MOLECULAR ECOLOGY

Molecular Ecology (2014) doi: 10.1111/mec.12962

Gene flow between nascent species: geographic, genotypic and phenotypic differentiation within and between *Aquilegia formosa* and *A. pubescens*

C. NOUTSOS,*† J. O. BOREVITZ*‡ and S. A. HODGES§

*Department of Ecology and Evolution, University of Chicago, 1101E 57th Street, Chicago, IL 60637, USA, †Cold Spring Harbor Lab, Cold Spring Harbor, NY 11724, USA, ‡Research School of Biology, The Australian National University, Canberra, ACT 0200, Australia, §Department of Ecology, Evolution & Marine Biology, University of California, Santa Barbara, CA 93106-9620, USA

Abstract

Speciation can be described as a reduction, and the eventual cessation, in the ability to interbreed. Thus, determining how gene flow differs within and between nascent species can illuminate the relative stage the taxa have attained in the speciation process. *Aquilegia formosa* and *A. pubescens* are fully intercompatible, yet occur in different habitats and have flowers specialized for pollination by hummingbirds and hawkmoths, respectively. Using 79 SNP loci, we genotyped nearly 1000 individuals from populations of both species in close proximity to each other and from putative hybrid zones. The species shared all but one SNP polymorphism, and on average, allele frequencies differed by only 0.14. However, the species were clearly differentiated using Structure, and admixed individuals were primarily identified at putative hybrid zones. PopGraph identified a highly integrated network among all populations, but populations of each species and hybrid zones occupied distinct regions in the network. Using either conditional graph distance (*cGD*) or Fst/(1-Fst), we found significant isolation by distance (IBD) among populations. Within species, IBD was strong, indicating high historic gene flow. IBD extended approximately 100 km in *A. pubescens* and 30 km in *A. formosa*. However, IBD between the species was very weak and extended only a few km beyond hybrid zones, suggesting little recent gene flow. The extensive sharing of SNP polymorphisms between these species suggests that they are very early in the speciation process while the low signal of IBD suggests that they have largely ceased gene exchange.

Keywords: columbines, hybridization, isolation by distance, population structure, SNPs

Received 5 November 2012; revision received 30 September 2014; accepted 3 October 2014

Introduction

The most commonly used definitions of speciation involve the evolution of reproductive isolation (Coyne & Orr 2004). When populations of different species no longer exchange genes (i.e. there is no gene flow), the entire genome of each species evolves independently. A decline in gene flow could be due simply to the physical distance between the nascent species, that is allopatric speciation. However, ecological speciation due to ecologically based divergent natural selection (Schluter 2009) may also reduce gene flow, even on a spatial scale where gene flow would otherwise occur. Recently, many studies have recognized that gene flow may occur during the speciation process and that genomic regions may differ in the degree of introgression between nascent species (Feder *et al.* 2012; Nosil 2008; Zheng & Ge 2010). Some genomic regions will be resistant to introgression because they harbour genes that cause reproductive isolation while other genomic regions can more easily introgress (Baack & Rieseberg

Correspondence: Scott A. Hodges, Fax: (805) 893 2266;
E-mail: scott.hodges@lifesci.ucsb.edu

2007). Regardless, as speciation progresses, gene flow between nascent species will decline relative to gene flow among populations within species. Thus, measures of how historic gene flow differs within and between nascent species *on the same spatial scale* will inform us about the average degree of reproductive isolation that has been achieved through the incipient ecological speciation process.

To measure how gene flow changes during speciation, it is important to consider taxa that have not fully achieved complete reproductive isolation because after this point, the taxa will change further for reasons that have nothing to do with the speciation process (Via 2009). Furthermore, to understand how ecological speciation affects gene flow, it is important to consider taxa that are still in close proximity, so that the landscape scale of populations of both nascent species does not preclude gene exchange. Such situations will allow differences in the scale of gene flow within and between species to be linked to ecological differences. Obvious candidates for such studies are taxa that currently form natural hybrid zones (Abbott *et al.* 2013). Usually, the frequency of advanced generation hybrids is interpreted as indicating the degree of gene flow between species, for example, large hybrid swarms suggest extensive gene flow while only occasional F1 hybrids imply rare or very little gene flow between the taxa (Abbott *et al.* 2013).

Gene flow can be estimated using cline theory and measurement of gametic disequilibrium (Barton 2000). However, it is difficult to quantify gametic disequilibrium when population differentiation is low (Sotka & Palumbi 2006). Another classic way to measure past gene flow is through the identification of a pattern of isolation by distance (IBD). Gene flow is expected to decline with physical distance as described by Wright (1943, 1946) and will produce a pattern of IBD whereby populations separated by greater distances will be more genetically differentiated than populations close together. Thus, identification of IBD has long been used to infer the scale of gene flow among populations (Chenoweth *et al.* 1998). Both one- and two-dimensional stepping stone models of migration have been used to predict patterns of IBD (Slatkin 1993) accounting for linear or exponential decay of genetic covariation with distance. Pairwise measures of Fst have traditionally been used as the measure of genetic differentiation when assessing IBD. A new measure of genetic distance, conditional graph distance (*cGD*), has been suggested to be superior in many instances (Dyer & Nason 2004; Dyer *et al.* 2010; Kluetch *et al.* 2012). This is because *cGD* is derived from a network approach that considers the data from all populations simultaneously to estimate population relatedness (Dyer *et al.* 2010).

Two species of *Aquilegia*, *A. formosa* and *A. pubescens*, are good candidates to compare landscape levels of gene flow. These two species have been studied for over 60 years to understand the processes of speciation and adaptation (Chase & Raven 1975; Grant 1952; Hodges *et al.* 2002), and they differ dramatically in floral pollination syndromes with differences in colour, spur length and flower orientation associated with preferences by hummingbirds or hawkmoths. In addition, both species are found in the southern Sierra Nevada mountains of California although they are separated ecologically from *A. formosa* usually occurring in shady sites below treeline with well-developed soils and *A. pubescens* occurring in poor, rocky soils above treeline. Despite these differences, natural hybrid zones with many advanced generation hybrids occur frequently when the two habitats come into close contact (Chase & Raven 1975; Grant 1952; Hodges & Arnold 1994) indicating that gene flow likely extends between the species. The two species are very closely related with *A. pubescens* being the sister taxon to a group of morphologically very similar species including *A. formosa* (Whittall & Hodges 2007), and remarkably few fixed genetic differences have been identified between the species indicating shared ancestral polymorphism or extensive gene flow (Cooper *et al.* 2010; Hodges & Arnold 1994; Whittall & Hodges 2007).

Aquilegia has been selected as a model genus for genomic studies of development, ecology and evolution (Kramer 2009; Kramer & Hodges 2010), and thousands of putative genetic markers have been identified from EST sequencing of an *A. formosa* × *A. pubescens* hybrid mapping population. Here we use this resource to first determine the degree that these two species have become genetically differentiated. We further asked whether we could estimate the landscape level of gene flow within these species and whether it differs from the landscape level of gene flow between the species. We sought to test for differentiation between the species using statistics summarizing these loci and to test for the signature of past gene flow within and between the species by identifying the spatial extent of admixed individuals and using tests of IBD at a fine geographic scale. We also measured IBD using both linearized Fst and *cGD*, as *cGD* may provide a better signal as has been predicted (Dyer *et al.* 2010; Kluetch *et al.* 2012).

Materials and methods

Samples collection and DNA isolation

We collected 1012 samples from the southern Sierra Nevada Mountains. We collected from a number of watersheds spanning approximately 115 km north to

south. Locations of specific sites can be found in Table 1. We sought to collect low-elevation *A. formosa*, phenotypic hybrids and high-elevation *A. pubescens* sites within a watershed. Within one watershed (Bishop Creek), we collected extensively with 749 plants sampled. As these collections were nearly continuous along the watershed, we grouped samples to maintain sample sizes of approximately 30 individuals for the Fst/*cGD* analyses and used the centre location of all samples in a group to determine its location and distance to other groups. Leaf tissue was dried in silica gel for later DNA extraction. Genomic DNA was isolated as previously described (Whittall & Hodges 2007).

Trait measurements

Within the Bishop Creek watershed, we measured nine floral traits from one flower per plant, for 646 plants. These traits were spur and blade chroma, hue, brightness and length, and the orientation of flowers (Hodges *et al.* 2002). At other populations, assignment as species or hybrids was based on visual inspection of floral traits. We used principle component analysis to assess overall morphological differentiation. PC1 explained >75% of the variation and is used to distinguish *A. pubescens*, hybrids and *A. formosa*. Individual trait measures and PCs are provided as supplemental data.

Marker development

We used EST sequences derived from Sanger sequencing of cDNA constructed from pooled tissues from an *A. formosa* × *A. pubescens* F2 population, described elsewhere (Kramer & Hodges 2010, DFCI Aquilegia Gene Index release 2.1), to identify potential SNPs as genetic markers. We made this F2 population by crossing two F1 plants, each with a different pair of parents. Thus, four alleles, two from *A. formosa* and two from *A. pubescens*, segregate within the F2 population. A set of custom perl scripts was developed to call the SNPs from multiple EST sequences that were part of the same contig. The following criteria were used: more than one read of nonconsensus sequence, a quality score higher than 20 (meaning that the specific base has 99.99 probability to be correct) and no other SNP in the contig within 50 bases up- or downstream of the potential SNP. The latter requirement was to facilitate primer design. To reduce the possibility of designing primers spanning intron/exon boundaries, we undertook an additional filter using the tblastx program with default parameters. We filtered for SNPs and their flanking regions that had at least 60% sequence identity with individual exons in Arabidopsis. In total, we identified 3349 SNPs with these criteria (supplemental material). To design primers, we used the SEQUENOM (San Diego,

Table 1 Collection sites. For each site, the number of individuals samples (N), floral morphology (*A. formosa* (F), *A. pubescens* (P) or hybrid (H)), elevation, watershed (Lundy (L), Rock Creek (R), Gable (G) Bishop Creek (B) and South Lake (S) and latitude & longitude are indicated. High-elevation *A. pubescens* sites near, but not actually in a watershed are included in parentheses.

SITE	Abbreviation	Morph	N	Elev ft/m	Watershed	Latitude	Longitude
BIG_HORN_LAKE	BHL	P	31	10 955/3339	L	37.949	−119.300
LUNDY_HYBRIDS	LH	H	51	9117/2779	L	38.009	−119.287
LUNDY_TRAIL	LUNT	F	18	8241/2512	L	38.020	−119.269
MONO_PASS	MONOP	P	32	11 955/3644	R	37.432	−118.768
MOSQ_FLAT	MF	F	17	10 272/3131	R	37.435	−118.748
MORGAN_PASS	MORP	P	32	11 033/3363	R	37.388	−118.747
GABLE_LAKE	GL	F	23	9760/2975	G	37.339	−118.689
PIUTE_TRAIL_U	PTU	P	84	11 404/3476	B	37.234	−118.674
LOCH_LEVEN_HYBRID	LLH	H	126	10 751/3277	B	37.232	−118.658
LOCH_LEVEN_CASCADE	LLC	H	91	10 279/3133	B	37.231	−118.648
BISHOP_CREEK_U	BCU	F	29	10 075/3071	B	37.229	−118.643
BISHOP_CREEK _M	BCM	F	31	9816/2992	B	37.228	−118.637
BISHOP_CREEK _L	BCL	F	28	9380/2859	B	37.227	−118.629
LAMARCK_TRAIL_U	LTU	P	47	11 633/3546	B	37.204	−118.650
LAMARCK_LAKES_U	LAMU	H	58	10 859/3310	B	37.215	−118.643
LAMARCK_LAKES_L	LAML	H	61	10 741/3274	B	37.214	−118.645
GRASS_LAKE_CASCADE	GLC	F	80	10 190/3106	B	37.217	−118.636
NORTH_LAKE_STREAM	NLS	F	32	9301/2835	B	37.228	−118.623
BISHOP_PARK	BP	F	38	8254/2516	B	37.244	−118.594
FALLEN__LOG	FL	F	43	7720/2353	B	37.260	−118.578
SLAKE_L	SL	F	14	9609/2929	S	37.175	−118.563
BISHOP_PASS	BPT	P	32	11 584/3531	S	37.120	−118.542

CA, USA) software, iPLEX Gold, with the standard parameters. We used the SBE High Multiplex, module to design primers that could be multiplexed to assay as many as 40 SNP markers per sample. Six multiplex assays were made totalling 180 markers. Thirty-two samples were genotyped twice to determine the repeatability of each SNP marker.

Analyses

To analyse population structure, we used STRUCTURE software (Falush *et al.* 2003; Pritchard *et al.* 2000). Specifically, we used the linkage model with λ defined to 0.5180, as it was calculated by Structure, following the instructions on the Structure manual and using default parameters. The allele frequencies were set to independent. In addition, we used 100 000 Burnins and 100 000 MCMC. We used several *K* starting from $K = 2$ up to $K = 10$ using three replications for each K. To estimate the best K, we used Structure Harvester (Evanno *et al.* 2005). DeltaK was highest for $K = 2$ by a wide margin (>700) separating the nascent species and thus $K = 2$ is presented throughout.

Genetic distances between localities were calculated as pairwise F_{ST} using the F_{STAT} software v. 2.9.3 (Goudet 1995). We also used PopGraph (Dyer & Nason 2004) to measure the conditional graph distance (*cGD*) derived from population networks (Dyer *et al.* 2010). We conducted analyses via Genetic Studio (http://dyerlab.bio.vcu.edu/software.html) and then viewed the population graph with a downloadable version of Graph (Dyer 2009). PopGraph is free of a priori assumptions about population geographic arrangements and uses a graph theoretical approach to determine the minimum set of edges (connections) that sufficiently explain the total among-population covariance structure of all of the populations.

We used simple and partial Mantel tests to determine and quantify isolation by distance and to estimate how species differences may influence gene flow. Spatial distances between collection sites were calculated using gdist function in the IMAP R package. Species distance was defined as 1 between populations of different species, 0 between populations of the same species (or between hybrid populations) and 0.5 between populations of one of the species and a hybrid population. Mantel tests were performed, to correlate spatial, species and genetic distances in R using *mantel* function from the ECODIST package (Dray *et al.* 2007). To assess statistical significance, we used 10 000 permutations. We first performed Mantel tests on the entire data set to determine whether there was evidence for IBD and whether population morphology was correlated with genetic distance. We conducted these analyses using both absolute distance and log (distance) to test whether correlations were greater under assumptions of a one- or two-dimensional model, respectively, as recommended by Rousset (1997). We also conducted these test using both Fst/(1-Fst) and *cGD* to test whether *cGD* provided stronger evidence for IBD as predicted. Partial Mantel tests were conducted to test how controlling for one factor influenced the correlation of the other two factors. To determine how IBD differs within and between populations of each species, we performed Mantel tests with matrices constructed for each species separately and matrices of pairs between the two species.

Results

Morphological variation

We quantified floral characters among individuals belonging to pure and hybrid populations. Measurements of blade and spur chroma and length, as well as orientation from 646 plants from the Bishop watershed (Table 1), showed bimodal distributions. In contrast, blade and spur hue and brightness displayed skewed unimodal distributions as expected from previous analyses of the species and F2 hybrids (Hodges *et al.* 2002). We quantified the composite floral morphological diversity with Principal Component Analysis. Of nine total principal components, PC1 explained more than 75% of the variation and showed a clear bimodal distribution distinguishing the species (Fig. S1A, Supporting information). Floral morphology, summarized by PC1, tracks with altitude illustrating hybrids between the lower forest dwelling *A. formosa* and alpine *A. pubescens* (Fig. 1B, Supporting information).

Genetic variation

Description of markers. We used six multiplex pools of approximately 30 SNP markers to genotype a total of 180 SNP markers from 1012 samples of *A. formosa*, *A. pubescens*, and natural hybrids (see Materials and Methods section). We trimmed the data set to exclude markers that had either low repeatability, low frequency of genotypes, low variability (minor allele frequency <0.02) or indications of a high frequency of null alleles. We also excluded fourteen individuals because of poor DNA quality. The final data set consisted of 998 individuals genotyped at 79 loci. Loci were genotyped on average for 89% of the individuals (range 71–97%), and individuals were genotyped on average for 89% of the loci (range 51–100%). For these 79 loci across all samples, the average minor allele frequency was 0.186 (range 0.021–0.498).

Population structure, morphology and elevation. Structure analysis was performed with $K = 2$ strongly differentiating

the majority of samples (Figs 1 and 2). Over 61% of individuals were assigned as 95% to one or the other class. Over 75% of individuals were assigned as 90% to one or the other class with these two classes essentially separating the two species. Assignment to one class correlated strongly to high-elevation *A. pubescens* and the other class to low-elevation *A. formosa*. Specifically, low-elevation *A. formosa* populations (high PC1 values) are genetically assigned almost exclusively to group 1 by Structure (Fig. 1, right side). In contrast, high-elevation *A. pubescens* populations (low PC1 values) are genetically assigned almost exclusively to group 0 by Structure (Fig. 1 left side). Individuals at intermediate elevations show varying degrees of Structure assignment (Fig. 1 (LH) and Fig. 2; see also Fig. 1C, Supporting information). Populations with a high degree of admixture show reduced correlation with floral morphology (Fig. 2). For example, there are individuals with *A. formosa* or *A. pubescens* morphology, but are assigned to a complete range of admixtures (Fig. 2). Thus, at and near morphological hybrid zones, genotype and phenotype associations are greatly reduced. This indicates advanced generation backcross and hybrid individuals where genes controlling traits are uncoupled from population genomic background.

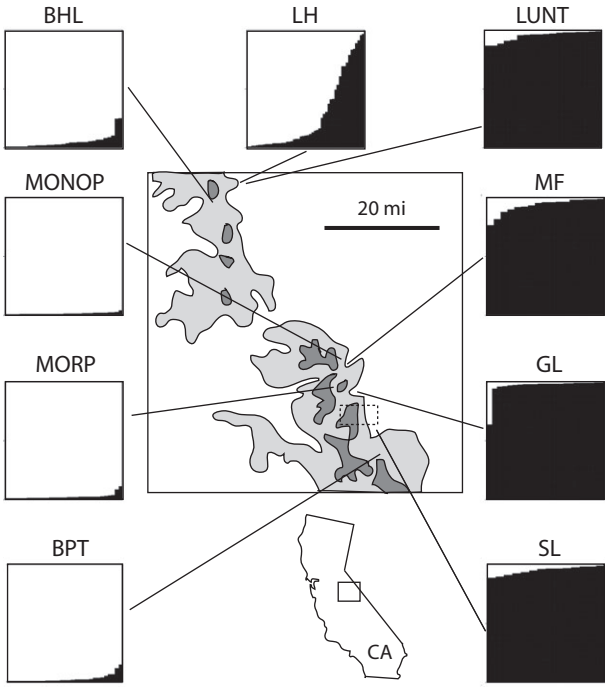

Fig. 1 Genetic structure and geography. Each box represents a population structure assignment ($K = 2$) to either species (*A. pubscens* white area, *A. formosa* black area). The map of California is shown and study area expanded in the box. Light-grey-shaded areas are between 10 000 and 12 000 feet while dark grey areas are above 12 000 feet. The dotted box is further expanded in Fig. 3.

Despite the positive ascertainment bias to find highly differentiated SNPs and the strong differentiation identified by Structure, no single SNP locus was completely differentiated between low-elevation *A. formosa* and high-elevation *A. pubescens*. The most differentiated SNP between $k = 2$ groups had an allele frequency difference of only 0.585 (TC14219), and the average difference was just 0.14.

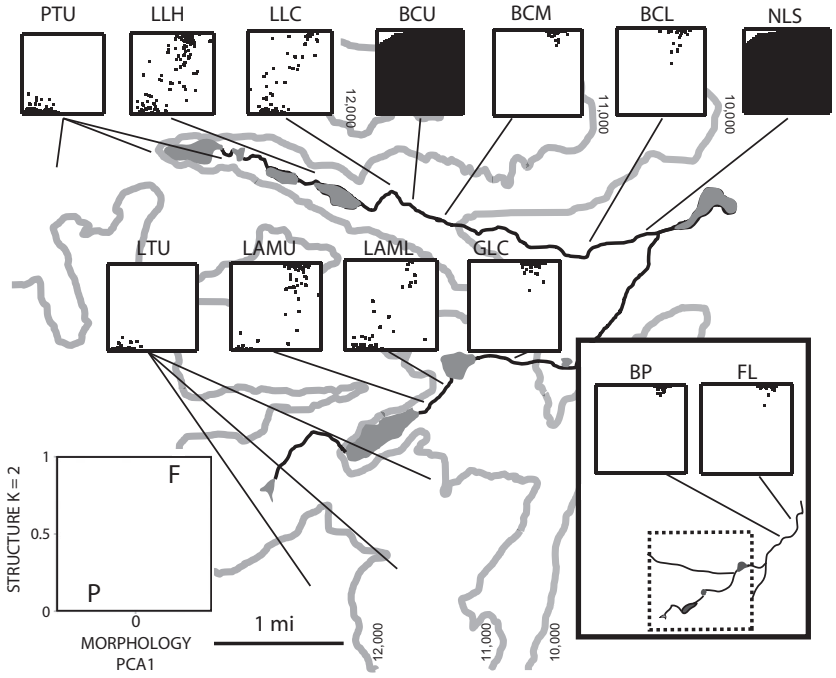

Fig. 2 Morphology, structure and local geography. The lower left exemplar plot illustrates *x*-axis as principal component 1 (PC1) floral morphology and *y*-axis as structure membership ($K = 2$) with *A. pubscens* (P) in the lower left and *A. formosa* (F) upper right, and hybrid plants are points in the middle. The lower right corner shows the larger watershed with BP and FL outlier populations with the location of blow-up in dashed box. Topographic lines are shown in light grey; rivers and lakes are shown in dark grey. The lines from each scatter plot box point to the approximate location of collection sites described in Table 1. Morphology data for BCU and NLS populations are missing, so only structure output is shown.

Isolation by distance

We next analysed the SNP data using PopGraph and found that all populations formed a single highly interconnected network. However, populations classified morphologically as *A. formosa*, *A. pubescens* and hybrids, occupied distinct regions within the network topology (Fig. 3). As expected, the hybrid populations had central positions in the topology largely separating the populations of the two species with links to both *A. formosa* and *A. pubescens* populations. *A. formosa* populations had greater genetic diversity than *A. pubescens* populations (indicated by larger circles in Fig. 3).

Using simple Mantel tests, genetic x geographic distances were always significantly correlated whether using Fst/(1-Fst) or *cGD* as the measure of genetic distance. However, the correlation coefficients for genetic x geographic distance were always higher when controlling for species differences in partial Mantel tests (Table 2) indicating that the morphology of two populations affects their genetic similarity. Furthermore, genetic distance was significantly positively correlated with species distance. These results suggest that there is significant isolation by distance, but this is strongly affected by whether or not the populations are morphologically similar. In addition, genetic and geographic distance correlations were higher when using *cGD* rather than Fst/(1-Fst) and when using log(distance) rather than absolute distance.

We next constructed matrices for each species separately or jointly and found evidence for isolation by distance in each comparison particularly using *cGD* (Table 3). Among populations of *A. pubescens*, the correlation was highest using absolute distance, while for *A. formosa* populations, log(distance) resulted in higher correlations perhaps reflecting differences in dispersal scale. In comparisons between the two species, the correlation was not significant or was much lower than within species. In general, populations of *A. pubescens* have less genetic distance among populations that are physical nearby, but differentiate at larger distances (up to 100 km). This is also true for *A. formosa*, although divergence occurs over a shorter distance (30 km)

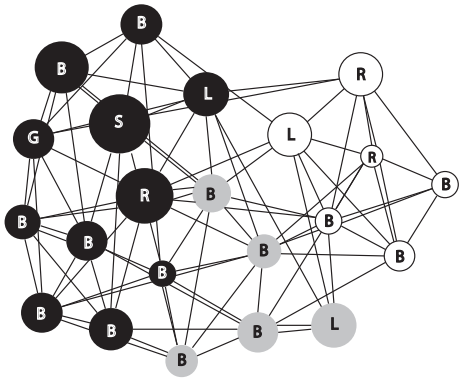

Fig. 3 PopGraph genetic network for all populations, designated by morphology to be *A. formosa* (black), *A. pubescens* (white) or hybrid (grey). Circle size reflects the levels of genetic variation within populations. Lines represent the between-population component of genetic variation due to connecting nodes. Letters within circles indicate the watershed for that population (Table 1).

Table 2 Mantel and partial Mantel results for the entire data set using either Fst/(1-Fst) or *cGD* as the measure of genetic distance (GEN) and physical distance (DIST) and species distance (Spp) as described in the text. For each test, the correlation coefficient (r), the significance level p and 95% confidence intervals are given below.

	Genetic distance measure	
TEST	Fst/(1-Fst)	*cGD*
Mantel		
GEN ~ DIST	0.18, $P = 0.13$ (0.09–0.31)	0.25, $P = 0.03$ (0.18–0.33)
GEN ~ log(DIST)	0.32, $P = 0.01$ (0.22–0.42)	0.42, $P = 0.00$ (0.34–0.49)
GEN ~ Spp	0.75, $P = 0.00$ (0.70–0.78)	0.61, $P = 0.00$ (0.53–0.67)
Partial Mantel		
GEN ~ Spp x DIST	0.25, $P = 0.02$ (0.13–0.35)	0.30, $P = 0.01$ (0.24–0.40)
GEN ~ Spp x log(DIST)	0.36, $P = 0.00$ (0.27–0.46)	0.45, $P = 0.00$ (0.39–0.52)

Table 3 Results of Mantel tests for specific sets of populations using Fst/(1−Fst) or *cGD* as the genetic distance measure as well as absolute distance or log(distance). For each test, the correlation coefficient (r), the two-tailed significance level p and 95% confidence intervals below are given. Significant ($P < 0.05$) correlations are indicated by bold.

		Geography	
Populations compared	Genetic distance measure	Absolute distance	Log(distance)
Formosa x pubescens	Fst/(1-Fst)	−0.052, $P = 0.70$ −0.18 to 0.11	0.04, $P = 0.69$ −0.095 to 0.23
	cGD	0.08, $P = 0.55$ −0.06 to 0.19	0.23, $P = 0.09$ 0.11 to 0.35
Formosa	Fst/(1-Fst)	0.15, $P = 0.59$ 0.03 to 0.40	**0.38, $P = 0.04$ 0.23 to 0.56**
	cGD	**0.41, $P = 0.01$ 0.28 to 0.59**	**0.59, $P = 0.00$ 0.45 to 0.73**
Pubescens	Fst/(1-Fst)	0.41, $P = 0.20$ 0.04 to 0.70	0.48, $P = 0.09$ −0.08 to 0.73
	cGD	**0.86, $P = 0.02$ 0.67 to 0.93**	**0.82, $P = 0.001$ 0.60 to 0.90**

(Fig. 4). However, comparisons between species show higher *cGD* across a range of physical distances indicating species barriers to gene flow even at close distances (Fig. 4).

Discussion

Here we have shown that isolation by distance, measuring historic gene flow, is substantially reduced between the nascent species *A. formosa* and *A. pubescens* despite the presence of advanced hybrid zones and there being no discernible intrinsic reproductive isolation barriers (Chase & Raven 1975; Grant 1952). While we did detect significant IBD between the species, the pattern was weak and populations close together physically were only slightly more similar genetically than those at much further distances. This suggests that while there is some genetic exchange between the species, it is largely confined to areas very close to hybrid zones and does not extend far into either species. Even at close spatial scales, on the order of 2–5 km, populations of the two species are nearly as differentiated as populations at much greater distances. Indeed, analysis with Structure also indicates that clearly admixed individuals are found nearly exclusively at intermediate elevations where morphological hybrids occur. Thus, these taxa represent largely genetically independent populations and are emerging species. This is remarkable given the frequent occurrence of natural hybrid zones and the extensive level of shared polymorphisms, which could be interpreted as indicating extensive gene flow between the nascent species.

Our analysis clearly shows that IBD is weak to nonexistent between the species at spatial scales where it is strong within both species (Fig. 4). This lack of introgression could be caused by intrinsic genetic incompatibilities although none have previously been identified (Chase & Raven 1975; Grant 1952). It is also possible that pollen competition influences the degree that hybrids are formed (e.g. Carney *et al.* 1996; Emms *et al.* 1996). However, the most prominent factors likely to affect interbreeding are the ecological differences between the species, including the stark differences in floral characters with their different attendant pollinators (Fulton & Hodges 1999; Grant 1952) and habitat associations (Chase & Raven 1975; Grant 1952; Hodges & Arnold 1994). As hummingbirds predominantly visit *A. formosa* and hawkmoths predominantly visit *A. pubescens*, gene flow between the species would be reduced compared to within-species gene flow. Any seed resulting from gene flow from the alternate species could also be selected against before becoming established in a population. One way to test for this possibility in the future would be to estimate IBD from seed produced in populations rather than from adult flowering individuals as well as the degree of admixture of seeds, seedlings and adult plants.

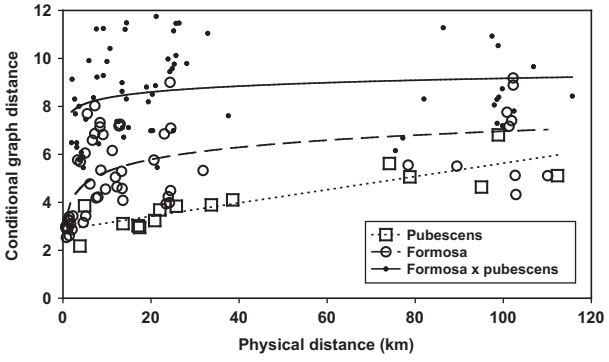

Fig. 4 Genetic isolation by distance. *cGD* is plotted against physical distance for each pair of populations. *A. pubescens* is fit with a linear response while *A. formosa* and hybrids are fit with an exponential log(distance).

Comparisons of IBD between closely related species can also indicate ecological factors that may influence gene flow within species. The two species, *A. formosa* and *A. pubescens*, differed in IBD at similar spatial scales with *A. pubescens* having very strong IBD that was linear across over 100 km while *A. formosa* has a logarithmic IBD at the scale of 20–40 km (Table 3, Fig. 4). Furthermore, there was much greater genetic differentiation among populations of *A. formosa* at the scale of a few to 40 km than among populations of *A. pubescens*. Speculating, we think that these findings may reflect the fact that *A. pubescens* occurs at very high elevations in the southern Sierra Nevada and there are usually contiguous corridors of suitable habitat among populations facilitating gene flow. In contrast, the pattern of IBD found for *A. formosa* may reflect the fact that *A. formosa* is generally found along stream sides between 6000 and 10 000 ft. Gene flow via passive seed dispersal may be restricted along watercourses as seeds may flow downstream resulting in greater isolation among populations in different watersheds. Small mammals, such as pika and marmots, which eat fruits and seeds of *Aquilegia,* may also largely restrict their movements and thus seed dispersal to within watersheds. In addition, *A. formosa* is pollinated primarily by hummingbirds, which often set up feeding territories. These territories may restrict gene flow as compared to hawkmoths, which are the primary pollinator of *A. pubescens* (Fulton & Hodges 1999). Gene flow may be further reduced as suitable habitat between watersheds is separated by unsuitable habitat both at higher and lower elevations, which may restrict gene flow due to both pollen and seed.

Our analysis also supports the contention that utilizing *cGD* instead of Fst/(1-Fst) as the measure of genetic

distance provides a better estimate of IBD (Dyer *et al.* 2010). PopGraph is a relatively new method to assay the distribution of genetic variation on the landscape. Rather than pairwise population comparisons with traditional F_{ST} statistics, populations are clustered using a population graph framework to demonstrate inferred ancestral relatedness and PopGraph has now been utilized in a number of studies (Giordano *et al.* 2007; Kluetch *et al.* 2012; Lowry *et al.* 2008; Murphy *et al.* 2010; Sork *et al.* 2010) . Simulations (Dyer *et al.* 2010) and one other study (Kluetch *et al.* 2012) suggest that *cGD* provides better estimates of IBD because the parameters are estimated from the complete data set rather than in a pairwise fashion when utilizing Fst. We found that utilizing *cGD* often resulted in significant and larger IBD coefficients as compared to Fst/(1-Fst); however, qualitatively, the results are consistent for either genetic distance measure (Tables 2 and 3).

Despite our finding of very limited IBD between *A. formosa* and *A. pubescens*, the two species show remarkable similarity in nucleotide polymorphisms indicating that these two species are at a very early stage in the speciation process (Via 2009). All but one SNP was polymorphic in both species (the SNP in TC14829 had a MAF of 0.12 in *A. formosa,* but this allele was not found in *A. pubescens*). Furthermore, the average allele frequency difference between the species was only 0.14 and the largest allele frequency difference was only 0.57 (the SNP in TC14219, 0.74 in *A. formosa* and 0.17 in *A. pubescens*). These findings are particularly remarkable, as we would have expected ascertainment bias in SNP discovery for markers that completely differentiate the species. Thus, these two species are surprisingly similar with shared genetic markers throughout the genome, even though they have large and distinct morphological and ecological differences. For example, flower colour is essentially a fixed difference between the species caused by a single QTL (Hodges *et al.* 2002; Whittall *et al.* 2006; Hodges & Derieg 2009). The discrepancy between morphological/ecological traits and genetic differentiation suggests that the loci responsible for these fixed phenotypic differences are under strong selection and that the genomic regions responsible for these traits will be detectable in genome scans as peaks of high differentiation (Cooper *et al.* 2010).

Recent studies in *Aquilegia* differ in their detection of IBD. Although the species shared nearly every polymorphism, here we were able to separate *A. formosa* and *A. pubescens* and their hybrids genetically using only 79 SNP markers. In contrast, Cooper *et al.* were unable to distinguish 40 geographically distant individuals of these same species using sequence data from 9 gene loci (containing 85 SNPs and 40 indels). The discrepancy between studies could be due to differences in power including our 20-fold larger and clustered sample and use of ascertained genome-wide markers. Similar to our study, Whittall & Hodges (2007) were able to genetically differentiate 15 *A. formosa* individuals from nine *A. pubescens* individuals using over 1500 polymorphic AFLP markers despite <1% of these being species specific. Populations of *A. coerulea* have been studied for isolation by distance using four microsatellite loci (Brunet *et al.* 2012). Here, populations at 100 km or less were more similar while those beyond this spatial scale do not further diverge. Studies in populations of *A. longissima* and *A. chrysantha* used cpDNA and thus only could assay gene flow due to seed dispersal, but found no evidence for IBD although most of the population pairs were more than 100 km apart (Strand *et al.* 1996). Thus, significant gene flow among populations of the adaptively radiating genus *Aquilegia* may largely be at a local scale of 100 km or less.

Here we have found that *Aquilegia formosa* and *A. pubescens* are remarkably young species in that they share a very high degree of genetic polymorphisms throughout their genomes. Furthermore, despite the presence of numerous hybrid zones, our study suggests that these species largely do not exchange genes and are thus predominantly evolving independently. We also note that we have found that the homogenizing effect of gene flow at and very near morphological hybrid zones is clearly evident (Figs 1 and 2), which suggests that genome-wide association studies of speciation genes in field-collected samples may be possible. We found plants in hybrid zones that were morphologically assigned to one species but had Structure assignments spanning those made to low-elevation *A. formosa* and high-elevation *A. pubescens* (Fig. 2). In these areas, plant phenotypes are largely decoupled from the background population structure that has already built up between the two species. This decoupling is a key to successful genome-wide association studies as demonstrated in Arabidopsis (Li *et al.* 2014), rice, maize (reviewed in Brachi *et al.* 2011) and sorghum (Morris *et al.* 2013). A clear future goal is to identify the genetic basis of the traits that are causing the reduced gene flow between *A. formosa* and *A. pubescens*. Genome scans in morphological hybrids of *Aquilegia* may be able to link genomic variation directly to specific morphological traits and identify speciation genes. Because hybrid zones between *A. formosa* and *A. pubescens* are fairly common at and near the treeline, the large samples necessary for such an analysis are obtainable.

Acknowledgements

The authors thank Brian Counterman and Steve Collie for isolating DNA samples and Nathan Derieg for fruitful

discussions. Mitch Cruzan and two anonymous reviewers provided comments that greatly improved the study. This research was funded through NSF grants DEB-9726272 to SAH and EF-0412727 to SAH and JOB.

References

Abbott R, Albach D, Ansell S *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.

Baack EJ, Rieseberg LH (2007) A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics & Development*, **17**, 513–518.

Barton NH (2000) Estimating multilocus linkage disequilibria. *Heredity*, **84**, 373–389.

Brachi B, Morris GP, Borevitz JO (2011) Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biology*, **12**, 232.

Brunet J, Larson-Rabin Z, Stewart C (2012) The distribution of genetic diversity within and among populations of the rocky mountain columbine: the impact of gene flow, pollinators, and mating system. *International journal of plant sciences*, **173**, 484–494.

Carney SE, Hodges SA, Arnold ML (1996) Effects of differential pollen-tube growth on hybridization in the Louisiana irises. *Evolution*, **50**, 1871–1878.

Chase VC, Raven PH (1975) Evolutionary and ecological relationships between *Aquilegia formosa* and *Aquilegia pubescens*. *Evolution*, **29**, 474–486.

Chenoweth SF, Hughesm JM, Keenan CP, Lavery S (1998) Concordance between dispersal and mitochondrial gene flow: isolation by distance in a tropical teleost, Lates calcarifer (*Australian barramundi*). *Heredity*, **80**, 187–197.

Cooper EA, Whittall JB, Hodges SA, Nordborg M (2010) Genetic variation at nuclear loci fails to distinguish two morphologically distinct species of *Aquilegia*. *PLoS One*, **5**, e8655.

Coyne JA, Orr AH (2004) *Speciation* Sinauer Associates.

Dray S, Dufour AB, Chessel D (2007) The ADE4 package – II: two-table and K-table methods. *R News*, **7**, 47–52.

Dyer R (2009) GeneticStudio: a suite of programs for spatial analysis of genetic-marker data. *Molecular Ecology Resources*, **9**, 110–113.

Dyer RJ, Nason JD (2004) Population Graphs: the graph theoretic shape of genetic structure. *Molecular Ecology*, **13**, 1713–1727.

Dyer RJ, Nason JD, Garrick RC (2010) Landscape modeling of gene flow: improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology*, **19**, 3746–3759.

Emms SK, Hodges SA, Arnold ML (1996) Pollen-tube competition, siring success, and consistent asymmetric hybridization in Louisiana irises. *Evolution*, **50**, 2201–2206.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

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.

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics: TIG*, **28**, 342–350.

Fulton M, Hodges SA (1999) Floral isolation between *Aquilegia formosa* and *Aquilegia pubescens*. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **266**, 2247–2252.

Giordano A, Ridenhour BJ, Storfer A (2007) The influence of altitude and topography on genetic structure in the long-toed salamander (*Ambystoma macrodactulym*). *Molecular Ecology*, **16**, 1627–1637.

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.

Grant V (1952) Isolation and hybridization between *Aquilegia formosa* and *Aquilegia pubescens*. *Aliso*, **2**, 341–360.

Hodges SA, Arnold ML (1994) Floral and ecological isolation between *Aquilegia formosa* and *Aquilegia pubescens*. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 2493–2496.

Hodges SA, Derieg NJ (2009) Adaptive radiations: from field to genomic studies. *Proceedings of the National Academy of Sciences*, **106**, 9947–9954.

Hodges SA, Whittall JB, Fulton M, Yang JY (2002) Genetics of floral traits influencing reproductive isolation between *Aquilegia formosa* and *Aquilegia pubescens*. *American Naturalist*, **159**, S51–S60.

Kluetch CFC, Dyer RJ, Misof B (2012) Combining multiple analytical approaches for the identification of population structure and genetic delineation of two subspecies of the endemic Arabian burnet moth *Reissita simonyi* (Zygaenidae; Lepidoptera). *Conservation Genetics*, **13**, 21–37.

Kramer EM (2009) *Aquilegia*: a new model for plant development, ecology, and evolution. *Annual review of plant biology*, **60**, 261–277.

Kramer EM, Hodges SA (2010) *Aquilegia* as a model system for the evolution and ecology of petals. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **365**, 477–490.

Li Y, Cheng R, Spokas KA, Palmer AA, Borevitz JO (2014) Genetic variation for life history sensitivity to seasonal warming in Arabidopsis thaliana. *Genetics*, **196**, 569–577.

Lowry DB, Rockwood RC, Willis JH (2008) Ecological reproductive isolation of coast and inland races of *Mimulus Guttatus*. *Evolution*, **62**, 2196–2214.

Morris GP, Ramu P, Deshpande SP *et al.* (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences of the United States of America*, **110**(2), 453–458.

Murphy MA, Dezanni R, Pilliod DS, Storfer A (2010) Landscape genetics of high mountain frog metapopulations. *Molecular Ecology*, **19**, 3634–3649.

Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology*, **17**, 2258–2275.

Noutsos C, Borevitz J, Hodges S (2014) Data from: gene flow between nascent species: geographic, genotypic and phenotypic differentiation within and between *Aquilegia formosa* and *A. pubescens*. *Dryad Digital Repository*. doi:10.5061/dryad.sj3vp

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

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Schluter D (2009) Evidence for ecological speciation and its alternative. *Science*, **323**, 737–741.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.
Sork VL, Davis FW, Westfall R *et al.* (2010) Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Nee) in the face of climate change. *Molecular Ecology*, **19**, 3806–3823.
Sotka EE, Palumbi SR (2006) The use of genetic clines to estimate dispersal distances of marine larvae. *Ecology*, **87**, 1094–1103.
Strand AE, Milligan BG, Pruitt CM (1996) Are populations islands? Analysis of chloroplast DNA variation in *Aquilegia*. *Evolution*, **50**, 1822–1829.
Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 9939–9946.
Whittall JB, Hodges SA (2007) Pollinator shifts drive increasingly long nectar spurs in columbine flowers. *Nature*, **447**, 706–709.
Whittall JB, Voelckel C, Kliebenstein D, Hodges SA (2006) Convergence, constraint and the role of regulatory genes during adaptive radiation: floral anthocyanins in *Aquilegia*. *Molecular Ecology*, **15**, 4645–4658.
Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.
Wright S (1946) Isolation by distance under diverse systems of mating. *Genetics*, **31**, 39–59.
Zheng XM, Ge S (2010) Ecological divergence in the presence of gene flow in two closely related Oryza species (*Oryza rufipogon* and *O. nivara*). *Molecular Ecology*, **19**, 2439–2454.

S.A.H., J.O.B. and C.N. designed the experiments. S.A.H. collected the samples and made morphological measurements. C.N. and J.B. performed the S.N.P. assays. C.N., J.O.B. and S.A.H. conducted the analyses and wrote the paper.

Data accessibility

Molecular Markers developed and Genotypes and Phenotypes have been uploaded as DRYAD entries (doi:10.5061/dryad.sj3vp), and the Supplementary figure is uploaded as online supporting material.

Supporting information

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

Fig. S1 Morphological and genetic variation.