

## ORIGINAL ARTICLE

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

Danny A. P. Hooftman,<sup>1,2</sup> Andrew J. Flavell,<sup>3</sup> Hans Jansen,<sup>4</sup> Hans C. M. den Nijs,<sup>1</sup> Naeem H. Syed,<sup>3</sup> Anker P. Sørensen,<sup>5</sup> Pablo Orozco-ter Wengel<sup>1,5</sup> and Clemens C. M. van de Wiel<sup>6</sup>

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

## Keywords

crop breeding, gene flow, genetic modifications, hitchhiking, introgression, *Lactuca*, linkage disequilibrium, transgenic plants.

## Correspondence

Centre for Ecology and Hydrology, Benson Lane, Crowmarsh Gifford, Wallingford, OX10 8BB, UK.

Tel.: +44 (0)1491 838800;

fax: +44 (0)1491 692424;

e-mail: dann1@ceh.ac.uk

Received: 6 August 2010

Accepted: 28 March 2011

First published online: 27 April 2011

doi:10.1111/j.1752-4571.2011.00188.x

## Abstract

Gene escape from crops has gained much attention in the last two decades, as transgenes introgressing into wild populations could affect the latter's ecological characteristics. However, different genes have different likelihoods of introgression. The mixture of selective forces provided by natural conditions creates an adaptive mosaic of alleles from both parental species. We investigated segregation patterns after hybridization between lettuce (*Lactuca sativa*) and its wild relative, *L. serriola*. Three generations of hybrids ( $S_1$ ,  $BC_1$ , and  $BC_{1S1}$ ) were grown in habitats mimicking the wild parent's habitat. As control, we harvested  $S_1$  seedlings grown under controlled conditions, providing very limited possibility for selection. We used 89 AFLP loci, as well as more recently developed dominant markers, 115 retrotransposon markers (SSAP), and 28 NBS loci linked to resistance genes. For many loci, allele frequencies were biased in plants exposed to natural field conditions, including over-representation of crop alleles for various loci. Furthermore, Linkage disequilibrium was locally changed, allegedly by selection caused by the natural field conditions, providing ample opportunity for genetic hitchhiking. Our study indicates that when developing genetically modified crops, a judicious selection of insertion sites, based on knowledge of selective (dis)advantages of the surrounding crop genome under field conditions, could diminish transgene persistence.

## 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 (Schierenbeck 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 (Baack 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 ( $S_1$ ,  $BC_1$ , and  $BC_{1S1}$ ) 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 involucre bracts are less able to be released.

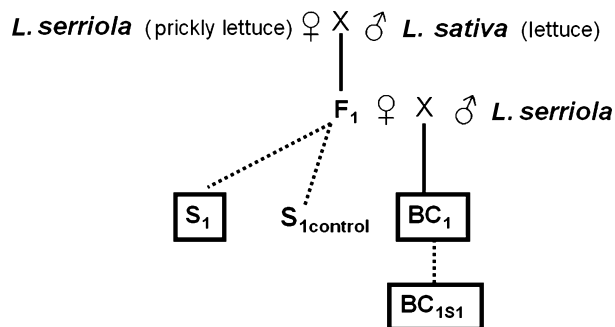
### Plant material and field experiments

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: F<sub>1</sub> 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 F<sub>1</sub> hybrids self-pollinate to obtain a generation of autogamous hybrid progeny (hereafter called S<sub>1</sub>). Backcrossing (BC<sub>1</sub>) was performed using F<sub>1</sub> hybrids maternally and *L. serriola* as pollen donors; doing so we mimic the most likely scenario of gene

escape into a wild population. Here, a single hybrid would be surrounded by wild plants leading to combined backcrossing and selfing.

We employed the genetic map of Syed et al. (2006) based on an S<sub>1</sub> 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 F<sub>1</sub> 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'), Sijbekarspel (Netherlands, N 52°42', E 05°00'), and Aachen (Germany, N 50°46', E 06°03'). At each location, two 100 m<sup>2</sup> plots were demarcated and ploughed, mimicking anthropogenic soil disturbance. In April 2003, seeds of *L. serriola*, *L. sativa*, and S<sub>1</sub>- and BC<sub>1</sub> 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 (BC<sub>1S1</sub>) of these BC<sub>1</sub> 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 BC<sub>1S1</sub> generation.



**Figure 1** 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 S<sub>1</sub> and BC<sub>1</sub> generations underwent one growing season of field exposure and the BC<sub>1S1</sub> generation two growing seasons. The control generation (S<sub>1control</sub>) underwent no field exposure.

### Sampling, molecular marker methods, and loci used

Leaf material was sampled in July 2003 (BC<sub>1</sub> and S<sub>1</sub>) and July 2004 (BC<sub>1S1</sub>) over the three experimental locations (92 plants generation<sup>-1</sup>). 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 S<sub>1</sub> 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 fitness. 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 dominantly). 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 homozygotes was performed with Xcelerator v.1.3.1. software package (proprietary software, Keygene N.V.). This program differentiates between those two classes depending on the intensity of the bands. To eliminate possible selfings that occurred during the BC<sub>1</sub> 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 BC<sub>1</sub> plant and one BC<sub>1S1</sub> plant were excluded because this analysis suggested 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 parental 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 BC<sub>1</sub> 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 individual 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 postzygotic selection. For this greenhouse-grown control (S<sub>1</sub>) 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 BC<sub>1</sub> and BC<sub>1S1</sub> 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; 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 implemented. In each simulation, 92 virtual diploid plants were created. In those plants, per locus, the state – homozygous 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 information: Table S1). For 17 loci with missing data in the control generation, we assumed neutral segregation. Subsequently, the obtained frequencies per locus of this sample of 92 plants were calculated. This procedure was repeated 50 000 times.

2 Confidence intervals for a given significance level ( $\alpha$ ) were obtained from the above obtained sampling distribution 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  $\alpha = (1/\# \text{ loci})$  per generation. To give an example, for 211 loci, the actual  $\alpha$  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 resulting  $\alpha$  values are the following: control: 0.0047 (for  $N = 213$ ); S<sub>1</sub>: 0.0045 ( $N = 220$ ); BC<sub>1</sub>: 0.0063 ( $N = 158$ ); and BC<sub>1S1</sub>: 0.0047 ( $N = 211$ ). The corresponding confidence 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 distributions

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).

$$LD_i = r_i = \frac{x_{11} - p_1q_1}{\sqrt{p_1p_2q_1q_2}} \quad (1)$$

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 ( $LD_i^M$ ) 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  $\alpha = 0.05$  as base value.

$$LD_i^M = \frac{LD_i^T}{\frac{1}{r} \times \sum_{i=1}^r \frac{LD_i^T}{LD_i^C}} \quad (2)$$

$$\chi^2 = \frac{(LD_i^M - LD_i^C)^2}{LD_i^C} + \frac{((100 - LD_i^M) - (100 - LD_i^C))^2}{(100 - LD_i^C)} \quad (3)$$

where  $LD_i^M$  = marginal LD for adjacent locus gap *i*;  $LD_i^T$  = observed LD in the tested generation for adjacent locus gap *i*;  $LD_i^C$  = 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). Included are three marker types of which AFLP were codominantly scored and NBS and SSAP dominantly. Four hybrid generations are included with 92 surviving plants generation<sup>-1</sup>: the control generation underwent no field exposure; the S<sub>1</sub> and BC<sub>1</sub> generations both underwent one growing season of field exposure; and the BC<sub>1S1</sub> generation two growing seasons.

Generation	Marker type/ total no. of loci	Distortion with prevalence for			
		<i>L. serriola</i> Allele	<i>L. sativa</i> allele	Hetero- zygotes	Directional <sup>†</sup>
AFLP					
Control	73	1	2	0	ns
S <sub>1</sub>	88	10	2	0	*
BC <sub>1</sub> <sup>‡</sup>	80 <sup>§</sup>	6		0	ns
BC <sub>1S1</sub>	72	6	6	0	ns
NBS					
Control	28	2	2		ns
S <sub>1</sub>	20 <sup>§</sup>	2	1		ns
BC <sub>1</sub> <sup>¶</sup>	9	1	1		ns
BC <sub>1S1</sub>	24 <sup>§</sup>	1	4		ns
SSAP					
Control	112	0	9		*
S <sub>1</sub>	112 <sup>§</sup>	5	2		ns
BC <sub>1</sub> <sup>¶</sup>	69	3	11		*
BC <sub>1S1</sub>	115	14	20		ns

<sup>†</sup>Significance (\**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.

<sup>‡</sup>Codominantly scored BC<sub>1</sub> loci (backcross with *L. serriola*) are either *L. serriola* homozygotes or heterozygotes.

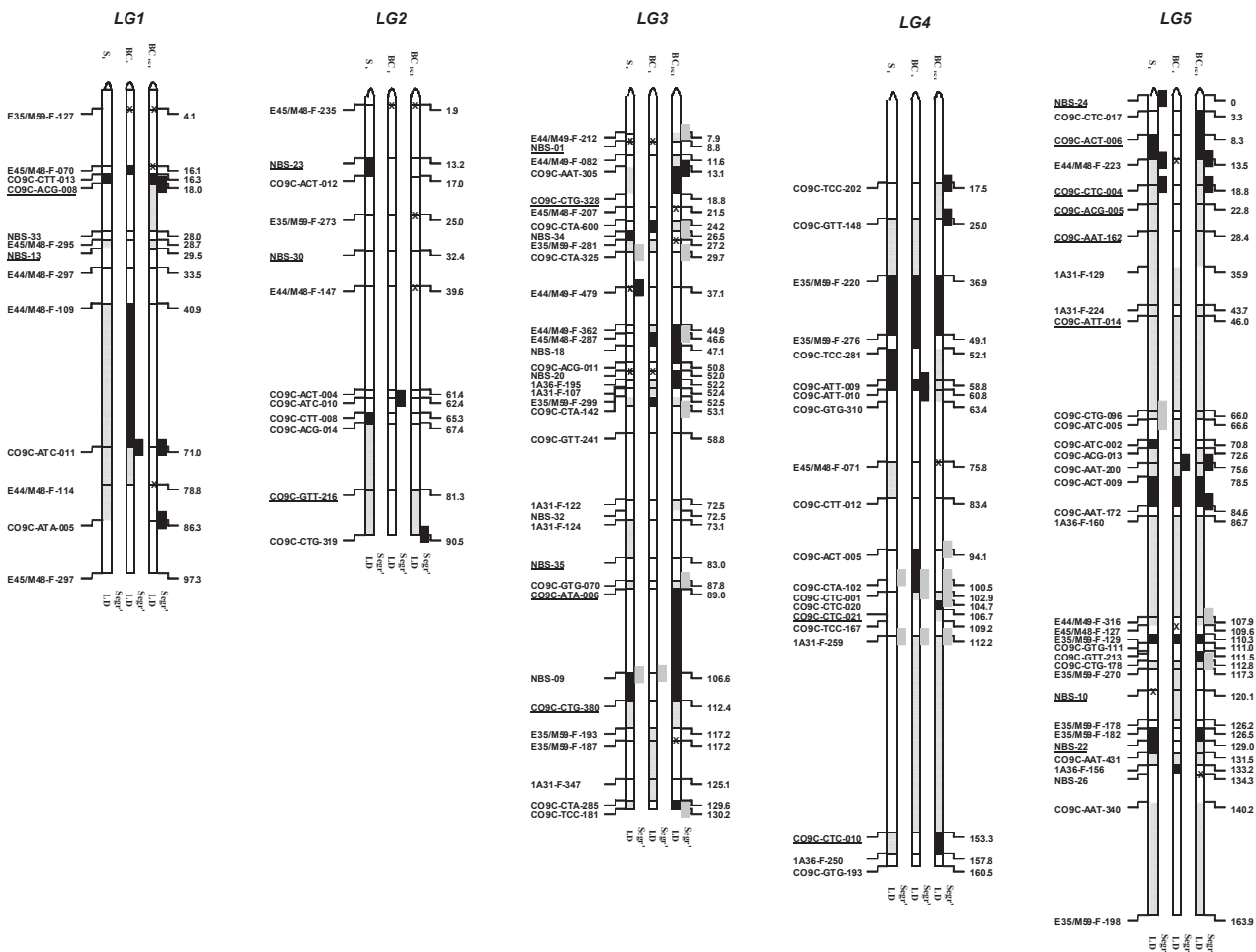
<sup>§</sup>Proportionally not more distorted than the control generation (*P* > 0.05; Chi-square, AFLP: df = 2; NBS/SSAP: df = 1).

<sup>¶</sup>*N* in the BC<sub>1</sub> 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 (BC<sub>1S1</sub>; 51 loci), the distortion was almost double compared with one year of field exposure (S<sub>1</sub>

and BC<sub>1</sub>, 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 S<sub>1</sub> 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 (S<sub>1</sub>, BC<sub>1</sub>, and BC<sub>1S1</sub>; columns 1–3, respectively), for each of nine linkage groups and an anonymous group of loci. Parentage, linkage group, and mapping 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 S<sub>1</sub>, greenhouse-grown, control generation ( $\alpha = 0.05$ , 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 ( $\alpha = 1/\text{loci per generation}$ ), 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 'CO'; NBS with 'NBS'; all other prefixes represent AFLP loci. 'X' represents missing data. Within the BC<sub>1</sub> 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.

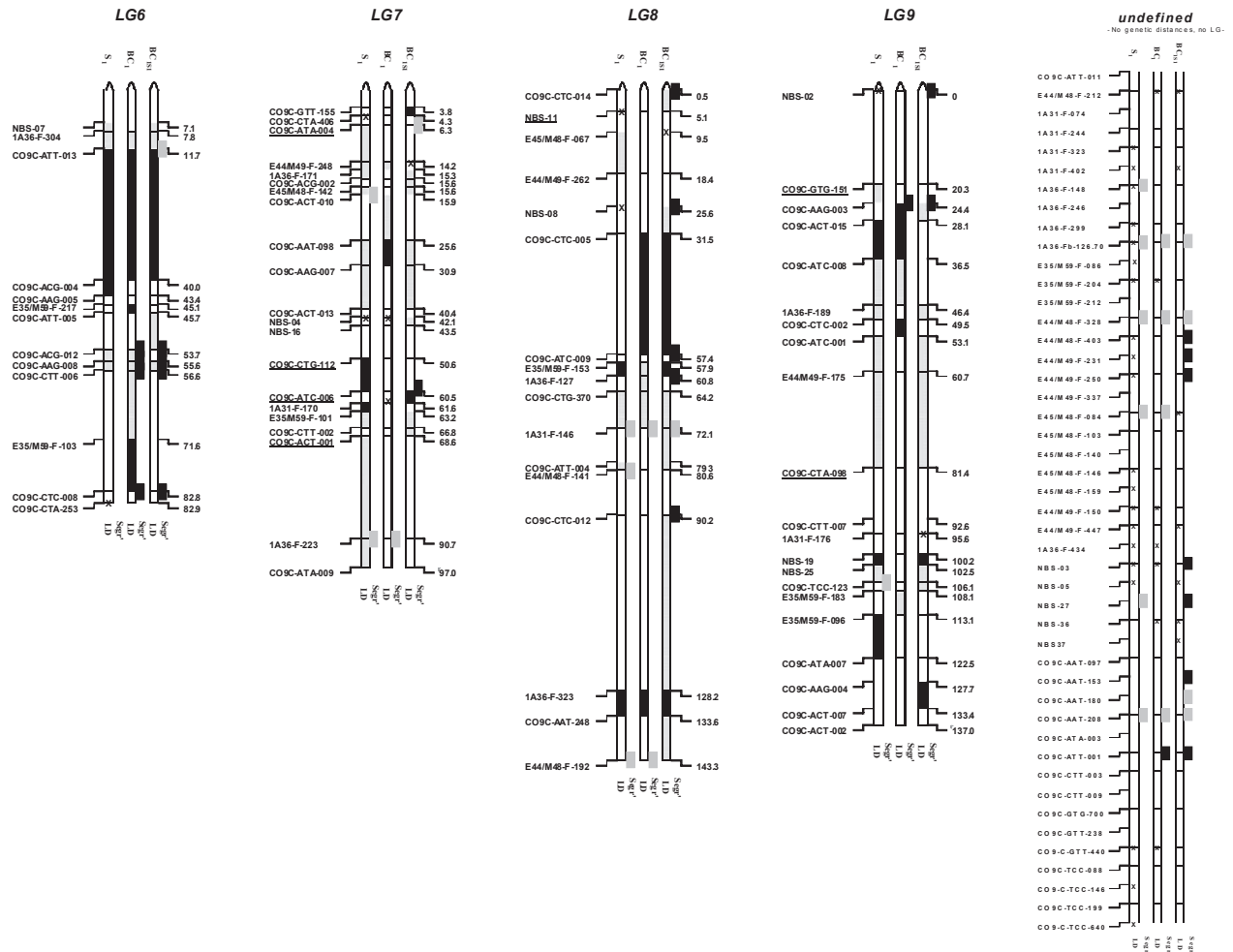


Figure 2b (Continued)

SSAP markers in the BC<sub>1</sub> 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 BC<sub>1</sub>, 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 (S<sub>1</sub>: 48; BC<sub>1</sub>: 69%, and BC<sub>1S1</sub>: 58%). For both selfing, field exposed, generations (S<sub>1</sub> and BC<sub>1S1</sub>), 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 (S<sub>1</sub> and BC<sub>1S1</sub>) 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) pro-

vided 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  $S_1$  generation for AFLP loci as well as a preference for crop alleles for SSAP loci in the  $BC_1$  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  $BC_3$  and  $S_4$  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,



**Table 2.** Direction of alteration of adjacent pairwise linkage disequilibrium (LD) under field exposure compared with a  $S_1$  control ( $\alpha = 0.05$ , corrected for multiple tests following Benjamini and Hochberg 1995). Comparisons are based on marginal, adjusted, values for the tested generations.

Generation	N	LD compared with ( $S_1$ -)control		
		Weakened (<)	Strengthened (>)	Unchanged
$S_1$	157	53***	23	81
$BC_1^\dagger$	77	31 <sup>ns</sup>	22	24
$BC_{1S1}$	145	57***	27	61

\*\*\*Significant ( $P < 0.001$ ) directional effect deviating from 1:1 weakened/strengthened (Chi-square,  $df = 1$ ).

<sup>†</sup>N in the  $BC_1$  generation is lower, as dominant loci with a dominant *L. serriola* allele will show all a band being either heterozygous or homozygous *L. serriola*.

<sup>ns</sup>No directional effect deviating from 1:1 weakened/ strengthened (Chi-square,  $df = 1$ )

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 effec-

tively 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  $BC_{1S1}$  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

**Table 3.** Among-generation comparison of the direction of adjacent pairwise Linkage disequilibrium alteration under field exposure compared with an  $S_1$  control ( $\alpha = 0.05$ , corrected for multiple tests following Benjamini and Hochberg 1995). Comparisons are based on marginal, adjusted, values for the tested generations. Three categories are provided: equal direction (=), opposite directions ( $\neq$ ), and both unchanged. N represents the total viable pairwise combinations available for among-generation comparison.

Gen'	$S_1$			Unchanged	$BC_1^\dagger$			Unchanged	$BC_{1S1}$			Unchanged
	N	=	$\neq$		N	=	$\neq$		N	=	$\neq$	
$S_1$					75***	22	4	17				
$BC_1^\dagger$									63 <sup>ns</sup>	17	8	8
$BC_{1S1}$	131***	44	6	36								

\*\*\*Significant ( $P < 0.001$ ) directional effect deviating from 1:1 equal/ opposite (Chi-square,  $df = 1$ ).

<sup>†</sup>N in the  $BC_1$  generation is lower, as dominant loci with a dominant *L. serriola* allele will show all a band being either heterozygous or homozygous *L. serriola*.

<sup>ns</sup>No directional effect deviating from 1:1 equal/opposite (Chi-square,  $df = 1$ ).

(top). 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, Maaik 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

- Ashraf, M. 2010. Inducing drought tolerances in plants: recent advances. *Biotechnology Advances* **28**:169–183.
- Baack, E. J., Y. Sapir, M. A. Chapman, J. M. Burke, and L. H. Rieseberg. 2008. Selection on domestication traits and quantitative trait loci in crop–wild sunflower hybrids. *Molecular Ecology* **17**:666–677.
- Barton, N. H. 2000. Genetic Hitchhiking. *Philosophical Transactions of the Royal Society B* **355**:1553–1562.
- Beddington, J. 2010. Food security: contributions from science to a new and greener revolution. *Philosophical Transactions of the Royal Society B* **365**:61–71.
- Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* **57**:289–300.
- Bhatnagar-Mathur, P., V. Vadez, and K. K. Sharma. 2008. Transgenic approaches for abiotic tolerance in plants: retrospect and prospects. *Plant Cell Reports* **27**:411–424.
- Burke, J. M., and M. L. Arnold. 2001. Genetics and fitness of hybrids. *Annual Review in Genetics* **35**:31–52.
- Burke, J. M., S. J. Knapp, and L. H. Rieseberg. 2005. Genetic consequences of selection during the evolution of cultivated sunflower. *Genetics* **171**:1933–1940.
- Campbell, L. G., A. A. Snow, P. M. Sweeney, and J. M. Ketner. 2009. Rapid evolution in crop–weed hybrids under artificial selection for divergent life histories. *Evolutionary Applications* **2**:172–186.
- Cao, Q. J., H. Xia, X. Yang, and B. R. Lu. 2009. Performance of hybrids between weedy rice and insect resistant transgenic rice under field experiments: implication for environmental risk assessment. *Journal of Integrative Plant Biology* **51**:1138–1148.
- Carter, R. N., and S. D. Prince. 1985. The geographical distribution of Prickly Lettuce (*Lactuca serriola*): a general survey of its habitat and performance in Britain. *Journal of Ecology* **73**:27–38.
- Chandler, S., and J. M. Dunwell. 2008. Gene flow, risk assessment and the environmental release of transgenic plants. *Critical Reviews in Plant Sciences* **27**:25–49.
- Chapman, M. A., and J. M. Burke. 2006. Letting the gene out of the bottle: the population genetics of genetically modified crops. *New Phytologist* **170**:429–443.

- Chapman, M. A., C. H. Pashley, J. Wenzler, J. Hvala, S. X. Tang, S. J. Knapp, and J. M. Burke. 2008. A genomic scan reveals candidates for genes involved in the evolution of cultivated sunflower. *Plant Cell* **20**:2931–2945.
- Collard, B. C. Y., and D. J. MacKill. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B* **363**:557–572.
- Cornman, R. S., and M. L. Arnold. 2009. Characterization and comparative analysis of sequence-specific amplified polymorphisms based on two subfamilies of IRRE retrotransposons in *Iris missouriensis* (Iridaceae). *Genetica* **135**:25–38.
- D'Andrea, L. 2006. Historical and Molecular Biogeography of Prickly Lettuce (*L. serriola* L.) and Evaluation of Hybridisation with the Crop (*L. sativa* L.). PhD thesis, University of Neuchâtel, Switzerland.
- Du, X. Y., Z. R. Zhang, and Z. R. Luo. 2009. Comparison of four molecular markers for genetic analysis in *Diospyros* L. (Ebenaceae). *Plant Systematics and Evolution* **281**:171–181.
- Ehrenreich, I. M., and M. D. Purugganan. 2007. The molecular genetic basis of plant adaptation. *American Journal of Botany* **93**:953–962.
- Ellstrand, N. C. 2003. Dangerous Liaisons? When Cultivated Plants Mate with their Wild Relatives. John Hopkins University press, Baltimore, Maryland.
- Flint-Garcia, S. A., J. M. Thornsberry, and E. S. Buckler. 2003. Structure of Linkage Disequilibrium in plants. *Annual Reviews of Plant Biology* **54**:357–374.
- Giannino, D., C. Nicolodi, G. Testone, E. Di Giacomo, M. A. Iannelli, G. Frugis, and D. Mariotti. 2008. Pollen mediated transgene flow in lettuce (*Lactuca sativa* L.). *Plant Breeding* **127**:308–314.
- Gressel, J., and B. E. Valverde. 2009. A strategy to provide long-term control of weedy rice while mitigating herbicide resistance transgene flow, and its potential use for other crops with related weeds. *Pest Management Science* **65**:723–731.
- Hails, R. S., and K. Morley. 2005. Genes invading new generations: a risk assessment perspective. *Trends in Ecology and Evolution* **20**:245–252.
- Heffner, E. L., M. E. Sorrells, and J. L. Jannink. 2009. Genomic selection for crop improvement. *Crop Science* **49**:1–12.
- Hegde, S. G., J. D. Nason, J. M. Clegg, and N. C. Ellstrand. 2006. The evolution of California's wild radish has resulted in the extinction of its progenitors. *Evolution* **60**:1187–1197.
- Hooftman, D. A. P., J. G. B. Oostermeijer, M. M. J. Jacobs, and J. C. M. den Nijs. 2005. Vital rates determine the performance advantage of crop-wild hybrids in Lettuce. *Journal of Applied Ecology* **42**:1086–1095.
- Hooftman, D. A. P., J. G. B. Oostermeijer, and J. C. M. den Nijs. 2006. Invasive behavior of *Lactuca serriola* (Asteraceae) in The Netherlands: spatial distribution and ecological amplitude. *Basic and Applied Ecology* **7**:507–519.
- Hooftman, D. A. P., M. J. de Jong, J. G. B. Oostermeijer, and J. C. M. den Nijs. 2007. Modelling the long-term consequences of crop-wild relative hybridization: a case study using four generations of hybrids. *Journal of Applied Ecology* **44**:1035–1045.
- Hooftman, D. A. P., J. G. B. Oostermeijer, E. Marquard, and J. C. M. den Nijs. 2008. Modelling the consequences of crop-wild relative gene flow: a sensitivity analysis of the effects of outcrossing rates and hybrid vigour breakdown in *Lactuca*. *Journal of Applied Ecology* **45**:1094–1103.
- Hooftman, D. A. P., Y. Hartman, J. G. B. Oostermeijer, and J. C. M. den Nijs. 2009. Existence of Vigorous Lineages of Crop-Wild Hybrids in Lettuce under Field Conditions. *Environmental Biosafety Research* **8**:203–217.
- Johansen-Morris, A. D., and R. G. Latta. 2006. Fitness consequences of hybridization between ecotypes of *Avena barbata*: hybrid breakdown, hybrid vigor, and transgressive segregation. *Evolution* **60**:1585–1595.
- Koopman, W. J. M., M. J. Zevenbergen, and R. G. van der Berg. 2001. Species relationships in *Lactuca* s.l. (Lactuceae, Asteraceae) inferred from AFLP fingerprints. *American Journal of Botany* **88**:1881–1887.
- Kraakman, A. T. W., R. E. Niks, P. M. M. van den Berg, P. Stam, and F. A. van Eeuwijk. 2004. Linkage disequilibrium of yield and yield stability in modern spring barley cultivars. *Genetics* **168**:435–446.
- Lebeda, A., I. Doležalová, E. Krístková, and P. Mieslerová. 2001. Biodiversity and ecogeography of wild *Lactuca* spp. in some European countries. *Genetic Resources and Crop Evolution* **48**:153–164.
- Mackay, I., and W. Powell. 2007. Methods for linkage disequilibrium mapping in crops. *Trends in Plant Science* **12**:57–63.
- Martin, N. H., A. C. Bouck, and M. L. Arnold. 2005. Loci affecting long-term hybrid survivorship in Louisiana Irises: implication for reproductive isolation and introgression. *Evolution* **59**:2116–2124.
- McHale, L. K., M. Truco, A. Kozik, T. Wroblewski, O. E. Ochoa, K. A. Lahre, S. J. Knapp et al. 2009. The genomic architecture of disease resistance in lettuce. *Theoretical and Applied Genetics* **118**:565–580.
- Meirmans, P. G., J. Bousquet, and N. Isabel. 2009. A metapopulation model for the introgression from genetically modified plants into their wild relatives. *Evolutionary Applications* **2**:160–171.
- Milne, R. I., and R. J. Abbott. 2000. Origin and evolution of invasive naturalised material of *Rhododendron ponticum* L. in the British Isles. *Molecular Ecology* **9**:541–556.
- Moon, H. S., Y. Li, and C. N. Stewart. 2010. Keeping the genie in the bottle: transgene containments by excision in pollen. *Trends in Biotechnology* **28**:3–8.
- Morjan, C. L., and L. H. Rieseberg. 2004. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology* **13**:1341–1356.
- Nagata, R. T. 1992. Clip-and-wash method of emasculating for lettuce. *Hortiscience* **27**:907–908.
- Ridley, C. E., S. C. Kim, and N. C. Ellstrand. 2008. Bidirectional history of hybridization in California wild radish, *Raphanus sativus* (Brassicaceae), as revealed by chloroplast DNA. *American Journal of Botany* **95**:1437–1442.
- Rose, C. W., R. J. Millwood, H. S. Moon, M. R. Rao, M. D. Halfhill, P. L. Raymer, S. I. Warwick et al. 2009. Genetic load and transgenic mitigating genes in transgenic *Brassica rapa* (field mustard) x *Brassica napus* (oilseed rape) hybrid populations. *BMC Biotechnology* **9**:93.
- Schierenbeck, K. A., and N. C. Ellstrand. 2009. Hybridization and the evolution of invasiveness in plants and other organisms. *Biological Invasions* **11**:1093–1105.
- Snow, A. A., D. A. Andow, P. Gepts, E. M. Hallerman, A. Power, J. M. Tiedje, and L. L. Wolfenbarger. 2005. Genetically engineered organisms and the environment: current status and recommendations. *Ecological Applications* **15**:377–404.
- Snow, A. A., T. M. Culley, L. G. Campbell, P. M. Sweeney, S. G. Hegde, and N. C. Ellstrand. 2010. Long-term persistence of crop alleles in weedy populations of wild radish (*Raphanus raphanistrum*). *New Phytologist* **186**:537–548.
- Stewart, C. N., M. D. Halfhill, and S. I. Warwick. 2003. Transgene introgression from genetically modified crops to their wild relatives. *Nature Reviews Genetics* **4**:806–817.

- Sweet, J. B. 2009. The 10th International Symposium on the Biosafety of Genetically Modified Organisms (ISBGMO), Wellington, New Zealand, November 2008. *Environmental Biosafety Research* **8**:161–181.
- Syed, N. H., S. Suresh Sundar, M. J. Wilkinson, B. S. Bhau, J. J. V. Cavalcanti, and A. J. Flavell. 2005. Ty1-copia retrotransposon-based SSAP marker development in Cashew (*Anacardium occidentale* L.). *Theoretical and Applied Genetics* **110**:1195–1202.
- Syed, N. H., A. P. Sørensen, R. Antonise, C. van de Wiel, C. G. van der Linden, W. van't Westende, D. A. P. Hooftman *et al.* 2006. A detailed linkage map of Lettuce based on SSAP, AFLP and NBS markers. *Theoretical and Applied Genetics* **112**:517–527.
- Tanaka, Y. 2007. Introgressive hybridization as the breakdown of postzygotic isolation: a theoretical perspective. *Ecological Research* **22**:929–939.
- Truco, M. J., R. Antonisse, D. Lavelle, O. Ochoa, A. Kozik, H. Witsenboer, S. B. Fort *et al.* 2007. A high-density, integrated genetic linkage map of Lettuce (*Lactuca* spp.). *Theoretical and Applied Genetics* **115**:735–746.
- Tutin, T. G., V. H. Heywood, N. A. Burges, D. M. Moore, D. H. Valentine, S. M. Walters, and D. A. Webb. 1976. *Flora Europaea*, vol. 4. Cambridge University Press, Cambridge, UK.
- Uwinama, B., C. van der Wiel, R. Smulders, and R. Visser. 2010. A genetic analysis of the introgression process from crops to wild relatives under abiotic stress conditions: the case of Lettuce. Proceedings of the 11th ISBGMO conference. Buenos Aires, Argentina.
- Van der Linden, C. G., D. C. A. E. Outers, V. Mahaska, E. Z. Kochieva, M. J. M. Smulders, and B. Vosman. 2004. Efficient targeting of plant disease resistance loci using NBS profiling. *Theoretical and Applied Genetics* **109**:384–393.
- Varshney, R. K., and A. Dubey. 2009. Novel genomic tools and modern genetic and breeding approaches for crop improvement. *Journal of Biochemistry and Biotechnology* **18**:127–138.
- Whitney, K. D., R. A. Randell, and L. H. Rieseberg. 2006. Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. *The American Naturalist* **167**:794–807.
- Wills, D. M., and J. M. Burke. 2007. Quantitative Trait Locus analysis of the early domestication of sunflower. *Genetics* **176**:2589–2599.
- Yamasaki, M., S. I. Wright, and M. D. McMullen. 2007. Genomic screening for artificial selection during domestication and improvement in Maize. *Annals of Botany* **100**:967–973.

## 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 BC<sub>1S1</sub> 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.