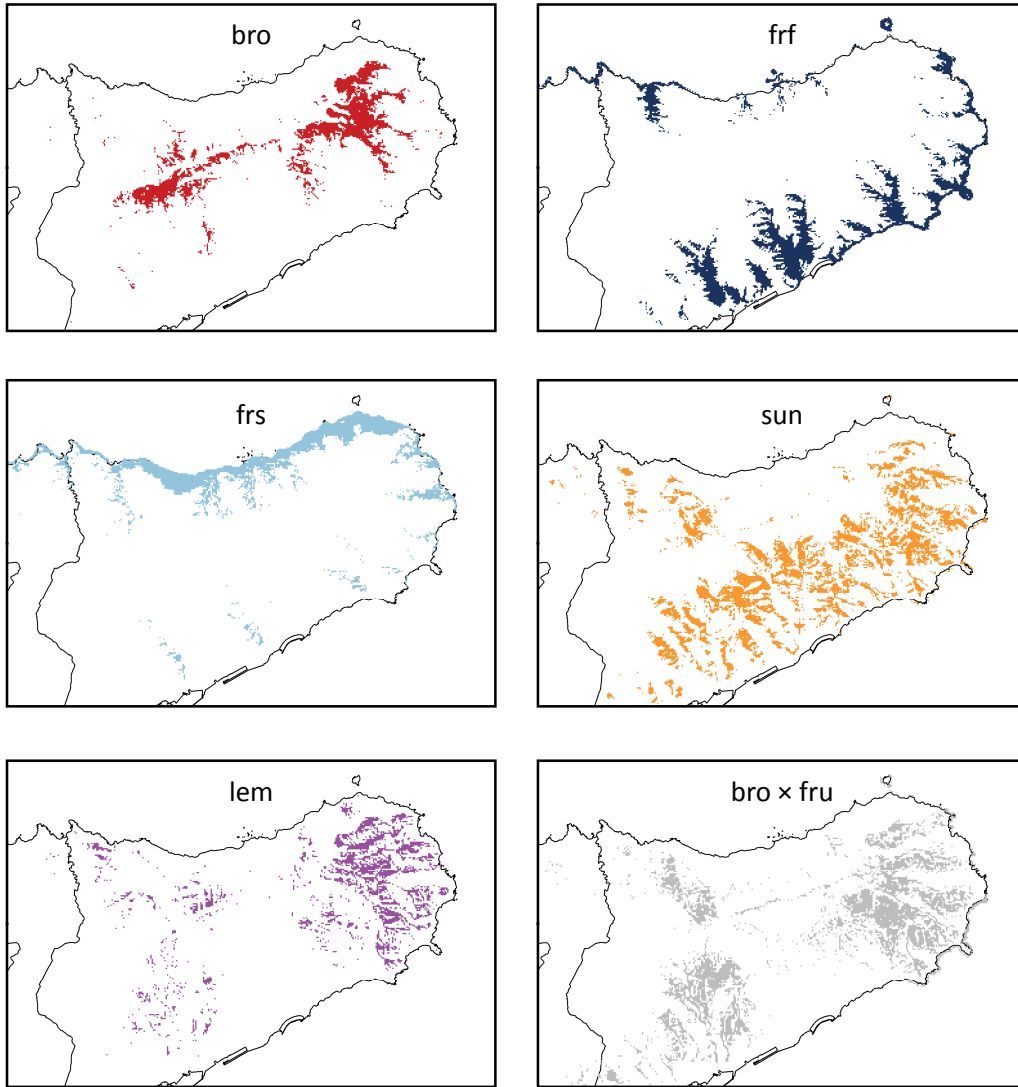
- △ *A. broussonetii* (bro)
- *A. sundingii* (sun)
- *A. broussonetii* x *A. frutescens* (bro x fru)
- *A. lemsii* (lem)
- △ *A. frutescens* subsp. *frutescens* (frf)
- △ *A. frutescens* subsp. *succulentum* (frs)

B

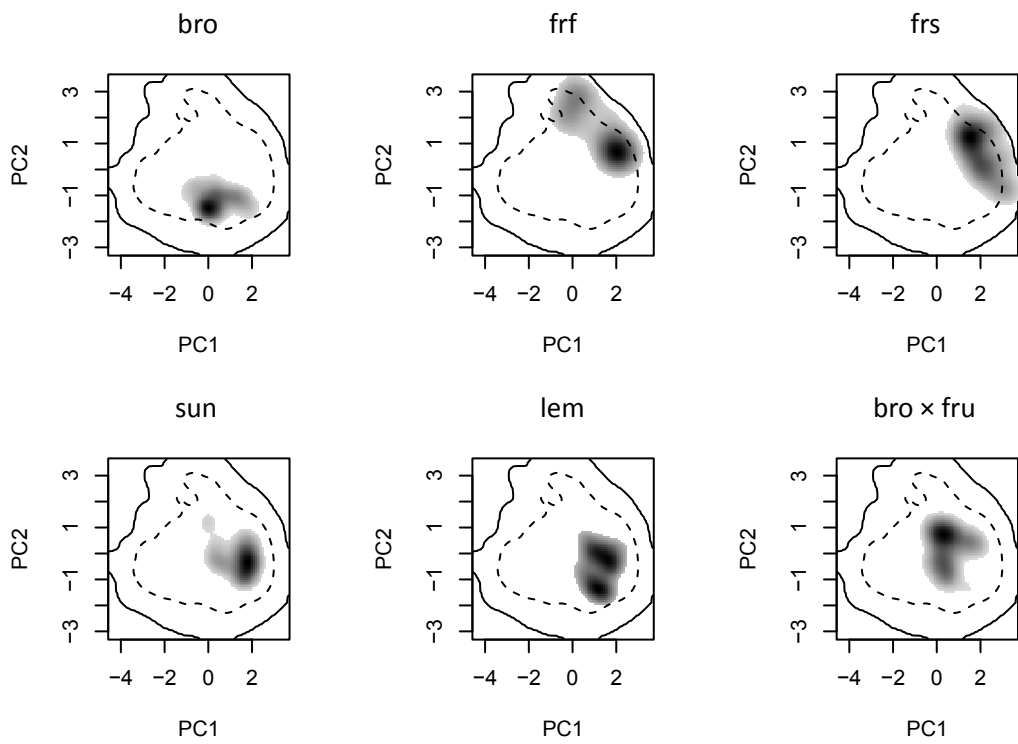


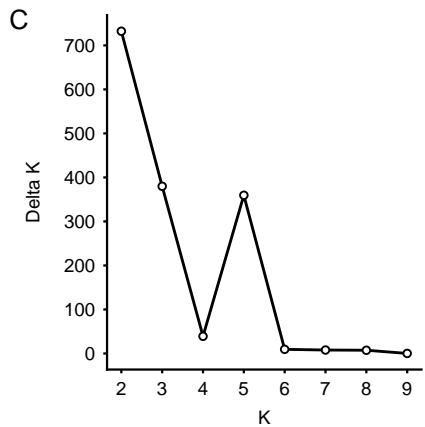
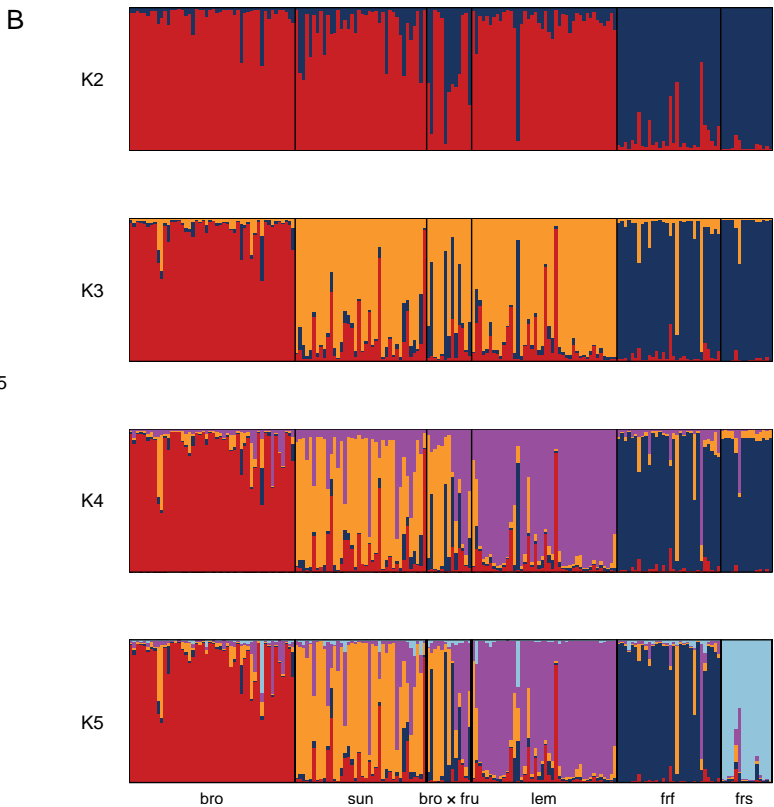
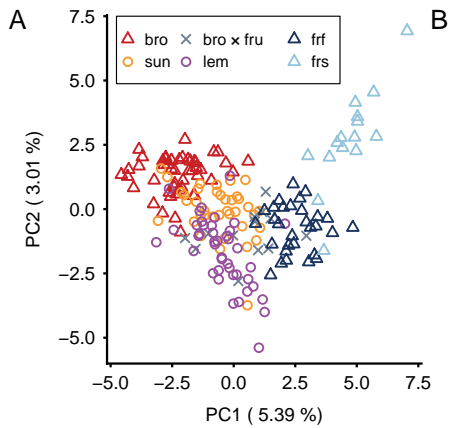


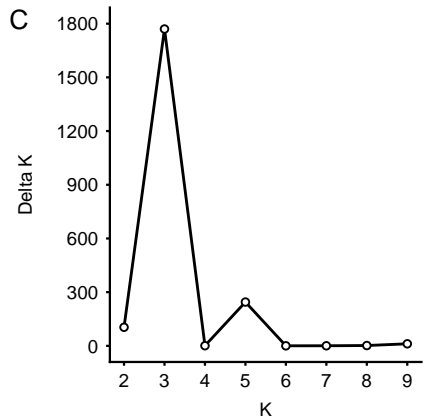
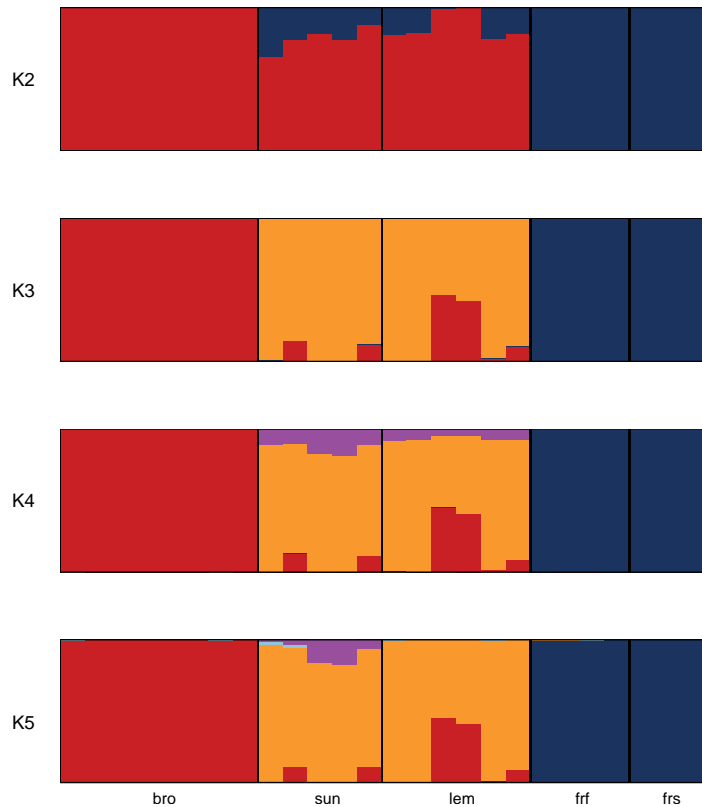
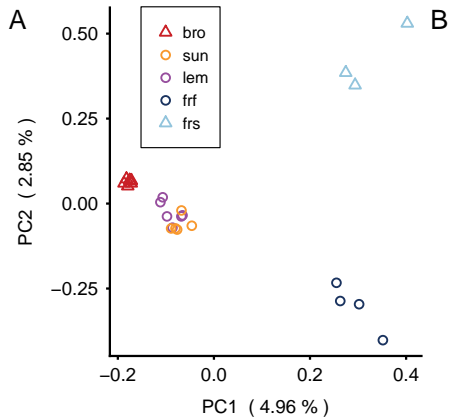
A



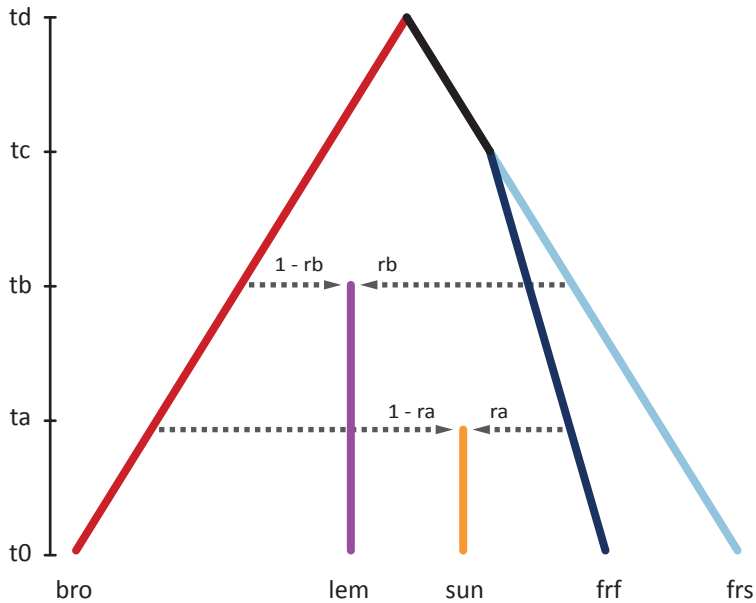
B



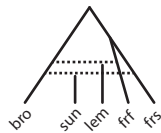




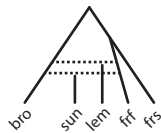
1



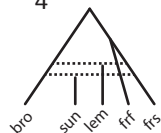
2



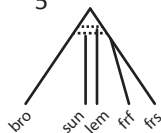
3



4



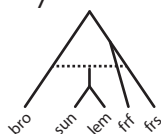
5



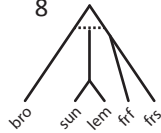
6



7



8



9

