Role of Gene Interactions in Hybrid Speciation: Evidence from Ancient and Experimental Hybrids

Loren H. Rieseberg,* Barry Sinervo, C. Randall Linder, Mark C. Ungerer, Dulce M. Arias

The origin of a new diploid species by means of hybridization requires the successful merger of differentiated parental species’ genomes. To study this process, the genomic composition of three experimentally synthesized hybrid lineages was compared with that of an ancient hybrid species. The genomic composition of the synthesized and ancient hybrids was concordant (r² = 0.68, P < 0.0001), indicating that selection to a large extent governs hybrid species formation. Further, nonrandom rates of introgression and significant associations among unlinked markers in each of the three synthesized hybrid lineages imply that interactions between coadapted parental species’ genes constrain the genomic composition of hybrid species.

Reduced hybrid fertility or viability appears to result from unfavorable interactions between parental species’ genomes (1). As a result, species’ genomes are considered to be coadapted and thereby resistant to the introgression of alien genes (1). However, the successful origin of new diploid species by means of hybridization raises the possibility that interactions between parental species’ genes are not universally unfavorable (2). Little is known about the strength and fitness consequences of gene interactions in hybrids or their role in hybrid speciation (3). Here we compare

REFERENCES AND NOTES

19. The gene encoding mouse BMPR-IB was subjected to site-directed oligo mutagenesis (Gonbetch) to change Lys531 to Arg. This was cloned into a ClaI/Nco I site of the vector and then into the replication-competent avian retroviral vector PCAS6 (18). Primary chicken embryo fibroblasts were transfected with the use of lipofectin (Gibco-BRL), culture supernatant was collected on days 6 to 10, and the virus was concentrated by centrifugation, resuspended in a small volume, aliquoted, and stored at −70°C. The concentration of virus was estimated on the basis of reverse transcriptase assays (~1 × 10^6 units of reverse transcriptase per microliter of concentrated stock), which correlates to ~1 × 10^11 plaque-forming units per milliliter of concentrated stock as titrated on our indicator virus. Four virus preparations were used for these studies with similar results.
32. We are grateful to K. Miyazono for the gene encoding mouse BMPR-IB, J. Massagué and F. Ventura for receptor construct design and for performing the in vitro kinase assay, and K. Manova of the Memorial Sloan-Kettering Cancer Center Molecular Cytolgy Facility. Probes were kindly provided by B. Robert (MSKCC) and M. Gomez, P. Brickell (BM72 and BM14), B. Houston (BM77), and J.-C. Iziusía-Belmonte (HOX21). We thank our lab members, B. Hogan and J. Massagué, for critical reading of the manuscript and S. Norrby for advice on the scale-to-feather transformation. This work was supported by Memorial Sloan-Kettering Cancer Center Support Grant and American Cancer Society Junior Faculty Research Award.
2. 2 January 1996; accepted 29 March 1996.
the genomic composition of three experimentally synthesized hybrid lineages (Helianthus annuus × H. petiolaris) with that of an ancient hybrid species (H. anomala) (4) to determine whether the genetic factors governing the genomic composition of artificial hybrids are similar to those involved in hybrid speciation. We then analyze parental marker segregation in the experimentally synthesized hybrid lineages to detect interactions among genes that affect hybrid fitness and, indirectly, hybrid genomic composition.

The three sunflower species studied here are self-incompatible annuals native to North America (5). The parents, H. annuus and H. petiolaris, are common in the western United States (5). Although differing in karyotype, morphology, and habitat requirements, they frequently grow together, and hybrid swarms are common (6). First generation hybrids are semisterile, but full fertility can be regained in later generation hybrids (6). The stabilized hybrid, H. anomala, is restricted to xeric habitats in northern Arizona and southern Utah, well within the range of its parental species (7). Although morphologically distinct (7), it combines parental ribosomal DNA-repeat units, allozymes, chloroplast DNA haplotypes, and random amplified polymorphic DNA (RAPD) markers (4, 8). The near absence of ribosomal DNA and chloroplast DNA sequence divergence between H. anomala and its parents suggests a recent origin of the hybrid species, probably within the last 170,000 years (8). Reproductive isolation between H. anomala and its parental species has been facilitated by rapid karyotypic evolution (4).

To study the genetic processes accompanying hybrid speciation, we synthesized three hybrid lineages between H. annuus and H. petiolaris: lineage I, P-F₁-BC₁-BC₂-F₂-F₃; lineage II, P-F₁-F₂-BC₁-BC₂-F₃; and lineage III, P-F₁-F₂-F₃-BC₁-BC₂ (9). For each hybrid lineage, total DNA was isolated from 56 or 58 progeny from the final generation (10), and these 170 DNAs were surveyed for 197 H. petiolaris RAPD markers of known genomic location (4, 10). On completion of the marker survey, H. petiolaris markers present in the hybrid lineage were plotted onto the genomic map of H. annuus, thereby generating a graphical genotype (11) for each hybrid lineage (Fig. 1).

Concordance in genomic composition among the three synthesized hybrid lineages was tested by comparing distributions of introgressed H. petiolaris markers (Fig. 1). Patterns of introgression were strongly correlated among the three synthesized lineages (12); (P < 0.0001 for all combinations), indicating that to a large extent selection rather than chance governs the genomic composition of hybrids between H. annuus and H. petiolaris.

To determine whether the genomic composition of the synthesized hybrids was concordant with that of the ancient hybrid species, H. anomala, we compared the distribution of species-specific markers in the H. anomala genome (4) to the distribution of parental genomic regions in the synthesized hybrids (Fig. 2). The genomic composition of the ancient hybrid species was recognizable similar to that of experimental hybrid lineages (13); (rₛ = 0.68, N = 140, P < 0.0001), suggesting that genomic structure and composition of hybrid species may be essentially fixed within a few generations after the initial hybridization event and remain relatively static thereafter (14).

Several genetic factors appear to govern hybrid genomic composition. First, the parental species are chromosomally divergent, and certain rearranged linkage blocks (for example, linkage group T) appear to resist recombination in both the experimental and ancient hybrids (Figs. 1 and 2). Moreover, because all backcrosses in the synthesized hybrids were in the direction of H. annuus, chromosomal rearrangements to a large extent account for the low frequency of H. petiolaris markers.
in the rearranged portion of the genome (10) (Fig. 1 and Table 1). However, chromosomal rearrangements alone cannot explain the concordance between the genomic composition of the synthesized and ancient hybrids because strong correlations are observed within both the collinear [(13); $r_e = 0.66, P < 0.0001$] and rearranged [(13); $r_e = 0.69, P < 0.0001$] portions of the genome (Fig. 1).

Gene interactions provide a more compelling explanation for concordance in genomic composition, particularly in collinear genomic regions. Evidence for the importance of gene interactions comes from two sources: marker frequencies and associations. Most H. petiolaris markers (71 to 85%) introgressed at significantly lower than expected frequencies in the synthesized hybrids (Table 1), suggestive of unfavorable interactions between loci tightly linked to these markers and H. annuus genes. By contrast, favorable interspecific gene interactions are implied by the significantly higher than expected rates of introgression observed for 5 to 6% of H. petiolaris markers. These markers represent 10 of the 17 linkage groups and show significant linkage to adjacent markers. In light of these genome-wide patterns of introgression, alternative explanations for higher than expected rates of introgression—such as meiotic drive or gene conversion—are untenable. Concordance of marker frequency across the three synthesized hybrid lineages further suggests that these interactions to a large extent remain constant, regardless of hybrid genealogy (15).

Analyses of associations among segregating parental markers allow detection of specific interactions among chromosome segments that affect hybrid fitness rather than the general interactions inferred from the frequency data. The rationale for this approach is that interacting genes may have nonadditive (epistatic) effects on hybrid fitness (16, 17). This fitness epistasis should be detectable as two-way, three-way, and higher order associations among unlinked markers that are themselves tightly linked physically with loci that influence hybrid fitness and, indirectly, hybrid genomic composition. To test for these interactions, we analyzed the locus × progeny array from each hybrid lineage for significant negative and positive associations between every two- and three-way combination of unlinked loci (18). Analyses of higher order interactions were precluded by the number of combinatorial possibilities.

For hybrid lineages I, II, and III, we observed 10, 15, and 10 significant ($\alpha \leq 0.0001$; Fig. 3, A to C) two-way associations, respectively, whereas 0 were expected by chance, given the total number of pairwise comparisons. In the more powerful three-way analysis, we observed 21, 29, and 15 three-way associations ($\alpha \leq 0.0001$; Fig. 3, D to F), whereas 0, 1, and 2, respectively, were expected by chance. Significantly, even at the high stringency levels used here, many of the same two- and three-way associations were observed in multiple hybrid lineages (Fig. 3). Because the hybrid lineages were generated independently (9), selection rather than random population bottlenecks must account for these associations. Moreover, markers with epistatic interactions were more likely to be found in all three lineages than markers lacking epistasis (19), suggesting that these interactions influence hybrid genomic composition.

In all, 35 H. petiolaris loci representing 15 linkage groups exhibited significant two- and three-way associations, generating the complex epistatic webs shown in Fig. 3. Furthermore, because much of the H. petiolaris genome was eliminated from the hybrid lineages in early generations not analyzed here, the epistatic interactions implied by these marker associations represent only the H. petiolaris coadapted gene complexes that have neutral or favorable interactions with the H. annuus genomic background. Analyses of negatively selected markers might reveal additional evidence for epistatic interactions such as those reported among male sterility genes in Drosophila (17).

A major question that remains concerns how these gene interactions affect the fitness of sunflower hybrids. Lowered
fitness of hybrids between *H. annuus* and *H. petiolaris* appears to result from reduced fertility; *F₁*′s exhibit pollen viabilities of less than 10% and seed set less than 1% (6). In this experiment, uniformly high fertility (>90% pollen viability) was recovered by the fifth generation in all three synthesized hybrid lineages. Thus, many of the gene interactions reported may affect hybrid fertility. Interspecific competition among viable pollen grains also serves as a reproductive barrier between *H. annuus* and *H. petiolaris* (20), so some of the epistatic gene combinations observed here possibly affect pollen tube growth rates. Further study will focus on identifying fitness traits physically linked to these markers.

Interactions between divergent species’ genomes are often viewed as uniformly disharmonious (1), resulting in hybrid inviability or sterility. The data presented here suggest that although the majority of interspecific gene interactions are indeed unfavorable or neutral, a small percentage of alien genes do appear to interact favorably in hybrids. These favorable gene interactions might provide the raw material for adaptive evolution in hybrid taxa. Although we detected gene interactions in interspecific crosses, it seems plausible that with much larger progenies, this approach might also be useful for estimating genome-wide fitness epistasis within species or populations.

**Fig. 3.** Epistatic interactions among introgressed *H. petiolaris* markers in synthesized hybrid lineages. (A through C) Scatter plots of observed and expected *p*’s, a measure of association (18), for two-way epistatic interactions in hybrid lineages I (A), II (B), and III (C). *N* is the number of two-way associations: (A) *P* < 0.0001; (C) *P* = 0.01; (E) *P* = 0.01; (C) *P* = not significant. Symbols (A, O, C) above the cluster of nonsignificant interactions (·) are positive associations (21); symbols below are negative associations (22). *(D through F)* ’Webs’ of significant three-way epistatic interactions (*P* ≤ 0.0001) for hybrid lineages I (D), II (E), and III (F) as indicated by triangles connecting three unlinked markers. Marker designations indicate primer number and linkage block (compare with Fig. 1). Any three markers that interacted epistatically are connected as a triangle. (●) Markers involved in three-way epistatic interactions in all three hybrid lineages; (□) markers involved in three-way epistatic interactions in two hybrid lineages; (●) markers involved in three-way epistatic interactions in one hybrid lineage. Lengths connecting markers are arbitrary. Positive associations are indicated by solid lines and negative associations by dashed lines. For example, the three unlinked markers 241, 149, and 149 (both at the base of all three webs) (D through F), and marker 376 (to interior) were positively associated in all three hybrid lineages, as indicated by the connecting triangle.

**Table 1.** Observed and expected proportions of markers introgressed into 0%, 1 to 25%, 26 to 50%, and >50% of individuals.

<table>
<thead>
<tr>
<th>Percentage of individuals in which markers introgressed</th>
<th>Entire genome (197 markers)</th>
<th>Collinear portion (56 markers)</th>
<th>Rearranged portion (193 markers)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Expected</td>
<td>Expected</td>
</tr>
<tr>
<td>Hybrid lineage I</td>
<td>0.85</td>
<td>0.0009</td>
<td>0.0006</td>
</tr>
<tr>
<td></td>
<td>1 to 25</td>
<td>0.9662</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>26 to 50</td>
<td>0.06</td>
<td>0.0329</td>
</tr>
<tr>
<td></td>
<td>&gt;50</td>
<td>0.03</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hybrid lineage II</td>
<td>0.76</td>
<td>0.0001</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>1 to 25</td>
<td>0.08</td>
<td>0.0868</td>
</tr>
<tr>
<td></td>
<td>26 to 50</td>
<td>0.12</td>
<td>0.1131</td>
</tr>
<tr>
<td></td>
<td>&gt;50</td>
<td>0.05</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hybrid lineage III</td>
<td>0.71</td>
<td>&lt;0.0001</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>1 to 25</td>
<td>0.17</td>
<td>0.5080</td>
</tr>
<tr>
<td></td>
<td>26 to 50</td>
<td>0.05</td>
<td>0.4920</td>
</tr>
<tr>
<td></td>
<td>&gt;50</td>
<td>0.07</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

*Expected values were calculated with the mean SD calculated from the SDs of 100 simulations of unrestricted marker introgression for the entire genome, as well as within the collinear and rearranged portions of the genome (19). Hybrid lineage I: 0.156 ± 0.0504 (mean ± SD); hybrid lineage II: 0.1878 ± 0.0515; hybrid lineage III: 0.2487 ± 0.0578.*

**REFERENCES AND NOTES**

9. The initial interspecific cross was *H. annuus* (oms/HA98; female) × *H. petiolaris* subsp. petiolaris (PET-PET-1741; male). Backcrosses were in the direction of *H. annuus*, the maternal parent, because more of the *H. annuus* genome is derived from *H. annuus* than from *H. petiolaris*. As a result of self-incompatibility, self-pollination rather than selfing was used for the F₁ generation. At least 20 plants were used for each generation. Crosses were performed by applying pooled pollen from all plants to a given generation to stigma of the same individual. All achenes produced from each generation were pooled, and 30 achenes were arbitrarily chosen as founders of the next generation. With the exception of the parents and F₁, each hybrid lineage was generated independently.
12. Spearman’s rank correlation coefficient was calculated for each pairwise combination of the three hybrid lineages. Patterns of introgressed and nonintrogressed markers were compared, such that if *H. petiolaris* markers present in at least one of the progeny from a given lineage were scored as introgressed. Lineages I × II, *r* = 0.89, *P* = 0.0001; lineages I × III, *r* = 0.89, *P* < 0.0001; lineages II × III, *r* = 0.83, *P* < 0.0001.
13. Concordance between the genomic composition of the experimental and ancient hybrids was calculated with the use of Spearman’s rank correlation coefficient for the entire, collinear, and rearranged portions of the genome. For the experimental hybrids, linkage blocksintrogressed in one or more individuals in any
of the synthesized hybrids were considered H. peli-
colaris genomic regions (1), whereas linkage blocks
lacking H. peliolaris markers were designated H. an-
nuus regions (0), thereby generating the composite
distribution of parental genomic regions shown in Fig.
2. Parental markers in H. anuuus either matched
(0.0 or 1.1) or did not match (0.1 or 1.0) the corre-
sponding genomic region in the experimental hybrids.

14. E. M. McCarthy, M. A. Amussen, W. W. Anderson,
Heredit 74, 502 (1989); V. Grant, Genetics 54, 1189
(1966).

15. Spearman’s rank correlation coefficient was cal-
culated for each pairwise combination of the three hy-
bred lineages. The percentage at which a marker introgressed for a given lineage was scored in the
following manner: 0.0% = 0; 1 to 25% = 1; 26 to
50% = 2; 51 to 100% = 3. Lineages I × II, rS = 0.79,
P < 0.0001; II × III, rS = 0.74, P < 0.0001; III × IV,
rS = 0.73, P < 0.0001.


27, 283 (1994).

18. The following equation was used to compute the test
statistic (p) for two-way epistatic interactions among
unlinked loci (Np×p, number of unlinked loci; Np×p, pro-
geny number of progeny tested; and loci(p), is if the H.
peliolaris marker is absent and 1 if present):

\[ p_{x,y} = \frac{\sum_{p=1}^{Np} \prod_{i=1}^{loci(p)} x_{i,p}}{\sum_{p=1}^{Np} \prod_{i=1}^{loci(p)} x_{i,p} \times \sum_{p=1}^{Np} \prod_{i=1}^{loci(p)} y_{i,p} x_{i,p}} \]

where 0 ≤ p_{x,y} ≤ 1

(1)

Equation (1) can be generalized to N-way epistatic
interactions:

\[ p_{1,2,\ldots,n} = \frac{\sum_{p=1}^{Np} \prod_{i=1}^{loci(p)} x_{i,p}}{\sum_{p=1}^{Np} \prod_{i=1}^{loci(p)} x_{i,p} \times \sum_{p=1}^{Np} \prod_{i=1}^{loci(p)} y_{i,p} x_{i,p} \times \ldots \times \sum_{p=1}^{Np} \prod_{i=1}^{loci(p)} z_{i,p} x_{i,p}} \]

where 0 ≤ p_{x,y} ≤ 1

(2)

Significance for each two- or three-way association
was tested by comparing p_{observed} with p_{expected}
computed by bootstrap randomization of the
observed data (N = 10,000 per association) (Fig. 3).

19. Seventy-four percent of epistatic markers were found
in all three lineages in contrast to 10% of nonepistatic
markers (G2 = 30.0, df = 2, P < 0.0001). However,
part of this correlation may be due to the greater
power of the p-test statistic in detecting associations
among markers of intermediate frequency.

J. Bot. 82, 515 (1995).

21. Positive two-way associations occur when unlinked
H. peliolaris markers appear together within individu-
als of the progeny array more often than would be
expected by chance, suggesting nonadditive, posi-
tive fitness effects (for example, increased pollen vi-
bility) when these markers appear together. By con-
trast, negative associations occur when unlinked H.
peliolaris markers appear together less often than
would be expected by chance, suggesting nonad-
divitiive negative fitness effects when these marks ap-
ppear together.

22. Positive three-way associations are similar to two-
way associations (21), except that three H. peliolaris
markers are involved. Negative three-way associa-
tions may consist of negative two-way associations
only or a combination of negative and positive two-
way associations.

23. We thank M. Debacon, G. Valladeres, and D. Salinas
for technical assistance and J. Whitton, C. Lively,
and N. Barton for helpful comments on an earlier
version of the manuscript. Supported by National
Science Foundation (NSF) grant BSR-9419263 and
USDA 92C337907590 (to L.H.R.) and NSF BIR-
9411728 (to C.R.L).

18 December 1995; accepted 29 February 1996

**TECHNICAL COMMENTS**

**Origin of Replication of Mycoplasma genitalium**

The complete sequence of the genome of Mycoplasma genitalium was reported and an-
alyzed by Claire M. Fraser et al. (1). The origin of replication of the chromosome was not
localized precisely because of the lack of consensus patterns, “DnaA boxes,” in this
species, but it was suggested that the origin might be in an untranscribed AT-rich
region between dnaA and dnaN. This location can be confirmed using a new method based
on the results of the mathematical analysis of the model of DNA evolution under no-
strand-bias conditions (2)—that is, when there is no strand-bias for the mutation pro-
cess or for the selective process between the two strands of DNA.

Under no-strand-bias conditions, the

 equilibrium point is such that the base fre-

quencies in each strand always respect

\[ [A]=[T] \text{ and } [C]=[G] \]

equalities, regardless of the initial state of the DNA sequence and of details of the substitution patterns.

Any significant deviation from the intra-
strand rules \([A]=[T]\) and \([C]=[G]\) is an

indication that there is an inequality in the

substitution patterns between the two

strands of DNA. The null hypothesis for the
detection of such an inequality is then

the appearance of intra-strand equi-

frequencies \([A]=[T]\) and \([C]=[G]\). Because

the mechanisms for DNA replication dif-

fer between the leading strand and the

lagging strand (3), at least in vitro, muta-
tion patterns could differ depending on

---

**Fig. 1.** Origin of replication in M. genitalium, showing switch of polarity of base composition asymmetries. Moving window size and step: top 10 kb, 1 kb; bottom 1 kb, 0.1 kb. Each data point is at the middle of its window.