Genetic Variation at Nuclear Loci Fails to Distinguish Two Morphologically Distinct Species of *Aquilegia*

Elizabeth A. Cooper[1]*, Justen B. Whittall[2], Scott A. Hodges[3], Magnus Nordborg[4]

1 Department of Molecular and Computational Biology, University of Southern California, Los Angeles, California, United States of America, 2 Biology Department, Santa Clara University, Santa Clara, California, United States of America, 3 Department of Ecology, Evolution, and Marine Biology, University of California Santa Barbara, Santa Barbara, California, United States of America, 4 Gregor Mendel Institute of Molecular Plant Biology, Vienna, Austria

Abstract

Aquilegia formosa and *pubescens* are two closely related species belonging to the columbine genus. Despite their morphological and ecological differences, previous studies have revealed a large degree of intercompatibility, as well as little sequence divergence between these two taxa [1,2]. We compared the inter- and intraspecific patterns of variation for 9 nuclear loci, and found that the two species were practically indistinguishable at the level of DNA sequence polymorphism, indicating either very recent speciation or continued gene flow. As a comparison, we also analyzed variation at two loci across 30 other *Aquilegia* taxa; this revealed slightly more differentiation among taxa, which seemed best explained by geographic distance. By contrast, we found no evidence for isolation by distance on a more local geographic scale. We conclude that the extremely low levels of genetic differentiation between *A. formosa* and *A.pubescens* at neutral loci will facilitate future genome-wide scans for speciation genes.

Introduction

The genetic mechanisms underlying the process of speciation are of critical interest to evolutionary biologists. In order to unravel this process, it is necessary to both identify the genes responsible for existing reproductive barriers and to consider what demographic and selective forces have shaped these traits. In particular, many recent studies have focused on the role of gene flow during the speciation process [3–10], even though the more traditional (allopatric) view of speciation posits that genetic exchange must be rare in order for species to remain distinct [11]. These studies have shown that adaptive differences between species can be maintained even in the face of significant amounts of introgression, especially if only a few genes or genomic regions control the traits that lead to reproductive isolation [4]. Genome-wide analyses of many species have shown that levels of introgression can vary across the genome, with divergent selection playing an active role in preventing gene flow at the loci underlying adaptive traits, but not acting at other areas in the genome [3,4,12]. Incipient species will also show varying levels of differentiation across the genome, with the most differentiated regions also being the most likely to contain genes that restrict random mating [3,7]. These species can appear almost identical at many loci, even in the complete absence of genetic exchange.

Whether or not gene flow is a factor, closely related taxa offer an excellent opportunity to study the genetic changes and processes that lead to reproductive isolation, since genome-wide scans should be able to pinpoint loci with higher levels of differentiation, and these loci are most likely to be under the influence of natural selection [3,13–15]. While the identification of potential speciation genes will not definitively prove a particular speciation model, comparing the pattern of variation in these loci with the pattern of shared variation in neutral loci will provide much more insight into the question of whether or not two species have diverged in the face of gene flow [3,7]. Thus, it is especially important to identify pairs or groups of species that maintain high levels of shared polymorphism over much of their genomes.

The columbine genus *Aquilegia* [Ranunculaceae] is an excellent example of a recent, rapid adaptive radiation [1], and thus should provide an opportunity to identify the genetic changes important for speciation. The genus is comprised of approximately 70 outcrossing species that occupy a wide variety of habitats in North America, Europe, and Asia [16] and that differ substantially in floral morphology [16,17]. Despite these differences, species are usually cross-compatible [18,19].

Two species, *Aquilegia formosa* and *A. pubescens*, have long been studied for the purpose of understanding the factors controlling reproductive isolation between them [20–23]. *A. formosa* is found throughout mountainous regions of western North America while *A. pubescens* is restricted to the southern Sierra Nevada range [22]. The species exhibit distinct differences in floral characters that have been shown to influence pollinator preference, thereby restricting gene flow between them [22,23](Figure 1). Additionally, they prefer different habitats: *A. formosa* populations typically occur in moist areas with well-developed soils at lower elevations (below 3,000m), whereas *A. pubescens* populations are found in drier, poorly developed soils at higher elevations (3,000–4,000m) [20,21,24]. However, the two species are highly interfertile, and

Figure 1. Striking differences in floral morphology between (A) the hummingbird pollinated *Aquilegia formosa* **and (B) the hawkmoth pollinated** *Aquilegia pubescens*.
doi:10.1371/journal.pone.0008655.g001

form natural hybrid zones at mid elevations where the two habitats co-occur [22]. Molecular markers exhibit more introgression than morphological characters near these zones, suggesting that gene flow could be extensive between these species for neutral markers [22].

Previous studies have uncovered limited DNA sequence variation between *A. formosa* and *A. pubescens* in both chloroplast and nuclear sequences [1,2]. However, these previous studies showed either low sequence variation across a wide range of *Aquilegia* species [1] or few individuals were sampled [2] and therefore do not address the degree of genetic differentiation between these species. Other studies suggest that intraspecific sequence variation may be quite similar in *A. formosa* and *A. pubescens* and thus that they may be especially useful for identifying speciation genes. For instance, microsatellite loci have similar numbers of alleles and size ranges [25], and another study including over 850 AFLP markers polymorphic in a small sample of both species found only one marker that showed complete differentiation [17]. Because these previous studies did not assess variation at the DNA sequence level or use relatively large population samples, we sought to gain insight into the inter- and intraspecific patterns of genetic variation in these species by sequencing nine nuclear loci from a total of 80 individuals from several populations. As a comparison, we also assessed variation among all of the North American species in the genus (plus some Eurasian taxa) by sequencing two of the nine nuclear loci. By

examining loci that are not believed to be involved in the maintenance of reproductive isolation, we sought to assess the potential of using genome-wide scans for speciation genes in these species by determining levels of neutral variation and population structure.

Results

Polymorphism Levels and Linkage Disequilibrium

The counts of segregating sites found in each fragment are given in Table 1. Estimates of both θ_W and Π generally fell in the range of 0.004 to 0.006 per base pair (Figure 2). Overall, these estimates are slightly lower than estimates of θ in other outcrossing plant species such as maize ($\theta \approx 0.0096$) [26] and sunflowers ($\theta \approx 0.0094$) [27], similar to estimates in the model species *Arabidopsis thaliana* [28], and higher than estimates found in soybeans ($\theta \approx 0.00097$) [29]. However, values of θ and Π varied across the 9 fragments, and one of them, *UF3GT*, was substantially more polymorphic than the others, with a value of θ between 0.01 and 0.02 (Figure 2). The estimates of θ for all fragments are strikingly correlated across species; as we shall demonstrate in the next section, this is because almost all variation is shared.

Linkage disequilibrium (LD) was not extensive in any of the 9 regions that were sequenced, with average r^2 values ranging between 0.1 and 0.2 for most of the fragments. When values of r^2 are plotted against physical distance between SNPs, the relationship is weak (Figure S7). The fragments with the highest levels of polymorphism show evidence for a rapid decay of LD (within about 1 kb or less). The combined fragment data show low LD values overall, and our estimate of ρ was 0.009, which is higher than estimates of ρ in humans [30], suggesting a relatively high rate of recombination in these species (Table 2).

Genetic Differentiation

When we compared the minor allele counts for each species, we found that few high frequency SNPs corresponded to species-specific polymorphisms (Figure 3 and Table 1). In fact, at sites with a minor allele frequency $>5\%$, there were more than twice as many shared polymorphisms (61) as species-specific polymorphisms (24). We found no fixed differences in any of the 9 sequences.

Table 1. Polymorphism Counts for Each Fragment For each count in bold, the number of sites represents the number of SNPs plus the number of indels treated as single SNPs. Numbers in parentheses represent the number of sites with a Minor Allele Frequency (MAF) $>5\%$ and $>10\%$, respectively.

Fragment	Total S_n	Indels	A. formosa Exclusive	A. pubescens Exclusive	Shared	Fixed Differences
Acetyl	**5** (2, 2)	0	**1** (0, 0)	**2** (0, 0)	**2** (2, 2)	0
DEFEN	**20** (8, 6)	3	**9** (1,0)	**4** (0, 0)	**7** (7, 6)	0
Gapc	**44** (14, 8)	14	**18** (1,0)	**11** (0,0)	**15** (13, 8)	0
H3	**18** (7, 2)	2	**5** (0, 0)	**5** (1, 0)	**8** (6, 2)	0
Heat	**20** (7, 4)	1	**7** (0, 0)	**9** (4, 1)	**4** (3, 3)	0
AP3	**29** (16, 12)	10	**10** (4, 3)	**11** (4, 2)	**8** (8, 7)	0
LFY	**12** (4, 2)	4	**6** (2, 0)	**3** (0, 0)	**3** (2, 2)	0
Pist	**15** (8, 4)	2	**10** (4, 1)	**2** (1, 0)	**3** (3, 3)	0
UF3GT	**34** (19, 16)	4	**9** (2, 0)	**6** (0, 0)	**19** (17, 16)	0

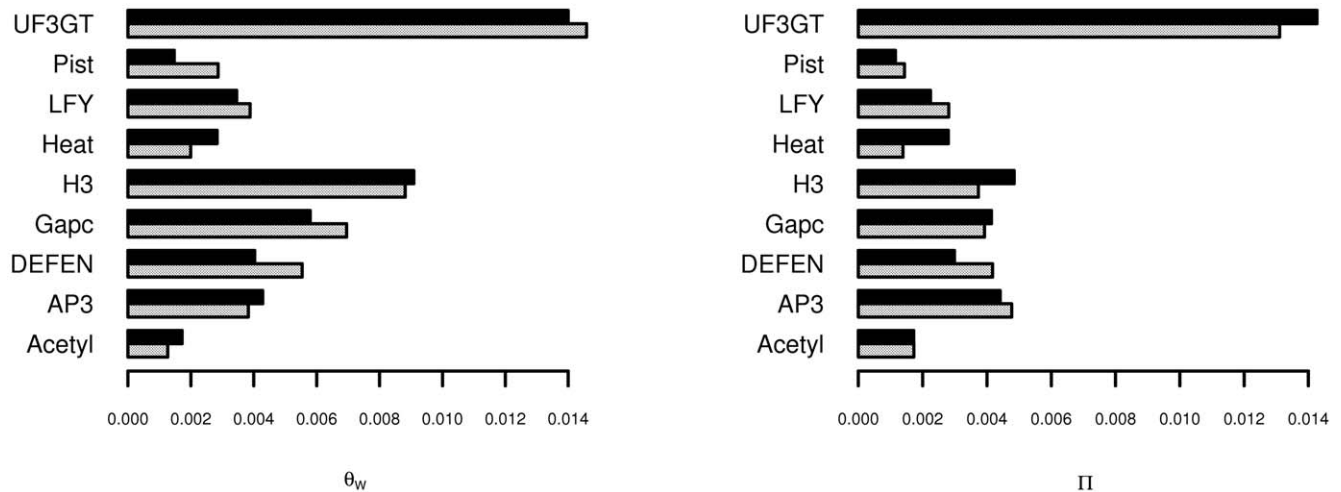doi:10.1371/journal.pone.0008655.t001

Figure 2. Variation among levels of polymorphism for each species. The sequences of each of the 9 fragments were grouped according to species, and θ_W and Π were estimated separately for each group. Grey bars represent estimates for *A. formosa*, and black bars represent *A. pubescens*.
doi:10.1371/journal.pone.0008655.g002

STRUCTURE was unable to cluster individuals according to species under the naïve assumption of K = 2 (Figure S1). The most likely number of clusters appeared to be around 11 (Figure S2), based on when the estimated probability and the average clusteredness stopped (consistently) increasing (Figures S3 and S4). Although the pattern of clustering does not correspond perfectly to the sample populations, it does not seem to be entirely random, especially among the more well-defined clusters (where individuals tend to have membership coefficients >0.5). We found that pairs of individuals from the same population tended to cluster together $\approx 15\%$ of the time, whereas pairs of individuals from different populations only clustered together $\approx 9.5\%$ of the time ($p = 0.0035$ in χ^2 test). Similarly, $\approx 11\%$ of same-species pairs were found in the same cluster, whereas only $\approx 8\%$ of different-species pairs were clustered together ($p = 0.03$).

Average F_{ST} between the two species was approximately 0.0388 (with 95% C.I. between 0.0383 and 0.0393), which is low, but statistically different from zero. When populations were randomly assigned to 2 groups (regardless of species), we achieved very similar results: a mean F_{ST} of 0.0349 with a 95% C.I. between 0.0344 and 0.0353. Although these estimates are technically

statistically different, they do not suggest that much of the observed differentiation is due to species differences.

In order to determine whether such a high degree of shared polymorphism was common in the *Aquilegia* genus or unique to *A. formosa* and *A. pubescens*, we also calculated F_{ST} in a broader sample of 32 taxa using two gene regions (*Gapc* and *UF3GT*). We estimated F_{ST} between pairs of populations and obtained a mean estimate of 0.247. Because the sample of 32 taxa encompassed a broader geographical range, we tested the relationship between geographic distance and genetic differentiation across pairs of populations (Figure 4). Results of the Mantel test indicated that
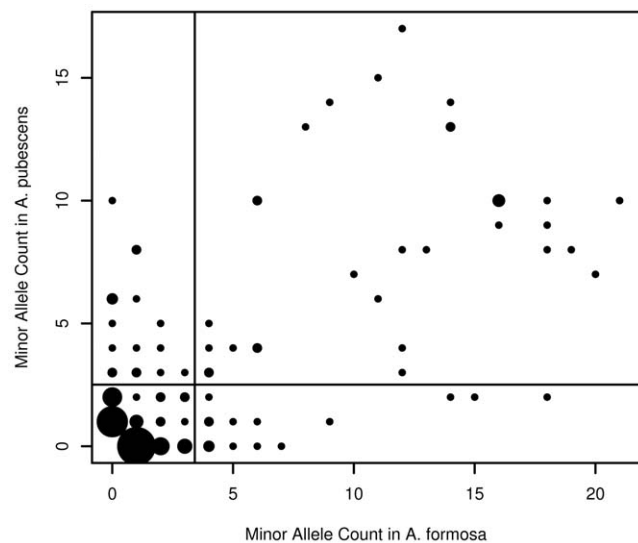
Table 2. Estimates of recombination rate for each fragment.

Fragment Name	R_M	ρ
Acetyl	1	0.081
AP3	11	0.003
Defen	8	0.133
Gapc	12	0.004
H3	3	0.00
Heat	6	0.126
LFY	4	0.006
Pist	4	0.001
UF3GT	15	0.023
Combined		0.009

doi:10.1371/journal.pone.0008655.t002

Figure 3. Comparison of minor allele counts in *A. formosa* and *A.pubescens*. The horizontal line represents the mean allele count in *A. pubescens*, while the vertical line represents the mean allele count in *A. formosa*. Point size reflects the number of comparisons at that point. Species-specific polymorphisms correspond to the points along either the very bottom or the far left of the plot. All other sites correspond to a shared polymorphism. There are no fixed differences. The average minor allele frequency for any species-specific polymorphism was 0.105, while the average frequency for any shared allele was 0.424.
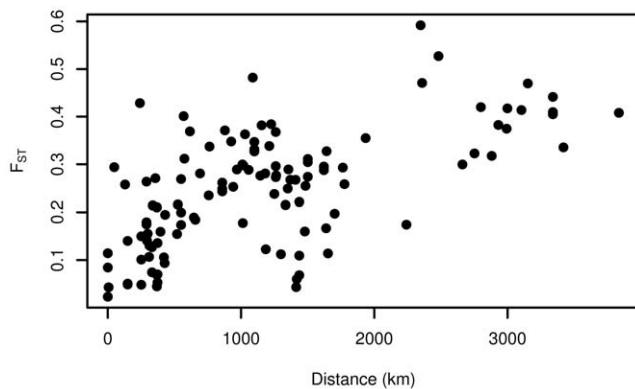doi:10.1371/journal.pone.0008655.g003

Figure 4. Relationship between geographic distance and genetic distance. Each dot represents a comparison between 2 populations of at least 5 individuals. For populations where there were more than 5 individuals, estimates of F_{ST} were bootstrapped to ensure that the larger sample size did not cause any bias in the estimate.
doi:10.1371/journal.pone.0008655.g004

geographic distance had a significant relationship with genetic differentiation within the *Aquilegia* genus ($r = 0.619$, $P = 0.0008$)-more so than any of the other factors we examined (Figure S5). However, on a more local scale, we do not find evidence for isolation by distance either within on between *A.formosa* or *A. pubescens* populations ($r = 0.171$, $P = 0.307$) (Figure 5).

Isolation-Migration

When MIMAR was run with the migration rate fixed at 0, the time since the split between *A. formosa* and *A. pubescens* is estimated as approximately 0.062 in coalescent time units (Figure S6). If we assume the mutation rate to be 6×10^{-9}, then this is equivalent to 55,784.5 generations. If the actual mutation rate in *Aquilegia* is higher than we assumed, then the estimated number of generations since the split will be lower, and if the actual mutation rate is lower, than the number of generations will be higher. The generation time in *Aquilegia* is not known, but a very rough estimate can be calculated as 10 years, based on the observation that the plants seem to produce seeds in the wild for about 20 years. If we assume the generation time is around 10 years, then
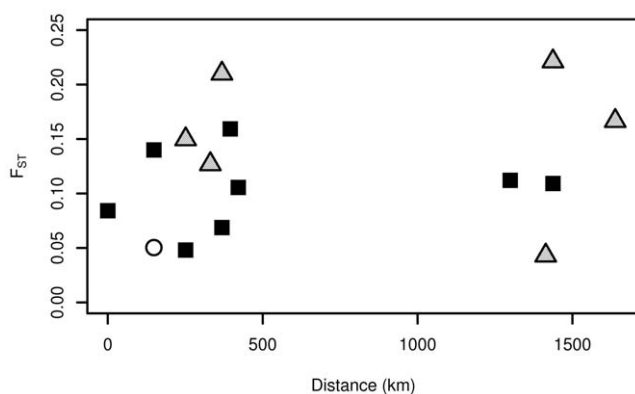

Figure 5. Relationship between geographic distance and genetic distance for *A. formosa* and *A.pubescens* only. Black squares represent comparisons between *A. formosa* and *A. pubescens* populations; gray triangles are comparisons among populations of *A. formosa*; white circles are comparisons among populations of *A. pubescens*.
doi:10.1371/journal.pone.0008655.g005

the MIMAR results suggest that *A. formosa* and *A. pubescens* diverged 557,845 years ago. When migration was incorporated into the model, the estimate for the time since the split rose slightly (to 660,860 years). Both of these estimates seem reasonable, given that the diversification of the North American *Aquilegia* clade is believed to have occurred less than 2 million years ago [31].

Although we obtained believable estimates of the divergence time, MIMAR was not able to converge on an estimate for the migration rate, despite the fact that the model seemed to be mixing well and the estimates of θ corresponded to our earlier calculations (data not shown). Using Wright's F_{ST}–based estimator of migration rate, we calculated that the average number of migrants between populations per generation ($M = 4N_e m$) was 6. Because the MIMAR analysis suggests that *A. formosa* and *A. pubescens* have diverged recently, it is reasonable to assume that at least some of the shared variation is due to ancestral polymorphism, and is not solely the result of gene flow between the two species. Therefore, this estimate of 6 migrants per generation should be considered as a maximum possible value for M.

Discussion

We used direct sequencing to compare levels of intra- and interspecific variation in *Aquilegia*, and found that our genetic data could not distinguish *A. formosa* and *A. pubescens*. Not only were values of θ strikingly similar across species for every fragment, but estimates of F_{ST} were also extremely low, indicating that almost all polymorphism is shared between species. This is a remarkable finding given that these two species are strongly differentiated both ecologically and morphologically.

Several studies of other species have uncovered the same phenomenon. Different species of wild sunflowers exhibit strong ecological differentiation, but it has been found that there are few fixed differences between the species, despite very high levels of intraspecific variation (higher than what we observed in *Aquilegia*) [4]. Hybridization also occurs between these species, and there is evidence for long-term introgression since their divergence one million years ago [4]. Gene flow has also played a role in shaping the patterns of genetic divergence among species in the Hawaiian silversword alliance, which (like *Aquilegia*) is another example of an adaptive radiation in plants [32]. Finally, African cichlid fishes represent one of the most dramatic examples of an adaptive radiation, and many of the more than 2,000 unique species in this group have arisen via sympatric speciation and are still capable of forming viable hybrid offspring, despite many ecological, mor-phological, and behavioral differences [33].

As in the above examples, it is known that hybrid zones form between *A. formosa* and *A. pubescens* [20–22]. There are also some genetic markers which suggest introgression beyond the hybrid zones [22], which makes it tempting to speculate that gene flow between the species has been occurring since their divergence. Our implementation of the isolation-migration model [34,35] produced an estimate of the divergence time that fit well with the model of recent speciation, but since it could not simultaneously converge on an estimate for the migration rate, we cant be sure that gene flow is still occurring. This may be the result of too little data in general, or it may also be the result of having zero fixed differences in the sample.

The patterns of population structure were also unclear in our sample; geographic distance between populations has a clear correlation with genetic differentiation in the broad sample of North American *Aquilegia* taxa, but there is not a clear relationship when only *A. formosa* and *A. pubescens* are examined on a more local scale. At the same time, the clustering of individuals in

STRUCTURE does not seem entirely random, with two individuals being more likely to cluster together if they are from the same species and the same population than if they are not. It is possible that these results are a reflection of a pre-exisiting population structure in the common ancestor, or that migration between populations has made the structure harder to discern.

We believe that finding the loci responsible for reproductive isolation will help us to gain a clearer understanding of how speciation has occurred in *Aquilegia*. A relatively recent scan of genome-wide patterns of interspecific differentiation in two species of European oaks led to the identification of a few genomic regions which seem to underlie species divergence [36]. Like *Aquilegia*, these oak species were closely related and highly interfertile, despite exhibiting significant differences in ecology and morphology. The overall low levels of interspecific variation in these species facilitated the identification of highly differentiated regions. The primary goal of this study was to assess the feasibility of a similar type of genome-wide scan for highly differentiated loci in *Aquilegia*. Our results have shown that despite reasonable levels of intraspecific polymorphism, genetic differentiation is incredibly low at neutral loci, which should make it easier to distinguish putative speciation genes.

Materials and Methods

Sample Collection and Preparation

Leaf tissue was collected from individual plants found in different locations along the west coast of North America. Samples were taken from 40 individuals of each species, for a total sample size of 80 individuals. *A. formosa* samples were taken from 9 different populations, ranging from California, Nevada, Washington state, British Columbia, and Alaska. The number of individuals in each of these populations varied between 1 and 10, but most populations had 5 individuals. There were only 3 populations of *A. pubescens*, and all of them were from California. There were between 4 and 16 individuals in each of these populations (see also Table S1 for a description of the sampling). Because the *A. pubescens* populations were less geographically dispersed than the *A. formosa* samples, there was some concern that *A. pubescens* might falsely appear to be less polymorphic than *A. formosa*. However, as was discussed in the Results section, the same level of polymorphism was found in both species, so sampling bias was not an issue.

DNA extractions were performed using Qiagen's DNeasy Plant Mini Extraction Kits. Due to limited sample amounts, extracted DNA was used directly in only 5 out of the 9 amplifications (*Acetyl*, *Defen*, *H3*, *LFY*, and *UF3GT*). For the remaining 4 amplifications, the extracted DNA was first amplified using Qiagen's REPLI-g Mini Kit and corresponding whole genome amplification protocol.

Additional leaves were collected from thirty-two *Aquilegia* taxa (including *A. formosa* and *A. pubescens*) [17]. Twenty-five of these are also native to North America, while the remaining 7 are found in Europe and Asia. For each species, between 1 and 3 populations were sampled, with an average of 5 individuals per population (Table S1). The majority of individuals came from western North America. DNA extractions were performed as described above.

Fragment Amplification and Sequencing

Nine short regions of the *Aquilegia* genome were amplified in the original sample via PCR using 3′-UTR anchored primers (Table S2). These primers were originally designed by Whittall et al. [2] to reconstruct a species-level phylogeny for several members of the *Aquilegia* genus (including *A. formosa*, but excluding *A. pubescens*). None of these regions are expected to be involved in the evolution

of reproductive barriers. Two of the 9 regions were also amplified in the broader sample of 32 species (*Gapc* and *UF3GT*). All of the sequences contained some non-exonic DNA (Table S3).

All PCR amplifications were done in a total volume of $25 \mu L$, with $20 \mu L$ Promega PCR Master Mix ($2 \times$: 50 units per mL of *Taq* polymerase, $400 \mu M$ dATP, $400 \mu M$ dGTP, $400 \mu M$ dCTP, $400 \mu M$ dTTP, 3mM $MgCl_2$), $3 \mu L$ of forward and reverse primers ($10 \mu M$ each), and approximately 20 ng of DNA template. Although the annealing temperature varied slightly among primer pairs, the cycling conditions were generally as follows: $92°C$ for 2 minutes, followed by 35 cycles of: $92°C$ for 45 seconds, $61°C$ for 30 seconds, $72°C$ for 1.5 minutes, and a final extension step at $72°C$ for 10 minutes.

Sequencing for the original sample of 80 individuals was performed in both directions using the Beckman-Coulter CEQ 2000 platform. Purifications and sequencing reactions were all done as recommended by the Beckman-Coulter protocols. PCR products were purified using Promega's Wizard MagneSil PCR Clean-Up System. Eight microliters of purified template were mixed with $1 \mu L$ CEQ $10 \times$ Buffer, $1 \mu L$ CEQ QuickStart Mix, $2.8 \mu L$ water, and $0.25 \mu L$ of either forward or reverse primer (for a total reaction volume of $13 \mu L$). The sequencing reaction mixtures were then subjected to the following cycling conditions: $96°C$ for 20 seconds, $50°C$ for 20 seconds, and $60°C$ for 4 minutes for a total of 40 cycles, followed by holding at $4°C$. The reaction products were cleaned up using the Beckman-Coulter protocol for "Ethanol Plate Precipitation in a CEQ sample plate," and then finally loaded into the CEQ 2000 for sequencing. Sequencing for the broader sample was performed on the Li-Cor System.

Sequence Alignment and Editing

Sequences obtained from the CEQ 2000 were aligned using phredPhrap [37,38], and visualized in Consed [39]. All alignments were edited manually with the aid of MABCW (program written by T. Hu; scripts and more information available upon request). The indel polymorphisms that we were able to identify were all relatively short, and we only observed two alleles at each of these sites. We were not able to characterize individuals that were heterozygous at these sites, and treated these sequences as missing data during our analyses. For homozygous individuals, indels were analyzed as biallelic SNPs.

For each fragment, the set of segregating sites was identified using alignments of all sequences from both species. The sites in this set were then subsequently characterized as either exclusive to one species or shared based on whether or not they were still segregating in an alignment of sequences from only one species. At each SNP position, the derived allele was determined by using a draft assembly of the *Aquilegia coerulea* (Goldsmith) genome as an outgroup (Joint Genome Institute (JGI) *Aquilegia* Sequencing Project, unpublished data).

For the purpose of linkage disequilibrium analyses, haplotypes were reconstructed using PHASE 2.0.2 [40,41]. For all other analyses, (estimation of θ, F_{ST}, MIMAR, and population structure), we used the un-phased genotype data directly.

Analysis

The population mutation parameter (θ) was estimated using Watterson's estimator (θ_W) [42] and the average number of pairwise differences (Π) [43]. Using in-house scripts (available upon request), both of these statistics were determined for each of the 9 sets of sequences and then scaled by the length of the sequence in order to get a per base pair value. The reading frame for each fragment was assumed based on alignment with cDNA sequences available in Genbank (accession numbers: DQ286961,

DQ224264, DQ224271, DQ217409, DQ286960, DQ224258, AY162852, and DQ286959). Estimates of θ for different classes of sites were scaled by the total number of silent sites or nonsynonymous sites in each sequence (see Table S4 for results). The number of silent sites (S) and the number of nonsynonmous sites (N) were calculated based on a simple Jukes-Cantor model of substitution [44], with the following equations: $S = \frac{L_2}{3} + L_4$, $N = \frac{2L_2}{3} + L_0$, where L_0 is the number of non-degenerate sites, L_2 is the number of twofold degenerate sites, and L_4 is the number of fourfold degenerate sites.

Linkage disequilibrium (LD) between SNPs was quantified using r^2, the squared correlation coefficient. For each fragment, r^2 was plotted as a function of the distance between SNPs (measured in base pairs). The population recombination parameter (ρ) was estimated by fitting the equation given in [45,46]:

$$E(r^2) = \frac{10+\Gamma}{22+13\Gamma+\Gamma^2} \times \left[1 + \frac{(3+\Gamma)(12+12\Gamma+\Gamma^2)}{n(22+13\Gamma+\Gamma^2)^2}\right]$$

where $\Gamma = \rho*(\text{distance})$. For all analyses of recombination, low frequency (MAF $\leq 10\%$) polymorphisms were removed, since they provide little information about the overall pattern of LD.

Estimates of Wright's F_{ST} were calculated based on estimates of Π [47] using the following equation:

$$F_{ST} = \frac{\Pi_{between} - \Pi_{within}}{\Pi_{between}}$$

where $\Pi_{between}$ refers to the average pairwise difference between individuals from different species, and Π_{within} is the average pairwise difference within species. Confidence intervals were obtained by using 10,000 bootstrap replicates.

Analysis of Isolation By Distance (IBD) was performed using a Mantel test [48] with 10,000 replications as implemented by the R package "ade4" [49,50]. The genetic distance matrix was composed of estimates for F_{ST} while the geographic distance matrix was measured in kilometers between populations.

Population structure was inferred directly from the sequence data using the program STRUCTURE 2.0, which implements a model-based clustering approach [51]. STRUCTURE was run under the "linkage model" with "correlated allele frequencies." Specifying correlated allele frequencies enhances the ability of the algorithm to detect distinct clusters even among a sample of very closely related populations [52], which is well suited to the *Aquilegia* data set. Although geographic sampling information was available, initial STRUCTURE runs suggested that geographic location did not correspond well with the genetic data, so we did not use the "prior population information" model to assist in clustering. The program was run with a burn-in length of 50,000 and a run length of 20,000. This was done several times for each K value (ranging from 2 to 15) in order to ensure that results were consistent. Plots of the STRUCTURE output were generated using *distruct* [53]. The average "clusteredness" of individuals was calculated for each STRUCTURE run according to the equation presented by Rosenberg et al. [52].

In order to estimate divergence time and migration rate, the data were analyzed using the program MIMAR [35], which can incorporate recombination into an "isolation–migration" model. The mutation rate, μ, was assumed to be 6×10^{-9}, based on an estimation of the average substitution rate in nuclear DNA in plants [54]. The intralocus recombination rate was set at $r = 0.009$, based on the estimation of the population recombination rate from

linkage disequilibrium data. θ_1, θ_2, and θ_A were all sampled from a uniform prior distribution $U[0.002,0.009]$. The time since split, T, measured in generations, was sampled from the prior distribution $U[0,100000]$. Migration was either fixed at 0, or drawn from a prior range between 0.135 and 7.39. The program was run for 1.1×10^5 recorded steps, and 1×10^4 burnin steps.

We also estimated the migration rate using Wright's equation [55] for an n-island population model, which is based on F_{ST}:

$$F_{ST} = \frac{1}{1 + 4N_e m}$$

Supporting Information

Figure S1 STRUCTURE cannot cluster individuals according to species. Each individual is indicated by a thin line, where the two colors represent the estimated membership coefficients for the 2 clusters. The clusteredness score for this plot was estimated as 0.26.
Found at: doi:10.1371/journal.pone.0008655.s001 (0.04 MB PDF)

Figure S2 Inferred population structure for 80 *Aquilegia* individuals. The results from STRUCTURE are plotted for K = 11, which had an average clusteredness score of ≈ 0.52. Each individual is represented by a thin horizontal line, with corresponding population and species information given on either side.
Found at: doi:10.1371/journal.pone.0008655.s002 (0.06 MB PDF)

Figure S3 Probability of different K estimates. The estimated log probability of the data (as calculated by STRUCTURE) is plotted against different K values. For each K value, STRUCTURE was run 3 times, and the plotted value is the average of those 3 runs.
Found at: doi:10.1371/journal.pone.0008655.s003 (0.03 MB PDF)

Figure S4 Average clusteredness for different K values. For each K value, the average clusteredness measures the extent to which each individual belongs to a single cluster rather than to multiple clusters, so the higher the clusteredness the "better" the clusters.
Found at: doi:10.1371/journal.pone.0008655.s004 (0.03 MB PDF)

Figure S5 Other factors influencing F_{ST} in *Aquilegia*. In all panels, red dots indicate comparisons where both populations were the same for the factor being considered, while gray dots indicate comparisons where the two populations were different. Panel (A) shows F_{ST} vs distance both within and between species, with the green diamonds indicating comparisons between either *A. formosa* or *A. pubescens* and one of the natural hybrid populations. Panel (B) shows F_{ST} vs distance with the same and different pollinator syndrome, while Panel (C) shows the same comparisons for habitat type.
Found at: doi:10.1371/journal.pone.0008655.s005 (0.09 MB PDF)

Figure S6 MIMAR estimates of time since divergence
Found at: doi:10.1371/journal.pone.0008655.s006 (0.41 MB PDF)

Figure S7 R^2 versus distance for the combined data.
Found at: doi:10.1371/journal.pone.0008655.s007 (0.07 MB PDF)

Table S1 Summary of *Aquilegia* samples used in this study.
Found at: doi:10.1371/journal.pone.0008655.s008 (0.04 MB PDF)

Table S2 Primer pairs used to amplify the 9 nuclear loci.
Found at: doi:10.1371/journal.pone.0008655.s009 (0.03 MB PDF)

Table S3 Positions of introns, exons, and UTRs in each locus.
Found at: doi:10.1371/journal.pone.0008655.s010 (0.02 MB PDF)

Table S4 Levels of polymorphism for synonymous and non-synonymous sites.
Found at: doi:10.1371/journal.pone.0008655.s011 (0.04 MB PDF)

Author Contributions

Conceived and designed the experiments: EAC JBW SAH MN. Performed the experiments: EAC JBW. Analyzed the data: EAC. Contributed reagents/materials/analysis tools: JBW SAH. Wrote the paper: EAC.

References

1. Hodges S, Arnold M (1994) Columbines: A geographically widespread species ock. Proceedings of the National Academy of Sciences 91: 5129–5132.
2. Whittall JB, Medina-Marino A, Zimmer EA, Hodges SA (2006) Generating single-copy nuclear gene data for a recent adaptive radiation. Molecular Phylogenetics and Evolution 39: 124–134.
3. Hey J (2006) Recent advances in assessing gene ow between diverging populations and species. Current Opinion Genetics and Development 16: 592–596.
4. Strasburg JL, Rieseberg LH (2008) Molecular demographic history of the annual sunowers Helianthus annuus and H. petiolaris-large effective population sizes and rates of long-term gene ow. Evolution 62: 1936–1950.
5. Crow KD, Munehara H, Kanamoto Z, Balanov A, Antonenko D, et al. (2007) Maintenance of species boundaries despite rampant hybridization between three species of reef fishes (Hexagrammidae): implications for the role of selection. Biological Journal of the Linnean Society 91: 135–147.
6. Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH (2007) Rampant gene exchange across a strong reproductive barrier between the annual sunowers, Helianthus annuus and H. petiolaris. Genetics 175: 1883–1893.
7. Butlin RK, Galindo J, Grahame JW (2008) Sympatric, parapatric or allopatric: the most important way to classify speciation? Philosophical Transactions of the Royal Society B: Biological Sciences 363: 2997–3007.
8. Mallet J (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. Philosophical Transactions of the Royal Society B: Biological Sciences 363: 2971–2986.
9. Noor MA, Feder JL (2006) Speciation genetics: evolving approaches. Nature Reviews Genetics 7: 851–861.
10. Seehausen O (2004) Hybridization and adaptive radiation. Trends in Ecology and Evolution 19: 198–207.
11. Coyne JA, Orr HA (2004) Speciation. SunderlandMA U.S.A.: Sinauer Associates, Inc.
12. Turner TL, Hahn MW, Nuzdhin SV (2005) Genomic islands of speciation in Anopheles gambiae. PLoS Biology 3: 1572–1578.
13. Clarke B, Johnson M, Murray J (1996) Clines in the genetic distance between two species of island land snails: how 'molecular leakage' can mislead us about speciation. Philosophical Transactions of the Royal Society B: Biological Sciences 351: 773–784.
14. Wang RL, Wakeley J, Hey J (1997) Gene ow and natural selection in the origin of Drosophila pseudoobscura and close relatives. Genetics 147: 1091–1106.
15. Machado CA, Kliman RM, Markert JA, Hey J (2002) Inferring the history of speciation from multilocus DNA sequence data: The case of Drosophila pseudoobscura and close relatives. Molecular Biology and Evolution 19: 472–488.
16. Munz P (1946) Aquilegia: the wild and cultivated columbines. Gentes Herbarum 7: 1–150.
17. Whittall JB, Hodges SA (2007) Pollinator shifts drive increasingly long nectar spurs in columbine owers. Nature 447: 706–709.
18. Prazmo W (1965) Cytogenetic studies on the genus Aquilegia. IV. fertility relationships among the Aquilegia species. Acta Soc Bot Poloniae 34: 667–685.
19. Taylor R (1967) Interspecific hybridization and its evolutionary significance in the genus Aquilegia. Brittonia 19: 379–390.
20. Grant V (1952) Isolation and hybridization between Aquilegia formosa and A. pubescens. Aliso 2: 341–360.
21. Chase V, Raven P (1975) Evolutionary and ecological relationships between Aquilegia formosa and A.pubescens (Ranunculaceae), two perennial plants. Evolution 29: 474–486.
22. Hodges S, Arnold M (1994) Floral and ecological isolation between Aquilegia formosa and Aquilegia pubescens. Proceedings of the National Academy of Sciences 91: 2493–2496.
23. Fulton M, Hodges S (1999) Floral isolation between Aquilegia formosa and Aquilegia pubescens. Proceedings of the Royal Society B: Biological Sciences 266: 2247–2252.
24. Hodges SA, Fulton M, Yang JY, Whittall JB (2004) Verne grant and evolutionary studies of Aquilegia. New Phytologist 161: 113–120.
25. Yang JY, Counterman BA, Eckert CG, Hodges SA (2005) Cross-species amplification of microsatellite loci in Aquilegia and Semiaquilegia (Ranunculaceae). Molecular Ecology Notes 5: 317–320.
26. Tenaillon M, Sawkins M, Long A, Gaut R, Doebley J, et al. (2001) Patterns of DNA sequence polymorphism along chromosome I of maize (Zea mays ssp. mays L.). Proceedings of the National Academy of Sciences 98: 9161–9166.
27. Kolkman JM, Berry ST, Leon AJ, Slabaugh MB, Tang S, et al. (2007) Single nucleotide polymorphisms and linkage disequilibrium in sunower. Genetics 177: 457–468.
28. Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, et al. (2005) The pattern of polymorphism in Arabidopsis thaliana. PLoS Biology 3: 1289–1299.
29. Zhu Y, Song Q, Hyten D, Tassell CV, Matukumalli L, et al. (2003) Single-nucleotide polymorphisms in soybean. Genetics 163: 1123–1134.
30. Ptak SE, HInds DA, Koehler K, Nickel B, Patil N, et al. (2005) Fine-scale recombination patterns differ between chimpanzees and humans. Nature Genetics 37: 429–434.
31. Kay KM, Whittall JB, Hodges SA (2006) A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. BMC Evolutionary Biology 6.
32. Lawton-Rauh A, Robichaux R, Purugganan M (2007) Diversity and divergence patterns in regulatory genes suggest differential gene ow in recently derived species of the Hawaiian silversword alliance adaptive radiation (Asteraceae). Molecular Ecology 16: 3995–4013.
33. Kocher TD (2004) Adaptive evolution and explosive speciation: The Cichlid fish model. Nature Reviews Genetics 5: 288–298.
34. Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D. persimilis. Genetics 167: 747–760.
35. Becquet C, Przeworski M (2007) A new approach to estimate parameters of speciation models with application to apes. Genome Research 17: 1505–1519.
36. Scotti-Saintagne C, Mariette S, Porth I, Goicoechea PG, Barreneche T, et al. (2004) Genome scanning for interspecific differentiation between two closely related oak species [Quercus robur L. and Q. petrea (Matt.) Liebl.]. Genetics 168: 1615–1626.
37. Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Research 8: 186–194.
38. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Research 8: 175–185.
39. Gordon D, Abajian C, Green P (1998) Consed: A graphical tool for sequence finishing. Genome Research 8: 195–202.
40. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. American Journal of Human Genetics 68: 978–989.
41. Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. American Journal of Human Genetics 73: 1162–1169.
42. Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theoretical Population Biology 7: 256–276.
43. Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105: 437–460.
44. Jukes T, Cantor C (1969) Evolution of protein molecules. New York, chapter: Academic Press. pp 21–123.
45. Weir B, Hill W (1986) Nonuniform recombination with the human beta-globin gene cluster. American Journal of Human Genetics 38: 776–778.
46. Cutter AD, Baird SE, Charlesworth D (2006) High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of Caenorhabditis remanei. Genetics 174: 901–913.
47. Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene ow from DNA—sequence data. Genetics 132: 583–589.

48. Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Research 27: 209–220.

49. Team RDC (2007) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, URL http://www.R-project.org.

50. Dray S, Dufour A (2007) The ade4 package: implementing the duality diagram for ecologists. Journal of Statistical Software 22: 1–20.

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

52. Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, et al. (2005) Clines, clusters, and the effect of study design on the inference of human population structure. PLoS Genetics 1: 660–671.

53. Rosenberg NA (2002) Distruct: A program for the graphical display of structure results. http://rosenberglabbioinformaticsmedumichedu/distructhtml.

54. Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the National Academy of Sciences 84: 9054–9058.

55. Wright S (1951) The genetical structure of populations. Annals of Eugenics 15: 323–354.