Published in final edited form as:

Am Nat. 2007 June ; 169(6): 725–737.

Reconstructing the History of Selection during Homoploid Hybrid Speciation

Sophie Karrenberg^{1,2,*}, Christian Lexer³, and Loren H. Rieseberg^{2,4}

¹ Institute of Integrative Biology, Eidgenössische Technische Hochschule Zürich, 8092 Zürich, Switzerland

² Department of Biology, Indiana University, Bloomington, Indiana 47405 ³ Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, United Kingdom ⁴ Department of Botany, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

Abstract

This study aims to identify selection pressures during the historical process of homoploid hybrid speciation in three *Helianthus* (sunflower) hybrid species. If selection against intrinsic genetic incompatibilities (fertility selection) or for important morphological/ecological traits (phenotypic selection) were important in hybrid speciation, we would expect this selection to have influenced the parentage of molecular markers or chromosomal segments in the hybrid species' genomes. To infer past selection, we compared the parentage of molecular markers in high-density maps of the three hybrid species with predicted marker parentage from an analysis of fertility selection in artificial hybrids and from the directions of quantitative trait loci effects with respect to the phenotypes of the hybrid species. Multiple logistic regression models were consistent with both fertility and phenotypic selection in all three species. To further investigate traits under selection, we used a permutation test to determine whether marker parentage predicted from groups of functionally related traits differed from neutral expectations. Our results suggest that trait groups associated with ecological divergence were under selection during hybrid speciation. This study presents a new method to test for selection and supports earlier claims that fertility selection and phenotypic selection on ecologically relevant traits have operated simultaneously during sunflower hybrid speciation.

Keywords

ecological selection; fertility selection; homoploid hybrid speciation; *Helianthus*; permutation test; quantitative trait loci (QTLs)

A major goal of evolutionary biology is to assess the relative importance of evolutionary factors responsible for present-day phenotypes and species differences. Studies of contemporary populations can identify traits currently under selection and connect the direction and intensity of selection to environmental variables (Lande and Arnold 1986). These kinds of studies are necessarily short-term, however, and their utility for reconstructing the history of selection is questionable because of temporal variation in the environment and in genotype-environment interactions (Agrawal et al. 2001; Hoekstra et al. 2001; Grant and Grant 2002). Comparative studies of phenotypic differences between species or higher taxa, on the other hand, typically fail to distinguish between phenotypic changes caused by selection and those resulting from neutral processes. Likewise, studies on the genetic footprints of selection can reveal genomic targets of recent selective sweeps but do not necessarily connect these targets to phenotypes or to the selective forces involved (reviewed in Nielsen 2005; but see Orr 1998).

* Corresponding author; e-mail: sophie.karrenberg@env.ethz.ch.

Homoploid hybrid species offer a unique opportunity for estimating the relative importance of drift and selection in evolution as well as for identifying traits or trait groups that are the direct targets of selection. During homoploid hybrid speciation, hybridization between chromosomally or genetically divergent species leads to stabilized and fertile derivatives of the same ploidal level with partial or complete reproductive isolation from the parental species (Grant 1981; Abbott 1992; Rieseberg 1997). Because this type of speciation appears to occur rapidly (Ungerer et al. 1998), it is sometimes possible to experimentally replicate the early stages of the hybrid speciation process. Comparison of experimental hybrid lineages with those that originated naturally has provided a powerful means for reconstructing the speciation process (Rieseberg et al. 1996a; Lexer et al. 2003; Gross et al. 2004; Ludwig et al. 2004; Gross and Rieseberg 2005; Hegarty and Hiscock 2005; Rosenthal et al. 2005).

Three of the best-investigated examples of homoploid hybrid species are annual sunflowers. Molecular evidence indicates that independent hybridization events between the annual sunflowers *Helianthus annuus* and *Helianthus petiolaris*, both widespread throughout North America, gave rise to three hybrid species, *Helianthus anomalus*, *Helianthus deserticola*, and *Helianthus paradoxus* (Rieseberg 1991), all with restricted geographical distributions in the southwestern United States. This process appears to have occurred recently, roughly within the past 200,000 years (Schwarzbach and Rieseberg 2002; Welch and Rieseberg 2002; Gross et al. 2003). Furthermore, in the *Helianthus* system, detailed molecular marker maps (Rieseberg et al. 2003), information on marker transmission in artificial crossing lines (Rieseberg et al. 1996b; Rieseberg 2000), and map positions for numerous quantitative trait loci (QTLs) are available (Rieseberg et al. 2003).

The parental species, *H. annuus* and *H. petiolaris*, have divergent karyotypes (Rieseberg et al. 1995), and early-generation hybrids are semisterile, with low pollen viability and seed set (Rieseberg 2000). This sterility appears to be a consequence of both karyotypic differences and gene incompatibilities (Lai et al. 2005). Artificial crossing studies revealed that fertility is regained in advanced-generation hybrid lines when a more stable genome composition is reached (Rieseberg et al. 1995; Rieseberg 2000; Lai et al. 2005). These advanced-generation artificial hybrids can be crossed easily with each other and with all three ancient hybrid species, but crosses with the parental species have reduced fertility (Rieseberg 2000; Lai et al. 2005). Molecular marker mapping indicates that the artificial crossing lines have genome compositions similar to each other and to the hybrid species *H. anomalus* (Rieseberg et al. 1996b), suggesting that selection against intrinsic genetic incompatibilities to increase hybrid fertility has played an important role in shaping hybrid genomic composition (fertility selection). In the two other hybrid species, *H. deserticola* and *H. paradoxus*, these comparisons have not been made to date, but we hypothesize that such fertility selection has also influenced marker parentage in these species.

Theoretical models indicate that intrinsic postzygotic barriers alone are insufficient for the origin of homoploid hybrid species and that ecological selection is also required for their successful establishment (McCarthy et al. 1995; Buerkle et al. 2000). Ecological differences between sunflower hybrid species and their parental species are large and well documented. While the parental species occur in relatively mesic habitats—*H. annuus* is found on clay-based soils and *H. petiolaris* on sandy soils—the three hybrid species are able to grow in extreme environments. *Helianthus anomalus* is found on desert sand dunes, *H. deserticola* also occurs in the desert but on more stabilized deposits, and *H. paradoxus* grows in salt seeps (Heiser et al. 1969; Rogers et al. 1982). Recent field experiments have shown that ecological selection likely maintains species differences between hybrid and parental species, suggesting that ecological selection was also involved in hybrid speciation. The hybrid species *H. anomalus* (Ludwig et al. 2004) and *H. paradoxus* (Lexer et al. 2003) survived at a higher rate in their natural habitat than did their parental species. For the third hybrid species, *H.*

deserticola, higher fitness in the natural habitat in terms of seed production was likely but not experimentally verified (Gross et al. 2004). Furthermore, selection gradients or differentials for ecologically important traits in artificial, early-generation hybrid populations between the parental species, grown in the habitats of *H. paradoxus* and *H. deserticola*, were mostly in the predicted direction for generating phenotypes of the hybrid species (Lexer et al. 2003; Gross et al. 2004). In the habitat of *H. anomalus*, selection gradients differed between artificial hybrid lines, and selection toward the hybrid species' phenotype was only partly supported (Ludwig et al. 2004). Although the above studies provide evidence of current selection, their implications for historical selection are less clear because we do not know how closely the experiments mimicked ancestral habitats and genotypes.

In this study, we ask which selective forces shaped the genomic composition of the three homoploid sunflower hybrid species. Our rationale is that if selection against intrinsic genetic incompatibilities (fertility selection) or for important morphological/ecological traits (phenotypic selection) were important in hybrid speciation, we would expect this selection to have influenced the parentage of molecular markers or chromosomal segments in the hybrid species' genomes. Fertility selection is inferred by comparing marker transmission in advanced-generation artificial hybrid lines (Rieseberg et al. 1996b; Rieseberg 2000) with marker parentage in high-density maps of the hybrid species. Phenotypic selection, on the other hand, is investigated by comparing actual marker parentage with predictions based on the directions of QTL effects with respect to the phenotypes of the hybrid species. At first glance, this latter approach appears similar to Orr's (1998) sign test that determines whether a significant number of QTLs associated with particular traits exhibit allelic effects in a consistent direction, suggesting adaptive divergence between the parents in that trait. However, unlike Orr, we are concerned with whether QTL directions in relation to phenotypes predict marker parentage in the hybrid species correctly, independent of the actual directions of QTL effects. Thus, rather than testing for adaptive divergence in parental phenotypes, as Orr's test does, we test for selection during hybrid speciation.

In a first step, we compare marker parentage predictions related to fertility selection and to phenotypic selection. We hypothesize that the two types of selection acted simultaneously during hybrid speciation and that both were required for the establishment of the hybrid species. If this were true, both predictors of marker parentage should be retained in a joint statistical model. If, however, fertility selection were more important and phenotypic selection acted only within the constraints imposed by fertility selection, the effect of phenotypic predictions, even if a significant predictor alone, should be eliminated in a joint model. The reverse would be expected if phenotypic selection were more important than fertility selection.

In the second part of this study, we look at phenotypic parentage predictions in more detail and ask which kinds of traits may have influenced marker parentage. Here, we developed a permutation test to investigate whether QTL parentage predictions for groups of functionally related traits are consistent with a random process. If a random process is rejected, we infer phenotypic selection on the corresponding traits to be involved in determining marker parentage near the QTL.

Material and Methods

Marker Parentage Predictions to Infer Fertility Selection and Phenotypic Selection

Marker parentage predictions were made using intrinsic genetic incompatibilities (fertility selection) and QTL directions with phenotype comparison (phenotypic selection). To infer fertility selection, data from an earlier study using three crossing lines of the parental species *Helianthus annuus* and *Helianthus petiolaris* (Rieseberg et al. 1996b; Rieseberg 2000) were used. For these lines, marker parentage was assessed after five generations when fertility was

mostly recovered and was used as a predictor of molecular marker parentage in the hybrid species' genomes (Rieseberg et al. 1996b). To infer phenotypic selection on the same marker loci, we made parentage predictions following Rieseberg et al. (2003). Predictions were first made for individual QTLs related to traits with transgressive or parental-like phenotypes in the hybrid species. For these, the hybrid species were predicted to have the QTL alleles (derived from either *H. annuus* or *H. petiolaris*) that would be needed to generate the hybrid species' phenotype from an ancestral F₁ generation with intermediate trait values. QTLs were investigated in a backcross family (BC₂) toward *H. petiolaris* (Rieseberg et al. 2003) with QTL directions coded such that the heterozygote state (with the *H. annuus* allele) leads to a change in the given direction. Thus, where the hybrid species had increased trait values, positive QTLs were assigned the *H. annuus* allele and negative QTLs the *H. petiolaris* allele and vice versa (cf. fig. 1). Because QTLs in both directions were detected for most traits (Rieseberg et al. 2003), the alleles that would move a trait value in a given direction usually were a mixture of parental alleles. These single-locus predictions were then assigned to the 1-logarithm-of-odds (LOD) interval of each QTL, and where more than one 1-LOD interval overlapped for a locus to be predicted, a majority rule based on summed likelihood ratios was used (Rieseberg et al. 2003). In this analysis, the number of predictions used equals the number of loci for which fertility predictions were available and that fell into the 1-LOD interval of at least one QTL (table 1). Note that these integrated phenotypic parentage predictions are influenced by all QTLs near a particular locus and do not allow conclusions about selection on individual traits. Congruence between actual and predicted marker parentage was estimated by comparison with detailed genetic maps of each of the three hybrid species that included species-specific parental markers (table 1; Rieseberg et al. 2003).

We first evaluated whether each predictor alone had more correct cases than would be expected given the percentages (expected probabilities) of the two species in actual and predicted marker parentage. The expected probability of correct cases was calculated from the frequencies (p) of the two species (*H. annuus* [H_a] and *H. petiolaris* [H_p]) in the markers (m) of known parentage and in the parentage predictions (pp) as:

$$p(m=pp)=p(m=H_a) \times p(pp=H_a) \\ + p(m=H_p) \times p(pp=H_p).$$

These values were compared with the confidence intervals for the observed percentages of correct predictions.

Logistic regression analyses (Venables and Ripley 2002) were then used to evaluate the joint effect of fertility predictions and integrated phenotypic predictions on the actual marker parentage in the three hybrid species. Because physical linkage could render individual markers in close proximity nonindependent, differences in apparent parental linkage block size in the genomes of the three hybrid species were investigated by tabulating map distances and whether or not marker parentage was identical for all possible two-locus pairs within linkage groups. Although non-independence in the predictors would not violate regression analysis assumptions, we used the same method to describe the resolution of fertility and phenotypic parentage predictions to aid the interpretation of regression results. Logistic regression analyses were carried out using the entire data set and, more conservatively, using 1,000 randomly generated subsets of the data restricted to marker distances above 3 and above 5 cM. Marker distances, total map length, and markers in the data subsets are given in table 1.

Assumptions of Marker Parentage Analysis

The following main assumptions are made in marker parentage predictions: (1) the gene or genes underlying QTLs map closely to predicted positions and thus are transmitted together with flanking markers, (2) QTL effects are additive, and (3) marker parentage was mostly

determined during rather than after hybrid speciation (genome stabilization). Assumption 1 depends both on the breadth of QTLs and on the scale of recombination in the hybrid species' genomes. For broad QTLs, this assumption is likely to be violated, so the test of selection presented here is conservative (Rieseberg et al. 2003). It is not possible to test assumption 2 at the present time because QTLs were determined in a backcross design where one cannot discriminate between additive and dominance effects (Rieseberg et al. 2003). Assumption 3 is probable for fertility selection because marker parentage stabilizes quickly in artificial hybrid lines that have been subject to selection for increased fertility (Rieseberg et al. 1996b; Rieseberg 2000). The timing of ecological selection and its effect on genome stabilization are less clear. However, theoretical models of hybrid speciation require strong ecological selection early in the speciation process.

Phenotypic Selection on Trait Groups

Traits that were likely to experience the same selection pressures were placed into 10 groups based on functional relatedness (fig. 2) because often there were not enough QTLs for individual traits to test for selection. Within each trait group, QTLs for strongly correlated traits (e.g., leaf length and leaf area) were reduced to one entry if their predicted positions were within 5 cM in the same marker interval or if they had the same nearest marker. Unlike the first analysis, we used one prediction per QTL (fig. 1) and took into account linkage among QTLs of the same trait group but disregarded all other QTLs; that is, the number of marker parentage predictions per trait group equaled the number of QTLs in that group.

QTLs with larger effects might be under stronger selection pressures (Orr 1998). For this reason, we tested whether QTL magnitude, expressed as percent variation explained (PVE), influenced congruence in this data set. In a logistic regression model (Venables and Ripley 2002), QTL magnitude had no effect on congruence in any of the three hybrid species (analysis of deviance, $P > \chi^2 = .366$ for *Helianthus anomalus* [number of QTLs analyzed is $N = 69$], $P > \chi^2 = .832$ for *Helianthus deserticola* [$N = 104$], and $P > \chi^2 = .322$ for *Helianthus paradoxus* [$N = 103$]). Furthermore, PVE values were indistinguishable across the trait groups, as indicated by standard ANOVA procedures, making it unlikely that QTL magnitude had an influence on permutation test results ($F = 1.096$, $df = 5, 63$, $P = .371$ for *H. anomalus*; $F = 1.684$, $df = 7, 96$, $P = .122$ for *H. deserticola*; $F = 1.415$, $df = 6, 96$, $P = .216$ for *H. paradoxus*). Thus, QTL magnitude was not included in the permutation analysis.

To infer whether the observed numbers of QTL loci with correctly predicted parentage in each group were significantly higher than would be expected under the null hypothesis of selective neutrality, we constructed a permutation test (Venables and Ripley 2002) modified to take into account linked QTLs within trait groups. To simulate the genotypes of the three hybrid species at QTL positions under the null hypothesis, we assigned the two parental species to QTL positions with a probability of 0.5 because in all three hybrid species, the parentage of markers did not differ significantly from 50% for each parental species in high-density linkage maps used to infer actual parentage (*H. anomalus* [427 species-specific markers]: 53.16% *H. annuus*, $P < \chi^2 = .208$; *H. deserticola* [290 markers]: 52.4% *H. annuus*, $P < \chi^2 = .445$; *H. paradoxus* [325 markers]: 51.6% *H. annuus*, $P < \chi^2 = .579$; one-sample proportions test with continuity correction; Dalgaard 2002). For QTLs positioned on the same linkage group, we set the probability of assigning the same parental species to both to 1 minus the recombination fraction. The recombination fraction was calculated from QTL map positions using the Kosambi mapping function (Lynch and Walsh 1998), which was also employed to generate the QTL map (Rieseberg et al. 2003). Note that this single-generation recombination fraction was used as a proxy for linkage here and does not reflect possible recombination events during hybrid speciation over multiple generations. One thousand simulated genotypes were generated for each trait group, except for groups with fewer than 10 loci. For groups with nine and eight

loci, 500 and 250 genotypes were simulated, respectively, to avoid exceeding the number of unique permutations. Fewer than 250 permutations yielded unreliable results, and therefore we considered only trait groups with at least eight QTLs. For all simulated multilocus genotypes, the number of QTLs at which the parentage was predicted correctly was evaluated. We obtained *P* values of the observed data as the fraction of simulated genotypes with more correct predictions than found in the observed data. All statistical analyses were carried out in R, version 2.1.0 (Dalgaard 2002; Venables and Ripley 2002).

Results

Fertility Selection versus Phenotypic Selection

For *Helianthus anomalus* and *Helianthus paradoxus* (but not *Helianthus deserticola*), fertility data predicted that a slightly higher proportion of genomic regions should be derived from *Helianthus annuus* than from *Helianthus petiolaris*. However, confidence intervals for the *H. annuus* percentage still included 50% in all three species (fig. 3A). Integrated phenotypic predictions of marker parentage, on the other hand, tended to favor the other parent, *H. petiolaris* (fig. 3A). In *H. anomalus*, only 33% of the loci had phenotypic predictions of *H. annuus*, a significant difference from 50%. In the two other species, confidence intervals for the percentage of *H. annuus* markers from phenotypic predictions included 50%. For comparison, the actual marker parentage of predicted loci is also shown in figure 3A; in all species, confidence intervals included 50%.

In all three hybrid species, marker parentage was predicted correctly in the majority of cases by both fertility and phenotypic data, with phenotypic predictions slightly outperforming fertility predictions (fig. 3B). The percentage of correct predictions was significantly different from the expected percentages based on random allocation of parental segments to hybrids (cf. confidence intervals and line for expected percentages in fig. 3B), but many correct predictions were made by both predictors (fig. 3B). Nonetheless, both predictors were retained in a joint logistic regression model when using the full data sets for all three hybrid species, as indicated by the significant *P* values for both predictors (table 2). In general, the results of these unconstrained analyses using all markers available were similar to the more conservative analysis with data subsets restricted to marker distances above 3 cM (table 2), with more than 95% of the 100 replicate data sets analyzed having significant effects of both fertility and phenotypic predictions (table 2). For the even more restrictive analysis of loci with pairwise distances over 5 cM, the effect of fertility predictions was not significant in 12.5% of the randomly generated data sets for *H. anomalus*. In the two other species, both predictors had significant effects in more than 95% of the data sets.

Despite the presence of several large parental chromosomal blocks in the genome of *H. anomalus* (Rieseberg et al. 1995; Ungerer et al. 1998), average block sizes were smaller than the marker distances (fig. 4A). A similar pattern was observed for *H. paradoxus* (fig. 4C). In these species, the percentage of instances where two markers on the same linkage group had identical parentage did not change with marker distance and was close to 50%. In *H. deserticola*, however, marker distances smaller than 3 cM had a higher percentage of identical parentage in two-locus comparisons than did larger distances, indicating that there might be interspecific linkage disequilibrium at this scale (fig. 4B). For fertility and phenotypic predictions, the average size of parental blocks was much greater than that for the ancient hybrid genomes, with percentages of identical predictions for marker distances under 3 cM and in some cases under 5 cM reaching over 90% (fig. 4). Note that the difference in block size is a predicted consequence of the number of generations of recombination in the natural versus artificial hybrids (see “Discussion”).

Phenotypic Selection on Trait Groups

We observed large differences in the congruence of predicted and actual parentage in hybrid species for the 10 trait groups (table 3; fig. 1). For the physiological trait groups describing ion uptake (*H. deserticola* and *H. paradoxus*) and photosynthesis (*H. anomalus*), the null hypothesis of a random process could be rejected (table 3), allowing us to infer a role of natural selection. Plant size appeared to be under selection in all three hybrid species, as did flower morphology in the two species tested (*H. deserticola* and *H. paradoxus*). We found evidence for selection on leaf morphology only in *H. anomalus* and for selection on ligule morphology only in *H. deserticola*. There also was no evidence of selection for the trait groups describing phyllary morphology (all three species), plant shape (*H. deserticola*), phenology (*H. deserticola*), and seed size (*H. anomalus*).

Discussion

Joint Action of Fertility Selection and Phenotypic Selection

Our results support the hypothesis that selection for both increased hybrid fertility and important morphological/ecological traits was necessary for the formation of sunflower hybrid species, as suggested by Grant (1981) and predicted by simulation studies (McCarthy et al. 1995; Buerkle et al. 2000). Both effects were significant predictors of marker parentage in three *Helianthus* hybrid species, even in conservative multiple logistic regression analyses of markers with pairwise distances over 3 and 5 cM, except for fertility predictions in *Helianthus anomalus* in the 5-cM-distance data set (table 2). This might be due to differences in the data sets and not to the reduction in power in the restricted data set because *H. anomalus* had the largest number of markers both in the full data set and in the restricted data sets. Thus, despite the small sizes of parental chromosomal blocks in the genomes of the hybrid species (fig. 4), we can still detect the signature of selection.

Phenotypic parentage predictions generally had larger percentages of correct predictions and larger multiple logistic regression coefficients (fig. 3; table 2), suggesting that phenotypic selection may have been a stronger force than fertility selection during hybrid speciation. The difference between the two predictors does not appear to be related to differences in the resolution of the two types of parentage predictions, although parental chromosome blocks in artificial crosses are greater in size than apparent blocks in the genomes of the hybrid species (fig. 4). Fertility and phenotypic predictions were obtained from synthetic hybrids after just five and two generations of recombination, respectively (Rieseberg et al. 1996a; Rieseberg 2000), and parental blocks are expected to be larger in this material than after the estimated 60 generations apparently required for hybrid species formation (Ungerer et al. 1998). The stronger effect of phenotypic selection as compared with fertility selection is surprising, given apparent constraints in the parental chromosomal combinations required to generate fertile recombinant genotypes (Rieseberg 2000). This observation implies that genes under strong ecological selection may affect fixation probabilities even in regions that cause reduced fertility in early hybrid generations.

Targets of Selection

Reconstructing the history of selection on trait groups using QTLs presents a number of significant challenges, including widespread linkage and/or pleiotropy for QTLs underlying functionally related traits, imprecision of estimated QTL positions, and limited resolution of hybrid species' genetic marker maps (see assumptions in "Material and Methods"). Differences in numbers of QTLs per trait group are of particular concern to this study because our ability to detect selection is reduced for trait groups with low numbers of QTLs. Thus, trait groups with few QTLs, many of which predict marker parentage correctly (e.g., plant size and plant

shape in *Helianthus paradoxus*), may have been under selection during hybrid speciation, but such selection may be undetectable in this study.

Despite these difficulties, we were able to detect selection on a number of trait groups that were in good agreement with ecological expectations. In contrast, one taxonomically important trait group, phyllary morphology, which differs widely between species (fig. 2), consistently failed to show evidence of selection (table 3; fig. 2). Trait groups under selection (table 3) contained numerous transgressive phenotypes (fig. 2), consistent with an important role of transgressive segregation in homoploid hybrid speciation (Rieseberg et al. 1999; Seehausen 2004; Schwarz et al. 2005). The converse, however, that transgressive traits are always strongly selected, does not hold true. For example, phyllary morphology is highly transgressive, but a significant effect of selection could not be detected. Hence, transgressive phenotypes in hybrid sunflowers may not result exclusively from directional selection on those traits but might arise through other processes such as pleiotropy, hitchhiking, or drift.

Selection on photosynthesis traits, leaf morphology, and plant size appears to have influenced marker parentage in the hybrid species *H. anomalus* (table 3). This species grows on unstable cold desert sand dunes that differ from the habitats of the parental species in terms of low substrate stability, elevated temperatures, and water and nutrient limitations (Danin 1996). Selection for a higher photosynthetic rate and associated lower intercellular CO₂ concentrations as found in *H. anomalus* (fig. 2) is to be expected for nonsucculent annuals in a desert habitat where illumination is not limiting (Gibson 1996). An earlier study by Schwarzbach et al. (2001) did not support higher photosynthetic activity in *H. anomalus* as compared with its parental species, but that study was conducted under low illumination. Data employed in this study were gathered under high-light conditions (Rosenthal et al. 2002; Rieseberg et al. 2003), implying that the photosynthetic rate of *H. anomalus* might surpass that of its parental species only under well-lit conditions, such as those found in the field. Leaf morphology QTLs found to be under selection in this study cause smaller and narrower leaves in *H. anomalus* (fig. 2). Desert plants often have small and narrow leaves, and these leaf traits are known to facilitate transpirational cooling (Givnish 1979).

Marker parentage in *Helianthus deserticola* was best predicted by QTLs for ion uptake, ligule morphology, flower morphology, and plant size (marginally significant at $\alpha < .07$). *Helianthus deserticola* occurs in a desert environment, but unlike *H. anomalus*, this species is not found on dunes but on more stabilized deposits (Heiser et al. 1969; Rogers et al. 1982). In this habitat, *H. deserticola* faces potentially toxic concentrations of boron, sodium, and other elements as a result of excessive evapotranspiration, in addition to extremes of temperature, drought, and nutrient limitation (Blamey et al. 1997; Smith et al. 1997). When compared with its parental species (fig. 2), *H. deserticola* exhibits reduced uptake of boron, possibly a mechanism to avoid boron toxicity. Leaf morphology QTLs, although predicting the expected phenotypic shift toward smaller and narrower leaves, are not found to be a significant predictor of marker parentage ($P = .096$). In contrast, ligule morphology traits appeared to be under selection in *H. deserticola*, and QTLs for ligule morphology that are present in *H. deserticola* reduce ligule size and number. One explanation is a water conservation strategy through reduced transpiration in smaller flowers (Galen 1999). Reduced plant size appears to be under selection as well (fig. 2). In this early-flowering species, it seems reasonable to associate smaller size with a rapid reproductive cycle, a common drought avoidance strategy of desert plants (Guterman 2002).

For *H. paradoxus*, which is found in salt seeps, our study supports ecological selection on ion uptake traits, flower morphology, and plant size (marginally significant at $\alpha = .07$). QTL alleles present in *H. paradoxus* lead to increased concentrations of sodium in the leaves, while potassium content is reduced (fig. 2), constituting a typical salt stress response (Volkmar et al.

1998; Wyn Jones and Gorham 2002; Lexer et al. 2003; Karrenberg et al. 2006). Plant size, in turn, might be reduced because of the cost associated with salt tolerance mechanisms (Volkmar et al. 1998).

Broader Implications

As alluded to in the introduction to this article, homoploid hybrid speciation offers a unique opportunity to identify and order the evolutionary factors that contribute to diversification. The key advantage associated with hybrid speciation is that it occurs quickly, and it is therefore possible to replicate the early stages of the speciation process experimentally. In this study, we extend this one step further by using results from experimental studies to predict the marker parentage of ancient hybrid species. Our findings corroborate suggestions from simulation studies that fertility and ecological selection operate together during hybrid speciation, and we further identify targets of phenotypic selection that are consistent with ecological expectations.

Two main questions arise from these conclusions. First, do our results have broader implications for other homoploid hybrid species? We suspect that they may be broadly predictive. In particular, the finding that ecological selection plays a more important role than fertility selection in hybrid speciation might be general because most other hybrid taxa have far fewer fertility problems than those associated with early-generation sunflower hybrids (e.g., Seehausen 2004; Howarth and Baum 2005; James and Abbott 2005; Schwarz et al. 2005; Ma et al. 2006). Also, all well-documented examples of homoploid hybrid speciation exhibit evidence of ecological divergence, but only a subset of these are also isolated by intrinsic postzygotic barriers (Gross and Rieseberg 2005).

A second question is whether the approach employed here (predictions of marker parentage) can be realistically applied to other examples of homoploid hybrid speciation. There are logistical and biological difficulties. Rieseberg and colleagues have spent the past 18 years (off and on) identifying and mapping species-specific molecular markers (Rieseberg et al. 1993, 1995, 2003; Ungerer et al. 1998), and even now, with > 400 such markers on the *H. anomalus* map, the resolution is barely adequate. However, new high-throughput methods for the discovery and mapping of nucleic acid polymorphisms have greatly reduced the time and expense required for such a study (Borevitz et al. 2003; Stanssens et al. 2004; Altshuler et al. 2005). Thus, we think the basic logistical hurdles have been greatly reduced since the first attempt to estimate the genomic composition of ancient hybrids (Rieseberg et al. 1993).

A major biological issue is whether the sizes of parental chromosomal segments that become stabilized in hybrid lineages will be of sufficient size for testing with predictions from QTL or fertility selection experiments. Even in sunflower, block sizes seem to be only barely large enough to allow such a method to work; fine-mapping and sequencing studies are under way to more precisely estimate the scale of recombination. In other instances of hybrid speciation, where there are few or no chromosomal rearrangements, genome stabilization may take place over a longer time period, making it more difficult to identify parental chromosomal segments. Nonetheless, we have been delighted to see many new examples of homoploid hybrid species identified in both plants and animals (e.g., Howarth and Baum 2005; James and Abbott 2005; Salazar et al. 2005; Schwarz et al. 2005; Meyer et al. 2006), and we look forward to results of experimental and comparative genomic studies (e.g., Albertson and Kocher 2005) that test some of the generalizations/predictions made here.

Acknowledgements

We thank S. P. Otto for suggestions on the analysis. We are grateful to E. J. Baack, S. P. Otto, and K. D. Whitney for helpful comments on this manuscript. This research was supported by postdoctoral fellowship grant 1777/1-1 of the Deutsche Forschungsgemeinschaft to S.K. and by National Institute of Health grant R01 G59065 to L.H.R.

Literature Cited

- Abbott RJ. Hybridization and the evolution of new plant taxa. *Trends in Ecology & Evolution* 1992;7:401–405.
- Agrawal AF, Brodie ED, Rieseberg LH. Possible consequences of genes of major effect: transient changes in the G-matrix. *Genetica* 2001;112:33–43. [PubMed: 11838774]
- Albertson R, Kocher T. Genetic architecture sets limits on transgressive segregation in hybrid cichlid fishes. *Evolution* 2005;59:686–690. [PubMed: 15856710]
- Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P. A haplotype map of the human genome. *Nature* 2005;437:1299–1320. [PubMed: 16255080]
- Blamey, FPC.; Zollinger, RK.; Schneitner, AA. Sunflower production and culture. In: Schneitner, AA., editor. *Sunflower technology and production*. American Society of Agronomy; Madison, WI: 1997. p. 595-670.
- Borevitz JO, Liang D, Plouffe D, Chang HS, Zhu T, Weigel D, Berry CC, Winzeler E, Chory J. Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Research* 2003;13:513–523. [PubMed: 12618383]
- Buerkle CA, Morris RJ, Asmussen MA, Rieseberg LH. The likelihood of homoploid hybrid speciation. *Heredity* 2000;84:441–451. [PubMed: 10849068]
- Dalgaard, P. *Introductory statistics with R*. Springer; New York: 2002.
- Danin, A. *Plants of desert dunes: adaptations of desert organisms*. Springer; Berlin: 1996.
- Galen C. Why do flowers vary? the functional ecology of variation in flower size and form within natural populations. *BioScience* 1999;49:631–640.
- Gibson, AC. *Structure-function relations of warm-desert plants*. Springer; Berlin: 1996.
- Givnish, TJ. On the adaptive significance of leaf form. In: Solbrig, OT.; Jain, S.; Johnson, GB.; Raven, PH., editors. *Topics in plant population biology*. Columbia University Press; New York: 1979. p. 373-407.
- Grant PR, Grant BR. Unpredictable evolution in a 30-year study of Darwin's finches. *Science* 2002;296:707–711. [PubMed: 11976447]
- Grant, V. *Plant speciation*. Columbia University Press; New York: 1981.
- Gross BL, Rieseberg LH. The ecological genetics of homoploid hybrid speciation. *Journal of Heredity* 2005;96:241–252. [PubMed: 15618301]
- Gross BL, Schwarzbach AE, Rieseberg LH. Origin(s) of the diploid hybrid species *Helianthus deserticola* (Asteraceae). *American Journal of Botany* 2003;90:1708–1719.
- Gross BL, Kane NC, Lexer C, Ludwig F, Rosenthal DM, Donovan LA, Rieseberg LH. Reconstructing the origin of *Helianthus deserticola*: survival and selection on the desert floor. *American Naturalist* 2004;164:145–156.
- Gutterman, Y. *Survival strategies of annual desert plants*. Springer; Berlin: 2002.
- Hegarty MJ, Hiscock SJ. Hybrid speciation in plants: new insights from molecular studies. *New Phytologist* 2005;165:411–423. [PubMed: 15720652]
- Heiser CB, Smith DM, Clevenger S, Martin WC. The North American sunflowers (*Helianthus*). *Memoirs of the Torrey Botanical Club* 1969;22:1–218.
- Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, Hoang A, Hill CE, Beerli P, Kingsolver JG. Strength and tempo of directional selection in the wild. *Proceedings of the National Academy of Sciences of the USA* 2001;98:9157–9160. [PubMed: 11470913]
- Howarth DG, Baum DA. Genealogical evidence of homoploid hybrid speciation in an adaptive radiation of *Scaevola* (Goodeniaceae) in the Hawaiian Islands. *Evolution* 2005;59:948–961. [PubMed: 16136795]
- James JK, Abbott RJ. Recent, allopatric, homoploid hybrid speciation: the origin of *Senecio squalidus* (Asteraceae) in the British Isles from a hybrid zone on Mount Etna, Sicily. *Evolution* 2005;59:2533–2547. [PubMed: 16526502]
- Karrenberg S, Edelist C, Lexer C, Rieseberg L. Response to salinity in the homoploid hybrid species *Helianthus paradoxus* and its progenitors *H. annuus* and *H. petiolaris*. *New Phytologist* 2006;170:615–629. [PubMed: 16626481]

- Lai Z, Nakazato T, Salmaso M, Burke JM, Tang SX, Knapp SJ, Rieseberg LH. Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. *Genetics* 2005;171:291–303. [PubMed: 16183908]
- Lande R, Arnold SJ. The measurement of selection on correlated characters. *Evolution* 1986;37:1210–1226.
- Lexer C, Welch ME, Raymond O, Rieseberg LH. The origin of ecological divergence in *Helianthus paradoxus* (Asteraceae): selection on transgressive characters in a novel hybrid habitat. *Evolution* 2003;57:1989–2000. [PubMed: 14575321]
- Ludwig F, Rosenthal DM, Johnston JA, Kane N, Gross BL, Lexer C, Dudley SA, Rieseberg LH, Donovan LA. Selection on leaf ecophysiological traits in a desert hybrid *Helianthus* species and early-generation hybrids. *Evolution* 2004;58:2682–2692. [PubMed: 15696747]
- Lynch, M.; Walsh, B. *Genetics and the analysis of quantitative traits*. Sinauer; Sunderland, MA: 1998.
- Ma XF, Szmidt AE, Wang XR. Genetic structure and evolutionary history of a diploid hybrid pine *Pinus densata* inferred from the nucleotide variation at seven gene loci. *Molecular Biology and Evolution* 2006;23:807–816. [PubMed: 16446291]
- McCarthy EM, Asmussen MA, Anderson WW. A theoretical assessment of recombinational speciation. *Heredity* 1995;74:502–509.
- Meyer A, Salzburger W, Scharl M. Hybrid origin of a swordtail species (Teleostei: *Xiphophorus clemenciae*) driven by sexual selection. *Molecular Ecology* 2006;15:721–730. [PubMed: 16499697]
- Nielsen R. Molecular signatures of natural selection. *Annual Review of Genetics* 2005;39:197–218.
- Orr HA. Testing natural selection vs. genetic drift in phenotype evolution using quantitative trait locus data. *Genetics* 1998;149:2099–2104. [PubMed: 9691061]
- Rieseberg LH. Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. *American Journal of Botany* 1991;78:1218–1237.
- Rieseberg LH. Hybrid origins of plant species. *Annual Review of Ecology and Systematics* 1997;28:359–389.
- Rieseberg LH. Crossing relationships among ancient and experimental sunflower hybrid lineages. *Evolution* 2000;54:859–865. [PubMed: 10937259]
- Rieseberg LH, Choi HC, Chan R, Spore C. Genomic map of a diploid hybrid species. *Heredity* 1993;70:285–293.
- Rieseberg LH, Vanfossen C, Desrochers AM. Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature* 1995;375:313–316.
- Rieseberg LH, Arias DM, Ungerer MC, Linder CR, Sinervo B. The effects of mating design on introgression between chromosomally divergent sunflower species. *Theoretical and Applied Genetics* 93 1996a:633–644.
- Rieseberg LH, Sinervo B, Linder CR, Ungerer MC, Arias DM. Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. *Science* 1996b;272:741–745. [PubMed: 8662570]
- Rieseberg LH, Archer MA, Wayne RK. Transgressive segregation, adaptation and speciation. *Heredity* 1999;83:363–372. [PubMed: 10583537]
- Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Murphy JL, Schwarzbach AE, Donovan LA, Lexer C. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 2003;301:1211–1216. [PubMed: 12907807]
- Rogers, CE.; Thompson, TE.; Seiler, GE. *Sunflower species of the United States*. National Sunflower Association; Fargo, ND: 1982.
- Rosenthal DM, Schwarzbach AM, Donovan LA, Raymond O, Rieseberg LH. Phenotypic differentiation between three ancient and hybrid taxa and their parental species. *International Journal of Plant Sciences* 2002;163:387–398.
- Rosenthal DM, Rieseberg LH, Donovan LA. Recreating ancient hybrid species' complex phenotypes from early-generation synthetic hybrids: three examples using wild sunflowers. *American Naturalist* 2005;166:26–41.
- Salazar CA, Jiggins CD, Arias CF, Tobler A, Bermingham E, Linares M. Hybrid incompatibility is consistent with a hybrid origin of *Heliconius heurippa* Hewitson from its close relatives, *Heliconius*

- cydno* Doubleday and *Heliconius melpomene* Linnaeus. *Journal of Evolutionary Biology* 2005;18:247–256. [PubMed: 15715831]
- Schwarz D, Matta BM, Shakir-Botteri NL, McPheron BA. Host shift to an invasive plant triggers rapid animal hybrid speciation. *Nature* 2005;436:546–549. [PubMed: 16049486]
- Schwarzbach AE, Rieseberg LH. Likely multiple origins of a diploid hybrid sunflower species. *Molecular Ecology* 2002;11:1703–1715. [PubMed: 12207721]
- Schwarzbach AE, Donovan LA, Rieseberg LH. Transgressive character expression in a hybrid sunflower species. *American Journal of Botany* 2001;88:270–277. [PubMed: 11222249]
- Seehausen O. Hybridization and adaptive radiation. *Trends in Ecology & Evolution* 2004;19:198–207. [PubMed: 16701254]
- Smith, SD.; Monson, RK.; Anderson, JE. *Physiological ecology of desert plants*. Springer; Berlin: 1997.
- Stanssens P, Zabeau M, Meersseman G, Remes G, Gansemans Y, Storm N, Hartmer R, et al. High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. *Genome Research* 2004;14:126–133. [PubMed: 14707174]
- Ungerer MC, Baird SJE, Pan J, Rieseberg LH. Rapid hybrid speciation in wild sunflowers. *Proceedings of the National Academy of Sciences of the USA* 1998;95:11757–11762. [PubMed: 9751738]
- Venables, WN.; Ripley, BD. *Modern applied statistics with S-Plus*. Springer; New York: 2002.
- Volkmar KM, Hu Y, Steppuhn H. Physiological responses of plants to salinity: a review. *Canadian Journal of Plant Science* 1998;78:19–27.
- Welch ME, Rieseberg LH. Patterns of genetic variation suggest a single, ancient origin for the diploid hybrid species *Helianthus paradoxus*. *Evolution* 2002;56:2126–2137. [PubMed: 12487344]
- Wyn Jones, G.; Gorham, J. Intra- and inter-cellular compartmentation of ions. In: Laüchli, A.; Lüttge, U., editors. *Salinity: environment–plants–molecules*. Kluwer; Dordrecht: 2002. p. 159–180.

Figure 1.

Illustration of phenotypic marker parentage predictions for individual qualitative trait loci (QTLs) in hybrid species using phenotype comparisons and QTL directions. Note that the integrated phenotypic predictions used in this study were also done on this principle, with the difference that a prediction for a given locus was obtained using all QTLs for which the 1-logarithm-of-odds intervals (*bars on QTL diamonds*) overlapped with the locus to be predicted.

Figure 2.

Phenotype comparisons between the annual sunflower species *Helianthus annuus* (A) and *Helianthus petiolaris* (P) and their hybrid derivatives *Helianthus anomalus* (ANO), *Helianthus deserticola* (DES), and *Helianthus paradoxus* (PAR) in 10 trait groups; data from Rieseberg et al. (2003). Previously used trait abbreviations are given in parentheses. Parental species with lower and higher values for each trait are identified on the left and right sides of the table, respectively. Filled triangles indicate transgressive phenotypes in the direction of the arrow (significantly different from both parental species), open triangles show parental-like phenotypes in the direction of arrow (significantly different from other parent), and circles are used for intermediate hybrid phenotypes. Evidence for selection on trait groups as indicated by a permutation test is denoted by boxes: *solid line*, significant at $\alpha = .05$; *dashed line*, marginally significant ($\alpha < .07$); *shaded boxes*, trait group not used because of insufficient number of quantitative trait loci.

Figure 3.

Observed marker parentage in three *Helianthus* hybrid species (A) and percentage of correct marker parentage predictions according to fertility predictions or integrated phenotypic predictions (B). Values are given with 95% confidence intervals for comparison with dotted reference lines at the expected values for a random allocation of parental segments to hybrid species genomes. The solid lines in B refer to the percentage of loci for which parentage was predicted correctly by both predictors.

Figure 4.

Percentage of molecular markers with identical parentage in the genome or identical marker parentage predictions according to fertility or phenotypic predictions in all possible two-locus comparisons across classes of map distances in the three *Helianthus* hybrid species *Helianthus anomalus* (A), *Helianthus deserticola* (B), and *Helianthus paradoxus* (C). Numbers of comparisons for each class are given above columns in each marker parentage graph (left column).

Table 1
 Characteristics of parentage prediction data sets and of high-density molecular marker maps in three *Helianthus* hybrid species

	<i>Helianthus anomalus</i>	<i>Helianthus deserticola</i>	<i>Helianthus paradoxus</i>
Parentage prediction data sets:			
Full data set:			
Marker number	146	87	96
Total map length (cM)	843.0	454.9	584.3
Median marker distance	3.80 (.9–24.9)	3.45 (1.7–19.2)	3.60 (.8–81.5)
Subset >3 cM:			
Median and range of marker numbers	110 (106–113)	69 (68–70)	77 (75–78)
Subset >5 cM:			
Median and range of marker numbers	86 (77–92)	60 (57–61)	52 (47–56)
High-density marker maps:			
All markers:			
Marker number	1,019	672	771
Species-specific markers:			
Marker number	427	290	325
Total map length (cM)	1,887.2	1,208.4	1,413.8
Median and range of marker distances	3.80 (0–21.5)	3.40 (0–19.2)	3.60 (0–24.8)

Table 2
Multiple logistic regression results for marker parentage predictions for three *Helianthus* hybrid species

	Helianthus anomalus		Helianthus deserticola		Helianthus paradoxus	
	Estimates	<i>P</i> value/fraction (<i>P</i> > α)	Estimates	<i>P</i> value/fraction (<i>P</i> > α)	Estimates	<i>P</i> value/fraction (<i>P</i> > α)
Full data set:						
Fertility	1.36 (.56–2.23)	.001	1.90 (.76–3.19)	.002	1.00 (.13–1.90)	.027
Phenotype	2.44 (1.59–3.38)	<.0001	2.45 (1.37–3.65)	<.0001	1.52 (.64–2.47)	.001
Subsets:						
>3 cM:						
Fertility	2.07 (1.46–2.76)	.011	1.69 (1.49–1.76)	.000	2.14 (1.96–2.35)	.000
Phenotype	2.51 (1.92–3.26)	.000	2.37 (2.01–2.49)	.000	2.46 (2.31–2.56)	.000
>5 cM:						
Fertility	2.00 (1.18–3.07)	.125	1.89 (1.38–2.34)	.047	1.99 (1.56–2.56)	.018
Phenotype	2.44 (1.59–3.45)	.004	2.12 (1.66–2.57)	.000	2.37 (1.89–3.15)	.000

Note: For the full data set, estimates for regression coefficients are given with 95% confidence intervals; for the subsets (1,000 replicates, randomly selected), the median of the regression coefficients with 0.025 and 0.975 quantiles is given together with the fraction of analyses where *P* values were not significant at $\alpha = .05$.

Table 3
Permutation analysis of parentage predictions in the homoploid hybrid species *Helianthus anomalus*, *Helianthus deserticola*, and *Helianthus paradoxus* using QTL directions and phenotypic differences

Trait group	H. anomalus			H. deserticola			H. paradoxus		
	LG	QTL	P	LG	QTL	P	LG	QTL	P
Ion uptake	3	4 (4)		13	21 (18)	.002**	14	29 (20)	.015*
Photosynthesis	7	9 (8)	.032 ^a *	6	6 (5)		6	6 (4)	
Leaf morphology	9	12 (11)	.011*	8	12 (9)	.096	10	15 (7)	.704
Ligule morphology	8	10 (8)	.093	11	16 (15)	.001**	10	13 (9)	.144
Phyllary morphology	11	13 (8)	.306	11	13 (9)	.153	9	11 (8)	.124
Plant shape	2	2 (1)		6	8 (4)	.640 ^b	7	9 (7)	.088 ^a
Plant size	11	16 (13)	.021*	11	18 (14)	.051 ⁺	8	10 (8)	.064 ⁺
Phenology	5	7 (6)		5	8 (6)	.187 ^b	5	7 (5)	
Flower morphology	5	5 (4)		8	8 (7)	.034, ^b *	11	16 (13)	.010*
Seed mass	9	9 (6)	.274 ^a		na			na	

Note: LG = number of linkage groups containing quantitative trait loci (QTLs); QTL = number of QTLs in the trait group, with the number of congruent QTLs in parentheses. P values were determined by 1,000 permutations, unless stated otherwise. na = not available.

^a500 permutations.

^b250 permutations.

* $P < .05$.

** $P < .01$.

⁺ $P < .07$.