MOLECULAR ECOLOGY

Molecular Ecology (2009) **18**, 4049–4060 doi: 10.1111/j.1365-294X.2009.04349.x

Population genetic analysis reveals a homoploid hybrid origin of *Stephanomeria diegensis* (Asteraceae)

NATASHA A. SHERMAN and JOHN M. BURKE

Department of Plant Biology, University of Georgia, Athens, GA, 30602, USA

Abstract

Homoploid hybrid speciation has generally been viewed as a rare evolutionary phenomenon, with relatively few well-documented cases in nature. Here, we investigate the origin of *Stephanomeria diegensis*, a diploid flowering plant species that has been proposed to have arisen as a result of hybridization between *S. exigua* and *S. virgata*. Across the range of *S. diegensis*, all individuals share a common chloroplast haplotype with *S. virgata* while showing a greater affinity for *S. exigua* in terms of nuclear genetic diversity. A prinicipal coordinates analysis (PCO) based on the nuclear data revealed that *S. diegensis* is most similar to each parent along different axes. Moreover, a Bayesian clustering analysis as well as a hybrid index-based analysis showed evidence of mixed ancestry, with approximately two thirds of the *S. diegensis* nuclear genome derived from *S. exigua*. These results provide strong support for a homoploid hybrid origin of *S. diegensis*. Finally, contrary to the finding that homoploid hybrid species are typically multiply-derived, our results were most consistent with a single origin of this species.

Keywords: genetic variation, hybridization, reproductive isolation, reticulate evolution, speciation

Received 20 June 2009; revision received 29 July 2009; accepted 6 August 2009

Introduction

Under the Biological Species Concept, speciation is defined as the development of reproductive isolating mechanisms between lineages that share a common ancestor (Mayr 1942). Because the path to complete cross-incompatibility is often lengthy, lineages that have begun to diverge may experience secondary reproductive contact. There are a number of possible evolutionary outcomes when incompletely isolated taxa come back into reproductive contact. For example, the taxa might merge into a single, highly variable population (e.g. Grant 1963), or one taxon might drive the other to extinction via genetic assimilation (e.g. Wolf *et al.* 2001). If, on the other hand, the incipient reproductive barriers are sufficiently strong and hybrid progeny perform poorly, selection against individuals that mate with the 'wrong' type might result in the reinforcement of prezygotic reproductive barriers (reviewed in Howard 1993; Servedio & Noor 2003). Alternatively, hybridization can have creative outcomes, including the introgression of alleles from one taxon into another (reviewed in Rieseberg & Wendel 1993; see also Bailey *et al.* 2009; Currat *et al.* 2008; Gagnaire *et al.* 2009; Kawakami *et al.* 2009; Kim *et al.* 2008; Lepais *et al.* 2009; Wood & Nakazato 2009), the origin of novel adaptations (reviewed in Arnold 2004), and even hybrid speciation (reviewed in Rieseberg 1997).

The most common form of hybrid speciation is allopolyploidy, in which a cross between two species gives rise to a hybrid lineage that carries the full complement of chromosomes from both parental species. In such cases, the hybrid lineage will be reproductively isolated from its parents due to the increase in chromosome number (reviewed in Ramsey & Schemske 1998; Otto & Whitton 2000). Alternatively, hybrid speciation can occur without a change in chromosome number – a phenomenon known as homoploid hybrid speciation (reviewed in Rieseberg 1997; see also Rieseberg *et al.* 2003; Gross & Rieseberg 2005). The most widely accepted model of homoploid hybrid speciation is the recombinational model of Stebbins (1957) and Grant (1958), wherein reproductive isolation results from the

Correspondence: John M. Burke, Fax: 706-542-1805;
E-mail: jmburke@uga.edu

production of novel combinations of sterility factors following hybridization between chromosomally or genetically divergent parental species. The resulting hybrids are interfertile with one another but are at least partially isolated from both of their parents, and thus free to evolve independently. Alternatively, hybridization between two ecologically distinct taxa can result in the production of a unique homozygous recombinant type isolated by external, rather than internal, barriers (Grant 1981).

While hybridization is widespread in plants (Mallet 2005), the conditions under which homoploid hybrid speciation can occur are stringent (McCarthy *et al.* 1995; Buerkle *et al.* 2000, 2003). This mode of speciation has thus generally been regarded as rare. Nonetheless, the taxonomic literature includes a number of proposed instances of homoploid hybrid speciation (Arnold 1997; Rieseberg 1997; Gross & Rieseberg 2005). Unfortunately, homoploid hybrid species are difficult to unambiguously identify, and only a handful of cases have been substantiated using molecular approaches. In some instances, these hybrid species have very restricted ranges (e.g. Arnold 1993), whereas in others they have achieved a more widespread distribution, either through a single initial hybridization event followed by range expansion, or via multiple origins (e.g. Schwarzbach & Rieseberg 2002; Welch & Rieseberg 2002; James & Abbott 2005). Here we investigate the origin of *Stephanomeria diegensis*, a putative homoploid hybrid species with a range that spans much of coastal southern California.

The genus *Stephanomeria*, one of *c.* 1100 genera within the sunflower family (Compositae or Asteraceae; Jeffrey 1993) is composed of six annual and ten perennial species and has a base chromosome number of $n = 8$. In his initial systematic treatment of the genus, Gottlieb (1971) identified the annual *S. diegensis* as a putative hybrid species, having most likely arisen as a result of interbreeding between *S. exigua* and *S. virgata*. These latter species are polytypic annuals that are composed of five and two subspecies, respectively. Both species have broad ranges in the western United States, and F_1 hybrids between them exhibit *c.* 10% pollen fertility (Gottlieb 1969).

The possibility of a hybrid origin of *S. diegensis* was first invoked because this species combines several morphological characters that distinguish *S. exigua* and *S. virgata*, though it also exhibits some transgressive (i.e. extreme) characters (e.g. the number of florets per inflorescence and length of lateral pinnae along the pappus bristles). In terms of geographic distribution, the range of *S. diegensis* falls entirely within the region of overlap between *S. exigua* and *S. virgata*. Finally, crossability with both of its putative parents is low. In fact, pollen fertility in F_1 hybrids between *S. diegensis* and its putative parents indicate that it is the most strongly reproductively isolated of all annual *Stephanomeria* species (Gottlieb1969). Interestingly, this pattern of increased reproductive isolation in hybrid species has been documented in other study systems (e.g. Rieseberg 2000; Lai *et al.* 2005).

Gottlieb's (1969, 1971) initial conclusions regarding the origin of *S. diegensis* were later supported by allozyme data (Gallez & Gottlieb 1982). More specifically, the *S. diegensis* gene pool was found to include alleles from both *S. exigua* and *S. virgata*, and there was a paucity of unique *S. diegensis* alleles. Unfortunately, the majority of loci surveyed in that work were uninformative because the putative parental species shared the same majority allele. Thus, while the results were consistent with a hybrid origin, they were far from conclusive. Beyond this, Gallez & Gottlieb (1982) relied on relatively limited geographic sampling and only included three of the five *S. exigua* subspecies in their study. As noted above, *S. exigua* and *S. virgata* co-occur and hybridize throughout the distribution of *S. diegensis*, yet the possible role of multiple origins in producing the widespread range of their putative hybrid daughter species has never been investigated. Moreover, the identity of the parents of *S. diegensis* (in terms of particular subspecies of *S. exigua* and *S. virgata*) remains unknown. Here we use chloroplast and nuclear DNA markers to test the hypothesis of a hybrid origin of *S. diegensis*, investigate the possibility of multiple origins, and identify the most likely parental subspecies.

Materials and methods

Plant materials and DNA extractions

Buds and/or achenes (single-seeded fruits) were collected from twenty-six populations of *S. exigua*, twelve populations of *S. virgata*, and nine populations of *S. diegensis* spanning the known range of this species. The collection sites ranged from southern San Diego County, California north through Mariposa County, California (Fig. 1; Table S1). Because *S. exigua* comprises five subspecies, and *S. virgata* comprises two subspecies, care was taken to include samples from all taxa throughout the range. Buds collected in the field were preserved in a saturated sodium chloride, 30% CTAB solution, shipped back to the lab, and stored at −20°C until extractions were completed (Rogstad 1992). Voucher specimens were collected for one or two individuals per population and were deposited at the University of Georgia herbarium. Total genomic DNA was isolated from 219 *S. exigua*, 79 *S. virgata*, and 79 *S. diegensis* individuals. DNA was isolated from either the preserved

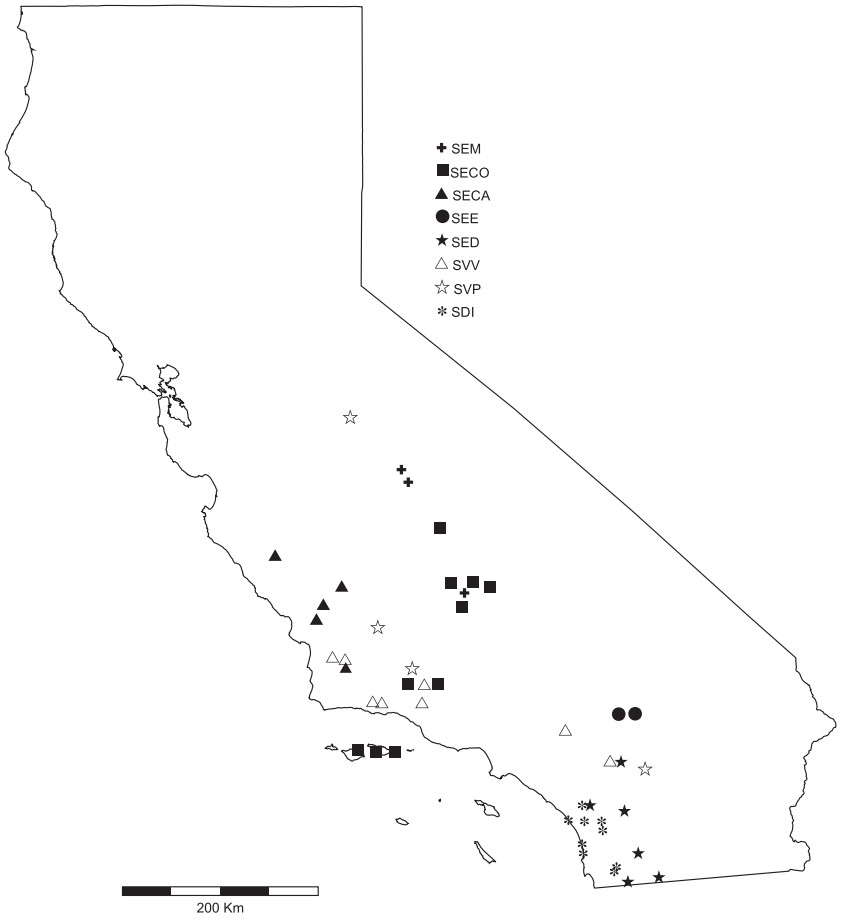

Fig. 1 Map of California, USA, showing the locations of *Stephanomeria* populations sampled for this study. SEM, *S. exigua* ssp. *macrocarpa*; SECO, *S. exigua* ssp. *coronaria*; SECA, *S. exigua* ssp. *carotifera*; SEE, *S. exigua* ssp. *exigua*; SED, *S. exigua* ssp. *deanei*; SDI, *S. digenesis*; SVV, *S. virgata* ssp. *virgata*; SVP, *S. virgata* ssp. *pleurocarpa*.

buds or from buds of plants grown in the greenhouse grown from field-collected achenes using a modified CTAB-based protocol (Doyle & Doyle 1990).

Chloroplast DNA analysis

Seven regions were amplified from the chloroplast genome using the polymerase chain reaction (PCR) and then sequenced for three individuals of each species (GenBank Accession no.: GQ429013-GQ429054). The regions amplified were: *atp*B-*rbc*L, *psb*B-*psb*F, *trn*K-*rpl*32, *rpl*36-*inf*A-*rps*8, *pet*N1-*psb*M2 section of *trn*C-*trn*D, *trn*L(UAA)5′exon-*trn*F(GAA), *trn*L-*rpl*32, (Taberlet *et al.* 1991; Johnson & Soltis 1995; Chiang *et al.* 1998; Hamilton 1999; Lee & Wen 2004; Kress *et al.* 2005; Timme *et al.* 2007). Reaction volumes were 20 μL. Each reaction contained 20 ng of template DNA, 30 mM Tricine pH 8.4-KOH, 50 mM KCl, 2 mM MgCl, 100 μM each deoxynucleotide triphosphate, 0.1 μM forward primer, 0.1 μM reverse primer, and one unit of *Taq* DNA polymerase. Cycling conditions followed a touchdown protocol to reduce non-specific binding (Don *et al.* 1991), as follows: initial denaturation at 95°C for 3 min; followed by 10 cycles of 30 s at 94°C, 30 s at 65°C (annealing temperature was reduced by 1° per cycle), and 45 s at 72°C; followed by 30 cycles of 30 s at 94°C, 30 s at 55°C, and 45 s at 72°C; and a final extension time of 20 min at 72°C. To prepare for DNA sequencing, 10 μL of each PCR product was incubated at 37°C for 45 m with 0.8 units of Shrimp Alkaline Phosphatase and 4 units of Exo-nuclease I (USB, Cleveland, OH). Enzymes were then denatured by heating to 80°C for 15 min. Purified PCR products (0.5–2 μL depending on approximate concentration) were then sequenced with the primers used for the initial amplification. DyeNamic (Amersham) chemistry was used for the sequencing following the manufacturers' protocols with minor modifications.

Unincorporated dyes were removed from the sequencing reactions with Sephadex (Amersham) cleanup and sequences were resolved on a Basestation (MJ Research) automated DNA sequencer. Sequences were aligned using Sequencher 4.7 (Gene Codes Corp.), and putative species-specific sequence differences were identified. All sampled individuals were analyzed via PCR-RFLP of the *trn*L(UAA)5′exon-*trn*F(GAA) region, which exhibited a putative species-specific *Rsa*I

restriction difference between the *S. exigua* and *S. virgata* samples (see below). PCR conditions were as described above, and restriction digestions were conducted as suggested by the manufacturer (Promega Corp.). Restriction fragments were then separated by gel electrophoresis and visualized by staining with ethidium bromide.

Simple sequence repeat markers and genotyping

Simple sequence repeat (SSR) primers were designed from 288 SSR-bearing *Stephanomeria* expressed-sequence tags (N.A. Sherman & J.M. Burke, unpublished data). These primer pairs were then tested for amplification and polymorphism on DNA samples from eight individuals across the three species (three each from *S. exigua* and *S. virgata*, and two from *S. diegensis*). This resulted in the identification of seventeen SSR markers that amplified reliably across taxa, exhibited polymorphism, and were easily scorable (Table 1).

All seventeen loci were amplified using a modification of the three-primer PCR protocol outlined by Schuelke (2000; see Wills *et al.* 2005). PCR reaction volumes were 14 μL. Each reaction contained 10 ng of template DNA, 30 mM Tricine pH 8.4-KOH, 50 mM KCl, 2 mM MgCl, 100 μM each deoxynucleotide triphosphate, 0.02 μM forward primer [with an M13 -29 sequence tail (5′-CACGACGTTGTAAAACGACA-3′)], 0.1 μM reverse primer, 0.1 μM fluorescently-labelled M13 -29 primer, and one unit of *Taq* DNA polymerase. The fluorescent labels used were HEX and TET, and cycling conditions were as above. Amplicons were diluted 1:50 or 1:150 (depending on product intensity in the original screen) and visualized on an ABI 3730xl DNA sequencer (Applied Biosystems) with MapMarker 1000 ROX size standards (BioVentures) included in each lane to allow for accurate fragment size determination. Alleles were called using the software package GeneMarker v. 1.70 (SoftGenetics).

Table 1 Locus names, primer sequences, and repeat motifs for each SSR marker

Locus	Primer sequences (5′–3′)	Repeat Motif
Steph_0279	F: TGGTGAGTATGGTGGTGGTC R: TCTGCCATACCTGGTTCTCC	GAT
Steph_0227	F: CCCAGTTCGATTCCTCTTCC R: AAATTTAAACACGCGGATCG	GCA
Steph_0283	F: TTCATTCACGACTTTGATCATTC R: AATTCTGCACACCCATGTTG	CGG
Steph_0196	F: GACAACTGAGATGAACATTGA AGATGG R: ACAATCGGAGCCTTGAAATG	GAT
Steph_0094	F: ACCCGGATCAAACGAAATAC R: GCACACTCCACCCTATCTCC	GGC
Steph_0024	F: GGGAGGAGAGAGAGAAGAG AGAG R: TTCATCGTCAAATCCAGGTTC	GA
Steph_0103	F: GTCCACCACCCATGAACAAG R: AGCAACATACTCAAACCACA AAG	TC
Steph_0140	F: AATATTCACCCACGCTGACG R: TTGCACCGTGTGGTCTTTAG	CGA
Steph_0202	F: AGAACGGAGGAGGATGCAAG R: TTCATCAAGTTCAGTCGCTATC	TG
Steph_0004	F: ACAGAGGCATGTGGTTTTCC R: TTCAGTAGATGAAAATGGTT CAAAG	TAT
Steph_0226	F: CCCACTTGAAGAACCCTACC R: TAATGCTACCTGCGGAAACC	AAG
Steph_0237	F: GAGTAACCGTGCAGCATTCC R: GAATCTCCAGAGCAGCAACC	CAT
Steph_0231	F: CGTACCAATTTCCACCAACC R: ACGCATTCGTCTTCTTGGAG	TGG
Steph_0288	F: GCCTTGACCTTGTTCATGTG R: TCCGCCCATCAGTATATTCC	AAG
Steph_0078	F: CCGAGTTTCTGCAAATTTCTC R: CCTGGAGACACCTGAACTGG	GA
Steph_0195	F: AACCATGGAGAACGAGAACG R: CGACCGAATTTGCATAACAC	CAC
Steph_0072	F: ATGCAGGTGCTGCTACTGTG R: TTTCAAGATTTGGGCAGAATG	TG

Data analyses

Utilizing the SSR data, descriptive population genetic statistics were calculated for each taxon using GenAlEx v. 6.1 (Peakall & Smouse 2006). These values included percentage of polymorphic loci, mean number of alleles per locus, and gene diversity [calculated as Nei's (1978) unbiased expected heterozygosity; H_e]. Relationships amongst taxa were then graphically assessed via principal coordinate analysis (PCO; again using GenAlEx) using the covariance matrix with data standardization of genetic distance. Neighbour-joining trees were constructed in PHYLIP 3.67 (Felsenstein 2007) using a distance matrix from MSAnalyzer (Dieringer & Schlotterer 2003) based on Nei *et al.*'s (1983) genetic distance (calculated from the allele frequency data). Nodal support was calculated with 1000 bootstrap replicates using the CONSENSE program in PHYLIP 3.67 (Felsenstein 2007).

The hypothesis of a hybrid origin of *S. diegensis* was next investigated using the admixture model of the Bayesian clustering program STRUCTURE ver. 2.2 (Pritchard *et al.* 2000, Falush *et al.* 2003) following the approach of James & Abbott (2005). Given that the neighbour-joining results for *S. exigua* and *S. virgata* showed these species to be distinct (Fig. 2 and see below), the STRUCTURE analysis was performed with $K = 2$ clusters and individuals of the two parental

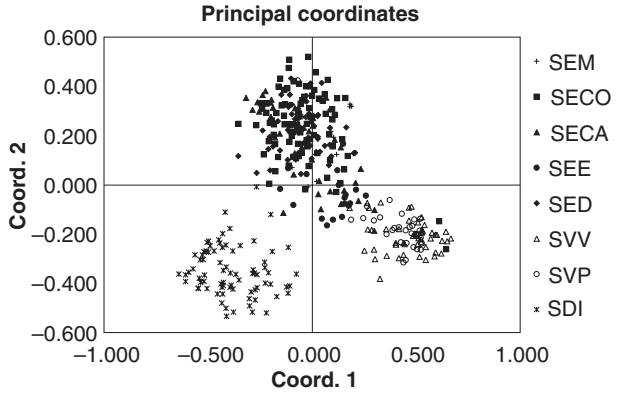

Fig. 2 Results of the principal coordinate (PCO) analysis of *S. exigua*, *S. virgata*, and *S. diegensis* individuals used in this study based on genotypic data from seventeen SSR loci.

species were treated as 'learning samples' with the USEPOPINFO feature turned on. For the purpose of this analysis, *S. diegensis* was treated as an unknown, and STRUCTURE was used to infer the ancestry of all individuals in the dataset. This analysis was performed assuming independent allele frequencies across populations with an initial burn-in period of 20 000 replicates and 50 000 MCMC iterations. This analysis was repeated and the results were found to be consistent across five runs.

The genomic composition of *S. diegensis* individuals was further investigated using a maximum-likelihood hybrid index score estimated using the program HINDEX (Buerkle 2005) following the methods of Gross *et al.* (2007). Briefly, each *S. diegensis* individual was assigned a hybrid index score, ranging from 0 (*S. virgata*-derived) to 1 (*S. exigua*-derived) based on its genotype and the allele frequencies of its putative parents. Because a hybrid species will be a stabilized mosaic of the two parental genotypes, one might expect a subset of the loci to be derived from one parent with the balance being derived from the other parent, rather than observing segregating variation at each locus. Hybrid index scores were calculated on a per-locus basis such that a putative parental origin could be assigned to each locus. Per-locus hybrid index scores were then averaged for each population and loci were considered to be *S. exigua*-derived if the value was greater than 0.60 or *S. virgata*-derived if the value was less than 0.40. These locus assignments were only made if they were consistent across six or more of the nine *S. diegensis* populations. To double-check the single locus results, hybrid indices were re-calculated for all *S. diegensis* individuals using either: (1) the putative *exigua*-derived loci, or (2) the putative *virgata*-derived loci. This allowed us to test whether or not the suites of loci assigned to each parent provided consistent results across all *S. diegensis* populations.

Finally, in order to investigate the particular parental subspecies that may have been involved in the origin of *S. diegensis*, neighbour-joining trees were constructed separately for *S. diegensis* along with each of its putative parents. These analyses utilized the subset of loci that were assigned to the parent of interest using HINDEX. In other words, *S. exigua* and *S. diegensis* were analyzed using the *S. exigua*-derived markers, whereas *S. virgata* and *S. diegensis* were analyzed using the *S. virgata*-derived markers. These trees were constructed as described above.

Results

Chloroplast DNA variation

Sequence analysis of the seven cpDNA regions revealed very low levels of polymorphism. Across a combined 5 kb of sequence, there were two polymorphic sites as well as five indels spanning 1–31 bp each. Overall, only two sites showed an apparent species-specific difference between *S. exigua* and *S. virgata*, and only one of these (in *trn*L(UAA)5′exon-*trn*F(GAA)) corresponded to a readily available restriction enzyme recognition site. Expanded genotyping of 374 individuals across multiple populations of all subspecies of both *S. exigua* and *S. virgata* confirmed the species-specific nature of this polymorphism, with 100% of *S. exigua* individuals harbouring an additional cut site in this region, and 0% of *S. virgata* individuals harbouring this cut site (Table 2). Genotyping of *S. diegensis* revealed that all 79 individuals across the nine sampled populations carried the *S. virgata* haplotype.

Nuclear variation

With the exception of *S. exigua* ssp. *macrocarpa*, all taxa exhibited polymorphism across all loci (Table 2). The mean number of alleles per locus in *S. exigua* ranged from a low of 3.1 across loci (range 1–6) in *S.exigua* ssp. *macrocarpa* to a high of 13.1 (range 7–24) in *S. exigua* ssp. *coronaria*. In *S. virgata*, there was an average of 9.1 (range 3–19) and 9.6 (range 4–15) alleles per locus in ssp. *virgata* and ssp. *pleurocarpa*, respectively. *Stephanomeria diegensis* exhibited an average of 11.5 (range 5–24) alleles per locus. Due to differences in sampling depth, however, variation in allele number must be interpreted with caution. In terms of overall levels of genetic diversity, H_e in *S. exigua* ranged from a low of 0.42 ± 0.07 (mean ± SE) in *S. exigua* ssp. *macrocarpa* to a high of 0.71 ± 0.03 in *S. exigua* ssp. *carotifera*. In *S. virgata*, the corresponding values were 0.68 ± 0.05 and 0.72 ± 0.05 in ssp. *virgata* and ssp. *pleurocarpa*, respectively. *Stephanomeria diegensis* had an average gene diversity of 0.64 ± 0.05.

Table 2 Summary of sample sizes and the results of the chloroplast/nuclear genotyping

	Subspecies Designation*	# of inds. (pops.)	% with cpDNA restriction site	Mean alleles per locus	Mean H_e†	% Polymorphic loci
S. exigua	SEM	13 (3)	100.0%	3.1 (0.4)	0.42 (0.07)	82.4%
	SECO	82 (10)	100.0%	13.1 (1.3)	0.69 (0.04)	100.0%
	SECA	41 (5)	100.0%	9.6 (1.1)	0.71 (0.03)	100.0%
	SEE	18 (2)	100.0%	6.0 (0.7)	0.57 (0.07)	100.0%
	SED	62 (6)	100.0%	12.7 (1.7)	0.69 (0.05)	100.0%
S. virgata	SVV	49 (8)	0.0%	9.6 (1.1)	0.68 (0.05)	100.0%
	SVP	30 (4)	0.0%	9.1 (0.9)	0.72 (0.05)	100.0%
S. diegensis	SDI	79 (9)	0.0%	11.5 (1.5)	0.64 (0.05)	100.0%
Grand Total		374 (47)				

*See Fig. 1 legend for definition of subspecies designations.
†Refers to Nei's (1978) unbiased expected heterozygosity averaged across loci.

The principal coordinate analysis (PCO) revealed that the three species (*S. exigua*, *S. virgata*, and *S. diegensis*) form relatively distinct groups (Fig. 2). The subspecies within each of the putative parental species, however, exhibit extensive overlap. Overall, *S. diegensis* was more similar to *S. exigua* along the first coordinate (which explains 30.8% of the variation) and more similar to *S. virgata* along the second coordinate (which explains 24.3% of the variation).

The neighbour-joining tree based on data from all seventeen loci revealed that *S. exigua* and *S. virgata* are genetically distinct from one another (with 53.5% bootstrap support; Fig. 3). When *S. diegensis* was added to this analysis, it formed a well-supported (99.9% bootstrap support), monophyletic group most closely related to the *S. exigua* ssp. *exigua* populations (its position is indicated by the star in Fig. 3). The addition of *S. diegensis* did not otherwise change the topology of the tree.

Consistent with the hypothesis of a hybrid origin, *S. diegensis* showed a signature of shared ancestry in the STRUCTURE analysis with the average *S. diegensis* individual exhibiting ca. 65% membership in the *S. exigua* group and ca. 35% membership in the *S. virgata* group (Fig. 4). These results were generally consistent across populations. However, there were individuals within both *S. exigua* and *S. virgata* that did not cluster true to their presumptive species type. Note that, while 'population' information was provided to STRUCTURE in the form of species identifications for individuals of *S. exigua* and *S. virgata*, this information was only used to train the algorithm. As such, these a priori designations can be (and in some cases were) overridden by the genetic data. Because the morphology and cpDNA profiles of these 'mis-assigned' parental individuals were all consistent with their species designation, the unexpected placement of these individuals most likely reflects either introgressive hybridization or the retention of ancestral polymorphism.

Mean hybrid index scores for *S. diegensis* populations (estimated with all seventeen loci) ranged from 0.67 to 1.0, with an average of 0.87 ± 0.03 (mean ± SE). Eleven loci were identified as putatively *S. exigua*-derived, four loci were putatively *S. virgata*-derived, and two loci could not be assigned to either parent. In one case, this was due to a paucity of shared alleles whereas, in the other case, the hybrid index value fell between 0.40 and 0.60. The pooled *S. exigua*-derived loci gave a per-population average hybrid index score for *S. diegensis* of 0.81–1 with an overall average value of 0.96 ± 0.02 across the species, and the pooled *S. virgata*-derived loci gave an average hybrid index score of 0.02–0.29 with an overall average value of 0.13 ± 0.03 across the species.

In the neighbour-joining trees constructed with the eleven *S. exigua*-derived loci, *S. diegensis* formed a monophyletic group with 97.8% bootstrap support (Fig. 5a). This group was most closely associated with populations of *S. exigua* ssp. *exigua* and *S. exigua* ssp. *deanei*, though the overall topology of the tree was not well supported. For the neighbour-joining trees constructed with the four *S. virgata*-derived loci, *S. diegensis* once again formed a monophyletic group with 80.8% bootstrap support (Fig. 5b). In this case, however, the most closely related subspecies could not be determined.

Discussion

Taken together, the results of this study point to a hybrid origin of *S. diegensis*. The SSR-based neighbour-joining tree shows a split between *S. exigua* and *S. virgata*, with *S. diegensis* clustering with *S. exigua* (Fig. 3, see star). The chloroplast data, however, clearly show that *S. diegensis* carries an *S. virgata*-like cpDNA haplotype. This pattern of non-concordance between nuclear and cytoplasmic data is one of the hallmarks of reticulate evolution (Arnold 1997) and provides strong support for the hypothesis of a hybrid origin of *S. diegensis*.

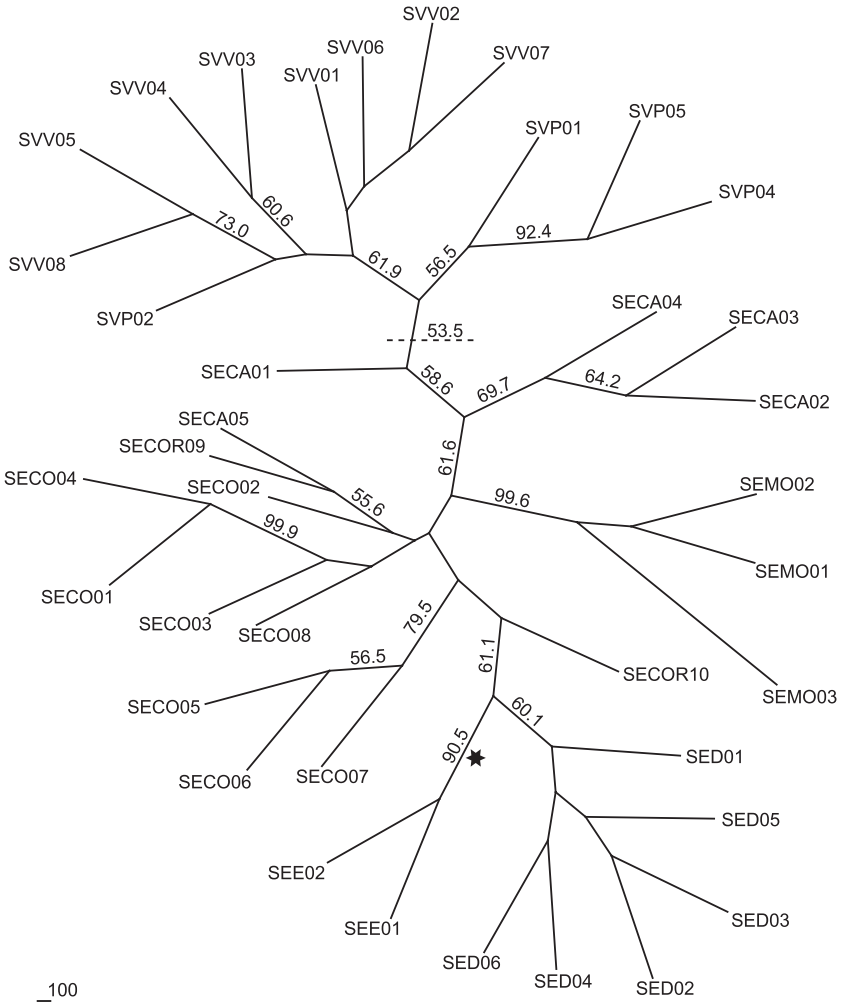

Fig. 3 Unrooted neighbour-joining dendrogram of *S. exigua* and *S. virgata* constructed using genotypic data from all seventeen SSR loci. The asterisk (*) indicates the location of the *S. diegensis* cluster when populations of this species are included. Numbers along branches represent bootstrap support after 1000 replicates. Only bootstrap values greater than 50 are shown.

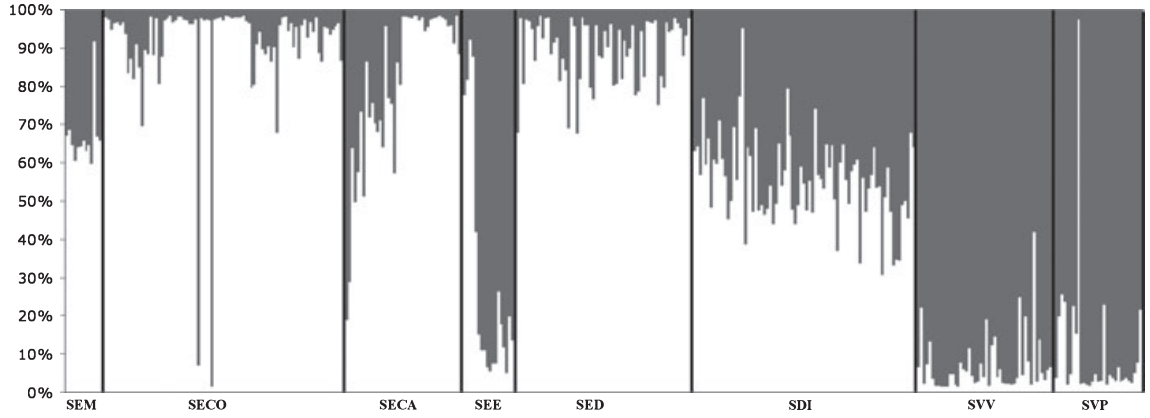

Fig. 4 Results of the STRUCTURE analysis with $K = 2$ corresponding to *S. exigua* and *S. virgata* and *S. diegensis* treated as unknowns. Bars for each individual reflect the average result across five independent runs.

Moreover, subsequent analyses (discussed in detail below) revealed that the *S. diegensis* nuclear genome is a mosaic of the *S. exigua* and *S. virgata* genomes, which is consistent with a hybrid origin, potentially via recombinational speciation. Finally, assuming maternal transmission, the cpDNA data indicate that *S. virgata* likely served as the seed parent in the initial hybridization event(s).

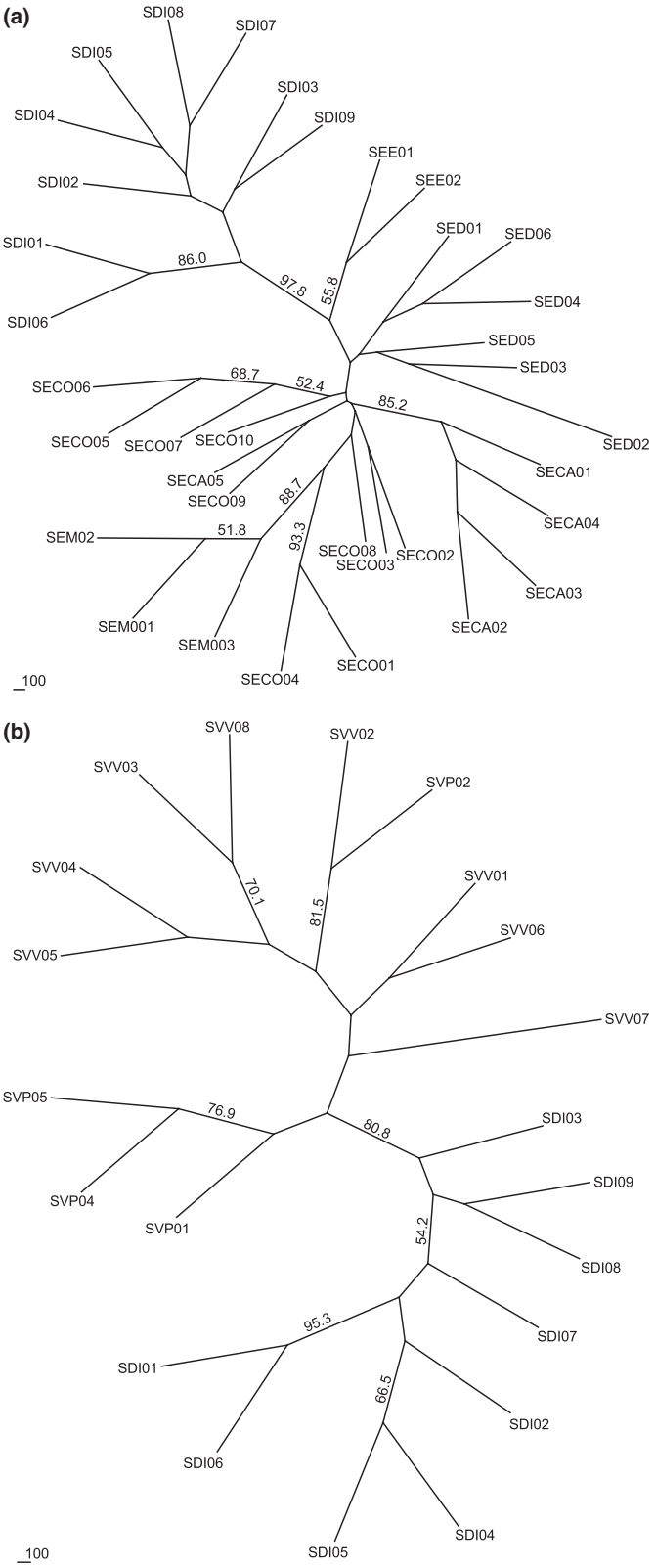

Fig. 5 Unrooted neighbour-joining dendogram of (a) *S. exigua* and *S. diegensis* using genotypic data from the eleven *S. exigua*-derived SSR loci, and (b) *S. virgata* and *S. diegensis* using genotypic data from the four *S. virgata*-derived SSR loci. Numbers along branches represent bootstrap support based on 1000 replicates. Only bootstrap values over 50 are shown.

In terms of genomic composition, the PCO plot revealed that *S. diegensis* is most similar to *S. exigua* along the first coordinate, and to *S. virgata* along the second coordinate, suggesting a stronger affinity for the former as compared to the latter (Fig. 2). Consistent with this finding, the STRUCTURE analysis (Fig. 4) revealed that *S. diegensis* individuals exhibit ca. 65% identity with *S. exigua* and ca. 35% identity with *S. virgata*, and HINDEX classified 11 of 17 of the SSR loci (64.7%) as being *S. exigua*-derived and 4 of the 17 SSR loci (23.5%) as *S. virgata*-derived. These results were found to be largely consistent across populations, suggesting that the *S. diegensis* populations surveyed trace back to a common origin. Interestingly, despite showing a stronger affinity for *S. exigua* in terms of nuclear genome composition, *S. diegensis* exhibits the *S. virgata* karyotype for the two chromosomes (out of a haploid number of $n = 8$) that can be visibly distinguished based on banding differences between *S. exigua* and *S. virgata* (Gottlieb 1971). Overall, these data accord well with the earlier suggestions that *S. diegensis* is a homoploid hybrid species (Gottlieb 1971; Gallez & Gottlieb 1982), and further suggest that its genome is disproportionately derived from *S. exigua*. It is noteworthy that Gallez & Gottlieb (1982) also found *S. diegensis* to be more closely allied with *S. exigua* based on allozyme data.

Beyond providing evidence of a hybrid origin, our data are suggestive of a single origin of *S. diegensis*. All *S. diegensis* individuals exhibited similar genomic composition based on the STRUCTURE analysis regardless of their population of origin, and all also carried the *S. virgata* chloroplast haplotype. Moreover, *S. diegensis* formed a single cluster in all of the neighbour-joining analyses, including those in which the markers were subdivided by presumptive species of origin (Figs 3 and 5). We cannot, however, entirely rule out the possibility of multiple origins, particularly if the same parental subspecies were involved in each case. While the HINDEX results were, as noted above, largely consistent across populations, there were a small number of instances in which locus assignments differed between populations. While this finding might be superficially consistent with multiple origins, there was no apparent pattern in terms of the population/locus combinations that gave conflicting results. It is thus more likely that these inconsistencies reflect the challenges associated with reliably assigning individual loci to a particular parental species (Gross *et al.* 2003, 2007). While the lack of cpDNA polymorphism limited our ability to use those data to test for multiple origins, as has previously been done in studies of hybrid speciation in *Helianthus* (Schwarzbach & Rieseberg 2002; Welch & Rieseberg 2002; and Gross *et al.* 2003), the balance of our data were fully consistent with a single origin with *S. virgata* having served as the seed parent.

It appears from the hybrid speciation literature that instances of singly-derived, homoploid hybrid plant species are relatively rare. This conclusion is, however, based on a rather small number of well-documented cases of homoploid hybrid speciation under natural conditions. In four of these five cases (*Helianthus anomalus*, *H. deserticola*, *Pinus densata*, and *Argyranthemum sundingii*) the hybrid species was found to have been multiply-derived (Brochmann *et al.* 2000; Wang *et al.* 2001; Schwarzbach & Rieseberg 2002; Gross *et al.* 2003, 2007; Song *et al.* 2003). In the fifth case, *Helianthus paradoxus* was found to trace back to a single origin, perhaps because genetic constraints related to its adaptation to a unique salt marsh habitat have limited the potential for multiple origins (Welch & Rieseberg 2002). While *Senecio squalidus* has likewise been found to have had a single origin, the stabilization of this species appears to have required long distance, human-mediated dispersal of hybrids from Italy to the British Isles (James & Abbott 2005).

Given the apparent tendency for homoploid hybrid species to be multiply-derived, the observation that *S. diegensis* likely traces back to a single origin requires explanation. One factor that might reduce the likelihood of multiple origins is a lack of opportunity for hybrid speciation. However, *S. exigua* and *S. virgata* overlap broadly, and hybridization occurs throughout the region of overlap (Gottlieb 1971; Sherman pers. obs.), making this an unlikely explanation. Alternatively, as noted above, Welch & Rieseberg (2002) have argued that genetic constraints related to the adaptation of a hybrid neo-species to an extreme habitat might decrease the likelihood of multiple origins. However, the habitat preferences of *S. diegensis* appear to overlap with those of its parents, which are likewise relatively similar to each other (Gottlieb 2006; N. Sherman, pers. obs.). Conversely, it has been argued that ecological differentiation resulting in spatial isolation dramatically increases the likelihood of homoploid hybrid speciation (Buerkle *et al.* 2000). Perhaps the relative lack of ecological divergence between these species has limited the potential for the stabilization of multiple hybrid lineages in this case.

In terms of the parentage of *S. diegensis*, the *S. exigua*-derived portion of the genome appears most closely related to either *S. exigua* ssp. *exigua* (which, like *S. exigua* ssp. *macrocarpa*, shows evidence of possible past admixture based on the STRUCTURE analysis) or *S. exigua* ssp. *deanei* (Fig. 5a). In concordance with our findings, sequence analysis of the ITS region has revealed that *S. diegensis* is most similar to *S. exigua* ssp. *deanei* and *S. exigua* ssp. *exigua* (Lee *et al.* 2002), and a recent analysis of sequence diversity at the *PgiC* locus places *S. diegensis*

with *S. exigua* ssp. *deanei*, to the exclusion of all other subspecies of *S. exigua* (Ford *et al.* 2006). With regard to the *S. virgata* parental subspecies, there was insufficient resolution to make any clear inferences. This may be due, at least in part, to the low number of *S. virgata*-derived loci in our data. Furthermore, *S. virgata* ssp. *virgata* and *S. virgata* ssp. *pleurocarpa* are very similar to each other, perhaps due to relatively weak isolation between these subspecies as compared to the barriers between *S. exigua* subspecies (Gottlieb 1969, 1971). It is also possible that the formation of *S. diegensis* pre-dated the divergence of the subspecies within *S. exigua* and *S. virgata*.

Conclusions and future directions

Our data provide strong support for a hybrid origin of *S. diegensis*. The nuclear genome of this species is a mosaic of the *S. exigua* and *S. virgata* genomes, with a greater proportion derived from the former as compared to the latter. Despite this closer alliance with *S. exigua*, *S. diegensis* is karyotypically more similar to *S. virgata*, and all individuals also carry the *S. virgata* chloroplast haplotype. When the nuclear and chloroplast results are combined, the picture that emerges is one in which pollen flow from *S. exigua* to *S. virgata* gave rise to the hybrid neospecies, and that subsequent backcrossing likewise involved pollen flow from *S. exigua*. The fact that *S. diegensis* appears to be singly-derived places it in the minority of homoploid hybrid species that have been analyzed to date. While it has been suggested that ecological divergence plays a major role in determining the likelihood of homoploid hybrid speciation, the role of ecological divergence in promoting or limiting hybrid speciation in *Stephanomeria* requires further investigation. A better understanding of the timing of the origin of *S. diegensis*, as well as of the divergence of subspecies within each of the parental species, will also provide important insights into the origin and parentage of this hybrid species.

Acknowledgements

The authors thank Leslie Gottlieb for general guidance and assistance with taxonomy in this group; Jason Strever and Eli Noblitt for assistance in the greenhouse and lab; Dirk Rodriguez, Sarah Chaney, Marilee Sherman, Chris Grantham for assistance in the field; and Wendy Zomlefer, Mark Chapman, Jennifer Dechaine, John Havla, Stepanie Pearl, and Evan Staton for comments on the manuscript.

References

Arnold ML (1993) *Iris nelsonii* (Iridaceae): origin and genetic composition of a homoploid hybrid species. *American Journal of Botany*, **80**, 577–583.

Arnold ML (1997) *Natural Hybridization and Evolution*. Oxford University Press, New York.

Arnold ML (2004) Transfer and origin of adaptations through natural hybridization: were Anderson and Stebbins right? *The Plant Cell*, **16**, 562–570.

Bailey JK, Schweitzer JA, Ubeda F *et al.* (2009) From genes to ecosystems: a synthesis of the effects of plant genetic factors across levels of organization. *Philosophical Transactions of the Royal Society B-Biological*, **364**, 1607–1616.

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.

Buerkle CA (2005) Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes*, **5**, 684–687.

Buerkle CA, Morris RJ, Asmussen MA, Rieseberg LH (2000) The likelihood of homoploid hybrid speciation. *Heredity*, **84**, 441–451.

Buerkle CA, Wolf DE, Rieseberg LH (2003) The origin and extinction of species through hybridization. *Ecological Studies*, **165**, 117–144.

Chiang TY, Schaal BA, Peng CI (1998) Universal Primers for amplification and sequencing a non-coding spacer between the atpB and rbcL genes of chloroplast DNA. *Botanical bulletin of Academia Sinica*, **39**, 245–250.

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: Massive introgression by local genes. *Evolution*, **62**, 1908–1920.

Dieringer D, Schlotterer C (2003) Microsatellite Analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, **3**, 167–169.

Don R, Cox P, Wainwright B, Baker K, Mattick J (1991) Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research*, **19**, 4008.

Doyle JJ, Doyle JL (1990) Isolation from plant DNA from fresh tissue. *Focus*, **12**, 13–15.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Felsenstein J (2007) *PHYLIP (Phylogeny Inference Package) Version 3.67*. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle. http://evolution.gs.washington.edu/phylip.html.

Ford VS, Lee J, Baldwin BG, Gottlieb LD (2006) Species divergence and relationships in Stephanomeria (Compositae): PgiC phylogeny compared to prior biosystematic studies. *American Journal of Botany*, **93**, 480–490.

Gagnaire PA, Albert V, Jonsson B, Bernatchez L (2009) Natural selection influences AFLP intraspecific genetic variability and introgression patterns in Atlantic eels. *Molecular Ecology*, **18**, 1678–1691.

Gallez GP, Gottlieb LD (1982) Genetic evidence for the hybrid origin of the diploid plant *Stephanomeria digenesis*. *Evolution*, **36**, 115–1167.

Gottlieb LD (1969) *The Role of Hybridization in the Annual Species of Stephanomeria (Compositae)*. Thesis. The University of Michigan, Ann Arbor.

Gottlieb LD (1971) Evolutionary relationships in the outcrossing diploid annual species of *Stephanomeria* (Compositae). *Evolution*, **25**, 312–329.

Gottlieb LD (2006) Stephanomeria. In: Flora of North America Editorial Committee ed. 1993+. Flora of North America North of Mexico. 12+ vols. New York and Oxford. Volumes 19, 20, 21.

Grant V (1958) The regulation of recombination in plants. *Cold Spring Harbor Symposium on Quantitative Biology*, **23**, 337–363.

Grant V (1963) *The Origin of Adaptations*, Columbia University Press, New York.

Grant V (1981) *Plant Speciation*, 2nd Edn. Columbia University Press, New York.

Gross BL, Rieseberg LH (2005) Ecological genetics of homoploid hybrid speciation. *Journal of Heredity*, **96**, 241–252.

Gross BL, Schwarzbach AE, Rieseberg LH (2003) Origin(s) of the diploid hybrid species *Helianthus deserticola* (Asteraceae). *American Journal of Botany*, **90**, 1708–1719.

Gross BL, Turner KG, Rieseberg LH (2007) Selective sweeps in the homoploid hybrid species *Helianthus deserticola*: evolution in concert across populations and across origins. *Molecular Ecology*, **16**, 5246–5258.

Hamilton MB (1999) Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology*, **8**, 521–523.

Howard DJ (1993) Reinforcement: origin, dynamics, and fate of an evolutionary hypothesis. In:*Hybrid Zones and the Evolutionary Proces* (ed. Harrison RG). pp. 46–69, Oxford University Press, New York.

James JA, 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*, **59**, 2533–2547.

Jeffrey C (1993) Asterales. In: *Flowering Plants of the World*, updated edition. (ed. Heywood VH), pp 263–268. Oxford University Press, New York.

Johnson LA, Soltis DE (1995) Phylogenetic inference on Saxifragaceae susu stricto and Gilia (Polemoniaceae) using matK sequences. *Annals of the Missouri Botanical Garden*, **82**, 149–175.

Kawakami T, Butlin RK, Adams M, Paull DJ, Cooper SJB (2009) Genetic analysis of a chromosomal hybrid zone in the Australian Morabine grasshopper (Vandiemenella, Viatica species group). *Evolution*, **63**, 139–152.

Kim M, Cui ML, Cubas P *et al.* (2008) Regulatory genes control a key morphological and ecological trait transferred between species. *Science*, **322**, 1116–1119.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the USA*, **102**, 8369–8374.

Lai Z, Nakazato T, Salmaso M *et al.* (2005) Extensive chromosomal repatterning and the evolution of sterlility barriers in hybrid sunflower species. *Genetics*, **177**, 291–303.

Lee C, Wen J (2004) Phylogeny of *Panax* using chloroplast *trn*C-*trn*D intergenic region and the utility of *trn*C-*trn*D in interspecific studies of plants. *Molecular Phylogentics and Evolution*, **31**, 894–903.

Lee J, Baldwin BG, Gottlieb LD (2002) Phylogeny of *Stephanomeria* and related genera (Compositae-Lactuceae) based on analysis of 18S-26S nuclear rDNA ITS and ETS sequences. *American Journal of Botany*, **89**, 160–168.

Lepais O, Petit RJ, Guichoux E *et al.* (2009) Species relative abundance and direction of introgression in oaks. *Molecular Ecology*, **18**, 2228–2242.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, **20**, 229–237.

Mayr E (1942) *Systematics and the Origin of Species*, Harvard University Press, Boston, USA.

McCarthy EM, Asmussen MA, Anderson WW (1995) A theoretical assessment of recombinational speciation. *Heredity*, **74**, 502–509.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution*, **19**, 153–170.

Otto SP, Whitton J (2000) Polyploid incidence and evolution. *Annual Review of Genetics*, **34**, 401–437.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–956.

Ramsey J, Schemske D (1998) Pathways, mechanisms, and rates of polyploidy formation in flowering plants. *Annual Review of Ecology and Systematics*, **29**, 467–501.

Rieseberg LH (1997) Hybrid origins of plant species. *Annual Review of Ecology and Systematics*, **27**, 259–389.

Rieseberg LH (2000) Crossing relationships among ancient and experimental sunflower hybrid lineages. *Evolution*, **54**, 859–865.

Rieseberg LH, Wendel JF (1993) Introgression and its consequences in plants. In:*Hybrid Zones and the Evolutionary Process* (ed. Harrison RG). pp. 70–109, Oxford University Press, New York.

Rieseberg LH, Raymond O, Rosenthal DM *et al.* (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, **301**, 1211–1216.

Rogstad SH (1992) Saturated NaCl-CTAB solution as a means of field preservation of leaves for DNA analyses. *Taxon*, **41**, 701–708.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, **18**, 233–234.

Schwarzbach AE, Rieseberg LH (2002) Likely multiple origins of a diploid hybrid sunflower species. *Molecular Ecology*, **11**, 1703–1715.

Servedio MR, Noor MAF (2003) The role of reinforcement in speciation: theory and data. *Annual Review of Ecology Evolution and Systematics*, **34**, 339–364.

Song BH, Wang XQ, Wang XR, Ding KY, Hong DY (2003) Cytoplasmic composition in *Pinus densata* and population establishments of the diploid hybrid pine. *Molecular Ecology*, **12**, 2995–3001.

Stebbins GL (1957) The hybrid origin of microspecies in the *Elymus glaucus* complex. *Cytologia Supplement*, **36**, 336–340.

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, **17**, 1105–1109.

Timme RE, Kuehl JV, Boore JL, Jansen RK (2007) A comparative analysis of the Lactuca and Helianthus

(Asteraceae) plastid genomes: Identification of divergent regions and categorization of shared repeats. *American Journal of Botany*, **94**, 302–312.

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.

Welch ME, Rieseberg LH (2002) Patterns of genetic variation suggest a single, ancient origin for the diploid hybrid species *Helianthus paradoxus*. *Evolution*, **56**, 2126–2137.

Wills DM, Hester ML, Liu A, Burke JM (2005) Chloroplast SSR polymorphisms in the Compositae and the mode of organellar inheritance in *Helianthus annuus*. *Theoretical and Applied Genetics*, **110**, 941–947.

Wolf DE, Takebayashi N, Rieseberg LH (2001) Prediction the risk of extinction through hybridization. *Conservation Biology*, **15**, 1036–1053.

Wood T E, Nakazato T (2009) Investigating species boundries in the Giliopsis group of *Ipomopsis* (Polymoniaceae): string discordance among molecular and morphological markers. *American Journal of Botany*, **96**, 853–861.

NAS is a plant population geneticist whose interests focus on speciation and hybridization. JMB is an evolutionary geneticist whose research interests focus on crop domestication and the evolution of reproductive isolation in plants. This work is part of NAS's dissertation research at the University of Georgia.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 List of population locations and sampling

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.