



UvA-DARE (Digital Academic Repository)

Genomic regions under selection in crop-wild hybrids of lettuce: implications for crop breeding and environmental risk assessment

Hartman, Y.

[Link to publication](#)

Citation for published version (APA):

Hartman, Y. (2012). *Genomic regions under selection in crop-wild hybrids of lettuce: implications for crop breeding and environmental risk assessment.*

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: <https://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.

6

General discussion

In the debate surrounding the potential ecological risks of genetically modified (GM) organisms, the focus has shifted from research on gene flow to hybrid fitness and on methods to minimize transgene escape. This shift is mainly caused by the realization that many crops hybridize with their wild relatives; even species that were previously thought to be of low risk (Ellstrand 2003; Armstrong et al. 2005; Kwit et al. 2011). Moreover, the spread of conventional crop alleles as well as transgenes from commercially grown crops to the wild has in recent years been molecularly proven (Kwit et al. 2011; Ellstrand in press; Uwimana et al. 2012a), for example, for a transgene for herbicide resistance in oilseed rape (Warwick et al. 2008). Therefore, the most important factors governing the spread of transgenes after a hybridization event are the fitness of a transgene and natural selection acting upon (Chapman and Burke 2006).

Currently, Environmental Risk Assessment (ERA) procedures are performed on a case-by-case basis and the approval of new transgenic crops is very time-consuming and stringent, in particular in Europe (EFSA 2011). It is difficult to generate general protocols or guidelines, since data available to evaluate the potential of transgenes to increase invasiveness and/or weediness are still scarce (Warwick et al. 2009). Moreover, there is an overall lack of information regarding genes or genomic blocks under selection in the field (Hails and Morley 2005). For ERA, it seems to be a need to know in which crop–wild systems there are regions in the crop genome that are more or are less likely to introgress. In addition, it is important to know to what extent genomic selection patterns can be generalized across different cultivars and different environments (EFSA 2011).

In this thesis, we focused on the identification of crop genomic regions under selection to investigate the likelihood of introgression of crop alleles to wild relatives and the applicability of Transgene Mitigation (TM) strategies. In TM strategies, transgenes are placed in close linkage to regions or genes that confer a lower fitness to crop–wild hybrids in the wild and therefore would be less likely to spread into a wild population (Chapter 1; Gressel 1999; Stewart et al. 2003). We performed Quantitative trait loci (QTL) analyses on fitness and fitness-related traits measured under general conditions and under specific abiotic and competition stresses, in a series of greenhouse and field experiments. We used Recombinant Inbred Lines (RILs) selfed for nine generations of a cross between a crop *Lactuca sativa* cv. Salinas (Crisphead) and a Californian wild relative *L. serriola* (UC96US23; Johnson et al. 2000; Argyris et al. 2005; Zhang et al. 2007) in all greenhouse and field experiments. In the field fitness experiment, we also employed Backcross lines selfed for one generation (BC_1S_1) of a cross between a crop *L. sativa* cv. Dynamite (Butterhead) and a Dutch wild relative *L. serriola* (Eys). To our knowledge, only a few field studies on crop–wild hybrids have used QTL analysis to identify crop genomic regions under selection (Baack et al. 2008; Dechaine et al. 2009). In the following paragraphs, we discuss the implications of our results for the main research questions posed in Chapter 1.

Do crop genes (*Lactuca sativa*) confer a selective advantage or disadvantage under controlled and field conditions?

Crop genes are partly advantageous. Our results show that in lettuce hybrids crop genes conferred a selective advantage for a substantial amount of QTL and traits in controlled, as well as field conditions. For example, in greenhouse conditions crop alleles induced higher germination rates, more branches, and capitula that might be beneficial in the wild habitat (Chapter 2). In addition, QTL results indicated that the crop genomic background increased hybrid vigor, indicated by an increase in biomass, under drought, nutrient limitation, and salt stress (Uwimana et al. 2012b; Chapter 4). Similarly, under selective conditions in the field, crop alleles conferred the selective advantage for a number of fitness-related traits at certain genomic locations, including for the main fitness trait Seeds Produced per Seed Sown (SPSS).

However, for QTL at other locations it was the wild allele that conferred an advantage in the hybrids (Chapter 3 and 5).

Our results and other experimental and molecular work on lettuce hybridization, demonstrate that crop domestication genes are not necessarily bad for performance of hybrid plants under natural conditions. In a series of earlier field experiments with lettuce, it was demonstrated that at least four generations of hybrids on average had higher germination and survival rates than the wild parent (Hooftman et al. 2005, 2007, 2009). Moreover, allele frequencies were skewed in the direction of the crop allele for various loci, suggesting that crop alleles can indeed confer selective advantage in specific genomic regions (Hooftman et al. 2011). The fact that some crop genes are advantageous suggests that gene flow from crops to wild species could lead to permanent introgression of various crop genes in the wild species. This is substantiated by a recent molecular study that indicated that a substantial part of wild *L. serriola* plants in Europe (7%) were offspring of hybridization events between *L. serriola* and *L. sativa* (Uwimana et al. 2012a).

The fact that the crop genomic background contributes positively to hybrid fitness in the wild habitat has also been shown in other species, notably sunflower (Mercer et al. 2007; Baack et al. 2008) and radish (Snow et al. 2010). This contrasts to the long standing assumption that crop genes do not spread readily into the wild because domestication genes confer some selective disadvantage to the hybrid individuals, since they were originally selected for a good performance in an agricultural setting (Stewart et al. 2003). Indeed, the underlying crop alleles of certain domestication traits are likely to have a negative impact on hybrid fitness, such as in lettuce a decrease in shattering of seeds, loss of spines, and a delay in flowering compared to the wild relative (de Vries 1997). However, crops are also often selected for higher germination rates and a higher seed production that might also be beneficial in the wild habitat; hence, changing the life history of the natural species to some degree. Our results indeed suggest that in lettuce parts of the crop genomic background contributes to higher hybrid fitness and, therefore, potentially increases the likelihood of transfer of putative transgenes to the wild population.

How important are linkage drag (hitchhiking effects) for the potential spread of transgenes?

Linkage drag could be important indeed. After a single hybridization event, several processes play a role: heterosis, epistasis, transgressive segregation, and hitchhiking effects due to linkage drag all interact to determine hybrid fitness (Stewart et al. 2003; Johansen-Morris and Latta 2006), and so influence the introgression chances of crop (trans)genes. Introgression of crop genes into a recipient population starts with F_1 hybrids, with equal amounts of crop–wild genomes, and genome-wide heterozygosity and linkage disequilibrium (LD). In a selfing species, such as lettuce, this initial LD in first generation hybrids decays slowly and selection will effectively act on large genomic blocks rather than on individual loci (Flint-Garcia et al. 2003).

Our results show that the probability for a putative transgene spreading from a crop into the wild population indeed depends strongly on the insertion location. In the field fitness experiments, both the BC and RILs had two genomic regions with fitness QTL that were consistent across both sites. For RILs, we identified a region on linkage group (LG) 5 (bottom) where the crop allele consistently conferred a selective advantage, whereas on LG7 (top) a region conferred a selective disadvantage, mainly through delaying flowering (Chapter 3). For BC lines, we also detected the fitness region on LG7, whereas we identified a different second region on LG6 (bottom). In this case, the wild allele conferred the selective advantage at both

genomic locations. These results indicate that introgression chances of crop alleles extrapolated from the genetic location might differ between crosses, because of the different genetic makeup of the different parental lines (Mercer et al. 2006; Muraya et al. 2012). Moreover, as a result of these genomic selection patterns some crop alleles and/or regions have a higher chance to introgress, while other crop alleles remain rare or do not introgress at all. Several empirical studies recently showed that the speed of introgression indeed can vary for different crop genomic regions (Scascitelli et al. 2010; Snow et al. 2010; Hooftman et al. 2011).

At the same location as the before mentioned fitness QTL on LG7, a QTL cluster was detected in the greenhouse competition experiment including shoot dry weight QTL for control, drought, and salt treatments (Chapter 4), as well as several QTL connected to the speed of development (Chapter 2). We found it was the wild allele (*L. serriola*) that increased aboveground biomass and proportion dry weight values under competition stress and that the wild allele induced early bolting and flowering. In addition, the genetic correlations indicated that the greenhouse competition treatments had the highest correlation with the field stress treatments. This indicates that competition rather than a specific abiotic stress was an important influential factor for field fitness and that the wild species, *L. serriola*, might be a better competitor compared to the cultivated varieties (Chapter 4). This suggest that in situations with high plant density and a seasonal flowering period, lettuce hybrids with a crop genomic background at LG7 probably have a higher likelihood to be outcompeted by their wild relatives and die before reproduction. Such would explain the observed patterns in earlier full life cycle experiments with lettuce (Hooftman et al. 2005, 2007).

Can small-scale contained experiments with transgenes be used to assess potential ecological consequences?

Using small-scale experiments is challenging. We show considerable differences between QTL patterns of controlled greenhouse and field experiments; these comparisons were done for fitness-related traits under general conditions and for biomass under abiotic and competition stress with RILs. For example, greenhouse QTL for fitness-related traits showed almost no overlap with those detected in the field fitness experiment, except for traits such as days to first flower, shoot number, and seed output (Chapter 3). In addition, not one of the abiotic stress QTL detected in the greenhouse coincided with any of the field stress QTL (Chapter 4). One of the most important differences between greenhouse and field QTL patterns involved a region with a fitness QTL on LG5 where crop alleles were positively selected in the field that was previously not identified in the greenhouse (Chapter 3). This fitness region, where crop alleles have a higher chance of introgressing into the wild population, would have gone undetected with only small-scale controlled greenhouse experiments. Similar major differences, as we have described here, have been found in other studies when QTL patterns between greenhouse and field were compared (Weinig et al. 2002; Malmberg et al. 2005; Martin et al. 2006; Gardner and Latta 2008).

We see a number of reasons for the large differences between greenhouse and field QTL patterns. First, heritability values are lower in the field due to an increase in environmental variance (Latta et al. 2007; Gardner and Latta 2008). This could affect the threshold at which QTL are statistically detectable, although sample sizes used in the field were higher to compensate for this. For the majority of traits and treatments family-mean heritability values were high, indicating a high genetic component underlying the variation and so a good ability to locate QTL (Hyne et al. 1995).

Second, unavoidable differences in experimental set-up may have caused variation in the QTL results, making generalizations of results difficult. This is especially the case for

the detection of abiotic stress QTL, where many genes, proteins, and metabolic pathways are involved (Knight and Knight 2001; Roy et al. 2011) and, consequently, many different genomic regions causing small to medium-sized effects (Chapter 4). Therefore, small differences in experimental set-up or differences in plant age and initial growing conditions could already cause different expression patterns in response to stress (Collins et al. 2008). Genetic correlations between abiotic stress treatments indeed indicated that genetic correlations across treatments were often higher among different stress treatments within the same experiment (greenhouse or field), than among the same type of stress applied in different experiments (Chapter 4).

Third, and most important, selection pressures on different traits might differ between greenhouse and field (Weinig et al. 2002; Martin et al. 2006), and this can subsequently cause changes in which alleles are selected for in different environments (Mercer et al. 2007). Such Genotype \times Environment ($G \times E$) and QTL by Environment interactions have been observed in many species, such as *Arabidopsis thaliana* (Weinig et al. 2002), slender wild oat (Gardner and Latta 2006; Johansen-Morris and Latta 2008), and Iris hybrids (Martin et al. 2006). Through mortality, if hybrid variability interacts with the environmental and seasonal variation in the field (Weinig et al. 2003; Hails and Morley 2005), different lineages will be selected in the field compared to greenhouse conditions (Campbell et al. 2009; Hooftman et al. 2009). Moreover, fitness distributions of both RIL and BC hybrid populations indicated that introgression of crop alleles through hybridization might be more likely to occur in novel habitats as opposed to the natural wild habitat of the wild parent (Chapter 5). A greenhouse environment might also provide growing conditions that lead to an overestimation of hybrid fitness compared to the original wild habitat (Johansen-Morris and Latta 2008).

Implications for Environmental Risk Assessment and crop breeding

Our results will have important consequences for the methodology of Environmental Risk Assessment (ERA) and for the applicability of Transgene Mitigation (TM) strategies. We will first discuss the implications for ERA and end with our recommendations for TM strategies.

Most research on transgenes is conducted in short-term controlled greenhouse environments or under agricultural field conditions (Hails and Morley 2005; Vinocur and Altman 2005), testing for the effectiveness of the transgene and to predict the potential effects of transgene escape in either. However, we have shown that there was a low correspondence between field QTL and those detected in the greenhouse, suggesting that greenhouse studies are likely unreliable indicators of field fitness (Chapter 3 and 4). In addition, results can be cultivar specific (Chapter 5) and selection pressure can differ across time and place (Mercer et al. 2007). Hence, ideally fitness studies for ERA should be conducted with selective, competitive conditions including a range of wild parents (Muraya et al. 2012), several locations, and multiple years (Hails and Morley 2005). Clearly, as we experienced ourselves, this involves field experiments that require a huge amount of time and labor, and clear guidelines are needed when this is really called for, given the potential likelihood of introgression and associated hazard. Finally, including specific abiotic stresses in field fitness experiments is not always informative: the expression patterns of abiotic stress QTL was poorly generalisable across different stress experiments (Chapter 4), despite the fact that the mean growth of plant was reduced by the applied stress.

On a more positive note, these experiments can pinpoint specific genomic regions with similar selection effects in various field environments by following plants through their entire life cycle, as we and others have shown (Baack et al. 2008; Dechaine et al. 2009; Chapter 3 and 5). This is particularly important because, in spite of a lower average hybrid fitness, there is a high chance that particular hybrid genotypes will outperform the parental lines under certain

Chapter 6

environmental conditions (Burke and Arnold 2001; Johansen-Morris and Latta 2008; Hooftman et al. 2009; Chapter 5).

Results on genomic selection patterns can be used in simulation models predicting the fate of transgenes after hybridization events and in TM strategies. The basis for such TM strategy is the fact that after a hybridization event the chances for introgression of transgenes into a wild relative depend mainly on the fitness of initial hybrids and the fitness effect of transgenes in the wild genomic background (Ellstrand 2003). If a transgene is placed in close linkage with a gene or genomic block that causes a lower fitness in the wild habitat compared to the wild relative, it is more likely to be purged from the wild population (Gressel 1999; Stewart et al. 2003). This has been tested in transgenic tobacco and oilseed rape hybrids by placing the transgene in linkage with a dwarfing gene, leading to a dramatic reduction in survival capacity of hybrids carrying the dwarfing gene especially under competition stress (Al-Ahmad et al. 2005; Rose et al. 2009). Techniques for targeted insertion of transgenes into specific regions in the genome are currently being developed (Lombardo et al. 2011; Nandy and Srivastava 2011).

Our results and those of various studies on radish (Campbell and Snow 2007; Campbell et al. 2009; Snow et al. 2010) indicate that for leafy vegetables, selected for a delay in bolting and flowering, genomic regions coding for delay in or even prevention of flowering might be good candidates for such TM strategies (Gressel 1999). These traits are likely deleterious in the wild habitat, especially in situations with high density competition and a seasonal flowering period. We identified such a genomic region on LG7, where the crop allele conferred a selective disadvantage through delayed flowering, lower survival, and a lower fitness (Chapter 3 and 5), and possibly through a lower competitive ability (Chapter 4). In this region, effects were stable across cultivars and the environments of our field experiments and could therefore be used in TM strategies. It should be noted that even tight linkage of a transgene to a negative block might not always prevent transgene escape when opportunity for hybridization is high (Linder et al. 1998). As time passes, linkage disequilibrium (LD) breaks up due to recombination events and thus separating the transgene from the negatively selected block (Lee and Natesan 2006). However, before LD is broken, a strong directional selection could already have led to the rapid purging of deleterious genomic blocks via lineage sorting, just as it can lead to the rapid spread of favorable alleles (Rieseberg et al. 2002; Morjan and Rieseberg 2004) during selective sweeps. Under stressful conditions this selection can happen in a few generations (Campbell et al. 2009; Fakheran et al. 2010; Hooftman et al. 2011). Whether or not the detrimental effect of delayed flowering is strong enough to prevent crop (trans)gene escape in lettuce, will be explored in future research with simulation models (Meirmans et al. in prep.) using these empirical field data.

In conclusion

Although we did not employ transgenic hybrids, this lettuce crop–wild model system allowed us to study selection patterns on specific genomic blocks in different genetic backgrounds and for different environmental conditions. In addition, it allowed us to experimentally test if these selection patterns implied that potential transgenes inserted in different parts of the genome differ in the likelihood that they become persistent in the field. Following from our results, these are the main conclusions from this thesis important for ERA:

- Under selective field conditions, various lettuce crop alleles conferred a selective advantage to hybrids, including QTL for traits as germination, flower production, and the main fitness trait, total seed yield (seeds produced per seed sown).

- The performance of crop parental varieties under natural conditions in the field was very poor; however, this does not mean that crop–wild hybrids automatically had a low fitness.
- Fitness distributions indicate that there is a high likelihood that particular hybrid genotypes outperform the wild parent under certain environmental conditions. Therefore, lower average hybrid fitness does not preclude introgression from crops to their wild relatives.
- In general, there was a low correspondence between field QTL and those detected in the greenhouse, suggesting that greenhouse studies are of limited value as predictors for field fitness.
- The expression patterns of abiotic stress QTL was very variable across different stress experiments, therefore it seems difficult to make general statements regarding the fate of a transgene under diverse abiotic stress conditions.
- By following plants through their entire life cycle, we identified specific crop genomic regions under negative, as well as positive selection that were consistent in different field environments.
- Genomic selection patterns varied in lines derived from different crop–wild parents: This could have been due to genetic differences between the two wild parents or two wild relatives used, but in any case it shows that such variation needs to be taken into account.
- We identified one genomic region (top LG7) where the crop genomic background conferred a selective disadvantage, across different sites and crosses, mainly through delayed flowering. Further research, such as modeling studies, should determine whether this region is applicable in transgene mitigation strategies.
- Overall, these results indicate that the probability for a putative transgene spreading indeed depends strongly on the insertion location.

Some afterthoughts

The decision whether or not to allow cultivation of a certain GM crop is not an easy one to make. Besides risks of transgene escape and the potential ecological impacts, there are other issues to consider such as nutritional value, impact on public health, ethics, land use, and the impact on the world food market, where the highest profits might be for a few companies that own patents of the GM crops. On the other hand, GM crops modified for disease and stress resistance can have clear benefits, such as cultivation in areas that are otherwise unsuitable for agriculture, higher yields, and reduction in fertilizers, fungicides, pesticides, and water usage (Snow et al. 2005; James 2010), ultimately leading to a higher food security and profits for farmers.

We have shown that after a hybridization event there is a high likelihood that particular hybrid genotypes will outperform the wild parent under certain environmental conditions thereby providing an avenue for transgene escape. With this in mind and realizing that the process of introgression can be irreversible, it seems wise to continue to apply the precautionary principle, where a GM crop is only allowed if the impact of a potential outbreak is expected to be very small, if the wild relative is not present in the area of cultivation, or if it has several TM strategies implemented minimizing the chances for transgene escape. However, the risks need to be balanced against potential benefits (see e.g., ACRE 2007), given the world's increasing human population and the decrease in agriculturally available lands due to global warming (UN 2011). Some GM crops might provide an important contribution to provide food security for the future world population (Cominelli and Tonelli 2010) and regulators therefore will be increasingly faced with the choice between possible benefits and potential risks.