An AFLP-Based Interspecific Linkage Map of Sympatric, Hybridizing Colias Butterflies

Baiqing Wang1 and Adam H. Porter

Program in Organismic and Evolutionary Biology and Department of Entomology, University of Massachusetts, Amherst, Massachusetts 01003

Manuscript received February 25, 2004
Accepted for publication May 18, 2004

ABSTRACT

Colias eurytheme and C. philodice are sister species with broad syrampthy in North America. They hybridize frequently and likely share a significant portion of their genomes through introgression. Both taxa have been ecologically well characterized and exploited to address a broad spectrum of evolutionary issues. Using AFLP markers, we constructed the first linkage map of Colias butterflies. The map is composed of 452 markers spanning 2541.7 cM distributed over 51 linkage groups (40 major groups and 11 small groups with 2–4 markers). Statistical tests indicate that these AFLP markers tend to cluster over the map, with the coefficient of variation of interval sizes being 1.236 (95% C.I. is 1.234–1.240). This nonrandom marker distribution can account for the nonequivalence between the number of linkage groups and the actual haploid chromosome number (N = 31). This study presents the initial step for further marker-assisted research on Colias butterflies, including QTL and introgression analyses. Further investigation of the genomes will help us understand better the roles of introgression and natural selection in the evolution of hybridizing species and devise more appropriate strategies to control these pests.

The sulfur butterflies, Colias philodice and C. eurytheme (Pieridae), are economic pests of alfalfa and clover crops. They are sympatric and widely distributed over large areas of the United States and Southern Canada (Ferris and Brown 1981; Opler 1992). These two butterflies can be distinguished easily by their wing color (wings of C. philodice are yellow, whereas those of C. eurytheme are orange). They also differ in several other traits including body size, the size of outer wing band, male ultraviolet reflectance pattern, phenome production, and female mating preferences (Silberglied and Taylor 1973, 1978; Grula and Taylor 1979, 1980a,b). Remarkably, almost all of these traits map genetically to the X chromosome (Grula and Taylor 1979, 1980a,b). This so-called “large X-effect” may play a major role in the evolution of Lepidoptera, in which females are the heterogametic (XY; often referred to as ZW) sex (Sperling 1994; Prowell 1998).

Colias butterflies have drawn researchers’ attention for decades, especially in the field of evolutionary and functional ecology. Much is known of their basic biology, including population ecology (Watt et al. 1977, 1979; Tabashnik 1980), phylogenetic relationships (Brunton 1998; Pollock et al. 1998), thermoregulation (Kingsolver 1983; Kingsolver and Watt 1983, 1984; Tsuji et al. 1986), seasonal and spatial variation (Hoffmann 1974, 1978; Ellers and Boggs 2002), oviposition behavior (Stern and Smith 1960; Stanton 1979, 1982, 1984), larval ecology (Austin 1958; Sherman and Watt 1973), mating system (Gerould 1946; Hovanitz 1949; Silberglied and Taylor 1973; Graham et al. 1980; Grula et al. 1980; Grula and Taylor 1980b; Rutowski 1980; Boggs and Watt 1981; Rutowski et al. 1981; Marshall 1982a,b), and adaptation at allozyme loci (Watt 1977, 1983, 1992; Watt et al. 1983, 1985; Carter and Watt 1988). These studies of Colias have melded ecology, genetics, and physiology and have significantly contributed to our understanding of natural selection and adaptation in field settings.

One important feature of Colias butterflies is that several sympatric species pairs, including C. philodice and C. eurytheme, hybridize wherever they are found together (Gerould 1946; Taylor 1972). Although this species pair generally has strong assortative mating, driven largely by female mate choice (Silberglied and Taylor 1978; Grula and Taylor 1980b), hybridization occurs when females, right after eclosing, have not hardened sufficiently to reject heterospecific males (Taylor 1972; Silberglied and Taylor 1978). Hybridization rate is therefore density dependent, increasing to near random at high densities (Taylor 1972). At more typical densities, hybrids constitute ~2–10% of the combined natural population. Laboratory studies showed F1 intercross families usually have lower fitness, which rebounds in most F2 and backcross families (Grula and Taylor 1980b). These results clearly demonstrate that a pathway exists for introgression, allowing the sharing
of neutral and adaptive traits between the taxa. The Colias system, therefore, offers a great opportunity to study the roles of introgression and selection in the organization of closely related genomes.

Although Colias butterflies have been studied for nearly a century and have already been ecologically well characterized, information about nuclear DNA markers, the molecular basis of their species-diagnostic traits, and genetic linkage is still lacking. Amplified fragment length polymorphism (AFLP; Vos et al. 1995), a powerful DNA marker system that has recently gained attention (reviewed by Blears et al. 1998; Mueller and Wolfenbarger 1999), presents an ideal method to study less-known organisms such as Colias. This PCR-based technique is highly reliable and informative, allowing a large number of DNA markers to be generated rapidly with no prior knowledge of DNA sequence. AFLP markers are highly reproducible, with low between-laboratory errors (Jones et al. 1997), and effective in a wide range of organisms. The method has been applied broadly in a variety of studies and proven useful for resolving genetic differences and relatedness among not only individuals or populations, but also independently evolving lineages, especially closely related taxa (Mueller and Wolfenbarger 1999).

The goal of our study was to develop polymorphic AFLP markers and build a comprehensive linkage map of C. eurytheme and C. philodice. We performed the linkage mapping on the basis of a backcross design, given the feasibility of obtaining large F1 and backcross families. Such an interspecific map serves as the foundation for a simultaneous analysis of both genomes. It will allow us to identify markers exhibiting strong species differentiation and to locate genes involved in maintaining species identity. This study provides the initial step toward carrying out further marker-assisted genetic and population studies on a genome-wide basis, which will help us better understand roles of ecological forces and natural selection in the population dynamics of these economic pests and the evolution of adaptive genetic variation.

MATERIALS AND METHODS

Insect materials: Our genetic analysis was performed on a backcross (BC) family since backcross designs give the best resolution for assessing dominant marker frequencies (Hawthorne 2001; Tan et al. 2001) and therefore yield the most reliable maps. In addition, the F2 method suggested by Yasukochi (1998) is not valid here since, unlike many other Lepidoptera species, female Colias do not recombine (Carter and Watt 1988) and therefore maternal-derived markers do not give the chromosome print as seen in silkworm mapping studies. Populations of C. eurytheme and C. philodice were established from wild females collected locally at Amherst and Sunderland, Massachusetts. Individuals from pure-breeding families of the two species were hybridized to produce F1 families. One of the F1 males (of eurytheme female \times philodice male) was crossed to a virgin C. eurytheme female to generate the BC1 population for mapping. The offspring were reared in petri dishes supplied with fresh-cut alfalfa as a food source. The rearing conditions were 27°C with a photoperiod of 14 L:10 D. Adult butterflies were frozen right after they emerged and kept at −80°C until DNA isolation. A total of 58 backcross individuals (31 females and 27 males) were genotyped using the AFLP method. This number was found to be suitable for linkage analysis with a sufficient level of statistical confidence (Ky et al. 2000; Tan et al. 2001).

AFLP analysis: Genomic DNA was isolated using a modified CTAB method as described by Del. Sai. et al. (1989) and Gustinuch et al. (1991). The DNA was purified by extraction with phenol/chloroform, precipitated by ethanol, and resuspended in TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). AFLP analysis was conducted according to the procedures and reaction conditions described by Vos et al. (1995) with modifications as described below.

Restriction/ligation reactions were performed in an 11-µl single-tube reaction with 250 ng of genomic DNA, 0.5 units of MseI, 5 units of EcoRI, 4 units of T4 DNA ligase (New England Biolabs, Beverly, MA), 4.5 µM MseI adapter, 0.45 µM EcoRI adapter, 0.05 mg/ml BSA, 50 mM NaCl, and T4 ligase buffer (50 mM Tris-HCl, 10 mM MgCl2, 10 mM DTT, 1 mM ATP, and 25 µg/ml BSA) for 3 hr at 37°C. Preamplifications were run in 20-µl reactions containing 3 µl of the diluted (1:4) restriction/ligation product, 300 µM of both primer EcoRI+A and primer MseI+C, 1 unit Tag polymerase, 0.2 mM dNTPs, and PCR buffer (15 mM MgCl2, 500 mM KCl, and 100 mM Tris-HCl, pH 8.3). The cycling conditions for preamplifications were 20 cycles of 94°C for 30 sec, 56°C for 30 sec, and 72°C for 2 min. Selective amplifications were run in 20-µl reactions containing 3 µl of the diluted (1:10) preamplification reaction product, 50 mM of EcoRI+1 primer (labeled with fluorescent dyes), 250 mM of MseI+3 primer, 1 unit AmpliTag Gold polymerase (Applied Biosystems, Foster City, CA), 0.2 mM dNTPs, and Geneline PCR buffer (15 mM MgCl2, 500 mM KCl, 100 mM Tris-HCl, pH 8.3, and 0.01% w/v gelatin). Selective PCR conditions were 10 cycles of 94°C for 30 sec, 65°C (−1°C/cycle) for 30 sec, and 72°C for 2 min followed by 35 cycles of 94°C for 30 sec, 56°C for 30 sec, and 72°C for 2 min. All PCR reactions were carried out in a Perkin Elmer thermocycler 9700 (Applied Biosystems). PCR products from three primer pairs labeled with different dyes were pooled and analyzed on a 96-lane sequencing gel using an ABI Prism 377 DNA sequencer (Applied Biosystems).

Genotype analysis: AFLP gels were visualized and analyzed using GeneScan software v. 2.1 (Applied BioSystems). Fragment size data were downloaded and sorted using BinThere software (N. Garnhart, University of New Hampshire). All the data points were then checked manually to match the corresponding bands on the gel. Misassigned data were corrected and ambiguous fragments were counted as missing data. Therefore, all markers were ultimately scored as presence (+) or absence (−) of the amplification (band), or unknown (missing data).

AFLP markers were sorted into three categories with independent segregation patterns: (1) amplification present only in the mother (C. eurytheme), (2) amplification present only in the father (F1 hybrid), and (3) amplification found in both parents. Results of linkage analysis using markers from only the second category are reported here since those markers give the linkage of both autosomes and the X chromosome (in Lepidoptera, females are the heterogametic sex) and thus give the best representation of the genome. Information on other markers and their linkage will be provided by the authors upon request.

Linkage map construction: Segregating AFLP markers were tested for deviation from expected 1:1 segregation ratios by
chisquare analysis ($P < 0.05$), with significance levels corrected for multiple comparisons (Rice 1989). AFLP data showing no significant deviation were used to calculate linkage. Their segregation type was coded as backcross with the banded genotype as heterozygous (H) and nonbanded as homozygous recessive (A). Linkage analysis was performed using the mapping software MapMaker/Exp v. 3.0 (Lander et al. 1987) and confirmed using Map Manager QTX (Manly et al. 2001). Linkage was determined with the criteria of LOD $\geq 3.0$ ($P = 0.001$ for Map Manager QTX) and a maximum recombination fraction of 0.55 (see Lander et al. 1987; Beckmann 1994; and Liu 1997 for discussion of the criteria used). Marker orders were estimated using the Kosambi (1944) mapping function. To study the marker distribution along the map, we used the chi-square test for goodness of fit as described by Roupp van der Voort et al. (1997) to test if the AFLP markers were randomly distributed within a linkage group. We also used the Kolmogorov-Smirnov and Lilliefors one-sample test (Kolmogoroff 1941; Lilliefors 1967) (on standardized data) to compare the distribution of marker intervals between consecutive loci along the map against the null expectation that they follow a normal distribution.

**RESULTS AND DISCUSSION**

AFLP genotypes and segregation distortion: A total of 1242 bands (69.3%) showed presence/absence of polymorphisms, indicating a substantial amount of molecular variation. High levels of polymorphism were previously observed at several allozyme loci within wild Colias populations, also suggesting Colias are highly polymorphic (Burns and Johnson 1967; Watt et al. 1985; Watt 1992). Among those polymorphic AFLP markers, 510 followed amplification pattern 2, with father banded and mother unbanded. Of these, 470 showed no deviation from the expected 1:1 ratio and were selected for the linkage analysis. For each primer combination, the number of bands used for mapping varied from 1 to 24 (Table 1), with an average of 7 mappable bands per primer combination.

The overall frequency of our AFLP loci showing segregation distortions was $\sim 12\%$ (149/1242), lower than the rate reported between strains in silkworm (Tan et al. 2001). Skewed segregation ratios have been observed commonly in AFLP loci, but the frequency of distorted loci is highly variable (Kocher et al. 1998; Vink et al. 1998; Ky et al. 2000; Katengam et al. 2002; Liu et al. 2003). Segregation distortion of AFLPs can be caused by constringing fragment complexes, linkage to lethal genes, variations in genomic DNA isolation, or simply vagaries of PCR (Vink et al. 1998; Nakaido et al. 1999; Ky et al. 2000; Tan et al. 2001). Lack of modifiers to suppress meiotic drive in hybrids might also be responsible for some of the distortions observed in interspecific crosses (Dermitzakis et al. 2000; Schwarz-Sommer et al. 2003; Wilkinson et al. 2003). In Colias butterflies, this scenario is perhaps less likely to occur since the genomes of these two species are homogenized by frequent hybridization. Introggression and high similarity between the two species may also account for the low distortion rate we observed.

Genetic linkage map: Of the 470 segregating markers tested, 452 (96%) showed detectable linkage to another polymorphism, and 18 markers remained unassigned.
The final linkage map comprised 51 linkage groups (Figure 1), encompassing a total of 2541.7 cM. Our linkage groups ranged in size from 0 to 183.4 cM (mean, 49.8 cM). The number of AFLP markers per group varied from 2 to 23, with an average of 9 markers. There were 40 major linkage groups with 5–21 markers and 11 small groups with <5 markers. The mean distance between adjacent loci was 6.3 cM (±0.4). There were seven gaps >30 cM in length distributed among 7 linkage groups. The longest was 34.7 cM on linkage group 37, approaching the maximum recombination fraction of 0.35.

The genome of *C. eurytheme* and *C. philodice* is characterized by numerous small chromosomes, often similar in size (Maeki and Remington 1960; Grula and Taylor 1980a). Mapping genomes that have a large number of chromosomes is often difficult (Yasukochi 1998; Liu et al. 2003). Both taxa have a haploid number of 31, typical of the Lepidoptera (Remington 1954; Maeki and Remington 1960), but much smaller than our AFLP map of 51 linkage groups. Obviously, some of the linkage groups are located on the same chromosomes and large gaps exist between those groups; additional markers are needed to bridge those gaps. Matching the linkage group number to a high chromosome number usually requires many more markers or a combination of dominant and codominant markers (Yasukochi 1998). In some cases, an insufficient number of markers may still give a similar number of linkage groups to the actual chromosomal number, but likely without one-to-one correspondence due to missing information on a subset of the chromosomes.

Since our linkage map contains 20 linkage groups more than the actual haploid chromosomal number, the complete recombination length of the genome should be higher than the map length (2541.7 cM) after

---

**Figure 1.**—Linkage map of *Colias* butterflies constructed from 58 backcross hybrids derived from the interspecific cross [*eurytheme* × (*eurytheme* × *philodice*)]. A total of 452 AFLP markers are described in terms of the selective nucleotides used and the fragment size; for example, AACCTA339—the *Eco*RI primer (AAC), the *Mse*I primer (CTA), and the size of the band (339 bp). Recombination distances from the origin (in centimorgans) are given on the left side of each linkage group (LG) and marker names are to the right.
adding the flanking regions of the extra linkage groups. This accounts for at least an additional 700 cM (20 LGs × 35 cM/LG). Therefore, it seems that the total recombination length of Colias is substantially higher than that of the silkworm, which was estimated to be ~2000 cM (Yasukochi 1998). The longer length shown in Colias may reflect a larger genome size or, we suspect, higher rates of crossing over than in silkworm. High rates of recombination in some regions of the genome may also generate large gaps, causing markers on the same chromosomes to be spuriously assigned to two or more linkage groups.

**AFLP marker distribution:** Our AFLP markers placed on the linkage map deviated significantly from a random distribution as suggested by two different statistical tests. Clustering of AFLP markers on linkage groups 2, 15, and 25 was revealed by the chi-square test for goodness of fit ($P \leq 0.05$). The test, however, is effective only for groups with >10 markers, which is not the case for a majority of our linkage groups. Therefore, clustering of our AFLP markers may occur more frequently than shown by the test. The Kolmogorov-Smirnov and Lilliefors one-sample test indicated that the distribution of intervals between consecutive markers significantly deviated from a normal distribution ($P \leq 0.001$, Figure 2). The coefficient of variation of interval sizes was 1.236 (95% C.I. is 1.234–1.240, calculated using a jackknife method), >1. This also suggests our AFLP markers were not distributed at random but aggregated spatially over the map.

AFLP markers tend to cluster around regions where recombination is suppressed, usually corresponding to
centromere and telomere regions (Tanksley et al. 1992; Roupppe van der Voort et al. 1997; Alonso-Blanco et al. 1998; Qi et al. 1998; Miklas et al. 2001). Since Lepidopteran chromosomes are holocentric (centromeres spread over >70% of the genome; Murakami and Imai 1974; Padhy 1986; Marec et al. 2001), dense clusters are expected to be observed in large portions of the map if suppressed recombination occurs at the centromeric regions. To obtain a saturated map of a 3000-cM genome (probability of coverage, 0.95), only 279 polymorphic markers would be needed if they were randomly distributed (Krutovskii et al. 1998). But for markers that are not distributed at random, a substantially higher number is required to achieve the same level of coverage. Therefore, it is not surprising that with 470 markers placed on the Colias map, we still observed a large number of gaps, causing our number of linkage groups to be more than the actual number of chromosomes.

**Perspectives:** Our AFLP map of Colias butterflies is one of the few linkage maps that have been reported in Lepidoptera. Although this order represents an extremely diverse and economically important group of insects, mapping studies have been conducted only in the silkworm, *Bombyx mori*. Several highly saturated linkage maps (Doira 1992; Yasukochi 1998; Tan et al. 2001) and a draft whole-genome sequence (Mita et al. 2004) of the silkworm have been published. Mapping studies are currently ongoing in at least two other groups of Lepidoptera (McMillan et al. 2002), which will further increase our knowledge about lepidopteran genomes.

This study presents the starting point for further molecular-based research on Colias butterflies. The map builds the foundation for thoroughly exploring the entire genome represented by a large number of mapped AFLP markers. It creates a framework for anchoring morphological or other molecular markers and identifying quantitative trait loci (QTL) for taxon-diagnostic, geographically varying, and economically important traits. This map also can be utilized to locate genes of interest and to develop DNA probes, SNPs, STS/expressed sequence tags, or other codominant DNA markers with a wider range of applications. Further in-
Interspecific Linkage Map of Colias

<table>
<thead>
<tr>
<th>LG 16</th>
<th>LG 17</th>
<th>LG 18</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>ACTCTA194</td>
<td>0.0</td>
</tr>
<tr>
<td>4.1</td>
<td>ACTCTG237</td>
<td>3.7</td>
</tr>
<tr>
<td>11.1</td>
<td>ACTCAC183</td>
<td>6.0</td>
</tr>
<tr>
<td>16.7</td>
<td>ACACGT223</td>
<td>19.1</td>
</tr>
<tr>
<td>19.1</td>
<td>AGGCAG158</td>
<td>21.5</td>
</tr>
<tr>
<td>25.5</td>
<td>ACCGAT178</td>
<td>23.6</td>
</tr>
<tr>
<td>33.7</td>
<td>ACCGCT3178</td>
<td>27.1</td>
</tr>
<tr>
<td>39.3</td>
<td>ACTCAA239</td>
<td>56.2</td>
</tr>
<tr>
<td>39.4</td>
<td>ACACAC97</td>
<td>65.6</td>
</tr>
<tr>
<td>46.7</td>
<td>AGGCACA206</td>
<td></td>
</tr>
<tr>
<td>48.9</td>
<td>AGGCAG460</td>
<td></td>
</tr>
<tr>
<td>61.8</td>
<td>ACACCT115</td>
<td>58.2</td>
</tr>
<tr>
<td>65.6</td>
<td>ACGCTA223</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LG 19</th>
<th>LG 20</th>
<th>LG 21</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>AGGCCTT270</td>
<td>0.0</td>
</tr>
<tr>
<td>5.7</td>
<td>ACACCTT173</td>
<td>9.1</td>
</tr>
<tr>
<td>7.6</td>
<td>ACCGAG157</td>
<td></td>
</tr>
<tr>
<td>22.9</td>
<td>ACTCA185</td>
<td>16.7</td>
</tr>
<tr>
<td>27.0</td>
<td>ACCGAG338</td>
<td>21.2</td>
</tr>
<tr>
<td>28.7</td>
<td>AGGCAG308</td>
<td></td>
</tr>
<tr>
<td>34.0</td>
<td>ACGGTG210</td>
<td></td>
</tr>
<tr>
<td>34.5</td>
<td>ACTCTA89</td>
<td></td>
</tr>
<tr>
<td>39.1</td>
<td>ACGCTA104</td>
<td>45.4</td>
</tr>
<tr>
<td>42.5</td>
<td>AGGCACA168</td>
<td></td>
</tr>
<tr>
<td>51.1</td>
<td>AGCCAA90</td>
<td>54.9</td>
</tr>
<tr>
<td>51.9</td>
<td>ACGGCT339</td>
<td></td>
</tr>
<tr>
<td>55.5</td>
<td>ACGGCT339</td>
<td></td>
</tr>
<tr>
<td>55.1</td>
<td>ACGCAG230</td>
<td></td>
</tr>
<tr>
<td>55.2</td>
<td>ACGCAG232</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1.**—Continued.

--

vestigation of the Colias genome will allow us to identify factors that maintain their species integrity, to understand the trade-offs between introgression and adaptation, and to measure quantitatively the species boundary. Ultimately, it will help us to predict the population and evolutionary dynamics of these hybridizing agricultural pests and design more appropriate control strategies.

We thank G. Gibson, C. Jiggins, L. Katz, B. Normark, and W. Watt for helpful discussions and critical reading of the manuscript. We also thank M. Henshaw for providing AFLP technical assistance, E. Levin and M. Sei for assisting with butterfly rearing, and an anonymous reviewer for additional constructive suggestions. This work was supported by grants from the National Science Foundation (DEB-0129234) and the U.S. Department of Agriculture, Cooperative State Research Extension and Education Service, awarded through the Massachusetts Agricultural Experiment Station under project no. MAS00789, paper no. 3345.

### LITERATURE CITED


Manly, K. F., R. H. Cudmore, Jr. and J. M. Meek, 2001 Map Manager


**Figure 2.**—Size distribution of linkage intervals between two successive markers in the interspecific linkage map of Colias butterflies. $m$, mean distance between adjacent markers $\pm$ SE; $N$, number of intervals; $P$, probability in the Kolmogorov-Smirnov and Lilliefors one-sample test (on standardized data) to compare the distribution of marker intervals against the null expectation of a normal distribution.


**Watt, W. B., R. C. Cassin and M. S. Swan, 1983** Adaptation at specific loci. III. Field behavior and survivorship differences among Colias PGI genotypes are predictable from *in vitro* biochemistry. Genetics 103: 725–739.


