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

Hartman, Y.

Citation for published version (APA):
1

General introduction
In the past decades, the use of Genetically Modified (GM) crops has risen dramatically, from 1.7 million hectares in 1996 to 148 million hectares in 29 countries in 2010 (James 2010). Tomato with delayed ripening and canola with altered oil content were the first commercially released transgenic crops in the United States in 1994 (Pilson and Prenderville 2004). Most transgenic crops commercially cultivated are genetically modified for herbicide tolerance or insect resistance (Snow et al. 2005); the most cultivated GM crop being herbicide tolerant soybean (*Glycine max*) occupying 73.3 million hectares or 50% of total GM production (James 2010). Since the commercial release of GM crops there has been an ongoing debate about the environmental safety of transgenic crops. This has led to stringent regulations for the commercial production of GM crops. Before a transgenic crop is allowed for commercially release, companies have to perform an elaborate Environmental Risk Assessment (ERA) (EFSA 2011). The ERA must address both the likelihood of an escape of the transgene by means of hybridization with wild relatives or as feral plants, as well as the potential impact of such an event. In this thesis, we focus on the likelihood of introgression of crop alleles into wild relatives and the applicability of Transgene Mitigation (TM) strategies that take advantage of genomic regions in the crop under negative selection in the wild, a method suggested to minimize transgene escape chances (Gressel 1999; Stewart et al. 2003).

**GM crops: benefits and risks**

The world population will increase with an additional 2 billion to 9 billion in 2050 (UN 2011). This coincides with a decrease of available land for agriculture due to salinization and drought, caused by climate change. To meet the growing demand for food of future world population and contrast detrimental effects of climate change on crop yields, it is imperative to develop new crops that, for example, have improved water use efficiency and have improved resistance to drought stress (Cominelli and Tonelli 2010). As genetic and physiological knowledge on stress tolerance increases, genetic engineering becomes more attractive as a fast method to improve crops for abiotic and biotic stresses (Bhatnagar-Mathur et al. 2008; Cominelli and Tonelli 2010).

GM crops improved for disease and stress resistance provide several benefits. For example, crops tolerant to salinity or drought can be cultivated in areas that otherwise would be unsuitable for agriculture. In addition, increasing efficiency of water usage will have a major impact on conservation and availability of water globally (James 2010). Along the same lines, crops tolerant to nutrient limitation need less nutrient input, reducing costs and pollution. Cultivation of insect and disease resistant crops may reduce the use of fungicides and pesticides, and a higher yield can put less pressure on natural areas because less area needs to be cultivated (Snow et al. 2005; James 2010). It should be noted though that the biggest increase in yield is expected from better management practices for existing crops in underyielding nations, including efficient use of fertilizer and water (Tilman et al. 2011). On the other hand, there are several potential risks to the use of GM crops. One of the potential negative ecological effects is increased weedingness, if the transgene contributes to a higher fitness or competitiveness of the wild relative (Pilson and Prenderville 2004; Warwick et al. 2009), leading to higher costs for agricultural weed control. Another possibility is that the wild relative can invade new habitats at the cost of other native species. For example, acquisition of a transgene increasing drought or salt tolerance could expand the typical habitat range of the wild relative (Andow and Zwahlen 2006; Warwick et al. 2009). In addition, continuous influx and introgression of crop alleles into the wild population could lead to genetic displacement of the diversity of the wild recipient species (Hedge et al. 2006). It should be noted that there are still no studies showing evidence for any of these potential negative ecological effects resulting from gene flow from the present transgenic crops to their wild relatives (Kwit et al. 2011).
Factors determining transgene escape

Many crop species have been demonstrated to hybridize with their wild relatives in at least part of their distribution range (Ellstrand 2003; Ellstrand et al. 2011). Moreover, for non-transgenic crop alleles, molecular evidence for transfer of crop alleles to wild relatives has been found for a variety of crop species (Kwit et al. 2011). There is no reason to expect this to be different for transgenic crops. However, the chances of permanent incorporation, introgression, in the genome of wild populations depend mostly on fitness effects rather than gene flow and initial hybridization rates (Ellstrand 2003; Chapman and Burke 2006). Model studies have shown that a selectively advantageous gene will eventually spread even if there is low gene flow pressure (Huxel 1999; Haygood et al. 2004).

Fitness effects include the adaptive fitness of the transgene itself but also that of the genes surrounding the transgene (Fig. 1 adapted from Stewart et al. 2003). After hybridization, linkage disequilibrium in first generation hybrids decays slowly, especially in selfing species, and selection will affect large genomic blocks rather than individual loci (Flint-Garcia et al. 2003). If a transgene, that is selectively neutral or even slightly deleterious in the wild habitat, is in close linkage with a genomic region under positive selection, this transgene could still spread through genetic hitchhiking (Stewart et al. 2003).

Such linkage and hitchhiking effects can also be employed to minimize trangene escape chances. In a TM strategy, first proposed by Gressel (1999), a transgene is placed in close linkage with a gene or genomic block that confers a negative fitness effect to hybrid individuals in the wild habitat (Fig. 1). This was experimentally tested in tobacco (Al-Ahmad et al. 2004, 2005) and oilseed rape (Rose et al. 2009) by placing a transgene in close linkage with a dwarfing gene, causing a great reduction in the survival of these transgenic individuals compared to transgenic individuals without the TM strategy. Techniques for targeted insertion of transgenes to specific regions in the genome, that make TM strategies possible, are under development (Lombardo et al. 2011; Nandy and Srivastava 2011; Shukla et al. 2009).

In general, there is little information on the genes or genomic blocks under selection in the field of crop–wild hybrid individuals (Hails and Morley 2005). In this thesis, we use Quantitative Trait Loci (QTL) analysis that allows pinpointing of traits and genomic areas under selection (Fig. 2a). In the last decade more and more genetic resources, such as genetic maps and markers, are becoming available, making it easier to study genomic selection patterns for an increasing amount of crop–wild complexes (Collard and MacKill 2008). So far, only a few studies, for example, on crop–wild sunflower hybrids have performed QTL analysis to

Figure 1. Schematic representation of a chromosome and chromosomal blocks under selection. Chromosomal blocks are either under positive or negative selection and therefore either have a higher or a lower likelihood of being transferred into the genome of related species. In a transgene mitigation strategy, a transgene placed in a chromosomal block of the crop that is negatively selected for is less likely to be introgressed into the genome of the wild relative. Adapted from Stewart et al. (2003).
detect such patterns (Baack et al. 2008; Dechaine et al. 2009). For ERA (e.g., EFSA 2011), it would be a need to know whether these differences in selection patterns across the genome are present among more crop–wild systems. Therefore, we focus on the following main research questions in this thesis: (i) Do crop genes confer a selective advantage under controlled and field conditions? (ii) How important are linkage drag (hitchhiking effects) for the potential spread of transgenes? (iii) Can small-scale contained experiments with transgenes be used to assess potential ecological consequences?

**Lettuce as a model species**

To study if different genomic regions have different introgression chances and the applicability of TM strategies, we use the crop lettuce (*Lactuca sativa* L.) and its wild relative prickly lettuce (*Lactuca serriola* L.) as a non-transgenic crop–wild model system.

The genus *Lactuca* belongs to the Compositae (Asteraceae), one of the largest plant families, but there are only two major crops within this family, namely sunflower and lettuce (Dempewolf et al. 2008). Cultivated lettuce probably originated somewhere in South-West Asia from where it spread to Egypt and where it appears on old wall paintings and reliefs of tomb walls dating 2500 B.C. (de Vries 1997). *Lactuca sativa* L. is an annual crop that has broad almost circular leaves with a low latex content and without any spines, without basal reproductive shoots and with delayed bolting (de Vries 1997). In addition, natural seed (achene) shattering is reduced due to erect involucral bracts surrounding the seeds (de Vries 1997). Lettuce is mostly harvested in the rosette stage and used as a leaf vegetable in salads, but there are varieties that are produced for the stalk or for oil extraction from the seeds (de Vries 1997). Its closest wild relative and probably one of the direct ancestors is *L. serriola* L. (Koopman et al. 2001).

*Lactuca serriola* L., wild prickly lettuce, is a common annual weed occurring mainly in pioneer and disturbed ruderal habitats, nowadays often roadsides, railways and construction sites, but also in agricultural fields (Lebeda et al. 2001; Hooftman et al. 2006; D’Andrea et al. 2009). It has a broad distribution range from Europe, North America, South Africa, Asia to Australia. In Western Europe the distribution range has expanded dramatically over the last 50–60 years (Hooftman et al. 2006; D’Andrea et al. 2009). Two forms are distinguished based on the leaf shape. These are *L. serriola f. integrifolia* with entire leaves and *L. serriola f. serriola* with serrate leaves, the type used in this study. In contrast to cultivated lettuce, leaves contain white bitter latex and plants frequently develop 2 mm long spines on the stem base and on downside leaf midribs. Individuals show almost no head formation, but instead bolt and flower early. Plants develop many basal and cauline reproductive shoots from axillary meristems, and produce capitula that have reflexed involucral bracts.

This crop–wild model system has several advantages to study hybridization and introgression processes. Cultivated and wild lettuce species are fully interfertile, without any known pre- or post-zygotic barriers (Kesseli et al. 1991; Koopman et al. 2001). Lettuce is predominantly selfing, but also insect pollinated with a reported outcrossing rate of approximately 1–5% (D’Andrea et al. 2008; Giannino et al. 2008). Cultivated and wild lettuce have overlapping geographic regions and flower simultaneously. Although commercially cultivated lettuce is usually harvested at the rosette stage, sometimes cultivated lettuce reaches the flowering stage in privately owned gardens, allowing the change for outcrossing. In a series of field experiments, Hooftman et al. (2005, 2007, 2009) showed that at least four generations of artificially created crop–wild hybrids on average had higher germination and survival rates than the wild parent, so potentially through hybridization crop alleles can introgress into the wild population.
Figure 2a. Construction of recombinant inbred lines (RILs) and principles of Quantitative Trait Loci (QTL) analysis. Solid lines indicate backcrossing and dotted lines indicate selfing. *Lactuca serriola* (UC96US23) and *Lactuca sativa* cv. Salinas differing in the occurrence of spines were crossed to form an F₁ with intermediate spine formation. An F₁ individual was selfed to form a population of F₂ individuals. Each F₂ is subsequently selfed for nine additional generations by single-seed descent, ultimately forming several RILs. Each RIL is almost completely homozygous for a section of a parental chromosome. The RILs are scored for several genetic markers, as well as for the occurrence of spines. The arrows mark those RILs that have almost no spines and also have inherited a genomic block only from the parent without spines (*L. sativa*), indicating that this chromosomal region probably contains a QTL for this trait. Adapted from Mauricio (2001).

Figure 2b. Construction of Backcross (BC) lines. Solid lines indicate backcrossing and dotted lines indicate selfing. Adapted from Hooftman et al. (2011).
In recent years, a large set of genotyped Recombinant Inbred Lines (RILs) has become available from the Compositae Genome Project (http://compgenomics.ucdavis.edu). RILs are homozygous lines, each with a unique combination of chromosomal segments from the original parents, which can be multiplied without any genetic change occurring (Collard et al. 2005). This means that the same genotypes can be used several times in different experiments, locations (greenhouse and field), and years.

In this thesis, we use two crossing populations. First, a RIL population resulting from a cross between the cultivar Crisphead lettuce (\textit{L. sativa} cv. Salinas) and its wild relative prickly lettuce, collected in California, USA (Fig. 2a). Second, a Backcross (BC) population resulting from a cross between the cultivar Butterhead lettuce (\textit{L. sativa} cv. Dynamite) as a pollen donor, and its wild relative prickly lettuce, collected in Eys, the Netherlands (Fig. 2b). The progeny of this cross was further backcrossed to the wild relative to mimic the most likely scenario after a hybridization event.

**Outline thesis**

In this thesis, we have measured domestication and/or fitness-related traits of lettuce hybrids in several greenhouse and field experiments (Fig. 3) and performed QTL analyses to locate the associated genomic regions. In each chapter, we highlight the likelihood of introgression of crop genes to wild relatives, the potential use of TM strategies, and/or the implications for designing new transgenic crops. We also discuss the implications of our findings for ERA for lettuce as well as possible generalizations to other (vegetable or seed) crops.

Chapter 2 addresses the genetic architecture of domestication in Crisphead lettuce and how this compares to trends found in other species. In a greenhouse experiment, we measured domestication and fitness-related traits during the entire life cycle of 114 RILs from a cross between the cultivar \textit{Lactuca sativa} cv. Salinas (Crisphead) and a Californian wild relative \textit{L. serriola} (UC96US23). Specifically, we look at effect size of QTL, clustering of QTL, and directionality of allelic effects.

Chapter 3 addresses selection on different genomic crop regions under field conditions and compares this to results obtained under greenhouse conditions (described in Chapter 2). We performed a field experiment at two locations in the Netherlands with contrasting soil types and measured fitness and fitness-related traits during the entire life cycle for a core group of 98 RILs. With the main fitness trait seeds produced per seed sown, we have been able to identify genomic crop regions under positive and negative selection.

In Chapter 4, we focus on the consistency of QTL for the same abiotic stress from different experiments as well as the interaction between abiotic stress and competition. We performed a series of stress experiments by subjecting plants to drought, salinity, and nutrient limitation in (i) a controlled, non-competitive greenhouse environment, (ii) a controlled, competitive greenhouse environment, and (iii) a non-competitive field situation. Experiments were performed with core groups of 60 RILs (experiment i) and 90 RILs (ii and iii). At the end of each experiment, we measured above ground biomass when the plants where still in the rosette stage and used this as a proxy for plant performance.

In Chapter 5, we extend the results presented in Chapter 3 with those of Backcross (BC) lines and, in addition, studied the performance of individual hybrid lineages for both crossing types. We conducted the fitness field experiment of Chapter 3 also with 98 BC\textsubscript{1} families from a cross between the cultivar \textit{L. sativa} cv. Dynamite (Butterhead) and a wild prickly lettuce, \textit{L. serriola} f. \textit{serriola}, collected in Eys, in the Netherlands. BC\textsubscript{1} individuals were genotyped,
Figure 3. Clockwise starting at top left: crossing work to create Backcross (BC) family lines; close-up flower; individuals of three recombinant inbred lines (RILs) of a control (at the back), drought (in the middle), and recovery (at the front) treatment; field experiment with RILs and BC₁S₁ families in 2010; salt and nutrient limitation stress experiment in 2009; germination experiment in 2008.
Chapter 1

allowed to self for one generation and subsequently their BC₁S₁ offspring was used in the experiments. This design allowed us to study differences in genomic selection patterns between different lettuce cultivar–wild crosses, hybrid performance in early- and late-generation hybrids, and environmental influence on hybrid fitness distributions.

Finally, in Chapter 6, we summarize our findings focusing on the main research questions. First, we answer if crop genes conferred selective advantages or disadvantages under controlled and field conditions. Second, we focus on the importance of linkage drag (hitchhiking effects) for the potential spread of transgenes. Third, we answer if small-scale contained experiments with transgenes can be used to assess potential ecological consequences in the field.

Funding
This study is funded by Dutch Organization for Scientific Research (NWO) as part of the ERGO program (838.06.042). The Compositae Genome Project (http://compgenomics.ucdavis.edu), that developed the RILs used in this study, is supported by the NSF Plant Genome Program; award #0820451.