Locus-dependent selection in crop-wild hybrids of lettuce under field conditions and its implication for GM crop development

Hooftman, D.A.P.; Flavell, A.J.; Jansen, H.; den Nijs, J.C.M.; Syed, N.H.; Sørensen, A.P.; Orozco-ter Wengel, P.; van de Wiel, C.C.M.

Published in:
Evolutionary Applications

DOI:
10.1111/j.1752-4571.2011.00188.x

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Locus-dependent selection in crop-wild hybrids of lettuce under field conditions and its implication for GM crop development

Danny A. P. Hooftman,1,2 Andrew J. Flavell,3 Hans Jansen,4 Hans C. M. den Nijs,1 Naeem H. Syed,3 Anker P. Sørensen,5 Pablo Orozco-ter Wengel1,5 and Clemens C. M. van de Wiel6

1 Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam; Amsterdam, The Netherlands
2 Centre for Ecology and Hydrology, Wallingford, UK
3 University of Dundee at SCRI, Dundee, UK
4 Biometris, Wageningen UR, Wageningen, The Netherlands
5 Keygene N.V, Wageningen, The Netherlands
6 Wageningen UR Plant breeding, Wageningen, The Netherlands

Introduction

New crop cultivars are continuously generated to deal with the challenges we face as a consequence of human population growth and a changing environment (Beddington 2010). In case of genetically modified ('transgenic') crops, one of the main foci is on developing cultivars tolerant not only to biotic but also to various abiotic stresses (Bhatnagar-Mathur et al. 2008; Ashraf 2010). At the same time, there is considerable concern that genes escaping from crops could result in establishment ('introgression') of transgenes from crops into wild relatives. Potentially, this would facilitate range expansions of hybrid taxa and thus affect natural plant populations and habitats (Ellstrand 2003; Chandler and Dunwell 2008). Therefore, assessing ecological consequences of crop-wild hybridization has become an important aspect in the development of risk assessment strategies for transgenic crops (Snow et al. 2005; Chapman and Burke 2006).

Not all genes that escape from crops have the same likelihood of introgressing into wild relatives (Stewart et al. 2003; Baack et al. 2008). Linkage Disequilibrium (LD) – the nonrandom association of alleles at separate loci – in hybrid genomes could lead to genetic hitchhik-
ing of transgenes with native crop genes that do confer a selective advantage under natural field conditions (MacKay and Powell 2007). On the other hand, LD could be directionally disrupted through experiencing opposing selective pressure upon physically linked loci (Stewart et al. 2003). Under field conditions, the result would be a segregation pattern that is systematically different from neutral expectations in progeny as many, less fit, genotypes are purged from populations (Burke and Arnold 2001). The extra-agricultural environment, into which crop genes escape, does likely provide a variety of strong stresses. Depending on the selective strength exerted by the environment, distinct changes in LD could rapidly take place around selected genes. At the same time, other un- or less-selected genes in a positively selected area of the genome could introgress, through genetic hitchhiking, at higher than expected rates (Barton 2000; Tanaka 2007). Therefore, genomic segments under strong selection as a whole would behave like basic units to be considered when assessing the potential of introgression of crop alleles into wild relatives, rather than individual genes. Identifying hotspots and cold areas within the genome related to the likelihood of crop gene introgression could be the first step in employing such information for targeting transgene insertion to the safest genomic areas with regard to transgene persistence in wild populations (Stewart et al. 2003).

Gene flow is seen as providing evolutionary useful new variation to plant populations (Schierebeck and Ellstrand 2009). This adaptive potential of introgressed hybrid swarms has often been emphasized as a large-scale phenomenon (Burke and Arnold 2001; Chapman and Burke 2006). The latter process is suggested to have happened in just tens of years in Californian Raphanus (Hegde et al. 2006; Ridley et al. 2008) and British Rhododendron (Milne and Abbott 2000). However, the fate of individual parts of the parental genomes – and associated characteristics – after a hybridization event is as yet less obvious (Baaek et al. 2008; Rose et al. 2009). Therefore, predicting the likelihood of introgression of specific genes in a hybrid genome is a tough task, which should start with adequate quantitative estimations of the likelihood of different genomic regions to introgress into wild species (Stewart et al. 2003; Martin et al. 2005). Our study is aiming at providing the first steps in exactly doing that, employing nontransgenic crop-wild relative hybrids in lettuce as model system.

Hooftman et al. (2006) hypothesized that crop-wild hybridization could have contributed to the recent invasiveness of prickly lettuce (Lactuca serriola) in Europe. Experimentally, an advantage in survival capacity was identified for at least four generations of synthetic hybrids between lettuce (L. sativa) and wild L. serriola over wild-type plants (Hooftman et al. 2005, 2007). However, increased survival could not readily be related to specific morphological or reproductive traits, indicating complex trait interactions (Hooftman et al. 2005). We suggest that a sorting of genotypes by mortality, favoring distinct genomic segments from one of the parents, under field conditions could be responsible (Hooftman et al. 2009). Under experimentally applied selection targeted at leaf length and flowering time, phenotypes have been shown to change rapidly in Raphanus (Campbell et al. 2009). Therefore, in this contribution to the series of papers into hybrid fitness in Lettuce, we will study the underlying genetic mechanisms of these fitness differences in more depth.

We will test whether selection attributable to natural field conditions leads to preferred combinations of chromosomal elements originating from both species instead of one preferential parent. Putatively, such mosaic could provide the suggested new additive interactions. We employ progeny derived from a cross between a lettuce butterhead cultivar (L. sativa var. capitata) and two Dutch/German wild L. serriola accessions. Three generations of hybrids (S1, BC1, and BC1S1) were grown under natural field conditions, providing nonartificial and a likely wide variety of different selective forces. A control generation was grown in the greenhouse. Markers generated by three different molecular genetic methods (AFLP, SSAP, and NBS profiling) were used, following the association map of Lactuca made for this identical cross (Syed et al. 2006). Our study is based upon two analyses, which in combination could provide tentative indications of genomic hotspots and cold areas for crop allele introgression under natural field conditions.

1 Changes in allele frequencies. Following ‘conventional wisdom’, alleles originating from crop taxa in hybrids are less favorable under natural conditions (Hails and Morley 2005), which would lead to selective retention of alleles from the wild taxon under natural conditions. We estimate changes in allele frequencies among surviving hybrids after field exposure relative to frequencies observed in the absence of such exposure (‘distortion’).

2 Linkage disequilibrium. Introgression likelihood of individual genes does not only depend on their own fitness associations but also on the strength of their linkage with other loci (Stewart et al. 2003). Here, we test for consistent postzygotic preservation and disruption of LD.

Materials and methods

Study organisms

For extensive descriptions of the two parental taxa, we refer to Tutin et al. (1976), as well as Hooftman et al. (2005, 2006). In short, Lactuca serriola L. (Asteraceae,
2n = 18) is a common annual weed found in anthropogenically disturbed habitats, such as roadsides, railways, and ruderal sites in urban areas. The species is native to North Africa, Western Asia, and Europe (Carter and Prince 1985). Nowadays, it has been reported as well in South Africa and the Americas (Lebeda et al. 2001). Plants are predominantly autogamous with outcrossing rates of approx. 1% via insect vectors. Seeds of L. serriola are suited for wind dispersal having a parachute-like pappus (D ’Andrea 2006).

Lactuca sativa L. (Asteraceae, lettuce, 2n = 18) is a common annual crop, which flowers frequently in allotments, largely simultaneously and often sympatrically with L. serriola. It is considered conspecific with, and derived from, L. serriola. Both taxa are fully interfertile, with no known pre- or postzygotic barriers (Koopman et al. 2001). The outcrossing rate of L. sativa is similar to that reported for L. serriola (0.5%; Giannino et al. 2008). Seeds of L. sativa have similar parachute-like pappus as L. serriola, but through cultivated erect involucral bracts are less able to be released.

Plant material and field experiments

The employed hybrid generations and their performances are described in Hooftman et al. (2005, 2007); their pedigree is depicted in Fig. 1. Explaining the hybridization procedure in short: F1 hybrids were created using L. serriola as mother plants and butterhead L. sativa cultivar (var. capitata cv. Dynamite) as pollen donors. The crossing technique followed the protocol by Nagata (1992) with minor alterations. We let F1 hybrids self-pollinate to obtain a generation of autogamous hybrid progeny (hereafter called S1). Backcrossing (BC1) was performed using a similar experimental design. Through this setup, the BC1S1 generation two growing seasons. The control generation of autogamous hybrid progeny (BC1S1) of these BC1 plants was collected and resown in April 2004 at the site of production utilizing the same DNA extracts. In further contrast to the field sampling, young plants were collected, which

![Pedigree of hybrid generations used. Solid lines indicate (back)crossing, and dotted lines indicate selfing. Boxed generations were exposed to nonartificial natural field conditions. The S1 and BC1 generations underwent one growing season of field exposure and the BC1S1 generation two growing seasons. The control generation (S1control) underwent no field exposure.](image)

Selection in lettuce hybrids

We employed the genetic map of Syed et al. (2006) based on an S1 generation from one of these interspecific crosses to obtain hybrid populations for field testing. This map consists of nine linkage groups (LGs), which are assumed to correspond to the nine chromosomes of lettuce. Based on preliminary analysis by both SSRs and AFLPs (Keygene N. V., unpublished data), the L. serriola paternal material employed here consisted of four genotypes, with possible heterozygosity identified for 40 loci. Those genotypes were present in the wild populations from which L. serriola seeds for F1 creation were derived (The Netherlands, N50°49’; E05°55’; Germany, N53°32’ E10°54’).

The experiments were split over three locations: Amsterdam (Netherlands, N 52°21’, E 04°58’), Sibekarpel (Netherlands, N 52°42’, E 05°00’), and Aachen (Germany, N 50°46’, E 06°03’). At each location, two 100 m² plots were demarcated and ploughed, mimicking anthropogenic soil disturbance. In April 2003, seeds of L. serriola, L. sativa, and S1- and BC1 hybrids were sown individually with 30-cm spacing. The grid positions of seeds within plots were fully randomized. A pioneer vegetation was allowed to emerge spontaneously next to Lactuca, exerting conditions similar to many natural habitats of L. serriola. The resulting mortality rate of Lactuca plants was approximately 90%, offering a window of opportunity for postzygotic selection. In August 2003, autogamous progeny (BC1S1) of these BC1 plants was collected and resown in April 2004 at the site of production using a similar experimental design. Through this setup, field exposure comprising two growing seasons was applied to this BC1S1 generation.

Sampling, molecular marker methods, and loci used

Leaf material was sampled in July 2003 (BC1 and S1) and July 2004 (BC1S1) over the three experimental locations (49 plants generation⁻¹). Sampled plants included almost all surviving individuals with seed-set for these generations. Hooftman et al. (2005, 2007), using these very same plants, did not observe significant interactions between site and fitness, including survival and germination rates as well as fecundity. Therefore, all sites were pooled to obtain sufficient statistical power for the main analysis (but see Table S4 for an analysis per field).

The control S1 generation (91 plants) was grown under fully controlled greenhouse conditions; individual plants were the same as used for the map of Syed et al. (2006), utilizing the same DNA extracts. In further contrast to the field sampling, young plants were collected, which
could include genotypes with low or no reproductive fit-
ness. Consequently, the control generation represents only
very limited postzygotic selection and hence can be used
as baseline for segregation patterns.

DNA was extracted using a modified CTAB extraction
method (Syed et al. 2005). All subsequent procedures
were identical as used in Syed et al. (2006), performed by
the same laboratories and technical staff – i.e., Keygene
N.V. for AFLP loci (scored codominantly), SCRI, Dundee
for SSAP loci, and Plant Breeding, Wageningen for SSAP
and NBS profiling loci (SSAP and NBS scored domi-
nantly). NBS and SSAP are the more recently developed
marker types. NBS loci linked to pathogen resistance (R)
genes actively used in breeding (Van der Linden et al.
2004). SSAPs are retrotransposon-based markers, expected
to behave as alleged neutral markers like AFLP (e.g.,
Cornman and Arnold 2009; Du et al. 2009). In AFLPs,
differentiating between heterozygotes and dominant ho-
mozygotes was performed with Xcelerator v.1.3.1. soft-
ware package (proprietary software, Keygene N.V.). This
program differentiates between those two classes depend-
ing on the intensity of the bands. To eliminate possible
selfings that occurred during the BC1 crossing process
(see Hooftman et al. 2005), we conducted a SAHN
(Sequential Agglomerative Hierarchical Nested) cluster
analysis using UPGMA parameters (Unweighted Pair-
Group Method, arithmetic Average). One BC1 plant and
one BC1S1 plant were excluded because this analysis sug-
gested an autogamous origin instead of backcrossing for
these plants.

We preselected 272 loci from the map in Syed et al.
(2006) for our analysis, aiming at an optimal coverage of
the genome and the nine linkage groups. Afterward, 40
loci with an alleged L. serriola band presence state were
discarded because of the possible presence of wild paren-
tal alleles other than the one used for the map. We
checked this by detecting false negatives –i.e., unexpected
and theoretically impossible homozygotes for the alleged
crop allele (band absence) – in the BC1 generation.

The cleaned data set consisted of 232 informative loci
containing 89 AFLP loci, 115 SSAP loci, and 28 NBS loci;
186 of those loci were directly located on one of the nine
linkage groups, with an average distance of 5.9 cm. We
included an additional anonymous set of 46 loci yet to be
assigned to one of the LGs. This latter group included
AFLP loci from both parentage and SSAP and NBS loci
originating from L. sativa. All analyses were performed
separately for each of the three hybrid generations.
Codominant loci failing to differentiate among hetero-
and homozygotes were regarded as not interpretable. Loci
for which less than half of the plants per generation had
interpretable banding patterns were regarded as missing
data.

Changes in allele frequencies in selected progeny
Significance levels for changes in allele frequencies were
tested by estimating the bias in the segregation of individ-
ual loci in progeny being exposed to field conditions,
conditions that provided a variety of nonartificial selective
forces. This was tested against patterns observed in the
control generation, which were not subjected to postzyg-
otic selection. For this greenhouse-grown control (S1)
generation, the distortion of loci was established through
comparing to expected neutral segregation. The observed
(baseline) distortion was subsequently extrapolated per
individual locus to hypothetical BC1 and BC1S1 expected
marker distributions (Supporting information: Tables S1
and S2). We will further refer to this significant bias as
‘segregation distortion’. We created confidence intervals
using Monte Carlo simulations adjusting the significance
threshold based on the number of tests. This analysis and
the subsequent LD calculations were written in Matlab;
the code is available as Supporting information (Data S1).
The two-step analysis proceeded as follows:

1 To take into account the random loss of alleles with
low frequencies because of the sampling of all 92 available
plants in the field, a Monte Carlo approach was imple-
mented. In each simulation, 92 virtual diploid plants were
created. In those plants, per locus, the state – homozy-
gous for one of the parents or heterozygous – was drawn
from the expected distribution of these states (Supporting
information: Table S2). The probabilities of obtaining
states per locus were based on the allele frequencies
extrapolated from the control generation, i.e., not from
Mendelian except for the control itself (Supporting infor-
mation: Table S1). For 17 loci with missing data in the
control generation, we assumed neutral segregation. Sub-
sequently, the obtained frequencies per locus of this sam-
ple of 92 plants were calculated. This procedure was
repeated 50 000 times.

2 Confidence intervals for a given significance level (α)
were obtained from the above obtained sampling distri-
bution of state frequencies per locus. To correct for the
number of tests (=loci) performed, we allowed for a type
I error of one single locus, consequently α = (1/# loci)
per generation. To give an example, for 211 loci, the
actual α would be 0.0047, and the confidence interval
would represent the 49 764th value in the sorted array
per locus from the above bootstrap procedure. The result-
ing α values are the following: control: 0.0047 (for
N = 213); S1: 0.0045 (N = 220); BC1: 0.0063 (N = 158);
and BC1S1: 0.0047 (N = 211). The corresponding confi-
dence limits per locus per generation are provided in
Supporting information (Table S3). A locus with an
observed state frequency above the confidence interval
is labeled as significantly distorted. The P-value distribu-
tions
for all tests are provided in the Supporting information. These distributions show a bias in the direction of either high or low p-values of field-exposed generations, indicating a strong distinction between loci with segregation distortion and loci without distortion (Fig. S1).

Increased and decreased LD

We tested for nonrandom associations among loci, using their correlation coefficients ($r$; Flint-Garcia et al. 2003), and we will subsequently use the term linkage disequilibrium (LD) when referring to this estimate (Eqn 1).

$$\text{LD}_i = r_i = \frac{x_{11} - p_1 q_1}{\sqrt{p_1 p_2 q_1 q_2}}$$

where $X_{11}$ as the frequency of homozygous $L$. sativa alleles for adjacent locus gap $i$; $p_1$ and $p_2$ the frequency of $L$. sativa and $L$. serriola alleles on the first haploid, respectively; $q_1$ and $q_2$ the frequency of $L$. sativa and $L$. serriola alleles on the second haploid, respectively.

Pairwise codominant/dominant combinations were integrated through transforming codominant scoring into dominant by assuming the band presence state being dominant in heterozygotes. To test for alterations in LD among generations, the observed levels of LD were compared with an expected level of LD, expressed as percentage. This comparison was achieved through calculating the ratio of LD per LG of control and field-tested generation, averaging over the adjacent pairwise combinations. We use the marginal LD ($\text{LD}_i^M$)values per generation as our estimate, i.e., the value for each pairwise combination adjusted for the level of LD in the control generation according to Eqn 2. In this way, we are able to compare per adjacent pairwise combination the relative decrease or increase in the control generation. A Chi-square test was included to test for statistical significance (df = 1; Eqn 3). We employed a correction for multiple testing following Benjamini and Hochberg (1995), with $z = 0.05$ as base value.

$$\text{LD}_i^M = \frac{\text{LD}_i^T}{\frac{1}{n} \sum_{i=1}^{n} \text{LD}_i^T}$$

$$\chi^2 = \frac{(\text{LD}_i^M - \text{LD}_i^C)^2}{\text{LD}_i^C} + \frac{((100 - \text{LD}_i^M) - (100 - \text{LD}_i^C))^2}{(100 - \text{LD}_i^C)}$$

where $\text{LD}_i^M$ = marginal LD for adjacent locus gap $i$; $\text{LD}_i^T$ = observed LD in the tested generation for adjacent locus gap $i$; $\text{LD}_i^C$ = observed LD in the control generation for adjacent locus gap $i$, and $r =$ number of adjacent locus gaps in the Linkage Group.

Genomic hotspots

We will in our discussion tentatively identify both genomic hotspots and cold areas for introgression of crop alleles into wild generations (Stewart et al. 2003). Such areas are defined by unidirectional distorted chromosomal parts of 2 loci or more, co-occurring with a higher than expected LD. We will, speculatively, indicate general genomic regions but refrain from being very specific. Further studies should narrow these regions.

Results

Single-locus allele frequency changes

Of the 232 informative loci, 66 showed allele frequencies significantly different from expected in one or more field-

### Table 1. Number of loci with significant segregation distortion after field exposure for either parental species compared with expected segregation based on an extrapolated control generation (Supporting information Table S1).

<table>
<thead>
<tr>
<th>Generation</th>
<th>Marker type</th>
<th>total no. of loci</th>
<th>L. serriola allele</th>
<th>L. sativa allele</th>
<th>Heterozygotes</th>
<th>Directional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>AFLP</td>
<td>73</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>ns</td>
</tr>
<tr>
<td>S1</td>
<td>NBS/SSAP</td>
<td>80</td>
<td>6</td>
<td>0</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>BC1</td>
<td>AFLP</td>
<td>72</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>ns</td>
</tr>
<tr>
<td>BC1S1</td>
<td>NBS/SSAP</td>
<td>80</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>ns</td>
</tr>
<tr>
<td>S1</td>
<td>NBS/SSAP</td>
<td>20</td>
<td>2</td>
<td>1</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>BC1</td>
<td>AFLP</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>BC1S1</td>
<td>AFLP</td>
<td>24</td>
<td>1</td>
<td>4</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>NBS</td>
<td>112</td>
<td>0</td>
<td>9</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>S1</td>
<td>NBS</td>
<td>112</td>
<td>5</td>
<td>2</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>BC1</td>
<td>AFLP</td>
<td>69</td>
<td>3</td>
<td>11</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>BC1S1</td>
<td>AFLP</td>
<td>115</td>
<td>14</td>
<td>20</td>
<td>ns</td>
<td></td>
</tr>
</tbody>
</table>

$^*P < 0.05$ indicates a noneven, i.e., directional, distribution of distorted loci among parental species, Chi-square, AFLP: df = 2; NBS/SSAP: df = 1.

$^a$Codominantly scored BC1 loci (backcross with L. serriola) are either L. serriola homozygotes or heterozygotes.

$^b$Proportionally not more distorted than the control generation ($P > 0.05$; Chi-square, AFLP: df = 2; NBS/SSAP: df = 1).

$^c$N in the BC1 generation is lower, as dominant loci with a dominant L. serriola allele will show all a band being either heterozygous or homozygous L. serriola.
exposed generations, compared with the control generation (Table 1, Fig. 2). This included a large number of loci for which crop alleles were in excess of expectations. We attribute this distortion to the actual field exposure, as only few distorted loci were identified in the control series, in which the postzygotic possibility for selection was very limited (16 loci; Table 1).

We identified substantial differences among the hybrid generations, as well as differences among the three different types of markers employed. After 2 years of field exposure (BC1S1; 51 loci), the distortion was almost double compared with one year of field exposure (S1 and BC1, 22 and 22 loci, respectively). This increase is attributed to repeated selection in combination with an additional recombination possibility. For all but two SSAP loci, at LG5 and LG6 – CO9C-CTC-004 and CO9C-AAG-008 – this preference for one parental allele per locus was based on at least two fields being distorted in the same direction (Supporting information, Table S4). In general, there was no overall bias for one of the parental species for all the three marker types. Two exceptions were identified: in the S1 generation, distortion was significantly biased in the direction of the L. serriola alleles for AFLP markers. In contrast, for

---

**Figure 2** Segregation distortion and direction of linkage disequilibrium (LD) change. Three hybrid generations of 92 plants each are included (S1, BC1, and BC1, columns 1–3, respectively), for each of nine linkage groups and an anonymous group of loci. Parentage, linkage group, and map distance in cm are according to Syed et al. (2006). Differences in gray-shading among loci within columns indicate LD alterations, based on marginal LD compared with the S1, greenhouse-grown, control generation (x = 0.05, corrected for multiple testing following Benjamini and Hochberg 1995): black segments indicate increased LD, and hatched segments indicate lowered LD. The allelic bias (‘distortion’) is indicated next to the loci per column (x = 1/loci per generation), compared with the segregation as observed in the (extrapolated) control generation. Dark-colored loci contain significantly more crop (L. sativa) alleles than expected over 92 plants; light-colored loci contain significantly more wild relative (L. serriola) alleles. SSAP loci are prefixed with ‘CD’, NBS with ‘NBS’. ‘X’ represents missing data. Within the BC1 generation, loci either with a L. serriola band presence or dominantly scored provided no segregation and LD signal (underlined loci). For the ease of readability, distances smaller than 1.8 cm are not scaled.
SSAP markers in the BC1 generation, we found a significant preference for *L. sativa* (Table 1). However, for none of the marker types, after 2 years of exposure, did we identify a significant overall preference for one of the parents. This suggests that elements of both parents are favored depending on the putative fitness association at the specific location. Also, no overall preferential distortion signal was found in the NBS loci (Table 1). For NBS markers, the amount of distorted loci was not larger than in the control generation, except for the BC1, but with low numbers (2). Strikingly, no tendency in favor of heterozygotes was identified for all generations (Table 1).

**Pairwise linkage disequilibrium**

For over half of the adjacent pairwise combinations, LD was significantly altered and putatively influenced by the field exposure (Table 2). In approximately half or more of the pairwise combinations, the level of LD was affected in field-exposed generations, compared with the extrapolated baseline from the control generation (Table 2). This proportion was different among the generations (S1: 48; BC1: 69%, and BC1S1: 58%). For both selfing, field exposed, generations (S1 and BC1S1), the ratio of alteration between lowered and elevated LD was about 2:1, significantly different from random (P < 0.001; Table 2). The direction of alterations from the baseline per pairwise combination is indicated in Fig. 2, marked by different gray shading.

Many pairwise combinations are altered in the same direction (either lower or higher) for all generations (Table 3; Fig. 2). Very strongly significant, the two generations with an included selfing event (S1 and BC1S1) show that 44 combinations are altered in the same direction, whereas only six were not (P < 0.001; Table 3). Such
strong bias makes it very unlikely that these patterns are caused by nondirectional, i.e., random, variation.

Throughout the *Lactuca* genome, there are clear areas in which LD is consistently strengthened even for all generations like in parts of LG4 (36.9–49.1 cM), LG5 (78.5–84.6 and 110.3–111.0), and LG6 (11.7–40). The locus starting this segment at LG6 is biased in the direction of *L. serriola*, making it a potential favorable spot for transgene insertion. However, the opposite, a consistent loosening of LD among all three generations, is found on various places throughout the genome. Putatively, this is caused by opposing forces, selecting toward different parental directions. Obvious examples can be found in LG5 (46.0–66.0) and LG8 (64.2–72.1).

**Discussion**

Our study together with its related work on hybrid fitness in natural field conditions (Hooftman et al. 2005, 2007, 2009) follows the fate of parental genomes through several generations after hybridization between wild and cultivated lettuce. This larger framework included the construction of a genetic map for the cross used in this study (Syed et al. 2006). Our study is among the first to empirically test the genomic structure in hybrids between crops and wild relatives as affected by natural, i.e., nonartificial, field conditions, others being *Brassica* and *Helianthus* in Rose et al. (2009) and Baack et al. (2008). Genomic hotspots and cool areas for introgression can be identified, as performed below, through the identification of distorted regions in which linkage disequilibrium (LD) is consistently strengthened through multiple generations. Such regions indicate that specific recombinants are favored under these field conditions. We consider such information very useful for plant breeding. Through marker-assisted breeding or related techniques (Varshney and Dubey 2009), plants could be bred having transgene insertion sites in genomic regions providing the highest likelihood of being purged after escape into wild relative populations (Stewart et al. 2003; Sweet 2009).

**Allele frequency shifts**

For a large number of loci, we identified a clear change in allele frequencies in favor of the retention of alleles specific to one or the other parental taxon in the natural field. This included many loci for which crop alleles were in excess of expectations for band present states of both *L. sativa* and *L. serriola*. The nonuniform distribution of \( P \)-values, skewed to contain many high values in field-exposed generations, indicates that the applied correction for multiple tests (Benjamini and Hochberg 1995) provided conservative results rather than speculative ones. Our results contrast to a view quite generally held until few years ago that alleles originating from the crop were considered deleterious for survival in the wild, assuming a general loss of adaptive capacity for crop genes in wild habitats (Hails and Morley 2005). It appears that at least part of the alleles, which have been (co–) selected by breeders, may confer advantages upon hybrid plants under field exposure, depending on the type of environment. This is a conclusion that we share with recent work in sunflower (Baack et al. 2008; Chapman et al. 2008), rice (Cao et al. 2009), *Brassica* (Rose et al. 2009), and *Raphanus* (Snow et al. 2010). In our study, the driving force seems to be directional selection. The approximately 90% mortality rate per growing season postzygotically removes many less-fit hybrid lineages (Whitney et al. 2006). Consequently, positively selected crop genes will promote subsequent introgression of many less advantageous crop alleles, including transgenes, via ‘genetic hitchhiking’ (Barton 2000; Stewart et al. 2003; Mackay and Powell 2007).

**Differences between marker types**

Distortion – i.e., changes in allele frequencies from those expected under neutrality – showed a tendency to the selection for many wild-type alleles after one growing season of field exposure in the \( S \) generation for AFLP loci as well as a preference for crop alleles for SSAP loci in the \( BC \) generation. No preferences were seen in the \( BC_{1S1} \) generation. Whether there is any biological explanation behind these opposite directions remains to be identified with further research. SSAPs have recently been successfully employed (e.g., Cornman and Arnold 2009; Du et al. 2009). However, the drawbacks and caveats of SSAPs are not yet completely known. Therefore, it is hard to tell whether there is any contribution from a marker-type bias in favor of a specific category of linked trait, which could explain the difference between AFLPs and SSAPs between the two generations.

**Fitter *Lactuca* hybrids**

In earlier work, we identified these identical hybrid plants to be substantially fitter than the wild parent (relative growth rate = 3). This difference is mostly because of a larger survival capacity at the seedling stage (Hooftman et al. 2007). The fitness advantage gradually declined through multiple generations but was still present in the \( BC \) and \( S \) generations as well as in the \( BC_{1S1} \) plants used in this study (relative growth rate = 1.7). However, as a group, hybrids were not morphologically distinguishable from the wild parent (Hooftman et al. 2005). Therefore,
any crop genes connected to this fitness advantage seem not to be correlated with easily visible morphological characteristics. Furthermore, in light of the fitness advantage of these hybrid generations, it is remarkable that being heterozygous was hardly advantageous. Prior to this study, we expected to find a larger effect of heterosis after releasing these conspecific taxa from their inbreeding load by outcrossing (Johansen-Morris and Latta 2006). Accordingly, we now strongly suggest that additive interactions among genes rather than heterosis could be responsible for the increased fitness of these *Lactuca* hybrids compared with both parental species.

### Fixation and selective strength

LD among loci could have led to coselection of large numbers of linked genes among which only one or few genes actually confer a fitness benefit, the rest being effectively neutral (Burke et al. 2005). Such segments could start functioning as independent units when experiencing opposing selective forces with respect to introgression. As a result, neighboring segments will become more and more separated and introgression will appear to occur in chromosomal ‘units’ surrounding an important locus associated with plant fitness (Stewart et al. 2003). Identification of such quantitative trait loci (QTLs) in this lettuce cross is currently underway (e.g., Uwinama et al. 2010). The various areas we identified with less than expected LD could be examples of areas that are more prone to tear apart in opposite parental directions under nonagricultural conditions. Rapid fixation of whole genomic areas could be the result, providing selection pressures do not vary much. The evidence that fixation of large segments is a distinct possibility is provided through the signature of domestication patterns by agriculture (Ehrenreich and Purugganan 2007; Wills and Burke 2007; Yamasaki et al. 2007). By applying repeated similar bottlenecks, fixation is rapid. After selecting for hundreds of generations with a constant and fairly similar environmental pressure provided by agricultural conditions, LD could be still significant up to 10 cm, such as reported in barley (Kraakman et al. 2004).

### Hotspots and cold areas for introgression

We can tentatively identify genomic hotspots and cold areas for the introgression of crop alleles into wild generations (Stewart et al. 2003). The BC<sub>1S1</sub> generation shows the clearest results for this having been exposed to 2 years of selection pressure. Being speculative, few suggestions for a judicious choice of insertion of a new (trans)gene to lettuce could be onto parts of LG3 (middle and bottom) or LG4 (lower middle), where manifold distortion to containing *L. serriola* segments combined with strengthened LD occurs, with a smaller similar suitable area on LG6.
When derived from *L. sativa*, those parts provide in general a higher likelihood of being purged from the hybrid genome after initial hybridization. In contrast, LG5 (top and middle) and LG8 (upper middle) seem to be better avoided, favoring *L. sativa* genomic segments. A combination with one or more of the proposed other biosafety containment constructs (e.g., Gressel and Valverde 2009; Moon et al. 2010) would decrease the likelihood of introgression even more.

**Implications for future crop development and risk assessment**

Our results are not only applicable to this single underlying association study of Syed et al. (2006), but to lettuce breeding in general. The majority of employed AFLP markers are integrated in the larger map of Truco et al. (2007). That map is based on an Iceberg lettuce cultivar and a Californian *L. serriola* accession and has been a prime outcome of the Compositae Genome Project (http://compgenomics.ucdavis.edu/). Currently, this database contains already over 42,000 lettuce unigenes (McHale et al. 2009).

Gene flow from crops to wild relatives has occurred and will continue to occur (Ellstrand 2003; Chandler and Dunwell 2008). Subsequently, LD induced by selection may cause genetic hitchhiking, which could be an important factor facilitating the introgression of specific genes (Stewart et al. 2003; Chapman and Burke 2006). Modeling studies have already shown that segments with only limited fitness advantage in the hybrid plants can provide rapid fixation of introduced genes (Morjan and Rieseberg 2004; Hooftman et al. 2008). Where recombination rates between a transgene and beneficial QTLs are low, introgression of such genes is either facilitated or mitigated, depending on the trait association (Stewart et al. 2003; Sweet 2009).

As a negative message from our study, we conclude that genes, should they be undesirable in natural settings (such as certain transgenes), could introgress into a wild background without conferring a direct fitness benefit themselves. More positively, a judicious choice of transgene insertion sites may inhibit transgene persistence because of selection against the segment in a nonagricultural, natural, habitat. Our observations of these mechanisms could have clear implications for developing transgene containment strategies based on LD. The first step is an adequate assessment of the likelihood of (trans) gene introgression over the whole genome (Stewart et al. 2003). The way forward could be using QTL approaches and related association studies (Mackay and Powell 2007; Heffner et al. 2009), employing the vastly expanding number of genetic maps (Collard and MacKill 2008). Such information could then be combined with developing appropriate predictive modeling techniques (Meirmans et al. 2009; Sweet 2009).

**Acknowledgements**

We thank Peter van Tienderen, Gerard Oostermeijer, Eric Schranz, Thure Hauser, Rene Smulders, Jeroen Rouppe van der Voort, Gerard van der Linden, as well as all members of ANGEL for discussions. Two anonymous reviewers and the associate editor are acknowledged for their constructive comments improving the manuscript. We also thank the technical staffs at the University of Amsterdam (Mirjam Jacobs, Maaike de Jong), SCRI, Keygene (Dick Lensink & Rudie Antonise), and PRI (Wendy van’t Westende). This study was funded by EU-QLK3-2001-01657.

**Literature cited**


Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Histograms of P-value distributions.
Table S1. Calculations for locus based expected segregation ratios.
Table S2. Expected segregation per locus.
Table S3. Confidence intervals for distortion.
Table S4. P-values per field for the BC1S1 generation.
Data S1. Data and Matlab code used.

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