MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 521–541 doi: 10.1111/j.1365-294X.2009.04504.x

Genome scan of hybridizing sunflowers from Texas (*Helianthus annuus* and *H. debilis*) reveals asymmetric patterns of introgression and small islands of genomic differentiation

M. SCASCITELLI,* K. D. WHITNEY,† R. A. RANDELL,‡ MATTHEW KING,* C. A. BUERKLE§ and L. H. RIESEBERG*‡

*Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4, †Department of Ecology and Evolutionary Biology, Rice University, Houston, TX 77005, USA, ‡Department of Biology, Indiana University, Bloomington, IN 47405, USA, §Department of Botany, University of Wyoming, Laramie, WY 82071, USA

Abstract

Although the sexual transfer of genetic material between species (i.e. introgression) has been documented in many groups of plants and animals, genome-wide patterns of introgression are poorly understood. Is most of the genome permeable to interspecific gene flow, or is introgression typically restricted to a handful of genomic regions? Here, we assess the genomic extent and direction of introgression between three sunflowers from the south-central USA: the common sunflower, *Helianthus annuus* ssp. *annuus*; a near-endemic to Texas, *Helianthus debilis* ssp. *cucumerifolius*; and their putative hybrid derivative, thought to have recently colonized Texas, *H. annuus* ssp. *texanus*. Analyses of variation at 88 genetically mapped microsatellite loci revealed that long-term migration rates were high, genome-wide and asymmetric, with higher migration rates from *H. annuus texanus* into the two parental taxa than *vice versa*. These results imply a longer history of intermittent contact between *H. debilis* and *H. annuus* than previously believed, and that *H. annuus texanus* may serve as a bridge for the transfer of alleles between its parental taxa. They also contradict recent theory suggesting that introgression should predominantly be in the direction of the colonizing species. As in previous studies of hybridizing sunflower species, regions of genetic differentiation appear small, whether estimated in terms of F_{ST} or unidirectional migration rates. Estimates of recent immigration and admixture were inconsistent, depending on the type of analysis. At the individual locus level, one marker showed striking asymmetry in migration rates, a pattern consistent with tight linkage to a Bateson–Dobzhansky–Muller incompatibility.

Keywords: adaptive introgression, annual sunflowers, interspecific hybridization, selective sweep

Received 16 September 2009; revision received 23 November 2009; accepted 25 November 2009

Introduction

In his classic monograph on introgressive hybridization, Anderson (1949, p. 102) argued that 'The more imperceptible introgression becomes, the greater its biological importance'. This argument makes sense, at least if biological importance is measured in terms of the contribution of introgression to adaptive evolution. Rampant introgression is readily documented, but in most cases it probably reduces the fitnesses of the hybridizing populations and may even contribute to their extinction (Ellstrand 1992; Rieseberg *et al.* 1995; Levin *et al.* 1996; Wolf *et al.* 2001; Muhlfeld *et al.* 2009; Takakura *et al.* 2009). In contrast, the introgression of a smaller number of advantageous alleles, while difficult to detect, should

Correspondence: M. Scascitelli, Fax: 604 822 6089;
E-mail: mscascit@gmail.com

increase population fitness and rates of adaptive evolution (Anderson 1949; Stebbins 1959; Barton 2001). This paradox may contribute to an apparent imbalance in the literature: hybridization is frequently shown to be of conservation concern (Rieseberg *et al.* 1989; Levin *et al.* 1996; Rhymer & Simberloff 1996; Field *et al.* 2009; Genovart 2009; Morgan-Richards *et al.* 2009; O'Brien *et al.* 2009), but examples where it has contributed positively to adaptive evolution (i.e. adaptive introgression) are rare. Even when 'adaptive introgression' is well documented (e.g. Grant & Grant 1996, 2008; Emms & Arnold 1997; Whitney *et al.* 2006; Castric *et al.* 2008; Kim *et al.* 2008; Rieseberg 2009), we know little about its genomic extent. Has only one or a handful of genes or traits introgressed or do we see evidence of introgression across much of the genome? How large are genomic islands of differentiation? Is introgression mainly in the direction of colonizing species as suggested by recent theory (Currat *et al.* 2008) or in the direction of local endemics, as implied by earlier studies discussing the possible conservation consequences of hybridization (Rieseberg 1991; Ellstrand 1992; Levin *et al.* 1996; Rhymer & Simberloff 1996)?

A well-studied example of adaptive trait introgression involves the colonization of Texas by the common sunflower, *Helianthus annuus* (Heiser 1951; Rieseberg *et al.* 1990; Kim & Rieseberg 1999; Whitney *et al.* 2006). The adaptive introgression hypothesis was inspired by observations that individuals of *H. annuus* collected from central and southern Texas were similar to a local endemic species, *Helianthus debilis* var. *cucumerifolius*, in several morphological traits, including deeply serrated leaves, purple mottled stems, basal branching, smaller flowering heads and achenes, fewer ray flowers and earlier flowering (Heiser 1951, 1954). The two species hybridize in Texas leading Heiser (1951) to argue that introgression with *H. debilis* facilitated the southward expansion of *H. annuus* through the provision of locally adapted alleles. Because populations from *H. annuus* in Texas appeared to be exclusive to human-disturbed habitats, Heiser (1951) further suggested that the colonization most probably took place during the Holocene.

While populations of *H. annuus* from southern Texas are morphologically distinctive, they are fully interfertile with other populations of *H. annuus* (Heiser 1951, 1954). Hence, Heiser (1954) considered them to be a subspecies of *H. annuus* (ssp. *texanus* hereafter) rather than a hybrid species *per se*. Populations of *H. annuus*, including those of ssp. *texanus*, are karyotypically divergent from *H. debilis* ssp. *cucumerifolius*, and first generation hybrids average just 3.6–6.7% viable pollen (Heiser 1951). The two putative parental species also appear to have different ecological requirements, with *H. annuus* (ssp. *annuus*) found mostly on moister, clay-based soils whereas *H. debilis* is restricted to drier, sandy soils. Transient hybrid swarms, which vary widely in morphology and fertility, can be found in heavily disturbed sites (Heiser 1951).

Previous molecular analyses are consistent with the adaptive introgression hypothesis. Populations of ssp. *texanus* contain chloroplast and nuclear ribosomal DNA markers diagnostic for *H. debilis* (Rieseberg *et al.* 1990). Genetic analyses of interspecific hybrids identified just two quantitative trait loci (QTLs) for pollen sterility, implying that much of the genome should be permeable to introgression (Kim & Rieseberg 1999). Interestingly, QTLs for phenotypic differences were highly clustered, indicating that the recovery of much of the ssp. *texanus* phenotype could be achieved with the introgression of as few as three small chromosomal segments.

An investigation of amplified fragment length polymorphisms (AFLPs) markers associated with QTLs underlying phenotypic differences between ssp. *annuus* and *H. debilis* (Rieseberg *et al.* 2007) revealed that 12 of 15 *H. debilis* markers assayed exhibited low or neutral levels of introgression. However, three markers were significantly over-represented in populations of ssp. *texanus*, and two of these markers were associated with QTLs underlying morphological traits that varied in the direction of *H. debilis*, as predicted by the adaptive introgression hypothesis.

Finally, analyses of biotic stress response (Whitney *et al.* 2006) showed that ssp. *texanus* had higher fitness, in terms of seed production, than did parental ssp. *annuus*, when grown in the environment of central and southern Texas. For some fitness-related traits, particularly for damage traits because of herbivores common in Texas, the *texanus* phenotype was shifted away from ssp. *annuus*, towards the locally adapted species *H. debilis*. A parallel analysis of BC1 hybrids between ssp. *annuus* and *H. debilis* showed that the same *H. debilis*-like resistance traits were favoured by natural selection. Taken together, these experiments imply that the introgression of biotic resistance traits facilitated that adaptation of *H. annuus* to the environment of central and southern Texas.

In this study, we analyse natural populations of ssp. *annuus*, *H. debilis* and of their putative hybrid, ssp. *texanus*, with 88 microsatellite markers of known genomic location. Our main goals are to determine the genomic extent of introgression between these taxa, as well to analyse in detail the intensity and direction of interspecific gene flow. In addition, we use these data to test a new model of introgression, which predicts that introgression between a colonizing species and a local congener will be asymmetric and predominantly in the direction of the invader (Currat *et al.* 2008). Finally, we ask whether any of the 88 markers, many of which are

derived from the flanking regions of expressed genes, shows evidence for selection in one or more of the taxa analysed.

Materials and methods

Plant collection

Achenes were collected from four or five populations of each of the three taxa targeted by this study: *Helianthus annuus* ssp. *annuus*, *H. annuus* ssp. *texanus* and *Helianthus debilis* (Table 1). The collections cover most of the native range of *H. debilis* var. *cucumerifolius* as well as that of ssp. *texanus* (Fig. 1). For ssp. *annuus*, we sampled one population from Texas as well as three populations from Kansas, Nebraska and Oklahoma. In addition, one population of an allopatric variety of *H. debilis* (var. *debilis*) was collected from Florida. Our rationale for sampling allopatric populations of *H. annuus* and *H. debilis* was to ensure that we had 'pure' parental populations that could provide a baseline for interpreting patterns of genetic variation in the sympatric, hybridizing populations.

Achenes were germinated at Indiana University, and DNA was extracted from the young leaves of a total of 378 individuals using the Qiagen DNeasy 96 Plant Kit.

DNA amplification and genotyping

Microsatellite loci were chosen for genotyping based on two criteria. First, they had to have been previously mapped in a BC1 mapping population between *H. annuus* and *H. debilis* (K. Whitney, R. Randell, & L. Rieseberg, unpublished). Second, DNA fragment profiles had to be clear, unambiguous and easily interpretable as a single locus with co-dominant allelic inheritance. A total of 88 loci satisfied these criteria and were employed for genotyping (Table S1). Loci were either derived from libraries enriched for dinucleotide repeats (Tang *et al.* 2002) or from expressed sequence tag libraries generated for *H. annuus* by the Compositae Genome Project (Heesacker *et al.* 2008). For genotyping in this study, forward primers were directly labelled with a fluorescent dye or indirectly labelled through the addition of a

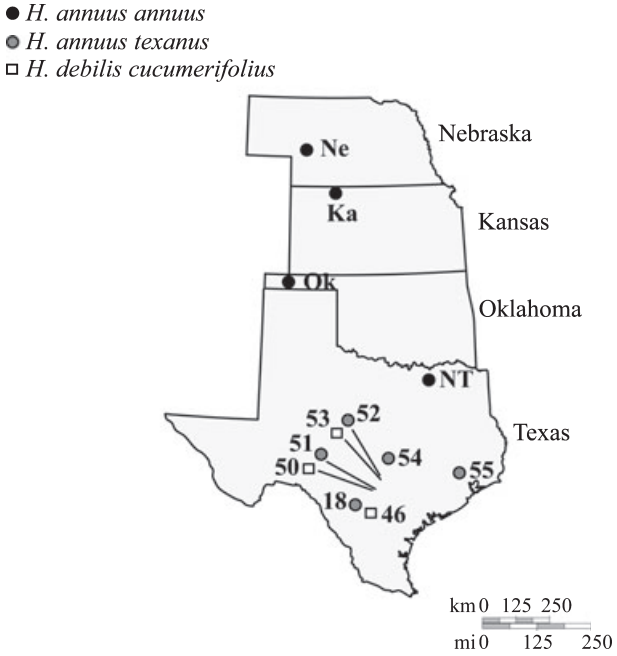

Fig. 1 Collection localities of the annual sunflowers used in this study (see Table 1 for details). Note that a population of *Helianthus debilis debilis* from Florida was also studied, but is not shown here.

Table 1 Collection localities for wild annual sunflowers *Helianthus annuus* and *Helianthus debilis*

Species	Taxon	*n*	State	Population code (previous code)	Latitude, longitude
Helianthus annuus	ssp. *annuus*	16	Nebraska	Ne (LHR 1238)*	41°07′30″N, 101°23′ 24″W
	ssp. *annuus*	16	Oklahoma	Ok (11)†	36°45′36″N 102°09′30″W
	ssp. *annuus*	16	Kansas	Ka (12)†	39°48′48″N 99°55′24″W
	ssp. *annuus*	44	Northern Texas	NT (RAR 59)*	33°31′26.4″N, 96°24′10.8″W
	ssp. *texanus*	30	Texas	18 (RAR 18)*	29°06′07.02″N, 98°56′24″W
	ssp. *texanus*	30	Texas	51 (RAR 51)*	29°26′24″N, 98°02′20.4″W
	ssp. *texanus*	30	Texas	52 (RAR 52)*	29°31′40.08″N, 97°54′32.04″W
	ssp. *texanus*	30	Texas	54 (RAR 54)*	29°50′42″N, 97°31′04.08″W
	ssp. *texanus*	30	Texas	55	29°43′09.02″N, 95°24′20.06″W
Helianthus debilis	*H. debilis cucumerifolius*	30	Texas	46 (RAR 46)*	29°03′21.06″N, 98°16′40.08″W
	H. debilis cucumerifolius	29	Texas	50 (RAR 50)*	29°25′30″N, 98°05′60″W
	H. debilis cucumerifolius	29	Texas	53 (RAR 53)*	29°30′50.04″N, 97°50′16.08″W
	H. debilis debilis	48	Florida	Fl	GRIN collection

*Whitney *et al.* (2006); †Harter *et al.* (2004).

fluorescently labelled M13 adaptor with homology to the 5′ tail of the microsatellite-specific primers (Schuelke 2000).

Polymerase chain reactions (PCR) were performed using a touch-down method (Don *et al.* 1991). Details about PCR conditions for primer pairs indirectly labelled using the M13 adaptor method can be found in Gross *et al.* (2007) and Kane & Rieseberg (2007). Similar conditions were employed for the directly labelled primers, except that primer concentrations were the same for forward and reverse primers (0.25 μM) and the final annealing temperature varied between 51 and 60 °C, according to the specific primer melting temperature.

Amplified microsatellite fragments were analysed with ABI 3730 capillary sequencer (Applied Biosystems) for a subset of individuals from each taxon. Once the size range of each fragment was determined, markers that differed in size or fluorescent labelling were pooled and analysed simultaneously for the rest of the genotyping. The indirect labelling of some gene-specific primers allowed us to vary the fluorescent dye, which facilitated pooling.

Fragments were scored with Genemapper 3.7 (Applied Biosystems) as described elsewhere (Gross *et al.* 2007; Kane & Rieseberg 2007). Allele sizes obtained for each locus were checked with MSA (Dieringer & Schlötterer 2003) and verified (or corrected) if potential errors were detected.

Analyses of population structure

The most likely number of discrete populations, as well as the assignment of individuals to one or more populations, was inferred according to the degree of admixture of their multilocus genotype using STRUCTURE 2.2, a Bayesian clustering method (Pritchard *et al.* 2000; Falush *et al.* 2003). The following parameters/assumptions were employed: (i) the number of discrete populations (K) was allowed to vary between 2 and 13, which is the number of collections analysed; (ii) 50 000 generations of 'burn-in' and 100 000 Markov chain Monte Carlo (MCMC) generations were used for each value of K; and (iii) individuals were assumed to have a mixed ancestry, with correlated allele frequencies among populations. Simulations were repeated three times for each value of K, and the resulting matrices of estimated cluster membership coefficients were permuted with CLUMPP (Jakobsson & Rosenberg 2007) to account for differences between the three runs. The final matrix for each K value was visualized with DISTRUCT (Rosenberg 2004). To assess patterns of introgression at individual loci in relation to their genomic location, we repeated the analyses described above using the 'Site-by-Site' option with a linkage model.

Analyses of long-term migration and effective population size

There are two general classes of methods for inferring migration rates (Faubet *et al.* 2007): (i) coalescent approaches that employ the genealogical data inherent in some kinds of molecular markers such as DNA sequence data or microsatellites, and (ii) multilocus genotypic methods that are based on patterns of gametic disequilibrium. Coalescent methods provide long-term estimates of migration rates and other migration parameters, whereas multilocus genotypic methods provide short-term estimates for many of the same parameters.

To estimate long-term migration rates and effective population sizes, we employed the computer program MIGRATE-N 2.4.3 (http://popgen.sc.fsu.edu/Migrate-n.html). MIGRATE-N uses maximum-likelihood or Bayesian inference to jointly estimate migration parameters under the coalescent model (Beerli & Felsenstein 1999, 2001; Beerli 2004). Values of M ($M = m/\mu$, where m is the migration rate and μ is mutation rate per site per generation) describe the impact of immigration relative to mutation in introducing new variants into a population or population group as identified by STRUCTURE. MIGRATE-N also estimates θ ($\theta = 4N_e\mu$), the mutation-scaled effective population size. A Metropolis-Hastings algorithm was used to explore genealogies and estimate demographic parameters. This allowed us to account for shared ancestry and migration in the population's history as well as to more accurately quantify the coalescent effective population size (Beerli 2006). The MIGRATE-N method is more appropriate in this case than the isolation with migration approach because our analyses include multiple lineages all experiencing different levels of gene flow, which violates many of the isolation with migration model assumptions (Hey & Nielsen 2007).

Three population groups were compared in these analyses: all populations of *H. annuus* ssp. *annuus*; all populations of *H. debilis cucumerifolius*; and all populations of ssp. *texanus* except for population '55', which is considered be to a discrete population by STRUCTURE with $K \geq 4$ (Fig. 2b). *Helianthus debilis debilis*, which represented the fifth cluster when $K \geq 5$, was also excluded because of its complete geographic isolation and lack of admixture with all the other populations analysed (Fig. 2b). $K = 5$ was chosen to define groups because it is the lowest clustering number able to distinguish between populations of ssp. *annuus* and ssp. *texanus* (see Results for further details).

We conducted two identical maximum-likelihood analyses with different starting seeds and a subset of 38 markers because of software limitations (Table S2).

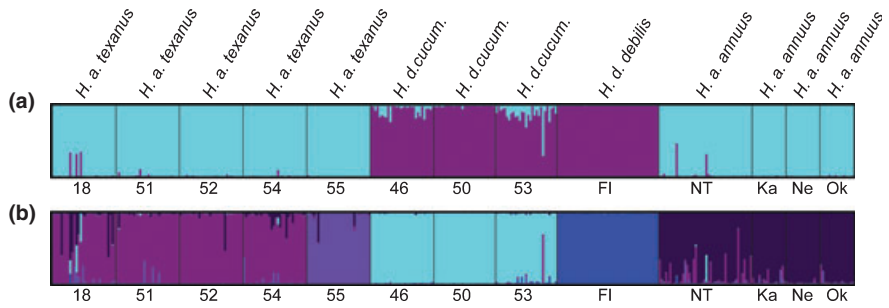

Fig. 2 Analysis with STRUCTURE for $K = 2$ (a) and $K = 5$ (b). Each colour represents a different cluster. Black lines separate populations. Within each population, individuals are represented by lines partitioned into coloured segments according to their assigned membership in the two inferred clusters. In this plot, *Helianthus debilis* is split into the two clusters that coincide with taxonomy and geography: *H. debilis cucumerifolius* from Texas and *H. debilis debilis* from Florida. Within ssp. *texanus*, population 55, which is a geographic outlier, also appears to be a genetically distinct.

Analyses assumed a Brownian motion model, which represents a continuous approximation to the stepwise mutation model for microsatellite evolution (Kimura & Ohta 1978). The Markov Chain settings consisted of 10 short and 3 long chains per run, with an increment of 20, 500 (short chain) and 1000 (long chain) sampled genealogies, with a burn-in of 1000. In addition, symmetric or no migration models were tested with the likelihood ratio test (Beerli & Felsenstein 1999), as implemented in MIGRATE-N: all possible combinations of symmetric or no migration between each population group were tested and compared with the likelihoods associated with the null-hypothesis. For each pairwise comparison ($i \rightarrow j$), the number of immigrants from population i into population j ($N_e m_{(i \rightarrow j)}$) was estimated as: $N_e m = \theta_i{*}M_{(i \rightarrow j)}/4$, where θ represents the mutation-scaled effective population size ($\theta = 4N_e\mu$).

To determine whether there were differences in migration rates among linkage groups, we repeated these analyses for all markers on each of 17 linkage groups. As before, we compared migration rates among the three main population groups: ssp. *annuus*, ssp. *texanus* and *H. debilis cucumerifolius*. However, to reduce run times, we reduced the number of long chain sampled genealogies to 500 and a burn-in of 500. We compared the results of the reduced sampling strategy in the linkage group 2 data set to an analysis where the sampling strategy was identical to the combined MIGRATE-N analysis. The reduction of sampled genealogies did not substantially change the estimated M and θ for linkage group 2, however, the confidence intervals were much larger.

These analyses also yielded estimates of M for each individual marker, which allowed us to test for correlations in migration rates of genetically linked loci. In hybridizing populations, we expect migration rates (and genetic distances) of linked markers to be correlated with the strength of correlation inversely proportional to genetic map distance (Scotti-Saintagne *et al.* (2004). Spatial autocorrelation was performed using unidirectional migration rates between ssp. *annuus* and *H. debilis cucumerifolius* and between ssp. *texanus* and *H. debilis cucumerifolius* as implemented in GeneAlex6 (Peakall & Smouse 2006). Distance intervals of 1, 2 and 10 centimorgans (cM) were explored. Observed correlations for each distance class were compared with the null distribution constructed from 1000 permutations. We repeated this analysis for both Nm and F_{ST} values (Weir & Cockerham 1984), calculated by GENEPOP 4.0 (Raymond & Rousset 1995), using the same groups and distance interval employed for the analyses of unidirectional migration rates. Lastly, we employed orthogonal regression of unidirectional migration rates (M) at individual loci to test whether values of M in one direction are correlated (i.e. symmetric) with those in the other direction. Correlations would be expected if most loci are neutral in the different genetic backgrounds or if selection acts symmetrically. Orthogonal regression fits lines that adjust for variability in X as well as Y, whereas standard least square fitting assumes that the X variable is fixed. Orthogonal regression was implemented in JMP version 5.0.1 (SAS Institute, Cary, NC, USA) and analysed the same population groupings used for spatial autocorrelation.

Analyses of recent immigration

To estimate the direction and rate of recent immigration, we employed BAYESASS1.3, which is a Bayesian multilocus genotyping procedure implemented with MCMC methods (Wilson & Rannala 2003). Unlike estimators of long-term gene flow, BAYESASS makes relatively few assumptions about demography and can be applied to populations that are not in mutation–drift or Hardy–Weinberg equilibrium. Thus, while complementary to MIGRATE-N with respect to timescale, analyses

with BAYESASS can also provide greater confidence that long-term migration rates result from post-divergence gene flow rather than the sorting of ancestral polymorphism (Kane *et al.* 2009). By using a Bayesian framework and a MCMC sampling scheme, BAYESASS is able to estimate recent immigration frequencies. While STRUCTURE uses a Bayesian probabilistic model to assign individuals to clusters, BAYESASS uses a Bayesian assignment algorithm to estimate the posterior probability of the individual's migration history. Thus, STRUCTURE and BAYESASS analyses provide complementary information about recent gene flow (Kane & King 2009).

The same populations used for the estimation of the long-term migration rate were included in the BAYESASS analysis. BAYESASS analyses were run for a total of 3 millions steps with 1 million of those being burn-in. Runs were carried out in duplicate to assess convergence of the MCMC after 3 millions steps.

Detection of loci under selection

To investigate the possibility of recent positive selection, we used the *ln*RV and *ln*RH test statistics of Schlötterer (2002) and Schlötterer & Dieringer (2005), which test for locus-specific reductions in gene diversity and variance, respectively, among population groups. The rationale for these tests is that the variability of loci that have undergone recent selective sweeps should be significantly reduced relative to the rest of the genome (Schlötterer 2002). To minimize false positives, Schlötterer & Dieringer (2005) recommend that candidates for selection be significant outliers in both statistics. The mean unbiased estimator of gene diversity (H_E) and the variance in the number of microsatellite repeats (VarRN) for each locus were calculated with MSA (Dieringer & Schlötterer 2003). Comparisons were made between populations or groups of populations as defined by STRUCTURE when $K = 5$ (Fig. 2B).

Evidence of selection can also be obtained by analyses of patterns of genetic divergence rather than losses of genetic variation. In this group of methods, searches are made for 'outlier loci': loci that behave differently from other loci in a given sample, presumably due to selection (Lewontin & Krakauer 1973). Here, we employed a simulation-based approach developed by Beaumont & Nichols (1996), as implemented in LOSITAN (Antao *et al.* 2008), to perform a widely used F_{ST}-outlier detection method. The method employs coalescent simulations under a uniform, finite island model and provides an expected neutral distribution of F_{ST} values, controlling for the corresponding expected heterozygosity value (H_E). Loci whose distribution lies outside the expected range are considered candidates of selection.

We performed an initial run with 50 000 simulations and all loci, using the mean neutral F_{ST} as a preliminary value. A more accurate estimate of the mean neutral F_{ST} was obtained following the first run by excluding all loci lying outside the 99% confidence interval, since their distribution could be the result of selection rather than neutral evolution. This refined estimate was used for a final set of 50 000 simulations over all loci.

To control for multiple comparisons, we employed a False Discovery Rate of 5% (Storey & Tibshirani 2003). This means that 5% of outlier loci detected by our analyses are expected to be false positives. Note that the False Discovery Rate differs from other approaches to the multiple comparisons problem (such as the Bonferroni method), which control for the chance of any false positives. *Q* values, which correspond to the observed *P*-value distribution without the expected false positive results, were calculated using the *Q*-value software (http://genomics.princeton.edu/storeylab/qvalue/) with default settings.

Results

Population structure

The inferred number of clusters (i.e. the grouping with the highest posterior probability among all the values tested, with *K* ranging from 2 to 13) was coincident with the number of sampled populations ($K = 13$). At $K = 2$, '*debilis*' and '*annuus*' clusters were recovered (Fig. 2a) with the putative hybrid, *Helianthus annuus* ssp. *texanus*, placed in the '*annuus*' cluster as predicted based on morphology and reproductive compatibility (Heiser 1951).

When higher numbers of clusters were allowed, a more complicated scenario emerged (Fig. 2b, $K = 5$). Within the '*debilis*' cluster, there was clear separation between the two varieties sampled (i.e. *Helianthus debilis cucumerifolius* from Texas and *H. debilis debilis* from Florida), noted also when $K = 3$ and 4 (data not shown). In addition, as predicted from their geographic distributions, there was no evidence of admixture between *H. debilis debilis* and either *H. debilis cucumerifolius* or *H. annuus*. In contrast, some individuals of *H. debilis cucumerifolius* appeared to contain genetic material of *H. annuus* (Fig. 2b).

A similar situation was found within the '*annuus*' cluster, in which three discrete population groups were recovered at $K = 5$: *H. annuus* ssp. *annuus*, ssp. *texanus*, and population 55 from eastern Texas, which was classified as ssp. *texanus* based on morphology. As with *H. debilis*, the geographically distant ssp. *annuus* populations (Ka, Ne, Ok in Fig. 2b) exhibited less admixture than did the geographically proximal populations from Texas.

The pattern of admixture can be seen more clearly in the site-by-site analysis (K = 2, Fig. 3). Within both species, the only individuals with an admixture coefficient (Q) significantly different from zero derive from Texas populations. Two of the three populations from *H. debilis cucumerifolius* (populations 46 and 53, Fig. 1) have the highest average admixture coefficients with an estimated average membership coefficient to the '*annuus*' cluster of Q_{annuus} = 5.7% and 12.5%, respectively, while ssp. *annuus* from Texas (population NT, Fig. 1) has a $Q_{debilis}$ of 2.3%. Although the putative hybrid lineage, ssp. *texanus*, was expected to have a predominately *H. annuus*-like genetic background, the degree of admixture was much lower than anticipated ($Q_{debilis}$ = 2%). Indeed, only 3 of 150 ssp. *texanus* individuals had an admixture coefficient significantly greater than zero. Interestingly, the pattern of admixture was not associated with chromosomal position (Fig. 3, which displays for each marker the difference in the probabilities that an individual is homozygous for alleles from the '*annuus*' or '*debilis*' cluster). Instead, introgression appeared to be scattered across the genome when it did occur.

Long-term estimates of migration and effective population size

The history of migration between *H. annuus* ssp. *annuus*, *H. annuus* ssp. *texanus* and *H. debilis cucumerifolius* was reconstructed through maximum-likelihood estimation of the effective population size and immigration rate under the coalescent theory. Both symmetric and absence of migration models were tested to explore the pattern of gene exchange that probably took place among these taxa.

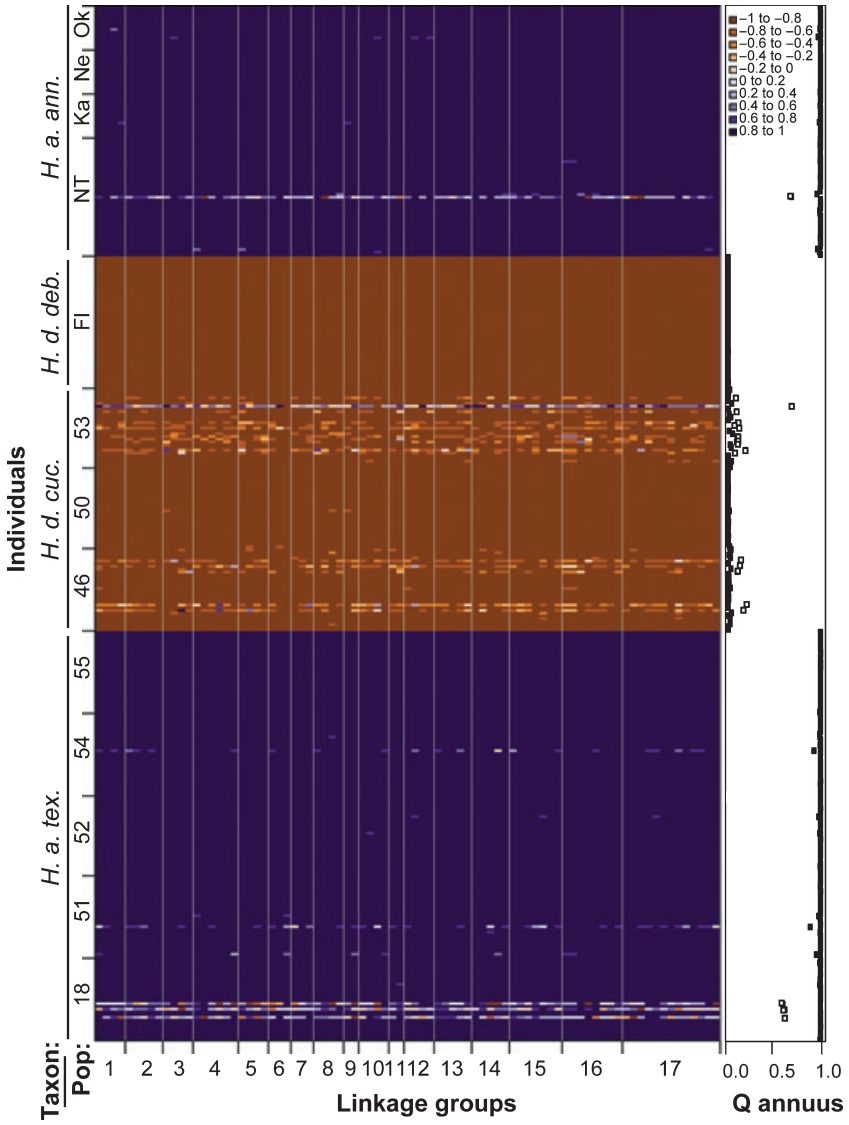

Fig. 3 Site-by-Site analysis obtained with STRUCTURE for K = 2. Genetic markers are ordered horizontally according to linkage group. Each individual is represented by a horizontal line partitioned into coloured segments according to the difference in probabilities that an individual is homozygous for alleles from the *H. annuus* or from the *H. debilis* cluster (p_{aa}–p_{dd}, range: −1 to 1). A scale showing correspondence between the difference in probabilities and colours is at the top right. The coefficient of membership in the *H. annuus* cluster (Q) for each sample is plotted to the right of the figure with white rectangles representing samples with an admixture coefficient (Q) significantly different from zero.

The hypothesis of absence of migration was rejected in all comparisons ($P < 0.00001$ in all the tests), and values of M between taxa were surprisingly high, ranging from 4.1 to 7.3 (Fig. 4). Thus, immigration appears to make a greater contribution to standing genetic variation than does mutation.

There was no difference in the direction of migration between ssp. *annuus* and *H. debilis cucumerifolius* (Fig. 4). However, the model of symmetric migration was rejected in the two tests involving ssp. *texanus* ($P < 0.000001$ in both tests). In both cases, there was greater introgression from ssp. *texanus* into its putative parental taxa than *vice versa*. This pattern is inconsistent with predictions of Currat *et al.* (2008) that migration between local congeners and colonizing species should be predominantly in the direction of the invader.

Estimates of the mutation-scaled effective population size indicate that ssp. *annuus* has a larger effective population size than *H. debilis cucumerifolius* ($\Theta = 2.0$ and 1.7, respectively, Fig. 4). This makes sense given the larger geographic range of the former. The observation that ssp. *texanus* has a larger effective population size ($\Theta = 2.7$) than either parental taxon is less easily accounted for, but might be owing to introgression from *H. debilis*, which should increase levels of genetic diversity.

Variation in migration rates within and among linkage groups

Estimates of M did not vary significantly among linkage groups for any of the intraspecific (*H. annuus* ssp. *annuus* vs. ssp. *texanus*) or interspecific (ssp. *annuus* or ssp. *texanus* vs. *H. debilis cucumerifolius*) comparisons that were made (Table S3). However, dramatic differences in migration were observed among individual loci (Appendix I; Table S2). This variation was only weakly correlated with genetic map position and virtually identical whether using M or Nm, so the weak correlation cannot be due to variation in the mutation rate among loci. In only one of four interspecific comparisons were significant correlations in unidirectional migration rates observed among markers in the closest distance (0–1 cM) class (Fig. S1). Oddly, significant correlations were observed among loci in the next two distance classes (1.1–2.0 cM and 2.1–3.0 cM) across several comparisons (Fig. S1). We do not have a reasonable biological explanation for this pattern and suspect that it is an artefact of the small number of the comparisons in these distance classes ($n = 8$ for both). No significant autocorrelations were observed at larger distance intervals or for F_{ST} in any distance category.

No correlations were observed between unidirectional migration rates at individual loci ($r < 0.04$ over all comparisons). That is, migration in one direction at a given locus was not correlated with migration in the other direction at that locus.

Recent immigration

Analyses of recent immigration with BAYESASS (Wilson & Rannala 2003) suggest that between 14% and 27% of individuals within each of the populations studied represent recent immigrants from another taxon (Fig. 5). Bayesian posterior standard deviations of immigration frequencies were zero for all populations and support for the assignment of immigrant alleles was similarly strong. However, according to a recent simulation study (Faubet *et al.* 2007), accurate estimates of immigration rates with BAYESASS require significant genetic differentiation among populations ($F_{ST} > 0.05$)

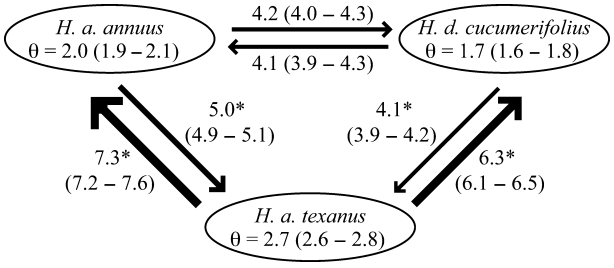

Fig. 4 Maximum-likelihood estimates and 95% confidence intervals (in parentheses) of the long-term migration rate (M) and the mutation-scaled effective population size (θ) from MIGRATE-N. Thickness of arrows corresponds to migration rate estimates. $M = m/\mu$, where m is the migration rate and μ is mutation rate per site per generation. $\theta = 4N_e\mu$, where N_e is the effective population size. *Model of symmetric migration rejected.

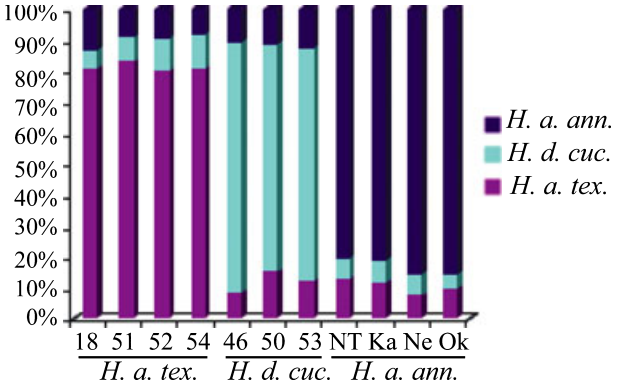

Fig. 5 Percentage of recent immigrants into populations of ssp. *annuus*, ssp. *texanus* and *Helianthus debilis cucumerifolius*, as estimated by BAYESASS. Coloured segments indicate the proportion of individuals within each population derived from the source populations or from immigration.

as well as fairly low migration rates ($m < 0.1$). F_{ST} values, measured over all loci (Weir & Cockerham 1984) with GENEPOP 4.0 (Raymond & Rousset 1995), range from 0.12 to 0.14 for comparisons between *H. annuus* and *H. debilis cucumerifolius* vs. 0.02 to 0.03 for comparisons of ssp. *annuus* with ssp. *texanus*. Thus, despite the apparent statistical confidence in all of our estimates of immigration frequencies, those involving interspecific comparisons are probably considerably more accurate than those for intraspecific comparisons (e.g. between populations of ssp. *annuus* and ssp. *texanus*). Another caveat is that posterior assignment probabilities have been shown to poorly predict the accuracy of allelic assignments (Faubet *et al.* 2007); therefore, interspecific allelic assignments are likely to be more accurate than intraspecific assignments.

Detection of loci under selection

A single marker (HT0414) was a significant outlier in both the lnRV and lnRH analyses ($P_{lnRV} < 0.005$; $P_{lnRH} < 0.01$) (Table 2; Table S4; Fig. S2) because of reduced variability in populations of ssp. *annuus* relative to either ssp. *texanus* or *H. debilis cucumerifolius*. The reduction in variability is restricted to populations of ssp. *annuus* from outside Texas. With the exception of a single heterozygous individual from Oklahoma, all plants north of Texas are homozygous for a single allele. HT0414 is located within contig2135 from Compositae Genome Project 1 (http://compgenomics.ucdavis.edu/compositae_index.php). This contig also includes a DNAJ heat-shock N-terminal domain-containing protein (*Arabidopsis* best hit AT4G37750), which represents the most likely candidate gene for the apparent selective sweep reported here.

The F_{ST} simulation approach yielded four outliers (HT0760, HT0429, HT0326, HT0538; Table S4) in both interspecific comparisons, but none in the intraspecific comparison (Table S4). After correction for multiple comparisons using an False Discovery Rate 5%, a total of three outliers remained significant in the comparisons between ssp. *annuus* and *H. debilis cucumerifolius* (HT0429; HT0760; Table 2; Table S4) or ssp. *texanus* and *H. debilis cucumerifolius* (HT0429; HT0326; Table 2; Table S4). The three outliers showed species-specific patterns in the distribution of allelic frequencies: all *H. annuus* ssp. *annuus* and ssp. *texanus* populations carry the same most representative alleles, which can be found at very low frequencies at least in one of the populations of *H. debilis cucumerifolius*. An opposite pattern was found for alleles with high frequencies in *H. debilis cucumerifolius*, since they were at low frequencies in all *H. annuus* populations (HT0326), or at least in some ssp. *texanus* populations (HT0760 and HT0429). HT0760 is weakly similar to an *Arabidopsis thaliana* photosystem II 5 kD protein, whereas HT0429 is similar to an *A. thaliana* protein with unknown function; no information about homologous genes with a good annotation was available for HT0326.

Discussion

Population structure and migration

The two species of annual sunflowers, *Helianthus annuus* ssp. *annuus* and *Helianthus debilis*, are thought to have given rise to a distinct and stabilized lineage, *H. annuus* ssp. *texanus*, through introgressive hybridization (Heiser 1951, 1954). However, both the extent of admixture and its genomic distribution have been unclear. Our analyses of 88 genetically mapped microsatellite loci provide two seemingly conflicting assessments of the extent of introgression. Analyses of genomic composition with STRUCTURE imply that admixture is extremely limited, with an average admixture coefficient lower than 2%. Moreover, most of the admixture derives from three individuals, which may be early generation hybrids.

In contrast, coalescent analyses of long-term migration rates imply that interspecific migration has been

Table 2 Outlier markers that may have been targeted by selection

	H. a. ann. and *H. d. cuc.*			*H. a. tex.* and *H. d. cuc.*				
	H_E	F_{ST}	$P_{(lnRHvslnRV//FDIST2)}$	H_E	F_{ST}	$P_{(lnRHvslnRV//FDIST2)}$	LG	cM
HT0414	0.80	0.43	*/–	0.79	0.36	–/–	17	124.3
HT0760	0.98	0.52	–/*	0.96	0.48	–/–	4	37.8
HT0429	0.99	0.56	–/*	0.96	0.54	–/*	14	10.3
HT0326	0.76	0.72	–/–	0.75	0.69	–/*	15	15.3

MarkerHT0414 was a significant outlier in both lnRH and lnRV test, whereas HT0760, HT0429 and HT0326 were detected using an F_{ST}-based outlier method. Linkage group (LG) and genetic map positions in cM are also reported.
H. a. ann., *Helianthus annuus annuus*; *H. d. cuc.*, *Helianthus debilis cucumerifolius*; *H. a. tex.*, *H. annuus texanus*.
–$P > 0.01$ and <0.05; *$P < 0.01$.

more important than mutation in providing new genetic variation in ssp. *texanus* populations. Immigration from *H. debilis cucumerifolius* into ssp. *texanus* (M = 4.05; N_em = 1.74) is higher than median estimates of intraspecific migration rates in plants (Morjan & Rieseberg 2004) and only slightly lower than immigration from ssp. *annuus* into ssp. *texanus* (M = 5.00). Immigration rates from ssp. *texanus* into *H. debilis cucumerifolius* and ssp. *annuus* are still higher (M = 6.30 and 7.29, respectively), implying that the putative hybrid lineage serves as a bridge for migration between the parental species.

Estimates of recent immigration from BAYESASS are high as well, with greater than 14% of individuals representing recent immigrants from another taxon (Fig. 5). Consistent with the MIGRATE-N results, there are more recent immigrants from ssp. *texanus* into its putative parental taxa than *vice versa*, the same pattern observed for long-term migration rates. One puzzle is that the estimated immigration rate between the allopatric taxa (*H. debilis cucumerifolius* and ssp. *annuus*) is approximately comparable with that between the sympatric taxa (*H. debilis cucumerifolius* and ssp. *texanus* ssp. *texanus*). This observation may point to the effectiveness of ssp. *texanus* as a bridge for the transfer of alleles between *H. debilis* and ssp. *annuus*. An alternative explanation is that the BAYESASS estimates are unreliable because homoplasy of microsatellite alleles is high in taxa with large-effective population sizes (Estoup *et al.* 2002) such as the annual sunflowers studied here.

The seemingly contradictory results from the STRUCTURE and MIGRATE-N analyses can be reconciled by consideration of the timescale over which these analyses effectively detect immigration. STRUCTURE is powerful at detecting recent admixture, whereas values of M reflect persistent migration over hundreds or thousands of generations. However, this explanation cannot account for the difference between the BAYESASS results and those of STRUCTURE since both are multilocus genotypic approaches designed to detect recent migration. It might be that BAYESASS has more power, but greater sensitivity to homoplasy, since it estimates the probability that an allele is an immigrant, whereas STRUCTURE estimates the probability that an individual is an immigrant. As a consequence, BAYESASS may be over-estimating and STRUCTURE under-estimating recent immigration rates. This possibility is supported by reports that BAYESASS consistently over-estimated immigration rates between laboratory-cultured populations of nematodes (Mardulyn *et al.* 2008).

Despite the apparent discrepancy between the results from STRUCTURE and BAYEASS, several conclusions can be made. Analyses with STRUCTURE support the distinctness of ssp. *texanus*, as well as its clear inclusion within *H. annuus*, as initially postulated by Heiser (1951). Likewise, the STRUCTURE, MIGRATE-N and BAYESASS analyses all confirm previous report of introgression between *H. annuus* and *H. debilis* (Rieseberg *et al.* 1990, 2007) and suggest that the introgression is genome-wide when it occurs. Contrary to the Heiser's scenario of a recent Holocene origin of ssp. *texanus*, however, our results are more consistent with a much longer history of contact between *H. annuus* and *H. debilis*. Otherwise, it is difficult to account for the presence of significant long-term migration between currently allopatric populations of ssp. *annuus* and *H. debilis cucumerifolius*. This revised scenario makes sense in the light of numerous glacial-interglacial cycles over the past million years (Hewitt 2000), which probably resulted in intermittent contact between *H. annuus* and *H. debilis*, with the last contact likely occurring during the Wisconsin glaciation, 18 000 BP. The current contact between the species differs from the previous ones in that it appears to have been human-aided. Possibly, some of the molecular evidence of introgression, particularly into allopatric populations, stems from past periods of contact.

Direction of introgression

A recent simulation study by Currat *et al.* (2008) predicted that during invasions, most neutral introgression should be from the local to the invading species. The reason for this pattern is that local alleles which introgress when the invading population is at low density can be amplified as a consequence of logistic growth of the invader. In contrast, local populations are probably already at carrying capacity, so introduced alleles cannot be amplified by population growth. Patterns of introgression in many invasions do indeed follow this general prediction (Currat *et al.* 2008; Petit & Excoffier 2009; Zalapa *et al.* 2009), although probably not so in situations where the invading species is diploid and the native is tetraploid (Stebbins 1971; Kim *et al.* 2008).

So why do we see the opposite pattern in this study, with greater introgression into the local (*H. debilis cucumerifolius*) rather than the invading taxon ssp. *texanus*? The 'Currat' effect requires that interbreeding events be frequent when the invading population is still at low density. This requirement is probably violated in our system, for two reasons. First, in central Texas, the microhabitat favoured by ssp. *texanus* (clay soils) is much more extensive than that favoured by *H. debilis* (sandy soils), with small patches of the latter located in a matrix of the former (pers. obs.). Thus, if early contact between the two species was in this region, abundances of the invading ssp. *annuus* were probably much higher than those of the local *H. debilis*. Second, the reproductive barrier between ssp. *annuus* and *H. debilis cucumerifolius* is strong: F1 fertility averages 3.7–6.7% (Heiser

1951). As a consequence, the pattern of introgression is probably influenced more by the current abundance of the invading vs. local species, which favours ssp. *annuus* in this case.

Analyses of the direction of introgression also offers the opportunity to detect 'speciation genes' (Payseur *et al.* 2004; Currat *et al.* 2008; Teeter *et al.* 2008; Gompert & Buerkle 2009), since Bateson–Dobzhansky–Muller (BDM) incompatibilities are predicted to generate asymmetric gene flow (Coyne & Orr 2004). In the BDM model, diverging populations accumulate hybrid incompatibilities without loss of fitness. In a simple two locus case, an ancestral genotype, aabb, can give rise to a genotype of AAbb in population 1 and aaBB in population 2. While the AAbb and aaBB genotypes are viabile and fertile, the A and B alleles are incompatible and result in a loss of fitness in hybrid gentypes (AaBb). However, the incompatibilities are asymmetric. The 'a' allele is compatible in the background of population 1 and the 'b' allele is compatible in the background of population 2. Thus, we expect introgression to occur in the direction of population 1 at the 'a' locus and in the direction of population 2 at the 'b' locus.

In this study, we detected one locus (ORS0885; LG 16) that may be tightly linked to a BDM incompatibility: the immigration rate from *H. annuus* ssp. *annuus* or ssp. *texanus* into *H. debilis cucumerifolius* is zero, whereas estimates of *M* in the opposite direction were very high (391.1 and 78.2, respectively; Table S2). No such asymmetry was observed for this locus in the intraspecific comparison between ssp. *annuus* and ssp. *texanus*. Follow-up studies are planned to ask whether this locus is associated with a loss of fertility as well as to fine map this region to identify the gene and site responsible for the asymmetry in gene flow.

The unit of genetic isolation

It is now widely recognized that the genomes of hybridizing species often are permeable to interspecific gene flow (Barton & Hewitt 1985; Rieseberg *et al.* 1999; Wu 2001; Payseur *et al.* 2004; Baack & Rieseberg 2007; Good *et al.* 2008; Lexer & Widmer 2008; Schwenk *et al.* 2008; Nolte *et al.* 2009). However, there is less agreement about how much of the genome is likely to experience gene flow or about the lengths of chromosomal segments protected from introgression (Barton & Hewitt 1985; Charlesworth *et al.* 1997; Coyne & Orr 2004; Turner *et al.* 2005; Yatabe *et al.* 2007; Via & West 2008; Wood *et al.* 2008). Theory indicates that the rate of introgression near a selected site will be inversely proportional to the selection:recombination ratio (Barton 1979). Thus, if only a small number of factors contribute to reproductive isolation, then most of the genome should be permeable to introgression. However, for most species we know too little about the numbers and genomic distribution of factors that contribute to reproductive isolation to make realistic predictions about genomic extent of introgression. The *H. annuus/H. debilis* system represents an exception to this general rule since an earlier QTL study of hybrid sterility involving these species detected only two QTLs, leading to the prediction that most of the genome would be permeable to gene flow (Kim & Rieseberg 1999). This prediction was confirmed by this study. These results augment growing evidence that species' genomes are more porous and interspecific gene flow is more widespread than previously suspected (Arnold *et al.* 1990; Lexer *et al.* 2005; Yatabe *et al.* 2007; Städler *et al.* 2008; Strasburg & Rieseberg 2008; Hertwig *et al.* 2009; Hoban *et al.* 2009; Kane *et al.* 2009; Lepais *et al.* 2009; Nolte *et al.* 2009).

Even less is known about the lengths of genetically differentiated chromosomal segments in hybridizing species. Simulation studies show that with population subdivision, local selection can extend regions of high F_{ST} for very long distances (on the order of the selection coefficient) because of a stable hitchhiking effect on either side of a locus under disruptive natural selection (Charlesworth *et al.* 1997). For example, a selection coefficient of 0.1 can result in increased differentiation up to 10 cM from the selected site. Regions of differentiation should be even larger if multiple, linked loci are under disruptive selection (Barton & Hewitt 1989) or if the selected locus occurs in a region of low recombination (Rieseberg 2001). Unfortunately, comparable predictions are not available for BDM incompatibilities, or for ecologically important loci in which alleles are divergently selected in one habitat, but neutral or nearly so in the other. However, first principles suggest that genetic differentiation should be weaker and extend over shorter differences in these situations because unidirectional gene flow will still be possible.

Empirical data are difficult to interpret. Several genome scans have reported an association between outlier AFLPs loci and major QTLs (e.g. Rogers & Bernatchez 2005; Via & West 2008). The genetic distance between outlier loci and QTLs averages >10 cM in both studies suggesting that islands of differentiation may be large. An alternative interpretation is that these are gene-rich regions that contain multiple, smaller regions of differentiation (Smadja *et al.* 2008). Regions of differentiation between house mouse subspecies have been reported to be large as well, averaging >10 cM in length (Harr 2006a). However, many of the apparently differentiated regions result from single nucleotide polymorphism ascertainment bias (Boursot & Belkhir 2006; Harr 2006b)

and do not show significant differentiation in more detailed analyses of individual markers (Teeter *et al.* 2008).

Other studies have found regions of differentiation to be much smaller, and more on the scale reported here. A genome-wide analysis of closely related forms of *Anopheles* identified three regions of genomic differentiation ranging from 5 to 50 genes (Turner *et al.* 2005), all in low recombination regions of the genome (Pombi *et al.* 2006). In oaks and another hybridizing pair of sunflowers (*H. annuus* × *Helianthus petiolaris*), differentiated regions are typically <1 cM in length, except in areas of low recombination (Scotti-Saintagne *et al.* 2004; Yatabe *et al.* 2007).

Via & West (2008) questioned the approach taken in the sunflower and oak studies because information from all loci is used to define differentiated regions as opposed outliers to only, thereby potentially confounding the effects of migration and retention of ancestral polymorphisms. However, the small scale of genomic differentiation in *H. annuus* × *H. petiolaris* has since been confirmed by sequence analyses of 77 loci distributed across the genome of the two species (Strasburg *et al.* 2009). In this study, we employed a coalescent approach to minimize the problem of ancestral polymorphism and considered the direction of introgression to account for loci with asymmetric effects on fitness (i.e. most BDM incompatibilities). However, we failed to find correlations between genetically linked loci in either the unidirectional migration rates or $F_{ST,}$ even in the smallest distance class of 1 cM. Also, we found no evidence that outlier loci were clustered: all occurred on different linkage groups. Thus, islands of low migration/high differentiation appear to be small in populations of *H. annuus* × *H. debilis*.

Similar results have been reported for different forms of the marine gastropod, *Littorina saxatilis* (Wood *et al.* 2008). Sequence data from outlier loci revealed that regions of differentiation were highly localized (<10 Kb). Via & West (2008) suggest that regions of genetic differentiation that accumulate in allopatry will be smaller than those in sympatry since the lack of mating between allopatric populations does not induce 'divergence hitchhiking', which would reduce effective recombination rates near QTLs under divergent selection. While this explanation may partially explain for the small scale of genomic differentiation in sunflower, it cannot account for the *Littorina* data, since any period of allopatry must have been short (Smadja *et al.* 2008). An alternative explanation is that loci experiencing strong disruptive selection are rare. If effective population sizes are large, such as those reported for sunflower (Strasburg & Rieseberg 2008), then even very weak selection ($s > m$) can produce significant divergence, leading to very small islands of differentiation (Smadja *et al.* 2008).

Selective sweeps

Only four loci (HT0414, HT0760, HT0429 and HT0326) analysed in this study showed evidence of having experienced significant selection. Interestingly, all four were derived from the coding region of expressed sequence tags, whereas no outliers (even when using relaxed significance thresholds) came from the dinucleotide repeat library, despite the fact that over half of the markers analysed were derived from the latter. This pattern is consistent with the hypothesis that the outlier loci differ from other loci because of selection.

One of the four outlier loci, HT0414, is associated with a heat-shock protein and was previously shown to have been the target of selection in *H. annuus* ssp. *annuus* salt marsh populations from the state of Utah (Kane & Rieseberg 2007), as well as in weedy sunflower populations from several different locations across the USA (Kane & Rieseberg 2008). In this study, populations of ssp. *annuus* from Kansas, Nebraska and Oklahoma monomorphic for a single allele, whereas populations of *H. debilis* and other populations of ssp. *annuus* are considerably more polymorphic. These results imply that HT0414 may be involved in adaptation to a range of different habitats or to conditions shared by several different habitats (Kane & Rieseberg 2008). For the remaining three outliers, only HT0760 showed a weak similarity to an *Arabidopsis thaliana* known protein, whereas no function is known for HT0429 and HT0326).

Conclusions

Our results confirm previous reports of introgression between the widespread common sunflower (*Helianthus annuus* ssp. *annuus*) and a local congener from Texas, *Helianthus debilis cucumerifolius*. However, our analyses imply that introgression has occurred over a much longer timescale than suspected by Heiser (1951). Long-term migration rates are high and appear to be genome-wide, confirming predictions from previous mapping studies, which reported a simple genetic architecture of reproductive isolation between these two species (Kim & Rieseberg 1999). The extent of recent immigration/admixture is less clear because of disagreement between the STRUCTURE and BAYEASS results.

There is strong asymmetry in migration rates both genome-wide and at the scale of individual loci. Contrary to recent theoretical predictions (Currat *et al.* 2008), there was more immigration into the local

species, *H. debilis cucumerifolius*, than into the invader (and putative hybrid lineage), ssp. *texanus*. This pattern can be explained by a more restricted microhabitat preference in the local species, presumably leading to a population size advantage for the invader even during early contact between the species, and by the strong reproductive barrier between *H. debilis* and *H. annuus*, which may have limited interspecific migration during the initial stages of the invasion. Both factors probably reduced the potential for population growth to amplify introgressed alleles to high frequency in the invader. Unidirectional migration rates were also significantly higher from ssp. *texanus* into ssp. *annuus* than *vice versa*, implying that ssp. *texanus* serves as a conduit for the movement of genetic material between the putative parental taxa. At the individual locus scale, one marker showed no migration from ssp. *annuus* into *H. debilis*, but very high migration rates in the opposite direction, a pattern consistent with tight linkage to a BDM incompatibility.

Finally, we tested whether correlations between unidirectional migration rates and genetic map distances would reveal longer regions of genetic isolation than did similar analyses with F_{ST}. However, in agreement with previous studies of hybridizing sunflower species (e.g. Yatabe *et al.* 2007; Strasburg *et al.* 2009), regions of differentiation between these species appear to be small, providing additional support for the porosity of reproductive barriers between hybridizing sunflower species.

Acknowledgements

We thank Rose Andrew for assistance with the spatial autocorrelation analyses and the Rieseberg laboratory, in particular Nolan Kane, for valuable suggestions on data analysis. This work was funded by NSF Grant 0716868 to K.D.W. and National Sciences and Engineering Research Council of Canada Grant 327475 to L.H.R.

References

Anderson E (1949) *Introgressive Hybridization*. John Wiley, New York.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a F_{ST} -outlier method. *BMC Bioinformatics*, **9**, 323.

Arnold ML, Bennett BD, Zimmer EA (1990) Natural hybridization between *Iris fulva* and *Iris hexagona*: pattern of ribosomal DNA variation. *Evolution*, **44**, 1512–1521.

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

Barton NH (1979) The dynamics of hybrid zones. *Heredity*, **43**, 341–359.

Barton NH (2001) The role of hybridization in evolution. *Molecular Ecology*, **10**, 551–568.

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

Barton NH, Hewitt GM (1989) Adaptation, speciation and hybrid zones. *Nature*, **341**, 497–503.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **263**, 1619–1626.

Beerli P (2004) Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Molecular Ecology*, **13**, 827–836.

Beerli P (2006) Comparison of Bayesian and maximum-likehood inference of population genetic parameters. *Bioinformatics*, **22**, 341–345.

Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*, **152**, 763–773.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4563–4568.

Boursot P, Belkhir K (2006) Mouse SNPs for evolutionary biology: beware of ascertainment biases. *Genome Research*, **16**, 1191–1192.

Castric V, Bechsgaard J, Schierup MH, Vekemans X (2008) Repeated adaptive introgression at a gene under multiallelic balancing selection. *PLoS Genetics*, **4**, e1000168.

Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, **70**, 155–174.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, Massachusetts.

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

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

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

Ellstrand NC (1992) Gene flow by pollen: implications for plant conservation genetics. *Oikos*, **63**, 77–86.

Emms SK, Arnold ML (1997) The effect of habitat on parental and hybrid fitness: reciprocal transplant experiments with Louisiana irises. *Evolution*, **51**, 1112–1119.

Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, **11**, 1591–1604.

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.

Faubet P, Waples RS, Gaggiotti OE (2007) Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. *Molecular Ecology*, **16**, 1149–1166.

Field DL, Ayre DJ, Whelan RJ, Young AG (2009) Molecular and morphological evidence of natural interspecific

hybridization between the uncommon *Eucalyptus* aggregata and the widespread *E. rubida* and *E. viminalis*. *Conservation Genetics*, **10**, 881–896.
Genovart M (2009) Natural hybridization and conservation. *Biodiversity and Conservation*, **18**, 1435–1439.
Gompert Z, Buerkle CA (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology*, **18**, 1207–1224.
Good JM, Hird S, Reid N *et al.* (2008) Ancient hybridization and mitochondrial capture between two species of chipmunks. *Molecular Ecology*, **17**, 1313–1327.
Grant BR, Grant PR (1996) High survival of Darwin's finch hybrids: effects of beak morphology and diets. *Ecology*, **77**, 500–509.
Grant BR, Grant PR (2008) Fission and fusion of Darwin's finches populations. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **363**, 2821–2829.
Gross BL, Turner KG, Rieseberg LH (2007) Selective sweeps in the homoploid hybrid species *Helianthus deserticola*: evolution in concert across populations and across origins. *Molecular Ecology*, **16**, 5246–5258.
Harr B (2006a) Genomic islands of differentiation between house mouse subspecies. *Genome Research*, **16**, 730–737.
Harr B (2006b) Regions of high differentiation-worth a check. *Genome Research*, **16**, 1193–1194.
Harter AV, Gardner KA, Falush D *et al.* (2004) Origin of extant domesticated sunflowers in eastern North America. *Nature*, **430**, 201–205.
Heesacker A, Kishore VK, Gao W *et al.* (2008) SSRs and INDELs mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility. *Theoretical and Applied Genetics*, **117**, 1021–1029.
Heiser CB (1951) Hybridization in the annual sunflowers: *Helianthus annuus* x *H. debilis* var. *cucumerifolius*. *Evolution*, **5**, 42–51.
Heiser CB (1954) Variation and subspeciation in the common sunflower, *Helianthus annuus*. *American Midland Naturalist*, **51**, 287–305.
Hertwig ST, Schweizer M, Stepanow S *et al.* (2009) Regionally high rates of hybridization and introgression in German wildcat populations (*Felis silvestris*, Carnivora, Felidae). *Journal of Zoological Systematics and Evolutionary Research*, **47**, 283–297.
Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.
Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Science of the United States of America*, **104**, 2785–2790.
Hoban SM, McCleary TS, Schlarbaum SE, Romero-Severson J (2009) Geographically extensive hybridization between the forest trees American butternut and Japanese walnut. *Biology Letters*, **5**, 324–327.
Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.
Kane NC, King MG (2009) Using parentage analysis to examine gene flow and spatial genetic structure. *Molecular Ecology*, **18**, 1551–1552.
Kane NC, Rieseberg LH (2007) Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *Helianthus annuus*. *Genetics*, **175**, 1823–1834.
Kane NC, Rieseberg LH (2008) Genetics and evolution of weedy *Helianthus annuus* populations: adaptation of an agricultural weed. *Molecular Ecology*, **17**, 384–394.
Kane NC, King MG, Barker MS *et al.* (2009) Comparative genomic and population genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. *Evolution*, **63**, 2061–2075.
Kim SC, Rieseberg LH (1999) Genetic architecture of species differences in annual sunflowers: implications for adaptive trait introgression. *Genetics*, **153**, 965–977.
Kim M, Cui ML, Cubas P *et al.* (2008) Regulatory genes control a key morphological and ecological trait transferred between species. *Science*, **322**, 1116–1119.
Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Sciences of the United States of America*, **75**, 2868–2872.
Lepais O, Petit RJ, Guichoux E *et al.* (2009) Species relative abundance and direction of introgression in oaks. *Molecular Ecology*, **18**, 2228–2242.
Levin DA, Francisco-Ortega J, Jansen RK (1996) Hybridization and the extinction of rare plant species. *Conservation Biology*, **10**, 10–16.
Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.
Lexer C, Widmer A (2008) The genic view of plant speciation: recent progress and emerging questions. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **363**, 3023–3036.
Lexer C, Fay MF, Joseph JA, Nica MS, Heinze B (2005) Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. *Molecular Ecology*, **14**, 1045–1057.
Mardulyn P, Vaesen MA, Milinkovitch MC (2008) Controlling population evolution in the laboratory to evaluate methods of historical inference. *PLoS ONE*, **3**, e2960.
Morgan-Richards M, Smissen RD, Shepherd LD *et al.* (2009) A review of genetic analyses of hybridisation in New Zealand. *Journal of the Royal Society of New Zealand*, **39**, 15–34.
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.
Muhlfeld CC, Kalinowski ST, McMahon TE *et al.* (2009) Hybridization rapidly reduces fitness of a native trout in the wild. *Biology Letters*, **5**, 328–331.
Nolte AW, Gompert Z, Buerkle CA (2009) Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Molecular Ecology*, **18**, 2615–2627.
O'Brien J, Devillard S, Say L *et al.* (2009) Preserving genetic integrity in a hybridising world: are European Wildcats (*Felis silvestris silvestris*) in eastern France distinct from sympatric feral domestic cats? *Biodiversity and Conservation*, **18**, 2351–2360.
Payseur BA, Krenz JG, Nachman MW (2004) Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution*, **58**, 2064–2078.

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

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology and Evolution*, **24**, 386–393.

Pombi M, Stump AD, Della Torre A, Besansky NJ (2006) Variation in recombination rate across the X chromosome of *Anopheles gambiae*. *American Journal of Tropical Medicine and Hygiene*, **75**, 901–903.

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

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rhymer JM, Simberloff DS (1996) Genetic extinction through hybridization and introgression. *Annual Review of Ecology and Systematics*, **27**, 83–109.

Rieseberg LH (1991) Homoploid reticulate evolution in *Helianthus*: evidence from ribosomal genes. *American Journal of Botany*, **78**, 1218–1237.

Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, **16**, 351–358.

Rieseberg LH (2009) Evolution: replacing genes and traits through hybridization. *Current Biology*, **19**, R119–R122.

Rieseberg LH, Zona S, Aberbom L, Martin T (1989) Hybridization in the island endemic, *Catalina mahogany*. *Conservation Biology*, **3**, 52–58.

Rieseberg LH, Beckstrom-Sternberg S, Doan K (1990) Helianthus *annuus* ssp. *texanus* has chloroplast DNA and nuclear ribosomal RNA genes of *Helianthus debilis* ssp. *cucumerifolius*. *Proceedings of the National Academy of Sciences of the United States of America*, **87**, 593–597.

Rieseberg LH, Van Fossen C, Desrochers AM (1995) Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature*, **375**, 313–316.

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.

Rieseberg LH, Kim SC, Randell RA *et al.* (2007) Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica*, **129**, 149–165.

Rogers SM, Bernatchez L (2005) Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*). *Molecular Ecology*, **14**, 351–361.

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Schlötterer C (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics*, **160**, 753–763.

Schlötterer C, Dieringer D (2005) A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity. In: *Selective Sweep* (ed. Nurminsky D) pp. 55–64 Springer, New York.

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

Schwenk K, Brede N, Streit B (2008) Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **363**, 2805–2811.

Scotti-Saintagne C, Mariette S, Porth I *et al.* (2004) Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics*, **168**, 1615–1626.

Smadja C, Galindo J, Butlin R (2008) Hitching a lift on the road to speciation. *Molecular Ecology*, **17**, 4177–4180.

Städler T, Arunyawat U, Stephan W (2008) Population genetics of speciation in two closely related wild tomatoes (Solanum section Lycopersicon). *Genetics*, **178**, 339–350.

Stebbins GL (1959) The role of hybridization in evolution. *Proceedings of the American Philosophical Society*, **103**, 231–251.

Stebbins GL (1971) *Chromosomal Evolution in Higher Plants*. Edward Arnold, London, UK.

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 9440–9445.

Strasburg JL, Rieseberg LH (2008) Molecular demographic history of the annual sunflowers *Helianthus annuus* and *H. petiolaris*–large effective population sizes and rates of long-term gene flow. *Evolution*, **62**, 1936–1950.

Strasburg JL, Scotti-Saintagne C, Scotti I, Lai Z, Rieseberg LH (2009) Genomic patterns of adaptive divergence between chromosomally differentiated sunflower species. *Molecular Biology and Evolution*, **26**, 1341–1355.

Takakura K, Nishida T, Matsumoto T, Nishida S (2009) Alien dandelion reduces the seed-set of a native congener through frequency-dependent and one-sided effects. *Biological Invasions*, **11**, 973–981.

Tang S, Yu J-K, Slabaugh MB, Shintani DK, Knapp SJ (2002) Simple sequence repeat map of the sunflower genome. *Theoretical and Applied Genetics*, **105**, 1124–1136.

Teeter KC, Payseur BA, Harris LW *et al.* (2008) Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Research*, **18**, 67–76.

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, **3**, e285.

Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Whitney KD, Randell RA, Rieseberg LH (2006) Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. *The American Naturalist*, **167**, 794–807.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177–1191.

Wolf DE, Takebayashi N, Rieseberg LH (2001) Predicting the risk of extinction through hybridization. *Conservation Biology*, **15**, 1039–1053.

Wood HM, Grahame JW, Humphray S, Rogers J, Butlin RK (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Molecular Ecology*, **17**, 3123–3135.

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

Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH (2007) Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics*, **175**, 1883–1893.

Zalapa JE, Brunet J, Guries RP (2009) Patterns of hybridization and introgression between invasive *Ulmus pumila* (Ulmaceae) and native *U. rubra*. *American Journal of Botany*, **96**, 1116–1128.

The laboratories of Loren Rieseberg, Alex Buerkle, and Ken Whitney study a broad array of topics in plant evolution, including the role of hybridization in adaptation and speciation. The present study was spearheaded by a postdoc in the Rieseberg lab, Moira Scascitelli, who is interested in hybridization, chromosomal evolution, and speciation in plants and animals. Rebecca Randell is a graduate student in the Rieseberg lab with interests in plant hybridization and conservation genetics. Postdoc Matthew King's interests from plant population genetics to evolutionary genomics.

Supporting information

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

Fig. S1 Spatial autocorrelational analysis of uniparental migration rates (*M*) between *Helianthus annuus* and *H. debilis cucumerifolius*.

Fig. S2 Plot of LnRH vs. lnRV statistics from pairwise comparisons among *Helianthus annuus annuus*, *H. debilis cucumerifolius* and *H. annuus texanus*.

Table S1 Details about the microsatellite markers

Table S2 Microsatellites, genetic map positions, and MCMC estimates of the number of migrants (*Nm*) calculated from Migrate-n results

Table S3 Maximum-likelihood estimates and 99% confidence intervals (in parentheses) of chromosome-specific long-term migration rates (*M*) between *H. annuus annuus* (*H. a. ann.*), *H. debilis cucumerifolius* (*H. d. cuc.*) and *H. annuus texanus* (*H. a. tex.*)

Table S4 Analyses to detect markers potentially under selection

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

Appendix I Microsatellites, genetic map positions, and MCMC estimates of long-term migration rates (*M*) and the mutation-scaled effective population size Θ from MIGRATE-N. $M = m/\mu$, where m is the migration rate and μ is mutation rate per site per generation. $\theta = 4N_e\mu$, where N_e is the effective population size. Underlined markers were used to calculate long-term migration rates among population groups (cf. Fig. 4).

			M [+=receiving population]							
		Locus*	*H. a. tex.*,+	*H. d. cuc.*,+	*H. a. ann.*,+					
LG	cM	*H. a. tex.*				LG		0.05	MLE	0.95
1	19.6	ORS0728	—	5.7363	8.2374	1	Θ_*H. a. tex.*	1.7837	1.8959	2.0177
	39.8	ORS0371	—	6.2009	9.2539		Θ_*H. d. cuc.*	1.1359	1.2455	1.3697
	57.1	ORS0543	—	0.6357	6.673		Θ_*H. a. ann.*	1.0149	1.0923	1.1778
		H. d. cuc.								
	19.6	ORS0728	7.1196	—	7.7288					
	39.8	ORS0371	7.8381	—	4.526					
	57.1	ORS0543	7.8679	—	2.4446					
		H. a. ann.								
	19.6	ORS0728	5.9438	4.0078	—					
	39.8	ORS0371	9.7326	6.2272	—					
	57.1	ORS0543	18.6335	1.5918	—					
		H. a. tex.								
2	0	ORS0229	—	4.6875	10.2633	2		0.05	MLE	0.95
	60.8	ORS0836	—	5.936	6.2919		Θ_*H. a. tex.*	1.7675	1.8542	1.9465
	63.4	ORS0203	—	1.2908	4.8347		Θ_*H. d. cuc.*	1.0864	1.1526	1.2244
	63.4	ORS0315	—	8.6969	9.1068		Θ_*H. a. ann.*	1.2753	1.3503	1.4311
	63.5	ORS1011	—	7.4229	6.5247					
		H. d. cuc.								
	0	ORS0229	11.3969	—	7.6375					
	60.8	ORS0836	13.7722	—	5.3523					
	63.4	ORS0203	21.1846	—	9.1198					
	63.4	ORS0315	10.0743	—	5.2046					
	63.5	ORS1011	6.493	—	4.2412					

Continued

			M [+=receiving population]							
		Locus*	*H. a. tex.*,+	*H. d. cuc.*,+	*H. a. ann.*,+					
		H. a. ann.								
	0	ORS0229	16.8002	4.2481	—					
	60.8	ORS0836	20.0854	10.4906	—					
	63.4	ORS0203	9.2263	1.6037	—					
	63.4	ORS0315	13.0812	5.1609	—					
	63.5	ORS1011	10.3045	6.4763	—					
		H. a. tex.						0.05	MLE	0.95
3	0	ORS0665	—	3.3686	7.306	3	Θ_*H. a. tex.*	1.6302	1.7189	1.814
	13.1	ORS1021	—	5.5322	9.0609		Θ_*H. d. cuc.*	1.1828	1.262	1.3489
	14.2	HT0745	—	4.8328	9.3562		Θ_*H. a. ann.*	1.2396	1.3231	1.4144
	38.1	ORS0488	—	7.5315	9.4107					
		H. d. cuc.								
	0	ORS0665	7.1154	—	4.8861					
	13.1	ORS1021	4.4232	—	2.2528					
	14.2	HT0745	7.7365	—	3.693					
	38.1	ORS0488	9.9103	—	6.3378					
		H. a. ann.								
	0	ORS0665	20.6081	2.9164	—					
	13.1	ORS1021	14.2042	5.7982	—					
	14.2	HT0745	10.3623	4.0125	—					
	38.1	ORS0488	17.1559	9.6249	—					
		H. a. tex.						0.05	MLE	0.95
4	0	ORS0271	—	4.9103	10.6422	4	Θ_*H. a. tex.*	1.5848	1.6567	1.7343
	36.7	HT0728	—	0.3552	9.1721		Θ_*H. d. cuc.*	1.2862	1.3677	1.4561
	37.4	ORS0341	—	2.5701	3.6935		Θ_*H. a. ann.*	1.1377	1.199	1.2646
	37.8	HT0760	—	0.5246	7.9589					
	38.1	ORS1091	—	1.9053	3.0447					
	63	HT0865	—	2.901	6.2892					
		H. d. cuc.								
	0	ORS0271	4.0995	—	2.7804					
	36.7	HT0728	2.5001	—	1					
	37.4	ORS0341	3.6324	—	9.3877					
	37.8	HT0760	1.2204	—	0.1877					
	38.1	ORS1091	5.0728	—	2.0646					
	63	HT0865	5.7477	—	7.2474					
		H. a. ann.								
	0	ORS0271	16.7642	3.7218	—					
	36.7	HT0728	19.6837	0.4832	—					
	37.4	ORS0341	10.1138	3.8761	—					
	37.8	HT0760	9.5416	0.8876	—					
	38.1	ORS1091	7.6522	2.3599	—					
	63	HT0865	12.7123	4.5875	—					
		H. a. tex.						0.05	MLE	0.95
5	0	ORS1120	—	5.9302	8.8465	5	Θ_*H. a. tex.*	1.4905	1.5844	1.6857
	8.7	ORS0547	—	7.7826	5.5048		Θ_*H. d. cuc.*	1.3618	1.4695	1.5881
	27.1	HT0311	—	4.463	4.7688		Θ_*H. a. ann.*	1.1179	1.2052	1.3022
		H. d. cuc.								
	0	ORS1120	10.561	—	9.1362					
	8.7	ORS0547	5.8837	—	7.8416					
	27.1	HT0311	9.0634	—	7.8049					
		H. a. ann.								
	0	ORS1120	12.8716	6.7419	—					
	8.7	ORS0547	10.5148	5.8948	—					
	27.1	HT0311	16.1994	11.2732	—					

Continued

			M [+=receiving population]							
		Locus*	*H. a. tex.*,+	*H. d. cuc.*,+	*H. a. ann.*,+					
		H. a. tex.						0.05	MLE	0.95
7	0	HT0938	—	7.987	7.2643	7	Θ_*H. a. tex.*	1.7914	1.9072	2.033
	14.8	HT0520	—	2.9443	9.3434		Θ_*H. d. cuc.*	0.9873	1.0601	1.1399
	17.5	ORS1246	—	3.0292	8.5054		Θ_*H. a. ann.*	1.1158	1.2046	1.3027
		H. d. cuc.								
	0	HT0938	11.077	—	4.1293					
	14.8	HT0520	2.9195	—	2.1574					
	17.5	ORS1246	1.8661	—	0.3035					
		H. a. ann.								
	0	HT0938	14.8586	5.6178	—					
	14.8	HT0520	10.3697	2.3549	—					
	17.5	ORS1246	11.5761	2.709	—					
		H. a. tex.						0.05	MLE	0.95
8	0	HT0854	—	2.8763	14.2636	8	Θ_*H. a. tex.*	1.8197	1.9346	2.0608
	13.3	ZH2114	—	2.2558	4.1847		Θ_*H. d. cuc.*	1.0155	1.0924	1.1773
	37.6	ORS0299	—	5.0968	7.1707		Θ_*H. a. ann.*	1.2433	1.3453	1.4589
		H. d. cuc.								
	0	HT0854	2.6254	—	3.6447					
	13.3	ZH2114	6.714	—	3.9666					
	37.6	ORS0299	17.7835	—	9.3113					
		H. a. ann.								
	0	HT0854	13.1997	3.4378	—					
	13.3	ZH2114	17.423	8.8379	—					
	37.6	ORS0299	18.0283	13.3482	—					
		H. a. tex.						0.05	MLE	0.95
9	0	ORS1034	—	4.9094	7.8078	9	Θ_*H. a. tex.*	1.7559	1.8589	1.9712
	6.3	ORS1076	—	2.4461	5.8145		Θ_*H. d. cuc.*	1.0487	1.1204	1.1988
	7.8	HT0834	—	1.7759	4.034		Θ_*H. a. ann.*	1.2413	1.3244	1.4151
	20.6	ORS1001	—	5.3699	6.7957					
		H. d. cuc.								
	0	ORS1034	12.7049	—	6.6875					
	6.3	ORS1076	12.5297	—	3.5901					
	7.8	HT0834	4.5322	—	7.0401					
	20.6	ORS1001	7.1771	—	8.7996					
		H. a. ann.								
	0	ORS1034	23.6782	6.9731	—					
	6.3	ORS1076	9.6109	4.126	—					
	7.8	HT0834	7.0565	3.8237	—					
	20.6	ORS1001	12.6869	7.5695	—					
		H. a. tex.						0.05	MLE	0.95
10a	0	HT0297	—	3.8251	10.8353	10a	Θ_*H. a. tex.*	1.6149	1.7437	1.886
	11.3	HT0872	—	7.8104	13.3178		Θ_*H. d. cuc.*	1.2294	1.3465	1.4788
		H. d. cuc.					Θ_*H. a. ann.*	1.2253	1.3386	1.4661
	0	HT0297	8.0089	—	4.682					
	11.3	HT0872	6.0821	—	5.0464					
		H. a. ann.								
	0	HT0297	17.3366	6.141	—					
	11.3	HT0872	12.6937	6.498	—					
10b		*H. a. tex.*				10b		0.05	MLE	0.95
	2.7	HT0419	—	7.666	7.6276		Θ_*H. a. tex.*	1.6071	1.7866	1.9665
	15.4	HT0692	—	5.3011	8.7093		Θ_*H. d. cuc.*	1.2468	1.3601	1.488
		H. d. cuc.					Θ_*H. a. ann.*	1.2244	1.3432	1.4779
	2.7	HT0419	8.401	—	5.0497					
	15.4	HT0692	4.729	—	5.4495					

Continued

			M [+=receiving population]							
		Locus*	*H. a. tex.*,+	*H. d. cuc.*,+	*H. a. ann.*,+					
		H. a. ann.								
	2.7	HT0419	11.0786	13.9899	—					
	15.4	HT0692	25.297	9.7371	—					
		H. a. tex.						0.05	MLE	0.95
11a	0	HT0544	—	6.2762	6.8639	11a	Θ_*H. a. tex.*	2.122	2.2915	2.4752
	6	ORS0457	—	3.5419	7.0267		Θ_*H. d. cuc.*	0.7338	0.8171	0.9137
		H. d. cuc.					Θ_*H. a. ann.*	1.3126	1.4371	1.5781
	0	HT0544	17.2287	—	11.5849					
	6	ORS0457	11.9092	—	9.7905					
		H. a. ann.								
	0	HT0544	20.5812	12.0788	—					
	6	ORS0457	18.2721	4.278	—					
11b		*H. a. tex.*				11b		0.05	MLE	0.95
	0	HT0732	—	7.0627	8.7897		Θ_*H. a. tex.*	1.7436	1.83759	1.9384
	1	HT0831	—	2.9565	10.0867		Θ_*H. d. cuc.*	1.211	1.29626	1.3894
	1.3	ORS0354	—	2.9498	8.6067		Θ_*H. a. ann.*	1.1956	1.27559	1.3629
	17	ORS0291	—	3.1354	8.0297					
		H. d. cuc.								
		HT0732	7.1694	—	3.5769					
		HT0831	5.924	—	3.6029					
		ORS0354	17.82	—	9.4837					
		ORS0291	7.6569	—	3.3253					
		H. a. ann.								
		HT0732	16.0309	10.28	—					
		HT0831	24.8971	5.1753	—					
		ORS0354	13.0248	3.5757	—					
		ORS0291	15.2999	2.899	—					
		H. a. tex.						0.05	MLE	0.95
13	0	ORS0215	—	0.3681	10.6382	13	Θ_*H. a. tex.*	1.6949	1.7772	1.8649
	1.2	ORS1042	—	1.8669	9.1684		Θ_*H. d. cuc.*	1.2341	1.3278	1.4311
	3.4	HT0295	—	2.2951	7.3926		Θ_*H. a. ann.*	1.3362	1.4132	1.4962
	30.7	HT0333	—	4.5838	4.3134					
	74.4	HT0382	—	8.7129	11.0122					
		H. d. cuc.								
	0	ORS0215	6.7192	—	5.7593					
	1.2	ORS1042	5.6436	—	5.6253					
	3.4	HT0295	4.0805	—	4.492					
	30.7	HT0333	9.694	—	4.227					
	74.4	HT0382	7.9678	—	6.8896					
		H. a. ann.								
	0	ORS0215	16.2755	0.3553	—					
	1.2	ORS1042	7.4681	0.867	—					
	3.4	HT0295	13.2537	2.1278	—					
	30.7	HT0333	9.6584	4.8178	—					
	74.4	HT0382	12.9363	8.9636	—					
		H. a. tex.						0.05	MLE	0.95
14	5.8	NO5468	—	5.0684	8.0437	14	Θ_*H. a. tex.*	1.4556	1.5295	1.6085
	10.7	ZH0636	—	6.2282	11.9562		Θ_*H. d. cuc.*	1.2041	1.2888	1.3826
	16.1	HT0429	—	2.0829	18.365		Θ_*H. a. ann.*	1.2968	1.3731	1.4556
	16.4	ORS1009	—	2.2969	8.0843					
	19.6	HT0823	—	3.6363	9.0035					

Continued

			M [+=receiving population]							
		Locus*	*H. a. tex.*,+	*H. d. cuc.*,+	*H. a. ann.*,+					
		H. d. cuc.								
	5.8	NO5468	4.5865	—	3.944					
	10.7	ZH0636	5.1297	—	6.3834					
	16.1	HT0429	1.9881	—	0.666					
	16.4	ORS1009	10.662	—	20.5307					
	19.6	HT0823	15.9241	—	7.5608					
		H. a. ann.								
	5.8	NO5468	22.5444	8.0885	—					
	10.7	ZH0636	8.4032	4.3867	—					
	16.1	HT0429	12.4627	0.1702	—					
	16.4	ORS1009	8.8201	2.2124	—					
	19.6	HT0823	6.6346	1.7147	—					
		H. a. tex.						0.05	MLE	0.95
15	0	ORS1215	—	0.7883	9.0305	15	Θ_*H. a. tex.*	1.6055	1.6726	1.7433
	14.6	ORS0239	—	2.7589	11.5789		Θ_*H. d. cuc.*	1.2134	1.2785	1.3484
	14.8	HT0284	—	4.1044	6.5657		Θ_*H. a. ann.*	1.1212	1.1783	1.2394
	15.3	HT0326	—	3.5153	7.1078					
	17.5	HT0290	—	5.3774	8.2675					
	32.8	HT0329	—	4.5536	7.0973					
	49.1	HT0528	—	7.3245	9.5326					
		H. d. cuc.								
	0	ORS1215	3.6097	—	1.9332					
	14.6	ORS0239	2.9804	—	3.8451					
	14.8	HT0284	7.3466	—	4.6522					
	15.3	HT0326	3.9819	—	2.8801					
	17.5	HT0290	4.563	—	1.5754					
	32.8	HT0329	11.345	—	5.2688					
	49.1	HT0528	7.8279	—	4.3237					
		H. a. ann.								
	0	ORS1215	19.1806	1.444	—					
	14.6	ORS0239	20.0231	2.5757	—					
	14.8	HT0284	11.9828	6.1188	—					
	15.3	HT0326	22.6116	6.3358	—					
	17.5	HT0290	16.7267	6.6682	—					
	32.8	HT0329	13.0883	7.8201	—					
	49.1	HT0528	20.1168	10.0289	—					
		H. a. tex.						0.05	MLE	0.95
16	0	ORS0245	—	3.3077	4.2242	16	Θ_*H. a. tex.*	1.6364	1.7032	1.7735
	0	ORS0946	—	4.6129	7.0216		Θ_*H. d. cuc.*	1.1957	1.2618	1.3328
	0.4	HT0915	—	3.6981	10.8058		Θ_*H. a. ann.*	1.1775	1.2367	1.3
	0.9	ORS1200	—	4.1829	6.647					
	5.4	ORS0788	—	3.3694	6.566					
	26.9	ORS0885	—	0	9.3958					
	36.4	HT0420	—	6.9063	3.1328					
		H. d. cuc.								
	0	ORS0245	9.1525	—	7.1747					
	0	ORS0946	13.5487	—	14.8611					
	0.4	HT0915	3.7186	—	1.3805					
	0.9	ORS1200	4.8156	—	4.2687					
	5.4	ORS0788	3.4377	—	3.7622					
	26.9	ORS0885	391.1038	—	78.2208					
	36.4	HT0420	5.4678	—	5.0225					

Continued

			M [+=receiving population]							
		Locus*	*H. a. tex.*,+	*H. d. cuc.*,+	*H. a. ann.*,+					
		H. a. ann.								
	0	ORS0245	10.5675	4.9126	—					
	0	ORS0946	16.2965	9.3065	—					
	0.4	HT0915	11.7249	2.8737	—					
	0.9	ORS1200	20.4539	7.668	—					
	5.4	ORS0788	17.7851	5.8359	—					
	26.9	ORS0885	19.005	0	—					
	36.4	HT0420	6.4956	4.454	—					
		H. a. tex.							0.05	MLE	0.95
17	0	HT0279	—	4.6482	10.6359	17	Θ_*H. a. tex.*		1.6583	1.7117	1.767
	23.8	NO2016	—	6.0835	4.4719		Θ_*H. d. cuc.*		1.2981	1.3527	1.4114
	93.4	ORS0169	—	6.9392	10.1178		Θ_*H. a. ann.*		1.0924	1.1348	1.1794
	114.9	ORS0386	—	1.3149	7.1678						
	120.2	ORS0561	—	1.1954	5.7915						
	123.6	HT0536	—	5.0984	7.3907						
	124.3	HT0414	—	6.8605	12.9574						
	124.8	ORS0495	—	5.9735	8.3954						
	130	HT0929	—	6.2172	7.8354						
	133.6	ORS0735	—	1.2289	6.9306						
	151.6	HT0289	—	3.7016	12.6755						
	152.9	HT0538	—	3.9135	6.8822						
		H. d. cuc.									
	0	HT0279	6.0602	—	3.0282						
	23.8	NO2016	8.8598	—	4.0411						
	93.4	ORS0169	12.0672	—	7.4533						
	114.9	ORS0386	14.9264	—	11.4718						
	120.2	ORS0561	17.7953	—	12.6383						
	123.6	HT0536	9.7854	—	8.57						
	124.3	HT0414	4.1389	—	2.133						
	124.8	ORS0495	8.7291	—	4.8036						
	130	HT0929	6.1823	—	4.048						
	133.6	ORS0735	3.9059	—	4.3852						
	151.6	HT0289	2.4204	—	2.0923						
	152.9	HT0538	2.1799	—	0.595						
		H. a. ann.									
	0	HT0279	14.3756	3.4483	—						
	23.8	NO2016	19.7283	9.3446	—						
	93.4	ORS0169	11.0103	5.8566	—						
	114.9	ORS0386	20.9214	0.5391	—						
	120.2	ORS0561	14.9934	1.7724	—						
	123.6	HT0536	15.3444	9.4069	—						
	124.3	HT0414	11.7502	5.7509	—						
	124.8	ORS0495	25.877	10.0357	—						
	130	HT0929	20.2168	10.3731	—						
	133.6	ORS0735	11.5361	1.9353	—						
	151.6	HT0289	25.4467	3.5986	—						
	152.9	HT0538	23.4226	2.9926	—						
Unlinked	ORS0297										
	ORS0826										

*ORS loci are derived from libraries enriched for dinucleotide repeats (Tang *et al.* 2002), whereas HT loci are derived from EST libraries (Heesacker *et al.* 2008).

H. a. ann., *Helianthus annuus annuus*; *H. d. cuc.*, *Helianthus debilis cucumerifolius*; *H. a. tex.*, *H. annuus texanus*