Influences of gene flow on adaptive speciation in the *Dubautia arborea – D. ciliolata* complex

By: David L. Remington and R. H. Robichaux

This is the peer reviewed version of the following article:

Remington, D.L., and R.H. Robichaux. 2007. Influences of gene flow on adaptive speciation in the *Dubautia arborea – D. ciliolata* complex. Molecular Ecology 16: 4014-4027.

which has been published in final form at https://doi.org/10.1111/j.1365-294X.2007.03447.x.

Abstract:

Mechanisms of reproductive isolation during plant speciation are often unclear because distinct species often experience high levels of gene flow and hybridization. Adaptive radiations such as the Hawaiian silversword alliance (HSA) provide unique opportunities to study the interactions of selection, gene flow and isolating mechanisms during the speciation process. We examined patterns of phenotypic and genetic differentiation in *Dubautia arborea* and *Dubautia ciliolata*, two parapatric HSA taxa that show marked morphological divergence but evidence of weak molecular differentiation, in order to estimate genome-wide differentiation and gene flow patterns. We scored 166 amplified fragment length polymorphism markers in a set of 89 plants from two populations each of *D. arborea* and *D. ciliolata* and phenotypically *D. arborea*-like and *D. ciliolata*-like plants from a natural hybrid zone. Analyses of population subdivision showed low levels of differentiation between the two species (F_{ST} = 0.089) and evidence that the phenotypically parental hybrid zone plants were largely of parental species rather than of hybrid origin. A Bayesian analysis of population ancestry identified a number of plants with admixed *D. arborea* and *D. ciliolata* ancestry, even in nonhybrid-zone populations. These results suggest that genome-wide low levels of differentiation between *D. arborea* and *D. ciliolata*are in part due to gene flow, and favour models of genic speciation and collective evolution in which gene flow has different effects on selected loci vs. nonselected genomic regions. We discuss ecological and climatic factors that may have shaped patterns of differentiation in this species complex.

Keywords: adaptive radiation | AFLP | Dubautia | gene flow | Hawaiian silversword alliance | speciation

Article:

Introduction

Understanding the evolutionary mechanisms responsible for speciation is a major topic in evolutionary biology. Key questions include the relative importance of genetic drift and selective agents in speciation, the extent and mechanisms of reproductive isolation, and the role of hybridization in either impeding or creating opportunities for speciation (Rieseberg *et al*. 2000; Schluter 2001; Turelli *et al*. 2001; Seehausen 2004). Ecological speciation, in which divergent selection in contrasting environments leads to reproductive isolation and speciation, has been assumed to play a prominent role in speciation but has often been difficult to distinguish empirically from nonadaptive explanations such as genetic drift in conjunction with ecological opportunity (Schluter 2000, 2001; Turelli *et al*. 2001). A number of possible mechanisms must be explored to dissect the dynamics of ecological speciation, including selection on various traits that may be adaptive in a given situation, various modes of sexual selection, reinforcement and other forms of character displacement, and direct selection for reproductive isolation (Schluter 2001).

Reproductive isolation has long been recognized as an essential component of speciation (Mayr 1942; Dobzhansky 1946; Dobzhansky 1951), and dissecting the nature of prezygotic and postzygotic barriers to hybridization is critical to understanding the speciation process (Rieseberg *et al*. 2000; Schluter 2001; Turelli *et al*. 2001). In plants, however, it is common for distinct species to be highly interfertile, and the processes that create and maintain species integrity in the face of gene flow are often unclear (Judd *et al*. 2002). Speciation may be especially dependent on adaptive evolution of morphological and physiological characters in plants lacking specialized pollinators, in which opportunities for isolation based on reproductive traits are relatively limited (Grant 1949; Schluter 2000). Moreover, hybridization may generate new avenues for speciation if the resulting novel allelic combinations provide increased fitness in new environments (Seehausen 2004). This process is facilitated in plants by the frequent occurrence of allopolyploidy, but instances of homoploid hybrid speciation are also well documented (Rieseberg 1997; Ungerer *et al*. 1998; Wolfe *et al*. 1998; Ferguson & Sang 2001). Adaptation of local populations to a novel environment is one factor that may lead to progenitor-derivative species relationships, evidenced in part by reduced genetic diversity in the derivative species relative to the progenitor (Gottlieb *et al*. 1985; Witter 1990; Gottlieb 2003; Baldwin 2005). In some instances, there is evidence that a small number of genes is responsible for major phenotypic modification in the derivative species, allowing selection to maintain species integrity in the face of ongoing gene flow (Rieseberg & Burke 2001; Gottlieb 2003).

Adaptive radiations, or the rapid multiplication of species and adaptive phenotypic differences within a single lineage, provide unique opportunities to investigate the interplay of ecological and demographic processes in speciation. The extent to which adaptive radiation has shaped the extant taxonomic and ecological diversity of the Earth's species is unknown, but mounting evidence suggests it has played a prominent role (Schluter 2000). In adaptively radiating complexes, it is frequently possible to find numerous divergent morphological forms at various stages of speciation, providing a natural laboratory for investigating the underlying processes. Genetic and molecular evolutionary studies in adaptively radiating lineages have provided insights on the evolution of molecular armour variation in threespine sticklebacks (Peichel *et al*. 2001; Shapiro *et al*. 2004; Colosimo *et al*. 2005), coloration and body form in East African cichlid fishes (Albertson *et al*. 2003; Streelman *et al*. 2003), and woodiness in Macaronesian *Pericallis* spp. (Panero *et al*. 1999), to name just a few examples. The Hawaiian

silversword alliance is one of the most spectacular examples of adaptive radiations in plants (Schluter 2000). The silversword alliance includes some 30 species that have diversified from a common ancestor introduced to the Hawaiian archipelago approximately 5 million years ago (Ma), encompassing a wide variety of growth forms and ecological adaptations (Robichaux *et al*. 1990; Baldwin & Robichaux 1995; Baldwin & Sanderson 1998). Closely related species within the silversword alliance frequently show few barriers to fertility, and even the most divergent members of the group can hybridize to some extent in spite of multiple chromosomal rearrangements (Carr & Kyhos 1986; Carr 2003). Progenitor-derivative relationships have been inferred for several species pairs in the silversword alliance based on allozyme data (Witter 1990).

Dubautia arborea and *Dubautia ciliolata* represent a species complex in the Hawaiian silversword alliance that provides unique opportunities to examine the dynamics and genetic basis of speciation. They are the only *Dubautia* spp. endemic to the Island of Hawaii, the most recent island in the Hawaiian archipelago which first formed ~0.6 Ma (Carr 1985). *D. arborea* occurs largely in parapatry with one subspecies of *D. ciliolata* (ssp. *glutinosa*), with the former mainly occurring in high-altitude dry woodlands on Mauna Kea, and the latter found commonly in a band of shrubland distinct from the woodlands containing *D. arborea*(Carr 1985; Robichaux *et al*. 1990). *D. arborea* and *D. ciliolata* are sharply divergent morphologically. *D. arborea* is a large shrub or tree up to 6 m tall, with elliptic leaves 30–90 mm long and 8–24 mm wide. By contrast, *D. ciliolata* ssp. *glutinosa* is a low and slow-growing shrub less than 1.8 m tall, with small lanceolate leaves 7–30 mm long and 2–6.5 mm wide (Carr 1985; Robichaux *et al*. 1990; Lawton-Rauh *et al*. 2007). Small narrow leaves and slow growth rates often represent adaptations to droughty environments such as the shrublands in which *D. ciliolata* is found (Reich *et al*. 1997; Ackerly & Reich 1999). Subspecies *ciliolata* of *D. ciliolata*, which occurs largely on newer Mauna Loa, Kilauea, and Hualalai substrates and generally at some distance from existing *D. arborea* populations, tends to have a less compact form than *D. ciliolata* ssp. *glutinosa*.

In spite of their sharp morphological divergence, *D. arborea* and *D. ciliolata* have been found to show weak molecular and cytogenetic differentiation. The two species are fully interfertile and lack cytogenetically detectable chromosome rearrangements (Carr & Kyhos 1986). Opportunities for gene flow occur in places along the woodland–shrubland ecotone, especially in gulches where microtopography creates a mosaic of woodland, shrubland and intermediate environments. A spectacular hybrid zone occurs in Waipahoehoe Gulch with hundreds of plants displaying a full range of *D. arborea*-like, *D. ciliolata*-like and various intermediate forms (Carr & Kyhos 1981; Carr 1985; 2003). The core hybrid zone is bordered immediately to the north and west by a population with *D. ciliolata* phenotypes, and immediately to the south and east by a population with *D. arborea* phenotypes, which may represent the source populations for the hybrid zone. DNA sequence analyses of several nuclear genes in *D. arborea* and *D. ciliolata* ssp. *glutinosa* reveal extensive sharing of nucleotide polymorphisms, and the two taxa are largely interspersed in haplotype trees or networks (Remington, unpublished data; Lawton-Rauh *et al*. 2007). The coefficient of genetic identity between *D. arborea* and *D. ciliolata* based on 10 allozyme loci was 0.944 (Witter & Carr 1988). Analyses using seven microsatellite loci show the average differentiation (F_{ST}) between populations of *D. arborea* and *D. ciliolata* to be no greater than that between populations within *D. ciliolata* (Friar *et al*. 2007). If the low level of

molecular differentiation were genome–wide, then association methods normally used for evaluating candidate genes for within-species phenotypic variation could be useful for interspecific analyses in the *D. arborea–D. ciliolata* complex if appropriate controls for population structure were applied (Pritchard *et al*. 2000b; Thornsberry *et al*. 2001).

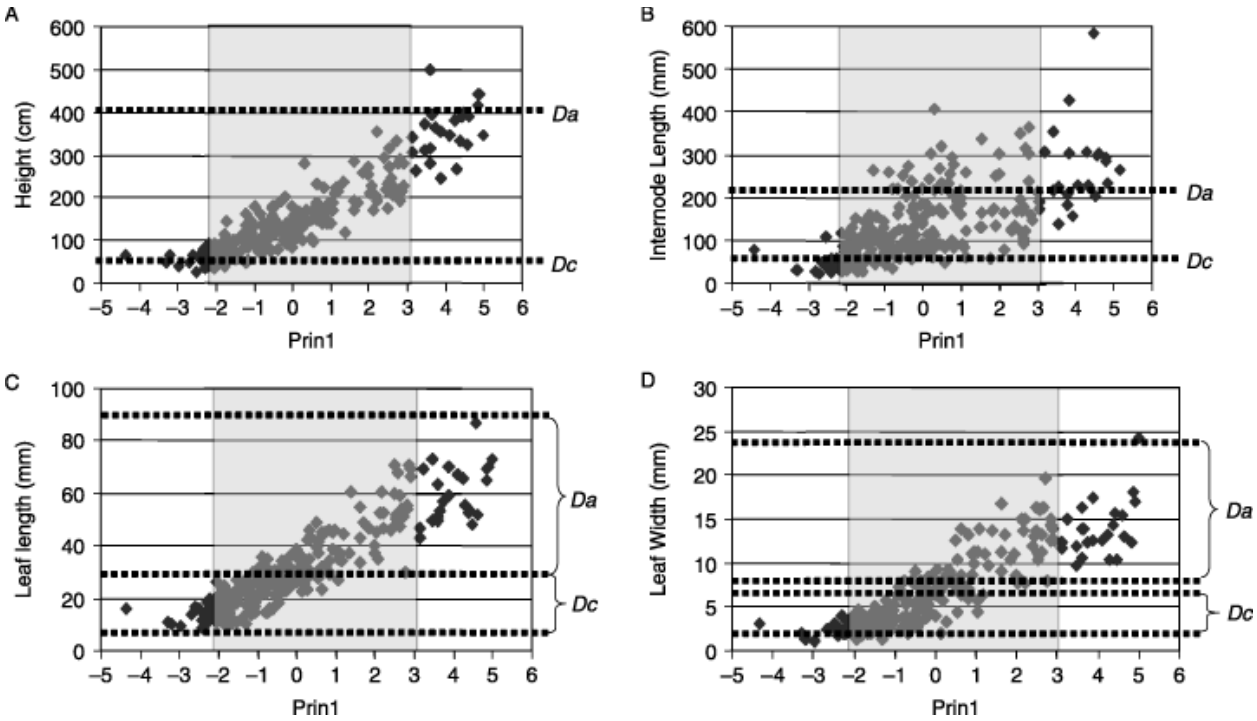

Figure 1. Plots of (A) plant height (B) length of 20 terminal internodes (C) average leaf length, and (D) average leaf width for 250 Waipahoehoe Gulch hybrid zone plants, plotted against the first principal component of six phenotypic traits (Prin1). Individuals to the left and right of the shaded box in each plot represent the *Dubautia ciliolata*-like and *Dubautia arborea*-like phenotypes, respectively, included in this study. Dotted horizontal lines in (A) and (B) show the mean trait values for a sample of *D. ciliolata* (*Dc*) from Pu'u Kanakaleonui and *D. arborea* (*Da*) from Pu'u La'au populations. Dotted horizontal lines in (C) and (D) show the range in trait values in *D. arborea* and *D. ciliolata* ssp. *glutinosa* from Carr (1985).

In a recent study, the core portion of the Waipahoehoe Gulch hybrid zone was exhaustively sampled for a total of 250 reproductively mature plants, which were scored for leaf length and width, height, crown width, internode lengths, and branchiness (Kirchoff *et al*. 2004). The first principal component of the six traits, which explained 63% of the total trait variation, was strongly correlated with a visual numerical rating of the degree of *D. arborea*-like vs. *D. ciliolata*-like appearance (Kirchoff *et al*. 2004). The ~10% of plants on each extreme of the first principal component distribution were rated as primarily *D. arborea*-like and *D. ciliolata*-like in appearance, and their trait distributions coincided closely with those of other *D. arborea* and *D. ciliolata* populations as described in the literature (Carr 1985) or estimated from our own sampling (Fig. 1). This observation leads to the intriguing hypothesis that the phenotypically extreme individuals in the core hybrid zone might be genetically recombinant hybrids in which the genotypes at key morphological loci differentiating the two species have been reconstituted. If this were true, the core hybrid zone might provide a valuable resource for isolating genomic

regions containing important speciation genes (Rieseberg *et al*. 1999, 2000; Rieseberg & Buerkle 2002)

In this study, we use data from 166 amplified fragment length polymorphism (AFLP) marker loci (Vos *et al*. 1995) in the *D. arborea–D. ciliolata* complex to investigate the dynamics of adaptive speciation and subsequent hybridization. Our objectives were fourfold. First, we wanted to test the hypothesis that the previously reported low level of differentiation in small sets of genetic loci represents a genome-wide phenomenon and that interspecific sharing of polymorphisms is extensive. This was tested by examining the level of molecular differentiation (F_{ST}) between the two species using the AFLP data and the extent to which AFLP polymorphisms are shared between the two species. Second, we wanted to test whether the relative levels of marker diversity in the two species are more consistent with divergence of sister species from a common ancestor or a progenitor-derivative mode of speciation. A recently formed progenitor-derivative pair would be expected to show reduced diversity and greater marker allele fixation in the derivative species, while diverged sister species would be expected to show similar levels of diversity in the absence of severe subsequent bottlenecks.

Third, we wanted to test the hypothesis that the phenotypically parental plants in the core Waipahoehoe Gulch hybrid zone are of hybrid origin. If this hypothesis is correct, there should be little genetic differentiation between the parental phenotypic classes (i.e. F_{ST}nearly zero). Under this scenario, loci that do exhibit differentiation between *D. arborea*-like and *D. ciliolata*-like phenotypes are likely to be linked to the genes actually responsible for the distinct phenotypic differences between the species. Alternatively, if phenotypes largely reflect the degree of *D. arborea* vs. *D. ciliolata* ancestry of each individual, F_{ST} values between the contrasting extremes would be close to the level occurring between nonhybrid-zone populations of the two species.

Finally, we wanted to examine whether there are effective barriers to gene flow between *D. arborea* and *D. ciliolata*. While individuals with hybrid features are occasionally found along gulches bridging the woodland–shrubland ecotone outside of the Waipahoehoe Gulch area, it is not clear that they contribute significantly to evolution of the species complex and thus to the extant population genetic structure. Bayesian clustering techniques that assign individuals to ancestral populations based on patterns of Hardy–Weinberg and/or linkage disequilibrium (Pritchard *et al*. 2000a; Corander & Marttinen 2006; Corander *et al*. 2006) provide a means of modelling the putative ancestry of individuals. We predicted that effective reproductive isolation subsequent to species divergence would result in strong support for two primary clusters that correspond to the two species, with little evidence of admixture. Further subdivision of the two main populations may in turn be supported to the extent that gene flow between conspecific populations in different localities is restricted. By contrast, we predicted that gene flow subsequent to speciation would result in evidence for a substantial number of individuals outside the hybrid zone that show interspecific admixture.

Materials and methods

Study populations

Leaf tissue was collected in 1998 and 2000 from *Dubautia arborea* populations at Pu'u La'au (DaL) and Pu'u Mali (DaM), from *Dubautia ciliolata* ssp. *glutinosa* populations at Pu'u Kaiwi'iwi (DcW) and Pu'u Kanakaleonui (DcK), and from plants in a zone of hybridization between *D. arborea* and *D. ciliolata* at Waipahoehoe Gulch (WHa and WHc). The DaL, DaM, DcW, and DcK populations were selected to sample genetic variation in the two species. These populations are relatively large and accessible, and were considered less likely to include plants of hybrid ancestry than the *D. arborea* and *D. ciliolata* populations proximate to Waipahoehoe Gulch. Some clusters of *D. arborea* in the vicinity of DaM have phenotypes that suggest hybrid origins, but the sampled population was chosen to minimize the likelihood of hybrid ancestry. Each of these populations has also been included in other studies (Lawton-Rauh *et al*. 2003, 2007; Friar *et al*. 2007). All sites are located at altitudes of 2100–2900 m above sea level on Mauna Kea (Fig. 2). The core portion of the Waipahoehoe Gulch population was exhaustively sampled for a total of 250 reproductively mature plants.

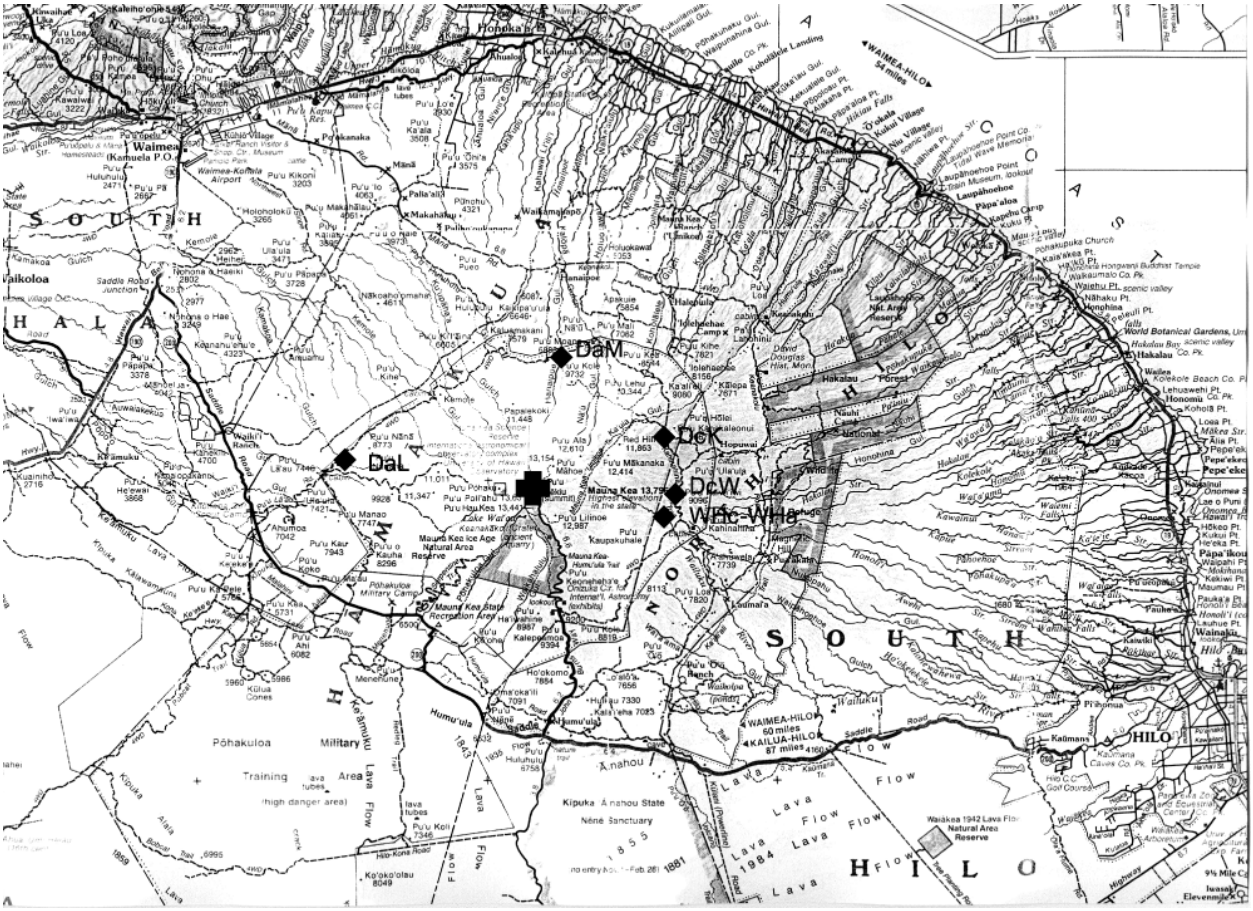

Figure 2. Map of the Mauna Kea region of Hawaii showing locations of sample populations (diamonds) and the summit of Mauna Kea (cross). DaL, Pu'u La'au; DaM, Pu'u Mali; Dc, Pu'u Kaiwi'iwi; DcK, Pu'u Kanakaleonui; WHc-WHa, Waipahoehoe Gulch *D. ciliolata*-like and *D. arborea*-like phenotypic extremes, respectively. (Base map reproduced from Map of Hawai'i: the Big Island, from Reference Maps of the Islands of Hawai'i, University of Hawai'i Press, used with permission.)

All Waipahoehoe Gulch plants were also measured for six traits: average length and width of three randomly chosen fully expanded leaves, plant height, crown width, length of the most terminal 20 elongated internodes on a vigorous vegetative shoot, and number of orders of branching in the terminal 30 cm of the same shoot. Personal observations had indicated that average values for these traits differ substantially between *D. arborea* and *D. ciliolata*. A principal components analysis was conducted on these six traits using PROC PRINCOMP of sas (SAS Institute Inc. 1999). The first principal component (PC1), which explained 63% of the total trait variation, was strongly correlated with a visual numerical rating of the degree of *D. arborea*-like vs. *D. ciliolata*-like appearance. Consequently, PC1 was used to rank the plants from most *D. ciliolata*-like to most *D. arborea*-like. Collection and analysis of measurements are described in more detail in Kirchoff *et al*. (2004). Only the 24 most *D. arborea*-like and the 24 most *D. ciliolata*-like plants (WHa and WHc, respectively) were used in the analyses reported in this study, because we were specifically interested in testing the hybrid status of hybrid-zone plants with parental species phenotypes. The remaining hybrid-zone plants showed various intermediate phenotypic combinations (Fig. 1), and their hybrid origin was not in doubt.

AFLP detection and scoring

Samples from 12 individuals each from the DcW, DcK, DaL and DaM populations, and the 24 individuals each from WHa and WHc were prepared for AFLP analysis. DNA was isolated from 25 to 50 mg ground frozen leaf tissue using the DNeasy Plant Mini-kit (QIAGEN). The isolation protocol was modified by adding 4 mg/mL caylase (provided by Elizabeth Friar, Rancho Santa Ana Botanic Garden) to the cell lysis Buffer A1, and by diluting the elution buffer AE 1:4 prior to elution of DNA in 200 µL total volume. The eluted DNA was concentrated to a volume of approximateky 50 µL in a centrifugal vacuum evaporator. AFLP template preparation and selective amplifications were carried out using LI-COR AFLP Template Preparation and Selective Amplification Kits (LI-COR Biosciences). AFLP kit protocols were modified by skipping the heat inactivation of restriction enzymes prior to ligation of adapters, decreasing the pre-amplification annealing time to 30 s, increasing the number of pre-amplification cycles to 25, reducing the selective amplification denaturation time to 10 s, and adding 1 s/cycle to the 1:00 extension time for the last 23 cycles of the monoplex selective amplifications. Selective amplification products from reactions using IRD700-labelled *Eco*RI primers were each pooled with products from reactions using IRD800-labelled *Eco*RI primers, and were concentrated to half the original volume in a centrifugal vacuum evaporator prior to adding formamide stop solution. AFLP products were resolved on LI-COR Model 4200 automated DNA Analysers. Fragments were scored using AFLP quantar software (Keygene) with manual checking and correction of all automatically scored fragments. All genotypes from individual primer combinations were selectively amplified in the same plate and resolved on the same sequencing gel to avoid systematic errors from plate-to-plate or gel-to-gel differences in fragment resolution. None of the samples were replicated in multiple runs, but images were examined for evidence of artefacts from poor-quality template, such as absence of monomorphic fragments due to allelic dropout and random variation in sample-to-sample intensity of individual fragments, before proceeding with scoring. Care was taken to identify and avoid scoring fragments that showed evidence of being amplification artefacts or comigrating fragments from different loci (e.g. slight mobility variation among samples or intensity variation greater than expected homozygote–heterozygote differences).

Data analysis

Estimates of the molecular differentiation among populations were obtained using hickoryversion 1.0 (Holsinger *et al*. 2002; Holsinger & Wallace 2004). hickory produces empirical Bayesian estimates of molecular differentiation in the allele frequencies underlying the observed marker frequencies by sampling from the distributions of θ^B, a measure of F_{ST}, and the inbreeding coefficient *f*. Posterior means and standard deviations (SD) of θ^B were obtained for both a full model in which *f* was also estimated and a reduced $f = 0$ model. The $f = 0$ model had higher deviance information criterion (DIC) values (Spiegelhalter *et al*. 2002) in all comparisons, suggesting poorer model fit, but the DIC difference between the two models was always less than the increase in the effective number of parameters estimated under the $f = 0$ model relative to the full model, providing little support for a positive value of *f* (Holsinger & Wallace 2004). Moreover, the estimates for *f* obtained under the full model (0.40–0.69 in the various comparisons) are biologically unrealistic for these self-incompatible taxa with weak interspecific and intraspecific differentiation. Consequently, we have chosen to use the results for the $f = 0$ model. θ^B values for the full model were in all cases slightly higher but followed the same pattern as those from the $f = 0$ model.

Bayesian analysis of population ancestry for each individual was conducted using the program structure version 2.2 (Pritchard *et al*. 2000a; Falush *et al*. 2003, 2007). Unlike previous versions of structure, the Markov chain Monte Carlo algorithm in version 2.2 correctly accommodates the genotypic uncertainty in data from dominant markers such as AFLPs by sampling band-present genotypes as either homozygotes or heterozygotes according to their posterior probabilities (Falush *et al*. 2007). Dominant AFLP marker data were entered by coding both alleles as '1' if band-present and as '0' if band-absent, and specifying '0' as a recessive allele for all markers as required (Pritchard *et al*. 2007). In order to eliminate possible spurious inferences due to unscored markers, 11 individuals that could not be scored for one or more AFLP primer combinations were eliminated from the Bayesian analysis. This resulted in inclusion of only five individuals from DaM, 22 individuals from WHc, and 15 individuals from WHa, along with the 12 individuals sampled from each of DcW, DcK, and DaL. Estimates were obtained under the admixture model using the correlated allele frequencies option and without using prior population information. The number of inferred populations (or 'source populations' as opposed to sampled populations) *K* was allowed to range from one to six, and two or more replicate runs were made for each value of *K* under each model. A burn-in of 100 000 cycles and data collection for 100 000 cycles were used in each run. The admixture model estimates the proportion of each individual's genome that is descended from each of the *K* inferred source populations. Molecular differentiation among $K = 2$ source populations was evaluated by assigning individual plants to populations based their predicted ancestries, and F_{ST} values were estimated for pairs of populations using hickory as described above.

Results

AFLP polymorphisms

We obtained AFLP templates that could be successfully amplified from eight individuals from the DaM population, 23 individuals from WHc, 22 individuals from WHa, and all 12 individuals from DcW, DcK, and DaL. AFLP analysis using five selective primer combinations generated a total of 166 scoreable polymorphic fragments as follows: E + AAG/M + CAC, 54 fragments; E + AAG/M + CTC, 15 fragments; E + AGG/M + CAT, 15 fragments; E + AAC/M + CAT, 44 fragments; and E + AGG/M + CTC, 38 fragments.

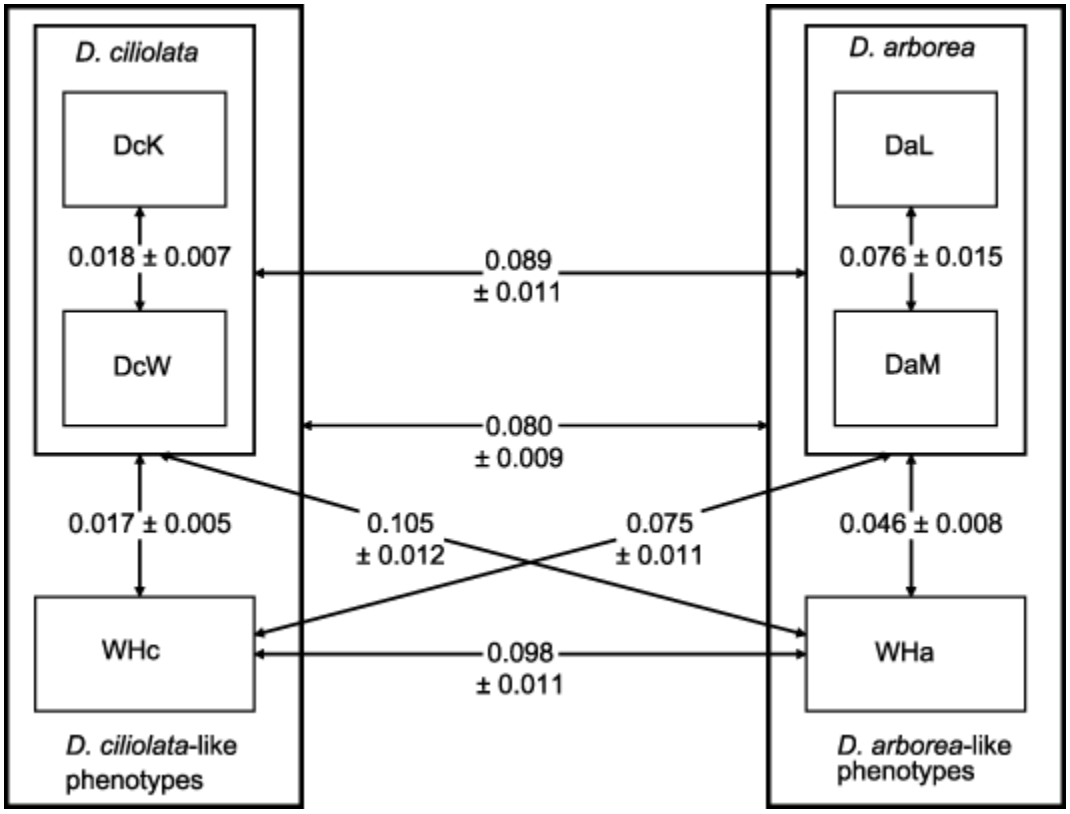

Figure 3. Genetic differentiation between species, populations, and sets of populations, estimated using Hickory vs. 1.0, f = 0 model. Pairwise F_{ST} values are estimated as the posterior mean (± SD) of θ^B. For abbreviations, see Fig. 2 legend.

Molecular differentiation between species and populations

We estimated the levels of molecular differentiation (F_{ST}) between *Dubautia arborea* and *Dubautia ciliolata* and between populations of each species (Fig. 3) using Bayesian estimates of θ^B from the software package hickory (Holsinger *et al*. 2002; Holsinger & Wallace 2004). The resulting estimate of differentiation between *D. arborea* and *D. ciliolata* (F_{ST} = 0.089) confirmed low levels of molecular differentiation between these species. Out of 166 polymorphic AFLP marker loci, 22 were monomorphic for either the band-absent or band-present class in one of the two species (14 in *D. ciliolata* and eight in *D. arborea*). The monomorphic marker class had an estimated allele frequency of less than 0.5 in the other species under Hardy–Weinberg assumptions at only three of these loci. The low level of interspecific molecular differentiation and extensive sharing of polymorphisms indicate that low levels of differentiation between *D. arborea* and *D. ciliolata* are a genome-wide phenomenon. Within *D. ciliolata*, the two populations showed very low levels of subdivision (F_{ST} = 0.018). By contrast,

the two *D. arborea* populations were almost as differentiated from each other (F_{ST} = 0.076) as were the two species.

When we compared the contrasting phenotypically parental groups of Waipahoehoe Gulch hybrid zone individuals (WHc and WHa), the degree of differentiation between the two extremes was comparable to that between the two species (F_{ST} = 0.098; Table 1 and Fig. 3). Moreover, the differentiation of WHc from *D. ciliolata* was almost identical to that between the two *D. ciliolata* populations (0.017), and WHa was less differentiated from *D. arborea* than the two *D. arborea* populations were from each other. The two species were about as differentiated from the opposite hybrid-zone phenotypic extreme as they were from each other (F_{ST} = 0.105 and F_{ST} = 0.075 for the two contrasts). These data were consistent with what would be expected if the WHc and WHa sample populations consisted of nonhybrid *D. ciliolata* and *D. arborea* individuals, respectively, rather than hybrids.

Table 1. Pairwise F_{ST} values, estimated as the posterior mean (± SD) of θ^B, for molecular differentiation between pairs of sampled populations

Population	**DcW (*n* = 12)**	**WHc (*n* = 23)**	**DaL (*n* = 12)**	**DaM (*n* = 8)**	**WHa (*n* = 22)**
DcK (*n* = 12)	0.0182 (± 0.0074)	0.0213 (± 0.0082)	0.1285 (± 0.0160)	0.0736 (± 0.0153)	0.1122 (± 0.0139)
DcW		0.0180(± 0.0065)	0.1281(± 0.0164)	0.0627(± 0.0150)	0.1005(± 0.0130)
WHc			0.1085(± 0.0139)	0.0545(± 0.0113)	0.0978(± 0.0114)
DaL				0.0760(± 0.0148)	0.0516(± 0.0109)
DaM					0.0732(± 0.0138)

We computed F_{ST} values for all remaining pairs of populations (Table 1). The DaM population showed only about half the level of differentiation from *D. ciliolata* group populations (including WHc) as did DaL. WHa showed intermediate levels of differentiation from the *D. ciliolata* group populations.

Because the differentiation patterns of the sampled WHc and WHa individuals suggested a predominantly nonhybrid origin for these groups, we repeated analyses of marker frequency distributions by combining the WHc and WHa data with the other two *D. ciliolata*and the other two *D. arborea* populations, respectively, to form two species groups. In the combined data set, only seven loci were monomorphic in one or the other species group, six of which were monomorphic in the *D. ciliolata* group. The single locus (E-AAG/M-CAC107) that was monomorphic in the *D. arborea* group, with all individuals band-present, had an estimated band-present allele frequency (under Hardy–Weinberg assumptions) of 0.49 in the *D. ciliolata* group. The monomorphic marker states at the other loci, however, were generally at high frequency in the other species group as well. Whether the WHa and WHc samples are included or excluded from the analysis, the number of monomorphic markers is greater in *D. ciliolata* than in *D. arborea*, but in both cases the differences are not statistically significant (P = 0.121 and P = 0.270, respectively, from two-tailed Fisher's exact test). Allelic diversity estimates from hickory for the combined data sets were also slightly lower for *D. ciliolata* (Hd = 0.372 ± 0.003 SD) than for *D. arborea* (Hd = 0.387 ± 0.003 SD).

Ancestry analysis

To test for gene flow between *D. arborea* and *D. ciliolata*, we used a Bayesian population clustering approach implemented in the program structure (Pritchard *et al*. 2000a; Falush *et al*. 2003) to infer the population ancestry (source population) or source population admixture of each plant in the study. (We use the term 'source populations' to describe the populations inferred from the ancestry analysis to distinguish these populations from the DcK, DcW, DaL, DaM, WHa, and WHc sampled populations.) The number of source populations was evaluated at values of K ranging from 1 to 6, with the maximum $K = 6$ corresponding to the number of sampled populations including WHa and WHc. Our objective was to test whether the two species form distinct clusters with minimal evidence of admixture, as predicted by a model of reproductive isolation between species, and secondarily partial isolation between sampled locations within each species. Inference of interspecific gene flow would require finding admixed individuals in the 'pure' populations of the two species, because we cannot rule out the possibility that some of the sampled plants from Waipahoehoe Gulch are hybrids.

The structure runs provided strong support for two source populations, with large and consistent improvements in the probability function [Ln P(D)] for runs with $K = 2$ relative to $K = 1$ (Fig. 4). When $K = 2$, many plants were supported as having nearly pure ancestry corresponding to their *D. arborea* and *D. ciliolata* identities, but some plants, especially in *D. arborea*, showed substantial interspecific admixture (Fig. 5). The lowest average percentage of admixture was actually found in the *D. ciliolata* phenotypes in the Waipahoehoe Gulch hybrid zone. Most of the plants with *D. arborea* phenotypes in Waipahoehoe Gulch showed evidence of admixture, but so did several of the plants from the 'pure'*D. arborea* population at Pu'u La'au. Surprisingly, four of the five *D. arborea* from Pu'u Mali included in the ancestry analysis were inferred to be of nearly pure *D. ciliolata* ancestry. One plant of *D. arborea*phenotype from Waipahoehoe Gulch was also estimated to be of nearly pure *D. ciliolata*descent.

There was poor statistical support for $K > 2$. Values of Ln P(D) did not improve for $K = 3$ or $K = 4$ relative to $K = 2$ (Fig. 4). Ln P(D) again improved modestly for $K = 5$, but $K = 6$ runs did not consistently produce further improvements. With four or more source populations, inferred source population admixtures differed somewhat between runs and some of the source populations were found only in highly admixed individuals. Moreover, the distribution of additional source populations among individuals did not appear to have a strong relationship to species or to individual-sampled populations (data not shown). Thus, there appears to be no clear statistical or biological basis for inferring the presence of more than two distinct source populations for our samples.

In order to estimate the effect of the detected admixture on molecular differentiation between *D. arborea* and *D. ciliolata*, we selected all effectively unadmixed individuals in each species group (estimated admixture < 10% in $K = 2$ runs), and estimated F_{ST} between these two groups in hickory. The resulting F_{ST} estimate using 39 unadmixed *D. ciliolata* and 13 *D. arborea* was 0.142 (± 0.0139 SD), compared to the overall differentiation of 0.080 between the two species groups. This suggests that gene flow has reduced the differentiation between *D. arborea* and *D. ciliolata* by nearly 44%. Even this figure must be considered a lower bound for the effect of gene flow on differentiation, as the source populations identified by structure may themselves have been influenced by older episodes of gene flow.

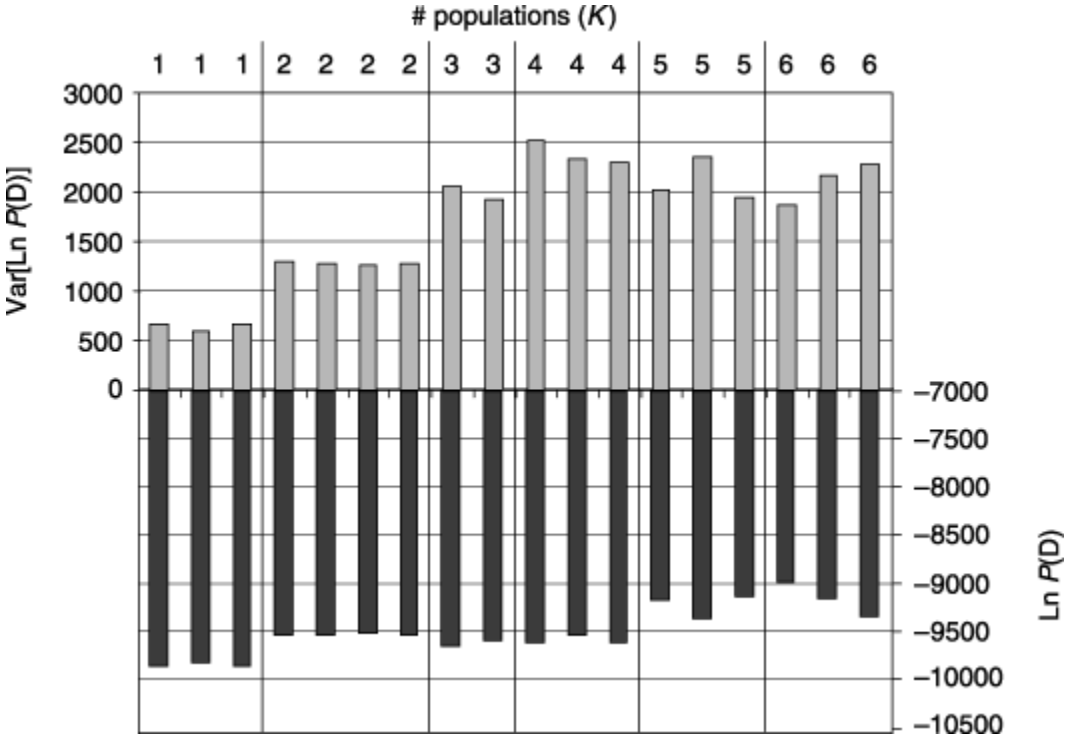

Figure 4. Estimated log probability of AFLP marker data inferred from structure version 2.2, given the number of source populations *K* (Ln *P*(D)) and its variance over 100 000 iterations after burn-in (Var [Ln *P*(D)]).

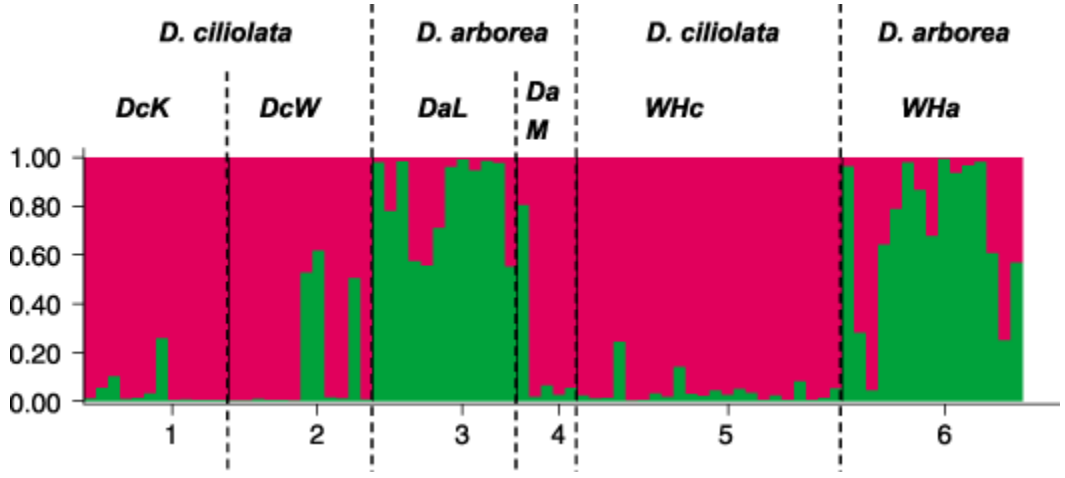

Figure 5. Results of Bayesian ancestry analysis from Structure version 2.2 under the admixture model for *K* = 2. Vertical columns represent individual plants, and the heights of bars of each colour are proportionate to the posterior means of estimated admixture proportions.

Discussion

Molecular and morphological differentiation between *D. arborea* and *D. ciliolata*

Our results indicate that the low level of interspecific differentiation is a genome-wide phenomenon. Our F_{ST} estimate of 0.080 for differentiation between *Dubautia arborea* and *Dubautia ciliolata* species groups is lower than those reported for a recent meta-analysis of differentiation between populations within individual species, in which a median F_{ST} of 0.14 among populations was found for outcrossing plants (Morjan & Rieseberg 2004). Such low levels of differentiation between species, however, are not unknown. A study using 389 molecular markers found an interspecific $F_{ST} = 0.036$ between the European oak species *Quercus robur* and *Quercus petraea* (Scotti-Saintagne *et al*. 2004). Nevertheless, the contrast between molecular and morphological divergence in *D. arborea* and *D. ciliolata* is striking. Distributions of trait values for leaf dimensions, plant height and internode lengths are almost completely disjunct for *D. arborea* and *D. ciliolata* (Fig. 1), but the two species share polymorphisms at 96% of the scored AFLP markers.

The paucity of markers fixed in either species suggests that selection at loci responsible for phenotypic differences between the two species did not generate large regions of selective sweep, as what might occur if speciation (or subsequent selection after hybridization) were accompanied by severe population bottlenecks that reduced the population recombination rate (e.g. Farnir *et al*. 2000; Maccaferri *et al*. 2005). None of the 166 AFLP fragments that were polymorphic in our sample displayed complementary fixation in *D. arborea* and *D. ciliolata*. Only seven of the polymorphic markers were monomorphic with respect to band presence or absence in either of the species groups, and four of these were fixed for the band-present phenotype that may include heterozygotes. Under Hardy–Weinberg assumptions, a band-present allele frequency as low as 0.88 would generate a probability greater than 0.5 of observing all 42 *D. arborea* or all 47 *D. ciliolata* individuals as band-present. Thus, our data may actually overestimate the proportion of the genome fixed in each species.

Other investigations have found low levels of molecular differentiation between *D. arborea*and *D. ciliolata* at several nuclear genes (Lawton-Rauh *et al*. 2007) and in allozymes (Witter & Carr 1988). In contrast, Friar *et al*. (2007) found much higher levels of differentiation in a set of seven microsatellite loci between some of the same populations that we sampled. However, they also found that most pairwise F_{ST} estimates between populations within species were as high as those for interspecific pairs of populations. Thus, it appears likely that the higher values of F_{ST} from microsatellites reflect differences in either the mutational processes or sampling properties of different marker systems rather than a higher estimate of interspecific reproductive isolation. Microsatellite loci generally have higher mutation rates than other classes of polymorphisms, and may thus sample a shorter time horizon (Remington *et al*. 2001; Jorgenson & Witte 2007). Genotyping errors could also affect estimates of differentiation (Bonin *et al*. 2004; Pompanon *et al*. 2005), possibly leading to underestimates of true values if the errors occur randomly with respect to population.

The similar levels of marker diversity observed in the two species fail to provide evidence of a progenitor-derivative relationship, in contrast with results obtained in other species pairs in the Hawaiian silversword alliance (Witter 1990). However, gene flow between the two species after their initial divergence would tend to increase genetic diversity in the derivative species, and thus obscure the evidence for the progenitor-derivative relationship. Because our ancestry analysis provides strong evidence of gene flow, conclusions about the mode of speciation between *D.*

arborea and *D. ciliolata* cannot be drawn on the basis of our data alone. F_{ST} estimates between populations within species showed substantially greater differentiation among *D. arborea* populations than among *D. ciliolata* populations whether or not the Waipahoehoe Gulch classes were included, which might suggest smaller population sizes in *D. arborea* or less gene flow between populations than in *D. ciliolata*. These differences appear significant, as the pairwise F_{ST} estimates between *D. arborea* populations and those between *D. ciliolata* populations are separated by more than two standard deviations in spite of the small sample sizes within populations (Table 1). Because of the small number of populations sampled per species, however, these results could easily be artefacts of the particular populations sampled. Moreover, the sampled *D. arborea*populations are more widely separated spatially than the *D. ciliolata* populations (Fig. 2), resulting in reduced opportunity for gene flow among these particular populations.

Hybridization and gene flow

We hypothesized that phenotypically *D. ciliolata*-like and *D. arborea*-like individuals in the Waipahoehoe Gulch hybrid zone are recombinant hybrids. Surprisingly, our data suggested that these classes were primarily of nonhybrid origin. The F_{ST} for WHc–WHa differentiation was comparable to that of the two species groups comprised of non-WH populations, and the differentiation of WHc and WHa from the other populations of the corresponding species was similar to the respective levels of DcK-DcW and DaL-DaM differentiation. The ancestry analysis results largely supported this inference. The WHc plants show little evidence of admixture. While most of the WHa plants are estimated to be admixed, it is unclear in the context of the overall results of the ancestry analysis whether this admixture is due to the ongoing hybridization in Waipahoehoe Gulch or was present in the parental *D. arborea* population. Sampling in the nearby *D. arborea* population presumed to be the source of the *D. arborea* phenotypes in the core hybrid zone could help resolve this question. These results suggest that the Waipahoehoe Gulch hybrid zone is very young, as the phenotypically extreme samples each comprise nearly 10% of all the sampled individuals. With just two generations of random mating in an admixed population with equal numbers of each species, the proportion of nonhybrid individuals remaining would be only ~6% (1/16) for each of the two species. Establishment of hybrid matings may have been gradual, however, due to progressive invasion of Waipahoehoe Gulch from opposite directions by *D. arborea* and *D. ciliolata* populations, so some advanced-generation hybrids could be present in the core hybrid zone.

Ironically, one of the other two *D. arborea* sampled populations (DaM) shows the highest level of hybrid admixture, with four of the five individuals included in the ancestry analysis estimated to be of nearly pure *D. ciliolata* ancestry. F_{ST} analyses, which included data from three more individuals, supported this inference, as the differentiation of DaM from the *D. ciliolata* populations was no greater than its differentiation from DaL and WHa, and only about half that of DaL from the *D. ciliolata* populations. Friar *et al*. (2007) also found evidence of *D. arborea* gene flow into the DaM population in their microsatellite data, consistent with our results. Populations of *D. ciliolata* occur not very far upslope from the broader *D. arborea* population that includes the sampled DaM population, which is spread out over a distance of several kilometres across the north slope of Mauna Kea. Some patches of obvious hybrids are present within this broader population. However, our sampling avoided individuals

showing phenotypic evidence of introgression, and the proximity of *D. ciliolata* to the sampled individuals is much less than at Waipahoehoe Gulch.

Some *D. arborea* individuals from DaL and *D. ciliolata* from DcW were also estimated to be highly admixed, even though there is neither phenotypic evidence of hybridization nor obvious opportunity for hybridization in these vicinities. This suggests that gene flow is not exclusively a recent phenomenon. Nevertheless, the clustering algorithm in structure was able to identify relatively distinct *D. arborea* and *D. ciliolata* source populations, and a number of individuals from all populations except DaM were estimated to be of relatively pure *D. ciliolata* or *D. arborea* origin. Thus, it appears that the overall extent of gene flow has not been high enough to obscure the genome-wide signal of differentiation between the two species. The comparison of F_{ST} estimates from the subsets of relatively pure *D. arborea* and *D. ciliolata* individuals and from the two species groups as a whole suggests that gene flow has reduced the interspecific molecular differentiation by nearly half and perhaps by more if the identified source populations have in turn been shaped by older episodes of gene flow. Statistical support for inferring more than two source populations is at best weak, so neither past hybridization events nor geographical isolation between intraspecific populations have generated source populations distinct enough to be detected by patterns of linkage disequilibrium. The analysis of Lawton-Rauh *et al*. (2007) using gene sequence data and different methodologies also suggests extensive gene flow has occurred between *D. arborea*and *D. ciliolata* subsequent to speciation.

The evolution of two such morphologically distinct species as *D. arborea* and *D. ciliolata* and their subsequent maintenance as distinct entities in the face of gene flow implies strong disruptive selection. The environmental factors acting as selective agents are unknown but may be related to the distinctness of the two ecosystems in which the two species occur. We have observed that the montane woodlands dominated by open stands of mamane (*Sophora chrysophylla*) and usually with a grassy understorey, are typically separated by relatively sharp ecotones on Mauna Kea from shrublands dominated by scattered pukiawe (*Styphelia tameiameiae*). *D. arborea* and *D. ciliolata* occur generally in somewhat patchy distributions within the respective woodland and shrubland ecosystems. Both *D. ciliolata*and *S. tameiameiae* have small narrow leaves that suggest adaptation to drought and/or cold conditions (Reich *et al*. 1997; Ackerly & Reich 1999), so differences in available moisture, possibly related to the typical locations of cloud banks shaped by prevailing trade wind patterns, may be a contributing factor.

Episodes of hybridization and gene flow between *D. ciliolata* and *D. arborea* may have been facilitated by the climatic history of Mauna Kea. The upper slopes of Mauna Kea have experienced repeated episodes of glaciation since the mid-Pleistocene, with the latest episode ending as recently as 9000 years ago. Glacial episodes coincided with lower air temperatures and a lowering of the snowline (Porter 1979). Associated shifts in the distribution of woodland and shrubland, and in the associated habitats for *D. arborea* and *D. ciliolata*, seem highly probable. It seems likely that populations of the advancing species during such shifts would have encountered and hybridized with populations of the other species rather than merely displacing them, with selection for the phenotypic characteristics that were adaptive in the changing environment of the hybrid population. Hybridization has also been suggested as a mechanism of species invasion in European oak species (Petit *et al*. 2004).

While this climate-fluctuation scenario is admittedly speculative, it provides a highly plausible explanation for the anomalous ancestry for the DaM population identified in our ancestry analysis and hinted at in the gene flow analysis of Friar *et al*. (2007). The DaM population could be the result of an upslope migration of *D. arborea* into an established population of *D. ciliolata* after the most recent glacial episode, followed by selection for *D. arborea* phenotypes. If so, the phenotypic selection process may still be incomplete, given the occasional hybrid phenotypes still present in the vicinity. Consequently, the DaM individuals were identified as admixed rather than representing a distinct population. A much larger sample from Pu'u Mali would be necessary to estimate the overall contribution of each species to the ancestry of this population, which would provide an indication of the relative sizes of the ancestral *D. arborea* and *D. ciliolata* populations. Studies of survival and physiological responses of *D. arborea*, *D. ciliolata*, and hybrids in contrasting environments will be required to verify the occurrence and strength of disruptive selection required by this scenario.

Dynamics of speciation

Our findings provide support for recent models of collective evolution and genic speciation (Rieseberg & Burke 2001; Wu 2001; Gottlieb 2003; Morjan & Rieseberg 2004). These models predict that genomes of incompletely isolated populations will consist of mosaics of differentiated and undifferentiated regions due to the combined effects of selection at adaptively important genes and gene flow. Theoretical considerations and recent empirical analyses suggest that low rates of gene flow among populations may be sufficient to allow adaptively favourable alleles to spread rapidly through a species, but not to break down differentiation in neutral regions of the genome (Barton & Hewitt 1985; Rieseberg & Burke 2001; Morjan & Rieseberg 2004). Meanwhile, gene flow between incipient species may homogenize their genomes at neutral loci while disruptive selection maintains or even increases differentiation at genes responsible for adaptive differentiation (Wu 2001; Gottlieb 2003; Scotti-Saintagne *et al*. 2004).

Both facets of this collective evolution model appear relevant to speciation in the *D. arborea–D. ciliolata* complex. At the species level, it appears that gene flow has been sufficient to reduce substantially the overall molecular differentiation between *D. arborea* and *D. ciliolata*, while disruptive selection has effectively negated interspecific gene flow at the still-unknown loci responsible for morphological differences. Moreover, the spread of favourable alleles from populations of one species to the other could have been rapid during climate-induced shifts in the locations of environments favourable to each species, even though the overall rate of gene flow between populations may have been insufficient to homogenize overall allele frequencies. This would result in populations with phenotypes of one species but genome-wide allele frequencies more similar to the other, as suggested by the Pu'u Mali samples in our ancestry analysis.

The number and effects of loci actually responsible for phenotypic differentiation between *D. arborea* and *D. ciliolata* remains to be explored. In cases where disruptive selection maintains interspecific phenotypic differences in the face of gene flow, the expected number of loci responsible for the differences would be small. Otherwise, the fraction of recombinant hybrids with parental allelic combinations would tend to be prohibitively low (Petit *et al*. 2004). Recent studies in sticklebacks indicate that independent episodes of selection on allelic variation in a

single gene have been responsible for parallel evolution of low-plated morphs in freshwater environments (Colosimo *et al*. 2005). In a hybridizing pair of European oak species, however, loci on nine separate chromosomes have been found to contribute to interspecific trait differences (Saintagne *et al*. 2004). Whether or not the growth habit differences between *D. arborea* and *D. ciliolata* are due to variation at a small number of genes of large effect has yet to be determined.

Implications for genetic studies

Natural hybrid zones have been suggested as a potentially useful alternative to controlled crosses for genetic mapping studies (Rieseberg *et al*. 2000; Rieseberg & Buerkle 2002). Multiple generations of hybridization may facilitate finer-scale resolution of quantitative trait loci (QTL), and simultaneous assessment of gene flow, fitness effects, and hybrid incompatibilities may be possible (Rieseberg & Buerkle 2002). Due to the presence of both parental and hybrid individuals in the Waipahoehoe Gulch hybrid zone, however, the effects of admixture would result in nonfunctional marker–trait associations. Thus, it seems doubtful that the hybrid zone offers any advantages over other *D. arborea* and *D. ciliolata*populations for initial investigation of marker–trait associations. Nevertheless, the Waipahoehoe Gulch population, including hybrids of intermediate morphological characteristics, may be a valuable resource for fine-scale evaluation of QTL detected in more conventional mapping studies. The combination of recent recombination events due to the ongoing hybridization and longer-term recombination during evolution of the *D. arborea–D. ciliolata* complex will have generated both large and small blocks of linkage disequilibrium, potentially allowing simultaneous coarse- and fine-structure mapping of QTL. Recently developed models that account for both estimated kinship among sampled individuals and population structure (Yu *et al*. 2006) may be especially useful for association mapping in such populations.

Acknowledgements

We express gratitude to Michael Purugganan for support and encouragement in carrying out this study, Amy Lawton-Rauh and Sola Haldorsdottir for field assistance, the Hawaii Department of Forestry and Wildlife for collection permits, Amy Lawton-Rauh for plant materials, and Elizabeth Friar, Amy Lawton-Rauh, Mitchell McGlaughlin, Dave McCauley, and three anonymous reviewers for helpful suggestions on the manuscript. Funding was provided by a National Science Foundation grant (DEB708540) to Michael Purugganan and R.H.R., an NIH Individual Postdoctoral Fellowship (5F32GM20554) to D.L.R., and a UNC-Greensboro Summer Excellence Award to D.L.R.

References

Ackerly, DD, Reich, PB (1999) Convergence and correlations among leaf size and function in seed plants: a comparative test using independent contrasts. *American Journal of Botany*, **86**, 1272– 1281.

Albertson, RC, Streelman, JT, Kocher, TD (2003) Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proceedings of the National Academy of Sciences, USA*, **100**, 5252– 5257.

Baldwin, BG (2005) Origin of the serpentine-endemic herb *Layia discoidea* from the widespread *L. glandulosa* (Compositae). *Evolution*, **59**, 2473– 2479.

Baldwin, BG, Robichaux, RH (1995) Historical biogeography and ecology of the Hawaiian silversword alliance (Asteraceae): new molecular phylogenetic perspectives. In: *Hawaiian Biogeography: Evolution on a Hot Spot Archipelago* (eds WL Wagner, VA Funk), pp. 259– 287. Smithsonian Institution Press, Washington, D.C.

Baldwin, BG, Sanderson, MJ (1998) Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proceedings of the National Academy of Sciences, USA*, **95**, 9402– 9406.

Barton, NH, Hewitt, GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113– 148.

Bonin, A, Bellemain, E, Bronken Eidesen, P *et al* . (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**, 3261– 3273.

Carr, GD (1985) Monograph of the Hawaiian Madiinae (Asteraceae): *Argyroxiphium, Dubautia,* and *Wilkesia*. *Allertonia*, **4**, 1– 123.

Carr, GD (2003) Hybridization in Madiinae. In: *Tarweeds and Silverswords: Evolution of the Madiinae (Asteraceae)* (eds S Carlquist, BG Baldwin, GD Carr), pp. 79– 104. Missouri Botanical Garden Press, St Louis, Missouri.

Carr, GD, Kyhos, DW (1981) Adaptive radiation in the Hawaiian silversword alliance (Compositae-Madiinae). I. Cytogenetics of spontaneous hybrids. *Evolution*, **35**, 543– 556.

Carr, GD, Kyhos, DW (1986) Adaptive radiation in the Hawaiian silversword alliance (Compositae-Madiinae). II. Cytogenetics of artificial and natural hybrids. *Evolution*, **40**, 959– 976.

Colosimo, PF, Hosemann, KE, Balabhadra, S *et al* . (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, **307**, 1928– 1933.

Corander, J, Marttinen, P (2006) Bayesian identification of admixture events using multi-locus molecular markers. *Molecular Ecology*, **15**, 2833– 2843.

Corander, J, Marttinen, P, Mäntyniemi, S (2006) Bayesian identification of stock mixtures from molecular marker data. Fishery Bulletin, **104**, 550– 558.

Dobzhansky, T (1946) Complete reproductive isolation between two morphologically similar species of *Drosophila*. *Ecology*, **27**, 205– 211.

Dobzhansky, T (1951) *Genetics and the Origin of Species*. Columbia University Press, New York.

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.

Falush, D, Stephens, M, Pritchard, JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Molecular Ecology Notes, **7**, 574– 578.

Farnir, F, Coppieters, W, Arranz, J-J *et al* . (2000) Extensive genome-wide linkage disequlibrium in cattle. *Genome Research*, **10**, 220– 227.

Ferguson, D, Sang, T (2001) Speciation through homoploid hybridization between alotetraploids in peonies (*Paeonia*). *Proceedings of the National Academy of Sciences, USA*, **98**, 3915– 3919.

Friar, EA, Cruse-Sanders, JM, McGlaughlin, ME (2007) Gene flow in *Dubautia arborea* and *D. ciliolata*: the roles of ecology and isolation by distance in maintaining species boundaries despite ongoing hybridization. *Molecular Ecology*, doi: 10.1111/j.1365-294X.2007.03423.x

Gottlieb, LD (2003) Rethinking classic examples of recent speciation in plants. *New Phytologist*,**161**, 71– 82.

Gottlieb, LD, Warwick, SI, Ford, VS (1985) Morphological and electrophoretic divergence between *Layia discoidea* and *L. glandulosa*. *Systematic Botany*, **10**, 484– 495.

Grant, V (1949) Pollination systems as isolating mechanisms in angiosperms. *Evolution*, **3**, 82– 97.

Holsinger, KE, Lewis, PO, Dey, DK (2002) A Bayesian approach to inferring population structure from dominant markers. *Molecular Ecology*, **11**, 1157– 1164.

Holsinger, KE, Wallace, LE (2004) Bayesian approaches for the analysis of population genetic structure: an example from *Planthera leucophaea* (Orchidaceae). *Molecular Ecology*, **14**, 887– 894.

Jorgenson, E, Witte, JS (2007) Reply: microsatellite markers for genome–wide association studies. *Nature Reviews Genetics*, **8**, http://www.nature.com/nrg/journal/v8/n2/pdf/nrg1962-c2.pdf [accessed on February 2007].

Judd, WS, Campbell, CS, Kellogg, EA, Stevens, PF, Donoghue, MJ (2002) *Plant Systematics: a Phylogenetic Approach*, 2nd edn. Sinauer Associates, Sunderland, Massachusetts.

Kirchoff, BK, Richter, SJ, Remington, DL, Wisniewski, E (2004) Complex data produce better characters. *Systematic Biology*, **53**, 1– 17.

Lawton-Rauh, A, Robichaux, RH, Purugganan, MD (2007) Diversity and divergence patterns in regulatory genes suggest differential gene flow in recently derived species of the Hawaiian silversword alliance adaptive radiation (Asteraceae). *Molecular Ecology*, **16**, 3995– 4013.

Lawton-Rauh, A, Robichaux, RH, Purugganan, MD (2003) Patterns of nucleotide variation in homoeologous regulatory genes in the allotetraploid Hawaiian silversword alliance (Asteraceae). *Molecular Ecology*, **12**, 1301– 1313.

Maccaferri, M, Sanguineti, MC, Noli, E, Tuberosa, R (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. *Molecular Breeding*, **15**, 271– 289.

Mayr, E (1942) *Systematics and the Origin of Species*. Columbia University Press, New York.

Morjan, CL, Rieseberg, LH (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology*, **13**, 1341– 1356.

Panero, JL, Francisco-Ortega, J, Jansen, RK, Santos-Guerra, A (1999) Molecular evidence for multiple origins of woodiness and a New World biogeographic connection of the Macaronesian Island endemic *Pericallis* (Asteraceae: Senecioneae). *Proceedings of the National Academy of Sciences, USA*, **96**, 13886– 13891.

Peichel, CL, Nereng, KS, Ohgl, KA *et al* . (2001) The genetic architecture of divergence between threespine stickleback species. *Nature*, **414**, 901– 905.

Petit, RJ, Bodenes, C, Ducousso, A, Roussel, G, Kremer, A (2004) Hybridization as a mechanism of invasion in oaks. *New Phytologist*, **161**, 151– 164.

Pompanon, F, Bonin, A, Bellemain, E, Taberlet, P (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, **6**, 847– 859.

Porter, SC (1979) Hawaiian glacial ages. *Quaternary Research*, **12**, 161– 187.

Yu, J, Pressoir, G, Briggs, WH *et al* . (2006) A unified mixed–model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, **38**, 203– 208.

Pritchard, JK, Stephens, M, Donnelly, P (2000a) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945– 959.

Pritchard, JK, Stephens, M, Rosenberg, NA, Donnelly, P (2000b) Association mapping in structured populations. *American Journal of Human Genetics*, **67**, 170– 181.

Pritchard, JK, Wen, X, Falush, D (2007) *Documentation for structure Software: Version 2.2*. Department of Human Genetics, University of Chicago, Chicago, Illinois.

Reich, PB, Walters, MB, Ellsworth, DS (1997) From tropics to tundra: global convergence in plant functioning. *Proceedings of the National Academy of Sciences, USA*, **94**, 13730– 13734.

Remington, DL, Thornsberry, JM, Matsuoka, Y *et al* . (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences, USA*,**98**, 11479– 11484.

Rieseberg, LH (1997) Hybrid origins of plant species. *Annual Review of Ecology and Systematics*, **28**,359– 389.

Rieseberg, LH, Baird, SJE, Gardner, KA (2000) Hybridization, introgression, and linkage evolution. *Plant Molecular Biology*, **42**, 205– 224.

Rieseberg, LH, Buerkle, CA (2002) Genetic mapping in hybrid zones. *American Naturalist*, **159**, S36– S50.

Rieseberg, LH, Burke, JM (2001) The biological reality of species: gene flow, selection, and collective evolution. *Taxon*, **50**, 47– 67.

Rieseberg, LH, Whitton, J, Gardner, K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, **152**, 713– 727.

Robichaux, RH, Carr, GD, Liebman, M, Pearcy, RW (1990) Adaptive radiation of the Hawaiian silversword alliance (Compositae-Madiinae): ecological, morphological, and physiological diversity. *Annals of the Missouri Botanical Garden*, **77**, 64– 72.

Saintagne, C, Bodenes, C, Barreneche, T *et al* . (2004) Distribution of genomic regions differentiating oak species assessed by QTL detection. *Heredity*, **92**, 20– 30.

SAS Institute Inc. (1999) *The SAS System for Windows*. SAS Institute, Cary, North Carolina.

Schluter, D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford, UK.

Schluter, D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution*, **16**, 372– 380.

Scotti-Saintagne, C, Mariette, S, Porth, I *et al* . (2004) Genome scanning for interspecific differentiation between two closely related oak species. *Genetics*, **168**, 1615– 1626.

Seehausen, O (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, **19**, 198– 207.

Shapiro, MD, Marks, ME, Peichel, CL *et al* . (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, **428**, 717– 723.

Spiegelhalter, D, Best, N, Carlin, B, Van Der Linde, A (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B, Biological Sciences*, **64**, 583– 639.

Streelman, JT, Albertson, RC, Kocher, TD (2003) Genome mapping of the orange blotch colour pattern in cichlid fishes. *Molecular Ecology*, **12**, 2465– 2471.

Thornsberry, JM, Goodman, MM, Doebley, J *et al* . (2001) *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genetics*, **28**, 286– 289.

Turelli, M, Barton, NH, Coyne, JA (2001) Theory and speciation. *Trends in Ecology & Evolution*, **16**,330– 343.

Ungerer, MC, Baird, S, Pan, J, Rieseberg, LH (1998) Rapid hybrid speciation in wild sunflowers. *Proceedings of the National Academy of Sciences, USA*, **95**, 11757– 11762.

Vos, P, Hogers, R, Bleeker, M *et al* . (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407– 4414.

Witter, MS (1990) Evolution in the Madiinae: evidence from enzyme electrophoresis. *Annals of the Missouri Botanical Garden*, **77**, 110– 117.

Witter, MS, Carr, GD (1988) Adaptive radiation and genetic differentiation in the Hawaiian silversword alliance (Compositae: Madiinae). *Evolution*, **42**, 1278– 1287.

Wolfe, AD, Xiang, Q-Y, Kephart, SR (1998) Diploid hybrid speciation in *Penstemon* (Scrophulariaceae). *Proceedings of the National Academy of Sciences, USA*, **95**, 5112– 5115.

Wu, CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851– 865.