

Genomic Patterns of Adaptive Divergence between Chromosomally Differentiated Sunflower Species

Jared L. Strasburg,* Caroline Scotti-Saintagne,*¹ Ivan Scotti,† Zhao Lai,‡ and Loren H. Rieseberg*§

*Department of Biology, Indiana University, Bloomington; †INRA–UMR ECOFOG, Campus agronomique—Avenue de France, Kourou, French Guiana; ‡Center for Genomics and Bioinformatics, Indiana University, Bloomington; and §Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada

Understanding the genetic mechanisms of speciation and basis of species differences is among the most important challenges in evolutionary biology. Two questions of particular interest are what roles divergent selection and chromosomal differentiation play in these processes. A number of recently proposed theories argue that chromosomal rearrangements can facilitate the development and maintenance of reproductive isolation and species differences by suppressing recombination within rearranged regions. Reduced recombination permits the accumulation of alleles contributing to isolation and adaptive differentiation and protects existing differences from the homogenizing effects of introgression between incipient species. Here, we examine patterns of genetic diversity and divergence in rearranged versus collinear regions in two widespread, extensively hybridizing sunflower species, *Helianthus annuus* and *Helianthus petiolaris*, using sequence data from 77 loci distributed throughout the genomes of the two species. We find weak evidence for increased genetic divergence near chromosomal break points but not within rearranged regions overall. We find no evidence for increased rates of adaptive divergence on rearranged chromosomes; in fact, collinear chromosomes show a far greater excess of fixed amino acid differences between the two species. A comparison with a third sunflower species indicates that much of the nonsynonymous divergence between *H. annuus* and *H. petiolaris* probably occurred during or soon after their formation. Our results suggest a limited role for chromosomal rearrangements in genetic divergence, but they do document substantial adaptive divergence and provide further evidence of how species integrity and genetic identity can be maintained at many loci in the face of extensive hybridization and gene flow.

Introduction

The roles of divergent selection in speciation (Coyne and Orr 2004) and in the development and maintenance of species differences (Rieseberg et al. 2004) have been longstanding questions in evolutionary biology. As more molecular data become available, it is becoming increasingly clear that positive selection has played a major role in both molecular and phenotypic evolutionary change (Rieseberg et al. 2002; Eyre-Walker 2006). Many examples of positive selection apparently acting on individual genes or gene families have been documented (Nei [2005] and references therein). Interest in documenting positive selection within species and adaptive divergence between species is motivated not only by the ongoing debate about the relative roles of neutral and selective processes in evolution but also by the promise that identifying genes under divergent selection between species can help us to understand both gene function and the nature of adaptive phenotypic differences (Steiner et al. 2007; Barrett et al. 2008).

As more and more genomic data, including dozens to hundreds of unlinked markers (Storz 2005) or complete genome sequences (Kosiol et al. 2008), have become available for an increasing number of species, there has been increased interest in the degree to which interspecific patterns of genetic variation reflect widespread positive selection throughout the genomes of related species. By comparing the distribution of genetic divergence or evolutionary rate

measures to a neutral model, outliers undergoing divergent or balancing selection can be identified (Lewontin and Krakauer 1973; Beaumont and Balding 2004).

In chromosomally divergent species, a related question concerning the role of chromosomal rearrangements in creating and maintaining species boundaries has been a topic of debate for close to a century (Sturtevant 1938; White 1978; Ayala and Coluzzi 2005). Traditionally, chromosomal rearrangements have been thought to contribute to speciation (White 1978; King 1993) or to the maintenance of species identities in hybridizing species (Barton and Bengtsson 1986; Levin 2002) via underdominant fitness effects associated with meiotic abnormalities and the creation of unbalanced gametes in chromosomal heterozygotes. However, this view of the role of chromosomal rearrangements in speciation and species boundaries suffers from significant theoretical and empirical difficulties (Rieseberg 2001). Rearrangements must be strongly underdominant for them to effectively reduce gene flow between incipient species, but such rearrangements are very unlikely to be fixed within either species, except under very restrictive conditions involving extremely small effective population sizes or strong meiotic drive (Wright 1940, 1941; Walsh 1982; Lande 1985). More weakly underdominant rearrangements are considerably easier to fix within species, but will also contribute less to reduced fitness of hybrids and thus isolation between species. (This theoretical difficulty may be partially overcome by invoking a cascade model in which numerous rearrangements that are weakly underdominant individually are strongly underdominant in concert; White 1978.) In addition, the fitness effects of rearrangements can be quite variable, and in many cases little reduction in fertility is seen (Sites and Moritz 1987; Coyne et al. 1991, 1993; Davisson and Akeson 1993).

¹ Present address: INRA–UMR ECOFOG, Campus agronomique—Avenue de France, Kourou, French Guiana.

Key words: species boundaries, chromosomal rearrangements, positive selection, hybridization, sunflowers, *Helianthus*.

E-mail: jstrasbu@indiana.edu.

Mol. Biol. Evol. 26(6):1341–1355. 2009

doi:10.1093/molbev/msp043

Advance Access publication March 10, 2009

Recently, a number of related models have been proposed in which chromosomal rearrangements can reduce gene flow and potentially contribute to speciation indirectly through the suppression of recombination rather than directly through underdominance. Rieseberg (2001) suggested that if a gene contributing to reproductive isolation between species is located within a chromosomal rearrangement, the isolating effects of that gene could extend much farther along the chromosome due to recombination suppression. Depending on the number of rearrangements by which species differ and how strongly recombination is suppressed within them, relatively few isolation genes could effectively block introgression for a significant portion of the genome. This model was partly motivated by evidence that pollen sterility quantitative trait loci (QTL) may cluster near chromosomal break points in sunflowers (although some sterility is probably due to the rearrangements themselves; Gardner et al. 2000; Lai, Nakazato, et al. 2005) and that introgression was suppressed across much larger linkage blocks in rearranged regions versus collinear regions (Rieseberg et al. 1999). Noor et al. (2001) proposed a closely related, nonmutually exclusive model in which asymmetrically acting isolation alleles, which would normally be removed from the incipient species by selection, may be maintained if such alleles from both species are found within the same chromosomal inversion. If these isolation alleles are prevented from recombining, neither allele or arrangement conferring sterility in the alternate genetic background can be removed by selection and neither arrangement can introgress across the incipient species boundary. This model is supported by both empirical data and theory. A number of asymmetric incompatibilities have been documented (Coyne and Orr 1989; Palopoli and Wu 1994), and loci conferring reproductive isolation between *Drosophila pseudoobscura* and *Drosophila persimilis* are located almost exclusively in genomic regions that show fixed chromosomal differences between the two species (Orr 1987; Noor et al. 2001a and 2001b). Navarro and Barton (2003a) modeled the effects of chromosomal rearrangements between hybridizing species and found that they could greatly facilitate the accumulation of incompatibility alleles. Kirkpatrick and Barton (2006) suggested a different mechanism by which adaptive species-specific differences may be preferentially found in inverted regions—if two or more alleles conferring fitness benefits in a particular environment or genetic background are captured within the same inversion, they become tightly linked due to recombination suppression within the inversion; such inversions can increase rapidly in frequency due to the cumulative effects of multiple adaptive alleles.

A common prediction of these recombination suppression models is that interspecific gene flow will be reduced across rearranged chromosomal regions. In species pairs with a history of gene flow, this should facilitate local adaptation and the preferential accumulation of genetic differences (including hybrid incompatibilities) on rearranged chromosomes (Noor et al. 2001b; Rieseberg 2001; Navarro and Barton 2003a; Kirkpatrick and Barton 2006). The widespread annual sunflowers of the genus *Helianthus* are ideal for testing predictions about the roles of chromosomal rearrangements and divergent selection in the main-

tenance of species integrity despite ongoing gene flow. *Helianthus annuus* and *Helianthus petiolaris* are strongly differentiated in karyotype (Burke et al. 2004; Lai, Nakazato, et al. 2005), morphology (Rosenthal et al. 2002, 2005), and ecological preference, most notably with respect to soil moisture and salt content (Heiser 1947; Gross et al. 2004; Karrenberg et al. 2006). At the same time, they hybridize extensively throughout their shared range, have high rates of complex backcross production (Rieseberg et al. 1999), and have long-term rates of gene flow of approximately $N_{em} = 0.5$ in each direction (Strasburg and Rieseberg 2008).

Here, we compare patterns of sequence diversity and differentiation between *H. annuus* and *H. petiolaris* at 77 loci distributed across collinear and rearranged chromosomes to test for correlations between chromosomal rearrangements and patterns of divergence and to examine patterns of natural selection across the two species' genomes. We also compare each of these two species to a third, historically allopatric species, *Helianthus argophyllus*, at a subset of these loci in order to address the degree to which long-term sympatry and hybridization have affected patterns of genetic differentiation among these species.

Materials and Methods

Collections and DNA Isolation

Achenes were collected from four *H. annuus* and three *H. petiolaris* populations distributed across the sympatric area of their distribution (fig. 1; locality information given in Yatabe et al. [2007]). Only locally allopatric populations were sampled to exclude early generation hybrids between the two species. A fourth *H. petiolaris* population was sampled but later determined to be of hybrid origin based on sequence and flow cytometry data (not shown); this population was excluded from all analyses. Achenes were also collected from three *H. argophyllus* populations. Achenes were germinated in greenhouses at Indiana University. DNA was extracted from leaf tissue for population genetic analyses from two individuals per population; in total, eight *H. annuus* and six *H. petiolaris* individuals were employed for sequencing. For *H. argophyllus*, leaf tissue was collected from two individuals each from three populations (fig. 1). All DNA extractions were performed using a QIAGEN Dneasy Plant kit (Qiagen, Valencia, CA).

DNA Sequencing

A transcript map was recently generated for sunflower based on the mapping of single-nucleotide polymorphisms (SNPs) for 243 expressed sequence tags (ESTs—Lai, Livingstone, et al. 2005). In the present paper, 139 mapped ESTs distributed across three collinear linkage groups (1, 3, and 10) and five rearranged chromosomes (12, 13, 14, 16, and 17) were tested for single band amplification in *H. annuus* and *H. petiolaris*. Of these, 92 ESTs were selected for sequencing tests and finally 77 were retained for the full analysis between these two species (33 collinear and 44 rearranged). Relationships among the five rearranged chromosomes between *H. annuus* and *H. petiolaris* that were

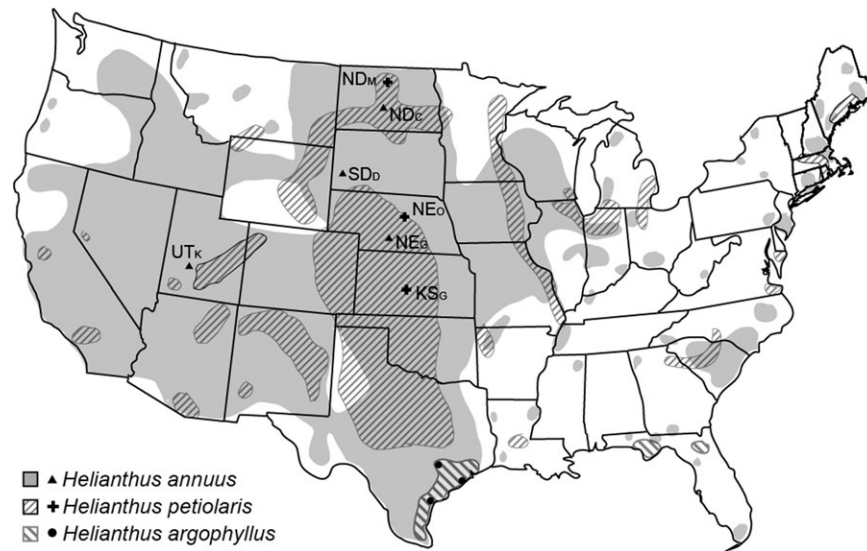


FIG. 1.—Collecting localities for *Helianthus annuus*, *Helianthus petiolaris*, and *Helianthus argophyllus* individuals used in this study. Two individuals were sampled from each population. Names for *H. annuus* and *H. petiolaris* populations correspond to those in table 2.

included in this study, as well as the locations of all 77 loci, are shown in figure 2. Twenty-two of these loci were also sequenced in *H. argophyllus*. Sequencing reactions were performed on polymerase chain reaction (PCR) products previously cleaned using EXOSAP-IT (USB, Cleveland, Ohio). Sequencing reactions were carried out in a total volume of 10 μ l containing 2 μ l of water, 3 μ l of 5 mM MgCl₂, 2 μ l of 2 μ M primer, 2 μ l of cleaned-up PCR product (equivalent from 10 to 20 ng of DNA), and 1 μ l of ABI Big Dye version 3.1. Sequencing reactions were then purified with the magnetic bead CleanSeq kit from Agencourt and loaded on an ABI 3730 capillary sequencer (Applied Biosystems, Foster City, CA).

Sequences were aligned using CodonCode Aligner version 1.2.0 (Codoncode Corporation, Dedham, MA) or Sequencher version 4.7 (Gene Codes Corporation, Ann Arbor, MI). Individuals heterozygous for indels were cloned using the Topo-TA cloning kit with DH5 α -T1^R one shot chemically competent cells (Invitrogen, Carlsbad, CA). DNA plasmid isolation was then performed using QIAprep Miniprep (Qiagen). Sequencing reactions were directly performed on purified plasmids using the same protocol as for PCR products. All sequences have been submitted to GenBank; accession numbers are given in supplementary table S1 (Supplementary Material online).

Data Analyses

Because PCR products were sequenced directly, it was not always possible to infer allelic phase. In addition, phase was not applicable for the concatenated sequences. Thus, all analyses considered genotypic rather than haplotypic data. We have included some measures of uncorrected average pairwise sequence diversity and divergence based on arbitrarily phased SNPs; these measures are not affected by SNP phase. We have taken care not to use measures involving full haplotypic data or corrected sequence diversity/

divergence, which would be affected by SNP phase. Because EST sequences were short (<500 bp), most calculations and tests were performed for each EST separately and for concatenated sequences of ESTs from collinear and rearranged linkage groups or of all ESTs together. DnaSP 4.50.1 software (Rozas et al. 2003) was used to calculate various measures of sequence diversity and divergence, including gross and net divergence between species (net divergence = gross divergence – average diversity within each species) as well as Pi(a)/Pi(s) and Ka/Ks ratios (of nonsynonymous to synonymous intraspecific diversity and interspecific divergence, respectively) and to perform McDonald–Kreitman (MK) tests (McDonald and Kreitman 1991). The total number of synonymous and nonsynonymous sites for a set of sequences was estimated by averaging the number of synonymous and nonsynonymous sites over all sequences when more than two sequences were compared. We attempted to correct for a possible loss of power in the MK test due to weakly deleterious variants possibly segregating at low frequency within each species (Fay et al. 2001) by performing the test on the entire data set, on only polymorphic sites with common variants (minor allele frequency >2/28) and on only polymorphic sites with rare variants (minor allele frequency \leq 2/28). For sites with more than two alleles segregating, the frequency of the rarest allele was used. We also attempted to correct for bias in the MK test due to past changes in selective constraints (due, e.g., to the significant increase in effective population size *H. annuus* and *H. petiolaris* have undergone since their initial divergence—Strasburg and Rieseberg 2008) by comparing rates of amino acid divergence above neutral expectations across loci. If lower selective constraints in the past due to smaller historical effective population sizes are the cause of the excess amino acid divergence, then those demographic changes should affect the entire genome, whereas positive selection is only expected to affect a subset of genes (Fay et al. 2002). We estimated the excess of amino acid differences between

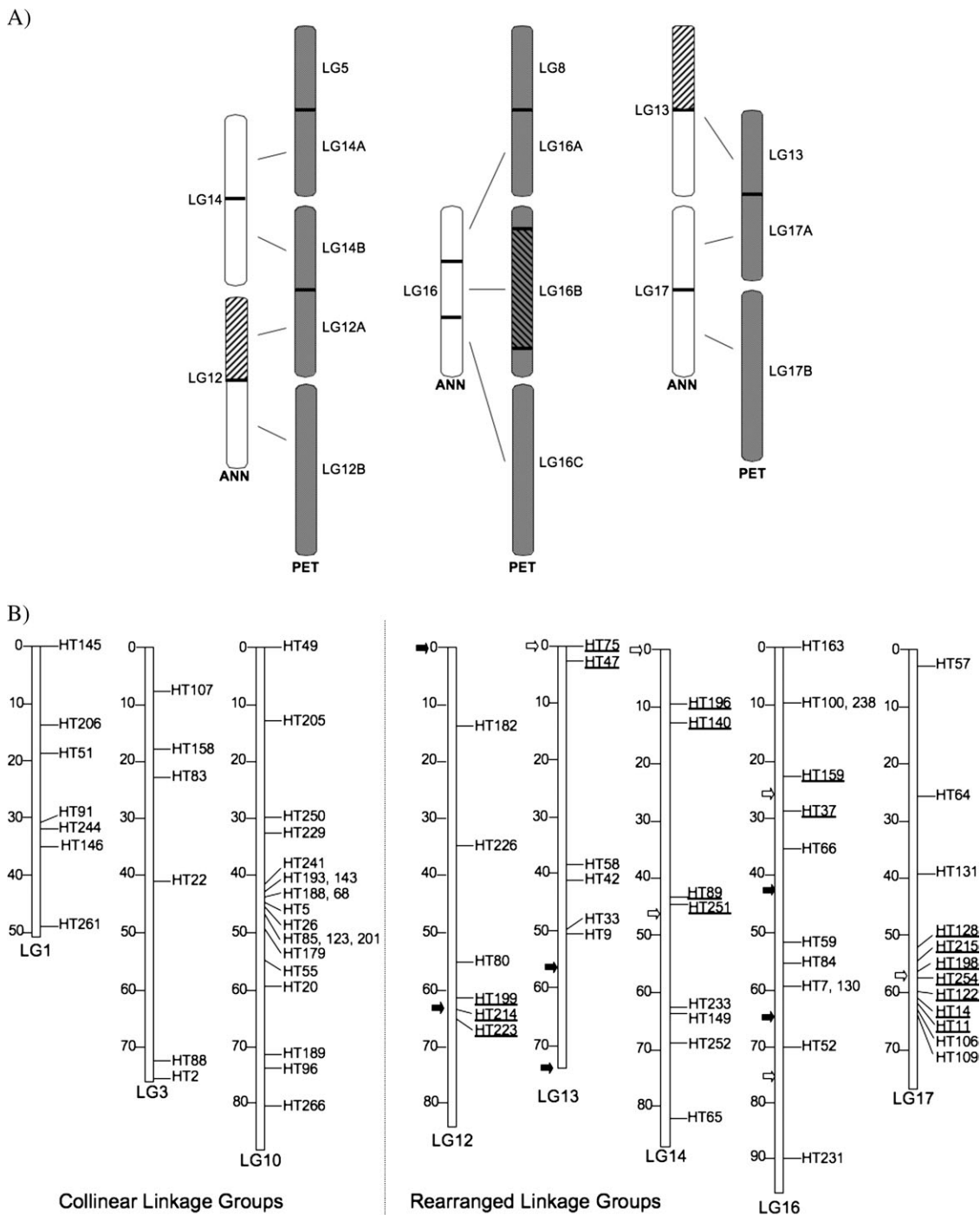


FIG. 2.—Chromosomal relationships and locations of sampled loci. (A) Relationships among five rearranged chromosomes between *Helianthus annuus* and *Helianthus petiolaris* that were included in this study. Comparative mapping information from Burke et al. (2004) and Lai, Nakazato, et al. (2005). (B) Locations of sequenced markers from both collinear and rearranged chromosomes. Filled arrows indicate the positions of inversion break points, whereas unfilled arrows refer to translocation break points. Markers within 5 cM of the break points are underlined. Note that markers HT196 and HT140 on LG14 map within 5 cM of a translocation break point in both the *H. petiolaris* mapping population and the composite mapping population used for comparative mapping (Lai, Nakazato, et al. 2005). Hence, they are considered to be within 5 cM of the break point for the purposes of this study. Map positions are from Lai, Livingstone, et al. (2005), except for LG13, which consisted of two smaller linkages in the initial analysis due to low marker density. Therefore, we employed a denser map for LG13 from the sunflower CMap database (<http://www.sunflower.uga.edu/cmap/>), developed by the laboratory of Steven J. Knapp (University of Georgia).

species over neutral expectations as the number of amino acid differences minus the number of silent differences times the ratio of amino acid to silent polymorphisms, $D_a - D_s$ • (P_a/P_s), after Fay et al. (2001).

Autocorrelation of genetic differentiation among loci along each linkage group was examined using the program SPAGeDi (Hardy and Vekemans 2002). Data from 99 mapped microsatellite loci (Yatabe et al. 2007) were also

Table 1
Uncorrected Sequence Diversity within Species

		π , all (%)	π , silent (%)
<i>Helianthus annuus</i>	All	1.14	2.04
	Collinear	1.08	2.09
	Rearranged	1.17	2.02
<i>Helianthus petiolaris</i>	All	1.07	2.07
	Collinear	0.86	1.74
	Rearranged	1.19	2.23
<i>Helianthus argophyllus</i>	All	0.55	0.94

included, and genetic differentiation was measured using the Gamma'st and G'st (Hedrick 2005) statistics for sequence and microsatellite data, respectively. Following Scotti-Saintagne et al. (2004), we tested for correlations between genetic differentiation and map distance using Moran's index (Sokal and Oden 1978); significance was tested using 20,000 permutations. A genome scan for divergent selection on individual SNPs was performed using the program BayesFst (Beaumont and Balding 2004), with all populations within a species combined and independent normal priors for all locus parameters. Each analysis was run for 3.2 million steps following a 640,000 step burn-in; 10,000 values were used to simulate the posterior distribution. Output from BayesFst was analyzed using the CODA package of R version 2.8.0 (<http://www.r-project.org/>), using a one-tailed 0.05 cutoff for sites showing evidence of divergent selection. Four runs with separate random number seeds gave identical results for each species pair.

Results

Genetic Diversity and Differentiation

Seventy-seven ESTs were sequenced in eight *H. annuus* individuals (two individuals per population from four populations) and six *H. petiolaris* individuals (two individuals per population from three populations). Among the ESTs, 17 were composed entirely of noncoding regions, 37 were composed entirely of coding regions, and 23 contained both. The average aligned sequence length was 154 bp, with a minimum of 53 bp and a maximum of 487 bp. The total aligned length of the concatenated sequences was 11,869 bp (4,052 bp and 7,817 bp for loci on collinear and rearranged chromosomes, respectively). Twenty-two of these loci were also sequenced in six *H. argophyllus* individuals (two individuals per population from three populations); the total aligned length of these 22 loci was 5,085 bp. Additional basic information and summary statistics for each locus individually are given in supplementary table S2 (Supplementary Material online).

Within-species diversity data are given in table 1. For both *H. annuus* and *H. petiolaris*, overall diversity is lower on collinear chromosomes than on rearranged chromosomes, although differences are fairly small. Both of these species have far more genetic diversity than *H. argophyllus*, as expected based on their respective ranges and on previous work (Strasburg and Rieseberg 2008, forthcoming).

Measures of genetic distance among populations within *H. annuus* and *H. petiolaris* and between the two

Table 2
Percent Net Sequence Divergence between Populations (above diagonal) and Nei's (1982) γ_{ST} Genetic Distance between Populations (below diagonal), Calculated from the Concatenated Sequence of All 77 Loci and for Loci on Collinear and Rearranged Chromosomes Separately

		<i>Helianthus annuus</i>				<i>Helianthus petiolaris</i>		
		SD _D	ND _C	NE _G	UT _K	KS _G	ND _M	NE _O
All Sequences (77 Loci)								
<i>H. annuus</i>	SD _D		0.17	0.15	0.26	0.89	0.98	0.93
	ND _C	0.22		0.20	0.24	0.84	0.93	0.89
	NE _G	0.22	0.23		0.27	0.84	0.91	0.89
	UT _K	0.26	0.24	0.26		0.78	0.89	0.86
<i>H. petiolaris</i>	KS _G	0.46	0.42	0.43	0.41		0.24	0.21
	ND _M	0.47	0.44	0.45	0.43	0.25		0.17
	NE _O	0.47	0.44	0.45	0.43	0.24	0.22	
Collinear (33 Loci)								
<i>H. annuus</i>	SD _D		0.16	0.37	0.34	0.97	1.07	0.89
	ND _C	0.23		0.33	0.32	0.88	1.03	0.87
	NE _G	0.32	0.31		0.47	0.96	1.07	0.95
	UT _K	0.30	0.30	0.36		0.71	0.87	0.70
<i>H. petiolaris</i>	KS _G	0.45	0.43	0.46	0.38		0.34	0.21
	ND _M	0.51	0.50	0.52	0.45	0.34		0.11
	NE _O	0.44	0.43	0.46	0.38	0.26	0.22	
Rearranged (44 Loci)								
<i>H. annuus</i>	SD _D		0.17	0.04	0.21	0.84	0.93	0.95
	ND _C	0.22		0.13	0.20	0.82	0.88	0.90
	NE _G	0.17	0.20		0.16	0.78	0.83	0.85
	UT _K	0.24	0.22	0.22		0.82	0.91	0.94
<i>H. petiolaris</i>	KS _G	0.38	0.34	0.34	0.34		0.18	0.21
	ND _M	0.38	0.34	0.34	0.35	0.22		0.20
	NE _O	0.41	0.36	0.36	0.38	0.23	0.23	

NOTE.—Population names correspond to those in figure 1.

species are given in table 2. Considering all 77 loci together, the genetic distances between populations of the same species are very similar based on Nei's (1982) γ_{ST} genetic distance (values range from 0.22 to 0.26) and on net sequence divergence (0.15–0.27%). Distances between populations of different species are also fairly consistent across population pairs, ranging from γ_{ST} of 0.41 to 0.47 and net sequence divergence of 0.78–0.98%. These results are consistent with previous studies showing high levels of gene flow and relatively little genetic structure within each species (Schwarzbach and Rieseberg 2002; Gross et al. 2003; Strasburg and Rieseberg 2008). When loci on collinear and rearranged chromosomes are considered separately, there is somewhat more variation in comparisons among individual populations, but no consistent trend toward increased divergence for either class of loci.

Sequence divergence among species pairs overall is given in table 3. Gross divergence between *H. annuus* and *H. petiolaris* for loci on rearranged chromosomes is somewhat higher than for loci on collinear chromosomes (although not significantly so), but this difference largely disappears in the comparison of net divergence, which takes into account diversity within each species. Interestingly, net sequence divergence between *H. annuus* and *H. argophyllus* is higher than between *H. annuus* and *H. petiolaris*, despite the fact that the former species pair is roughly 40% younger than the latter pair (1.1 vs. 1.8 My—Strasburg and Rieseberg, forthcoming). These results are consistent

Table 3
Uncorrected Sequence Divergence among Species Pairs

	Gross divergence (%)	Net divergence (%)	Shared polymorphisms	Fixed differences
<i>Helianthus annuus</i> / <i>Helianthus petiolaris</i> all	1.81	0.71	113	26
<i>H. annuus</i> / <i>H. petiolaris</i> collinear	1.64	0.68	36	7
<i>H. annuus</i> / <i>H. petiolaris</i> rearranged	1.90	0.73	77	19
<i>H. annuus</i> / <i>Helianthus argophyllus</i>	1.76	0.86	28	16
<i>H. petiolaris</i> / <i>H. argophyllus</i>	2.15	1.37	12	41

with genome-wide microsatellite divergence data showing *H. annuus* and *H. petiolaris*, which are broadly sympatric, to be more similar to each other than are *H. annuus* and *H. argophyllus*, which have historically been allopatric (Yatabe et al. 2007). *Helianthus annuus* and *H. argophyllus* have come into contact within the past ~100 years and are exchanging genetic material in their area of range overlap (Strasburg and Rieseberg, forthcoming), but the *H. annuus* populations sampled here are well outside the *H. argophyllus* range (see fig. 1).

Over the 11,869 bp distributed across eight chromosomes, only 26 fixed nucleotide differences (plus one fixed indel) were found (table 3) between *H. annuus* and *H. petiolaris*. An additional 649 mutations were polymorphic in one species and monomorphic in the other and 113 mutations were shared by the two species. Close to three times more fixed differences are found on rearranged chromosomes than collinear chromosomes, but the total length of sequence from rearranged chromosomes is almost twice as high, and the ratio of fixed differences to sequence length is not significantly different between the two groups ($\chi^2 = 0.32$, degrees of freedom [df] = 1, $P = 0.44$).

The distributions of gross and net sequence divergence values by locus are shown in figure 3. Loci found on collinear and rearranged chromosomes are not significantly different from each other for either average gross (one-tailed t -test, $t = 0.65$, df = 75, $P = 0.27$) or net ($t = 0.11$, df = 75, $P = 0.46$) divergence. Considering rearranged chromosomes only, nine loci map to regions within three inversions that distinguish the two species (see fig. 2), where effective recombination suppression is likely to be high, whereas 35 loci are outside of inversions, including areas within large translocations, where recombination suppression may be fairly limited except near break points (Grant 1975). Net sequence divergence is very similar for loci within inversions and loci outside of inversions (0.69% vs. 0.66%, $t = 0.09$, df = 42, $P = 0.47$). If a single dramatic outlier, HT080 (see fig. 3C), is excluded, loci within inversions show substantially lower divergence, and the difference is marginally significant (0.39% vs. 0.66%, $t = 1.64$, df = 41, $P = 0.052$). Sample sizes are obviously small here, but there does not appear to be any effect of inversions generally on increasing genetic divergence between the two species.

When we focus on loci that are closely linked (within 5 cM) to a chromosomal break point, where recombination suppression is expected to be highest, a marginally significant trend arises, in which the 18 loci near break points have greater net sequence divergence than the 26 loci

on rearranged chromosomes but not near break points (0.84% vs. 0.55%, $t = 1.44$, df = 42, $P = 0.071$). In this case, if the outlier, HT080 (see fig. 3D), is excluded, the difference becomes significant (0.84% vs. 0.45%, $t = 2.35$, df = 41, $P = 0.014$). If there is an effect of rearrangements on genetic divergence, it appears to be localized to areas very near break points.

For both collinear and rearranged genes, the majority show virtually no differentiation (net sequence divergence less than 0.5%), with a long tail of more strongly differentiated genes. This pattern is consistent with expectations under fairly high gene flow with diversifying selection at a number of loci (Le Corre and Kremer 2003; Latta 2004), as is the case with these two species. The farthest outlier for both gross and net divergence, HT123, is an unknown protein found on a collinear chromosome. Another outlier for net divergence, HT080, is found on a rearranged chromosome and contains 3 of the 27 fixed differences between the species (see below).

No autocorrelations were observed in levels of genetic differentiation along linkage groups for any map distance class (supplementary fig. S1, Supplementary Material online). This pattern holds true when all loci are analyzed together or when collinear and rearranged chromosomes are considered separately. It also holds when sequence data are analyzed alone or with the previously published microsatellite data (Yatabe et al. 2007).

Natural Selection

Results of MK tests are given in table 4. There is a dramatic excess of fixed amino acid differences between *H. annuus* and *H. petiolaris* compared with nonsynonymous polymorphism within each species when standardized by silent fixed differences/polymorphism, and this excess is largely explained by loci found on collinear chromosomes—half of all fixed amino acid differences are found on collinear chromosomes, compared with just 1/8 of all fixed synonymous or noncoding sequence differences. This pattern holds when only common polymorphisms or only rare polymorphisms are considered, indicating that it is not significantly affected by slightly deleterious nonsynonymous mutations segregating at low frequencies in either species (in fact, patterns of excess amino acid divergence are very slightly higher when only rare polymorphisms are considered). The silent substitutions in table 4 include both synonymous substitutions in protein-coding regions as well as substitutions in noncoding

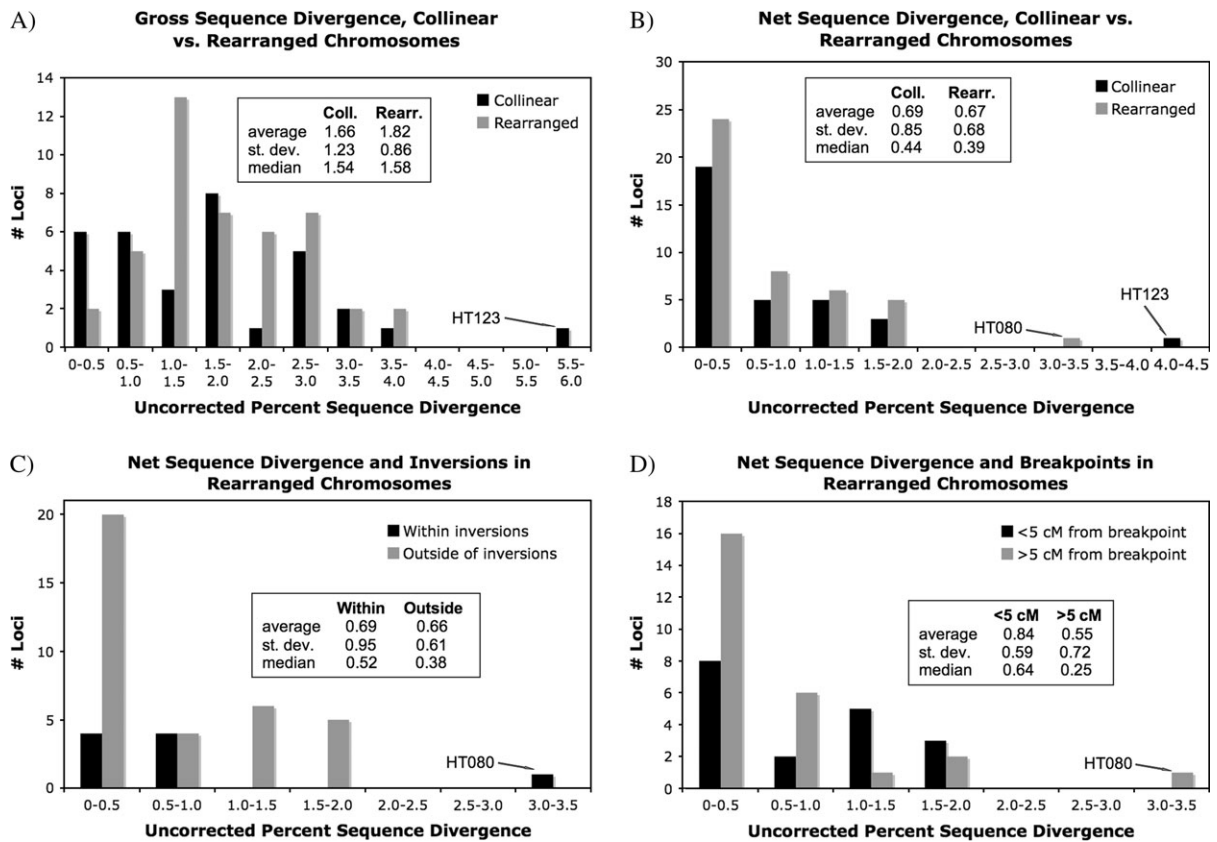


FIG. 3.—Distribution of sequence divergence values among loci on collinear and rearranged chromosomes. (A) Gross divergence, collinear versus rearranged. (B) Net divergence, collinear versus rearranged. (C) Net divergence, rearranged within inversions versus outside of inversions. (D) Net divergence, rearranged within 5 cM of a chromosomal breakpoint versus greater than 5 cM from a breakpoint.

regions. MK tests restricted to coding regions are qualitatively similar, with the same set of comparisons showing significant differences in patterns of amino acid divergence versus polymorphism (supplementary table S3, Supplementary Material online). Taking all sites together, the excess of amino acid differences between *H. annuus* and

H. petiolaris over neutral expectations is 7.5 or 75% of amino acid differences. Excess amino acid differences for loci on collinear and rearranged chromosomes are 4.6 and 2.9, respectively, despite the fact that rearranged loci contain almost 75% more coding sequence (5,138 vs. 2,948 bp).

Table 4
MK (1991) Tests for Selection between Each Species Pair

	Polymorphisms	Chromosomes	P_a	P_s	P_a/P_s	D_a	D_s	D_a/D_s	G -test value (P value)
<i>Helianthus annuus</i> / <i>H. petiolaris</i>	All	All	103	655	0.16	10	16	0.63	9.50 (0.0021)
		Collinear	34	182	0.19	5	2	2.5	10.30 (0.0013)
		Rearranged	69	473	0.15	5	14	0.36	2.42 (0.12)
	Rare	All	57	377	0.15	10	16	0.63	9.66 (0.0019)
		Collinear	14	87	0.16	5	2	2.5	10.81 (0.0010)
		Rearranged	43	290	0.15	5	14	0.36	2.28 (0.13)
	Common	All	46	278	0.17	10	16	0.63	8.39 (0.0038)
		Collinear	20	95	0.21	5	2	2.5	9.10 (0.0026)
		Rearranged	26	183	0.15	5	14	0.36	2.39 (0.12)
<i>H. annuus</i> / <i>H. argophyllus</i>	All	All	68	208	0.33	2	14	0.14	1.40 (0.24)
	Rare	All	37	107	0.35	2	14	0.14	1.55 (0.21)
	Common	All	31	101	0.31	2	14	0.14	1.12 (0.29)
<i>H. petiolaris</i> / <i>H. argophyllus</i>	All	All	42	200	0.21	16	25	0.64	8.87 (0.0029)
	Rare	All	27	120	0.23	16	25	0.64	7.13 (0.0076)
	Common	All	15	80	0.19	16	25	0.64	8.29 (0.0040)

NOTE.—Silent sites here include both synonymous sites and noncoding sites; analogous MK tests of coding sequence only are in supplementary table S3 (Supplementary Material online). Polymorphisms are classified as rare if the minor allele occurs at a frequency of 2/28 or less (2/24 or less for the *Helianthus petiolaris*/*Helianthus argophyllus* comparison). P_a and P_s are nonsynonymous and silent polymorphisms within species, respectively, and D_a and D_s are nonsynonymous and silent fixed differences between species.

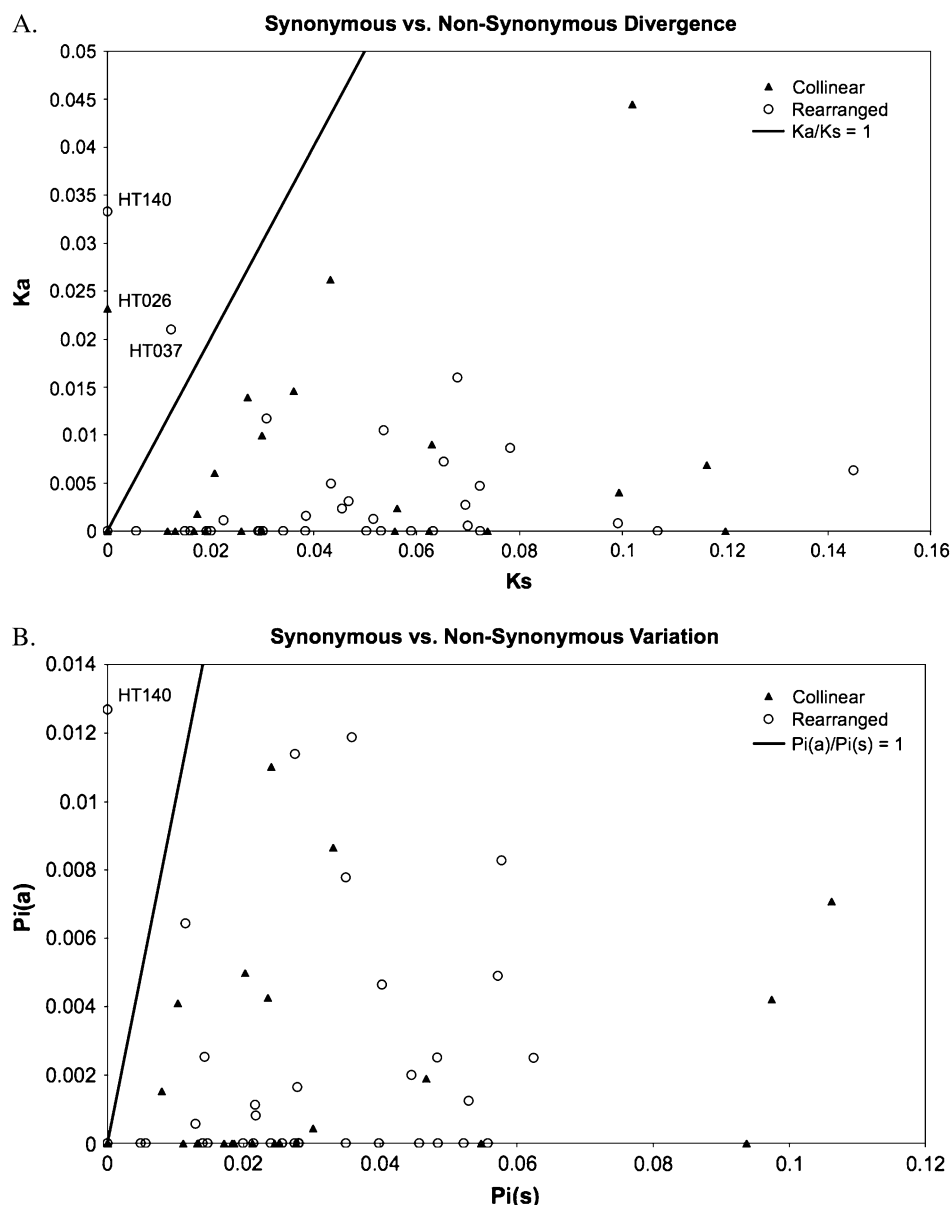


FIG. 4.—Ratios of nonsynonymous to synonymous (A) divergence and (B) polymorphism among loci on collinear versus rearranged chromosomes. Polymorphism data are averages between *Helianthus annuus* and *Helianthus petiolaris*.

Helianthus annuus and *H. argophyllus* show no evidence of divergence due to natural selection based on MK tests. In contrast, *H. petiolaris* and *H. argophyllus* show a significant excess of amino acid divergence. Ratios of nonsynonymous to silent divergence are virtually identical for *H. annuus*/*H. petiolaris* and *H. petiolaris*/*H. argophyllus*; the ratio of nonsynonymous to silent diversity is somewhat higher for *H. petiolaris*/*H. argophyllus* but not significantly so.

Plots of synonymous versus nonsynonymous divergence and diversity values for loci containing coding regions (25 and 35 loci on collinear and rearranged chromosomes, respectively) for *H. annuus* and *H. petiolaris* are shown in figure 4. Diversity values are averages between the two species. As expected, the large majority of loci have Ka/Ks or $Pi(a)/Pi(s)$ ratios less than one, indicating a prominent role for purifying selection; half of all

protein-coding loci show no amino acid divergence and more than half show no amino acid polymorphism. Ratios for individual loci must be interpreted carefully given their short length, but there are no obvious differences in the distribution of nonsynonymous to synonymous divergence or diversity ratios for loci in collinear versus rearranged chromosomes. Excluding loci with no synonymous divergence, the unweighted average Ka/Ks ratio for collinear loci, 0.13, is not significantly different from the average for rearranged loci, 0.10 ($t = 0.51$, $df = 53$, $P = 0.59$). Likewise, $Pi(a)/Pi(s)$ ratios are not significantly different (0.08 for collinear loci, 0.09 for rearranged loci; $t = 0.36$, $df = 53$, $P = 0.76$). However, when concatenated sequences are considered together, Ka/Ks is almost twice as high for collinear loci as for rearranged loci (0.18 vs. 0.10), whereas $Pi(a)/Pi(s)$ is 0.10 for both classes.

Two of the three loci with Ka/Ks ratios greater than one also have fixed differences between *H. annuus* and *H. petiolaris* that were found to be outliers in a genome scan for positive selection (table 5). All sites found to be significantly differentiated in the genome scan have fixed differences between *H. annuus* and *H. petiolaris*; 26 sites have fixed nucleotide differences (7 on collinear chromosomes and 19 on rearranged chromosomes), and 8 sites are part of a fixed 8-bp indel difference between the two species. Counting the indel, which is located on a collinear chromosome, as a single site, there is not a significant difference between the proportion of fixed differences on collinear chromosomes versus rearranged chromosomes compared with overall sequence length ($\chi^2 = 0.08$, $df = 1$, $P = 0.62$). Among the loci that are found on rearranged chromosomes, those within 5 cM of a chromosomal break point between *H. annuus* and *H. petiolaris* are marginally over-represented among loci showing fixed differences between the two species, although the trend is not significant. Six of the 10 loci showing fixed differences are within 5 cM of a break point, compared with 18 of 44 rearranged loci overall (Fisher's exact test, $P = 0.23$). Considering individual sites, 13 of 19 fixed differences are within 5 cM of break points, compared with 3,765 of 7,817 sites overall ($\chi^2 = 2.36$, $df = 1$, $P = 0.078$). These fixation events are not necessarily independent; in particular, four loci within 5 cM of the same break point on chromosome 17 account for 8 of the 19 fixed differences on rearranged chromosomes. As seen in the MK test results, amino acid differences are over-represented among fixed differences compared with the ratio of nonsynonymous to synonymous polymorphism within species, and this overrepresentation is largely due to fixed differences on collinear chromosomes; amino acid differences represent 63% of fixed differences on collinear chromosomes (including the 8-bp indel as a single difference) compared with just 26% of fixed differences on rearranged chromosomes. Interestingly, none of the eight fixed differences around the break point on chromosome 17 are nonsynonymous changes. Although a total of 16 and 41 sites show fixed differences between *H. annuus/H. argophyllus* and *H. petiolaris/H. argophyllus*, respectively, no sites were found to be more differentiated than neutral expectations based on Bayesian analyses. Presumably, this is because fewer sites were available for analysis and the average differentiation is higher for these two species pairs (especially *H. petiolaris/H. argophyllus*).

An excess of amino acid differences between species relative to amino acid polymorphism within species may be due to natural selection, but it may also be explained by reduced selective constraints due to smaller effective population sizes in the past (Eyre-Walker 2002). Both *H. annuus* and *H. petiolaris* have undergone significant population growth since their divergence (Strasburg and Rieseberg 2008), so such an explanation is plausible here. To disentangle the effects of demographic changes from divergent selection, we plotted the amount of amino acid divergence between *H. annuus* and *H. petiolaris* against the excess of amino acid divergence over neutral expectations, equal to $Ka - Ks \cdot [Pi(a)/Pi(s)]$ (Fay et al. 2001) for loci at which there is any amino acid variation. As seen in figure 5, there is a significant correlation between the rate

of amino acid divergence between the two species and the excess of amino acid divergence relative to amino acid polymorphism within species, indicating that not all loci have been affected equally by a genome-wide increase in selective constraint. We find that genes that are evolving more quickly have a significantly higher excess amino acid divergence, as Fay et al. (2002) found with their "fast" versus "neutral" genes in a comparison of *Drosophila melanogaster* and *Drosophila simulans*. The relationship between Ka and excess of Ka is highly significant ($r^2 = 0.286$, $df = 28$, $P = 0.0012$); and if the most divergent gene, an outlier that shows essentially no excess of amino acid divergence, is removed, the relationship becomes even stronger ($r^2 = 0.585$, $df = 27$, $P < 0.0001$). This pattern is more consistent with a subset of loci being under divergent selection than with a genome-wide excess of amino acid divergence due to lowered selective constraints at some point in the past.

Discussion

We have attempted to address a number of questions concerning the evolution and differentiation of two widespread, broadly sympatric annual sunflower species, *H. annuus* and *H. petiolaris*. 1) Is there molecular evidence for positive selection in the divergence of these two species? 2) What effect, if any, do chromosomal rearrangements have in the partitioning of sequence divergence generally and nonsynonymous divergence specifically between the two species? 3) How do genomic patterns of diversity and divergence between *H. annuus* and *H. petiolaris* compare with patterns between these two species and a third, historically allopatric species, *H. argophyllus*?

We found considerable evidence for divergent selection between *H. annuus* and *H. petiolaris*, with four times as many fixed amino acid differences as predicted under neutrality based on the ratio of synonymous to nonsynonymous polymorphism within each species. Our results add to a growing body of literature documenting significant divergent selection at the molecular level, at the level of both individual genes and large groups of genes or entire genomes (e.g., Fay et al. 2001; Barrier et al. 2003; Kosiol et al. 2008; Minder and Widmer 2008; Vamathevan et al. 2008). Our data indicate that approximately 75% of fixed amino acid differences between *H. annuus* and *H. petiolaris* were driven by natural selection (with an admittedly small sample size, given that the two species show very few fixed differences at nonsynonymous or silent sites). This number is within a range of values from a number of studies, mostly in *Drosophila* and primates, finding that anywhere from 35% to more than 90% of nonsynonymous substitutions are driven by positive selection (Fay et al. 2002; Sawyer et al. 2003, 2007; Bierne and Eyre-Walker 2004; Obbard et al. 2006). Comparably high rates of adaptive divergence have also been described in bacteria (Charlesworth and Eyre-Walker 2006) and viruses (Nielsen and Yang 2003). Other estimates have been considerably lower (Bustamante et al. 2005) and vary considerably among taxonomic groups (Eyre-Walker 2006) but overall rates of adaptive evolution are quite high, and many of these

Table 5
Outlier SNPs indicating positive selection in a genome scan of 809 polymorphic sites using the method implemented in BayesFst (Beaumont and Balding 2004)

Genome region	Linkage Group	Map distance (cM)	Locus	Best <i>Arabidopsis thaliana</i> Blast hit GenBank ID	Best <i>A. thaliana</i> Blast hit gene	Type of Substitution	Within 5 cM of chromosomal break point?	Linkage to species differences QTL	
Collinear	1	18.8	HT051	AT5G58787	Zinc finger (C3HC4-type ring finger) family protein	Synonymous	N/A	Floral shape; stem diameter	
		22.8	HT083	AT2G16350	Expressed protein	Amino acid	N/A		
	10	29.6	HT250	AT5G44120	CRA1 (Cruciferina); nutrient reservoir	Amino acid	N/A	Floral shape, stem diameter; leaf K, Mn concentration	
		32.8	HT229	AT3G51895	SULTR3; 1 (sulfate transporter 1)	Amino acid	N/A	Seed weight; stem diameter; leaf K concentration	
		44.9	HT026	AT4G24220	VEP1 (vein patterning 1)	Amino acid	N/A		
	Rearranged	12	46.6	HT085	AT5G11420	Expressed protein	Synonymous	N/A	
			54.6	HT055	AT1G01230	Expressed protein	8-bp indel	N/A	
			55.0	HT080	AT2G39700	ATEXPA4 (Expansin A4)	Synonymous	NO	
		13	2.2	HT047	AT1G72280	Disulfide bond formation protein related	Synonymous	NO	
			28.6	HT037	AT2G32540	Cellulose synthase-related protein	Noncoding	YES	Flower biomass; branch no.
16		36.3	HT066	AT2G18960	ATPase1, plasma membrane type	Amino acid	YES	Seed weight	
		70.2	HT052	AT1G13260	RAV1 (related to ABI3/VP1 1)	Amino acid	NO		
		90.2	HT231	AT5G66190	Ferredoxin-NADP reductase (adrenodoxin reductase)	Amino acid	NO		
		17	52.0	HT128	AT1G14360	ATUTR3/UTR3 (UDP-galactose transporter 3)	Synonymous	YES	
			57.4	HT254	AT1G62710	BETA-VPE (vacuolar processing enzyme beta); cysteine-type endopeptidase	Synonymous	YES	Seed weight
	60.5	HT014	AT4G35080	High-affinity nickel-transport family protein	Noncoding	YES	Seed weight; leaf toughness		
	61.7	HT011	AT5G47780	GAUT4 (galacturonosyltransferase 4)	Noncoding	YES	Seed weight; leaf toughness		

NOTE.—All substitutions are fixed differences between *Helianthus annuus* and *Helianthus petiolaris*. Chromosomal positions and locus names are from Lai, Nakazato, et al. (2005).

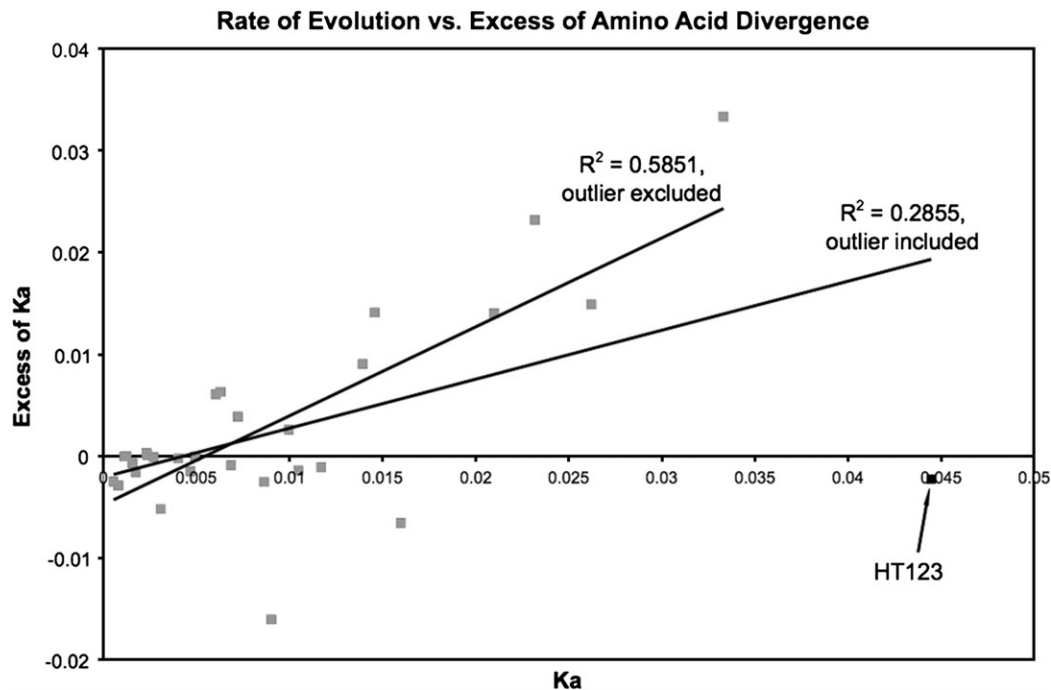


FIG. 5.—Comparison of rate of nonsynonymous evolution, K_a , with excess of nonsynonymous evolution compared with neutral expectations based on nonsynonymous and synonymous polymorphism. The single locus marked in black, HT123, is the outlier that is excluded from one of the regressions.

values are likely to even be underestimates (Charlesworth and Eyre-Walker 2008).

While there is substantial information available concerning morphological and ecological differentiation between *H. annuus* and *H. petiolaris* (Heiser 1947; Gross et al. 2004; Lexer et al. 2005; Karrenberg et al. 2006), we do not have adequate information to describe in detail the ecological or other adaptive significance of the specific amino acid differences documented here. We can, however, draw some tentative conclusions based on previous work. Fifteen of the 27 fixed differences between *H. annuus* and *H. petiolaris* are within 5 cM of a QTL for phenotypic differences between the two species (Rieseberg et al. 2003; Lexer et al. 2005), including seed shape (5 of the 17 loci, accounting for 11 of the 27 fixed differences), floral shape, tissue ion concentrations, and growth architecture (see table 5). In addition, 13 of the 19 fixed differences on rearranged chromosomes map near chromosomal break points, where QTLs for species differences and pollen sterility are preferentially located (Lai, Nakazato, et al. 2005). The extent to which these fixations are related to the QTL for species differences, proximity to chromosomal break points, or some other factor is not clear. But they represent interesting candidate genes for future study, especially those containing multiple fixed nonsynonymous differences. For example, a cellulose synthase on linkage group 16 has four fixed amino acid substitutions between the two species; and CRA1, a seed storage protein-encoding gene on linkage group 10, has two fixed nonsynonymous differences (one nonconservative) three codons apart. The only gene within an inversion showing fixed differences (one amino acid and two synonymous differences) is an ex-

pansin; expression changes in members of this gene family have been associated with both drought (Wu and Cosgrove 2000; Buchanan et al. 2005) and salt (Camacho-Cristobal et al. 2008; Kwon et al. 2008) stress.

Regarding the effects of chromosomal rearrangements differentiating *H. annuus* and *H. petiolaris*, we find little support for models that implicate a role for rearrangements in increasing differentiation via the suppression of recombination over large genomic distances. We also see little evidence for a particular role of inversions, which are likely to show stronger recombination suppression than large translocations. However, we do see evidence that recombination suppression very near chromosomal break points has a weak effect in increasing differentiation in those regions, as evidenced by marginally higher net sequence divergence and a marginally greater frequency of fixed sequence differences near break points. These results are similar to those reported by Yatabe et al. (2007), who found no difference in microsatellite differentiation on collinear versus rearranged chromosomes, but did find a weak effect of proximity to chromosomal break points on differentiation. Their interpretation, which is also consistent with our results, was that the unit of isolation between *H. annuus* and *H. petiolaris* is likely to be quite small and that genes or genomic regions contributing to isolation are readily decoupled from nearby regions, except in areas where recombination is suppressed most strongly, such as near chromosomal break points. Portions of rearranged chromosomes more distant from break points are likely to show more limited recombination suppression (Grant 1975; Schaeffer and Anderson 2005).

Other empirical tests confirm the predictions of recombination suppression models, at least in some systems

(reviewed in Hoffmann and Rieseberg 2008). They have largely been supported in *Drosophila* (Noor et al. 2001b; Schaeffer et al. 2003; Brown et al. 2004; Machado et al. 2007; Noor et al. 2007), tomatoes (Livingstone and Rieseberg 2004), *Anopheles* mosquitoes (Stump et al. 2007), and chromosome races of the *Sorex araneus* complex of shrews (Basset et al. 2006). However, although a number of studies comparing humans and chimps suggested increased sequence divergence, gene expression divergence, or evidence of selection in rearranged relative to collinear chromosomes (Navarro and Barton 2003b; Marques-Bonet et al. 2004), some researchers have questioned whether the sequence divergence and K_a/K_s differences documented by Navarro and Barton (2003b) can plausibly be explained by the theoretical model and hybridization scenario they suggest (Hey 2003; Lu et al. 2003); and reanalyses or studies done with larger data sets have brought these earlier results into doubt (Vallender and Lahn 2004; Zhang et al. 2004; Mikkelsen et al. 2005; Szamalek et al. 2007).

Interestingly, the bulk of excess amino acid divergence we see is found on collinear chromosomes rather than rearranged chromosomes; in fact, polymorphism and divergence ratios for rearranged chromosomes are not significantly different from neutral expectations. Lexer et al. (2005) also found that an excess of QTL for species differences were located on collinear chromosomes, a pattern generally consistent with our data. The reason for this pattern is not immediately clear. Although recombination suppression in rearrangements may limit the efficiency of positive selection due to Hill–Robertson interference (Hill and Robertson 1966), it is not expected to have an effect within each species if rearrangements are fixed, as is the case here (Burke et al. 2004), and recombination levels are not decreased. One highly speculative possibility is that the chromosomal variants that are currently fixed between the two species were still segregating within one or both species for some time following initial species divergence, and that during this period adaptive evolution was inhibited on chromosomes that were polymorphic for rearrangements but not on collinear chromosomes. Comparisons between these two species and *H. argophyllus* indicate that much of the adaptive divergence did occur soon after initial species divergence (see below), which is consistent with this hypothesis. Alternatively, this pattern may be a sampling artifact that will disappear when additional loci are sampled.

As far as we are aware, the degree to which segregating chromosomal rearrangements, especially inversions, might limit adaptive evolution due to recombination suppression has not been explored empirically or theoretically. Although recombination is generally very strongly reduced near chromosomal break points, gene flux can be fairly high near the center of inversions due to gene conversion and double crossing over (Andolfatto et al. 2001; Schaeffer and Anderson 2005). There is evidence of positive selection and rapid decay of linkage disequilibrium within inversions in *D. pseudoobscura*, suggesting that gene flux is enough to break up associations that are not maintained by epistatic selection (Schaeffer et al. 2003; Schaeffer and Anderson 2005). The degree of linkage disequilibrium within inver-

sions is a function of both the degree of recombination suppression and the age of the inversion. Unfortunately, it is difficult to put ages on the rearrangements between *H. annuus* and *H. petiolaris* because it is not clear what arrangements are derived or when the rearrangements occurred relative to species divergence.

A comparison of *H. annuus* versus *H. argophyllus* and *H. petiolaris* versus *H. argophyllus* MK tests is potentially informative with regard to the timing of adaptive divergence among these species. The sister species *H. annuus* and *H. argophyllus* show no evidence of adaptive divergence—polymorphism and divergence ratios indicate a 56% deficit of nonsynonymous fixations relative to neutral expectations, suggestive of a history of largely purifying selection following divergence for these two species. In contrast, *H. petiolaris* and *H. argophyllus* show patterns of adaptive divergence comparable to *H. annuus* and *H. petiolaris*; they show an excess of almost 11 nonsynonymous fixations relative to neutral expectations, indicating that close to 70% of their amino acid divergence was driven by positive selection. *Helianthus annuus* and *H. petiolaris* diverged roughly 700,000 years before *H. annuus* and *H. argophyllus* (Strasburg and Rieseberg 2008), and MK test results suggest either that the adaptive divergence between *H. annuus* and *H. petiolaris* largely occurred prior to *H. annuus*/*H. argophyllus* speciation or that it has occurred largely along the *H. petiolaris* lineage. Which scenario is more likely could be determined by polarizing each substitution to determine whether *H. annuus* or *H. petiolaris* has the derived allele. We have attempted to do so for coding substitutions using *Helianthus tuberosus* (Jerusalem artichoke) and *Taraxacum officinale* (dandelion) sequences from the Compositae Genome Project EST database (<http://www.compgenomics.ucdavis.edu/>). We were able to reliably polarize 14 of 19 coding substitutions; *H. annuus* had the derived allele in six cases, and *H. petiolaris* had the derived allele in eight cases. Although these sample sizes are far too small for anything more than a speculative answer, it appears to be the case that each species has undergone a comparable amount of divergent evolution, suggesting that most of it took place relatively soon after their initial divergence, prior to the divergence of *H. annuus* and *H. argophyllus*.

If this hypothesis is correct, it has implications for the mode of speciation for *H. annuus*/*H. petiolaris* and *H. annuus*/*H. argophyllus*. *Helianthus annuus* and *H. petiolaris* are currently broadly sympatric and have been hybridizing for most if not all of their existence (Strasburg and Rieseberg 2008). If these two species underwent parapatric or sympatric speciation, we would expect to find a signal of positive selection at loci directly or indirectly involved in reproductive isolation through divergent adaptation (Albertson et al. 2003; Blais et al. 2007; Via and West 2008). In contrast, although the early biogeographic history of *H. annuus* and *H. argophyllus* is not well understood, they have historically been allopatric and it is plausible that their speciation was allopatric as well. If that is the case, reproductive isolation may have been a byproduct of the accumulation of nonadaptive incompatibility alleles (Turelli et al. 2001), in which we would not necessarily expect to see a signature of divergent selection

during and soon after speciation. Although our data are consistent with these predictions and although it is tempting to speculate about how much of the divergence between *H. annuus* and *H. petiolaris* occurred during incipient speciation and was perhaps involved in the speciation process itself, our data and the biogeographic histories of these three species are not sufficiently detailed to address this question rigorously.

Conclusion

As genome-wide genetic diversity and divergence data become increasingly available, empirical support has increased for a genic, as opposed to genomic, view of the speciation and species divergence (Wu 2001; Lexer and Widmer 2008). Species that are strongly reproductively isolated at a number of genes and maintain morphological and ecological differences may still actively hybridize and exchange genetic material across much of their genomes (Strasburg and Rieseberg 2008). Adaptive divergence can proceed at the loci contributing to reproductive isolation and species differences even as much of the genome shows little or no genetic differentiation over millions of years. In species that are chromosomally differentiated, the role of rearrangements in promoting differentiation over large genomic regions may be limited depending on factors such as the nature of the rearrangements and amount of hybridization that occurs between species; but chromosomal break points, where recombination is suppressed most strongly, may permit both adaptive and nonadaptive divergence. Future work in this system will include analyses with more markers to allow more fine-scale characterization of genomic patterns of differentiation as well as more detailed population genetic and gene expression studies of genes near break points and other genes showing significant differentiation. This will help us to understand in more detail the degree of overall genomic differentiation between these species as well as the genetic basis of their isolation and phenotypic differentiation.

Supplementary Material

Supplementary tables S1–S3 and figure S1 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We would like to thank Eric Baack and Sophie Karrenberg for sharing their collections; Steve Wooley for writing a computer program to make filtered alignments and BayesFst input files; and Briana Gross, Ken Olsen, Nic Kooyers, Kate Waselkov, and Stuart McDaniel for comments on a previous version of the manuscript. This work was supported by a National Institutes of Health Ruth L. Kirschstein Postdoctoral Fellowship (5F32GM072409-02) to J.L.S. and grants from the National Science Foundation (DEB-0314654 and DBI0421630) and the National Institutes of Health (GM059065) to L.H.R.

Literature Cited

- Albertson RC, Streebman JT, Kocher TD. 2003. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proc Natl Acad Sci USA*. 100:5252–5257.
- Andolfatto P, Depaulis F, Navarro A. 2001. Inversion polymorphisms and nucleotide variability in *Drosophila*. *Genet Res*. 77:1–8.
- Ayala FJ, Coluzzi M. 2005. Chromosome speciation: humans, *Drosophila*, and mosquitoes. *Proc Natl Acad Sci USA*. 102:6535–6542.
- Barrett RDH, Rogers SM, Schluter D. 2008. Natural selection on a major armor gene in threespine stickleback. *Science*. 322: 255–257.
- Barrier M, Bustamante CD, Yu JY, Purugganan MD. 2003. Selection on rapidly evolving proteins in the *Arabidopsis* genome. *Genetics*. 163:723–733.
- Barton N, Bengtsson BO. 1986. The barrier to genetic exchange between hybridising populations. *Heredity*. 56:357–376.
- Basset P, Yannic G, Brunner H, Hausser J. 2006. Restricted gene flow at specific parts of the shrew genome in chromosomal hybrid zones. *Evolution*. 60:1718–1730.
- Beaumont MA, Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol*. 13:969–980.
- Bierne N, Eyre-Walker A. 2004. The genomic rate of adaptive amino acid substitution in *Drosophila*. *Mol Biol Evol*. 21: 1350–1360.
- Blais J, Rico C, van Oosterhout C, Cable J, Turner GF, Bernatch L. 2007. MHC adaptive divergence between closely related and sympatric African cichlids. *PLoS ONE*. 2(e734): 731–712.
- Brown KM, Burk LM, Henagan LM, Noor MAF. 2004. A test of the chromosomal rearrangement model of speciation in *Drosophila pseudoobscura*. *Evolution*. 58:1856–1860.
- Buchanan CD, Lim SY, Salzman RA, et al. (11 co-authors). 2005. *Sorghum bicolor*'s transcriptome response to dehydration, high salinity and ABA. *Plant Mol Biol*. 58: 699–720.
- Burke JM, Lai Z, Salmaso M, Nakazato T, Tang SX, Heesacker A, Knapp SJ, Rieseberg LH. 2004. Comparative mapping and rapid karyotypic evolution in the genus *Helianthus*. *Genetics*. 167:449–457.
- Bustamante CD, Fledel-Alon A, Williamson S, et al. (14 co-authors). 2005. Natural selection on protein-coding genes in the human genome. *Nature*. 437:1153–1157.
- Camacho-Cristobal JJ, Herrera-Rodriguez MB, Beato VM, Rexach J, Navarro-Gochicoa MT, Maldonado JM, Gonzalez-Fontes A. 2008. The expression of several cell wall-related genes in *Arabidopsis* roots is down-regulated under boron deficiency. *Environ Exp Bot*. 63:351–358.
- Charlesworth J, Eyre-Walker A. 2006. The rate of adaptive evolution in enteric bacteria. *Mol Biol Evol*. 23:1348–1356.
- Charlesworth J, Eyre-Walker A. 2008. The McDonald-Kreitman test and slightly deleterious mutations. *Mol Biol Evol*. 25:1007–1015.
- Coyne J, Orr HA. 2004. *Speciation*. Sunderland (MA): Sinauer Associates Inc. p. 545.
- Coyne JA, Orr HA. 1989. Patterns of speciation in *Drosophila*. *Evolution*. 43:362–381.
- Coyne JA, Aulard S, Berry A. 1991. Lack of underdominance in a naturally occurring pericentric inversion in *Drosophila melanogaster* and its implications for chromosome evolution. *Genetics*. 129:791–802.
- Coyne JA, Meyers W, Crittenden AP, Sniegowski P. 1993. The fertility effects of pericentric inversions in *Drosophila melanogaster*. *Genetics*. 134:487–496.

- Davisson MT, Akeson EC. 1993. Recombination suppression by heterozygous Robertsonian chromosomes in the mouse. *Genetics*. 133:649–667.
- Eyre-Walker A. 2002. Changing effective population size and the McDonald-Kreitman test. *Genetics*. 162:2017–2024.
- Eyre-Walker A. 2006. The genomic rate of adaptive evolution. *Trends Ecol Evol*. 21:569–575.
- Fay JC, Wyckoff GJ, Wu CI. 2001. Positive and negative selection on the human genome. *Genetics*. 158:1227–1234.
- Fay JC, Wyckoff GJ, Wu CI. 2002. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature*. 415:1024–1026.
- Gardner K, Buerkle A, Whitton J, Rieseberg LH. 2000. Inferring epistasis in wild sunflower hybrid zones. Pp. 264–279. In: Wolf JB, Brodie ED, Wade MJ, editors. *Epistasis and the evolutionary process*. Oxford: Oxford University Press.
- Grant V. 1975. *Genetics of flowering plants*. New York: Columbia University Press.
- Gross BL, Kane NC, Lexer C, Ludwig F, Rosenthal DM, Donovan LA, Rieseberg LH. 2004. Reconstructing the origin of *Helianthus deserticola*: survival and selection on the desert floor. *Am Nat*. 164:145–156.
- Gross BL, Schwarzbach AE, Rieseberg LH. 2003. Origin(s) of the diploid hybrid species *Helianthus deserticola* (Asteraceae). *Am J Bot*. 90:1708–1719.
- Hardy OJ, Vekemans X. 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes*. 2:618–620.
- Hedrick PW. 2005. A standardized genetic differentiation measure. *Evolution*. 59:1633–1638.
- Heiser CB. 1947. Hybridization between the sunflower species *Helianthus annuus* and *H. petiolaris*. *Evolution*. 1: 249–262.
- Hey J. 2003. Speciation and inversions: chimps and humans. *Bioessays*. 25:825–828.
- Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet Res*. 8:269–294.
- Hoffmann AA, Rieseberg LH. 2008. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annu Rev Ecol Evol Syst*. 39:21–42.
- Karrenberg S, Edelist C, Lexer C, Rieseberg L. 2006. Response to salinity in the homoploid hybrid species *Helianthus paradoxus* and its progenitors *H. annuus* and *H. petiolaris*. *New Phytol*. 170:615–629.
- King M. 1993. *Species evolution: the role of chromosome change*. Cambridge: Cambridge University Press.
- Kirkpatrick M, Barton N. 2006. Chromosome inversions, local adaptation and speciation. *Genetics*. 173:419–434.
- Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, Siepel A. 2008. Patterns of positive selection in six mammalian genomes. *PLoS Genet*. 4:e1000144.
- Kwon YR, Lee HJ, Kim KH, Hong SW, Lee H. 2008. Ectopic expression of Expansin3 or Expansin beta 1 causes enhanced hormone and salt stress sensitivity in *Arabidopsis*. *Biotechnol Lett*. 30:1281–1288.
- Lai Z, Livingstone K, Zou Y, Church SA, Knapp SJ, Andrews J, Rieseberg LH. 2005. Identification and mapping of SNPs from ESTs in sunflower. *Theor Appl Genet*. 111: 1532–1544.
- Lai Z, Nakazato T, Salmaso M, Burke JM, Tang SX, Knapp SJ, Rieseberg LH. 2005. Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. *Genetics*. 171:291–303.
- Lande R. 1985. The fixation of chromosomal rearrangements in a subdivided population with local extinction and colonization. *Heredity*. 54:323–332.
- Latta RG. 2004. Gene flow, adaptive population divergence and comparative population structure across loci. *New Phytol*. 161:51–58.
- Le Corre V, Kremer A. 2003. Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics*. 164:1205–1219.
- Levin DA. 2002. *The role of chromosomal change in plant evolution*. Oxford: Oxford University Press.
- Lewontin RC, Krakauer J. 1973. Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. *Genetics*. 74:175–195.
- Lexer C, Rosenthal DM, Raymond O, Donovan LA, Rieseberg LH. 2005. Genetics of species differences in the wild annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics*. 169:2225–2239.
- Lexer C, Widmer A. 2008. The genic view of plant speciation: recent progress and emerging questions. *Philos Trans R Soc B Biol Sci*. 363:3023–3036.
- Livingstone K, Rieseberg L. 2004. Chromosomal evolution and speciation: a recombination-based approach. *New Phytol*. 161:107–112.
- Lu J, Li WH, Wu CI. 2003. Comment on “Chromosomal speciation and molecular divergence—accelerated evolution in rearranged chromosomes”. *Science*. 302:988.
- Machado CA, Haselkorn TS, Noor MAF. 2007. Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*. 175:1289–1306.
- Marques-Bonet T, Caceres M, Bertranpetit J, Preuss TM, Thomas JW, Navarro A. 2004. Chromosomal rearrangements and the genomic distribution of gene expression divergence in humans and chimpanzees. *Trends Genet*. 20:524–529.
- McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the ADH locus in *Drosophila*. *Nature*. 351:652–654.
- Mikkelsen TS, Hillier LW, Eichler EE, et al. (67 co-authors). 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*. 437:69–87.
- Minder AM, Widmer A. 2008. A population genomic analysis of species boundaries: neutral processes, adaptive divergence and introgression between two hybridizing plant species. *Mol Ecol*. 17:1552–1563.
- Navarro A, Barton NH. 2003a. Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution*. 57:447–459.
- Navarro A, Barton NH. 2003b. Chromosomal speciation and molecular divergence - Accelerated evolution in rearranged chromosomes. *Science*. 300:321–324.
- Nei M. 1982. Evolution of human races at the gene level. In: Bonne-Tamir B, Goodman RM, editors. *Human Genetics. Part A: The Unfolding Genome*. New York: Alan R. Liss. p. 167–181.
- Nei M. 2005. Selectionism and neutralism in molecular evolution. *Mol Biol Evol*. 22:2318–2342.
- Nielsen R, Yang ZH. 2003. Estimating the distribution of selection coefficients from phylogenetic data with applications to mitochondrial and viral DNA. *Mol Biol Evol*. 20:1231–1239.
- Noor MAF, Garfield DA, Schaeffer SW, Machado CA. 2007. Divergence between the *Drosophila pseudoobscura* and *D. persimilis* genome sequences in relation to chromosomal inversions. *Genetics*. 177:1417–1428.
- Noor MAF, Grams KL, Bertucci LA, Almandarez Y, Reiland J, Smith KR. 2001a. The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* and *D. persimilis* via backcross hybrid males. *Evolution*. 55:512–521.
- Noor MAF, Grams KL, Bertucci LA, Reiland J. 2001b. Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci USA*. 98:12084–12088.

- Obbard DJ, Jiggins FM, Halligan DL, Little TJ. 2006. Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Curr Biol*. 16:580–585.
- Orr HA. 1987. Genetics of male and female sterility in hybrids of *Drosophila pseudoobscura* and *Drosophila persimilis*. *Genetics*. 116:555–563.
- Palopoli MF, Wu CI. 1994. Genetics of hybrid male sterility between *Drosophila* sibling species—a complex web of epistasis is revealed in interspecific studies. *Genetics*. 138:329–341.
- Rieseberg LH. 2001. Chromosomal rearrangements and speciation. *Trends Ecol Evol*. 16:351–358.
- Rieseberg LH, Church SA, Morjan CL. 2004. Integration of populations and differentiation of species. *New Phytol*. 161:59–69.
- Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzbach AE, Donovan LA, Lexer C. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*. 301:1211–1216.
- 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, Widmer A, Arntz AM, Burke JM. 2002. Directional selection is the primary cause of phenotypic diversification. *Proc Natl Acad Sci USA*. 99:12242–12245.
- Rosenthal DM, Rieseberg LH, Donovan LA. 2005. Re-creating ancient hybrid species' complex phenotypes from early-generation synthetic hybrids: three examples using wild sunflowers. *Am Nat*. 166:26–41.
- Rosenthal DM, Schwarzbach AE, Donovan LA, Raymond O, Rieseberg LH. 2002. Phenotypic differentiation between three ancient hybrid taxa and their parental species. *Int J Plant Sci*. 163:387–398.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*. 19:2496–2497.
- Sawyer SA, Kulathinal RJ, Bustamante CD, Hartl DL. 2003. Bayesian analysis suggests that most amino acid replacements in *Drosophila* are driven by positive selection. *J Mol Evol*. 57:S154–S164.
- Sawyer SA, Parsch J, Zhang Z, Hartl DL. 2007. Prevalence of positive selection among nearly neutral amino acid replacements in *Drosophila*. *Proc Natl Acad Sci USA*. 104:6504–6510.
- Schaeffer SW, Anderson WW. 2005. Mechanisms of genetic exchange within the chromosomal inversions of *Drosophila pseudoobscura*. *Genetics*. 171:1729–1739.
- Schaeffer SW, Goetting-Minesky MP, Kovacevic M, Peoples JR, Graybill JL, Miller JM, Kim K, Nelson JG, Anderson WW. 2003. Evolutionary genomics of inversions in *Drosophila pseudoobscura*: evidence for epistasis. *Proc Natl Acad Sci USA*. 100:8319–8324.
- Schwarzbach AE, Rieseberg LH. 2002. Likely multiple origins of a diploid hybrid sunflower species. *Mol Ecol*. 11:1703–1715.
- Scotti-Saintagne C, Mariette S, Porth I, Goicoechea PG, Barreneche T, Bodenès K, Burg K, Kremer A. 2004. Genome scanning for interspecific differentiation between two closely related oak species *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. *Genetics*. 168:1615–1626.
- Sites JWJ, Moritz C. 1987. Chromosomal evolution and speciation revisited. *Syst Zool*. 36:153–174.
- Sokal RR, Oden NL. 1978. Spatial autocorrelation in biology. 1. Methodology. *Biol J Linn Soc*. 10:199–228.
- Steiner CC, Weber JN, Hoekstra HE. 2007. Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biol*. 5:1880–1889.
- Storz JF. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol*. 14:671–688.
- Strasburg JL, Rieseberg LH. In review. The geography and dynamics of interspecific gene flow—high levels of recent gene flow between historically allopatric annual sunflower species *Helianthus annuus* and *H. argophyllus*. *Genetics*.
- 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.
- Stump AD, Pombi M, Goeddel L, Ribeiro JMC, Wilder JA, Torre AD, Besansky NJ. 2007. Genetic exchange in 2La inversion heterokaryotypes of *Anopheles gambiae*. *Insect Mol Biol*. 16:703–709.
- Sturtevant AH. 1938. On the origin of interspecific sterility. *Q Rev Biol*. 13:333–335.
- Szamalek JM, Cooper DN, Hoegel J, Hameister H, Kehrer-Sawatzki H. 2007. Chromosomal speciation of humans and chimpanzees revisited: studies of DNA divergence within inverted regions. *Cytogenet Genome Res*. 116:53–60.
- Turelli M, Barton NH, Coyne JA. 2001. Theory and speciation. *Trends Ecol Evol*. 16:330–343.
- Vallender EJ, Lahn BT. 2004. Effects of chromosomal rearrangements on human-chimpanzee molecular evolution. *Genomics*. 84:757–761.
- Vamathevan JJ, Hasan S, Emes RD, et al. (13 co-authors). 2008. The role of positive selection in determining the molecular cause of species differences in disease. *BMC Evol Biol*. 8.
- Via S, West J. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol Ecol*. 17:4334–4345.
- Walsh JB. 1982. Rate of accumulation of reproductive isolation by chromosome rearrangements. *Am Nat*. 120:510–532.
- White MJD. 1978. Modes of speciation. San Francisco: W.H. Freeman.
- Wright S. 1940. Breeding structure of populations in relation to speciation. *Am Nat*. 74:232–248.
- Wright S. 1941. On the probability of fixation of reciprocal translocations. *Am Nat*. 75:513–522.
- Wu CI. 2001. The genic view of the process of speciation. *J Evol Biol*. 14:851–865.
- Wu YJ, Cosgrove DJ. 2000. Adaptation of roots to low water potentials by changes in cell wall extensibility and cell wall proteins. *J Exp Bot*. 51:1543–1553.
- 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.
- Zhang JZ, Wang XX, Podlaha O. 2004. Testing the chromosomal speciation hypothesis for humans and chimpanzees. *Genome Res*. 14:845–851.

Jody Hey, Associate Editor

Accepted March 5, 2009