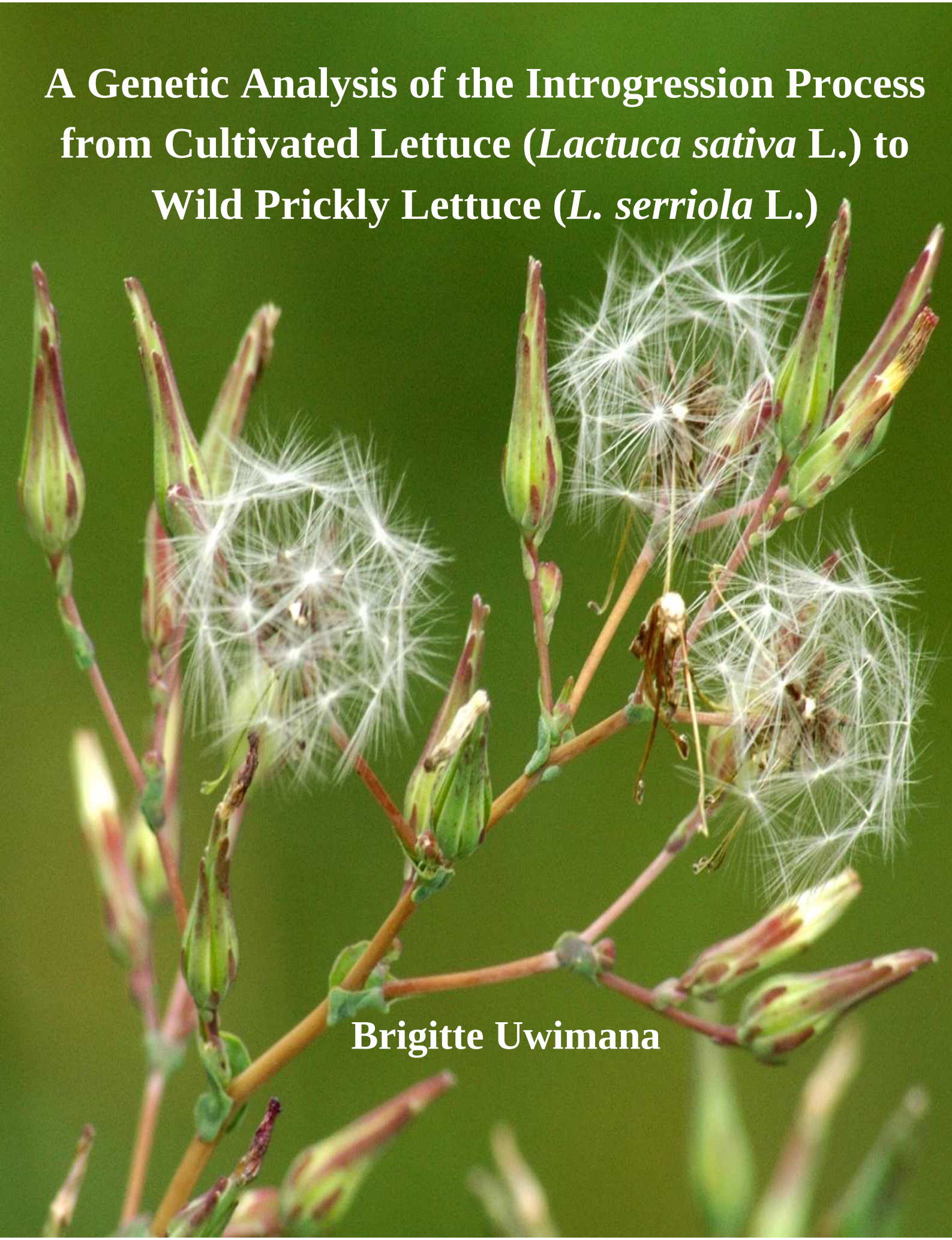
A Genetic Analysis of the Introgression Process from Cultivated Lettuce (*Lactuca sativa* L.) to Wild Prickly Lettuce (*L. serriola* L.)



Brigitte Uwimana

A Genetic Analysis of the Introgression Process from Cultivated Lettuce (*Lactuca sativa* L.) to Wild Prickly Lettuce (*L. serriola* L.)

Brigitte Uwimana

Thesis committee

Thesis supervisor

Prof. dr. Richard G.F. Visser
Professor of Plant Breeding
Wageningen University

Thesis co-supervisors

Dr. Marinus J.M. Smulders
Researcher, Plant Research International
Wageningen UR

Dr. Clemens C.M. van de Wiel
Researcher, Plant Research International
Wageningen UR

Other members

Dr. Jeremy B. Sweet, Independent Environmental Consultant, UK
Prof. dr. R. (Kuke) Bijlsma, University of Groningen, the Netherlands
Prof. dr. Peter H. van Tienderen, University of Amsterdam, the Netherlands
Prof. dr. Rolf F. Hoekstra, Wageningen University, the Netherlands

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences

A Genetic Analysis of the Introgression Process from Cultivated Lettuce (*Lactuca sativa* L.) to Wild Prickly Lettuce (*L. serriola* L.)

Brigitte Uwimana

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. dr. M.J. Kropff,
in the presence of the
Thesis Committee appointed by theAcademic Board
to be defended in public
on Friday 2 December 2011
at 1.30 p.m. in the Aula

Brigitte Uwimana
A Genetic Analysis of the Introgression Process from Cultivated Lettuce (*Lactuca sativa* L.) to Wild Prickly Lettuce (*L. serriola* L.), 144 pages.

Thesis, Wageningen University, Wageningen, the Netherlands (2011)
With references, with summaries in Dutch and English

ISBN 978-94-6173-031-2

Dedicated to my sister Joséphine Umulisa

Contents

Chapter 1

General Introduction

Hybridization in plants

Early botanists believed in the theory of fixed species, according to which plants cross within kinds, resulting in fixed species. The concept of hybridization between species was conceived by Carol Linnaeus in the 18th century who, after observing a plant he thought was *Linaria*, was convinced that it was a relative of *Linaria* but with a hybrid nature due to its different floral structure, and he named it *Peloria* (Larson 1968). He later found more proof of interspecific hybrid plant species; but with little scientific evidence to back up the hypothesis, interspecific hybridization received little attention in his days and later on (Larson 1968). It was not until in the 20th century that cytogenetic studies brought light to interspecific hybridization as an important evolutionary phenomenon (Baack and Rieseberg 2007). Nowadays interspecific hybridization is accepted as one of the most important phenomena through which new species evolve, not only in plants but in animals as well (Carson *et al.* 1975; Dowling and Secor 1997; Ellstrand *et al.* 1996; Smulders *et al.* 2008; van Tienderen 2004).

Crop-wild gene flow

Since the development and commercial release of genetically modified (GM) crop varieties in the late 1980s, hybridization in plants has become a persistent topic of discussion. GM crop varieties are developed to improve crop yield, increase crop resistance and tolerance to biotic and abiotic stress factors, lower the production costs, improve weed and pest control - hence making crop production more environment-friendly - and improve food, feed and pharmaceutical qualities (James 2003; Warwick *et al.* 2009). However, the release of GM varieties has raised concerns both in the general public and among scientists about the potential risks associated with their commercial and hence large scale cultivation. As far as the environment is concerned, three ecological concerns have been raised which are associated with hybridization between transgenic crops and their wild relatives. Firstly, transgene 'escape' through crop-wild hybridization may render weedy wild relatives weedier in agricultural areas, for instance by incorporating into the weed species genes that confer tolerance to herbicides. Secondly, it may result in hybrid swarms that are more fit than the wild parent in its natural habitats and could displace the latter, resulting in genetic erosion. Thirdly, the crop-wild hybrids may show phenotypes diverging from the wild parent to the extent that they may invade new habitats (Chapman and Burke 2006; Groot *et al.* 2003; Warwick *et al.* 2009).

In their often cited review, Ellstrand *et al.* (1999) put together documented evidence for crop-wild hybridization, showing that 12 out of the 13 most important crops in terms of cultivation area hybridize with their wild relatives somewhere in the world. In the Netherlands, 23 out of 42 domesticated species have small to substantial potential to hybridize with their wild relatives (De Vries *et al.* 1992). Because crops can hybridize with their wild relatives, and because once the transgene escapes to the wild the process could hardly be reversed, crop-wild gene flow has become a subject of scientific scrutiny (Conner *et al.* 2003; Snow *et al.* 2005; Tiedje *et al.* 1989).

After hybridization, the outcome of crop-wild gene exchange will depend on the performance of hybrids under natural conditions (Hails and Morley 2005). The net effect can be negative, for instance if crop genes reduce the hybrids' competitive ability under natural conditions, or positive, if crop genes confer fitness advantages. Natural selection will weed out maladapted genotypes, and could lead to the establishment of successful transgressive phenotypes with high fitness (Burke and Arnold 2001). The fate of the hybrids will therefore be determined by their individual genetic composition and the selection pressure defined by the prevailing environmental conditions. It is within such a context that the effects of transgenes should be evaluated: the baseline is the dynamics of the crop-wild hybridization process, and effects of introduced transgenes in such a system are superimposed upon this baseline. It thus follows that both knowledge of the baseline system of hybridization as well as the putative effects of transgenes (Chapman and Burke 2006; Stewart *et al.* 2003) are needed for an adequate assessment of the effects of hybridization. One instance about the dynamics of crop-wild hybridization and introgression is the effect of linkage. Selection does not apply on genes affecting fitness alone, but on the genomic block of which the gene affecting fitness is part (Kwit *et al.* 2011; Stewart *et al.* 2003). Therefore, a (trans)gene neutral to fitness or mildly deleterious can be introgressed if it is linked to a gene that affects fitness positively, a phenomenon known as genetic hitchhiking. In the same way, a (trans)gene could be selected against if it is inserted close to a gene that confers reduced fitness, a phenomenon known as background selection (Figure 1).

This study was initiated with the aim of understanding the basic dynamics of crop-wild hybridization and introgression. We used the crop-weed complex consisting of *Lactuca sativa* L and *L. serriola* L. to study the genetic process of introgression from crops to wild relatives. After crop-wild hybridization, the hybrids undergo selection by abiotic (drought, salinity, nutrient deficiency, cold, etc.) and biotic stress factors (disease, herbivore insects, competition). While some studies have looked into the effect While some studies have looked into the effect of

introgression of biotic stress resistance/tolerance genes (Cao *et al.* 2009; Hooftman *et al.* 2007c; Mason *et al.* 2003), abiotic stress factors have received little attention so far. We therefore evaluate the performance of the hybrids under the major abiotic stress conditions of salinity, drought and nutrient deficiency under greenhouse conditions.

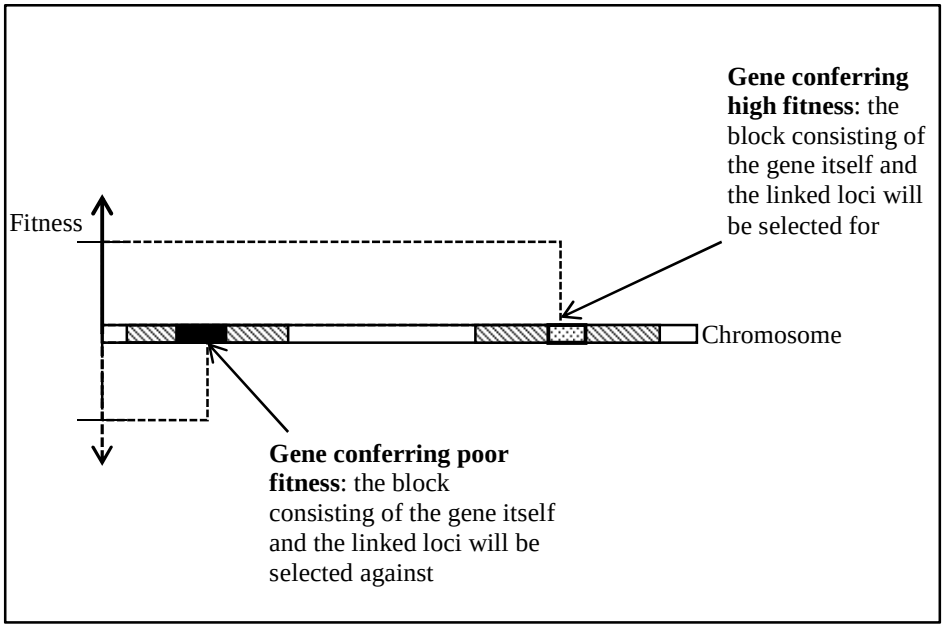


Figure 1 Selection and introgression apply on the genes affecting fitness and the linked loci (adapted from Stewart *et al.* 2003)

Lactuca serriola L. and *L. sativa* L.

Cultivated lettuce (*Lactuca sativa* L.) is one of the domesticated species with a possible ecological effect in the Netherlands due to its hybridization potential with its wild relative prickly lettuce (*L. serriola* L.) (De Vries *et al.* 1992). *L. sativa* is an annual vegetable crop from the family *Asteraceae* or *Compositae*. It has been domesticated as early as 2500 B. C., probably in South-West Asia (De Vries 1997; Frietema de Vries *et al.* 1994). It is mostly harvested at the rosette stage and consumed as salad, with some types harvested as seeds and used for oil extraction (oilseed lettuce) or for the stem base (stalk lettuce). Among the members of the large *Lactuca* genus, *L. serriola* is the closest relative of *L. sativa* of which it is considered to be a progenitor and part of the primary gene pool (Koopman *et al.* 1998).

L. serriola is a wild weed species original from the Mediterranean region but it has expanded its distribution worldwide (Lebeda *et al.* 2004). It thrives in anthropologically disturbed areas such

as roadsides, construction sites and agricultural fields, and along railways (Lebeda *et al.* 2001; Lebeda *et al.* 2004). *L. serriola* and *L. sativa* have the same number of chromosomes (2n=2x=18), are completely cross-compatible in the two directions (*L. serriola* x *L. sativa* and *L. sativa* x *L. serriola*), and the resulting hybrids are viable and fertile (D'Andrea *et al.* 2008; De Vries 1990; Koopman *et al.* 1993). Morphologically, *L. serriola* is mostly distinct from *L. sativa* (Figure 2) (de Vries and Raamsdonk 1994), but some accessions of the two species have overlapping traits, making it difficult to draw a distinguishing line between them (Frietema de Vries *et al.* 1994). The morphological overlap in combination with crossing experiments and genotypic analysis has led to the suggestion that the two species may be conspecific (Frietema de Vries *et al.* 1994; Koopman *et al.* 2001).

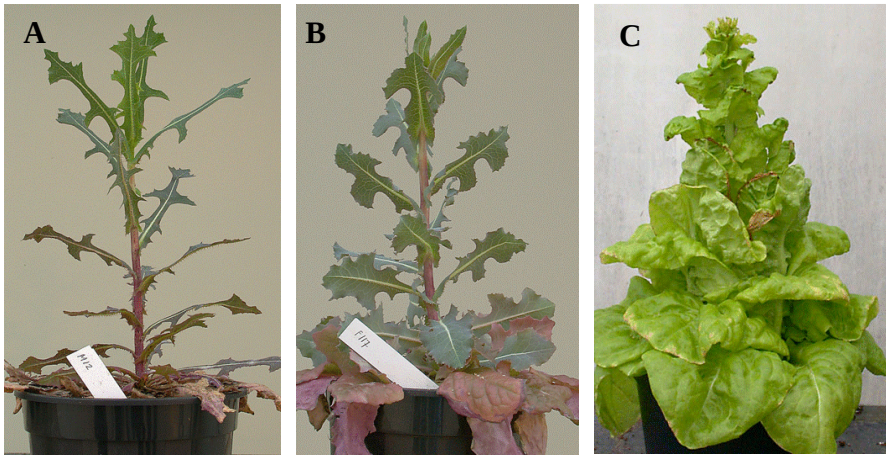


Figure 2 *Lactuca serriola* (A) and *L. sativa* (C) plants and an F_1 plant from a cross between the two species (B)

L. serriola has been expanding its geographical distribution in Europe from south to north (Bowra 1992; Hooftman *et al.* 2006; Lebeda *et al.* 2001). One of the hypotheses put forward for this invasion is the possibility that *L. serriola* has acquired new fitness traits from *L. sativa* through hybridization. Although *L. sativa* is mostly harvested before it produces seeds, the two species grow and flower sympatrically in many locations. For instance, seed production for *L. sativa* is done in open fields in certain regions in Southern Europe. In home gardens and the so-called "amateur" gardens in the Netherlands, plants are sometimes not harvested and they are left to flower and produce seeds. Moreover, cases have been reported where the low market price of lettuce heads has prompted farmers not to harvest their lettuce crop, leaving hundreds hectares of lettuce to flower and produce seeds in open fields (D'Andrea *et al.* 2009). Furthermore, previous

studies of *L. sativa* x *L. serriola* hybrids have shown that some of the hybrid lineages show improved vigour and fitness over the wild parent under field conditions and so may have the potential of displacing *L. serriola* in its natural habitats (Hooftman *et al.* 2009; Hooftman *et al.* 2007b; Hooftman *et al.* 2005; Hooftman *et al.* 2008).

Abiotic stress

As stated above, the establishment of crop genes in the wild will depend on the selection pressure exercised over the hybrid plants by the prevailing natural conditions. Abiotic stress factors such as drought, salinity, nutrient deficiency and extreme temperatures are the primary factors that affect the growth of the plants (Boyer 1982; Munns and Tester 2008; White and Brown 2010; Witcombe *et al.* 2008). With climate change and land degradation, such stress factors are expected to become more important in agriculture in the future as well. For instance, salinization is expected to affect 50% of the arable land by 2050 (Wang *et al.* 2003). Therefore, abiotic stress factors are likely to play an important role in determining the fate of the hybrids after crop-wild hybridization by acting as selection forces.

Conventional breeding for abiotic stress tolerance has been limited due to the complexity of the trait and the involved mechanisms (Cuartero *et al.* 2006; Farooq *et al.* 2009; Roy *et al.* 2011). Therefore, genetic modification is regarded as a potential solution to breeding for abiotic stress tolerant varieties (Bhatnagar-Mathur *et al.* 2008; Tester and Langridge 2010; Vinocurand Altman 2005; Zhang *et al.* 2000). Many studies are currently undertaken on the application of genetic modification to improve plant tolerance to abiotic stress factors, and the release of abiotic stress-tolerant varieties should be expected in the near future (Abdeen *et al.* 2010; Aharoni *et al.* 2004; Karaba *et al.* 2007). With the increasing intensity of the abiotic stresses, if an abiotic stress tolerant variety happens to hybridize with a wild relative, the genes conferring tolerance will increase the fitness of the hybrids containing them, hence presenting an advantage over the wild plants.

In this study we evaluated the performance of the hybrids under the major abiotic stress conditions of salinity, drought and nutrient deficiency. In order to decipher the performance of the hybrids under each of the mentioned stress factors, these factors were artificially created under greenhouse controlled conditions. The hybrids were also grown on the field where they were subject to natural conditions and they were evaluated for vigour at the rosette stage, vigour at the adult stage, survival and seed production.

Scope of the thesis

This project was part of the programme "Ecology Regarding Genetically modified Organism" (ERGO), an initiative of the Netherlands Organization for Scientific Research (NWO) established to fund and coordinate research on the ecological risks associated with the cultivation of GM varieties (http://www.nwo.nl/nwohome.nsf/pages/NWOA_6JNP94). ERGO programme focuses on three research themes, namely multitrophic interactions regarding gene-modified crops, effects of hybridization and introgression between crops and their wild relatives, and effects of gene-modified crops on soil ecosystem functioning. This project falls under the second research theme. It aimed at establishing a baseline about the genetic process of introgression from crops to wild relatives using *L. sativa* and *L. serriola* as crop-weed complex model with an emphasis on the contribution of the crop parent to the vigour and fitness of the hybrids and the effect of linkage on the likelihood of introgression of a specific crop genomic segment based on its genomic location. Because of the restrictions imposed on the cultivation of GM plant material, we did not use GM lettuce. Instead, research was carried out using conventionally bred lettuce varieties. Although crop-wild hybrids are bound to undergo selection under natural field conditions and greenhouse experiments are therefore likely to be less representative of the growing conditions of the hybrids, controlled greenhouse experiments offer certain advantages over field experiments in terms of the number of experiments that can be run in a certain period of time and the possibility to creating conditions that mimic a certain stress factor so that the tolerance or resistance of the plants to the stress can be deciphered. Combining greenhouse and field experiments can therefore give insight concerning the use of greenhouse results in predicting field conditions. We therefore carried out greenhouse experiments on specific stress factors and run field experiments under natural conditions. Based on the knowledge that selection takes place during the early growth stage of crop-wild lettuce hybrids (Hooftman *et al.* 2009; Hooftman *et al.* 2005), the greenhouse experiments concerned the vigour of the hybrids at the rosette, whereas the field experiments encompassed the whole life cycle of the plants, from germination till seed production.

The two major factors determining the outcome of crop-wild hybridization (hybrid genetic make-up and the environment) are addressed by studying three hybrid classes F_2, BC_1 and BC_2 resulting from one crop-wild hybridization event under the major abiotic stress conditions of drought, salinity and nutrient deficiency. We answer the questions (i) whether there is evidence

of spontaneous hybridization between *L. serriola* and *L. sativa*, (ii) whether crop genes confer any (dis)advantage to the crop-wild hybrids under controlled conditions of non-stress and abiotic stress conditions and under field conditions, (iii) whether the (dis)advantageous effects of the crop are dependent on environmental conditions, and (iv) whether we can identify genomic regions where transgenes could be inserted with the purpose of mitigating their persistence after crop-wild hybridization.

The second chapter of this thesis deals with the quantification of the occurrence of spontaneous crop-wild hybrids in natural populations of *L. serriola* both in Europe, the Middle East and Central Asia, using Bayesian methods of analysis and a set of simple sequence repeat markers (microsatellites). Based on the identified hybrids and their geographical localization, we discuss whether hybridization between *L. serriola* and *L. sativa* might have contributed to the recent spread of *L. serriola* in Europe.

In the third chapter we analyse the genetic basis of plant vigour in a crop-wild F_2 population under non-stress, drought, salt and nutrient deficiency conditions. Using Single Nucleotide Polymorphism (SNP) markers, we map Quantitative Trait Loci (QTLs) associated with vigour and define the role of the crop genome in the vigour of the hybrids. We define the genetic mode of action of the QTLs and their genomic localization.

In the fourth chapter we assess the genetic effect of introgression by studying the vigour of the hybrids in two backcross populations BC_1 and BC_2 generated by backcrossing F_1 progeny to the wild parent *L. serriola*, thus mimicking the introgression process of crop genomic segments into a wild genetic background. While the selfing pathway might be more common in lettuce due to its selfing nature, the backcross pathway may take place as a result of a higher frequency of "pure" wild plants than the hybrids. By conducting experiments like the ones described in chapter three under control and abiotic stress conditions of drought, salinity and nutrient deficiency, we establish the effect of introgression of smaller and fewer crop genome segments into the wild genetic background.

Because the experiments reported in chapters 3 and 4 were carried out under greenhouse controlled conditions, whereas spontaneous crop-wild hybridization takes place outside and hybrids are subject to natural conditions, chapter 5 deals with experiments with the BC_1 population conducted under semi-natural field conditions. In this chapter we establish a link between field and greenhouse conditions, and between plant vigour and fitness. We establish the

effect of genetic by environment (GxE) interaction on plant fitness, and investigate whether small-scale contained experiments could be used to assess potential ecological consequences in the field. In the general discussion (chapter 6) we discuss the above-mentioned research questions.

Chapter 2

A Bayesian analysis of gene flow from crops to their wild relatives: cultivated (*Lactuca sativa* L.) and prickly lettuce (*L. serriola* L.), and the recent expansion of *L. serriola* in Europe

Brigitte Uwimana[1], Luigi D'Andrea[2], François Felber[2], Danny A.P. Hooftman[3], Hans C.M. den Nijs[4], Marinus J.M. Smulders[1], Richard R.G.F. Visser[1] and Clemens C.M. van de Wiel[1]

[1] Wageningen UR Plant Breeding, Wageningen, The Netherlands

[2] Laboratoire de Botanique Evolutive, Université de Neuchâtel, Switzerland

[3] Centre for Ecology & Hydrology, Wallingford, UK

[4] IBED, UvA, Amsterdam, The Netherlands

Abstract

Interspecific gene flow can lead to the formation of hybrid populations that have a competitive advantage over the parental populations, even for hybrids from a cross between crops and wild relatives. Wild prickly lettuce (*Lactuca serriola*) has recently expanded in Europe and hybridization with the related crop species (cultivated lettuce, *L. sativa*) has been hypothesized as one of the mechanisms behind this expansion. In a basically selfing species such as lettuce, assessing hybridization in natural populations may not be straightforward. Therefore, we analysed a uniquely large dataset of plants genotyped with SSR markers with two programs for Bayesian population genetic analysis, STRUCTURE and NewHybrids. The dataset comprised 7738 plants, including a complete genebank collection, which provided a wide coverage of cultivated germplasm and a fair coverage of wild accessions, and a set of wild populations recently sampled across Europe. STRUCTURE analysis inferred the occurrence of hybrids at a level of 7% across Europe. NewHybrids indicated these hybrids to be advanced selfed generations of a hybridization event or of one backcross after such an event, which is according to expectations for a basically selfing species. These advanced selfed generations could not be detected effectively with crop-specific alleles. In the northern part of Europe, where the expansion of *L. serriola* took place, the fewest putative hybrids were found. Therefore, we conclude that other conditions than crop-wild gene flow, such as an increase in disturbed habitats and/or climate warming, are more likely explanations for this expansion.

Introduction

Gene flow through hybridization is a common phenomenon among closely related plant species. Recent studies have shown it to be more frequent between crop species and their wild relatives than assumed based on the supposition that domestication traits are likely to reduce fitness under natural conditions (Ellstrand 2003). For instance, gene flow was reported to occur between 12 of the 13 most important food crops and their respective wild relatives (Ellstrand *et al.* 1999). With the large-scale cultivation of genetically modified cultivars, gene flow from crops to their wild relatives has attracted public interest and concern, and has initiated research on gene escape and introgression in the framework of environmental risk assessment of transgenic plants (Pilson and Prendeville 2004; Snow *et al.* 2005; Chapman and Burke 2006; Warwick *et al.* 2008; Warwick *et al.* 2009).

Gene flow may have evolutionary impact, especially when particular genes from crops would increase the fitness, and thus possibly the invasiveness of weeds, for instance, by increasing their adaptability to various climatic or environmental conditions (Ellstrand and Schierenbeck 2000; Langevin *et al.* 1990; Magnussen and Hauser 2007). In this regard, an invasive trend was reported for wild (weedy) prickly lettuce (*Lactuca serriola*), the closest wild relative of cultivated lettuce (*L. sativa*) in many Mediterranean, Central and Western European countries (Frietema de Vries *et al.* 1994; Lebeda *et al.* 2004b). Hooftman *et al*. (2006) reported a sweeping spread of *L. serriola* also in the Netherlands since 1980 and lists four possible reasons for this recent invasiveness: (1) a change in environment due to global warming; (2) increased landscape disturbance by human activities, which produces more suitable habitat; (3) micro-evolution of the species towards extended adaptability, and (4) hybridization between *L. serriola* and cultivated *L. sativa*. The latter reason would be a direct consequence of gene flow through interspecific hybridization, leading to the transfer and introgression of genes from the crop to the wild lettuce that confer increased fitness to the resulting crop-wild hybrids.

Several aspects of gene flow in lettuce have been studied previously. De Vries (1990) showed that *L. sativa* can be crossed with *L. serriola* to form viable and fertile hybrids. Even though *L. sativa* and *L. serriola* are basically self-pollinators, Thompson *et al*. (1958) reported an out-crossing rate of 1-5% among *L. sativa* varieties, and D'Andrea *et al*. (2008) an interspecific hybridization rate of up to 2.5% between the two species. Hooftman *et al*. (2005; 2009) studied the performance of the hybrids resulting from manual crosses between *L. serriola* and *L. sativa*

and Hooftman *et al.* (2007) modelled the long-term consequences of hybridization between the two species. These two studies established that hybrids between *L. sativa* and *L. serriola* are viable, fertile and that the hybrid offspring may even be fitter than the wild parent. Based on single fitness components, hardly any significant differences were detected between prickly lettuce and the hybrid plants, and backcrossed hybrids were morphologically indistinguishable from their wild parent (*L. serriola*) (Hooftman *et al.* 2005). Hence, the fact that only very few occurrences of crop/wild hybrid lettuce in the field have been reported (*cf.* Frietema de Vries *et al.* 1994) is not necessarily proof of a lack of occurrence as it may, at least in part, be due to problems in recognizing putative hybrids.

In the present study, we aimed to quantify the spontaneous occurrence of gene flow between cultivated and wild lettuce in Europe. A number of methods can be used for the identification of hybrid plants in natural populations of wild relatives of crop species, including screening based on phenotypic traits (Ureta *et al.* 2008), tracking crop-specific markers (Morrell *et al.* 2005; Scurrah *et al.* 2008; Westman *et al.* 2001) and, in case of GM crops, tracking the transgene itself (Warwick *et al.* 2008). These methods do not work well in all cases. As already indicated, for the present study the use of morphological traits would be difficult because the hybrids resulting from crosses between *L. serriola* and *L. sativa* often look morphologically like *L. serriola* (Hooftman *et al.* 2005). The use of a transgene as a marker is also not applicable outside of contained conditions because, there is not (yet) any transgenic lettuce cultivar allowed for commercial cultivation. The "crop-specific" allele approach scans each locus for alleles with differences in occurrence between crop and wild relatives. Alleles far more frequent in crops are then used as indications for introgression from the crop when found in wild plants. This method has been regularly used to trace hybridization between crops and wild relatives, but suffers from two problems: (1) the definition of "crop-specific" alleles; and (2) how to distinguish their occurrence as a result of recent introgression from one as a result of a more ancient common ancestry (*e.g.* Van de Wiel *et al.* 2005). Thus, for this study, we used two Bayesian posterior probability-based methods, one as implemented in the software package STRUCTURE (Pritchard *et al.* 2000) and the other as implemented in NewHybrids (Anderson and Thompson 2002), to analyse two large datasets of lettuce samples, one from the gene bank collection of crop (*L. sativa*) and wild lettuce (*L. serriola*), and the other set comprising of *L. serriola* samples collected across Europe in the period of 2002-2005. Together, these datasets constitute an exceptionally broad set for a study of gene flow between crops and hybrids. When Smulders *et*

al. (2008) compared STRUCTURE with NewHybrids to detect hybridization with cultivated poplar hybrids in offspring of wild *Populus nigra* trees, they found that NewHybrids was more informative on the degree of hybridization. As lettuce is highly selfing, we anticipated that putative hybrids would have a high likelihood of being advanced selfed generations. We therefore applied STRUCTURE to identify potential hybrid plants and NewHybrids to infer the number of selfings or backcrossings after an initial hybridization event between *L. serriola* and *L. sativa*. STRUCTURE results were checked using the programme InStruct (Gao *et al.* 2007), which takes into account the divergence from the Hardy–Weinberg equilibrium due to self-fertilization. We compared the STRUCTURE results with a crop-specific allele approach, in order to assess to what extent the latter method still has its value with regard to its relative easy implementation for detecting gene flow between crop and wild relatives.

Materials and methods

Plant material

We studied the crop-weed complex of cultivated lettuce (*L. sativa*) and wild prickly lettuce (*L. serriola*). *L. serriola* is a weed plant, which thrives in anthropogenically disturbed areas (Lebeda *et al.* 2001), whereas *L. sativa* is a vegetable crop species. The two species are closely related and have the same number of chromosomes (2n = 18) (Koopman *et al.* 1993; Koopman *et al.* 2001). They are readily crossable without any known barrier, and their hybrids are viable and fertile (De Vries 1990). *L. serriola* is mostly distinct from *L. sativa* based on their morphological traits (De Vries and Van Raamsdonk, 1994), but their hybrids, especially those resulting from backcrosses to *L. serriola*, are generally not distinguishable from the latter (Hooftman *et al.* 2005).

We used plant material originating from two sources: the lettuce collection from the Centre for Genetic Resources, The Netherlands (CGN), and a recent collection of *L. serriola* from across Europe. CGN hosts the largest lettuce germplasm collection worldwide (http://documents.plant.wur.nl/cgn/pgr/ildb/download.htm), which has a comprehensive representation of genetic variation in cultivated lettuce, supplemented with a fair representation of wild relatives, particularly of *L. serriola* (Van de Wiel *et al.* 2010). This collection comprises lettuce accessions collected since 1940, with some over-representation of germplasm from Europe and the Middle East (Lebeda *et al.* 2004a). *L. serriola* accessions of this collection used

in this study are designated as "CGN *L. serriola*" and *L. sativa* accessions are designated "*L. sativa*".

The recent *L. serriola* collection was sampled from 2002 to 2005 within the EU project "Analysis of gene flow from crop to wild forms in lettuce and chicory and its population-ecological consequences in the context of GM-crop biosafety" (ANGEL, QLK3-CT-2001-01657, http://www.plant.wageningen-ur.nl/projects/angel/) (Van de Wiel *et al*. 2003). These *L. serriola* individuals are designated as "ANGEL *L. serriola*". They were collected in 17 European countries (Austria, Bulgaria, Croatia, Czech Republic, Denmark, France, Germany, Greece, Hungary, Italy, Luxembourg, Netherlands, Poland, Portugal, Slovakia, Spain, Sweden and Switzerland) from ruderal sites such as roadsides, along railways, vicinities of riverbanks, crop fields and construction sites, and in vegetable kitchen gardens. The kitchen garden locations included "amateur gardens" in the Netherlands where *L. sativa* plants are more often left to flower than in professional cultivations and thus hybridization may have a good likelihood of occurrence (Hooftman *et al*. 2005). The particular advantage of the ANGEL and CGN datasets is that CGN has a comprehensive representation of germplasm cultivated world-wide as well as wild material collected from a large part of the areas where *L. serriola* occurs, and the ANGEL samples added more details and density on recent wild populations across the European continent.

Genotyping

Ten SSR markers, originally described by Van de Wiel *et al*. (1999), were used to genotype the individual plants (Table 1). The genetic positions for these ten markers have been determined on the lettuce genetic map (Truco *et al*. 2007). Eight marker loci were located on 8 different linkage groups and two loci were located on the ninth chromosome but with a distance of 86 cM. Thus, the loci were considered as genetically unlinked. The CGN samples were genotyped under the EU project "Molecular markers for genebanks: Application of marker technology for the improvement of *ex situ* germplasm conservation methodology" (PL96.2062) using a gel-based ABI PRISM® 377 DNA Sequencer (Applied Biosystems) (Van Hintum 2003, Van Treuren *et al*. 2008). The ANGEL samples were genotyped using a capillary-based ABI PRISM® 3700 DNA Sequencer (Applied Biosystems). The two genotyping methods were checked for consistency and concordance by genotyping a random sample of CGN gel-scored accessions using the capillary method and by using three standard samples across all runs (Van Treuren *et al*. 2008). Five

individuals were included for each CGN *L. serriola* accession, and 2 individuals for each *L. sativa* accession as the crop accessions were expected to be more uniform than the wild accessions. Each ANGEL *L. serriola* collection site was represented by 30 individuals, which were all genotyped, and each site was considered as an ecological population. The loci amplified well in all the three data sets except for locus *LsD103* which had a poorer amplification in both ANGEL and CGN *L. serriola* samples than in CGN *L. sativa* samples, hence showing species-specificity and a potential for finding crop-specific alleles.

After genotyping, individuals with more than 50% missing data were removed, together with CGN *L. serriola* samples whose country of origin was not recorded in the CGN passport data. In total, 7738 individuals remained: 2456 ANGEL *L. serriola* samples, 2462 CGN *L. serriola* samples, and 2820 *L. sativa* samples. The ten markers used for genotyping resulted in 14 to 54 alleles per locus. The effective number of alleles per locus, calculated as $1/\sum P_i^2$, with P_i = allele frequency (Storme *et al*. 2004), ranged from 4 to 14 (Table 1), as many alleles were rare in both wild and cultivated lettuce.

Table 1 Information on the SSR markers used for genotyping

Locus[1]	Repeat motif	Linkage group	Observed number of alleles	Effective number of alleles
LsA001	$(GA)_{44}(GT)_{11}$	1	51	12.67
LsA004	$(GA)_{19}(GT)_7(GAGT)_4(GA)_{10}(GAGT)_2\ (GA)_{21}(GT)_{12}$	6	27	8.60
LsB101	$(GT)_{12}(AT)_5(GT)_{17}$	2	31	7.72
LsB104	$(GA)_5(GT)_7TATT(GT)_{12}\ (T)_4(GT)_8(GA)_{11}$	4	36	7.71
LsD103	$(TCT)_{17}$	9	14	5.56
LsD106	$(TCT)_{17}(T)_5(TCT)_2$	5	16	5.64
D108	$(TCT)_{35}$	4	48	11.65
D109	$(TCT)_{22}$	8	34	14.42
LsE003	$(TGT)_{24}(TA)(TGT)_{10}\ (TAT)_2(TGT)_3$	7	24	4.07
E011	$(TGT)_{26}$	3	24	4.61

[1] Originally described by van de Wiel *et al.* (1999)

Data analysis

Determination of population structure using STRUCTURE and InStruct

Analysis for population structure was performed on the 7738 individuals using STRUCTURE (Pritchard *et al.* 2000) version 2.2 (Falush *et al.* 2007). It uses a model-based Bayesian method to cluster the plant samples in a number of clusters (K) based on their genotypes. The ancestry model was set to admixture with correlated allele frequencies and lambda 1.0. No prior population information was used in the analysis. After a number of combinations of burn-in and

Markov Chain Monte Carlo (MCMC) runs, the final number of runs was chosen so that the differences in likelihood for each run [lnP(D)] between different K's was larger than the variability within runs of the same K. The length of the burn-in period was set to 80,000 with 150,000 MCMC replication runs after burn-in. In order to identify potential crop-wild hybrids, STRUCTURE should correctly differentiate *L. serriola* from *L. sativa* based on the 10 SSRs. To avoid any bias, we did not impose two clusters (K=2) on the program, but we let it run from K = 1 to 35 to check whether it would differentiate the two species or come to alternative subdivisions. The median of lnP(D) for six iterations of each K was considered (Saisho and Purugganan 2007), and the optimum number of clusters was determined by looking at the value of K with the highest likelihood (Pritchard *et al.* 2000) and the K value where the maximum number of information was gained in the analysis, i.e. where the lnP(D) value increased most from one K to the next (Evanno *et al.* 2005). To differentiate non-admixed and admixed (potentially hybrid) plants, we used a threshold posterior probability (Q value) of 0.90. Plants with Q value equal to or greater than 0.90 were considered as non-admixed; and those with Q value smaller than 0.90 were considered as admixed or potential hybrids (Vähä and Primmer 2006; Burgarella *et al.* 2009).

The data were also analysed with InStruct (Gao *et al.* 2007), a Bayesian-based program similar to STRUCTURE but specifically written for selfing species to account for divergence from the Hardy–Weinberg equilibrium due to self-fertilization. The length of the burn-in period was set to 100,000 and 200,000 MCMC runs after burn-in with K=1 to K=10 and five iterations for each K. The optimum number of clusters and the classification of individuals as non-admixed or hybrid were done as described above for the STRUCTURE results.

Inference of the hybrid generations using NewHybrids

NewHybrids (Anderson and Thompson 2002) is Bayesian model-based software to calculate the posterior probability that each plant belongs to a certain category of parents or hybrids based on the genotypic information of the plants. We used NewHybrids version 1.1 to infer the generations of the hybrid plants identified by STRUCTURE. Because of the limited capacity of the software, only a subset of the samples analysed with STRUCTURE was used with NewHybrids, namely all the hybrids identified by STRUCTURE and a randomly chosen set of non-admixed plants from the two *L. serriola* datasets and *L. sativa* plants. These were 706 ANGEL *L. serriola*, 617 CGN *L. serriola* and 677 *L. sativa* individuals, totalling 2000 individuals. *L. serriola* and *L. sativa*

being basically self-pollinating species, hybrid plants were expected to belong to advanced selfing generations after either one cross between the two species or one back-cross to any of the two parents. Therefore, the categories considered here were non-admixed *L. serriola* and non-admixed *L. sativa* considered as the parents of the hybrids (Parent 0 and Parent 1 respectively), early and advanced generations of selfing after one cross between the two parents (F1 and F2 as early generations and F7 as advanced generation) and in early and advanced generations of selfing after one backcross to either of the two parents (BC1 and BC1S1 as early generations and BC1S7 as advanced generation; Figure 1). The advanced hybrid generations should not be considered literally but as representative of various advanced inbred generations, as these cannot be distinguished reliably by the program in a dataset with only ten markers, due to little change in heterozygosity from one advanced generation to another. We ran NewHybrids using the uniform prior for both θ and the mixing proportion π and the program was left to run for 900,000 sweeps after burn-in. Because NewHybrids is less sensitive than STRUCTURE in differentiating between non-admixed and admixed individuals (Vähä and Primmer 2006), a threshold posterior probability (*P* value) of 0.70 was used to categorize an individual as belonging to a specific group.

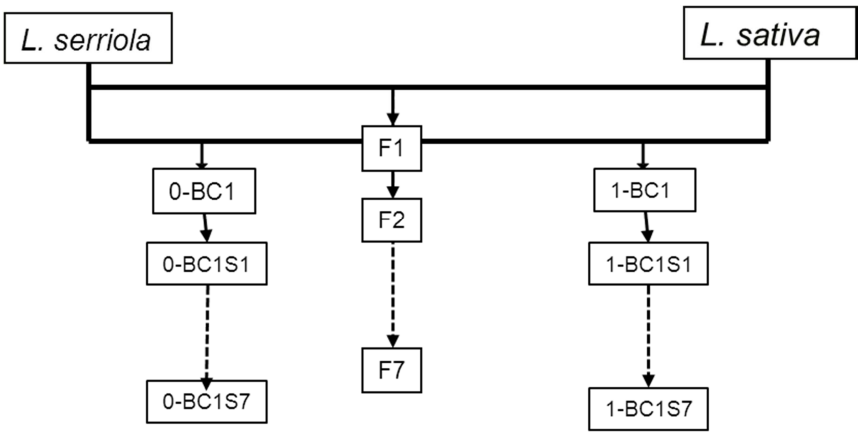


Figure 1 Hybrid classes used in NewHybrids: the hybrids were categorized into early and advanced generations of selfing after one cross between the two parents and early and advanced generations of selfing after one backcross to either of the two parents, with the advanced generations representing the inbred generations. Advanced inbred generations cannot be differentiated due to limited change in heterozygozity from one generation to the next.

Is crop-wild hybridization the cause of the spread of L. serriola in Europe?

To test whether crop-wild hybridization is the cause of the northward spread of *L. serriola* in Europe, a Pearson Chi-square test of independence was run on the most recent collection of *L. serriola* in Europe (ANGEL data set), testing whether the occurrence of the hybrids depended on

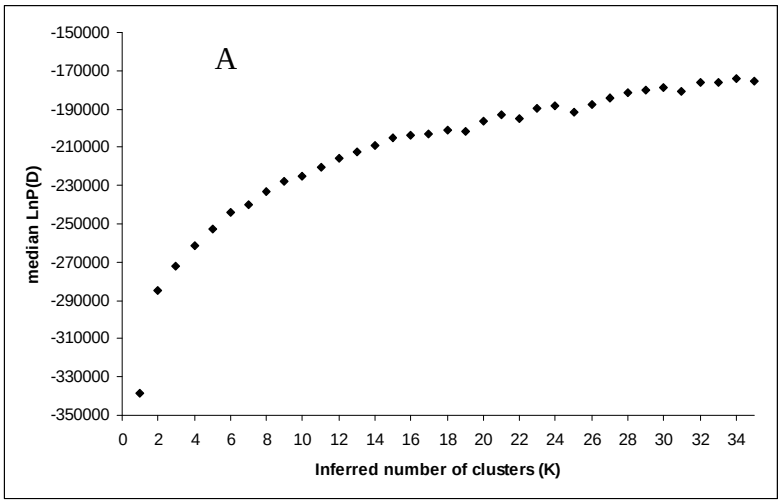
the region where the samples were collected. The origin of the samples was categorized in two groups based on the plant geographical regions of Europe (Frietema de Vries *et al.* 1994; Schaminée *et al.* 1992). The Southern region was represented by Portugal, Spain, south of France (below 45º of latitude), Italy, Switzerland, Hungary and Austria. The Northern region consisted of the remaining part of France, Luxembourg, the Netherlands, Germany, Czech Republic and Denmark. If crop-wild hybridization is responsible for the northward spread of *L. serriola*, we expect the Chi-square test to show a dependence between the occurrence of hybrids and the region (North and South) where the samples were collected and a bigger proportion of hybrids compared to non hybrids should be found in the Northern region.

Results

Distinction between wild L. serriola *and cultivated* L. sativa *by STRUCTURE*

With the STRUCTURE analysis, lnP(D) gradually increased and no clear peak was reached up to K=35 (Figure 2A). Therefore, the choice of the number of clusters, K, was not based on the highest lnP(D) value but on the value of K where the maximum information was gained in the analysis (Evanno *et al*. 2005), which was from K=1 to K=2 (Figure 2A). At K=2, STRUCTURE well differentiated *L. sativa* from *L. serriola*.

The plants with high posterior probability (>0.90) to one of the two groups coincided with *L. serriola* and the plants with high posterior probability to the other group coincided with *L. sativa*. "Admixed" plants with intermediate probabilities to both groups were considered as potential hybrids (Figure 3). To check whether the differentiation achieved by STRUCTURE between *L. serriola* and *L. sativa* was consistent, we checked K values larger than 2. At K=3, *L. serriola* remained distinct from *L. sativa*, with the third cluster arising by a split of the *L. sativa* cluster. At K=4, *L. serriola* samples were again clearly distinct from *L. sativa* samples, with both *L. sativa* and *L. serriola* being split into two clusters (Supplementary material, Figure S1). However, these clusters did not coincide with any recognizable biological or geographical group.



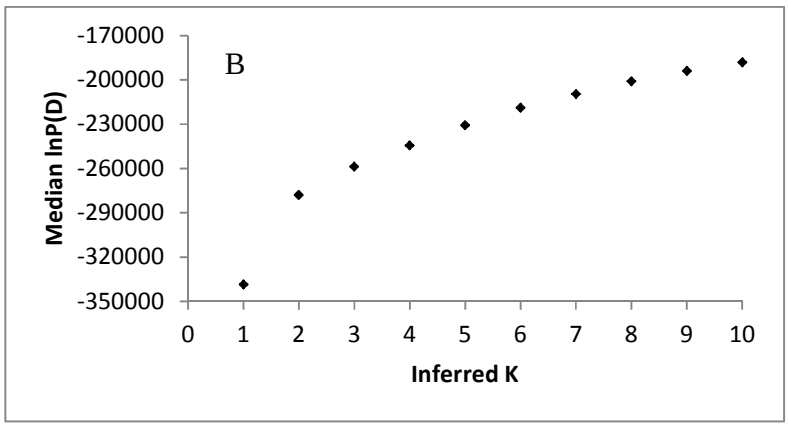


Figure 2 Analysis with STRUCTURE and InStruct. (A) STRUCTURE lnP(D) median as a function of the number of inferred clusters, K, up to K=35. (B) (C) InStruct lnP(D) median as a function of the number of inferred clusters K, up to K=10.

Occurrence of "admixed" (hybrid) plants

At K = 2, potential hybrids had intermediate probabilities to the two clusters that coincided with *L. serriola* and *L. sativa*. Because the two clusters mirrored each other (as a Q value of 0.90 for one cluster is equivalent to 0.10 for the other), samples with Q values smaller than 0.10 were regarded as non-admixed *L. sativa* and samples with Q values greater than 0.90 as non-admixed *L. serriola* in the remainder of this study. Ninety-three per cent of ANGEL *L. serriola* and 87% of CGN *L. serriola* individuals clustered in these groups, resulting in 7% potential hybrids (181 plants) among the ANGEL *L. serriola* individuals and 13% (312 plants) in the CGN *L. serriola* dataset (Figure 4A). The CGN dataset not only had a greater proportion of admixed individuals, but its putative hybrids also showed a more extended Q-value range (0.13 to 0.90) than the putative hybrids in the ANGEL dataset (0.32 to 0.90).

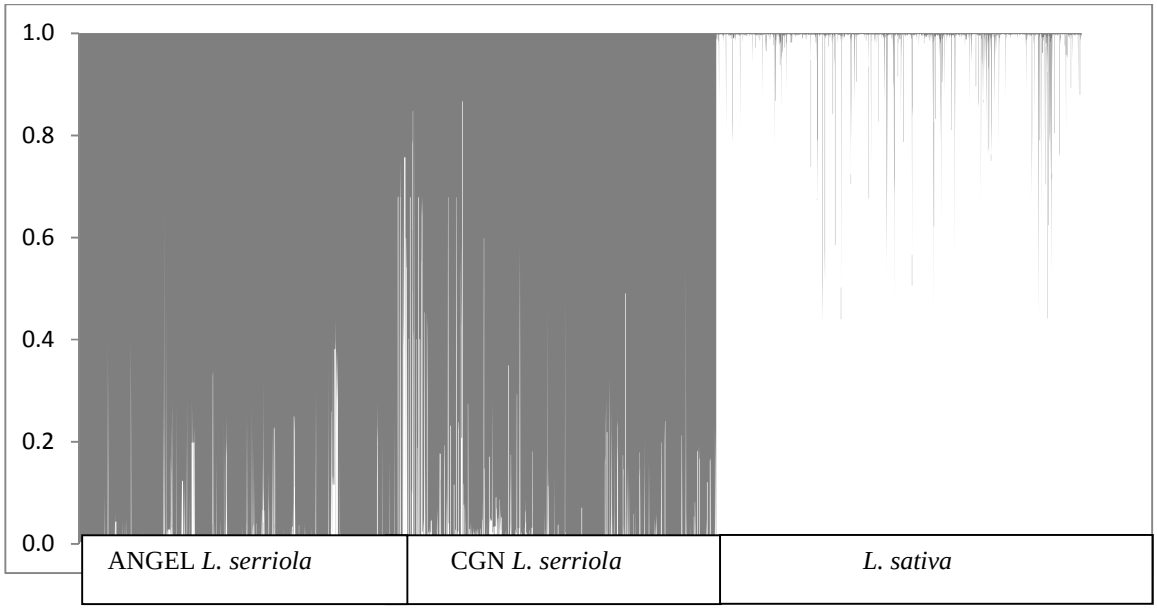


Figure 3 Q graph of all the samples at K=2. The Y axis represents the Q-values and the X axis represents the plants: each plant is given a membership proportion into two groups (grey and white), corresponding to *L. serriola* and *L. sativa*, respectively. Putative hybrids are discernible as partly belonging to both parental groups, *L. sativa* (right) and the combined data-sets of *L. serriola* (left).

On closer scrutiny of the CGN passport data the origin of the CGN accessions proved to be the major cause of the differences between the two datasets. Even though we had removed the samples whose origin was completely unknown before performing the analysis, CGN *L. serriola* hybrids with the lowest Q-values (0.13 to 0.35, so most "L. *sativa*"-like) were not obtained directly from their original habitats, but through research institutions or botanical gardens (Lebeda *et al.* 2004a). Such accessions have been shown to deviate genetically from material with an established origin in the same region, and some were even genetically identical to accessions from botanical gardens in other, distant, countries (Van de Wiel *et al.* 2010). By excluding all accessions without clearly established origins, the CGN *L. serriola* dataset became more similar to the ANGEL dataset: the lowest Q-value for CGN *L. serriola* hybrids increased to 0.40, and the proportion of hybrids dropped to 9% (Figure 4B). Among *L. sativa* 5% (147 plants, Figure 4C) clustered as putative hybrids, with Q values ranging from 0.11 (close to non-admixed *L. sativa*) to 0.57.

InStruct gave similar results as STRUCTURE. LnP(D) did not show any peak up to K=10 and the maximum information was gained from K=1 to K=2, making K=2 the optimum number of clusters (Figure 2B). At a threshold posterior probability of 0.90, the two programs classified

98% of the plants in the same categories (Table 2). The two per cent dissimilarity arose from InStruct identifying more *L. serriola* hybrids (0.16% among ANGEL and 3.25% among CGN *L. serriola*) and fewer *L. sativa* hybrids (0.39%) than STRUCTURE. These dissimilarities between InStruct and STRUCTURE were due to small differences in Q-values which ranged from 0.01 to 0.08. The results by STRUCTURE were more conservative, and hence we used them for further analysis (Arrigo *et al.* 2011).

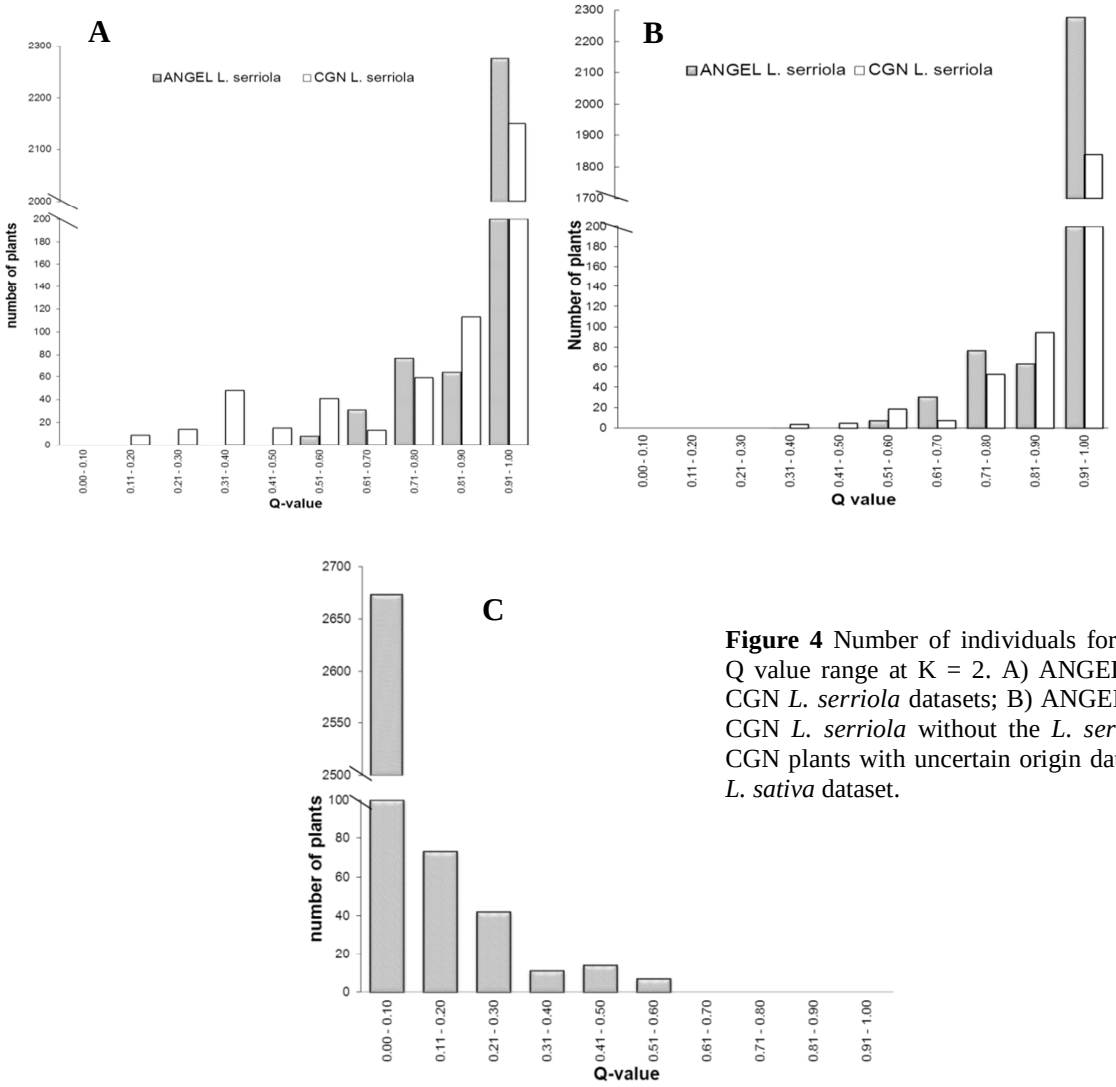


Figure 4 Number of individuals for each Q value range at K = 2. A) ANGEL and CGN *L. serriola* datasets; B) ANGEL and CGN *L. serriola* without the *L. serriola*-CGN plants with uncertain origin data; C) *L. sativa* dataset.

Table 2 Comparison between the classification of hybrids by STUCTURE and InStruct: the two programmes categorize 98% of the data in the same classes of non-admixed and potential hybrids

InStruct / STRUCTURE	ANGEL *L. serriola*		CGN *L. serriola*		*L. sativa*		Total STRUCTURE
	non-admixed	hybrids	non-admixed	hybrids	non-admixed	hybrids	
ANGEL *L. serriola*							
non-admixed[1]	2270	5					2275
hybrids	1	180					181
CGN *L. serriola*							
non-admixed			2059	91			2150
hybrids			11	301			312
L. sativa							
non-admixed					2654	19	2673
hybrids					30	117	147
Total InStruct	2271	185	2070	392	2684	136	7738

1 non-admixed: Q>0.90; hybrids: Q≤0.90

In Europe, the putative *L. serriola* hybrids were more frequent in the South (Figure 5). In the ANGEL dataset, most of the putative hybrids were found in Spain, Portugal, Italy and southern France: 141 out of the 181 ANGEL potential hybrids came from this region. Q-values in the region were as low as 0.32. In the more northerly country of the Netherlands, only 10 out of 152 samples (6%) were putative hybrids (Q-values 0.70-0.90). The 28 plants collected in the direct vicinity of amateur gardens did not indicate any increased likelihood of hybridization, as only 2 of these plants were identified as hybrids (with Q=0.70). These represented 7% of the hybrid occurrence, which was similar to that of the randomly sampled populations. Taken together, the proportion of hybrids compared to non-admixed individuals was 10% in the Southern region, while it was 2% in the Northern region. A Chi-square test of independence showed that the occurrence of the hybrids differed between these regions (P<0.001, Table 3). This indicates that crop-wild hybridization is not a likely cause of the spread of *L. serriola* in Northern Europe.

Table 3 Contingency table for the Chi-square test of independence between the occurrence of hybrids among ANGEL *L. serriola* individuals and the region where the samples were collected

		Hybrids	Non-admixed	Total
South	count	151	1135	1756
	expected	96	1190	
North	count	30	1119	700
	expected	85	1064	
Total		181	2277	2456
Chi-square test of independence				73.52 (P<0.001)

In the CGN *L. serriola* dataset, there was no difference in hybrid occurrence between Europe and the Middle East (9% and 10%, respectively). These figures could be biased due to the

overrepresentation of accessions from Europe in the genebank (Lebeda *et al.* 2004a). Outside of Europe, most of the *L. serriola* individuals from the wild habitats were collected from Israel and Turkey (89% of all Middle East individuals, Figure 5). The occurrence of hybrids in these two countries was 9%, which reflected the same pattern of hybrid occurrence as in the whole region.

Inferences about likely generations of putative hybrid plants by NewHybrids

NewHybrids classified non-admixed plants as *L. sativa* and *L. serriola*. Admixed plants belonged to advanced selfing generations after one hybridization event between *L. serriola* and *L. sativa* (represented by F7) and to advanced selfed generations after one backcross to either *L. serriola* or *L. sativa* (represented by 0_BC1S7 and 1_BC1S7, Figure 6).

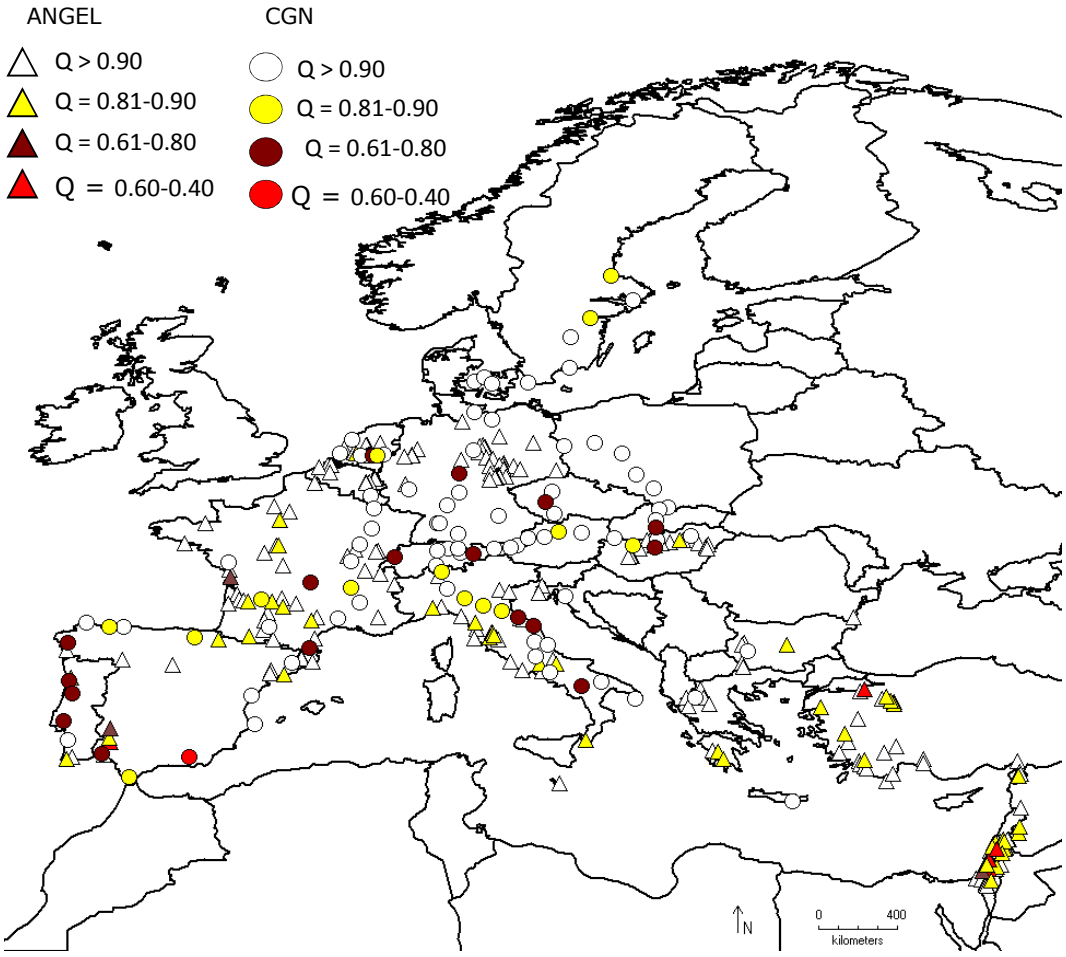


Figure 5 Geographical origin of non-admixed (white triangles and circles) and putative hybrid *L. serriola* plants (coloured triangles and circles) as identified by STRUCTURE in ANGEL and CGN data sets. At each location the lowest Q value is represented.

Other classes were not represented at all or were represented by probabilities smaller than 0.006. At a threshold P value of 0.70, NewHybrids classified 1265 of the 2000 plants in one of the five categories. The remaining 735 had probabilities divided between two or three categories (Table 4).

We compared the NewHybrids and STRUCTURE results (Table 4). NewHybrids recognized as hybrids all 181 ANGEL *L. serriola* plants identified as potential hybrids by STRUCTURE, and 97% of the 312 CGN *L. serriola* plants that STRUCTURE identified as potential hybrids. For *L. sativa*, NewHybrids recognized as hybrids 99% of the plants that STRUCTURE identified as hybrids. Conversely, NewHybrids also classified many of the STRUCTURE non-admixed plants as hybrids. Ten per cent of the ANGEL *L. serriola* plants identified as non-admixed by STRUCTURE were recognized as either hybrid or *L. serriola* (with higher probabilities for the *L. serriola* class, $P>0.45$), and 8% were classified as hybrids by NewHybrids. Of the CGN *L. serriola* plants identified by STUCTURE as non-admixed, NewHybrids classified 21% as undecided between non-admixed *L. serriola* and hybrids, and 34% as hybrids. Of the *L. sativa* plants identified by STRUCTURE as non-admixed, NewHybrids classified 10% as either non-admixed or hybrids, and 13% as hybrids.

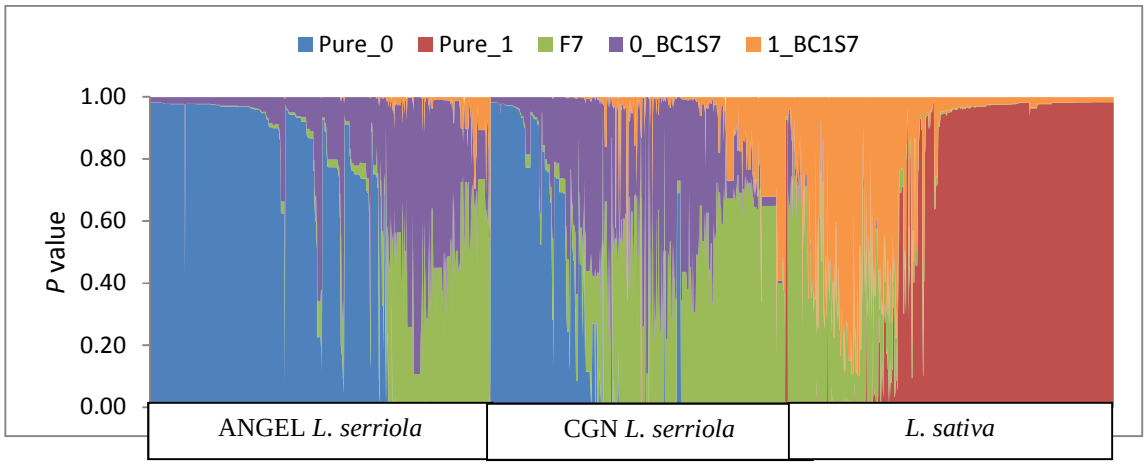


Figure 6 Average category probabilities of all analysed plants by NewHybrids. Each class is represented by a colour; on the X-axis are individual plants and the Y-axis represents the probability values, which add up to 1. The plants categorise as non-admixed *L. serriola* and *L. sativa* and advanced hybrid generations. The remaining classes included in the analysis were not represented at all or were represented by very small probability values (P<0.006).

Comparison between STRUCTURE and crop-specific allele method in identifying hybrids

To assign alleles as "crop-specific", we used the following criteria: 1) the frequency of the allele among *L. sativa* individuals is at least an order of two magnitudes higher than its frequency in the

L. serriola datasets (restricted to accessions with confirmed origin data for the CGN dataset) and 2) to attain a fair level of representativeness, the putative "crop-specific" alleles should occur in more than 10% of the accessions of the *L. sativa* dataset. Only 6 alleles from 5 loci conformed to our criteria of "crop-specificity" (*LsA001*-187, *LsD103*-263, *LsD103*-266, *LsD106*-191, *LsE003*-206 and *E011*-251); an additional 2 alleles (*D109*-251 and *E011*-254) conformed in the ANGEL *L. serriola* set only (Table S1, Supplementary material). For the CGN *L. serriola* set, both sets of alleles matched these criteria only when accessions from Europe were exclusively taken into account. The considerably higher frequency of the *D109*-251 and *E01*1-254 alleles in accessions from the Middle East and Central Asia could be related to this area being the most likely centre of origin of cultivated lettuce.

Table 5 shows a comparison of hybrid identification results from these crop-specific alleles with those from STRUCTURE. The number of plants with crop-specific allele(s) was higher in STRUCTURE hybrids than in STRUCTURE non-admixed L. *serriola* plants. However, the crop-specific allele method identified only 9% of ANGEL and CGN *L. serriola* hybrids plants, significantly fewer than STRUCTURE ($P_{\chi 2}<0.001$). This is a very small number, even when taking into account that our necessarily strict criteria for when an allele could be considered as crop-specific, was expected to lead to conservative estimates of hybrids. Limiting the use of the crop-specific alleles to Europe, which enables using all 8 alleles of Table S1 (Suppl. material), as in the ANGEL dataset, does not really change this situation, except for the absolute numbers (25 vs. 16 putative hybrids conforming to STRUCTURE and 16 vs. 8 not confirming to STRUCTURE, see Table 5).

Table 4 Comparison between NewHybrids and STRUCTURE results: NewHybrids classifies the hybrids as 7th generation of selfing after the initial cross between *L. serriola* and *L. sativa* (F7) or the 7th generation of selfing after one back-cross to either *L. serriola* (0-BC1S7) or *L. sativa* (1-BC1S7) (see Figure 1 for an overview of classes). The '7th generation' represents advanced selfed generations.

NewHybrids / STRUCTURE	*L. serriola* (Parent 0)	*L. sativa* (Parent 1)	F7	0-BC1S7	1-BC1S7	F7 or 0-BC1S7	*L. serriola* or F7or 0-BC1S7	*L. serriola* or 0-BC1S7	F7 or 1-BC1S7	*L. sativa* or 1-BC1S7	*L. sativa* or F7 or 1-BC1S7	Total STRUCTURE
ANGEL *L. serriola*												
Non-admixed	434			7		33	16	34				524
Hybrids			36	40		97			8			181
CGN *L. serriola*												
Non-admixed	135			7		89	25	40	9			305
Hybrids	2	7	40	33	2	129		8	91			312
L. sativa												
Non-admixed		411	4		25				39	42	9	530
Hybrids		1	29		52	6			59			147
Total NewHybrids	571	419	109	87	79	355	41	82	206	42	9	2000

Table 5 Frequency of crop-specific alleles[1] among *L. serriola* datasets categorized as potential hybrids and non-admixed using STRUCTURE

STRUCTURE groups	**ANGEL *L. serriola***	**CGN *L. serriola* from Europe**	**CGN *L. serriola* from outside of Europe**
Hybrids (Q≤0.90)			
Frequency STRUCTURE	181	84	93
Frequency of plants containing at least one crop-specific allele	16	2	14
Non-admixed (Q>0.90)			
Frequency STRUCTURE	2275	1105	831
Frequency of plants containing at least one crop-specific allele	8	1	2
Total	2456	1189	924
Chi-square value for goodness of fit between STRUCTURE and crop-specific alleles	150.44 (P<0.001)	80.04 (P<0.001)	67.11 (P<0.001)

[1] crop-specific alleles used: LsA001-187, LsD103-263, LsD103-266, Ls D106-191, E003-206, E011-251

Discussion

Even though crop-wild introgression is nowadays accepted as a common phenomenon, it is mostly recognized among cross-pollinating species, such as carrots (Magnussen and Hauser 2007; Rong *et al.* 2010), sunflower (Arias and Rieseberg, 1994; Whitton *et al.* 1997) and chicory (Kiær *et al.* 2009). In self-pollinating species with restricted levels of cross-pollination such as lettuce it is expected to occur (Ellstrand 2003, D'Andrea *et al.* 2008), but rarely and difficult to detect. Nevertheless, we found an occurrence of 7% of putative *L. sativa – L. serriola* hybrid plants from the wild habitats of *L. serriola* in Europe (recently sampled wild populations) and 9% from *L. serriola* accessions present in the CGN genebank collection. The identification of lettuce crop-wild hybrids in natural wild population implies that *L. serriola* does hybridize with *L. sativa*, and that the hybrid lineages persist along with *L. serriola* non-admixed plants. These results are different from those found in soybean (Kuroda *et al.*, 2010), which is a basically self-pollinating species as well: although these authors found evidence for crop-wild hybridization, the hybrids did not persist in the natural habitats of wild soybean. Our results are consistent with previous studies in lettuce which showed that *L. sativa-L. serriola* hybridization produces some hybrid lineages more vigorous and fit that the wild parent (Hooftman *et al.* 2005), and that the persistence of the hybrids depends on their relative fitness and the species outcrossing rate (Hooftman *et al.* 2007).

L. sativa and *L. serriola* are so closely related that some studies have labelled them as conspecific (Koopman *et al.*, 2001). Despite this close relatedness, using the ten SSR markers, STRUCTURE and InStruct differentiated the two species and identified intermediate plants, which were potential hybrids. Simko and Hu (2008) obtained similar results using STRUCTURE on a smaller set in which they could distinguish cultivated (*L. sativa*) from two wild lettuce species (*L. serriola* and *L. saligna*). The use of large datasets may improve the power and accuracy for the identification of hybrids (Burgarella *et al*. 2009), as shown here. NewHybrids recognized nearly all hybrids detected by STRUCTURE, but also several putative hybrids among the other lines. Vähä and Primmer (2006) encountered the same trend with NewHybrids, as the program classified some non-admixed individuals as hybrids. As, in addition, NewHybrids could not handle all the available data of our dataset, NewHybrids results were used here solely for the determination of the hybrid classes. NewHybrids categorized the *L. serriola* hybrid plants as identified with STRUCTURE into two hybrid classes: advanced selfed generations after hybridization between *L. serriola* and *L. sativa* (represented by F7), and advanced selfed

generations after one back-cross to *L. serriola* (represented by 0-BC1S7). In a fine-scale field study of the self-pollinating species *Medicago truncatula*, Siol *et al*. (2008) found comparable results: many of the genotyped plants represented recombinant inbred lines (advanced selfed generations after a hybridization event) between the most frequently occurring highly inbred lines. In studies on the detection of spontaneous hybrids in perennial, cross-pollinating woody species, the identified hybrids usually belonged to early hybrid generations such as F_2, BC_1 and BC_2 (Smulders *et al*. 2008; Schanzer and Kutlunina 2010).

The higher frequency of hybrids found in the southern part of Europe for both *L. serriola* datasets could be related to the occurrence of seed multiplication in open air in the Mediterranean area (e.g. Portugal, Spain and Italy, in particular the Emilio-Romagna region), which is rare (or mostly under glass) in the north. On the other hand, the small subset of samples taken in the Netherlands near amateur gardens, one of the few places in the north where bolting and small-scale seed multiplication might occur, did not show a higher likelihood of hybrid occurrence as compared to randomly sampled populations. At such sites, bolting may only be haphazard, but it cannot be excluded that this is related to a possibly lower cross-fertilization rate under climatic conditions of northern countries.

The posterior probability approach of STRUCTURE has shown to be a good tool to identify gene flow between closely related species using molecular markers for a wide array of organisms, such as in animals, *e.g*. the carnivorous marsupial *Antechinus flavipes* (Lada *et al.* 2008), in trees, *e.g.* various oak species (Burgarella *et al.* 2009), and also for gene flow between crop and wild forms in, for instance, alfalfa (Greene *et al.* 2008), beet, *Beta vulgaris* (Andersen *et al.* 2005), and chicory, *Cichorium intybus* (Kiær *et al.* 2009). The fact that the software uses genotypic information encompassing all the scored alleles and their frequencies enables it to obtain a more comprehensive picture of the individuals' genetic make-up, without any previous bias of *a priori* grouping information or alleles identified as specific for any of such groups. Indeed, our trial of using the crop-specific allele approach did not work well. At best, it may lead to a conservative estimate of hybrids which logically followed from our necessarily strict definition of "crop-specific" alleles, that is, only 6 alleles from 5 loci out of a total of 315 alleles from 10 loci could at most be used as such. Moreover, about a third of the hybrids indicated by the crop-specific alleles were identified as non-admixed plants by STRUCTURE. This could be attributed to small introgressions containing only one of the crop-specific alleles that were in the "noise" range of the more comprehensive analysis of STRUCTURE, but it could also be due to rare coincidental

occurrences of the allegedly crop-specific allele in non-admixed wild lettuce. Indeed, recent studies using "crop-specific" alleles often targeted more local situations with known combinations of crop cultivations and wild populations in the vicinity (*e.g*. Morrell *et al.* 2005 on introgression of sorghum into Johnson grass) or more widely different species combinations (e.g. Schulze *et al*. 2011 on garden strawberry *Fragaria* x *ananassa* and wild woodland strawberry *F. vesca* in Central Europe). Smulders *et al.* (2008), Rathmacher *et al.* (2010) and others successfully used species-specific alleles to detect gene flow and identify F1 hybrids between poplar species and hybrids. However, while crop-specific alleles are very effective in first generation hybrids, their power is lost in selfing and backcross generations, as each generation 50% of the offspring will by chance not inherit the allele and become indistinguishable from non-introgressed plants. Thus, while useful for detecting introgression in outcrossing, long-lived perennials, crop-specific alleles are not very effective in selfing annuals in which introgression is present in advanced inbred lines in the field.

Introgression from crops to wild relatives has been connected to the invasiveness of some wild species such as Johnson grass (*Sorghum halepense*) (De Wet and Harlan 1975), *Rhododendron ponticum* (Milne and Abbott 2000), and sunflower (Rieseberg *et al.* 2007). Hooftman *et al*. (2006) suggested that introgression from *L. sativa* to *L. serriola* could be one of the reasons behind the recently observed increase of the latter in Europe, whereas D'Andrea *et al*. (2009) argued that this spread may be attributed mainly to the expansion of the favourable habitat due to climate warming and anthropogenic habitat disturbance, and to seed dispersal due to transportation of goods. The results of this study do not support the hypothesis of Hooftman *et al*. (2006). If introgression were behind the spread of *L. serriola*, we would expect to find more putative *L. serriola* hybrids than non-admixed *L. serriola*, particularly in North-Western Europe where the new invasiveness of *L. serriola* was most obvious. Moreover, we would also expect to observe more hybrids among the more recently collected *L. serriola* ANGEL data set (collected between 2002 and 2005) than in the mostly older CGN genebank collection. Both expected patterns were not visible in our data. Although a number of putative *L. serriola* hybrids were found with STRUCTURE, these did not constitute the dominant proportion of the *L. serriola* plants, neither in the ANGEL nor in the CGN data set. Moreover, hybrids were particularly rare in northern Europe. Hence, we found no evidence that crop introgression conferred an increased invasiveness to wild lettuce. Therefore, the expansion of *L. serriola* in Europe and in the Netherlands in particular resulted most likely from the combination of factors indicated by

D'Andrea *et al.* (2009). Nevertheless, with lettuce being a basically self-pollinating species, the occurrence of 7% of crop-wild hybrids among natural *L. serriola* populations is relatively high and reveals a potential of transgene movement from crop to wild relatives also for self-pollinating crops. After hybridization, however, the fate of the transgene will depend on many factors including the survival and fertility of the hybrids, the fitness effect of the transgene, and the relative fitness effect of the genomic region where the transgene is inserted (Stewart *et al.* 2003). The fitness effects of the genomic background in relation to environmental conditions is the topic of on-going experimental and modelling research in a joint project of Wageningen UR, University of Amsterdam and Groningen University in the Netherlands.

Acknowledgements

Data for this publication were obtained in the context of the EU projects "Analysis of gene flow from crop to wild forms in lettuce and chicory and its population-ecological consequences in the context of GM-crop biosafety" (ANGEL, QLK3-CT-2001-01657) and "Molecular markers for genebanks: Application of marker technology for the improvement of *ex situ* germplasm conservation methodology" (PL96.2062). The data analysis was financially supported by the Netherlands' Organization for Scientific Research (NWO) through their research program "Ecology Regarding Genetically modified Organisms" (ERGO).

Chapter 3

Hybridization between crops and wild relatives: the contribution of cultivated lettuce to the vigour of crop-wild hybrids under drought, salinity and nutrient deficiency conditions

Brigitte Uwimana[1], Marinus J.M. Smulders[1], Danny A.P. Hooftman[2], Yorike Hartman[3], Peter H. van Tienderen[3], Hans Jansen[4], Leah K. McHale[5], Richard W. Michelmore[6], Clemens C.M. van de Wiel[1], Richard G.F. Visser[1]

1 Wageningen UR Plant Breeding, Wageningen, The Netherlands
2 Centre for Ecology and Hydrology, Wallingford, UK.
3 IBED, Universiteit van Amsterdam, Amsterdam, The Netherlands
4 Wageningen UR Biometris, Wageningen, The Netherlands
5 Department of Horticulture and Crop Science, Ohio State University, Columbus, USA
6 Genome Center, University of California Davis, USA

Abstract

With the development and commercial release of transgenic crop varieties, crop-wild hybridization has received exceptional consideration due to the feared potential of transgenes to be transferred to wild species. Although many studies have shown that many crops can hybridize with their wild relatives and that the resulting hybrids may show improved vigour and fitness over the wild parents, little is still known on the genetic contribution of the crop parent to the performance of the hybrids. In this study we investigated the vigour of lettuce hybrids using 98 $F_{2:3}$ families from a cross between cultivated lettuce (*Lactuca sativa* L.) and its wild relative prickly lettuce (*L. serriola* L.) under non-stress and abiotic stress conditions of drought, salinity and nutrient deficiency. Using Single Polymorphism Nucleotide markers, we mapped quantitative trait loci associated with plant vigour in the $F_{2:3}$ families and determined the allelic contribution of the two parents.

The vigour traits showed mild to high broad sense heritability across the treatments (H^2 ranging from 0.51 to 0.99). Seventeen QTLs associated with vigour and 6 QTLs associated with the accumulation of ions (Na^+, Cl^- and K^+) were mapped on the 9 linkage groups of lettuce. Seven of the vigour QTLs had a positive effect from the crop allele and 6 had a positive effect from the wild allele across treatments, and 4 QTLs had a positive effect from the wild crop in one treatment and positive for the wild allele in another treatment. The dominance effect of the QTLs was not significant for 16 QTLs, while epistasis as non-additive interaction among the mapped QTLs played a significant role on the vigour of the hybrids.

Introduction

Gene flow between crop species and their wild relatives may result in the introgression of crop genes into wild genomic background, or in the formation of new species through novel combinations of crop and wild genes (Burke and Arnold 2001; Hails and Morley 2005). The possibility of hybridization between transgenic crops and their wild relatives has brought interest on crop-wild gene flow to another level due to the potential ecological consequences, incited by the possibility that transgenes could also be introgressed into wild populations (Hall *et al.* 2000; Snow *et al.* 2005; Tiedje *et al.* 1989; Warwick *et al.* 2009; Wilkinson and Tepfer 2009).

Gene flow can lead to hybrid plants containing crop alleles. However, if crop alleles are selectively neutral, many crop-wild hybridization events would be required to increase their frequency in the wild population. In order for the crop alleles to be established in the population of the wild relative with few crop-wild hybridization events, they have to provide a selective advantage to the fitness of the hybrid plants and their offspring (Lee and Natesan 2006). In the introgression and speciation processes, the unit of selection in the first generations of hybrids is not the crop gene as such, but genomic blocks from the crop consisting of the gene under selection and the surrounding linked genomic region (Stewart *et al.* 2003). Consequently, linkage between genes plays a crucial role in the introgression process, because a gene (or transgene) that has no effect on fitness may become introgressed just by hitchhiking along with a gene that increases fitness. Conversely, a (trans)gene could be selected against due to its proximity to a gene which reduces fitness. Such linkage would provide a natural mechanism against introgression and escape of transgenes into wild populations (Kwit *et al.* 2011; Stewart *et al.* 2003).

Multiple studies have focused on the rate of hybridization between crops and wild relatives, and on the occurrence of hybrids and their fitness in relation to the fitness of the wild parent (Arias and Rieseberg 1994; D'Andrea *et al.* 2008; Giannino *et al.* 2008; Hoc *et al.* 2006; Hooftman *et al.* 2009; Kiær *et al.* 2009). However, few studies have been conducted with the aim of understanding the specific contribution of the crop and wild parents to the fitness of the hybrids, the role of the genomic locations of the genes (as for instance assessed through quantitative trait loci, (Baack *et al.* 2008; Rose *et al.* 2009), and the role of epistasis and genotype by environment interaction on the fitness or vigour of the hybrids. Such knowledge could be exploited to identify crop genomic regions with a higher or lower likelihood of introgression into wild populations

(Kwit *et al.* 2011; Stewart *et al.* 2003), thus assessing the possibility of containing transgenes by utilizing integration events in regions with lower likelihoods of introgression.

The combination of synthetic mapping populations and genetic linkage maps provides an excellent tool for studying the introgression process in an experimental set-up. It allows the determination of quantitative trait loci (QTL) affecting hybrid vigour or fitness, estimation of the contribution of each parent to the performance of the offspring under controlled or non-controlled conditions, and monitoring of specific genomic blocks in different generations after hybridization (Baack *et al.* 2008; Burke and Arnold 2001; Rieseberg *et al.* 2000; Stewart *et al.* 2003).

In this study we investigated the contribution of the crop alleles to the performance of a crop-wild hybrid population derived from a cross between cultivated lettuce (*Lactuca sativa* L.) and wild prickly lettuce (*Lactuca serriola* L.). Cultivated lettuce and wild prickly lettuce are interfertile species whose hybrids are viable and fertile (De Vries 1990; Hooftman *et al.* 2005). Experiments have shown that lettuce crop-wild hybrids are more vigorous than their parents (Hooftman *et al.* 2007a; Hooftman *et al.* 2005) and that this increased vigour may lead to improved fitness of their offspring (Hooftman *et al.* 2009). In this study, we investigated the genetic basis of improved hybrid vigour of lettuce hybrid plants. Under natural conditions the hybrids will most likely be subject to adverse conditions of abiotic stress such as drought, heat, cold, etc. Genetic modification presents a lot of potential for breeding for abiotic stress tolerance (Christou and Twyman 2004; Wang *et al.* 2003; Zhang *et al.* 2000). Tolerance to abiotic stress factors is a prominent goal of today GM breeding and evaluation, and the release of GM crop varieties tolerant to the major abiotic stress is expected in the near future for many crop species (Abdeen *et al.* 2010; Choi *et al.* 2011; Li *et al.* 2010). Therefore, we conducted experiments under controlled abiotic stress conditions of drought, salinity and nutrient deficiency in the progeny of a cross between *L. sativa* and *L. serriola*. We addressed the following questions: (i) how is the performance of the hybrids relative to the wild parent under non-stress and stress conditions? (ii) do crop alleles contribute an advantage or disadvantage to the crop-wild hybrids under non-stress and abiotic stress conditions (drought, salinity and nutrient deficiency)? (iii) how are the vigour QTLs distributed along the genome, and what is the nature of their allelic effects?

Materials and Methods

Lactuca serriola and *L. sativa*

Lactuca serriola, known as prickly lettuce, is a weedy species that thrives in ruderal, anthropogenic areas (Lebeda *et al.* 2001). It is the closest relative of cultivated lettuce (*L. sativa*) with which it is part of the primary gene pool and considered to be conspecific with *(Koopman et al. 1998; Koopman et al. 2001)*. *L. serriola* and *L. sativa* form a classic crop-weed complex perfect for introgression studies. The two species have the same number of chromosomes (2n = 2x = 18), are completely cross-compatible without any known crossing barrier, and the resulting hybrids are also viable and fertile (De Vries 1990; Hooftman *et al.* 2005; Koopman *et al.* 1993; Lindqvist 1960). Both species are autogamous with a limited rate of out-crossing by insects of 1 to 5% for *L. sativa* (Thompson *et al.* 1958) and an interspecific hybridization rate of up to 2.5% between the two species (D'Andrea *et al.* 2008). In an experimental set-up, D'Andrea *et al.* (2008) concluded that whenever *L. serriola* and *L. sativa* populations grow in sympatric proximity, cross-pollination between the two species should be expected to occur.

Development of hybrid plants

F_1 progeny was created by crossing *L. serriola* and *L. sativa* in the greenhouse. *L. serriola* parent was a progeny of the genotype collected from Eys (Province of Limburg, the Netherlands) designated as "cont83" in the description of *L. serriola* genotype distributions in Europe, and it represents a commonly occurring genotype of *L. serriola* in North-Western and Middle Europe (Van De Wiel *et al.* 2010). For *L. sativa* parent, we used the commercial cultivar Dynamite, a butterhead lettuce developed by Nunhems Zaden. It harbours genes for resistance to aphids, downy mildew and lettuce mosaic virus (Van der Arend *et al.* 1999), which represent the main breeding goals of lettuce cultivars. *L. sativa* was used as the pollen donor, mimicking a scenario of pollen flow from a crop to its wild relative. Crossing was done according to the protocols by Nagata (1992) and Ryder (1999) and as described in (Hooftman *et al.* 2005). F_2 seeds were produced by selfing of one F1 plant. F_2 seeds were sown and 200 seedlings were randomly chosen, transplanted and genotyped as described below. The plants were selfed and the resulting F_3 seeds were harvested per individual F_2 plant.

Genotyping and construction of the linkage map

The Compositeae Genome Project at UC Davis Genome Center has developed Single Nucleotide Polymorphism (SNP) markers from lettuce populations derived from crosses between closely related cultivars of *L. sativa* and between *L. sativa* and *L. serriola*. These SNPs were mined initially by re-sequencing PCR-amplified genes of interest between *Lactuca sativa* cv. Salinas and *L. serriola* acc. UC96US23 using Sanger sequencing (McHale *et al.* 2009) and by mining Illumina sequencing data aligned to reference EST assemblies (http://compgenomics.ucdavis.edu/compositae_SNP.php). cDNA libraries from parental lines were sequenced with Illumina Genome Analyzer II. These ESTs sequences encode genes for disease resistance and plant development. In this way, more than 10,000 SNPs were developed from 3,950 ESTs in four parental pair combinations, namely Salinas x Valmaine, Pavane x Parade, Emperor x El Dorado, and Thompson x Cisco. To improve the conversion success rate of bioinformatically identified SNPs to molecular markers, potential SNPs were filtered to 1083 SNPs that had been previously assayed and shown to be robust; were polymorphic in more than one of the four parental pair combinations; were not located in intron/exon splice sites; were limited to one SNP per contig; were candidate genes of interest; were evenly distributed based on previous mapping work and the ultra-dense lettuce map; and the surrounding sequence was suitable for oligonucleotide design for the Illumina GoldenGate assay. The selected 1083 SNPs were converted into Custom GoldenGate Panels (OPA) for genotyping, using an Illumina BeadXpress . From the 1083 SNPs, a customized OPA of 384 SNPs which were polymorphic between the F_2 parental lines was made specifically for the population.

DNA was extracted from freeze-dried leaf samples of the 200 F_2 and parent lines using the QIAGEN DNeasy 96 Plant Kit (QIAGEN, Venlo, the Netherlands) with slight modifications for dry plant tissue to obtain a minimum DNA concentration of 60 ng/µl. The DNA concentration was quantified using a NanoDrop 1000 Spectrophotometer V3.7 (Thermo Scientific). We genotyped 187 F_2 individuals and the parents using the customized 384 SNP OPA in a BeadXpress assay. Out of the 384 SNPs, 355 were successfully scored in the 187 F_2 and parental lines. Three hundred thirty-one markers were co-dominant, 16 were dominant for the *L. serriola* allele, and 8 were dominant for the *L. sativa* allele. The genotypes for the 187 F_2 individuals were used to build a genetic linkage map using JoinMap® 4 (Van Ooijen 2006). Segregation distortion was tested against the expected allele frequency ratio of 1:1, using the χ^2 test of goodness of fit with one degree of freedom. Markers within linkage groups were ordered using the maximum

likelihood option of JoinMap (Jansen *et al.* 2001). The linkage map was displayed using MapChart 2.2 (Voorrips 2002).

Greenhouse experiments

Based on the genotypes of the 187 F_2 individuals, we selected a set of 98 F_2 plants that optimized the number of different combinations of parental haplotype blocks, using the program "Genetic Distance Optimization" (GDOpt) (Odong *et al.* 2011). The program uses adapted K-medoids clustering (Kaufman and Rousseeuw 1990) in which one individual in each of the K clusters acts as cluster centre and clusters are formed by minimizing the total distance of all individuals to the nearest of the K individuals designated as cluster centres. In order to obtain a good starting point, the initial configuration of cluster centers was provided by a modified version of Genetic Distance Sampling (Jansen and van Hintum 2007).

$F_{2:3}$ families were derived from the genotyped F_2 plants by selfing, together with their parents in greenhouse experiments were used in greenhouse experiments in Wageningen, the Netherlands. We added to the experimental lines two additional lines, *L. serriola* acc. UC96US23 and *L. sativa* cv. Salinas which, together with the parental lines, were later used to estimate the environmental error. We carried out two experiments: (i) the "drought experiment" (March-April 2010), which comprised drought and control treatments and (ii) the "salt-nutrient experiment" (June-July 2010), which comprised salt, nutrient deficiency and control treatments. Each F_2 plant was represented by 12 $F_{2:3}$ seedlings per treatment. The parents and the two additional lines were also replicated 12 times per treatment.

During first establishment, the seedlings were irrigated twice a week for two weeks with water supplemented with nutrients. Subsequently, the treatments were started at the beginning of the 3rd week after transplanting of the seedlings, when the plants had 4 to 5 leaves. For the drought experiment, the plants in the control treatment were still watered twice a week, while the plants in the drought treatment were not given water at all . For the salt-nutrient experiment the plants were again irrigated twice a week but with added 100 mM of NaCl in the irrigation water. The plants under nutrient deficiency treatment received water to which no nutrients were added. The control plants received nutrients for the whole period of the experiment. Stress was applied for three weeks after which time the plants were harvested at the rosette stage, 35 days after transplanting. A photoperiod of 18ºC/16 hours of light and 15ºC/8 hours of darkness was maintained throughout the experiments by temperature control and application of artificial

lighting as needed. However, high summer temperatures influenced the greenhouse conditions during the salt-nutrient experiment when outside temperature reached as high as 35^{o}C.

Phenotypic measurements

For each plant, vigour was determined by fresh and dry shoot biomass and shoot height. Shoot dry weight was measured after these were dried at 80ºC for 3 days. We also calculated shoot relative moisture content as the ratio of the amount of water in the shoot to the total shoot weight [(fresh weight-dry weight)*100/fresh weight]. The ion content (Na^+, K^+, and Cl^-) for salt and control treatments of the salt-nutrient experiment was measured. Because ion content is measured based on dry matter, the 80ºC-dried shoots were dried again at 100ºC for 24 hours. The 12 plants per line per treatment were pooled, ground to fine powder, well mixed, and about 30 mg of dry matter was measured with the precise weight recorded. The ground samples were ashed at 545ºC for 5 hours, diluted in 3M formic acid, and further diluted 1000 times with extra-pure water. The diluted solutions were used in ion chromatography analysis on an 881 Compact IC pro (Metrohm AG, Herisau, Switzerland, Stolte *et al.* 2011).

Analysis of phenotypic data

Statistical analysis was performed using GenStat 13 (Payne *et al.* 2010). The drought and the salt-nutrient experiments were analysed separately. The significance of the different terms was determined by the analysis of variance, fitting the model:

$$Response = general\ mean + block + genotype + treatment + genotype.treatment + error$$

Broad sense heritability of the traits was estimated for each treatment separately as the proportion of the total variance accounted for by the genetic variance using the formula

$$H^2 = Vg_{(F2)}/(Vg_{(F2)}+Ve/r);$$

where $Vg_{(F2)}$ is the genetic variance among $F_{2:3}$ families, *Ve* is the environmental variance, and *r* is the number of replications (Chahal and Gosal 2002). $Vg_{(F2)}$ was estimated based on the restricted maximum likelihood (REML) method from the mixed model:

$$Response = general\ mean + block + \underline{F_2\ genotype} + error;$$

with the *Response* term representing the measured traits, and the term F_2 *genotype* taken random. *Ve* was the error variance derived from a one-way analysis of variance of the model:

Response = general mean + block + parents + error;

with the term *parents* representing the two parents of the F_2 plants and the two added lines (*L. serriola* acc. UC96US23 and *L. sativa* cv. Salinas).

Quantitative Trait Loci analysis

In order to effectively model genotype by environment interaction (GxE, with environments represented by the different treatments) through QTL by environment interaction (QTLxE), each trait was analysed individually using the single trait – multiple environment option of the program. Genome-wide association between markers and traits was decided based on a significance level of 0.05 corrected for multiple tests using the Li and Ji method (Li and Ji 2005). After the selection of the best variance-covariance model for the treatments (Malosetti *et al.* 2004), the candidate QTLs were determined by initial genome scan. Final QTL positions were determined by composite interval mapping taking into account co-factors. The allelic effect of the detected QTLs in each treatment, the effect of QTLxE and the explained phenotypic variance of each QTL per treatment were determined by running a backward selection on the candidate QTLs in a mixed linear model, taking the QTL effect in each treatment as fixed terms and the interaction between each hybrid family and the treatment as random (Mathews *et al.* 2008). In that way, each QTL detected in one treatment was tested for its effect and significance in the other treatments.

Epistasis was tested for the detected QTLs (Holland 2007). Each QTL region was represented by the genotypic scores of the most significant marker in a multiple regression model in GenStat. To avoid the effect of linkage, overlapping QTLs were represented by one SNP marker and no interaction was estimated for QTLs on the same linkage group even if they did not overlap. In each treatment, every trait was explained by the main effects of all the detected QTLs to which interaction between one pair of QTLs was added at a time. QTL x QTL interaction was decided significant at a level of 0.05 which was corrected for the number of traits by the Bonferroni method (Bland and Altman 1995).

Results

Phenotypic variation

The analysis of variance revealed significant genotypic variation for the measured vigour traits (plant height, fresh weight, dry weight and relative moisture content; $P_{genotype} < 0.001$), and there was significant genotype x treatment variation ($P_{genotype\ x\ treatment} < 0.001$). Broad sense heritability of the traits ranged from moderate to high ($0.51 \leq H^2 \leq 0.99$, Table 1), showing that the phenotypic variation among the $F_{2:3}$ families was mainly explained by genetic factors. Heritability depended on the treatment. In the drought experiment, it was lower for all the traits under the drought conditions than under the control conditions, except for relative moisture content. In the salt-nutrient experiment, the heritability of the traits was comparable under control and salt treatments, whereas it was lower under nutrient deficiency treatment. Crop-wild hybridization released genetic variance: even when the means of the parents were not significantly different, heritability was relatively high as observed for dry weight under control ($H^2 = 0.90$) and drought conditions ($H^2 = 0.66$) and for relative moisture content under nutrient deficiency conditions ($H^2 = 0.89$) (Table 1).

For each trait and under all the treatments, there were $F_{2:3}$ individuals whose measurements were equal to or greater than the means of the two *L. serriola* lines (Supplementary material Figure S2). Moreover, the mean for *L. serriola* parent was always comprised within the range of the means of $F_{2:3}$ families for all the traits and under all the treatments (Table 1). It is evident from these results that crop-wild lettuce hybrid families have potential increased vigour in comparison to the wild parent under the four tested conditions (non-stress, drought, salt and nutrient deficiency conditions).

Plant height positively correlated with biomass, except under salt treatment where fresh weight was negatively correlated with plant height (r=-18, Table 2). Under salt treatment, Na^+, Cl^- and K^+ negatively correlated with plant height. The correlation between ion content and plant biomass was apparently due to shoot moisture content as Na^+ and Cl^- positively correlated with fresh weight and relative moisture content, but did not correlate with dry weight ($r = 0.03$ for Na^+ and $r = 0.07$ for Cl^-). The lack of correlation between ion content and dry weight indicates that the accumulation of ions in the shoots was not related to the biomass of the plants under salt treatment.

Genotypic data

The linkage map comprised 345 SNPs (Figure 1) which, at a LOD score of 4, gave nine linkage groups (LG) representing the nine chromosomes of lettuce. These had a total length of 1312 cM, with an individual length of 105 to 174 cM per LG. Each LG had 33 to 48 markers, with a median distance between the markers of 1.2 to 3.2 cM, except for LG 9 that had 19 markers with a median distance between the markers of 4.2 cM.

Table 1 Mean, range values and heritability for measured traits of the $F_{2:3}$ families and their parents under drought, salinity, nutrient deficiency and non-stress conditions

		L. serriola	*L. sativa*	$F_{2:3}$ families			
Trait	Treatment	mean	mean	Mean	Min	Max	H^2
Plant height (cm)	Control-D[1]	35.88	25.69	32.95	26.49	44.99	0.84
	Drought	21.63	17.85	20.52	16.80	26.43	0.82
	Control- SN[1]	57.68	22.85	43.31	13.17	89.28	0.98
	Salt	27.08	14.72	25.68	13.34	53.53	0.99
	Nutrient deficiency	21.22	12.12	18.91	10.03	48.07	0.98
Fresh weight (g)	Control-D	44.76	72.55	53.14	31.91	69.55	0.90
	Drought	10.22	13.94	11.16	8.46	14.02	0.51
	Control- SN	34.51	55.18	42.25	28.5	53.44	0.86
	Salt	12.64	24.98	15.28	9.38	19.92	0.83
	Nutrient deficiency	7.46	10.73	8.10	5.62	10.53	0.66
Dry weight (g)	Control-D	3.08[2]	3.20	3.02	1.60	4.39	0.90
	Drought	1.83[2]	1.91	1.62	1.19	1.96	0.66
	Control-NS	2.98[2]	2.46	2.91	2.12	4.32	0.90
	Salt	1.33	1.97	1.54	1.09	2.21	0.80
	Nutrient deficiency	1.05	1.58	1.14	0.78	1.58	0.71
Relative moisture	Control-D	93.09	95.62	94.44	93.07	95.73	0.82
content (%)	Drought	81.47	85.49	84.67	79.38	88.50	0.89
	Control-SN	91.31	95.56	93.08	88.62	94.57	0.93
	Salt	89.41	92.10	89.84	85.99	91.78	0.96
	Nutrient deficiency	85.88[2]	85.32	85.81	81.53	88.76	0.89
Na^+ (µg/g dry weight)	Control- SN	11.02	13.24	9.19	3.18	20.70	-
	Salt	24.35	49.91	31.47	8.32	54.89	-
Cl^- (µg/g dry weight)	Control- SN	10.51	19.28	15.56	7.24	22.13	-
	Salt	56.37	78.47	67.78	20.65	105.67	-
K^+ (µg/g dry weight)	Control- SN	44.14	82.77	66.14	38.36	93.07	-
	Salt	49.91	39.22	49.10	23.39	72.92	-

[1] Control-D: control treatment in the drought experiment, Control-SN: control treatment in the salt-nutrient deficiency experiment

[2] *L. serriola* and *L. sativa* not significantly different

Based on the 331 co-dominantly scored SNPs in 187 F_2 plants, the whole crop genome was represented in the F_2 population. The average crop allele content in the F_2 plants was 50% as expected, with individuals comprising of 28% to 66% crop alleles. The selection of the 98 F_2 plants for the experiment did not alter the average crop genome content. Using a significance level of 0.05 corrected for multiple tests by the Bonferroni method (α=0.05/331, Bland and Altman 1995), 8 markers (2.4%) had crop/wild allele frequency ratios that significantly deviated from the expected 1:1 ratio (χ2 ranging from 14 to 65). Three of these markers could not be

placed on the map and the remaining five mapped on LG3 where they spanned a continuous segment of 76 cM, and they favoured the crop allele (Figure 1). The flanking markers had relatively high $\chi 2$ values as well ($P_{\chi 2}$=0.0015) on both sides of the segment, indicating a non-random effect of distortion of the segment.

Table 2 Pearson's coefficients of correlation among the traits

Trait	Treatment	Plant height	Fresh weight	Dry weight	Relative moisture content	Na^+	Cl^-
Fresh weight	Control-D*	0.28					
	Drought	0.50					
	Control-SN*	0.04^{ns}					
	Nutrient deficiency	0.40					
	Salt	-0.18					
Dry weight	Control-D	0.35	0.83				
	Drought	0.29	0.58				
	Control-SN	0.59	0.61				
	Nutrient deficiency	0.24	0.76				
	Salt	0.33	0.77				
Relative moisture content	Control-D	-0.24	-0.19	-0.69			
	Drought	0.31	0.64	-0.17			
	Control-SN	-0.68	0.13	-0.65			
	Nutrient deficiency	0.17	0.12	-0.52			
	Salt	-0.80	0.17	-0.47			
Na^+	Salt	-0.56	0.56	0.03^{ns}	0.69		
Cl^-	Salt	-0.77	0.65	0.07^{ns}	0.80	0.79	
K^+	Salt	-0.32	0.01^{ns}	-0.13^{ns}	0.23	-0.31	0.20

* Control-D: control treatment in the drought experiment, Control-NS: control treatment in the salt-nutrient deficiency experiment**;** ns: correlation coefficient not significant (P>0.05)

QTL analysis

Seventeen QTLs were mapped for vigour traits (plant height, fresh weight, dry weight and relative moisture content) and six QTLs were mapped for ion content traits (Na^+, Cl^- and K^+). The details about the detected QTLs under control and stress conditions are shown in Table 4 and their locations on the linkage map are presented in Figure 1. The QTLs were located on 8 linkage groups, with LG1 having no QTL. The dominance effects of the QTLs were not significant, except for two QTLs, one for fresh weight, and another one for Na^+ content, showing that the vigour of the hybrids was not mainly due to the heterozygous genotypes. QTL by environment interaction (here the environments represented by the treatments) was significant for all the vigour trait QTLs and Cl^- content QTLs. This non-additive QTL effect from one treatment to another was due to the presence of a QTL in one treatment and its absence in another or to a differential QTL allelic effect characterized by unequal or opposite allelic effect from one treatment to another.

Eleven QTLs were detected in the drought experiment and seven of them had a positive effect from the crop allele. Five of the QTLs were common in the control and drought treatments, while three were specific to the control treatment and three were specific to the drought treatment. Fifteen vigour QTLs were detected in the salt-nutrient experiment with five of them having a positive effect from the crop allele and three QTLs having a positive effect from the crop allele in either the control or salt treatment and a positive effect from the wild allele in the nutrient deficiency treatment. Plant height was solely inherited from the wild parent in all the treatments, while the other vigour traits were inherited from both the crop and the wild parents.

Although the QTLs were located on 8 out of 9 lettuce LGs, sixteen of the twenty-three detected QTLs were located on three LGs. These were LG4, 7 and 9 and they constituted QTL hotspots because the QTLs overlapped on the same segments (Figure 1). On LG7 six QTLs overlapped on a chromosome segment of 28cM and two more QTLs overlapped in a neighbouring region. Five QTLs overlapped on LG 9 and three QTLs overlapped on LG4.

Table 4 QTLs for vigour traits and ion content traits mapped under drought, salt, nutrient deficiency and non-stress conditions

					Additive and dominance effect per treatment [2] (% explained variance)				
Trait	QTL name[1]	Most significant marker	LG	QTL x E	C-D[3]	D	C-SN	S	N
Plant height (cm)	*L-4-1*	CLS_S3_Contig5228-3-OP4	4	yes	-1.89(16)	-0.54(4)	-4.84(8)		
	L-6-1	Contig13566-1	6	yes				-3.15(6)	-3.32(8)
	L-7-1	QGB11B18.yg-2-OP5	7	yes			-10.23(34)	-8.66(46)	-5.44(23)
	L-9-1	CLS_S3_Contig2201-5-OP5	9	yes			-6.06(12)	-3.97(10)	-5.34(22)
Fresh weight (g)	*FW-2-1*	CLS_S3_Contig3908-1-OP5	2	yes	-2.84				-0.77
	FW-3-1	Contig1242-5	3	yes			2.52(11)	1.09(12)	
	FW-6-2	CLS_S3_Contig4649-3-OP5	6	yes	-2.46(5)				
	FW-7-2	CLSS4482.b2_C18-6-OP5	7	yes		0.53(12)		0.91(8)	-0.35(6)
	FW-8-1	RHCLS_S3_Contig7957_5	8	yes	3.73(12)		2.85(14)	0.85(7)	0.37(6)
	FW-9-2	CLS_S3_Contig2201-5-OP5	9	yes	3.35(9)	0.39(6)		0.06(11)	-0.37(6)
Dry weight (g)	*DW-3-2*	Contig1242-5	3	yes				0.07(6)	
	DW-4-2	CLS_S3_Contig4328-2-OP5	4	yes	-0.34(18)	-0.07(10)	-0.32(20)	-0.10(10)	-0.06(5)
	DW-7-3	CLSS4482.b2_C18-6-OP5	7	yes		0.07(12)	-0.23(11)		
	RMC-4-3	CLS_S3_Contig5668-7-OP5	4	yes	0.41(25)	1.28(28)	0.30(4)		0.88(16)
Relative moisture content (%)	*RMC-5-1*	CLRY8019-1	5	yes	0.24(9)	0.40(3)			
	RMC-7-4	QGB11B18.yg-2-OP5	7	yes			0.72(23)	0.98(32)	-0.90(17)
	RMC-9-3	CLSM16121.b1_B24-1-OP5	9	yes		0.48(4)	0.40(7)	0.48(8)	
Na^+ (µg/g dry matter)	*Na-7-5*	CLSM4311.b1_M21-1-OP5	7	NA[4]	-	-		9.34/3.36(48)	-
Cl^- (µg/g dry matter)	*Cl-7-6*	CLSS4482.b2_C18-6-OP5	7	yes	-	-		10.29(24)	-
	Cl-9-4	CLS_S3_Contig2201-5-OP5	9	yes	-	-		6.71(8)	-
K^+ (µg/g dry matter)	*K-7-7*	CLS_S3_Contig4590-1-OP5	7	yes	-	-		-8.44(24)	-
	K-7-8	CLSS4482.b2_C18-6-OP5	7	No	-	-	5.23(8)	5.23(8)	-
	K-9-5	CLS_S3_Contig2201-5-OP5	9	No	-	-	3.68(5)	3.68(5)	-

1 QTL names are derived from the traits they determine followed by the linkage group on which they are located and the number of the QTL on that linkage group; 2 QTL effect for the crop allele: positive value: effect positive for the crop allele, negative value: effect positive for the wild allele).Underlined values: significant dominance effect; 3 C-D: control treatment of the drought experiment; C-SN: control treatment of the salt-nutrient experiment; D: drought treatment, S: salt treatment, N: nutrient deficiency treatment; 4 NA: not applicable because one QTL was detected per trait

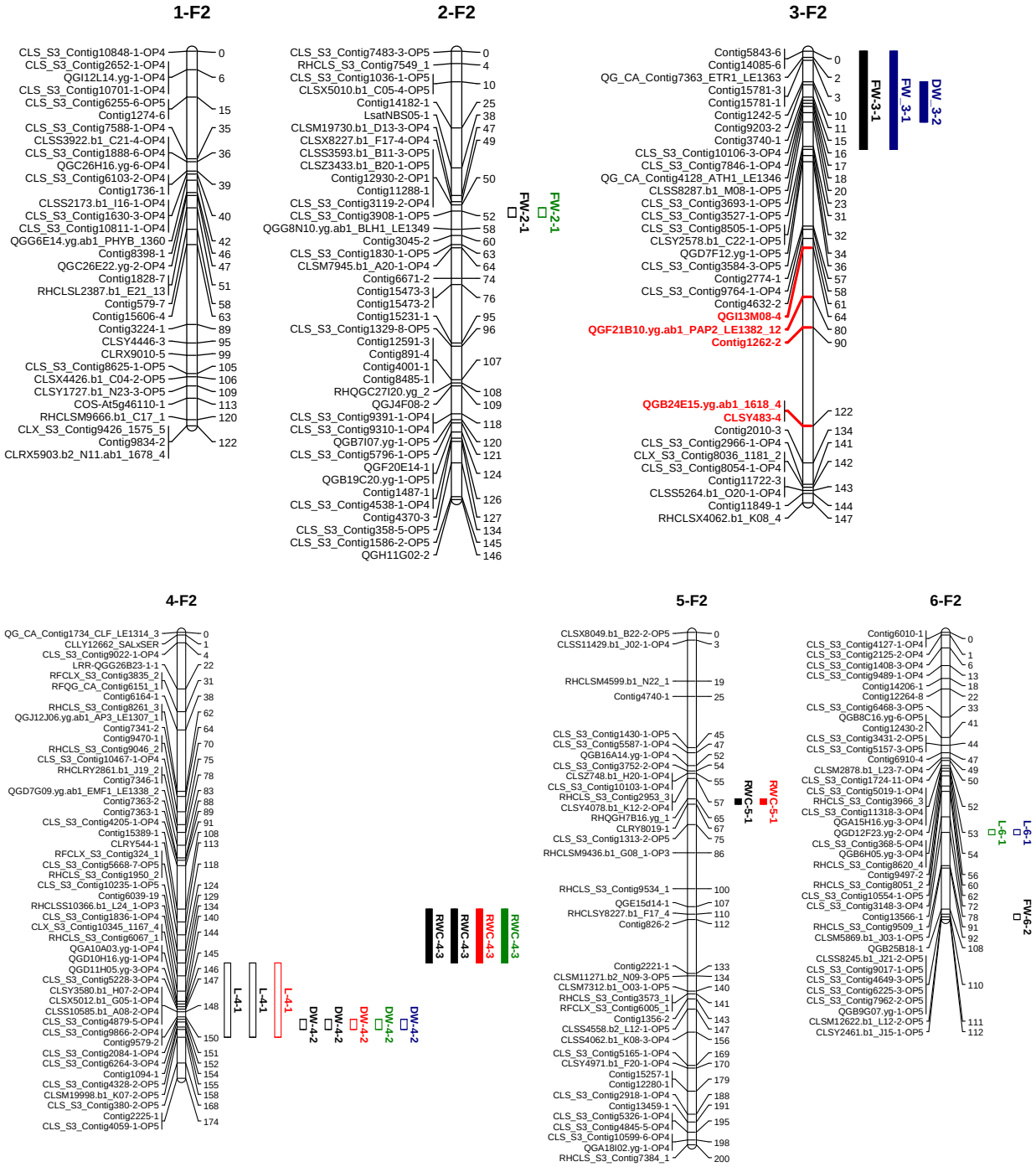


Figure 1 Linkage map of 345 SNPs based on 187 F_2 plants derived from a cross between *L. sativa* and *L. serriola*. The names of the markers are shown on the left of the LG bar and the distance is given on the right in centiMorgans. The markers with distorted segregation are shown in red (distortion towards the crop allele). The genomic localizations of the QTLs for plant height (L), fresh weight (FW), dry weight (DW), relative moisture content (RMC), sodium (Na), potassium (K) and chloride (Cl) as mapped under control (black), drought (red), salt (blue) and nutrient deficiency (green) conditions in 98 $F_{2:3\ families}$ are represented by the blocks. Solid QTL block: effect positive for the crop allele; open QTL block: effect positive for the wild allele.

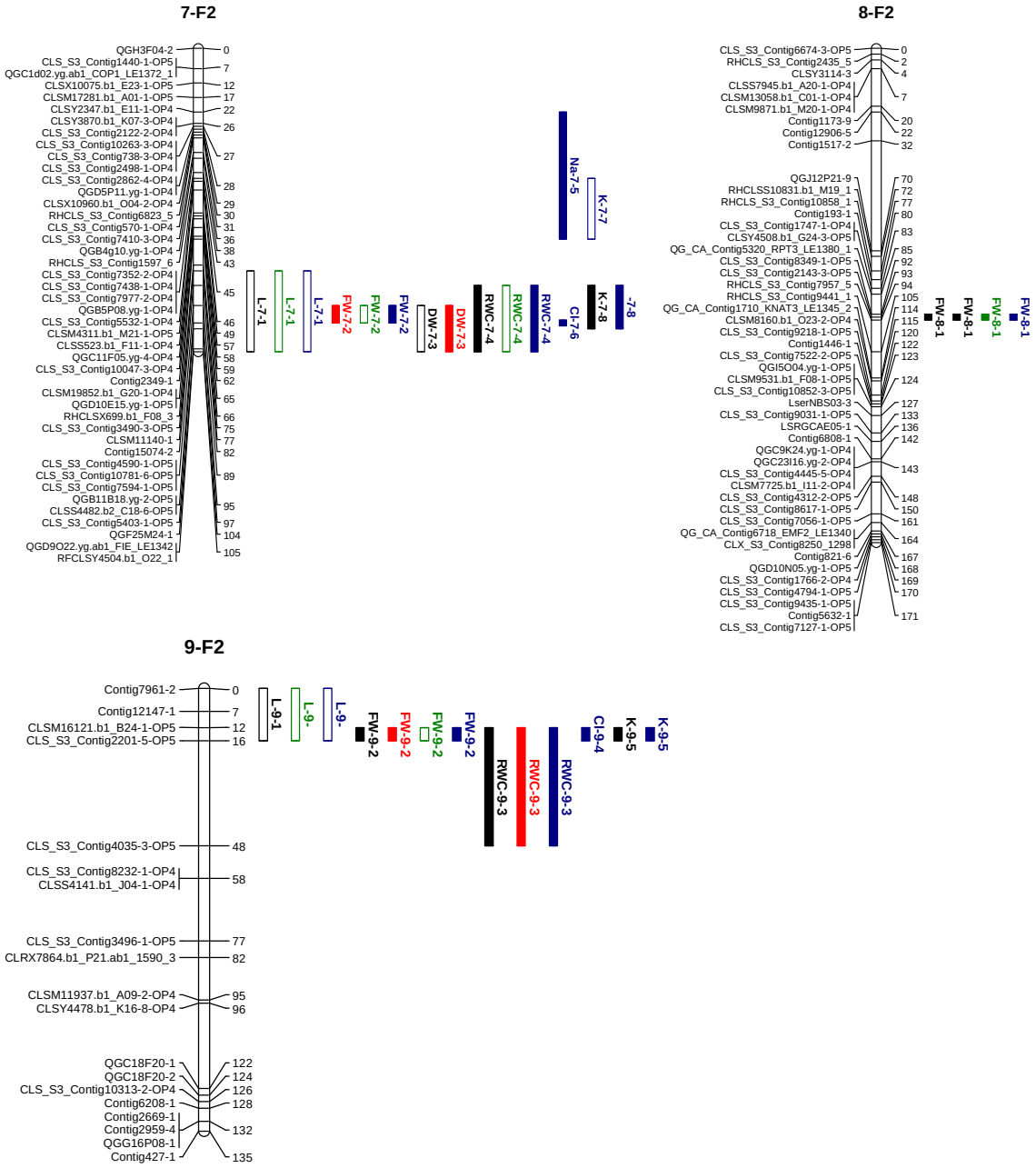


Figure 1-Continued

QTL epistatic effect

Twenty-one QTL pairs epistatically affected the traits under the five treatments, increasing the explained phenotypic variance by 6 to 12% (Table 5). Heterozygosity did not play an important role in the epistatic effect: for 18 QTL pairs, the predicted means for homozygous genotype

combinations were equal to or greater than the predicted means for the heterozygous combinations. Four of these QTL pairs were homozygous for the crop allele, six were homozygous for the wild allele, and eight of the QTL pairs were homozygous for the crop allele at one locus and homozygous for the wild allele at the other locus.

Table 5 Significant QTL x QTL interactions as detected by generalized linear model analysis fitting the main QTL effects and adding interaction between one pair of QTLs at a time

				Predicted genotypic means[1]								
Treat-ment[2]	Trait	QTL x QTL	% expl. variance	a/a	a/h	a/b	h/a	h/h	h/b	b/a	b/h	b/b
C-D	Plant height (cm)	*L-6-1* x *RMC-5-1*	11	32.7	31.8	35.5	30.1	33.4	34.5	34.0	32.6	31.9
	Dry weight (g)	*FW-6-2* x *DW-4-2*	8	2.9	2.7	2.6	2.5	3.2	3.1	2.5	3.2	3.5
	Relative moisture content (%)	*L-4-1* x *RMC-5-1*	7	94.8	94.7	93.5	94.6	94.4	94.2	94.7	94.5	94.4
		FW-2-1 x *RMC-5-1*	7	94.7	94.9	94.0	94.8	94.3	94.1	94.5	94.6	94.2
D	Plant height (cm)	*FW-8-1* x *DW-4-2*	12	18.2	21.3	20.4	19.7	20.8	20.2	20.7	19.4	21.8
	Dry weight (g)	*FW-3-1* x *RMC-4-3*	12	1.7	1.6	1.6	1.5	1.7	1.8	1.5	1.6	1.6
	Relative moisture Content (%)	*L-6-1* x *FW-8-1*	9	85.3	84.0	84.2	83.9	85.1	86.3	83.7	84.3	85.2
C-SN	Fresh weight (g)	*L-4-1* x *L-9-1*	11	43.2	41.6	47.18	44.4	43.1	38.8	39.51	42.5	38.1
		L-6-1 x *RMC-4-3*	11	45.7	41.0	42.9	35.8	43.1	43.5	39.2	42.1	42.0
	Dry weight (g)	*L-4-1* x *L-9-1*	9	2.9	2.8	3.2	2.8	3.1	2.8	2.4	3.0	3.1
		L-4-1 x *L-7-1*	8	2.8	2.9	3.4	2.7	2.9	3.2	1.9	3.0	2.8
		L-7-1 x *L-9-1*	11	2.7	2.6	2.4	2.7	3.1	3.0	2.7	3.2	3.6
		L-7-1 x *FW-8-1*	7	2.6	2.6	3.1	3.2	2.9	2.7	3.6	3.2	2.8
	Relative moisture content (%)	*L-4-1* x *L-9-1*	6	93.4	93.2	93.1	93.6	92.83	92.8	94.0	92.9	91.9
		L-9-1 x *DW-4-2*	6	93.2	93.6	93.7	93.3	93.0	92.7	93.3	92.9	91.6
		DW-4-2 x *RMC-5-1*	6	94.0	93.7	92.4	93.4	93.2	93.1	92.0	92.4	92.9
N	Plant height (cm)	*L-7-1* x *L-9-1*	6	11.2	7.7	14.0	13.1	20.0	25.4	15.6	25.7	31.6
	Dry weight (g)	*L-7-1* x *L-9-1*	12	1.3	1.2	1.1	1.1	1.2	1.2	1.0	1.1	1.3
S	Plant height (cm)	*L-4-1* x *L-9-1*	6	22.9	22.4	27.0	20.7	24.9	24.9	25.5	29.2	40.8
	Fresh weight (g)	*FW-6-2* x *RMC-5-1*	7	15.4	15.5	14.6	16.2	15.5	14.4	14.5	14.7	16.6
	Relative moisture content (%)	*L-9-1* x *DW-4-2*	6	90.6	90.1	90.9	89.7	89.5	90.7	89.3	89.5	88.7

1 a: homozygous for the crop allele, b: homozygous for the wild allele, h: heterozygous; 2 C-D: control treatment of the drought experiment, D: drought, C-SN: control treatment of the salt-nutrient experiment, N: nutrient, S: salt

Discussion

Early life stages of plants such as germination, seedling stage and vegetative growth are crucial phases as they determine the survival and reproduction of the plant, especially under stress conditions (Albacete *et al.* 2008; Donohue *et al.* 2010; Foolad 1996). In *Avena barbata*, early plant growth was found positively correlated to survival, fully grown plant biomass and plant fitness under field conditions (Latta and McCain 2009). In lettuce crop-wild hybrids, selection takes place on young plants, leading to surviving lineages with higher vigour and fitness than the wild genotypes (Hooftman *et al.* 2009; Hooftman *et al.* 2005). We studied the tolerance of young lettuce crop-wild hybrid plants to drought, salinity, and nutrient deficiency and mapped QTLs associated to plant vigour under those conditions in $F_{2:3}$ families derived from a cross between *L. serriola* and *L. sativa*.

Crop genome content in the hybrids

Interspecific crosses have been reported to result in high pre-zygotic segregation distortion in progeny (ranging from 22% to 90% of the markers) and to be associated with reproduction barriers (Jenni and Hayes 2009; Platt *et al.* 2010; Yue *et al.* 2009). The relatively low rate of distorted segregation in the F_2 population (2.4%) is consistent with the close relatedness of *L. serriola* and *L. sativa* and the complete fertility between the two species (De Vries 1990; Koopman *et al.* 1993; Koopman *et al.* 2001). In the same crop-wild cross, Hooftman *et al.* (2011) observed a segregation distortion of 7.5% under greenhouse (no mortality) conditions. Their results are similar to ours with the differences in percentage accountable to different methods of correcting the significance level for multiple tests. The region on LG3 where the distortion was located in our study could unfortunately not be compared with Hooftman *et al.* (2011) results due to the lack of common markers. The occurrence of genomic regions which favour one of the parental alleles may result in an increase in frequency of one parental allele at the expense of the other allele in subsequent generations. This has introgression consequences: on one hand, further selfing of the hybrids will lead to a rapid fixation of the crop alleles in regions such as on LG3 where the crop alleles are favoured over the wild alleles, regardless of the fitness effect of the crop (trans)genes. On the other hand, regions favouring the wild alleles will slow down the crop allele fixation, although none was identified in this cross. The identification of such genomic regions with pre- and post-zygotic segregation distortion could be exploited to minimize the introgression likelihood of transgenes. However, the use such regions in minimizing the escape of transgenes will depend on the stability of the distortion over generations and across genotypes.

Phenotypic variation and heritability of the traits

Hybridization between cultivated and wild lettuce resulted in a moderate to high heritability for the vigour traits and hybrids show improved vigour over the wild parent under non-stress and stress conditions. These results lend credence to previous experiments on lettuce which have shown that crop-wild hybrids could perform equally or better than the wild parent and that, depending on their fitness, hybrids could displace the wild taxon *L. serriola* in its natural habitat (Hooftman *et al.* 2005; Hooftman *et al.* 2008). The results also suggest that, if early vigour results in better fitness, lettuce hybrids could outperform the wild parent under stress conditions of salinity, drought and nutrient deficiency. These results are in line with the experimental results on radish (*Raphanus* spp.) (Campbell and Snow 2007; Campbell *et al.* 2006) in which hybrids

outperformed the wild parent in a new environment, hence indicating that hybrids could invade new ecological areas.

QTL effects

Hybridization brings together two species genomes and might result in the creation of new genotype combinations which will define the performance of the hybrids. Despite the close relatedness between *L. serriola* and *L. sativa*, a previous study has shown that the two species are molecularly distinguishable (Chapter 2). Improved hybrid vigour in early generations of hybrids has been associated with heterosis through dominance, overdominance and epistatic loci in repulsion phase (pseudo-overdominance) (Birchler *et al.* 2003; Burke and Arnold 2001). Hybrid vigour due to dominance and overdominance is expected to be short lived as it is associated with the advantage of the heterozygote genotypes which breaks down over subsequent generations due to selfing. In this study additivity was the major allelic action at 16 of the 17 vigour QTLs identified in the F_2 population. Dominance was significant for one vigour QTL (*FW-2-1*), hence dominance is likely not the most important genetic basis behind the improved vigour among $F_{2:3}$ families.

Conversely, epistasis as a result of non-additive effect of genotypes at two QTLs was significant for the traits under stress and control conditions. Despite the proven importance of epistasis on polygenic traits (Tisné *et al.* 2010; Yu *et al.* 1998), it is often underestimated due to the required large population size which is experimentally challenging to handle, combined with computational load, which makes it difficult to scan all pairs of loci, especially in highly heterozygous populations such as F_2 (Carlborg and Haley 2004). In a whole genome epistasis analysis, (Bai *et al.* 2010) found that the interaction between identified QTLs accounted only for 18% of all the interacting pairs of loci. We have probably also underestimated epistasis, as it was calculated only for those loci whose main effect was significant on their own and the background loci were not included in the interaction analysis. Despite the inclusion of only a subset of all loci in the analysis, the effect of epistasis was significant and it accounted for 6 to 12% of the phenotypic variance of the traits per pair.

Many of the vigour QTLs mapped to the same genomic regions in this study, notably on LG4, LG7 and LG9. Co-localizing QTLs were also obtained by Baack *et al.* (2008) for traits related to survival and morphology in a recombinant inbred line population of crop-wild sunflower hybrids. QTL co-localization may be due to a pleiotropic effect, if one QTL affects more than one trait,

but it is also possible that the QTLs are genetically linked and inseparable with the markers and recombination events observed in this study. The combination of QTL hotspots with QTL x treatment interaction through opposite allelic effect across treatments makes it difficult to choose which QTL region favours which parental allele. Nevertheless, these regions will remain under selection, positively or negatively, depending on to the prevailing conditions (optimum, dry, saline or nutrient deficient). The QTL region on LG7 corresponds to the QTL for germination under low and high temperature with a positive effect from the wild allele (Argyris *et al.* 2005). It also overlaps with the QTL for the number of lateral roots in the bottom length of the tap root with a positive effect from the wild allele (Johnson *et al.* 2000), indicating that the region could be under positive or negative selection. Therefore, in the process of creating genetically modified cultivars, such QTL regions should be avoided when selecting a transgene insertion event because, if the regions are under positive selection, leading to an increased frequency of linked loci through genetic hitchhiking, neutral crop alleles or transgenes in the population (Hooftman *et al.* 2011; Kwit *et al.* 2011; Stewart *et al.* 2003).

From F_1 progeny, the natural process of introgression in lettuce will continue with the creation of inbred lines through continued selfing or backcrosses to *L. serriola*, or a combination of the two. Therefore, further research encompassing generations of selfing and backcrossing to *L. serriola* is needed in order to establish the relevance of the detected QTL regions. Furthermore, this study was limited to plant vigour at an early stage of growth of the hybrid plants under controlled greenhouse conditions, while spontaneous crop-wild hybrids grow under natural field conditions. Additionally, greenhouse and field experiments are not always consistent (Gardner and Latta 2008; Latta and McCain 2009). Hence, the hybrids should be evaluated on the field in order to correlated early vigour with adulthood and reproduction, and link individual stress treatment with field conditions which may encompass multiple abiotic stress factors in combination with biotic stress factors such as diseases and herbivores.

Acknowledgements

This study was financially supported by the Netherlands Organization for Scientific Research (NWO) through their research programme “Ecology Regarding Genetically modified Organisms” (ERGO). The contributions of Leah K. McHale and Richard W. Michelmore were supported by the National Science Foundation Plant Genome Program Grant #: DBI0421630.

Chapter 4

Plant Vigour in Backcrossing Populations of Lettuce

Brigitte Uwimana[1], Marinus J.M. Smulders[1], Danny A.P. Hooftman[2], Yorike Hartman[3], Hans Jansen[4], Clemens C.M. van de Wiel[1], Richard G.F. Visser[1]

1 Wageningen UR Plant Breeding, Wageningen, The Netherlands
2 Centre for Ecology and Hydrology, Wallingford, UK.
3 IBED, Universiteit van Amsterdam, Amsterdam, The Netherlands
4 Wageningen UR Biometris, Wageningen, The Netherlands

Abstract

After crop-wild hybridization, some of the crop segments are established through selfing of the hybrids or introgressed by backcrossing to the wild parent or a combination of the two, depending on the reproductive system of the plants. In lettuce, selfing is the most likely pathway that the hybrids will follow due to its self-pollination nature. However, introgression through backcrossing to the wild parent is also likely to happen due to the high frequency of the wild plants relative to the hybrids. To test the effect of backcrossing on the vigour of the hybrids, two backcross populations of lettuce were created from a cross between *L. serriola* and *L. sativa* and backcrossed to *L. serriola* to generated BC_1 and BC_2 populations. Plant vigour was tested in the two populations under greenhouse conditions of non-stress and abiotic stress conditions of salinity, drought and nutrient deficiency. Despite the decreasing crop genomic composition in the backcross populations, the hybrids were characterized by a substantial genetic variation and hybrids were identified under non-stress and stress conditions that performed equally or better than the wild genotypes, indicating that two backcrossing events did not eliminate the effect of the crop. QTLs for plant vigour under non-stress and the mentioned stress conditions were detected in the two populations with a positive effect from the crop and the wild parents. Based on the location of the QTLs on the linkage groups, we suggest genomic regions where transgenes could be inserted in order to limit their persistence through genetic hitchhiking.

Introduction

One of the debated ecological risks associated with the commercial cultivation of genetically modified crop varieties is the possibility of introgression of transgenes from crops to their wild relatives through hybridization. It is feared that introgression may increase the weediness of the wild relatives in agricultural areas, cause genetic erosion of the latter in wild populations, or result in crop-wild transgenic wild lineages invading new habitats (Groot *et al.* 2003; Pilson and Prendeville 2004; Snow *et al.* 2005; Tiedje *et al.* 1989). Because hybridization between crops and their wild relatives is likely for most of the crop species (Ellstrand 2003; Ellstrand *et al.* 1999), the outcome of hybridization between crops and their wild relatives has become a subject of many research studies, using either transgenic or conventional crop varieties (e.g. (Baack *et al.* 2008; Dechaine *et al.* 2009; Hooftman *et al.* 2009; Snow *et al.* 2003).

The net effect of crop-wild hybridization may be negative, for instance if crop genes reduce the competitive ability under natural conditions; or positive, if hybrids inherit combinations of additive positive traits from the crop and the wild parents (Burke and Arnold 2001). If hybrids are viable and fertile, hybridization can result in a swarm of hybrids in which crop and wild genomes interactively define the hybrid phenotypes. From the F_1 progeny onwards, crop alleles can be fixed through selfing or through backcrossing to the wild parent. Natural selection will purge out maladapted genotypes, leaving those genotypes with the same or higher net fitness as the wild parent in the natural habitat of the wild taxon, or with broadened adaptation as a result of transgressive segregation (Lexer *et al.* 2003b; Rieseberg *et al.* 1999).

Initially, any crop gene in a hybrid plant will be in a chromosome segment comprising the gene itself and other genes linked with it, and the fitness effect will depend on the effect of the whole chromosome segment (Hooftman *et al.* 2011). In the course of crop allele fixation, a gene that confers a selective advantage may be introgressed, but it will do so along with other loci tightly linked to it, even if these are neutral to fitness. In the same way, a gene may be selected against due to its linkage to a deleterious gene (Barton 2000; Kwit *et al.* 2011; Stewart *et al.* 2003). It is within such a context that the dynamics of the process of introgression from crops to wild relatives constitute a baseline for understanding the effects of transgene escape and fixation into wild taxa (Baack *et al.* 2008; Chapman and Burke 2006) .

We have initiated a study in which we follow the genetic process of introgression from cultivated lettuce (*Lactuca sativa* L.) to its wild relative prickly lettuce (*Lactuca serriola* L.). The two species hybridize successfully, giving viable and fertile hybrids (De Vries 1990),

hence making a typical crop-weed complex. Despite the limited outcrossing rate in the two species (D'Andrea *et al.* 2008; Thompson *et al.* 1958), we have identified crop-wild hybrid plants among natural populations of *L. serriola* which are a result of spontaneous gene flow between the two species (see Chapter 2).

In a previous study we have explored the genetic basis of hybrid vigour in an F_2 population resulting from a synthetic cross between cultivated *L. serriola* and *L. sativa* (see Chapter 3). We mapped QTLs for plant vigour, which co-localized in a small number of chromosome regions, with additive QTL main effect and epistasis as the major genetic effects. After hybridization, the crop segments will be established in the wild background or eliminated by selection either through selfing of the hybrids or through backcrossing to the predominant wild plants, or a combination of the two processes. Selfing generations after a single hybridization event between the crop and the wild parents are characterized by crop genomic segments which constitute an average of 50% of the hybrid genome. In contrast, every backcross to the wild parent decreases the crop genome content by half, while the crop genome segments become smaller through recombination (Supplementary material Figure S3). In this way, crop segments that contribute to the vigour and fitness of the hybrids get introgressed with a decreasing number of hitchhiking loci with each backcross generation. Therefore, the fitness effects of a transgene in the context of its genomic location will differ in the selfing and backcrossing pathways.

In this study we follow up the crop-weed complex of L. *sativa* and *L. serriola* in a marker-assisted introgression study, and we focus on BC_1 and BC_2 generations in which *L. serriola* was the recurrent parent, hence mimicking the introgression process from crops to wild relatives through repeated backcrosses with wild populations. We want to get answers to the following questions: (1) Do the backcross generations exhibit transgressive segregation for vigour? (2) Are the vigour QTL regions that were identified in the selfing pathway (F_2 population) also detected in the backcross populations? (3) How does the contribution of the crop to the vigour of the hybrids change with the increasing wild genetic background?

Materials and Methods

Generation of BC_1 and BC_2 hybrid progenies and genotyping

The present study concerns two backcross populations, BC_1 and BC_2, back-crossed to *L. serriola* to mimic the natural introgression from a crop to its wild relative. Flowers from the F_1 hybrid plant resulting from a cross between *Lactuca serriola* (collected from Eys, the Netherlands) and *Lactuca sativa* (cv. Dynamite), which was also used to create the F_2

population (Chapter 3), were hand-pollinated with *L. serriola* pollen to generate BC_1 lines according to the lettuce pollination protocols by Nagata (1992) and Ryder (1999). By the same method, BC_2 plants were created using the same *L. serriola* parental line.

Because the BC_1 population was developed before the 1083 Single Nucleotide Polymorphism (SNP) markers were available (see chapter 3) 192 individuals of the BC_1 population were initially genotyped using the then available 384 SNP markers (OPA 1 and 2, http://compgenomics.ucdavis.edu/compositae_SNP.php). One hundred sixty-seven SNPs were successfully scored in the 192 BC_1 individuals but resulted in a sparse genetic linkage map. Based on the genotypes, 100 BC_1 individuals were selected to be used in greenhouse experiments, using the program "Genetic Distance Optimization program" (GDOpt) (Odong *et al.* 2011). Forty-five of the 100 BC_1 plants were backcrossed to *L. serriola* to generate BC_2 lines. At the same time, the BC_1 lines were left to self-pollinate to BC_1S_1 seeds (Supplementary material Figure S3). Six hundred BC_2 individuals (12 BC_2 plants for each of the 45 back-crossed BC_1 lines) were regenerated and selfed to produce BC_2S_1 seeds. When 768 SNP markers were developed and a customized GoldenGate genotyping panel of 384 SNPs which were polymorphic between the parents (*L. serriola*/Eys and *L. sativa* cv. Dynamite) was completed (Chapter 3), the 100 BC_1 individuals were genotypes again in order to improve the map density of the BC_1 population, along with 458 randomly chosen BC_2 individuals. Based on their genotypes, a selection of 100 BC_2 individuals was made using the program GDOpt and their BC_2S_1 progenies were used in greenhouse experiments

Greenhouse experiments

The BC_1S_1 and BC_2S_1 seeds of the selected 100 BC_1 and 100 BC_2 individuals were used in greenhouse experiments together with their parents (*L. serriola*/Eys and *L. sativa* cv. Dynamite). We also included two lines, *L. serriola* acc. UC96US23 and *L. sativa* cv. Salinas, which, together with the parents, were used to estimate the environmental error. The parents and the two additional lines were replicated 12 times per treatment, and each BC_1 and BC_2 individual was represented by 12 BC_1S_1 and BC_2S_1 seedlings per treatment respectively.

Experiments were conducted separately for the two populations, using the same set up as in the F_2 experiments (Chapter 3). For each population, two experiments were carried out, one comprising salt and nutrient treatments together with a control treatment and another experiment comprising a drought treatment together with a control treatment. The drought experiment for the BC_1 population was carried out in the period of February-March 2009, the salt and nutrient experiment for the BC_1 population was carried out in April-May 2009, the

drought experiment for the BC_2 population was carried out in November 2009-January 2010 and the salt and nutrient experiment of the same population was carried out in January-March 2010. After transplanting, the plants were given water twice a week for two weeks after which the stress treatments were applied. For the drought treatment, the plants were not given water for three weeks; for salt treatment, irrigation water was supplied with 100 mM NaCl, and for nutrient deficiency treatment, plants were irrigated with water without nutrients for three weeks. At the end of the fifth week after transplanting (at the rosette stage) we measured plant vigour for individual plants as shoot height, shoot fresh weight and shoot dry weight (after drying at 80ºC for 3 days). We calculated shoot relative moisture content as the ratio of the amount of water in the shoot to the total shoot weight [(fresh weight-dry weight)*100/fresh weight].

Construction of the linkage maps

Out of 384 SNP markers, 347 were successfully scored in the 100 BC_1 individuals and 348 in the 458 BC_2 individuals. Genetic linkage maps of the two populations were built separately using JoinMap® 4 (Van Ooijen 2006). The BC_2 population was handled as a back-cross population without selfing (BC_{b2F0}) and the expected genotype segregation was adjusted to 3:1. The marker grouping was kept the same as in the BC_1 and F_2 populations, and the order of the markers and their genetic distances were calculated based on recombination among the BC_2 individuals. The linkage maps were displayed using MapChart2.2 (Voorrips 2002).

Analysis of phenotypic data

Statistical analysis was performed using GenStat 13th Edition (Payne *et al.* 2010). The drought and the salt-nutrient experiments were analysed separately. The significance of the different terms was determined by the analysis of variance, fitting the model

Response = general mean + block + genotype + treatment + genotype.treatment + error; with the term *genotype* representing the hybrid families (BC_1S_1 or BC_2S_1). Broad sense heritability of the traits was estimated for each treatment in each population as the proportion of the total variance accounted for by the genetic variance using the formula:

$$H^2 = Vg/(Vg+Ve/r);$$

where *Vg* is the genetic variance for the BC_1S_1 or BC_2S_1 families, *Ve* is the environmental variance, and *r* is the number of replications (Chahal and Gosal, 2002). *Vg* was estimated based on the restricted maximum likelihood (REML) method from the mixed model:

Response = general mean + block + genotype + error;

with the term *genotype* taken random. Because BC_1S_1 and BC_2S_2 families were segregating, the term *Ve* was the error variance derived from a one-way analysis of variance of the model:

Response = general mean + block + parents + error;

with the term *parents* representing the two parents (*L. serriola*/Eys and *L. sativa* cv. Dynamite) and the two added lines (*L. serriola* acc. UC96US23 and *L. sativa* cv. Salinas).

Quantitative Trait Loci analysis

The genetic linkage map, the genotype scores and the phenotypic means were combined for QTL analysis using the QTL analysis function of GenStat 14th Edition (Payne *et al.* 2011). Each trait was analysed individually using the single trait – multiple environment option of the program. The BC_1 and BC_2 populations were analysed separately. The BC_2 population was handled as a BC_1 population. To adjust for the calculation differences caused by the marker gaps due to the additional recombination event in the BC_2 population, the gaps in the BC2 linkage map were filled with virtual markers which were given missing marker scores. Thirty-five virtual markers were added on LG1, 2, 3, 4, 5, 8 and 9, keeping a maximum distance of 12 cM between the markers (Figure 2).

In order to effectively model genotype by environment interaction (GxE, with environments represented by the different treatments) through QTL by environment interaction (QTLxE), each trait was analysed individually using the single trait – multiple environment option of the program. Genome-wide association between markers and traits was decided based on a significance level of 0.05 corrected for multiple tests using the Li and Ji method (Li and Ji 2005). After the selection of the best variance-covariance model for the treatments (Malosetti *et al.* 2004), the candidate QTLs were determined by initial genome scan. Final QTL positions were determined by composite interval mapping taking into account co-factors. The allelic effect of the detected QTLs in each treatment, the effect of QTLxE and the explained phenotypic variance of each QTL per treatment were determined by running a backward selection on the candidate QTLs in a mixed linear model, taking the QTL effect in each treatment as fixed terms and the interaction between each hybrid family and the treatment as random (Mathews *et al.* 2008). In that way, each QTL detected in one treatment was tested for its effect and significance in the other treatments.

To test for QTL epistatic effect (QTL x QTL), the phenotypic means were regressed against the genotypes of the most significant markers for each QTL in a generalized linear model.

One marker was considered for each QTL region, and no QTL interaction was estimated for QTLs on the same LG. For each treatment, every trait was explained by the main effects of all the detected QTLs to which interaction between one pair of QTLs was added at a time. The interaction effects of the QTL regions that were unique to the BC_1 population were also included in the QTLxQTL analysis in BC_2. QTLxQTL interaction was decided significant at a level of 0.05 corrected for the number of the traits using the Bonferroni method (Bland and Altman 1995).

Results

Phenotypic variance among the hybrid families

Backcrossing rendered the hybrid plants morphologically very similar to their wild parent, *L. serriola*. Vigour depended on the backcross families and varied between the treatments in the two hybrid populations as revealed by the significance of GxE ($P_{\text{genotype x treatment}}$ <0.001 for all traits). The BC_1S_1 and BC_2S_1 families showed a wide range of means for the vigour traits under stress and non-stress conditions (Table 1). Some trait-treatment combinations such as plant height under all the treatments and dry weight under control and drought conditions showed transgressive segregation over the two parents. For all traits and in both backcross generations the mean of the wild parent *L. serriola* was lower than the maximum mean of the hybrid families. In spite of a second generation of backcrossing from BC_1 to BC_2, for each trait-treatment combination individual BC_1S_1 and BC_2S_1 plants and families stood out that performed better than the two wild genotypes (*L. serriola*/Eys and *L. serriola* acc. UC96US23, Table 1 and Supplementary material Figures S4 and S5,), indicating that the BC_2 plants still contained crop genome segments which contributed positively to their vigour.

Genetic variation as expressed by broad sense heritability of the traits ranged from 0.44 to 0.95 in the BC_1 experiments, showing that a substantial part of the phenotypic variation was due to genetic factors (Table 1). In the drought experiment, heritability decreased from control to drought treatment for all traits. In the salt-nutrient experiment, the heritability decreased from control to stress treatments (salt and nutrient deficiency) for plant height, fresh weight and dry weight, but it increased for relative moisture content, with a greater increase in the nutrient deficiency treatment (from 0.64 to 0.90).

In the BC_2 population, heritability of the traits among BC_2S_1 families ranged from 0.43 to 0.85, which is comparable to the range found in the BC_1 population (Table 1). Also comparable to the BC_1 population is that the heritability decreased from control to drought in the drought experiment for all the traits in the BC_2 population.

In the salt-nutrient experiment, heritability decreased from control to salt for fresh weight and dry weight, while it slightly increased for plant height and relative moisture content. In the same experiment, heritability considerably decreased under nutrient deficiency conditions for plant height, from 0.85 under control to 0.43 under nutrient deficiency conditions. The genetic changes due to a second backcross to the wild parent from BC_1 to BC_2 led to a decrease of the heritability of the traits for most of the trait-treatment combinations. The most remarkable changes were under nutrient deficiency conditions, where heritability decreased by 50% for plant height, while it increased by 34% for fresh weight, 23% for dry weight and 18% for relative moisture content in the same treatment.

Table 1 Parental means and mean, minimum and maximum values and heritability of the BC_1S_1 and BC_2S_1families for vigour traits under non-stress, drought, salinity and nutrient deficiency conditions

Trait	Treatment	*L. serriola*	*L. sativa*	Hybrid families			H^2
				Mean	Min	Max	
				BC_1S_1			
Plant height (cm)	Control-D[1]	31.42	23.52	30.87	27.43	36.19	0.86
	Drought	16.58	13.19	16.05	13.53	18.53	0.74
	Control- SN	33.36	17.95	28.75	22.39	42.64	0.95
	Salt	16.72	14.70	17.18	13.40	24.93	0.95
	Nutrient deficiency	10.57	8.75	10.07	7.88	13.35	0.86
Fresh weight (g)	Control-D	39.20	68.11	46.56	25.25	62.26	0.87
	Drought	6.48	8.01	6.46	5.40	8.14	0.48
	Control- SN	25.52	39.15	27.24	21.76	32.27	0.79
	Salt	8.40	20.13	10.85	7.90	14.35	0.69
	Nutrient deficiency	2.60	4.82	3.16	2.46	3.92	0.44
Dry weight (g)	Control-D	2.42	3.14	2.87	1.51	4.09	0.90
	Drought	1.15	1.38	1.16	0.93	1.41	0.80
	Control-NS	2.01	2.56	2.13	1.61	2.75	0.75
	Salt	0.84	1.79	1.07	0.74	1.43	0.59
	Nutrient deficiency	0.50	0.90	0.61	0.42	0.86	0.62
Relative moisture content (%)	Control-D	93.80	95.41	93.89	93.07	94.83	0.75
	Drought	81.97	82.56	81.77	79.26	84.74	0.69
	Control-SN	92.11	93.50	92.24	91.25	93.21	0.64
	Salt	90.00	91.18	90.18	89.23	91.24	0.67
	Nutrient deficiency	81.03 [ns]	81.85	80.77	77.84	85.09	0.90
				BC_2S_1			
Plant height (cm)	Control-D[1]			29.17	24.09	37.01	0.85
	Drought	14.02	12.32	14.32	12.39	17.75	0.77
	Control- SN	21.02	16.63	21.51	17.51	28.04	0.80
	Salt	16.54	13.62	16.54	13.01	22.2	0.84
	Nutrient deficiency	11.62	10.07	11.33	9.69	14.05	0.43
Fresh weight (g)	Control-D	27.21	67.25	38.59	23.38	54.89	0.73
	Drought	5.24 [ns]	5.46	4.55	3.31	6.32	0.37
	Control- SN	13.70	31.32	17.64	13.19	26.39	0.72
	Salt	9.31	17.87	10.35	7.27	13.36	0.63
	Nutrient deficiency	4.87	7.14	5.34	4.26	7.65	0.59
Dry weight (g)	Control-D	2.08	3.34	2.76	1.81	3.95	0.80
	Drought	1.12 [ns]	1.22	1.06	0.88	1.27	0.50
	Control-NS	1.03	1.88	1.31	0.96	1.94	0.71
	Salt	0.84	1.29	0.92	0.65	1.21	0.61
	Nutrient deficiency	0.68	0.83	0.71	0.50	1.16	0.76
Relative moisture content (%)	Control-D	92.28	95.06	92.86	92.06	94.03	0.78
	Drought	77.94 [ns]	77.53	76.12	71.72	80.39	0.77
	Control-SN	92.52	94.06	92.61	91.79	93.79	0.73
	Salt	90.95	92.85	91.24	89.92	92.36	0.76
	Nutrient deficiency	85.88	88.26	86.70	84.49	89.48	0.79

1 Control-D: the control treatment of the drought experiment, Control-SN: the control treatment of the salt-nutrient experiment

Allelic composition of the hybrids and linkage maps

BC_1 individuals on average contained 26% crop genome with individual plants ranging from 11% to 39%. The population was characterized by long crop segments in a heterozygous state which sometimes spanned all the markers on a whole linkage group (Figure 1A). One additional backcross to the wild parent resulted in a reduction of crop genome content to 14%, varying among BC_2 individuals both in segment size and proportional amount, ranging from 3% to 29% (Figure 1B). Twenty-five markers (7%) showed a segregation distortion towards the crop allele in the BC_2 population. Because none of the markers was distorted in the BC_1 population, the segregation distortion in the BC_2 population was most probably due to the selection of the 45 BC_1 plants that were backcrossed to *L. serriola* to create the BC_2 population.

The linkage maps, shown in Figure 2, consist of nine linkage groups (LG) which represented the nine chromosomes of lettuce (Truco *et al.* 2007). The same marker order was obtained in the BC_1 and BC_2 populations. The BC_1 map was made of 347 markers spanning a total length of 1301 cM, while the BC_2 map had 348 markers with a total length of 1403 cM. The linkage groups contained 34 to 50 SNP markers, except LG9, which had 18 markers. As mentioned in the QTL analysis subsection of Materials and Methods, virtual markers were added on the BC_2 map to fill up the gaps for better QTL mapping results. These markers are underlined in the BC_2 linkage map (Figure 2).

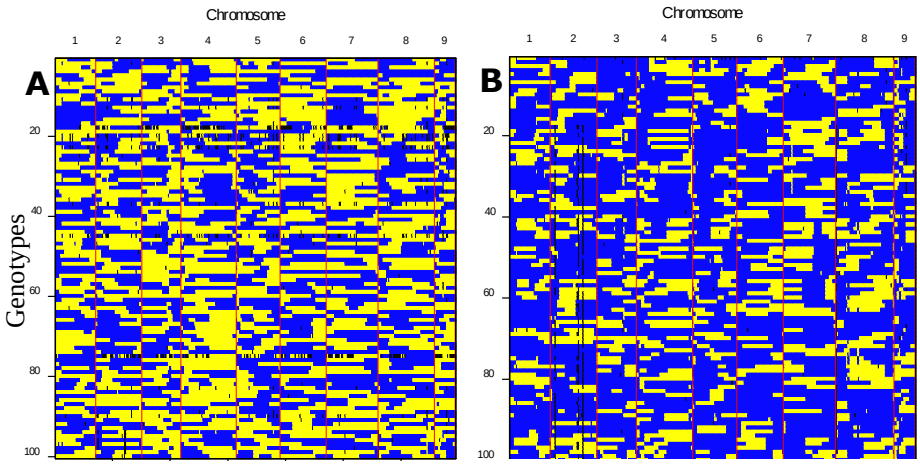


Figure 1 Allelic composition of the selected BC_1 (A) and BC_2 (B) genotypes. Blue: homozygous for the wild allele; yellow: hetereozygous; black: missing genotype scores. Backcrossing to the wild parent reduces the crop genome content in amount and segment size

Quantitative Trait Loci

Twenty QTLs associated with plant vigour were mapped in the BC_1 population, 5 for plant height, 4 for fresh weight, 4 for dry weight and 7 for relative moisture content (Table 2 and Figure 2). The QTLs were located on all linkage groups except LG2. Only three of these

QTLs had the same order of magnitude additive effect in all treatments. The remaining QTLs were significantly affected by QTLxE. QTLs for plant height had an additive effect positive for the wild allele in the two control treatments, drought and salt treatments. Under nutrient deficiency, two of the plant height QTLs had an additive effect positive for the wild allele, while three QTLs for the same trait were positive for the crop allele, including two QTLs (*L-3-3* and *L-7-1*) which had a positive effect for the wild allele in other treatments, hence showing opposite allelic effects from one treatment to another.

Fresh weight QTLs were inherited from the crop as three of the QTLs for this trait showed a positive additive effect for the crop allele. Dry weight was inherited from both the crop and wild parent as three of the QTLs for the trait had a positive additive effect for the crop allele, while one QTL for that trait showed a positive additive effect for the wild allele. Relative moisture content QTLs were inherited from both the crop and the wild parents. Four of the QTLs mapped for this trait had a positive additive effect for the crop allele, while the additive effect was positive for the wild allele for the remaining three QTLs.

Fewer QTLs were mapped in the BC_2 than in the BC_1 population (Table 2 and Figure 2). Thirteen QTLs were mapped in BC_2 for vigour-related traits. Four of the QTLs were significant in all the treatments with the same additive effect, hence having non-significant QTLxE effect, while the remaining nine had were significantly affected by QTLxE. Two of the QTLs for plant height had a positive additive effect for the wild allele and they were significant under the control treatment of the salt-nutrient experiment and under salt treatment. The other two had a positive additive effect for the crop allele.
The three fresh weight QTLs had a positive additive effect for the crop allele. For the dry weight QTLs, one had a positive additive effect from the crop allele and the other one was positive for the wild allele. Relative moisture content QTLs were inherited from both the wild and the crop parent.

QTL epistatic effects

QTL epistatic effects on the vigour traits were significant in the two hybrid populations and under stress and non-stress conditions. In the BC_1 population epistasis was estimated for 10 QTL pairs and it explained 4 to 9% of the phenotypic variance per individual QTL pair and up to 23% per trait. Nine QTL regions mapped in BC_2 and 6 QTL regions unique to the BC_1 population were used for QTL epistatic effect analysis and they explained 3 to 11% of the phenotypic variance per QTL pair and up to 27% per trait (Table 3). There were more QTL pairs in the BC_2 population showing significant interaction than in the BC_1 population, which

could be due to the fact that more QTL regions were included in QTLxQTL interaction analysis in the BC_2 population. None of the QTL pairs was significant in both populations. There was more consistency in QTL epistatic effect in BC_2 than in the BC_1 as only one QTL pair (*L-3-3* x *RMC-8-4*) was significant for the same trait (plant height) under control and nutrient deficiency conditions while one QTL pair could affect more than one trait in the BC_2 population.

Table 2 Quantitative trait loci mapped in 100 BC_1S_1 and 100 BC_2S_1 families for vigour traits under non-stress, drought, salt and nutrient deficiency conditions

					QTL x	Additive effect for the crop allele (% expl. variance)				
Trait	Pop.	QTL name	Most significant marker	LG	E	C-D[1]	D	C-SN	N	S
Plant height (cm)	BC_1	*L-7-1*	QGF25M24-1	7	yes			-2.1(6)	0.4(3)	
		L-9-1	CLS_S3_Contig2201-5-OP5	9	yes			-4.6(32)	-0.8(13)	-2.3(30)
		L-1-1	QGC26E22.yg-2-OP4	1	yes				0.7(12)	
		L-3-3	QGF21B10.yg.ab1_PAP2_LE1382_12	3	yes		-0.5(6)		0.39(3)	
		L-5-2	CLS_S3_Contig1313-2-OP5	5	yes	-1.2(11)		-1.7(4)	-0.5(5)	-1.5(13)
	BC_2	*L-7-1*	QGB11B18.yg-2-OP5	7	yes			-1.3(4)		-0.7(3)
		L-9-1	CLS_S3_Contig2201-5-OP5	9	yes					-0.5(3)
		L-4-4	Contig1094-1	4	no	0.4(1)	0.4(6)	0.4(1)	0.4(9)	0.4(2)
		L-8-2	CLX_S3_Contig8250_1298	8	yes	1.5(8)	0.7(14)	1.9(19)	0.4(6)	1.57(24
Fresh weight (g)	BC_1	*FW-7-2*	CLSS4482.b2_C18-6-OP5	7	yes	4.0(10)	0.2(4)	1.0(4)	0.3(27)	1.3(22)
		FW-9-2	CLS_S3_Contig2201-5-OP5	9	yes	2.4(3)			0.3(17)	0.5(3)
		FW-1-2	QGG6E14.yg.ab1_PHYB_1360	1	yes	4.3(11)		1.5(11)	0.1(4)	0.9(11)
		FW-4-5	Contig6039-19	4	yes	3.04(6)				
	BC_2	*FW-7-2*	CLS_S3_Contig7594-1-OP5	7	no	0.3(1)	0.3(9)	0.3(1)	0.3(5)	0.3(1)
		FW-1-2	CLSS3922.b1_C21-4-OP4	1	yes	2.7(7)		2.3(19)		1.0(19)
		FW-8-1	RHCLS_S3_Contig9441_1	8	no	0.3(1)	0.3(13)	0.3(1)	0.3(7)	0.3(2)
		RMC-7-4	QGF25M24-1	7	yes	0.2(7)			-0.6(9)	
		RMC-1-3	CLRX9010-5	1	yes		-1.3(19)	-0.2(5)	-0.6(11)	-0.2(8)
		RMC-5-3	Contig2221-1	5	yes		0.7(6)	-0.2(9)		
Dry weight (g)	BC_1	*DW-7-3*	CLSS4482.b2_C18-6-OP5	7	no	0.1(1)	0.1(15)	0.1(3)	0.1(19)	0.1(8)
		DW-4-6	CLRY544-1	4	yes	-0.4(16)		-0.1(4)	-0.1(4)	
		DW-8-3	QG_CA_Contig5320_RPT3_LE1380_1	8	no	0.1(1)	0.1(6)	0.1(1)	0.1(7)	0.1(3)
		DW-9-6	QGG16P08-1	9	yes			-0.1(4)		0.1(6)
	BC_2	*DW-7-3*	QGF25M24-1	7	no	0.1(1)	0.1(25)	0.1(4)	0.1(12)	0.1(9)
		DW-4-6	Contig7363-2	4	yes	-0.2(6)	-0.1(15)	-0.1(8)		
Relative moisture content (%)	BC_1	*RMC-4-3*	CLRY544-1	4	yes	0.2(9)	0.6(6)	0.2(4)	1.2(17)	
		RMC-5-1	RHCLSM9436.b1_G08_1-OP3	5	yes	0.2(10)				
		RMC-7-4	CLSS4482.b2_C18-6-OP5	7	yes		-0.7(9)	-0.2(6)	-0.9(9)	
		RMC-3-4	QGF21B10.yg.ab1_PAP2_LE1382_12	3	yes				0.8(8)	0.2(5)
		RMC-4-7	CLX_S3_Contig10345_1167_4	4	no	0.2(7)	0.2(1)	0.2(10)	0.2(1)	0.2(4)
		RMC-6-3	QGB25B18-1	6	yes				-1.2(17)	
		RMC-8-4	CLS_S3_Contig9218-1-OP5	8	yes		-0.8(11)		0.5(3)	-0.2(4)
	BC_2	*RMC-4-3*	Contig15389-1	4	yes	0.3(19)	1.2(14)	0.3(16)	0.9(23)	0.3(9)
		RMC-7-4	QGF25M24-1	7	yes	0.2(7)			-0.6(9)	
		RMC-1-3	CLRX9010-5	1	yes		-1.3(19)	-0.2(5)	-0.6(11)	-0.2(8)
		RMC-5-3	Contig2221-1	5	yes		0.7(6)	-0.2(9)		

1 C-D: control treatment of the drought experiment; D: drought, C-SN: control treatment of the salt-nutrient experiment, N: nutrient deficiency, S: salt

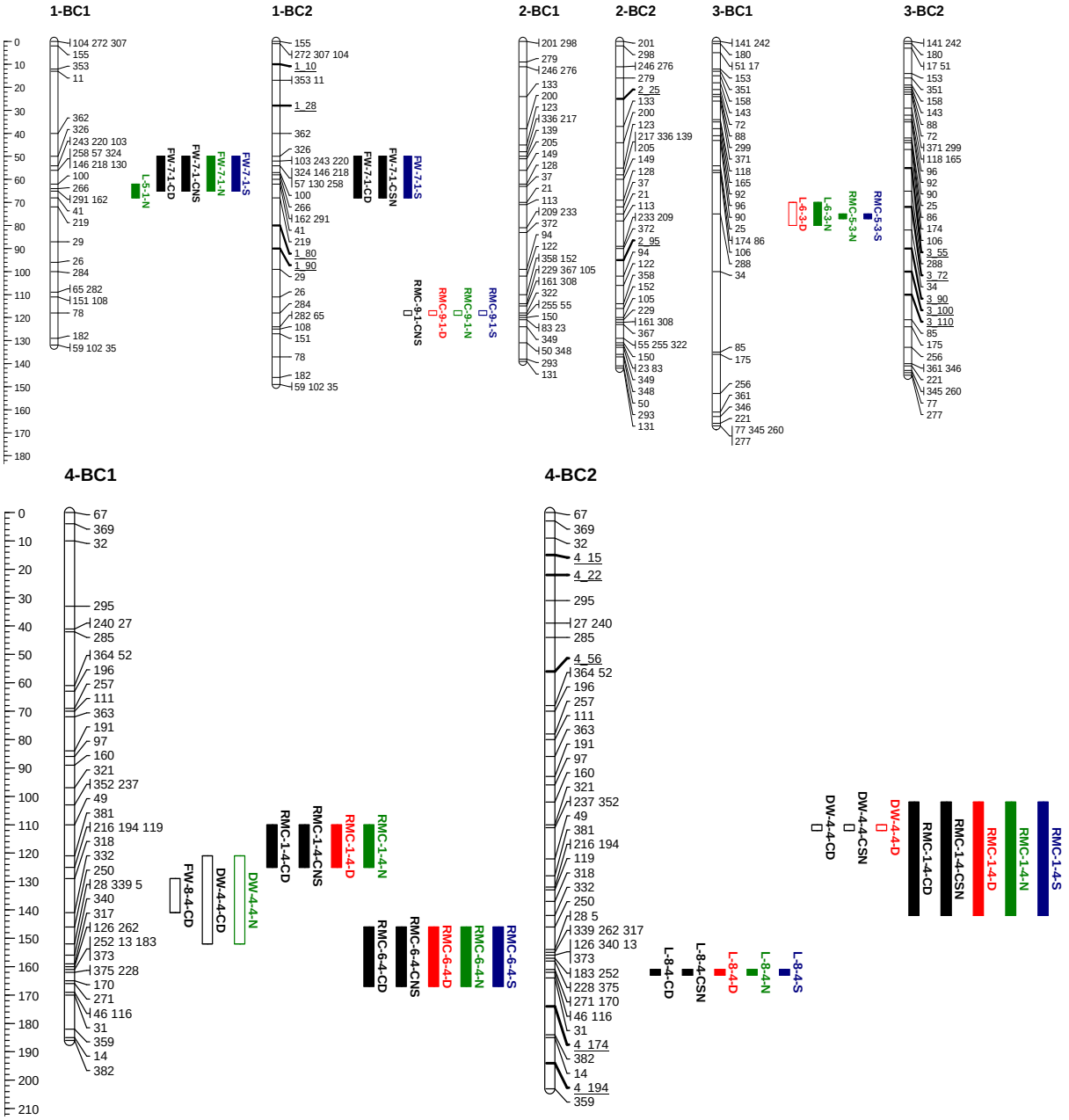


Figure 2 Linkage maps of the BC_1 and BC_2 populations based on. Markers (replaced by their indexes) are shown on the right of the bar. The added virtual markers on the BC_2 map with missing scores are underlined. Vigour QTLs as mapped in BC_1S_1 and BC_2S_1 families under non-stress (black), drought (red), salt (blue) and nutrient deficiency (green) conditions are shown next to the marker positions. Open QTL block indicate a positive additive effect for the wild allele, and closed QTL block indicate a positive additive effect for the crop allele. Trait abbreviations: L: plant length, FW: fresh weight, DW: dry weight, RMC: relative moisture content

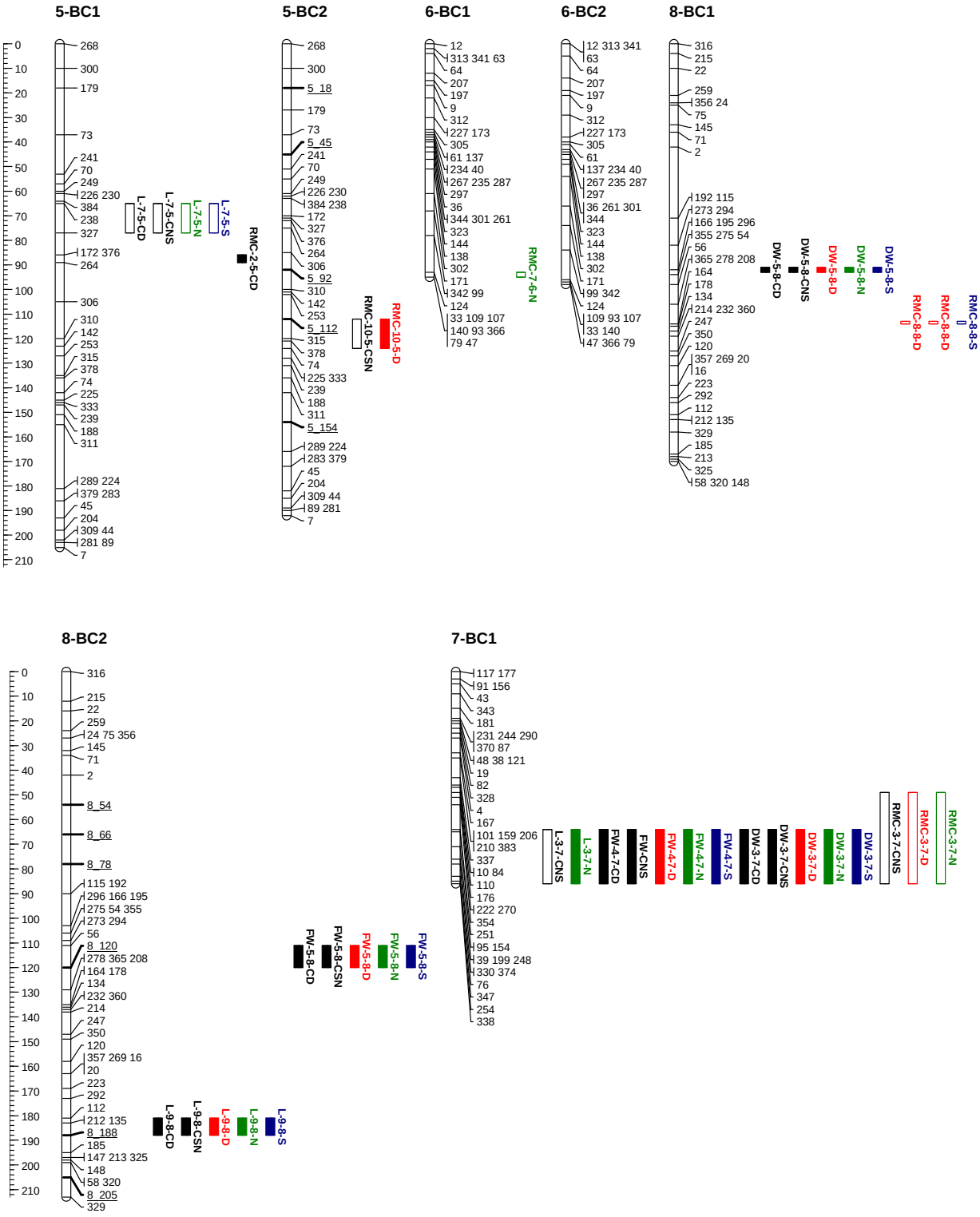


Figure 2 - continued

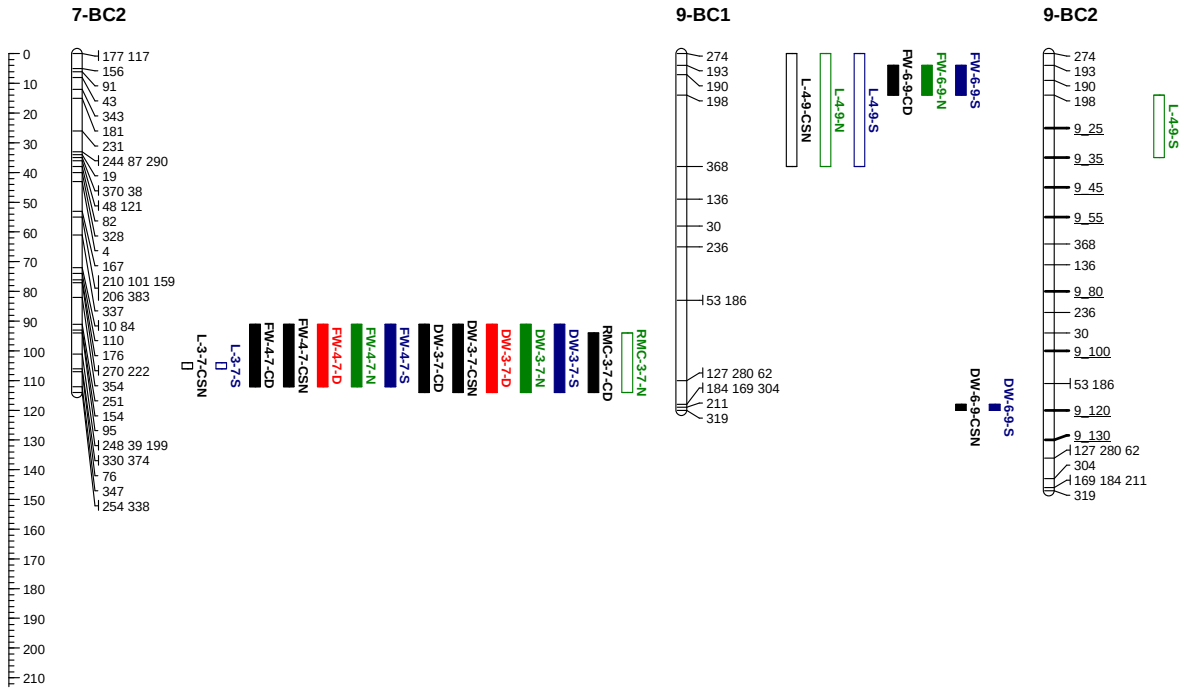


Figure 2 - continued

This was the case for *L-7-1* x *L-8-2* which affected plant height under the two control treatments and drought conditions with the highest mean associated with the hybrid-wild genotype combination (b/h) in the three cases. *L-9-1* x *DW-4-6* affected relative moisture content in the two control treatments and in the nutrient deficiency treatment with the highest mean associated with the hybrid genotype combination (h/h). Under nutrient deficiency conditions, *L-7-1* x *FW-1-2* affected three traits: plant height, fresh weight and relative moisture content.

While interacting QTLs for plant height had a higher mean for the crop-crop or wild-wild genotype combinations in the BC_1 population, the highest mean for the same trait was associated with crop-wild genotype combinations in the BC_2 population, showing the effect of the combination of QTLs inherited from the two parents in a repulsion phase. The genotype combination of a wild allele at the two epistatic loci (b/b) was associated with the highest mean for 3 out of 17 QTL pairs in BC_1 and 3 out of 23 QTL pairs in BC_2, indicating that the advantageous epistatic effect was mostly associated with the genotype combinations involving a crop allele at one of the two loci.

Co-localization of QTL regions

QTL regions on LG4 and LG7 were the most important in the two populations as they comprised most of the QTLs. Four QTLs were mapped on the same region on LG7 in the BC_1 and BC_2 populations, one for each of the measured vigour traits (Figure 2). The QTLs for

fresh weight and dry weight had the same allelic effect which was positive for the crop allele under all the treatments. However, the plant height and relative moisture content QTLs showed allelic specificity for treatments in the two populations. On LG4, four QTLs were mapped around the same region in BC_1 and the same region contained three QTLs in BC_2, including two QTLs that were common in the two populations. In total 8 QTLs were common in the BC_1 and BC_2 populations on LG1, LG4, LG7, and LG9. Additionally, a QTL region was found in both populations on LG8 but it contained QTLs for different traits in the two populations.

Table 3 Significant QTL x QTL interactions in the BC_1 and BC_2 populations, their explained phenotypic variance and the predicted means per genotype combination

				% expl.	Predicted mean per genotype combination2			
Pop.	Treatment[1]	Trait	QTLxQTL	variance	h/h	h/b	b/h	b/b
BC_1	Control-D	Plant height	*L-1-1* x *DW-4-6*	5	30.71	30.30	30.54	31.86
		Dry weight	*L-1-1* x *RMC-8-4*	4	2.947	3.10	2.832	2.578
		Relative moisture content	*L-7-1* x *L-5-2*	4	93.98	93.74	93.84	93.98
			L-7-1 x *RMC-5-1*	5	94.09	93.62	93.96	93.89
			L-1-1 x *L-5-2*	4	93.77	93.95	94.00	93.8
	Control-SN	Plant height	*L-9-1* x *RMC-5-1*	4	26.14	27.20	29.29	33.79
			L-3-3 x *RMC-8-4*	5	29.76	28.49	27.46	29.94
		Relative moisture content	*L-1-1* x *L-5-2*	7	92.18	92.40	92.30	92.11
	Salt	Plant height	*L-5-2* x *RMC-8-4*	4	17.85	16.61	17.01	17.68
		Fresh weight	*DW-9-6* x *RMC-5-1*	5	11.44	10.78	10.38	11.06
		Relative moisture content	*L-1-1* x *L-5-2*	7	90.07	90.34	90.25	90.02
			L-3-3 x *DW-8-3*	7	90.05	90.39	90.17	90.05
			L-3-3 x *RMC-8-4*	9	90.09	90.41	90.20	90.04
	Nutrient deficiency	Plant height	*L-3-3* x *RMC-8-4*	4	10.56	10.19	9.62	10.16
			DW-8-4 x *RMC-5-1*	5	9.85	10.41	9.18	10.79
		Dry weight	*L-9-1* x *DW-8-3*	4	0.67	0.59	0.58	0.58
		Relative moisture content	*DW-8-3* x *DW-9-6*	5	80.17	80.87	81.53	80.86
BC_2	Control-D	Plant height	*L-7-1* x *L-8-2*	5	28.87	29.04	31.06	28.69
		Fresh weight	*FW-1-2* x *DW-9-6*	5	37.80	40.99	41.41	37.33
			L-3-3 x *RMC-6-3*	8	29.61	40.00	39.01	38.89
			L-7-1 x *L-3-3*	8	33.83	39.58	41.74	38.49
		Relative moisture content	*L-9-1* x *DW-4-6*	11	93.43	92.82	92.82	92.77
			L-4-4 x *L-8-2*	4	92.77	93.04	92.85	92.78
	Drought	Plant height	*L-7-1* x *L-8-2*	5	14.42	14.29	15.11	14.02
	Control-SN	Plant height	*L-7-1* x *L-8-2*	5	21.20	20.69	23.71	21.16
		Fresh weight	*L-4-4* x *DW-9-6*	8	19.65	17.22	16.97	17.76
			L-3-3 x *RMC-6-3*	7	14.89	18.04	17.94	17.70
		Dry weight	*L-4-4* x *DW-9-6*	8	1.46	1.26	1.28	1.32
		Relative moisture content	*L-9-1* x *DW-4-6*	7	93.09	92.56	92.64	92.54
	Salt	Plant height	*L-9-1* x *RMC-5-3*	5	16.91	16.04	16.11	16.98
		Fresh weight	*DW-4-6* x *L-5-2*	4	11.50	10.34	9.88	10.44
		Dry weight	*L-8-2* x *FW-1-2*	5	0.93	0.95	0.98	0.87
	Nutrient deficiency	Plant height	*L-7-1* x *FW-1-2*	6	10.95	11.37	11.68	11.26
		Fresh weight	*L-7-1* x *FW-1-2*	6	5.42	5.70	5.47	5.06
			RMC-1-3 x *RMC-5-3*	5	5.73	5.20	5.19	5.41
		Relative moisture content	*L-9-1* x *DW-4-6*	6	87.87	86.45	86.84	86.56
			DW-4-6 x *RMC-1-3*	6	86.01	87.65	86.24	86.63
			L-7-1 x *RMC-1-3*	6	85.54	86.78	86.79	86.97
			FW-1-2 x *DW-9-6*	8	86.76	86.53	86.13	86.92

1.Control-D: control treatment of the drought experiment, Control SN: control treatment of the salt-nutrient experiment; 2 h: heterozygous genotype, b: homozygous for the wild allele

Discussion

Performance of crop-wild hybrid lines

Backcrossing of crop-wild hybrids to the wild parent is characterized by an increase of the genomic content of the wild parent accompanied by a decrease of the crop genome content in the form of fewer as well as shorter crop genome segments. As Hooftman *et al*. (2005) indicated, lettuce crop-wild hybrids are morphologically mostly not distinguishable from their wild parent *L. serriola*, and the likeness increases with further backcrossing to the wild parent. Despite the increasing resemblance with the wild parent among the BC_2S_1 individuals, the genetic variance for vigour-related traits did not change much from BC_1 to BC_2 as expressed by the broad sense heritability values for the traits in all the treatments, except for plant height under nutrient deficiency. This may indicate that the considered vigour traits are not genetically linked with the morphological traits which distinguish *L. sativa* and *L. serriola* parents (Hooftman *et al.* 2011).

Studies on introgression of crop genes into wild relative genomes have shown that although the average fitness of the hybrids might be lower than the fitness of the wild relative, individual hybrid plants could have similar or better fitness than their wild parent, showing a potential for introgression of advantageous crop genes (Hauser *et al.* 1998; Mercer *et al.* 2007). In our study the BC_1S_1 and BC_2S_1 families revealed lines showing transgressive segregation for vigour in the control and stress treatments, indicating that two generations of backcrossing to the wild parent did not eliminate the effect of the crop segments. The occurrence of BC_2S_1 families that outperform the wild parent shows that if vigour traits positively correlate with fitness under natural conditions, crop segments that confer improved vigour could be introgressed into the wild taxon, rendering it more vigorous under non-stress as well as under abiotic stress conditions.

QTL effects

Backcrossing has been applied in plant breeding for fine-mapping of QTLs and for the introgression of desired QTL alleles from wild donors into elite cultivars (Fulton *et al.* 2000; Ho *et al.* 2002; Robert *et al.* 2001). In crop-to-wild gene flow, repeated backcrossing to the wild parent is likely to take place as a result of the often much higher frequency of wild individuals compared to crop-wild hybrids. One of the direct consequences of repeated backcrossing to the wild species is the continuing decrease in crop genome segments, both in size as they become successively shorter and in frequency as each plant has fewer segments. Consequently, each backcrossing event is expected to reduce the detection power of QTL

analysis (Tanksley and Nelson 1996). Consistent with this, we detected more QTLs in the BC_1 population than in the BC_2 population for each of the considered vigour traits. However, despite the decreasing crop content, new QTLs with an additive effect from the crop allele were detected in the backcross populations compared to F_2. In the previous chapter (Chapter 3) plant height QTLs in the F_2 population were entirely inherited from the wild parent. In this study, two additional QTLs were mapped for the nutrient deficiency treatment (*L-1-1* and *L-3-3*) in the BC_1 population with an additive effect from the crop. In the BC_2 population we detected two more QTLs for plant height (*L-4-4* and *L-8-2*) with the same allelic effect in all the treatments which was positive for the crop allele, showing that the contribution of the crop to plant vigour could be underestimated depending on the population studied.

Seven QTLs were common between F_2, BC_1 and BC_2 populations on LG4, LG7 and LG9, three were common in at least two populations on LG1 and LG5, and one QTL was located in very close regions in the backcross populations on LG4 (Table 4). Detecting different QTLs in mapping populations of the same cross is a common ambiguity in plant breeding. The differences could be attributed to statistical power, especially with a number of lines in the population smaller than 200, to a combination of recessiveness, skewed linkage map (Jeuken *et al.* 2008), or to genetic variation between populations, with further backcrossing associated with decreasing QTL detection power (Tanskley and Nelson 1996). In the present study, the common QTLs were especially those with the greatest effects in terms of explained phenotypic variance per treatment and per trait, while the QTLs with small effect were mapped in one hybrid generation. Linkage groups 4, 7 and 9 were the most important in BC_1 and BC_2 populations as they showed regions that contained many and common QTLs in the two populations. The same regions were important in the F_2 population (see chapter 3).

Table 4 Recapitulation on common QTLs for vigour in the three hybrid populations F_2, BC_1 and BC_2 under non-stress (C), drought (D), salt (S) and nutrient deficiency (N) conditions. The positive sign shows a positive effect for the crop allele and the negative sign shown a positive effect for the wild allele

			F_2				BC_1				BC_2			
Trait	QTL	LG	C	D	N	S	C	D	N	S	C	D	N	S
Plant height	*L-7-1*	7	-		-	-	-		+		-			-
	L-9-1	9	-		-	-	-		-	-				-
Fresh weight	*FW-7-2*	7		+	+	+	+	+	+	+	+	+	+	+
	FW-8-1	8	+		-	+					+	+	+	+
	FW-9-2	9				+	+		+	+				
	FW-1-2	1					+		+	+	+			+
Dry weight	*DW-7-3*	7	-	+			+	+	+	+	+	+	+	+
	DW-4-6	4					-		-		-	-		
Relative moisture content	*RMC-4-3*	4	+	+	+		+	+	+		+	+	+	+
	RMC-5-1	5	+	+			+							
	RMC-7-4	7	+		-	+	-	-	-		+		-	

Despite the overlapping QTL regions across hybrid populations, some QTLs showed treatment specificity per population. For instance, *L-7-1* had a positive effect for the wild allele under nutrient deficiency conditions in the F_2 population, but the same QTL region showed a positive effect for the crop allele under the same treatment in the BC_1 population and it was not significant in the BC_2 population. Conversely, *RMC-4-3* was consistent across populations and treatments with a positive allelic effect from the crop, though it was not significant in the salt treatment of the F_2 population. Such QTL interactions suggest that the regions might contain different treatment-specific genes which contribute to the vigour of the plants. Moreover, QTLs for different vigour traits were mapped in those same regions with opposite allelic effect. Nevertheless, the involvement of the same regions in the vigour of the hybrids in three populations indicate that these regions will be under selection, either positive or negative, depending on the prevailing conditions, hence giving a hint on which regions to avoid when generating transgenic lettuce in order to avoid hitchhiking of neutral transgenes in the introgression process of crop genes into wild background. To our knowledge, this is the first study on introgression which combines a QTL analysis approach under different stress treatments to address the process of introgression.

QTL epistatic effect was significant for vigour traits in the two backcross populations. Epistasis has been suggested as one of the major allelic actions affecting fitness in self-pollinating species such as *Arabidopsis thaliana* (Malmberg *et al.* 2005) and rice (Mei *et al.* 2003). Epistatic QTL effects are expected to play a major role in selfing populations and to decline with further backcrossing as a result of decreasing genetic variation (Tanksley and Nelson 1996). Our results show that the vigour traits were affected by the epistatic effect of the QTLs under stress and non-stress conditions, and that positive epistatic effects were associated with genotype combinations involving the crop allele.

QTL epistatic effect in BC_1 and BC_2 populations emphasizes the genetic importance of the crop segments even after two backcrosses to the wild parent. Importantly, the combination of beneficial epistatic and additive allelic effects from two parents at different loci in repulsion phase has been associated with the origin of transgressive segregation that leads to the creation of superior or even ecologically diverging phenotypes (Latta *et al.* 2007; Lexer *et al.* 2003a; Lexer *et al.* 2003b). However, the fact that none of the QTL epistatic effect was common in the two populations makes the stability of the epistatic effect over generations questionable; makes make it difficult to predict the effect in further generations.

This study was carried out on plant vigour, based on the previous knowledge that lettuce crop-wild hybrids undergo selection at an early stage of growth (Hooftman *et al.* 2009). It was run on a narrow crop-wild genotypic range because the hybrids were from a cross between two single crop and wild genotypes. In addition, the study was conducted under greenhouse conditions, and measured growth and vigour-related traits only for the rosette stage of development. Therefore, the results should be considered as baseline rather than as conclusive. Future experiments will consider the whole life cycle of hybrid plants from seed germination to seed production under field conditions, hence covering early and late plant vigour, natural selection and survival, and reproduction.

Acknowledgements

This study was financially supported by the Netherlands' Organization for Scientific Research (NWO) through their research programme "Ecology Regarding Genetically modified Organisms" (ERGO). The development of the SNP markers and genotyping were carried out at the Compositeae Genome Project, UC Davis Genome Center and supported by the National Science Foundation Plant Genome Program Grant #: DBI0421630.

Chapter 5

Genetic Analysis of Vigour and Reproductive Traits in Lettuce Crop-Wild Hybrids under Field Conditions

Brigitte Uwimana[1], Marinus J.M. Smulders[1], Danny A.P. Hooftman[2], Yorike Hartman[3], Clemens C.M. van de Wiel[1], Richard G.F. Visser[1]

1 Wageningen UR Plant Breeding, Wageningen, The Netherlands

2 Centre for Ecology and Hydrology, Wallingford, UK.

3 IBED, Universiteit van Amsterdam, Amsterdam, The Netherlands

Abstract

While greenhouse experiments are useful in studying specific conditions such as stress factors, field experiments are preferred for ecological studies as they are closer to the natural conditions. In this chapter, the BC_1 population from a cross between *L. serriola* and *L.* sativa and backcrossed to *L. serriola* was used in two field experiments, in Wageningen and Sijbekarspel, the Netherlands. We studied the performance of the hybrids under natural conditions and determined the contribution of the two parents to germination, rosette vigour, adult vigour, survival and seed production of the crop-wild hybrids and the effect of genotype by environment interaction (GxE) on the traits. Under field conditions, the germination, vigour and reproductive traits were characterized by a moderate to high broad sense heritability (H^2 ranging from 0.41 to 0.89). All traits were affected by GxE except for the traits survival which was similar at the two sites. The crop contributed for the vigour of the hybrids at the rosette stage and for the number of branches which in is an important trait for seed production, and for the total number of seeds. QTL regions associated with the traits were mapped and suggestions of chromosome regions where to insert or not insert a transgene in order to minimize its likelihood of persistence under natural conditions were given.

Introduction

Since the mid 1980's, the development of genetically modified (GM) crop varieties has brought about a potential agricultural revolution which makes it possible to cope with the enormous challenge of feeding the increasing human population in spite of the increasing pressure on crop production due to biotic and abiotic stresses. However, the commercial cultivation of GM varieties has been associated with potential risks, both for humans and to the environment. Among the risks associated with the cultivation of GM varieties, the possibility of hybridization between GM crop varieties with their wild relatives has raised ecological concerns. On the one hand, the transfer of transgenes to wild relatives may produce more aggressive, and therefore difficult to manage weeds in agricultural areas. On the other hand, crop-wild hybrids resulting from hybridization may disturb the ecology and diversity of the wild relative taxon by displacing it in its natural habitats, hence causing genetic erosion (Auer 2008; Pilson and Prendeville 2004; Pirondini and Marmiroli 2008; van de Wiel *et al.* 2005).

The consequences of crop to wild gene flow have been argued to be negligible based on the logic that crop-wild hybridization would result in mal-adapted hybrids due to crop genes that are supposedly less fit under natural conditions (Hails and Morley 2005). However, various studies have reported successful establishment of crop-wild hybrids under natural conditions (Kiær *et al.* 2009; Morrell *et al.* 2005; Snow *et al.* 2010; Whitton *et al.* 1997). While crop species are evaluated according to their capacity to give the expected yield (e. g. grains, fodder, tubers, etc.), wild plants growing under natural conditions are under natural selection for their capacity to survive the harsh environmental conditions and to reproduce. Germination, survival and reproduction constitute the ultimate criteria that determine plant fitness under natural conditions because if a plant genotype does not germinate or survive or fails to produce seeds, it disappears from future populations. Crop-wild hybridization results not in a single, uniform hybrid, but rather in a swarm of hybrids and various offspring generations with a wide range of combinations of traits that may affect fitness. Hence, while selection by the prevailing natural conditions purges out less fit genotypes, those with fitness equal to or greater than the fitness of the wild parent have a chance to persist. Experimental studies on the consequences of crop-wild hybridization found individual hybrids that were just as fit as the wild parent (Gueritaine *et al.* 2002; Hauser *et al.* 1998; Hooftman *et al.* 2005). The effect of fitness-related traits might depend on the

experimental site, hence exhibiting genotype by environment interaction (GxE, Campbell *et al.* 2006; Dechaine *et al.* 2009).

Lactuca serriola L. is one of the wild species that is considered to be potentially affected by crop to wild gene flow due to its close relatedness with crop lettuce *L. sativa* and the crossability of the two species (De Vries 1990; Frietema de Vries *et al.* 1994). Therefore, *L. serriola* and *L. sativa* are one of the crop-weed complexes that have been the subject of crop-wild gene flow studies. These studies have established that *L. serriola* and *L. sativa* are completely cross-compatible and they produce viable and fertile hybrids (De Vries 1990; Hooftman *et al.* 2005), and that whenever the two species populations grow sympatrically, interspecific hybridization should be expected to occur (D'Andrea *et al.* 2008). Lettuce crop-wild hybrids undergo selection under natural conditions and the surviving hybrids are as fit as or more fit than the wild parent (Hooftman *et al.* 2009; Hooftman *et al.* 2005). In their modelling studies, Hooftman *et al.* (2007) and Hooftman *et al.* (2008) showed that, despite the low cross-pollination rate of lettuce, crop-wild hybrids could potentially displace *L. serriola* in its natural habitats.

This study aims at broadening the knowledge on lettuce crop-wild hybrids by investigating the genetic basis of the fitness of the hybrids and the contribution of the crop and wild parents to that fitness. Using a quantitative approach, we determined the contribution of the crop and wild genomic segments to the vigour, survival and reproduction of the hybrids under field conditions. In the previous chapters, we have focused on the vigour of the hybrids and the contribution of the two parents under non-stress and abiotic stress conditions of salinity, drought and nutrient deficiency conditions. We identified the major genomic locations which control hybrid vigour under those conditions, and determined the role of the two parents in selfing (F_2) and backcross (BC_1 and BC_2) generations of hybrids under controlled greenhouse conditions. The aim of the present study was to investigate the performance of the hybrid generations BC_1 under field conditions, by looking at the whole plant life cycle from germination through early plant vigour, adult plant vigour, and survival to seed production. Specifically, we will (i) determine how vigour at young plant stage correlates with adult plant vigour, reproduction and survival among the lettuce hybrids, (ii) explore the effect of genotype by environment interaction on germination, plant vigour, survival and seed production, (iii) determine whether the crop confers any (dis)advantage to the germination, vigour and reproduction of the hybrids under field conditions, and (iv) localize the QTLs for germination, vigour, survival and reproduction in the hybrids.

Materials and methods

Plant materials and experimental set up

Ninety-eight BC_1S_1 families, progeny of L. *serriola*/Eys x *L. sativa* cv. Dynamite and backcrossed to their *L. serriola* parent, were used in field experiments. These are the same families used in the greenhouse experiments, and the creation of the lines was described in Chapter 4. In the field, the BC_1S_1 families were grown together with their parents and another set of 98 lettuce recombinant inbred lines resulting from a cross between *L. serriola* acc. UC96US23 and *L. sativa* cv. Salinas (Truco *et al.* 2007) and their parents. For practical reasons, this chapter concerns the BC_1S_1 families and their parents only. The experiments were set up at two sites from spring until the end of fall of 2010 in the Netherlands. One site was in Wageningen (51° 59' North, 5° 39' East), where the soil was sandy; and the other site was in Sijbekarspel (52° 46' North, 06° 03' East) and was characterized by clayey soil (Soil data, Wageningen UR – Alterra, http://www.bodemdata.nl). The plots were ploughed to create disturbed soil, which is typical for the normal habitat of *L. serriola*. In Wageningen, no crop had been grown on the plot for the previous two years and a lot of weeds had been growing at the site. Therefore, the plot was sprayed with the herbicide Roundup three weeks before sowing. In Sijbekarspel no herbicide was applied. No fertilizers were applied at both sites. Seeds were sown directly on the field in 40 cm x 40 cm squares. The squares were spaced at 10 cm between squares in the same row and 20 cm between rows in Wageningen (10 cm in Sijbekarspel as the plot size was limited). The squares were set up in a randomized block design with 12 blocks. Each family/line occupied one square per block, resulting in 12 replicates. Thirty seeds per family/line per replicate were sown directly in the squares. In Sijbekarspel the seeds were sown on April 26, 2010 and in Wageningen the seeds were sown a week later on May 3, 2010. The plots were watered right after sowing so as to give the seeds a chance to germinate. After the seeds had germinated, the plots were hand-weeded in order to be able to score for germination.

The experimental period was characterized by a cold spring during the germination stage with temperatures as low as 0ºC at the two sites. The beginning of summer was dry and hot at the two sites with temperatures as high as 35ºC in Wageningen and 33ºC in Sijbekarspel. The rest of summer and the beginning of fall were mildly hot and wet, followed by chilly temperatures towards the end of fall (Figure 1).

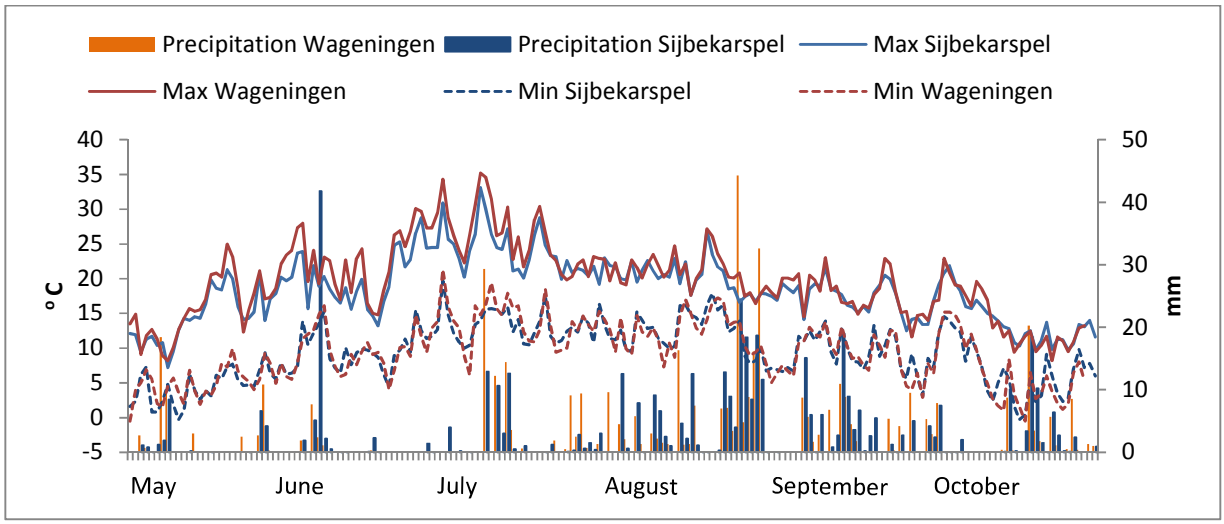


Figure 1 Temperature (left axis) and precipitation (right axis) in Wageningen and Sijbekarspel during the experimental period (data source: Wageningen UR Meteorology and Air Quality, Haarweg weather station, http://www.met.wau.nl; Koninklijk Nederlands Meteorologisch Instituut, Berkhout weather station, http://www.knmi.nl)

Data collection

Data were collected on seed germination, plant vigour at the rosette stage, plant survival, plant vigour at the adult stage and seed production (Table 1). Three weeks after sowing, germinated seedlings were counted for each square. After counting, the seedlings were thinned to 4 seedlings per square. Six weeks after sowing, two plants were collected per square and these were weighed after drying them at 80ºC for three days, giving the dry weight of the plants at the rosette stage. A week later, the plants were thinned again to one plant per square by choosing the plant closest to the centre of the square.

Two months after sowing, the plots were visited daily to check for the appearance of the first flower and first seeds for each plant, which were recorded as flowering time and seed set respectively. To assess for reproduction, seeds from 10 randomly chosen, mature capitula were collected per plant and counted and the average number of seeds per capitulum was derived. The number of basal shoots and branches per plant were counted as well. The total number of capitula per plant was estimated using the equation that Hooftman *et al.* (2005) developed:

$$\textit{Number of capitula} = (50.6 \times \textit{number of branches}) + (177 \times \textit{number of shoots}) - 5.3$$

The total number of seeds per plant was then estimated by multiplying the number of capitula per plant by the average number of seeds per capitulum. Survival was scored throughout the experimental period. In Wageningen, the height of each adult plant was measured as the length of the main shoot of the plant. Due to practical reasons, plant height was not measured in Sijbekarspel. By the end of October 2010 there were about 30 plants at each site that had not flowered yet and showed no sign of doing so because the temperatures were dropping gradually with the approaching winter period. For the analysis of seed set these plants were scored as not having survived.

Table 1 Traits measured in the field experiments

Trait	Stage of growth	Unit
Germination	Germination	count and %
Dry weight rosette	Vigour rosette	g
Survival	Survival	binary and %
Number of reproductive basal shoots	Vigour adult	number
Number of reproductive branches	Vigour adult	number
Adult plant height	Vigour adult	cm
Number of seeds per capitulum	Reproduction	number
Days to flowering	Reproduction	number
Total number of capitula	Reproduction	number
Total number of seeds	Reproduction	number

Analysis of phenotypic data

Phenotypic data were analysed using GenStat 13th edition (Payne *et al.* 2010). Continuous data (number of branches, number of basal shoots, days to flowering, number of seeds per capitulum, total number of capitula, and total number of seeds) were analysed by two-way ANOVA with blocking with all interactions, using the terms “genotype” and “site” as factors and “block” as a block factor. For germination and survival, an analysis of deviance was conducted with a logistic generalized linear model for germination data and an ordinal linear regression using a logistic link function for the binary survival data.

Broad sense heritability of the traits was estimated per site as a ratio between the genetic variance among the BC_1S_1 families and the total phenotypic variance:

$$H^2 = Vg/(Vg+Ve/r);$$

where Vg is the genetic variance among the BC_1S_1 families, Ve is the environmental variance, and r is the number of replications (Chahal and Gosal 2002). $Vg_{(BC1S1)}$ was estimated based on the restricted maximum likelihood (REML) method from the mixed model:

$$Response = general\ mean + block + \underline{BC_1S_1} + error;$$

with the *Response* term representing the measured traits, and the term BC_1S_1 taken random. *Ve* was the error variance derived from a one-way ANOVA of the model: *Response = general mean + block + parents + error*; with the term *parents* representing the two parents of the BC_1S_1 families and the two additional lines (*L. serriola* acc. UC96US23 and *L. sativa* cv. Salinas).

We performed a path analysis to determine how vigour at the rosette and adult stages correlates with survival and seed production. The programme was written in GenStat and it was run on the means for continuous data and on percentages for germination and survival. Path analysis converts correlation coefficients between a set of independent variates and one response variate into direct and indirect effects. It gives a path coefficient between each dependent variate and the response variate which is a standardized partial regression between the two variates (Agbicodo 2009). Path analysis was carried out separately for Wageningen and Sijbekarspel. The response variates were survival and the total number of seeds and the remaining traits were the independent variates.

QTL analysis

The linkage map and genotypes of the BC_1 individuals, which were determined in Chapter 4, were combined with the phenotypic data of the field experiment in a QTL analysis. The trait "total number of seeds" was analysed in two ways, first by considering that all dead plants produced no seeds and giving them the value "0", and second by considering them as missing, hence removing the bias imposed by survival. The single-trait multi-environment QTL mapping function of GenStat 14th edition was used for QTL analysis (Payne *et al.* 2011), using the procedure described by Mathews *et al*. (2008). After determining the genetic predictors and the best fitting variance-covariance model between the two sites (Malosetti *et al.* 2004), QTL candidates were determined by genome-wide scan (SIM) with a significance level of 0.05 corrected for multiple tests by the Li and Ji method (Li and Ji 2005). The final candidate QTLs were defined by composite interval mapping (CIM) using the candidate QTLs as cofactors. The significance of each detected QTL in a specific site was determined by fitting a multi-environment multi-QTL model and running a backward selection on the candidate QTLs (Mathews *et al.* 2008).

Results

Phenotypic variation and genotype by environment interaction

The phenotypic data are summarized in Table 2. The traits showed moderate to high broad sense heritability, H^2, ranging from 0.52 to 0.89 in Sijbekarspel and from 0.41 to 0.89 in Wageningen, indicating a substantial genetic variation of the traits at the two sites. The range of the BC_1S_1 families out-bounded the means for *L. serriola* and *L. sativa* parents for germination, vigour and reproduction traits at the two sites, showing a transgressive segregation of the traits over the two parents. Consequently, some BC_1S_1 families had higher means than the two wild *L. serriola* lines for vigour and reproduction traits, indicating improved vigour and fitness as a result of crop-wild hybridization. Seventy-nine BC_1S_1 families had a higher survival rate, 73 had more number of seeds, and 70 BC_1S_1 families had a higher or equal means for the two traits than or as the wild parent than *L. serriola*/Eys (their wild parent) in Sijbekarspel. In Wageningen, the numbers were lower, with 42 BC_1S_1 families having a higher or equal survival rate, 4 having more number of seeds and 4 having a higher means for the two traits. Comparing the hybrids with the two *L. serriola* lines, in Sijbekarspel 13 BC_1S_1 families had a higher or equal germination than the best of the two *L. serriola* lines, 62 had more number of seeds and 13 BC_1S_1 families had higher means than the two *L. serriola* lines. In Wageningen 22 BC_1S_1 had a higher or equal survival rate than the best of the two *L. serriola* lines, 4 had more number of seeds and one BC_1S_1 family had a higher mean than the two *L. serriola* lines for the two traits.

Temperatures play a key role in lettuce germination with the increase in temperature associated with poor germination (Argyris *et al.* 2005). The low temperatures that occurred after sowing ensured a good germination of the seed at the two sites, with a higher germination rate recorded in Wageningen. Although *L. sativa* lines (cv. Dynamite and Salinas) germinated better than *L. serriola* lines and had higher dry weight at the rosette, they did not survive the field conditions, as only one plant of cv. Dynamite produced seeds in Wageningen and both lines died before flowering in Sijbekarspel. Conversely, the two *L. serriola* parental lines had a higher survival rate, but mortality of *L. serriola*/Eys in Sijbekarspel was high (50% survival vs. 91% in Wageningen). For the BC_1S_1 families, survival ranged from 8 to 100% with 25% and 20% of the BC_1S_1 plants failing to produce seeds in Sijbekarspel and Wageningen, respectively, with the plants starting to die in the 12^{th} week after sowing.

Vigour and reproduction traits depended on the genotype of the BC_1 used to produce BC_1S_1 seeds ($P_{genotype}$ <0.001) and they differed from site to site ($P_{genotype.site}$ <0.001, Supplementary material Table S2), showing significant genotype by environment interaction (GxE) for those traits among the BC_1S_1 families. Conversely, survival, though dependent on site (P_{site}<0.001), did not show significant GxE ($P_{genotype.site}$=0.484 and $P_{genotype.site.block}$=1).

Table 2 Mean performance of the BC_1S_1 families, the parental lines and the additional two lines and broad sense heritability of the traits at Sijbekarspel and Wageningen

		BC_1S_1 families			*L. serriola* /Eys	*L. sativa* cv. Dynamite	*L. serriola* acc. UC96US23	*L. sativa* cv. Salinas
Trait	Site*	Mean	Range	H^2				
Germination (%)	Sbk	51.14	22.78 – 70.56	0.83	48.33	62.22	25.28	60.83
	Wgn	63.04	34.17 – 83.61	0.85	66.11	82.22	35.00	72.78
Dry weight rosette (g)	Sbk	0.86	0.52 – 1.37	0.63	0.76	1.01	0.54	1.08
	Wgn	1.24	0.66 – 1.98	0.72	0.99	1.34	0.96	1.54
Days to flowering	Sbk	109	88 – 137	0.89	116	128	95	116
	Wgn	96	81 – 113	0.84	103	122	82	104
Reproductive basal shoots (number)	Sbk	11	5 – 18	0.62	10	0	4.16	0
	Wgn	8	5 – 13	0.89	10	1	1.25	0
Reproductive branches (number)	Sbk	28	20 – 36	0.52	31	0	36	0
	Wgn	28	22 – 35	0.41	29	25	42	0
Total number of capitula	Sbk	3386	1992 – 4267	0.64	3392	0	3566	0
	Wgn	2897	2159 – 3757	0.68	3239	4**	2333	0
Seeds per capitula (number)	Sbk	12	8 – 17	0.88	13	0	9.96	0
	Wgn	16	13 – 22	0.90	18	13**	19.08	0
Total number of seeds	Sbk	30306	4202 – 55029	0.85	22480	0	26034	0
	Wgn	36643	5558 – 62538	0.88	53067	52**	45171	0
Adult plant height (cm)	Sbk	119.9971	93.58 – 156.58	0.89	155.50	72.78	-	-
Survival (%)	Wgn	72.53	16.67 – 100	-	50.00	0	100	0
	Sbk	80.10	8.33 – 100	-	91.67	8.33	100	0

* Sbk: Sijbekarspel, Wgn: Wageningen; **Absolute numbers for one *L. sativa* plant that survived and the 4 capitula that gave seeds

Correlations among the traits

Plant height was not included in the path analysis in Sijbekarspel because it was not measured. Despite this slight difference in traits included in the analysis, the path coefficients at the two sites were similar for most of the traits (Figure 2). In Wageningen, dry weight at the rosette stage positively correlated with seed germination, number of basal shoots and late flowering; but negatively correlated with the number of branches and plant height. In Sijbekarspel, dry weight positively correlated with seed germination rate and late flowering. Time of flowering showed a strong direct and indirect relationship with survival as late flowering plants tended to die at the two sites (Pa=-0.64 in Wageningen and -0.76 in Sijbekarspel), while plant height positively related to survival (Pa=0.31; Figure 2A). While germination rate positively correlated with

rosette vigour (r=0.38 in Wageningen and 0.45 in Sijbekarspel), germination rate was not related to survival or seed production at the two sites as the path coefficients were close to zero. Plants with high dry weight at the rosette stage tended to die as shown by the correlation coefficients between the plant height and survival (r = -0.31 in Wageningen and -18 in Sijbekarspel). However, the path coefficients between dry weight at the rosette stage and survival were less strong at the two sites (Pa=-12 in Wageningen and -0.03 in Sijbekarspel), probably due to the indirect correlations through other traits such as germination rate and the indirect path from the number of basal shoots to the height of the plants.

Survival had a strong relationship with the number of seeds (Pa=0.70 in Wageningen and 0.67 in Sijbekarspel), which was understandable because seed production is the eventual outcome of survival. Although dry weight at the rosette stage did not relate with the total number of seeds (Pa close to 0 at the two sites), it was positively correlated with the number of basal shoots which in turn positively correlated with the number of seeds, showing a direct path from vigour at the rosette stage and vigour at the adult stage. Time of flowering had an important relationship with survival and consequently with the total number of seeds as the plants that flowered early tended to have a better survival rate and a higher total number of seeds (Figure 2B).

QTL analysis

Twenty QTLs were detected for vigour, survival and reproduction traits in Sijbekarspel and Wageningen, and they were found on all linkage groups (LG) except for LG2, as summarized in Table 3 and Figure 3. For each trait one to three QTLs were found, except for the number of basal shoots and the total number of capitula for which no QTLs were detected. The crop allele contributed positively to dry weight at the rosette stage, to days from sowing to flowering and to the total number of seeds. All traits had been analysed both in Wageningen and Sijbekarspel with the exception of plant height. In total, thirteen QTLs were detected both in Wageningen and Sijbekarspel, always with the same direction of the allelic effect (from the crop or from the wild parent).

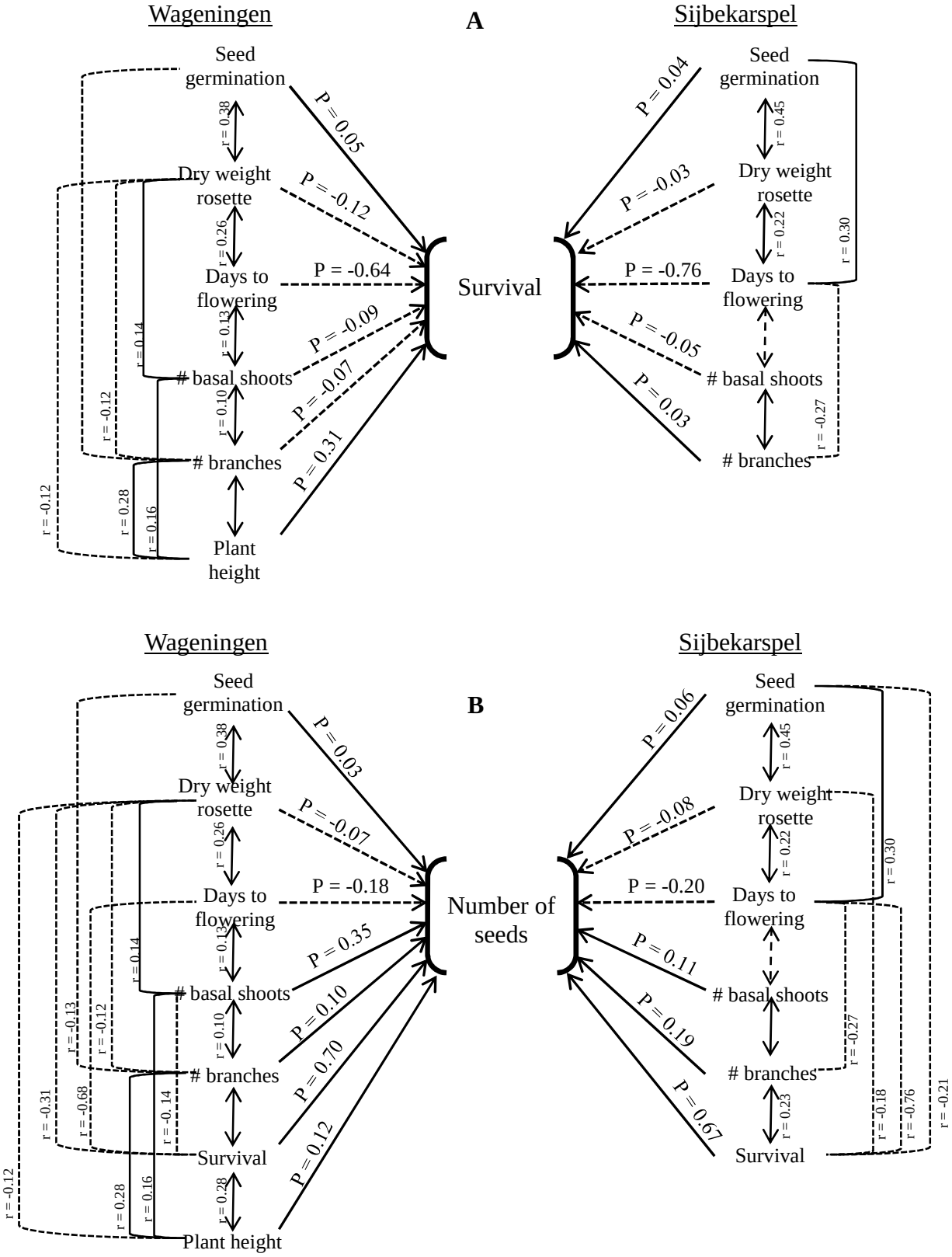


Figure 2 Path coefficients (Pa) between vigour traits and survival (A) and seed production (B) and the correlations coefficients among the vigour traits (r) at Sijbekarspel and Wageningen. Only the correlations

greater than 0.10 or smaller than -0.10 are shown and negative correlation and path coefficients are shown by dashed lines

Table 3 QTLs associated with germination, vigour, seed production and survival detected in the BC_1 population in Sijbekarspel and Wageningen

				Additive effect for the crop allele (% explained variance)	
Trait	**Closest marker**	**LG**	**QTLxE**	**Sijbekarspel**	**Wageningen**
Germination (%)	CLS_S3_Contig7056-1-OP5	8	n/a[1]	-7.13(13)	-5.58(6)
Rosette dry weight (g)	QGF25M24-1	7	n/a[1]	0.14(17)	0.26(20)
Survival (%)	QGB25B18-1	6	no	-21.16(26)	-21.16(33)
	QGB11B18.yg-2-OP5	7	no	-16.54(16)	-16.54(20)
Adult plant height (cm)	CLS_S3_Contig6255-6-OP5	1	n/a[1]	-	-7.76(43)
	CLS_S3_Contig2201-5-OP5	9	n/a[1]	-	-11.15(89)
Days to flowering	QGB25B18-1	6	n/a[1]	5.66(9)	5.38(15)
Number of branches	RHCLS_S3_Contig9046_2	4	n/a[1]	2.47(12)	1.75(10)
Number of seeds per capitulum	Contig1262-2	3	yes	-1.45(11)	-
	Contig2010-3	3	no	-0.88(4)	0.88(7)
	CLS_S3_Contig2201-5-OP5	9	no	-1.37(10)	-1.37(16)
Number of seeds-1 (dead plant given 0 seeds)	QGB25B18-1	6	no	-9475(18)	-9475(21)
	QGB11B18.yg-2-OP5	7	no	-9197(17)	-9197(20)
	CLS_S3_Contig7056-1-OP5	8	no	3776(3)	3776(3)
	CLS_S3_Contig2201-5-OP5	9	no	-5969(7)	-5969(9)
Number of seeds-2 (dead plant considered as missing)	QGF21B10.yg.ab1_PAP2_LE1382_12	3	yes	-6967(17)	-
	Contig4740-1	5	yes	-	-5138(13)
	CLS_S3_Contig2201-5-OP5	9	no	-5452(10)	-5452(14)

1 n/a: not applicable QTL x environment effect not applicable because one QTL was detected for the trait

GxE for the traits was observed through QTL by environment interaction (QTLxE) as well whenever there were more than one QTL detected per trait. QTLxE was characterized by a QTL detected at one site but not at the other or by a difference in magnitude of effect from one site to another.

QTLxE was not significant for two traits, namely survival and total number of seeds estimated by considering dead plants as plants producing zero seeds. For survival, the non-significance of QTLxE was in line with the phenotypic results as GxE was not significant for the same trait. For the three QTLs for the number of seeds calculated by assigning zero seed production to the dead plants, the lack of significant QTLxE indicated the effect of survival on that trait, because when the dead plants were considered as missing, three other QTLs were detected and their QTLxE effect was significant. The twenty mapped QTLs were located on 9 regions of 8 LGs with vigour QTLs overlapping with reproduction QTLs (Figure 3). On LG7, a dry weight QTL co-localized with a QTL for survival and a QTL for the total number of seeds. On LG4, a QTL for seeds per capitulum was close to a QTL for the number of branches, and on LG9, three QTLs related to seed production overlapped with a QTL for plant height.

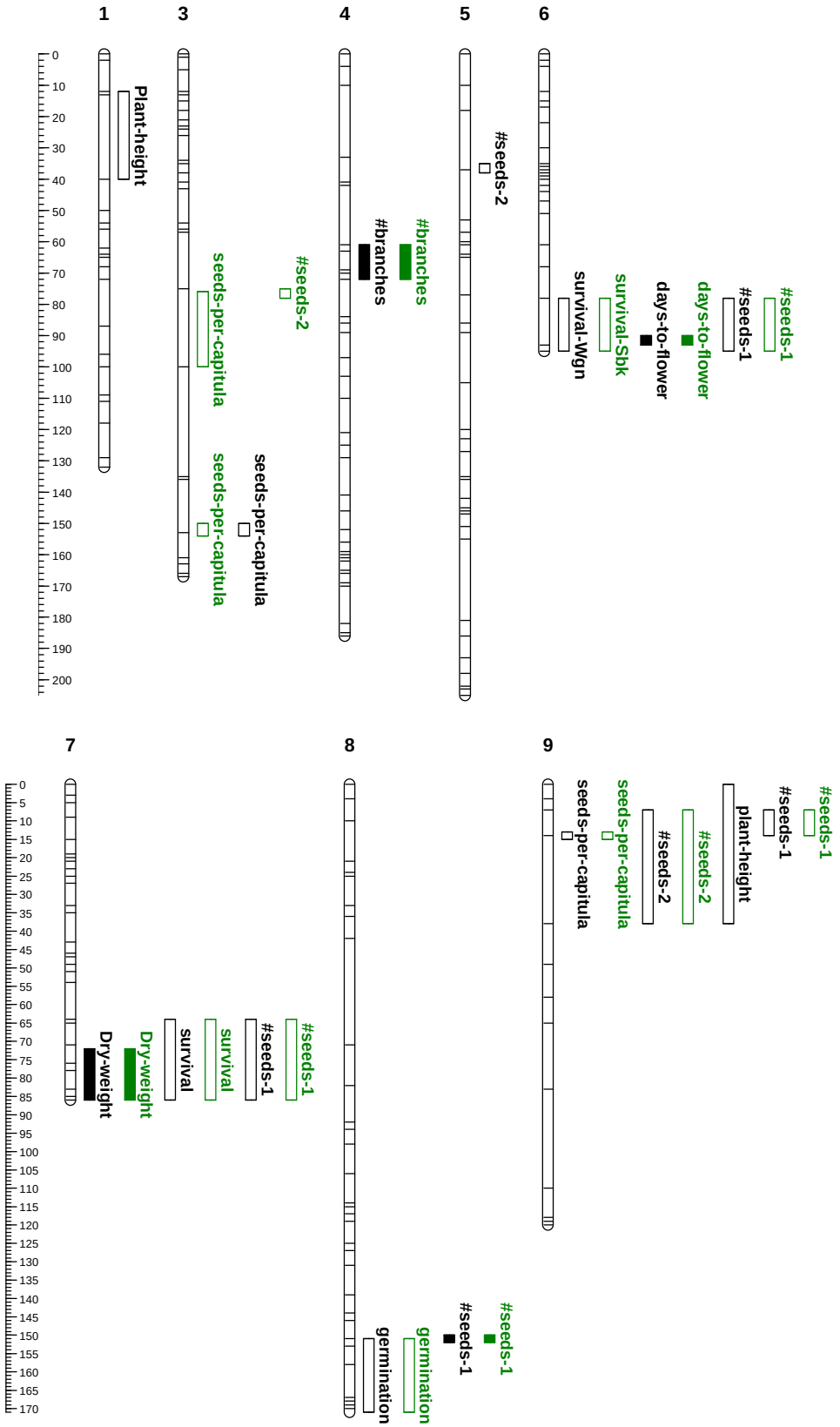


Figure 3 Localization of the QTLs detected for plant vigour, survival and reproduction in Wageningen (black blocks) and Sijbekarspel (green blocks) on the BC_1 linkage map. Open blocks: QTL positive for the wild allele; filled blocks: QTL positive for the crop allele; green; #seeds-1: total number of seeds estimated by including the dead plants; #seeds-2: total number of seeds estimated by excluding the dead plants

Table 4 Significant QTL epistatic effect on traits in the field experiments. QTLs are represented by the LGs on which they are located

			Predicted means per genotype combination[1]			
Trait	QTLxQTL		h/h	h/b	b/h	b/b
Wageningen						
Number of basal shoots	LG7xLG5	7	8.63	8.74	7.48	9.33
Total number of capitula	LG7xLG5	8	2926	2905	2696	3073
Plant height	LG9xLG4	5	120.50	108.60	125.70	124.50

1 h: heterozygous genotype, b: homozygous genotype for the wild allele

The epistatic effects of the QTLs on the traits were not significant in Sijbekarspel. In Wageningen, three traits were affected by QTL epistasis, namely the number of basal shoots, the number of capitula and the height of the plants (Table 4). QTL regions on LG5 interacted with the region on LG3 to affect the number of basal shoots and the number of capitula and increasing their phenotypes by 7 and 8% respectively. The highest value for these traits was associated with the homozygous genotypes for the wild allele at the two loci. Plant height was affected by the interaction between the QTL regions on LG4 and 9 with an increase of the phenotypic variance of 5% and a greater phenotype was associated with the combination of the homozygous genotype for the wild allele on LG9 and the heterozygous genotype on LG4.

Discussion

Despite the close relatedness between *L. serriola* and *L. sativa* (Frietema de Vries 1992; Koopman *et al.* 1998; Koopman *et al.* 2001), their hybrids were characterized by a substantial genetic variation for vigour and reproduction traits as indicated by the broad-sense heritability of the measured traits. This genetic variation resulted in transgressive segregation with some individual hybrids being as fit as or more fit than the wild genotypes in terms of survival and seed production. This contradicts the assumption that wild genotypes are already adapted to harsh natural conditions and therefore hybrids are bound to be less fit than them (Hails and Morley 2005; Latta and McCain 2009). In our study we observed at the two sites individual hybrid plants and hybrid families that performed equally well as or better than the wild genotypes in germination, survival and seed production, indicating that hybrids could compete or outperform the wild parent. Similar results were obtained by Hooftman *et al.* (2005) in field experiments on lettuce hybrids from the same parental lines and including one common field site (Sijbekarspel) among the ones used at the time, hence indicating the consistency of hybrid fitness over a wide range of environmental conditions represented by different years and sites (Wageningen). Hooftman *et al.* (2005) reported a gradual decrease in hybrid vigour and fitness with further generations of the hybrids. Despite this vigour and fitness breakdown, they also reported that the hybrids still showed better fitness than the wild plants in generations as advanced as the third generation backcrossed to the wild parent (BC_3)

and the fourth generation of selfing F_1, indicating that hybrid vigour and fitness was not completely due to heterosis.

The hybrid families showed similar survival rates in Wageningen and in Sijbekarspel, hence the non-significance of GxE and QTLxE for this trait, although the effect of GxE was significant on the vigour and reproduction traits. The non-significance of GxE on survival indicates a potential for predicting the persistence of the hybrids in different environments based on their genotypes. The remainder traits measured on the field were characterized by GxE effect. However, the QTLs for each trait were common in the two environments, with QTLxE effect arising from differences in the magnitudes of the QTL between Wageningen and Sijbekarspel (for 11 out of 14 QTLs). However, generalization of these results is limited by the fact that the experiments were run in one year and at locations with limited climatic differences (Figure 1), although the two sites differed in soil type, which greatly influences water retention capacity and the type of weeds that occur. Moreover, the hybrids resulted from a single crop-wild cross. In their multi-environment study on the relative fitness of crop-wild sunflower hybrids, which encompassed various crop-wild cross combinations and different wild populations, Mercer et al. (2006) reported that the relative fitness of hybrids depended on the population and on the local abiotic and biotic stress conditions. Therefore, the results of our study should be confirmed in multi-year-multi-environment experiments including hybrids from other lettuce crop-wild crosses.

Even though early stages of growth are important for the vigour and reproduction of the adult plants, whether early growth traits such as germination and early vigour predict survival and seed production depends on the developmental processes and the growth stage at which selection takes place (Donohue *et al.* 2010). Temperature is one of the major factors affecting seed germination in cultivated and wild lettuce as high temperatures are associated with poor germination (Marks and Prince 1982; Valdes *et al.* 1985). In our study, the environmental conditions of low temperature and moist soil were favourable for germination, leading to limited selection at this stage of growth, and the first plants died in the 12^{th} week after sowing when some plants had started to flower. The results contradict those obtained by Hooftman *et al.* (2005) who did not moisten the soil after sowing and observed a high selection during germination and a high mortality of the plants in the first month after sowing (Hooftman *et al.* 2009). The late selection in our experiments could explain the low correlations between early growth traits (germination and rosette dry weight) and adult vigour, survival and reproduction traits.

The same BC_1 population was used in greenhouse experiments for vigour at the rosette stage under drought, salt and nutrient deficiency conditions (Chapter 4). Dry weight at the rosette stage showed high correlations between field and greenhouse experiments with the highest correlations observed between nutrient deficiency conditions and the two sites (Wageningen and Sijbekarspel) followed by drought and Sijbekarspel (Table 5). The same QTL region on LG7 was identified as associated with dry weight under all the treatments in the greenhouse and on the field at the two sites with a positive effect from the crop allele. Therefore, early plant vigour in the greenhouse can be well linked to early plant vigour on the field. On the field, germination positively correlated with dry weight at the rosette stage and the latter positively correlated with the number of basal shoots at the adult stage, indicating a positive correlation between early growth and adulthood. However, dry weight at the rosette stage weakly correlated with survival and the number of seeds, making it difficult to establish a relationship between early vigour and survival and between early vigour and reproduction, although the relationship could be drawn indirectly vial the number of branches. Therefore, the greenhouse assessments only showed limited capabilities for predicting reproductive fitness in the field.

Table 5 Pearson's correlation coefficients for dry weight between field and greenhouse experiments in the BC_1S_1 families based on means

Field / Greenhouse	Wageningen (P-value)	Sijbekarspel (P-value)
Control drought	0.25 (0.010)	0.35 (<0.001)
Drought	0.28 (0.004)	0.49 (<0.001)
Control Salt-Nutrient	0.32 (0.001)	0.37 (<0.001)
Nutrient deficiency	0.53 (<0.001)	0.54 (<0.001)
Salt	0.46 (<0.001)	0.47 (<0.001)

On the field, crop alleles contributed positively for biomass at the rosette stage (QTL on LG7), for the number of branches at the reproductive stage of the plants (QTL on LG4) and for the total number of seeds (QTL on LG8) with the three QTLs detected both in Wageningen and Sijbekarspel. These regions are likely to undergo positive selection under natural conditions because they directly or indirectly influence the reproduction of the hybrids. However, the QTL for dry weight on LG7 is complex because it overlaps with QTLs for survival and total number of seeds which are the most relevant traits determining the fitness of the plants. The same region contains a QTL for the number of lateral roots at the deep end of the tap root with a positive effect from the wild allele (Johnson *et al.* 2000). With regards to genetic hitchhiking and background selection and the introgression likelihood of transgenes based on their genomic location, such regions should be avoided when inserting transgenes because the genes in their genomic neighbourhood are likely to be hitchhiked, regardless of their effect on fitness (Kwit *et al.* 2011; Snow *et al.* 2010; Stewart *et al.* 2003).

In addition to the above-mentioned traits likely to be advantageous under field selection, crop alleles were also associated with late flowering (QTL on LG6) which is a favourable trait in lettuce crop due to its association with late bolting (Carvalho Filho *et al.* 2009; Ryder and Milligan 2005; Waycott *et al.* 1995). Conversely, late flowering is undesirable under field conditions as it negatively influenced survival and the total number of seeds (Figure 2). The QTL conferring late flowering overlapped with QTLs for survival and the number of seeds whose effect was positively inherited from the wild parent. Therefore this QTL region (LG6) could be a good candidate for the insertion of transgenes because the crop allele will most likely be selected against under natural conditions, resulting in background selection on the genes linked to it. However, it should be stressed again that these results are based on one crop-wild cross and one-year experiment. To confirm these QTL regions, multi-year and multi-location experiments should be conducted. Multi-year experiments with multiple generations of the same hybrids should also be used to confirm the genetic hitchhiking and background selection around the regions.

Acknowledgements

The authors would like to thank Dr. Oene Dolstra (Wageningen UR Plant Breeding) for allowing them to use his programme for path analysis. Acknowledgements. This study was financially supported by the Netherlands' Organization for Scientific Research (NWO) through their research programme "Ecology Regarding Genetically modified Organisms" (ERGO). The development of the SNP markers and genotyping were carried out at the Compositeae Genome Project, UC Davis Genome Center and supported by the National Science Foundation Plant Genome Program Grant #: DBI0421630.

Chapter 6

General Discussion

General discussion

Hybridisation between crops and their wild relatives is a possible route along which transgenes could 'escape', i.e. disperse from a crop to related wild species and establish themselves in the environment. Studies on crop-wild hybridization have focused on various aspects of this process, namely the rate of hybridization (Arias and Rieseberg 1994; Ureta *et al.* 2008), the occurrence of spontaneous crop-wild hybrid plants (Bing *et al.* 1996; Magnussen and Hauser 2007), the performance of hybrids relative to the wild parent under laboratory, greenhouse and field conditions (Campbell *et al.* 2006; Hooftman *et al.* 2009; Mercer *et al.* 2006a; Snow *et al.* 2001), the persistence of hybrids under natural conditions (Kuroda *et al.* 2010; Kuroda *et al.* 2008) and modelling the long term fate of crop-wild hybrids (Hooftman *et al.* 2007a; Hooftman *et al.* 2008; Huangfu *et al.* 2011). However, the contribution of crop genomic blocks to the performance of the hybrids has received little attention, with a few exceptions. To our knowledge, only studies in *Helianthus* (Baack *et al.* 2008; Dechaine *et al.* 2009) and *Brassica* (Rose *et al.* 2009) have addressed this issue.

While the rate of dispersal of the gene will depend on the frequency of hybridisation, establishment will be determined by a combination of the genetic make-up of the hybrids, the environmental conditions and the population dynamics of the wild populations. In this regard, the transgene does not occur isolated but imbedded in a region of the chromosome of the crop. For several generations of inbreeding or backcrossing to the wild species any fitness effect will depend on the effect of such a genomic block, not on that of a single gene. This thesis focuses on the contribution of the crop and wild parent to the performance of the hybrids in terms of vigour, survival and reproduction using a genetic approach of mapping populations, molecular markers and QTL analysis.

The genomic location of a transgene has been suggested as one of the ways to limit its likelihood of introgression. If a transgene is located in a region close to a gene or QTL conferring reduced fitness, the transgene will have a lower likelihood of introgression once the crop hybridises with a wild relative because it will be selected against along the gene or QTL affecting fitness (Kwit *et al.* 2011; Stewart *et al.* 2003). This thesis aimed at determining the contribution of the crop to the vigour and fitness of the hybrids and the effect of linkage on the fate of crop segments in the hybrids using *L. serriola* and *L. sativa* as a model for crop-weed complex. Because of the importance of the abiotic stress factors in natural selection, and because abiotic stress factors are the subject of present and future genetic transformation, we conducted greenhouse experiments on the contribution of the crop on the vigour of the

hybrids under the major abiotic stress conditions (drought, salinity and nutrient deficiency). Specifically, the thesis aimed at answering the following research questions: (i) whether there is evidence of spontaneous hybridization between *L. serriola* and *L. sativa* under natural conditions, (ii) whether crop genes confer any (dis)advantage to the crop-wild hybrids under controlled non-stress conditions, under controlled conditions with abiotic stress, and under field conditions, (iii) whether the (dis)advantageous effects of the crop are dependent on environmental conditions, and (iv) whether we can identify genomic regions where transgenes could be inserted with the aim of mitigating their persistence based on the localization of the QTLs.

Crop-wild hybrids grow under natural field conditions in which conditions will vary during the growing season. Therefore, field experiments are preferred to greenhouse experiments. However, greenhouse experiments offer certain advantages over field experiments, such as the number of experiments that can be run in a certain period of time, and, more importantly, in mimicking a certain stress factor so that the tolerance or resistance of the plants to the stress can be deciphered (Latta *et al.* 2007; Latta and McCain 2009). We conducted greenhouse experiments to study the tolerance of the hybrids to abiotic stress conditions of salinity, drought and nutrient deficiency at the rosette stage, expressed as plant height, fresh weight, dry weight and relative moisture content (Chapters 3 and 4). Two field experiments were carried out (Chapter 5) in order to link vigour under greenhouse conditions with vigour, survival and reproduction of the hybrids under field conditions. Below, I will discuss the above-mentioned research questions with regards to the experimental results of the previous chapters.

Evidence of spontaneous hybridization between *L. serriola* and *L. sativa*

A number of methods have been used for the identification of hybrid plants. Among these methods is screening based on morphological traits (Ureta *et al.* 2008), tracking crop-specific markers (Arias and Rieseberg 1994; Scurrah *et al.* 2008; Westman *et al.* 2001) and, in case of GM crops, tracking the transgene itself (Warwick *et al.* 2008). However, these methods are not applicable in all cases. For lettuce hybrids resulting from a cross between *L. serriola* and *L. sativa*, the use of morphological traits would not be possible because the hybrids cannot always be distinguished with a good degree of certainty from their wild parent (Hooftman *et al.* 2005). The use of a transgene as a marker to find evidence of hybridization among natural populations in lettuce is also not applicable because no transgenic lettuce variety has been released so far.

To establish whether hybridisation occurred, and in which way to identify such hybrid plants, we genotyped *L. serriola* and *L. sativa* genebank accessions from the lettuce collection of the Centre for Genetic Resources, the Netherlands (CGN) and another large set of *L. serriola* populations collected from its natural habitats in Europe from 2002 until 2005 (ANGEL *L. serriola*), with 10 microsatellite markers. It was found (Chapter 2) that the application of single, crop-specific alleles in identifying potential crop-wild hybrids in lettuce is inadequate. However, Bayesian analysis, such as implemented in the programme STRUCTURE (Pritchard *et al.* 2000), identified potential intermediate hybrid plants which constituted 7% of the ANGEL *L. serriola* dataset and 9% of the CGN *L. serriola* collected from the wild. The programme NewHybrids (Anderson and Thompson 2002) categorized the *L. serriola* hybrid plants identified with STRUCTURE into two hybrid classes: advanced selfing generation after hybridization between *L. serriola* and *L. sativa* and advanced selfing generation after one back-cross to *L. serriola.*

Hybridization between *L. serriola* and *L. sativa* has been hypothesized as one of the reasons behind the recent northward spread of *L. serriola* in Europe (Frietema de Vries *et al.* 1994; Hooftman *et al.* 2006). Based on the proportion of the crop-wild hybrids among *L. serriola* populations relative to the "non-admixed" *L. serriola* plants and to the geographical location of the hybrids, this hypothesis was rejected in Chapter 2. If introgression was behind the spread of *L. serriola*, we would expect to find more putative *L. serriola* hybrids than non-admixed *L. serriola*, particularly in North-Western Europe where the new invasiveness of *L. serriola* was most obvious. Moreover, we would also expect to observe more hybrids among the recently collected (between 2002 and 2005) *L. serriola* ANGEL data set than among the CGN accessions, most of which were collected decades earlier. Both expected patterns were not visible in our data. Therefore, other causes such as climatic warming, increase in anthropogenically disturbed areas and the spread of seeds through transportation networks are more likely responsible for the invasive behaviour of *L. serriola* (D'Andrea *et al.* 2009; Hooftman *et al.* 2006; Lebeda *et al.* 2004).

This study provides evidence that spontaneous hybridization occurs among basically self-pollinating species with limited outcrossing rate such as *L. serriola* and *L. sativa*. The results are in line with those in wheat, another self-pollinating species for which evidence of gene flow from cultivated and wild species was found in the Mediterranean region (Arrigo *et al.* 2011). The results lend credence to D'Andrea *et al.* (2008) who, based on hybridization experiments, concluded that whenever *L. serriola* and *L. sativa* grow in sympatry, the two species should be expected to hybridize. The occurrence of crop-wild lettuce hybrids among

L. serriola populations also shows that natural selection does not completely purge crop alleles or genomic segments, or not fast enough compared to the incidence of new hybridisation events. In a modelling study, Hooftman *et al.* (2008) indicated that, although the outcrossing rate could have an effect on the composition of a population after crop-wild hybridisation, the fitness of the hybrids had a greater effect. Therefore, once transgenic lettuce varieties will have been developed and grown in the field, the low rate of outcrossing in lettuce will allow the occurrence of introgression of a transgene from transgenic *L. sativa* into *L. serriola* x *L. sativa* hybrids. However, whether the transgenic hybrids can subsequently persist under natural conditions will depend on the fitness of the hybrids which in turn will depend on the genetic makeup of the hybrid including the effect of the transgene itself and the surrounding genes, and the prevailing environmental conditions. That is the subject of the following sections where we look at the vigour and fitness of the hybrids and how the crop and wild parents may contribute to that under greenhouse and field conditions.

Performance of the hybrids

Under greenhouse conditions, the hybrids showed a wide range of phenotypic variation for vigour-related traits, which was mainly due to genetic factors as broad sense heritability values of the traits ranged from moderate to high under stress and non-stress conditions in the selfing ($F_{2:3}$, Chapter 3) and backcrossing (BC_1S_1 and BC_2S_1, Chapter 4) populations. In the F_2 population, individual hybrid plants and hybrid families were found that performed better than the two wild genotypes used in the experiments, among which the wild parent of the hybrid lines, hence showing that improved vigour of the hybrids over their wild parent is possible under non-stress conditions and stress conditions of salinity, drought and nutrient deficiency. Introgression of crop segments takes place through repeated backcrossing of the hybrids to the wild parent, hence we can expect a gradual decrease of crop genomic segments, both in frequency and size, in increasing wild genomic background (Baack and Rieseberg 2007). Despite the reduction in crop segments among the backcross hybrids (backcrossed to *L. serriola*), hybrids performing better than the wild parent were also observed among BC_1S_1 and BC_2S_1 hybrids, indicating that two backcrossing generations did not completely eliminate the effect of crop segments.

Under field conditions (Chapter 5), vigour and fitness traits were characterized by moderate to high broad sense heritability, indicating that crop-wild hybridization results in new genetic combinations. Moreover, for each of the field-measured traits (germination, rosette and adult vigour, flowering time, survival and seed production) there were BC_1S_1 hybrid families that had the same performance as or even a better performance than the two wild genotypes. For

survival and seed production, 13% of the BC1S1 families had equal or higher survival rate and 63% had a higher number of seeds than the two *L. serriola* lines in Sijbekarspel, whereas the numbers were 22% for survival rate and 4% for seed production in Wageningen. The results contradict the theory that crop-wild hybrids are less fit than the wild plants due to domestication alleles inherited from the cultivated species (Hails and Morley 2005). Previous studies have reported a decrease in fitness among lettuce hybrids over generations (Hooftman *et al.* 2005), which could be due to heterosis breakdown (Burke and Arnold 2001). However, the dominance effect of 16 out of 17 QTLs associated with vigour traits was not significant in the F_2 population, and neither were the heterozygous genotypes associated with higher predicted means for QTL epistatic effect in the same population (Chapter 3), indicating that heterozygosity was not the most likely cause of the vigour of the hybrids. Moreover, Hooftman *et al.* (2005) reported that, despite the reduced fitness among the hybrids over generations, hybrids belonging to the 3rd generation of backcrossing to the wild parent and the 4th generation of selfing of the F_1 plant were on average still fitter than their wild parent *L. serriola*. Therefore, although the differences in vigour between offspring of crosses and the wild parent gradually become smaller, the improved vigour and fitness of the hybrid as a result of additive main effects and additive interaction effects of the genes/QTLs are likely to persist in some offspring over several generations. Given sufficient selective advantage, and taking into account genetic drift in small weedy populations, this might locally result in the displacement of the wild parent (Hooftman *et al.* 2007a; Hooftman *et al.* 2008).

Contribution of the crop to the vigour and fitness of the hybrids

In the three hybrid populations (F_2, BC_1 and BC_2), the crop contributed to the vigour of the hybrids at the rosette stage under stress and non-stress conditions through additive and epistatic allelic effects as revealed by QTL analysis. In the F_2 population, the crop contributed to vigour through fresh weight and relative moisture content in the control, drought and salt treatments as the QTLs for these traits had a positive effect from the crop allele. In the nutrient deficiency conditions, some QTLs had a positive effect for the crop allele, while others had a positive effect from the wild allele. QTLs for dry weight were inherited from both parents under drought and salt conditions, whereas it was entirely inherited from the wild parent under control and nutrient deficiency conditions. The contribution of the two parents could be linked to domestication and selection effect. Plant weight has been the subject of positive selection in cultivated lettuce, as it is mostly harvested at the rosette stage to be consumed as salad (Ryder and Whitaker 1976). Conversely, plant height, a trait selected against in cultivated lettuce because of its association with early bolting (Carvalho Filho *et al.*

2009; Fukuda *et al.* 2011), was inherited from the wild parent under all the treatments. In the backcross populations (BC_1 and BC_2), fresh weight was again mainly inherited from the crop alleles under all the treatments. Dry weight was inherited from both parents under all the treatments except for the salt treatment under which the trait was inherited entirely from the crop. Relative moisture content was also inherited from both parents as QTLs were identified with a positive effect from the crop allele and others with a positive effect from the wild allele. Interestingly, for plant height, which was exclusively inherited from the wild allele in the F_2 population, additional QTLs for the trait were detected in the BC_1 and BC_2 populations with a positive effect from the crop allele, which could be the effect of genetic variation due to fewer crop segments in increasing wild genetic background (Tanksley and Nelson 1996).

Under field conditions, the crop contributed to the vigour of the BC_1S_1 hybrids at the rosette stage which was measured as dry weight. The dry weight QTL detected on LG7 in the BC_1 population in the greenhouse under the three stress and non-stress conditions was also detected in the field with an explained variance of 17 and 20%. Moreover, the crop contributed positively for the number of branches (QTL on LG4) which was positively correlated with the total number of seeds produced (Pa = 0.35 in Wageningen and 0.11 in Sijbekarspel) and for the total number of seeds (QTL on LG8). However, the QTLs for plant height, survival and the other QTLs for the number of seeds were inherited from the wild parent under field conditions. Therefore, the crop contributes partially to the vigour and fitness of the hybrids. The contribution of the crop to vigour and reproduction indicates that the crop alleles in the crop-wild hybrids are not bound to be purged by selection. In crop-wild sunflower a similar trend was obtained where crop alleles which contributed to vigour and reproductive traits were under positive selection under field conditions (Baack *et al.* 2008).

Effect of genotype by environment interaction

Fitness effects are relative to the environment in which they are being measured. For instance, Dechaine *et al*. (2009) observed in sunflower filed experiments that crop alleles conferring an increase in size were favoured in crop-wild hybrids, but at one field site only if three types of herbivores were present. Therefore, the effect of GxE is an important aspect in crop-wild hybridization as wild plants grow over a wide range of environmental conditions. If a QTL is present but with different magnitudes in different environments, this still allows a prediction of the performance of the hybrids but requires additional experiments, whereas the occurrence of QTLxE due to the presence/absence of the QTL or opposite allelic effect leads to a lack of correlation between environments, and renders any generalization impossible. In our experiments the vigour and reproductive traits depended on the environment under

greenhouse and field conditions. Under field conditions, survival was the only trait not affected by GxE and the QTLs associated with this trait had the same direction and magnitude (QTLxE not significant). The remaining traits (seed germination, dry weight at the rosette stage, number of branches, number of basal shoots, seeds per capitulum, total number of capitula and total number of seeds) were affected by GxE and the QTLxE effect was significant for the QTLs associated with those traits. The non-significance of GxE on survival means that the survival of the hybrids can be predicted based on their genotypes. In contrast, significant GxE means that the hybrids perform differently from one environment to another (based on the tested sites), hence making it difficult to generalize the results. Despite the significance of QTLxE for the rest of the traits, most of the QTLs were common between different treatments in the greenhouse and between the two field sites, with the same direction of the effect but with different magnitudes. Only few QTLs under greenhouse conditions had opposite allelic effect from one treatment to another (Chapters 3 and 4), and in the field a small number of QTLs (3 out of 16) was significant at one site but not significant at the other (Chapter 5).

QTLs for vigour and fitness and the likelihood of introgression of a transgene based on its genomic location

One way of limiting the likelihood of introgression of a transgene could be to target it to a chromosomal region where it is linked to a crop gene/QTL that generally confers reduced fitness. As the gene affecting fitness is selected against, the linked transgene will be reduced in frequency along with the gene in question, resulting in background selection against the transgene (Kwit *et al.* 2011; Stewart *et al.* 2003). Indeed, the chromosomal position of the transgene was speculated to negatively affect the fitness of the hybrids from a cross between transgenic oil seed rape and wild radish (Gueritaine *et al.* 2002). Up till recently, genetic transformation was a random event, without any prior knowledge on the genomic area where the transgene would be inserted. Advances in biotechnology have made it possible to transform a specific site (Cermak *et al.* 2011; Shukla *et al.* 2009; Townsend *et al.* 2009; Urnov *et al.* 2010). In the future, these technologies promise the possibility to target a transgene to a specific location in the genome. If we then could choose a region conferring reduced fitness in crop-wild hybrids, and be able to avoid regions containing a gene or QTL positively affecting fitness, this would be an additional mechanism to contain the transgene. But how well do we know where these regions are in lettuce, and how well can we predict that the putative regions identified will indeed confer the predicted fitness effect to the crop-wild hybrids and their offspring across a range of natural conditions?

Based on the current cross between *L. serriola*/Eys and *L. sativa* cv. Dynamite, we suggest some regions on the lettuce genome where a transgene could be inserted or not inserted with the aim of mitigating its persistence after crop-wild hybridization (see Figure 1). The QTL region on LG7 should be avoided when inserting a transgene. This region was associated with the vigour traits (plant height, fresh weight, dry weight and relative moisture content) at the rosette stage under greenhouse conditions of non-stress, salinity and nutrient deficiency in F_2, BC_1 and BC_2 populations. Under field conditions with the BC_1 population, the same region was associated with plant dry weight at the rosette stage, survival and number of seeds (Figure 1). However, the region had an antagonistic allelic effect, with plant height, dry weight and relative moisture content sometimes inherited from the crop and other times from the wild under greenhouse conditions. Under field conditions, dry weight was inherited from the crop while survival and the number of seeds were inherited from the wild allele. Therefore, it would be difficult to predict the direction of selection, whether the region will be characterized by hitchhiking or background selection, and the best option would be to avoid the region altogether. LG4 contained many vigour QTLs under greenhouse conditions (for plant height, fresh weight, dry weight and relative moisture content) with positive effects from both the crop and the wild alleles (Chapters 3 and 4). Under field conditions, one QTL for the number of branches with a positive effect from the crop allele was detected on LG4. However, the field QTL did not overlap with the greenhouse QTLs, resulting in a scatter of QTLs on LG4. Therefore, predicting the direction of selection on LG4 would be difficult, and the advice would be to avoid inserting a transgene on it as well. The same applies for the QTL region on LG8 and it should be avoided as well. The QTL region on LG9 contained common QTLs between greenhouse and field experiments. Under greenhouse stress and non-stress conditions, the crop allele contributed positively for fresh weight and relative moisture content, whereas the wild allele contributed for plant height. However, under field conditions only the wild allele contributed for plant height at the adult stage and for the number of seeds per capitula and the total number of seeds. Therefore, under natural conditions, the crop allele at the QTL region on LG9 is likely to be selected against, thus making the region a good candidate for the insertion of a transgene. Likewise, the QTL regions on LG1 and 6 are likely to favour the wild allele and are candidates for the insertion of a transgene. The region on LG1 contained a QTL for plant height under field conditions with a positive effect from the wild allele. Under greenhouse conditions, the region on LG6 contained QTLs for plant height and relative moisture content with a positive effect from the wild allele. Under field conditions, the region contains QTLs for survival and seed production with a positive effect from the wild allele and a QTL for flowering with late flowering associated with the crop

allele. Therefore, the region will favour the wild allele under field conditions and the crop will be selected against, leading to background selection of the (trans)genes linked to the region in the crop genome.

Based on segregation distortion and changes in linkage disequilibrium after the hybrids had been exposed to field conditions for two consecutive years, Hooftman *et al.* (2011) also suggested lettuce regions on which transgenes could be inserted with the aim of reducing their likelihood of persistence due to genetic hitchhiking. Unfortunately, a comparison between their results and ours was not possible due to a lack of common markers between their map and ours. An indirect comparison through the lettuce integrated map, which encompasses most of the SNPs used in the present study (Truco *et al.* 2007), was neither possible as common markers per LG were still limited to one or two. In the present study we used SNP markers because of the possibility for high throughput genotyping (http://dnatech.genomecenter.ucdavis.edu/illumina.html) and co-dominant scoring, and because they were developed from Expressed Sequence Tags (ESTs), making the detected QTL close to actual genes (McHale *et al.* 2009). Moreover, most of the SNPs are already part of the latest version of the lettuce integrated map which makes our results comparable to other crop-wild studies which used the said map such as Johnson *et al.* (2000) and Argyris *et al.* (2005) (Chapters 3 and 5), and any future studies which will use the same map.

Segregation distortion of a region towards one of the parents on a linkage map can be used to determine a chromosome region with limited likelihood of introgression (Hooftman *et al.* 2011). In the F_2 population a region on LG3 was identified with a segregation distortion towards the crop allele (Chapter 3). Such a distortion, if stable over generations, could also be exploited for limiting the introgression of transgenes. It would be better to avoid inserting a transgene into the region as it would increase the likelihood of persistence of the transgene once the crop hybridizes with a wild relative. However, the segregation distortion on LG3 was not observed in the backcross populations, which was most probably due to the fact that the distortion in F_2 was biased towards the homozygous genotype of the crop allele, and detecting the same distortion in the backcross population would not be feasible as the crop genome is represented by the heterozygous genotypes.

Can small-scale contained experiments predict potential ecological consequences?

Although field experiments are more representative of natural conditions, they are labour-intensive and depend on the prevailing environmental conditions so that only one experiment can be run per year in temperate regions for instance, and several years may be needed to

encounter various combinations of weather conditions across the growing season. Conversely, greenhouse experiments, though less representative of natural conditions, are easy to handle, can be manipulated to simulate a specific selection factor so as to understand its mechanism, and many experiments can be run in one location simultaneously and the whole year round. One important question that arises is whether and to what extent can small-scale contained greenhouse experiments predict field conditions. Our experiments can only partially answer this question because the greenhouse experiments were run on young plants (5 weeks after transplanting) only. Nevertheless, the greenhouse and field experiments showed some similarities based on the correlations of dry weight at the rosette stage between greenhouse and field measurement and the detection of the same QTL for dry weight at the rosette stage in the two environments (Chapter 5). Therefore, we can precautiously conclude that greenhouse experiments can explain at least some of the effects occurring under field conditions. Importantly, the rosette stage data did not completely predict the fitness of the plants in the field experiments, which underlines the notion that fitness-related experiments need to cover the whole life cycle of a plant. Therefore, one should ideally run complete life cycle experiments (from germination to seed production), in the field as well as in the greenhouse.

Prospective research on crop-wild hybridization in lettuce

Various crop-wild introgression studies have been carried out on the change of crop allele frequency among crop-wild hybrids over multiple years with the aim of predicting the fate of the crop alleles under natural conditions (Cummings *et al.* 2002; Snow *et al.* 2010). In lettuce, exposure of the hybrids to field conditions over two years resulted in post-zygotic segregation distortion of the alleles at specific loci, although the role of the loci with respect to fitness was not known (Hooftman *et al.* 2011). In the present study, QTL regions were identified that affected hybrid vigour under greenhouse stress and non-stress conditions of salinity, drought and nutrient deficiency as well as QTLs affecting vigour and those affecting fitness of the hybrids under field conditions. However, the present study was based on a single crop-wild cross, whereas QTLs are relative to the genetic background in which they are measured, hence making it as yet impossible to generalize our results on lettuce in general. In their study on the germination of hybrid seeds from multiple crop-wild crosses in sunflower, Mercer et al. (2006a) reported an increase in seed germination rate and decrease in dormancy due to crop-wild hybridization. However, germination depended on the crop-wild crosses and the environment, and the performance of the crop-wild hybrids relative to the wild species depended on the crop-wild cross, the wild populations and the environmental conditions

(Mercer *et al.* 2006b). In their experiments on radish crop-wild hybrids Campbell *et al.* (2006) found that the hybrids had greater survival and fecundity then the wild lines in a new environment, different from the natural habitat of the wild parent. Therefore, a follow-up of the present study is recommended which should encompass multi-year-multi-location experiments to study the consistency of the QTLs detected in the present study. The predicted genetic hitchhiking and background selection around the QTLs should be assessed through multi-year field selection followed by intensive genotyping with markers within and around the mapped QTLs. A study on the performance of lettuce crop-wild hybrids with another crop-wild cross is underway (http://home.medewerker.uva.nl/y.hartman), hence covering the multi-cross aspect. Their results will complement ours and enrich the knowledge on hybridization, introgression and the fate of the hybrids in lettuce.

Other natural selection aspects not considered in our experiments were seed germination in the following year after exposure to field conditions, herbivory, competition and diseases. Although germination was scored in the field, the seeds had been produced in the greenhouse and stored under optimum conditions of temperature and pressure. Moreover, the demographic life cycle of the plants was not complete in our study as the field-harvested seeds were not tested for germination. Competition was found significant on the fitness of radish crop-wild hybrids (Campbell and Snow 2007) but not significant in transgenic oilseed rape and wild radish crop-wild hybrids (Gueritaine *et al.* 2002). In our study, competition was not fully allowed in our experimental field plots as herbicide was applied in Wageningen and the two plots were weeded to enable the counting of seedlings, and herbivory (cf. (Dechaine *et al.* 2009) was not scored. Hooftman *et al.* (2007b) studied the introgression of downy mildew, the most important crop lettuce disease (Michelmore *et al.* 2008) and they found that the occurrence of the disease on lettuce hybrids and wild genotypes did not affect their reproductive fitness, suggesting that diseases constitute a major problem to crop lettuce but not to wild lettuce, and the introgression of disease resistance genes will not have any effect on the crop-wild hybrids. We therefore suggest the consideration of germination after seeds have remained on the soil under field conditions, herbivory and field competition in future studies in order to have a more complete picture of selection under field conditions.

Implications for lettuce breeding

Wild species of lettuce have been used in lettuce breeding. *Lactuca saligna* was found to be a good source for downy mildew resistance (Jeuken and Lindhout 2002; Zhang *et al.* 2009), *L. virosa* was used as a source for aphid resistance (Van der Arend *et al.* 1999) and *L. saligna* and *L. virosa* accessions have been identified that are resistant to northern root knot nematode

(Kaur and Mitkowski 2011). *L. serriola*, which is considered as part of the primary gene pool of *L. sativa* (Koopman *et al.* 1993) was identified as a potential source for lettuce germination QTLs under adverse conditions (Argyris *et al.* 2005) and various disease resistance genes (McHale *et al.* 2009). The QTLs identified in this study could benefit lettuce breeding for salinity, drought and nutrient deficiency tolerance. The moderate to high broad sense heritability values for the studied vigour traits under the said stress conditions show that lettuce tolerance to those stress factors can be improved at the vegetative growing phase through breeding and selection. However, in order to avoid linkage drag from *L. serriola*, fine mapping of the QTLs will be required to distinguish between pleiotropic and linked QTLs and, if possible, to separate unwanted from useful regions. One of such regions is on LG4 where a dry weight QTL under control, salinity, drought and nutrient deficiency with a positive effect from the wild allele overlaps with a QTL for plant height (Chapter 3) which is an undesired trait in cultivated lettuce.

Major conclusions

1. In spite of the low out-crossing rate in crop and wild lettuce, we found evidence of spontaneous hybridization between *L. serriola* and *L. sativa* among natural populations of *L. serriola.*
2. The geographical location and frequency of lettuce crop-wild hybrids led to the conclusion that crop-wild hybridization is not the main reason for the recent spread of *L. serriola* in Europe.
3. Cultivated lettuce contributes positively to the vigour of the hybrids at the rosette stage under non-stress, salinity, drought and nutrient deficiency conditions through additive and epistatic allelic effects.
4. This contribution is sustained over two backcrossing generations to *L. serriola*, indicating the potential of introgression of crop segments in the increasing wild genetic background.
5. Under field conditions, cultivated lettuce contributes positively to plant vigour at the rosette stage, to the number of branches and to the total number of seeds.
6. Although plant vigour and seed production traits were dependent on the environment per hybrid family, survival was not affected by GxE under field conditions, hence showing the potential to predict the survivorship of a plant based on its genotype.
7. Based on the location and allelic effect of the QTLs for germination, vigour, survival and reproductive traits, we have suggested some genomic regions where transgenes could be inserted in order to mitigate their persistence in crop-wild hybrids.

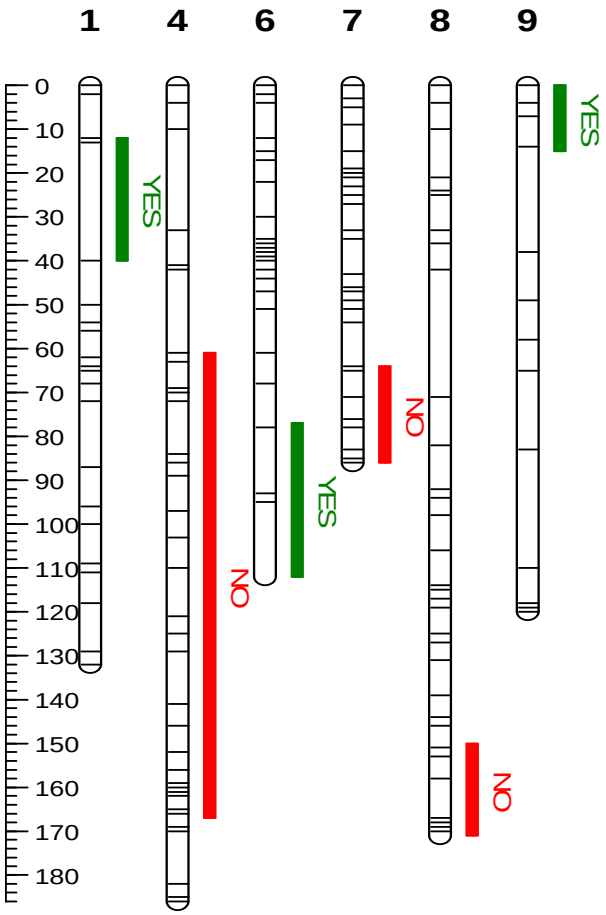


Figure 1 Suggestions of genomic regions where a transgene could be inserted or not inserted to mitigate its persistence after crop-wild hybridization based on the localization and allelic effect of vigour and fitness QTLs of this study. The shown map is that of the BC_1 population. The red bars indicate where a transgene should not be inserted because the segments are likely to undergo selection in favour of the crop allele, and the green bars indicate regions where a transgene could be inserted because the segments are likely to undergo selection in favour of the wild allele

References

Abdeen A, Schnell J, Miki B (2010) Transcriptome analysis reveals absence of unintended effects in drought-tolerant transgenic plants overexpressing the transcription factor ABF3. *BMC Genomics* **11**, 69.

Agbicodo EM (2009) *Genetic analysis of abiotic and biotic resistance in cowpea [Vigna unguiculata (L.) Walp.]*, Wageningen UR Plant Breeding.

Aharoni A, Dixit S, Jetter R, Jetter R, Thoenes E, van Arkel G, Pereira A (2004) The SHINE Clade of AP2 Domain Transcription Factors Activates Wax Biosynthesis, Alters Cuticle Properties, and Confers Drought Tolerance when Overexpressed in Arabidopsis. *The Plant Cell Online* **16**, 2463-2480.

Albacete AA, Martínez-Andújar C, Pascual JA, Acosta M, Pérez-Alfocea F (2008) Increasing vegetative growth, yield and seed quantity in tomato by inducing plant vigour at the earliest seedling stage. In: *Acta Horticulturae* (ed. Leskovar DI), pp. 265-271.

Andersen NS, Siegismund HR, Meyer V, Jørgensen RB (2005) Low level of gene flow from cultivated beets (*Beta vulgaris* L. ssp. *vulgaris*) into Danish populations of sea beet (*Beta vulgaris* L. ssp. *maritima* (L.) Arcangeli). *Molecular Ecology*, **14,** 1391-1405.

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, **160,** 1217-1229.

Argyris J, Truco MJ, Ochoa O, Knapp SJ, Still DW, Lenssen GM, Schut JW, Michelmore RW, Bradford KJ (2005) Quantitative trait loci associated with seed and seedling traits in Lactuca. *Theoretical and Applied Genetics* **111**, 1365-1376.

Arias DM, Rieseberg LH (1994) Gene flow between cultivated and wild sunflowers. *Theoretical and Applied Genetics*, **89,** 655-660.

Arrigo N, Guadagnuolo R, Lappe S, Pasche S, Parisod C, Felber F (2011) Gene flow between wheat and wild relatives: empirical evidence from *Aegilops geniculata*, *Ae. neglecta* and *Ae. triuncialis*. *Evolutionary Applications* **4**(5), 685-695.

Auer C (2008) Ecological risk assessment and regulation for genetically-modified ornamental plants. *Critical Reviews in Plant Sciences* **27**, 255-271.

Baack JE, Rieseberg HL (2007) A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics & Development* **17**, 513-518.

Baack EJ, Sapir Y, Chapman MA, Burke JM, Rieseberg LH (2008) Selection on domestication traits and quantitative trait loci in crop–wild sunflower hybrids. *Molecular Ecology* **17**, 666-677.

Bai W, Zhang H, Zhang Z, Teng F, Wang L, Tao Y, Zheng Y (2010) The evidence for non-additive effect as the main genetic component of plant height and ear height in maize using introgression line populations. *Plant Breeding* **129**, 376-384.

Barton NH (2000) Genetic hitchhiking. *Phil.Trans. R. Soc. Lond. B* , 355, 1553 - 1562.

Bhatnagar-Mathur P, Vadez V, Sharma K (2008) Transgenic approaches for abiotic stress tolerance in plants: retrospect and prospects. *Plant Cell Reports* **27**, 411-424.

Bing DJ, Downey RK, Racow GFW (1996) Hybridizations among *Brassica napus*, *B. rapa* and *B. juncea* and their two weedy relatives *B. nigra* and *Sinapis arvensis* under open pollination conditions in the field. *Plant Breeding* **115**, 470-473.

Birchler JA, Auger DL, Riddle NC (2003) In Search of the Molecular Basis of Heterosis. *The Plant Cell Online* **15**, 2236-2239.

Bland JM, Altman DG (1995) Multiple significance tests: the Bonferroni method. *BMJ* **310**, 170.

Bowra JC (1992) Prickly lettuce (*Lactuca serriola*): a population explosion in Warwickshire. *BSBI News* **12**, 14-16.

Boyer JS (1982) Plant Productivity and Environment. *Science* **218**, 443-448.

Burgarella C, Lorenzo Z, Jabbour-Zahab R, Lumaret R, Guichoux E, Petit RJ, Soto A, Gil L (2009) Detection of hybrids in nature: application to oaks (*Quercus suber* and *Q. ilex*). *Heredity*, **102,** 442-452.

Burke JM, Arnold ML (2001) Genetics and the fitness of hybrids. *Annual Review of Genetics* **35**, 31-52.

Campbell LG, Snow AA (2007) Competition alters life history and increases the relative fecundity of crop-wild radish hybrids (Raphanus spp.). *New Phytologist* **173**, 648-660.

Campbell LG, Snow AA, Ridley CE (2006) Weed evolution after crop gene introgression: Greater survival and fecundity of hybrids in a new environment. *Ecology Letters* **9**, 1198-1209.

Cao QJ, Xia H, Yang X, Lu BR (2009) Performance of hybrids between weedy rice and insect-resistant transgenic rice under field experiments: Implication for environmental biosafety assessment. *Journal of Integrative Plant Biology* **51**, 1138-1148.

Carlborg Ö, Haley CS (2004) Epistasis: too often neglected in complex trait studies? *Nature Review-Genetics* **5**, 618-625.

Carson HL, Nair PS, Sene FM (1975) Drosophila hybrids in nature: proof of gene exchange between sympatric species. *Science* **189**, 806-807.

Carvalho Filho JLSd, Gomes LAA, Biguzzi FA, Maluf WR, Ferreira S (2009) F4 families of crispleaf lettuce with tolerance to early bolting and homozygous for resistance to *Meloidogyne incognita* race 1. *Horticultura Brasileira* **27**, 335-339.

Cermak T, Doyle EL, Christian M, wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF(2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research* **39** (12), e82.

Chahal GS, Gosal SS (2002) Principles and Procedures of Plant Breeding: Biotechnological and conventional approaches. Alpha Science International Ltd., Harrow, UK.

Chapman MA, Burke JM (2006) Letting the gene out of the bottle: the population genetics of genetically modified crops. *New Phytologist* **170**, 429-443.

Choi J, Seo Y, Kim S, Kim W, Shin J (2011) Constitutive expression of CaXTH3, a hot pepper xyloglucan endotransglucosylase/hydrolase, enhanced tolerance to salt and drought stresses without phenotypic defects in tomato plants (*Solanum lycopersicum* cv. Dotaerang) *Plant Cell Reports* **30**, 867-877.

Christou P, Twyman RM (2004) The potential of genetically enhanced plants to address food insecurity. *Nutrition Research Reviews* **17**, 23-42.

Conner AJ, Glare TR, Nap J-P (2003) The release of genetically modified crops into the environment: Part II. Overview of ecological risk assessment. *The Plant Journal* **33**, 19–46.

Cuartero J, Bolarín MC, Asíns MJ, Moreno V (2006) Increasing salt tolerance in the tomato. *Journal of Experimental Botany* **57**, 1045-1058.

Cummings CL, Alexander HM, Snow AA, Rieseberg LH, Kim MJ, Culley TM (2002) Fecundity selection in a sunflower crop-wild study: Can ecological data predict crop allele changes? *Ecological Applications* **12**, 1661-1671.

D'Andrea L, Broennimann O, Kozlowski G, Guisan A, Keller-Senften J, Felber F (2009) Climate change, anthropogenic disturbance and the northward range expansion of *Lactuca serriola* (Asteraceae). *Journal of Biogeography* **36**, 1573-1587.

D'Andrea L, Felber F, Guadagnuolo R (2008) Hybridization rates between lettuce (*Lactuca sativa*) and its wild relative (*L. serriola*) under field conditions. *Environmental Biosafety Research* **7**, 61-71.

De Vries FT, Van der Meijden R, Brandenburg WA (1992) Botanical files: A study of the real chances for spontaneous gene flow from cultivated plants to the wild flora in the Netherlands. *Gorteria (Suppl.)* **1**.

De Vries IM (1990) Crossing experiments of lettuce cultivars and species (Lactuca sect. Lactuca, Compositae). *Plant Systtematics and Evolution* **171**, 233-248.

De Vries IM (1997) Origin and domestication of Lactuca sativa L. *Genetic Resources and Crop Evolution* **44**, 165-174.

De Vries IM, Van Raamsdonk LWD (1994) Numerical morphological analysis of Lettuce cultivars and species (*Lactuca* sect. *Lactuca*, *Asteraceae*). *Plant Systematics and Evolution*, **193**(1), 125-141.

De Wet JMJ, Harlan JR (1975) Weeds and domesticates: Evolution in the man-made habitat. *Economic Botany*, **29,** 99-107.

Dechaine JM, Burger JC, Chapman MA, Seiler GJ, Brunick R, Knapp SJ, Burke JM (2009) Fitness effects and genetic architecture of plant-herbivore interactions in sunflower crop-wild hybrids. *New Phytologist* **184,** 828-841.

Donohue K, Rubio de Casas R, Burghardt L, Kovach K, Willis GC (2010) Germination, Postgermination Adaptation, and Species Ecological Ranges. *Annu. Rev. Ecol. Evol. Syst.* **41**, 293-319.

Dowling TE, Secor CL (1997) The Role of Hybridization and Introgression in the Diversification of Animals. *Annual Review of Ecology and Systematics* **28**, 593-619.

Ellstrand NC (2003) Current knowledge of gene flow in plants: implications for transgene flow. *Philosophical Transactions of the Royal Society London, Series B*, **358,** 1163-1170.

Ellstrand NC, Prentice HC, Hancock JF (1999) Gene flow and Introgression from domesticated plants into their wild relatives. *Annu. Rev. Ecol. Syst.* **30**, 539-563.

Ellstrand NC, Schierenbeck KA (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences of the USA*, **97,** 7043-7050.

Ellstrand NC, Whitkus R, Rieseberg LH (1996) Distribution of spontaneous plant hybrids. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 5090-5093.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14,** 2611-2620.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578.

Farooq M, Wahid A, Kobayashi N, Fujita D, Basra SMA (2009) Plant drought stress: Effects, mechanisms and management. *Agronomy for Sustainable Development* **29**, 185-212.

Foolad MR (1996) Genetic analysis of salt tolerance during vegetative growth in tomato, Lycopersicon esculentum Mill. *Plant Breeding* **115**, 245-250.

Frietema de Vries F. T. (1992) The systematic relationship of *Lactuca sativa* and *Lactuca serriola*, in relation to the distribution of Prickly lettuce. *Acta Botanica Neerlandica* **43**, 79.

Frietema de Vries FT, van der Meijden R, Brandenburg WA (1994) Botanical files on lettuce (*Lactuca sativa*): on the chance for gene flow between wild and cultivated lettuce (*Lactuca sativa* L. including *L. serriola* L., Compositae) and the generalized implications for risk-assessments on genetically modified plants. *Gorteria Supplement*, **2**, 44 pp.

Fukuda M, Matsuo S, Kikuchi K, Kawazu Y, Fujiyama R, Honda I (2011) Isolation and functional characterization of the FLOWERING LOCUS T homolog, the LsFT gene, in lettuce. *Journal of Plant Physiology* **168**, 1602-1607.

Fulton TM, Grandillo S, Beck-Bunn T, Fridman E, Frampton A, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley SD (2000) Advanced backcross QTL analysis of a *Lycopersicon esculentum* ×*Lycopersicon parviflorum* cross. *TAG Theoretical and Applied Genetics* **100,** 1025-1042.

Gao H, Williamson S, Bustamante CD (2007) A Markov Chain Monte Carlo Approach for Joint Inference of Population Structure and Inbreeding Rates From Multilocus Genotype Data *Genetics* **176**, 1635-1651.

Gardner KM, Latta RG (2008) Heritable variation and genetic correlation of quantitative traits within and between ecotypes of Avena barbata. *Journal of Evolutionary Biology* **21**, 737-748.

Giannino D, Nicolodi C, Testone G, Di Giacomo E, Iannelli MA, Frugis G, Mariotti D (2008) Pollen-mediated transgene flow in lettuce (*Lactuca sativa* L.). *Plant Breeding* **127**, 308-314.

Greene SL, Kisha TJ, Dzyubenko NI (2008) Conserving alfalfa wild relatives: Is past introgression with Russian varieties evident today? *Crop Science*, **48,** 1853-1864.

Groot MHM, van de Wiel CCM, van Tienderen PH, den Nijs HCM (2003) Hybridisation and introgression between crops and wild relatives: Current knowledge and research priorities in lieu of impending introductions of GM crops. In: *COGEM research*.

Gueritaine G, Sester M, Eber F, Chevre AM, Darmency H (2002) Fitness of backcross six of hybrids between transgenic oilseed rape (*Brassica nap*us) and wild radish (*Raphanus raphanistrum*). *Molecular Ecology* **11**, 1419-1426.

Hails RS, Morley K (2005) Genes invading new populations: a risk assessment perspective. *TRENDS in Ecology and Evolution* **20**, 245-252.

Hall L, Topinka K, Huffman J, Davis L, Good A (2000) Pollen flow between herbicide-resistant Brassica napus is the cause of multiple-resistant B. napus volunteers. *Weed Science* **48**, 688-694.

Hauser TP, Jørgensen RB, Østergård H (1998) Fitness of backcross and F2 hybrids between weedy *Brassica rapa* and oilseed rape (*B. napus*). *Heredity* **81**, 436-443.

Ho, McCouch, Smith (2002) Improvement of hybrid yield by advanced backcross QTL analysis in elite maize. *TAG Theoretical and Applied Genetics* **105**, 440-448.

Hoc PS, Espert SM, Drewes SI, Burghardt AD (2006) Hybridization between wild and domesticated types of *Phaseolus vulgaris* L. (Fabaceae) in Argentina. *Genetic Resources and Crop Evolution* **53**, 331-337.

Holland JB (2007) Genetic architecture of complex traits in plants. *Current Opinion in Plant Biology* **10**, 156-161.

Hooftman DAP, Flavell AJ, Jansen H, Den Nijs HCM, Syed NH, Sørensen AP, Wengel PO, Van de Wiel CCM (2011) Locus-dependent selection in crop-wild hybrids of lettuce under field conditions and its implication for GM crop development. *Evolutionary Applications*, **4**(5), 648-659.

Hooftman DAP, Hartman Y, Oostermeijer JGB, Den Nijs HCM (2009) Existence of vigorous lineages of crop-wild hybrids in Lettuce under field conditions. *Environmental Biosafety Research* **4**, 203-217.

Hooftman DAP, Jong MJD, Oostermeijer JGB, Den Nijs HCM (2007a) 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 DAP, Nieuwenhuis BPS, Posthuma KI, Oostermeijer JGB, den Nijs HCM (2007b) Introgression potential of downy mildew resistance from lettuce to Lactuca serriola and its relevance for plant fitness. *Basic and Applied Ecology* **8**, 135-146.

Hooftman DAP, Oostermeijer JGB, den Nijs JCM (2006) Invasive behaviour of Lactuca serriola (Asteraceae) in the Netherlands: Spatial distribution and ecological amplitude. *Basic and Applied Ecology* **7**, 507-519.

Hooftman DAP, Oostermeijer JGB, Jacobs MMJ, Den Nijs HCM (2005) Demographic vital rates determine the performance advantage of crop–wild hybrids in lettuce. *Journal of Applied Ecology* **42**, 1086-1095.

Hooftman DAP, Oostermeijer JGB, Marquard E, Den Nijs HCM (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.

Huangfu CH, Qiang Sheng S, Song XL (2011) Performance of hybrids between transgenic oilseed rape (*Brassica napus*) and wild *Brassica juncea*: An evaluation of potential for transgene escape. *Crop Protection* **30**, 57-62.

James C (2003) Global review of commercialized transgenic crops. *Current Science* **84**, 303-309.

Jansen J, De Jong AG, Van Ooijen JW (2001) Constructing dense genetic linkage maps. *Theoretical and Applied Genetics* **102**, 1113-1122.

Jansen J, van Hintum TJL (2007) Genetic distance sampling: a novel sampling method for obtaining core collections using genetic distances with an application to cultivated lettuce. *Theoretical and Applied Genetics* **114**, 421-428.

Jenni S, Hayes RJ (2009) Genetic variation, genotype × environment interaction, and selection for tipburn resistance in lettuce in multi-environments. *Euphytica* **171**, 427-439.

Jeuken M, Lindhout P (2002) *Lactuca saligna*, a non-host for lettuce downy mildew (Bremia lactucae), harbors a new race-specific *Dm* gene and three QTLs for resistance. *Theoretical and Applied Genetics* **105**, 384-391.

Jeuken MJW, Pelgrom K, Stam P, Lindhout P (2008) Efficient QTL detection for nonhost resistance in wild lettuce: backcross inbred lines versus F-2 population. *Theoretical and Applied Genetics* **116**, 845-857.

Johnson WC, Jackson LE, Ochoa O, van Wijk R, Peleman J, St.Claire DA, Michelmore RW (2000) Lettuce, a shallow-rooted crop, and *Lactuca serriola*, its wild progenitor, differ at QTL determining root architecture and deep soil water exploitation. *Theoretical and Applied Genetics* **101**, 1066-1073.

Karaba A, Dixit S, Greco R, Aharoni A, Trijatmiko KR, Marsch-Martinez N, Krishnan A, Nataraja KN, Udayakumar M, Pereira A (2007) Improvement of water use efficiency in rice by expression of HARDY, an Arabidopsis drought and salt tolerance gene. *Proceedings of the National Academy of Sciences* **104**, 15270-15275.

Kaufman L, Rousseeuw PJ (1990) Finding groups in data. An introduction to cluster analysis. *Wiley, New York*.

Kaur P, Mitkowski NA (2011) Evaluation of lactuca germplasm for resistance to the northern root-knot nematode. *International Journal of Vegetable Science* **17**, 26-36.

Koopman WJM, Guetta E, Van De Wiel CCM, Vosman B, Van Den Berg RG (1998) Phylogenetic relationships among *Lactuca* (Asteraceae) species and related genera based on ITS-1 DNA sequences. *American Journal of Botany* **85**, 1517-1530.

Koopman WJM, Jong JH, de Vries IM (1993) Chromosome banding patterns in Lettuce species (*Lactuca* sect.*Lactuca*, Compositae). *Plant Systematics and Evolution* **185**, 249-257.

Koopman WJM, Zevenbergen MJ, Van Den Berg RG (2001) Species relationships in Lactuca s.l. (Lactuceae, Asteraceae) inferred from AFLP fingerprints. *American Journal of Botany* **88**, 1881-1887.

Kiær LP, Felber F, Flavell A Guadagnuolo R, hauser TP, Olivieri AM, Scotti I, Syed N, Vaschi M, Van de Wiel CCM, Jørgensen RB (2009) Spontaneous gene flow and population structure in wild and cultivated chicory, *Cichorium intybus* L. *Genetic Resources and Crop Evolution*, **56,** 405-419.

Kuroda Y, Kaga A, Tomooka N, Vaughan DA (2008) Gene flow and genetic structure of wild soybean (*Glycine soja*) in Japan. *Crop Science* **48**, 1071-1079.

Kuroda Y, Kaga A, Tomooka N, Vaughan D (2010) The origin and fate of morphological intermediates between wild and cultivated soybeans in their natural habitats in Japan. *Molecular Ecology*, **19**, 2346-2360.

Kwit C, Moon HS, Warwick SI, Stewart Jr CN (2011) Transgene introgression in crop relatives: molecular evidence and mitigation strategies. *Trends in Biotechnology* **29**, 284-293.

Lada H, Mac Nally R, Taylor AC (2008) Distinguishing past from present gene flow along and across a river: the case of the carnivorous marsupial (*Antechinus flavipes*) on southern Australian floodplains. *Conservation Genetics*, **9,** 569-580.

Langevin SA, Clay K, Grace JB (1990) The incidence and effects of hybridization between cultivated rice and its related weed red rice (*Oryza sativa* L.). *Evolution*, **44,** 1000-1008.

Larson JL (1968) The Species Concept of Linnaeus. *Isis* **59**, 291-299.

Latta RG, Gardner KM, Johansen-Morris AD (2007) Hybridization, recombination, and the genetic basis of fitness variation across environments in Avena barbata. *Genetica* **129**, 167-177.

Latta RG, McCain C (2009) Path analysis of natural selection via survival and fecundity across contrasting environments in Avena barbata. *Journal of Evolutionary Biology* **22**, 2458-2469.

Lebeda A, Doležalová I, Astley D (2004a) Representation of wild *Lactuca* spp. (Asteraceae, Lactuceae) in world genebank collections. *Genetic Resources & Crop Evolution* **51**, 167-174.

Lebeda A, Doležalová I, Feráková V, Astley D (2004b) Geographical distribution of wild *Lactuca* species (Asteraceae, Lactuceae). *Botanical Review* **70**, 328-356.

Lebeda A, Doležalová I, Křístková E, Mieslerová B (2001) Biodiversity and ecogeography of wild *Lactuca* spp. in some European countries. *Genetic Resources and Crop Evolution*, **48,** 153-164.

Lee D and Natesan E (2006) Evaluating genetic containment strategies for transgenic plants, *Trends in Biotechnology* **24**(3), 109-114.

Lexer C, Welch ME, Durphy JL, Rieseberg LH (2003a) Natural selection for salt tolerance quantitative trait loci (QTLs) in wild sunflower hybrids: Implications for the origin of Helianthus paradoxus, a diploid hybrid species. *Molecular Ecology* **12**, 1225-1235.

Lexer C, Welch ME, Raymond O, Rieseberg LH (2003b) The Origin of Ecological Divergence in Helianthus paradoxus (Asteraceae): Selection on Transgressive Characters in a Novel Hybrid Habitat. *Evolution* **57**, 1989-2000.

Li B, Li N, Duan X, Wei A, Yang A, Zhang J (2010) Generation of marker-free transgenic maize with improved salt tolerance using the FLP/FRT recombination system. *Journal of Biotechnology*, **145**(2), 206-213.

Li J, Ji L (2005) Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* **95**, 221-227.

Lindqvist K (1960) Cytogenetic studies in the Serriola group of Lactuca, *Hereditas* **46,** 75-151.

Magnussen LS, Hauser TP (2007) Hybrids between cultivated and wild carrots in natural populations in Denmark. *Heredity* **99,** 185-192.

Malmberg RL, Held S, Waits A, Mauricio R (2005) Epistasis for Fitness-Related Quantitative Traits in Arabidopsis thaliana Grown in the Field and in the Greenhouse. *Genetics* **171**, 2013-2027.

Malosetti M, Voltas J, Romagosa I, Ullrich SE, van Eeuwijk FA (2004) Mixed models including environmental covariables for studying QTL by environment interaction. *Euphytica* **137**, 139-145.

Marks MK, Prince SD (1982) Seed Physiology and Seasonal Emergence of Wild Lettuce Lactuca Serriola. *Oikos* **38**, 242-249.

Mason P, Braun L, Warwick SI, B. Z, Stewart CNJ (2003) Transgenic Bt-producing Brassica napus: Plutella xylostella selection pressure and fitness of weedy relatives. *Enviromental Biosafety research* **2**, 263-276.

Mathews KL, Malosetti M, Chapman S, McIntyre L, Reynolds M, Shorter R, Van Eeuwijk F (2008) Multi-environment QTL mixed models for drought stress adaptation in wheat. *Theoretical and Applied Genetics* **117**, 1077-1091.

McHale LK, Truco M-J, Kozik A, Wroblewski T, Ochoa OE, Lahre KA, Knapp SJ, Michelmore RW (2009) The genomic architecture of disease resistance in lettuce. *Theoretical and Applied Genetics* **118**, 565-580.

Mei HW, Luo LJ, Ying CS, Wang YP, Yu XQ, Guo LB, Paterson AH, Li ZK (2003) Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred rice population and two testcross populations. *TAG Theoretical and Applied Genetics* **107**, 89-101.

Mercer KL, Shaw RG, Wyse DL (2006a) Increased germination of diverse crop-wild hybrid sunflower seeds. *Ecological Applications* **16**, 845-854.

Mercer KL, Wyse DL, Shaw RG (2006b) Effects of competition on the fitness of wild and crop-wild hybrid sunflower from a diversity of wild populations and crop lines. *Evolution* **60**, 2044-2055.

Mercer KL, Andow DA, Wyse DL, Shaw RG (2007) Stress and domestication traits increase the relative fitness of crop–wild hybrids in sunflower. *Ecology Letters* **10**, 383-393.

Michelmore R, Ochoa O, Wong J (2008) Bremia lactucae and Lettuce Downy Mildew. In: *Oomycete Genetics and Genomics*, pp. 241-262. John Wiley & Sons, Inc.

Milne RI, Abbott RJ (2000) Origin and evolution of invasive naturalized material of *Rhododendron ponticum* L. in the British Isles. *Molecular Ecology*, **9,** 541-556.

Morrell PL, Williams-Coplin TD, Lattu AL, Bowers JE, Chandler JM, Paterson AH (2005) Crop-to-weed introgression has impacted allelic composition of johnsongrass populations with and without recent exposure to cultivated sorghum. *Molecular Ecology*, **14,** 2143-2154.

Munns R, Tester M (2008) Mechanisms of Salinity Tolerance. *Annual Review of Plant Biology* **59**, 651-681.

Nagata RT (1992) Clip-and-wash method of emasculation for lettuce. *Hortiscience* 27, 907-908.

Odong TL, van Heerwaarden J, Jansen J, van Hintum TJL, van Eeuwijk FA (2011) Statistical techniques for constructing reference sets of accessions and microsatellite markers *Crop Science* **51**, On line.

Payne RW, Murray DA, Harding SA, Baird DB, Soutar DM (2010) An Introduction to GenStat for Windows (13th Edition). VSN International, Hemel Hempstead, UK.

Payne RW, Murray DA, Harding SA, Baird DB, Soutar DM (2011) An Introduction to GenStat for Windows (14th Edition). VSN International, Hemel Hempstead, UK.
Pilson D, Prendeville HR (2004) Ecological effects of transgenic crops and the escape of transgenes into wild populations. *Annual Review of Ecology, Evolution, and Systematics*, **35,** 149-174.
Pirondini A, Marmiroli N (2008) Environmental risk assessment in GMO analysis. *Rivista di Biologia - Biology Forum* **101**, 215-246.
Platt A, Vilhjálmsson BJ, Nordborg M (2010) Conditions Under Which Genome-Wide Association Studies Will be Positively Misleading. *Genetics* **186**, 1045-1052.
Pritchard KJ, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155,** 945-959.
Rathmacher G, Niggemann M, Köhnen M et al. (2010) Short-distance gene flow in *Populus nigra* L. accounts for small-scale spatial genetic structures: implications for in situ conservation measures. *Conservation Genetics*, **11**, 1327-1338.
Rieseberg LH, Archer MA, Wayne RK (1999) Transgressive segregation, adaptation and speciation. *Heredity* **83**, 363-372.
Rieseberg LH, Baird SJE, Gardner KA (2000) Hybridization, introgression, and linkage evolution. *Plant Molecular Biology*, **42,** 205-224.
Rieseberg LH, Kim SC, Randell RA, Whitney KD, Gross BL, Lexer C, Clay K (2007) Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica*, **129,** 149-165.
Robert VJM, West MAL, Inai S, Caines A, Arntzen JK, Smith JK, StClair DA (2001) Marker-assisted introgression of blackmold resistance QTL alleles from wild Lycopersicon cheesmanii to cultivated tomato L. esculentum) and evaluation of QTL phenotypic effects. *Molecular Breeding* **8**, 217-233.
Rong J, Janson S, Umehara M, Ono M, Vrieling K (2010) Historical and contemporary gene dispersal in wild carrot (*Daucus carota* ssp. *carota*) populations. *Annals of Botany*, **106**, 285-296.
Rose CW, Millwood RJ, Moon HS, Rao M, Halfhill M, Raynmer P, Warwick S, Al-Ahmad H, Gressel J, Stewart CN (2009) Genetic load and transgenic mitigating genes in transgenic *Brassica rapa* (field mustard) × *Brassica napus* (oilseed rape) hybrid populations. *BMC Biotechnology* **9**(1), 93.
Roy SJ, Tucker EJ, Tester M (2011) Genetic analysis of abiotic stress tolerance in crops. *Current Opinion in Plant Biology* **14**, 232-239.
Ryder EJ (1999) Lettuce, Endive and Chicory. *CAB International, Wallingford, UK.*
Ryder EJ, Milligan DC (2005) Additional genes controlling flowering time in Lactuca sativa and L. serriola. *Journal of the American Society for Horticultural Science* **130**, 448-453.
Ryder EJ, Whitaker T (1976) Lettuce, *Lactuca sativa* (Compositae). In: *Evolution of Crop Plants* (ed. Simmonds NW), pp. 39-41, Longman, London.
Saisho D, Purugganan MD (2007) Molecular phylogeography of domesticated barley traces expansion of agriculture in the Old World. *Genetics*, **177,** 1765-1776.
Schaminée JHJ, van Duuren L, de Bakker AJ (1992) Europese en mondiale verspreiding van Nederlandse vaatplanten. *Gorteria* **18,** (57-96).
Schanzer IA, Kutlunina NA (2010) Interspecific hybridization in wild roses (*Rosa* L. sect. Caninae DC.). *Biology Bulletin*, **37**, 480-488.
Scurrah M, Celis-Gamboa C, Chumbiauca S, Salas A, Visser RGF (2008) Hybridization between wild and cultivated potato species in the Peruvian Andes and biosafety implications for deployment of GM potatoes. *Euphytica* **164**, 881-892.
Shukla VK, Doyon Y, Miller JC, DeKelever RC, Moehle EA, Worden SE, Mitchell JC, Arnold NL, Gopalan S, Meng X, Choi V, Rock JM, Wu Y-Y, Katibah GE, Zhifang G, McCaskill D, Simpson MA, Blakeslee B, Greenwalt SA, Butler HJ, Hinkley SJ, Zhang L, Rebar EJ, Gregory PD, Urnov FD (2009) Precise genome modification in the crop species Zea mays using zinc-finger nucleases. *Nature* **459**, 437-441.
Simko I, Hu J (2008) Population structure in cultivated lettuce and its impact on association mapping. *Journal of the American Society for Horticultural Science*, **133,** 61-68.
Siol M, Prosperi J, Bonnin I, Ronfort J (2008) How multilocus genotypic pattern helps to understand the history of selfing populations: a case study in *Medicago truncatula*. *Heredity*, **100**, 517-

525.
Smulders MJM, Beringen R, Volosyanchuk R, Van den Broeck A, Van der Schoot J, Arens P, Vosman B (2008) Natural hybridization between Populus nigra L. and P. x canadensis Moench. Hybrid offspring competes for niches along the Rhine river in the Netherlands. *tree genetics and genomes*.
Snow AA, Andow DA, Gepts P, Hallerman EM, Power A, Tiedje JM, Wolfenbarger LL (2005) Genetically engineered organisms and the environment: Current status and recommendations. *Ecological Applications* **15**, 377-404.
Snow AA, Culley TM, Campbell LG, Sweeney PM, Hedge SG, Ellstarnd NC (2010) Long-term persistence of crop alleles in weedy populations of wild radish (Raphanus raphanistrum). *New Phytologist* **186**, 537-548.
Snow AA, Pilson D, Rieseberg LH, Paulsen MJ, Pleskac N, Reagon MR, Wolf DE, Selbo SM (2003) A Bt Transgene reduces herbivory and enhances fecundity in wild sunflower *Ecological Applications* **13**, 279-286.
Snow AA, Uthus KL, Culley TM (2001) Fitness of hybrids between weedy and cultivated radish: Implications for weed evolution. *Ecological Applications* **11**, 934-943.
Stewart Jr CN, Halfhill MD, Warwick SI (2003) Transgene introgression from genetically modified crops to their wild relatives. *Nature Reviews Genetics* **4**, 806-817.
Stolte S, Steudte S, Markowska A, Arning J, Neumann J, Stepnowski P (2011) Ion chromatographic determination of structurally varied ionic liquid cations and anions-a reliable analytical methodology applicable to technical and natural matrices. *Analytical Methods* **3**, 919-926.
Storme V, van den Broeck A, Ivens B, Halfmaerten D, Van Slycken J, Castiglione S, Grassi F, Fossati T, Cottrell JE, Tabbener HE, Lefèvre F, Saintagne C, Fluch S, Krystufek V, Burg K, Bordacs S, Borovics A, Gebhardt K, Vornam B, Pohl A, Alba N, Agfflndez D, Maestro C, Notivol E, Bovenschen J, Van Dam BC, Van der Schoot J, Vosman B, Boerjan W, Smulders MJM (2004) Ex-situ conservation of Black poplar in Europe: genetic diversity in nine gene bank collections and their value for nature development. *Theoretical and Applied Genetics*, **108**, 969-981.
Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor Appl Genet* **92**, 191-203.
Tester M, Langridge P (2010) Breeding Technologies to Increase Crop Production in a Changing World. *SCIENCE* **327**, 818-822.
Thompson, R. C., Whitaker TW, Bohn GW, Van Horn CW (1958) Natural cross-pollination in lettuce. *American Society for Horticultural Science*, **72,** 403-409.
Tiedje JM, Colwell RK, Grossman YL, Hodson RW, Lenski RE, Mack RN, Regal PJ (1989) The planned introduction of genetically engineered organisms: ecological considerations and recommendations. *Ecology* **70**, 298-315.
Tisné S, Schmalenbach I, Reymond M, Dauzat M, Pervent M, Vile D, Granier C (2010) Keep on growing under drought: Genetic and developmental bases of the response of rosette area using a recombinant inbred line population. *Plant, Cell and Environment* **33**(11), 1875-1887.
Townsend JA, Wright DA, Winfrey RJ, Fu F, Maeder ML, Joung JK, Voytas DF (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* **459**, 442-445.
Truco MJ, Antonise R, Lavelle D Ochoa O, Kozik A, Witsenboer H, Fort SB, Jeuken MJM, Kesseli RV, Lindhout P, Michelmore RW, Peleman J (2007) A high-density, integrated genetic linkage map of lettuce (*Lactuca* spp.). *Theoretical and Applied Genetics*, **115,** 735-746.
Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. *Nature* **11**, 636-646.
Ureta MS, Cantamutto M, Carrera A Delucchi C, Poverene M (2008) Natural hybrids between cultivated and wild sunflowers (*Helianthus* spp.) in Argentina. *Genetic Resources and Crop Evolution*, **55**, 1267–1277.

Vähä J-P, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, **15,** 63-72.

Valdes VM, Bradford KJ, Mayberry KS (1985) Alleviation of thermodormancy in coated lettuce seeds by seed priming. *HortScience* **20**, 1112-1114.

Van de Wiel C, Arens P, Vosman B (1999) Microsatellite retrieval in lettuce (*Lactuca sativa* L.). *Genome*, **42,** 139-149.

Van de Wiel C, Groot M, Den Nijs H (2005) Gene flow from crops to wild plants and its population-ecological consequences in the context of GM-crop biosafety, including some recent experiences from lettuce. In: Wesseler J (Ed.) *Environmental Costs and Benefits of Transgenic Crops*, Springer, Dordrecht, pp. 97-110.

Van de Wiel C, Van der Linden G, Den Nijs H, Flavell A, Syed N, Jørgensen R, Felber F, Scotti I, Van der Voort JR, Peleman J (2003) An EU project on gene flow analysis between crop and wild forms of lettuce and chicory in the context of GMO biosafety: first results in lettuce. In: Van Hintum TJL, Lebeda A, Pink DA, Schut JW (Ed.) *EUCARPIA leafy vegetables 2003. Proceedings of the EUCARPIA Meeting on Leafy Vegetables Genetics and Breeding, Noordwijkerhout, Netherlands, 19-21 March 2003*, Centre for Genetic Resources, the Netherlands (CGN), Wageningen, pp. 111-116.

Van de Wiel CCM, Sretenović Rajičić T, Van Treuren R, Dehmer KJ, Van der Linden CG, Van Hintum TJL (2010) Distribution of genetic diversity in wild European populations of prickly lettuce (*Lactuca serriola*): implications for plant genetic resources management. *Plant Genetic Resources-Characterization and Utilization*, **8**, 171-181.

Van der Arendt AJM, Ester A, van Schijndel JT (1999) Developing an aphid resistant butterhead lettuce 'Dynamite'In: Lebeda, A. (Ed.) Kristkova, E. (Ed.) Proceedings of the Eucarpia meeting on leafy vegetables genetics and breeding, pp. 149-157. Olomouc, Czech Republic.

Van Hintum TJL (2003) Molecular characterisation of a lettuce germplasm collection. In: Van Hintum ThJL, Lebeda A, Pink D, Schut JW (eds) *Eucarpia leafy vegetables 2003, Proceedings of the Eucarpia meeting on leafy vegetables, Noordwijkerhout, The Netherlands, 19–21 March 2003*. Centre for Genetic Resources, The Netherlands (CGN), Wageningen, The Netherlands**,** pp. 99-104.

Van Ooijen JW (2006) JoinMapR 4, Software for the culculation of genetic linkage maps in experimental populations. *Kyazma B. V., Wageningen, Netherlands*.

Van Tienderen PH (2004) Hybridisation in nature: lessons for the introgression of transgenes into wild relatives. In: *In Introgression from Genetically Modified Plants into Wild Relatives* (eds. den Nijs HCM, Bartsch D, Sweet J). CABI Publishing.

Van Treuren R, Van Hintum TJL, Van de Wiel CCM (2008) Marker-assisted optimization of an expert-based strategy for the acquisition of modern lettuce varieties to improve a genebank collection. *Genetic Resources and Crop Evolution*, **55,** 319-330.

Vinocur B, Altman A (2005) Recent advances in engineering plant tolerance to abiotic stress: achievements and limitations. *Current Opinion in Biotechnology* **16**, 123-132.

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. *The Journal of Heredity* **93**, 77-78.

Wang W, Vinocur B, Altman A (2003) Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* **218**, 1-14.

Warwick SI, Beckie HJ, Hall LM (2009) Gene flow, invasiveness, and ecological impact of genetically modified crops. *Annals of the New York Academy of Sciences* **1168,** 72-99.

Warwick SI, Légère A, Simard M-J, James T (2008) Do escaped transgenes persist in nature? The case of an herbicide resistance transgene in a weedy *Brassica rapa* population. *Molecular Ecology*, **17,** 1387-1395.

Waycott W, Fort SB, Ryder EJ (1995) Inheritance of dwarfing genes in Lactuca sativa L. *Journal of Heredity* **86**, 39-44.

Westman AL, Levy BM, Miller MB, Gilles GJ, Spira TP, Rajapakse S, Tonkyn DW, Abbott AG (2001) The potential for gene flow from transgenic crops to related wild species: A case study in strawberry. *Acta Horticulturae*, **560,** 527-530.

White PJ, Brown PH (2010) Plant nutrition for sustainable development and global health. *Annals of Botany* **105**, 1073-1080.

Whitton J, Wolf DE, Arias, DM, Snow AA, Rieseberg LH (1997) The persistence of cultivar alleles in wild populations of sunflowers five generations after hybridization *Theoretical and Applied Genetics*, **95**, 33-40.

Witcombe JR, Hollington PA, Howarth CJ, Reader S, Steele KA (2008) Breeding for abiotic stresses for sustainable agriculture. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 703-716.

Yu S, Li J, Xu C, Tan Y, Gao Y, Li X, Zhang Q, Maroof MAS (1998) Epistasis plays an important role as genetic basis of heterosis in rice. *Science in China, Series C: Life Sciences* **41**(3), 293-302.

Yue B, Cai X, Yuan W, Vick B, Hu J (2009) Mapping quantitative trait loci (QTL) controlling seed morphology and disk diameter in sunflower [Helianthus annuus L. *Helia* **32**(50), 17-36.

Zhang J, Klueva NY, Wang Z, Wu R, Ho T-HD, Nguyen HT (2000) Genetic engineering for abiotic stress resistance in crop plants. *In Vitro Cell. Dev. Biol.-Plant* **36**, 108-114.

Zhang NW, Pelgrom K, Niks RE, Visser RGF, Jeuken MJW (2009) Three combined quantitative trait loci from nonhost lactuca saligna are sufficient to provide complete resistance of lettuce against bremia lactucae. *Molecular Plant-Microbe Interactions* **22**, 1160-1168.

Summary

Summary

Many plant species can hybridise and produce fertile offspring. This happens where species areas meet, when species invade the area of another species, and between crops and their wild relatives. The latter has raised concern with regard to GM crops, as it constitutes a possible route along which the transgene could disperse from crops into related wild species, establish itself in the natural population, and persist under natural conditions. This may cause unintended ecological consequences such as the formation of weeds that are difficult to manage in agricultural areas or the formation of more fit crop-wild hybrids that could displace the wild species, thus causing genetic erosion.

After crop-wild hybridization, the persistence of the hybrids and of the crop genes (including the transgenes) in later generations depend on their genetic make-up, which consists of specific combinations of wild and crop genomic blocks, and on the environmental conditions. Therefore, knowledge on the dynamics of crop-wild hybridization and introgression using conventional crop varieties is needed as it constitutes the baseline for putting into perspective the effects of transgene introgression under natural conditions.

The present study focused on understanding the genetic process of hybridization and introgression from cultivated to wild relative species using *Lactuca sativa* L. and *L. serriola* L. as a model crop-wild complex with an emphasis on the contribution of the crop genome to the performance of the hybrids. It aimed at answering the following questions: (i) whether there is evidence of spontaneous hybridization between *L. serriola* and *L. sativa* under natural conditions, (ii) whether crop genes confer any (dis)advantage to the crop-wild hybrids under controlled non-stress conditions, under controlled conditions with abiotic stress, and under field conditions, (iii) whether the (dis)advantageous effects of the crop are dependent on environmental conditions, and (iv) whether we can identify genomic regions where transgenes could be inserted with the purpose of mitigating their persistence after crop-wild hybridization.

A large dataset of *L. serriola* and *L. sativa* plants from the Centre for Genetic Resources, the Netherlands (CGN) and another set of recently collected (between 2002 and 2005) *L. serriola* samples from Europe (ANGEL) that were genotyped with 10 microsatellite markers, were analysed with the Bayesian-based programme STRUCTURE. In spite of the low outcrossing rate in lettuce, an occurrence of 7% of crop-wild hybrids was found among the natural populations of *L. serriola* in Europe (ANGEL dataset) and 9% among CGN *L. serriola* plants collected from the wild. From the occurrence and geographical localization of the hybrids we

concluded that hybridization between *L. sativa* and *L. serriola* is not a likely cause of the recent spread of *L. serriola* in Europe.

To test for the importance of the crop genomic segments to the performance of the crop-wild hybrids, F_1 hybrid progeny was created by crossing *L. serriola* collected from Eys (the Netherlands) with *L. sativa*, cv. Dynamite. As selfing and backcrossing constitute the two possible routes along which a crop gene could become established in a wild genetic background, three subsequent hybrid generations were created, namely F_2 (by selfing F_1) and two backcross populations, BC_1 and BC_2 by backcrossing to the *L. serriola* parent. The three populations were genotyped with Single Nucleotide Polymorphism (SNP) markers for genetic analysis. Because of the importance of the abiotic stresses as selection factors under natural conditions and the prospective of generating GM crop varieties with enhanced abiotic stress tolerance, the three populations were evaluated for vigour at the rosette stage under greenhouse conditions of non-stress, drought, salinity and nutrient deficiency. The BC_1 population was also evaluated under field conditions at two locations (Wageningen and Sijbekarspel, the Netherlands) for hybrid performance from germination to seed production.

Vigour under greenhouse conditions and germination, vigour and reproduction traits under field conditions were characterized by moderate to high broad sense heritability values, indicating that crop-wild hybridization generates genetic variation on which selection could act. Using a Quantitative Trait Loci (QTL) approach, we determined the contribution of the crop to the vigour and fitness of the hybrids. In the F_2 population, 17 QTLs for vigour were detected with most of the QTLs for biomass having a positive effect from the crop allele, showing that cultivated lettuce can contribute positively to the vigour of the hybrids at the rosette stage under non-stress, salinity, drought and nutrient deficiency conditions. The dominance effects of the QTLs were not significant for 16 out of the mapped 17 QTLs, nor were the heterozygous genotypes associated with the highest means for QTL epistatic effect. This suggests that heterozygosity did not play a major role to the vigour of the hybrids. Conversely, QTL epistatic effect was significant for many of the trait-treatment combinations. In the BC_1 and BC_2 populations, the crop contributed mostly to the biomass traits through additive and epistatic QTL effect, indicating a potential of introgression of crop segments in the increasing wild genetic background.

Under field conditions, cultivated lettuce contributed positively for the QTLs for plant vigour at the rosette stage, the number of branches and the total number of seeds among the hybrids, whereas QTLs for germination, survival and plant height had their positive allelic effect from

the wild allele. Genotype by environment (GxE) effect was significant on the vigour traits under greenhouse conditions of stress and non-stress and on germination, vigour and reproduction traits under field conditions. Survival was the only trait not affected by GxE and all the QTLs for the trait had the same direction and magnitude at the two sites. Therefore, one can predict the survival rate of lettuce crop-wild hybrids based on their genotypes. Despite the significance of QTLxE for the QTLs detected for the remainder of the traits, most of the QTLs were significant in more than one treatment with the same direction of allelic effect (either positive for the crop allele or for the wild allele) under greenhouse conditions or at the two sites under field conditions. Therefore, the significance of QTLxE for those QTLs would not prevent an estimation of the performance of a hybrid based on its genotype. Based on the location and allelic effect of the QTLs for germination, vigour, survival and reproductive traits in the current lettuce crop-wild cross, some genomic regions were suggested where transgenes could be inserted in order to mitigate their persistence in crop-wild hybrids through genetic hitchhiking and background selection on linkage groups 1, 6 and 9, and other regions in which it would be better not to insert a transgene on linkage groups 6, 7 and 8.

Samenvatting

Een aanzienlijk aantal plantensoorten kan hybridiseren en daarbij fertiele nakomelingen genereren. Dit gebeurt waar soortarealen elkaar ontmoeten, wanneer soorten binnendringen in het areaal van een andere soort, en tussen cultuurgewassen en hun wilde verwanten. Het laatste verschijnsel heeft de aandacht getrokken in verband met de opkomst van genetisch gemodificeerde (GG) gewassen, aangezien het een mogelijke route vormt waarlangs transgenen zouden kunnen verspreiden van gewassen naar verwante wilde soorten. Zo zouden ze zich kunnen vestigen in natuurlijke populaties en vervolgens zich handhaven onder natuurlijke omstandigheden. Dit zou weer onbedoelde ecologische gevolgen kunnen hebben zoals het ontstaan van onkruiden die moeilijk in bedwang te houden zijn in landbouwgebieden of het ontstaan van gewas-wild hybriden die de wilde soort verdringen, wat genetische erosie tot gevolg heeft.

Na gewas-wild hybridisatie hangt de persistentie van de hybriden en de gewasgenen (met inbegrip van de transgenen) in opvolgende generaties af van de genetische samenstelling, die bestaat uit combinaties van gewas- en wilde genoomstukken, en van de milieuomstandigheden. Zodoende is kennis nodig over de dynamiek van gewas-wild hybridisatie en introgressie aan de hand van conventionele cultuurvariëteiten, aangezien dit de “baseline” (het uitgangspunt) vormt waarmee de effecten van transgeenintrogressie onder natuurlijke omstandigheden kunnen worden vergeleken.

De onderhavige studie richtte zich op het begrijpen van het genetische proces in de hybridisatie en introgressie van gewas naar wilde verwant aan de hand van het gewas-wild complex in sla dat gevormd wordt door *Lactuca sativa* L. en *L. serriola* L., met de nadruk op de bijdrage van het gewasgenoom aan de prestaties van de hybriden. Het richtte zich op het beantwoorden van de volgende vragen: (i) of er aanwijzingen zijn voor spontane hybridisatie tussen *L. serriola* en *L. sativa* onder natuurlijke omstandigheden, (ii) of gewasgenen een voor- dan wel nadeel bieden aan de gewas-wild hybriden onder gecontroleerde omstandigheden van abiotische stress en geen stress, en onder veldomstandigheden, (iii) of voor- dan wel nadelige effecten van het gewas afhankelijk zijn van milieuomstandigheden, en of we genoomgebieden kunnen identificeren waar transgenen ingevoegd kunnen worden met het oog op het tegengaan van hun persistentie na gewas-wild hybridisatie.

Een grote dataset van *L. serriola* en *L. sativa* planten uit de slacollectie van het Centrum voor Genetische Bronnen Nederland (CGN) en een andere dataset van recent (tussen 2002 en 2005) verzamelde monsters uit Europa (EU onderzoeksproject ANGEL) die gegenotypeerd

zijn op basis van 10 microsatellietmerkers, zijn geanalyseerd met het Bayesiaanse softwareprogramma STRUCTURE. Ondanks het lage uitkruisingsniveau in sla vertoonde 7% van de planten uit natuurlijke populaties in Europa (ANGEL dataset) kenmerken van gewas-wild hybriden en hetzelfde werd gevonden voor 9% van de CGN *L. serriola* planten die oorspronkelijk in het wild verzameld waren. Op basis van het voorkomen en de geografische verspreiding van de hybriden concludeerden we dat hybridisatie tussen *L. sativa* en *L. serriola* geen aannemelijk verklaring vormt voor de recente uitbreiding van *L. serriola* in Europa.

Om het belang van de gewasgenoomsegmenten voor de prestaties van de gewas-wild hybriden vast te stellen werd een F_1 hybride nakomelingschap geproduceerd door het kruisen van *L. serriola* uit Eys (Limburg, Nederland) met *L. sativa*, cv. Dynamite. Aangezien zelfbevruchting en terugkruising de twee mogelijk routes uitmaken waarlangs een gewasgen zich zou kunnen vestigen in een wilde genetische achtergrond, werden er drie opvolgende hybride generaties geproduceerd, namelijk een F_2 (door zelfbevruchting van de F_1) en twee terugkruisingspopulaties, een BC_1 en een BC_2 door terugkruising met de *L. serriola*-ouderlijn. De drie populaties werden gegenotypeerd met SNP merkers (Single Nucleotide Polymorphism = variaties gebaseerd op enkelvoudige DNA-basenwijzigingen) voor de genetische analyse. Vanwege het belang van verschillende typen abiotische stress als selectiefactoren onder natuurlijke omstandigheden en het vooruitzicht van de introductie van GG gewasvariëteiten met verbeterde tolerantie voor abiotische stress werden de drie populaties geëvalueerd voor groeikracht ("vigour") in het rozetstadium onder kasomstandigheden zonder stress en met droogte, zout en nutriëntengebrek. De BC_1 populatie werd ook geëvalueerd onder veldomstandigheden op twee locaties (Wageningen en Sijbekarspel, Nederland) met betrekking tot hybridenprestaties vanaf zaadkieming tot zaadproductie.

Groeikracht ("vigour") onder kasomstandigheden en zaadkieming, groeikracht en reproductie-eigenschappen onder veldomstandigheden werden gekenmerkt door gemiddelde tot hoge waarden voor "broad-sense heritability" (erfelijkheidsfactor in brede zin), wat erop wijst dat gewas-wild hybridisatie genetische variatie genereert waarop selectie zou kunnen aangrijpen. Met gebruikmaking van een QTL-benadering (Quantitative Trait Loci = genoomposities gerelateerd aan kwantitatieve variatie in een eigenschap) bepaalden we de bijdrage van het gewas aan de groeikracht ("vigour") en de "fitness" (mate van aanpassing aan groeiomstandigheden) van de hybriden. In de F_2 populatie werden 17 QTLs voor groeikracht gedetecteerd, waarbij de meeste biomassa-QTLs een positief effect van het gewasallel vertoonden, hetgeen aangeeft dat cultuursla positief kan bijdragen aan de groeikracht van

hybriden in het rozetstadium onder omstandigheden zonder stress, en met zout, droogte en nutriëntengebrek. De dominantie-effecten waren niet statistisch significant bij 16 van de op de genetische kaart gezette 17 QTLs; evenmin was er een relatie tussen heterozygote genotypen en de hoogste gemiddelden voor het epistatische (interacties tussen genen) effect van QTLs. Dit suggereert dat de mate van heterozygotie geen overheersende rol speelde in de groeikracht van de hybriden. Daartegenover was het epistatisch effect van QTLs significant bij veel van de combinaties van eigenschap en groeiomstandigheden.

Onder veldomstandigheden droegen allelen van de cultuursla positief bij in de QTLs voor plantgroeikracht in het rozetstadium, het aantal vertakkingen en het aantal zaden in de hybriden, terwijl QTLs voor zaadkieming, overleving van de plant en planthoogte een positief effect vertoonden van het wilde allel. Het GxE effect ("genotype by environment" = interactie tussen genotype en milieuomstandigheden) was statistisch significant bij de groeikrachteigenschappen gemeten onder kasomstandigheden met en zonder stress en bij zaadkieming, groeikracht en reproductie-eigenschappen gemeten onder veldomstandigheden. Plantoverleving op het veld was de enige eigenschap die geen invloed van GxE effecten vertoonde, en alle QTLs voor deze eigenschap hadden dezelfde richting en grootte op de twee veldlocaties. Zodoende zou men de mate van overleving van gewas-wild hybriden in sla kunnen voorspellen op basis van hun genotypen. Niettegenstaande de significantie van QTLxE effect (interactie tussen QTL en milieuomstandigheden) voor de QTLs voor de rest van de eigenschappen waren de meeste QTLs significant in meer dan één behandeling met dezelfde richting van het allelische effect (positief voor ofwel het gewasallel ofwel voor het wilde allel) onder kasomstandigheden of op de twee veldlocaties. Daarom zou het significante QTLxE effect voor deze QTLs een schatting van de prestaties van een hybride op basis van het genotype niet in de weg staan. Uitgaande van de locatie en het allelische effect van de QTLs voor zaadkieming, groeikracht ("vigour"), plantoverleving en reproductieve eigenschappen in de onderhavige kruising tussen gewas en wild konden enkele genoomgebieden gesuggereerd worden waar een transgen ingevoegd zou kunnen worden teneinde de persistentie in het milieu door middel van genetisch "meeliften" en achtergrondselectie tegen te gaan, te weten op ""linkage groups" (groepen van verbonden genetische merkers, genetische kaartequivalenten van chromosomen) 1, 6 en 9. Ook konden genoomgebieden worden aangewezen waar met dit doel een transgen beter niet ingevoegd zou kunnen worden, namelijk op "linkage groups" 6, 7 en 8.

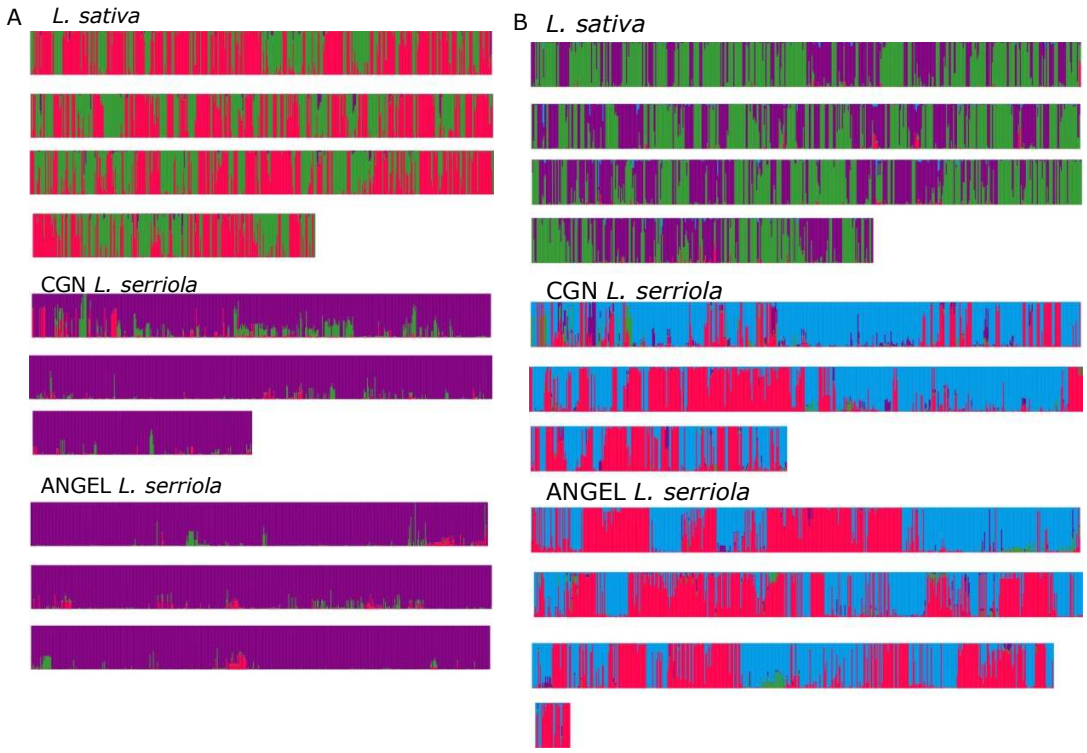
Supplementary Material

Chapter 2

Table S1 Frequency of "crop specific" alleles in lettuce datasets

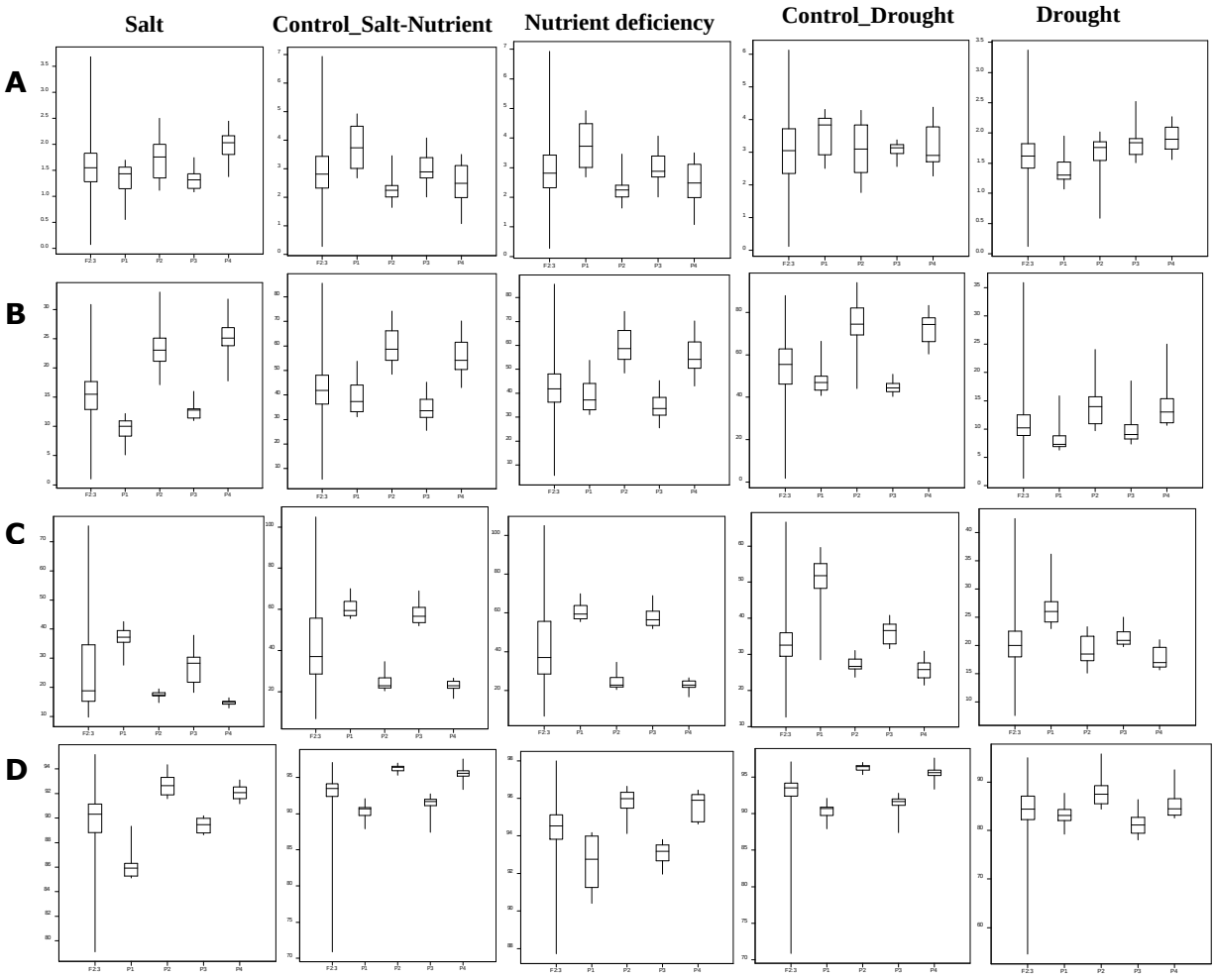
SSR allele	*L. sativa*	ANGEL *L. serriola*	CGN *L. serriola* Europe	CGN *L. serriola*
LsA001-187	0.1159	0.002	0	0.002
D103-263	0.5286	0.000	0.0008	0.001
D103-266	0.1813	0.004	0.0017	0.001
D106-191	0.4122	0.000	0	0.003
D109-251	0.1601	0.003	0	0.028
LsE003-206	0.8888	0.000	0	0.003
E011-251	0.1253	0.005	0	0.001
E011-254	0.7904	0.006	0	0.163

Figure S1 STRUCTURE Q graph of all datasets at A) K=3 and B) K=4: At higher K (>2) *L. serriola* remains distinct from *L. sativa*



Chapter 3

Figure S2 Boxplots showing phenotypic variation among $F_{2:3}$ plants (position 1), *L. serriola* acc. UC96US23 (position 2), *L. sativa* cv. Salinas (position 3), *L. serriola*/Eys (position 4) and *L. sativa* cv. Dynamite (position 5) for dry weight (A), fresh weight(B), plant height (C) and relative moisture content (D) under the five treatments



Chapter 4

Figure S3 Crossing and experimental scheme of the study on introgression process from cultivated to wild lettuce. The selfing pathway (F_2) was in chapter 3The back-crossing pathway (BC_1 and BC_2 populations) is the subject of chapter 4

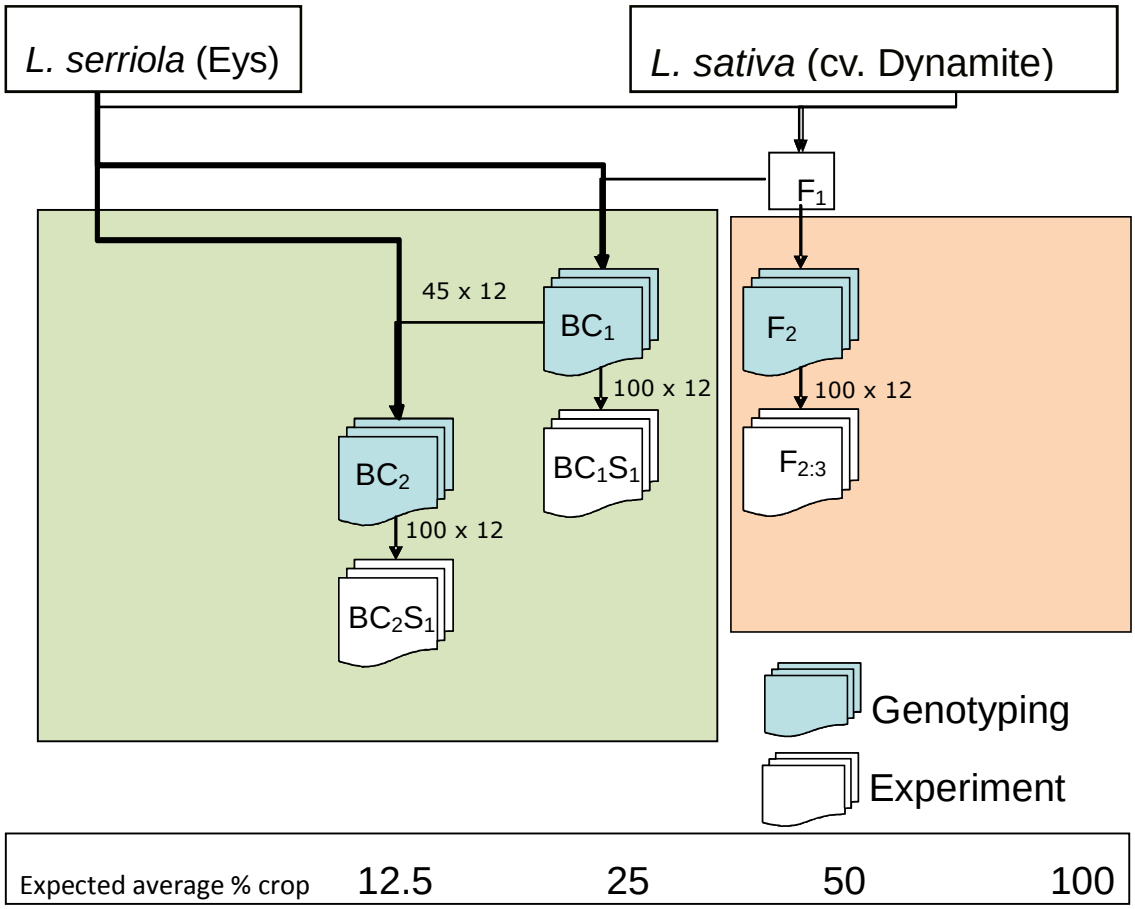


Figure S4 Boxplots representing the phenotypic variation among BC_1S_1 (position 1) relative to *L. serriola* acc. UC96US23 (position 2), *L. sativa* cv. Salinas (position 3), *L. serriola*/Eys (position 4) and *L. sativa* cv. Dynamite (position 5) for vigour traits dry weight (A), fresh weight (B), plant height (C) and relative moisture content (D) under the five treatments

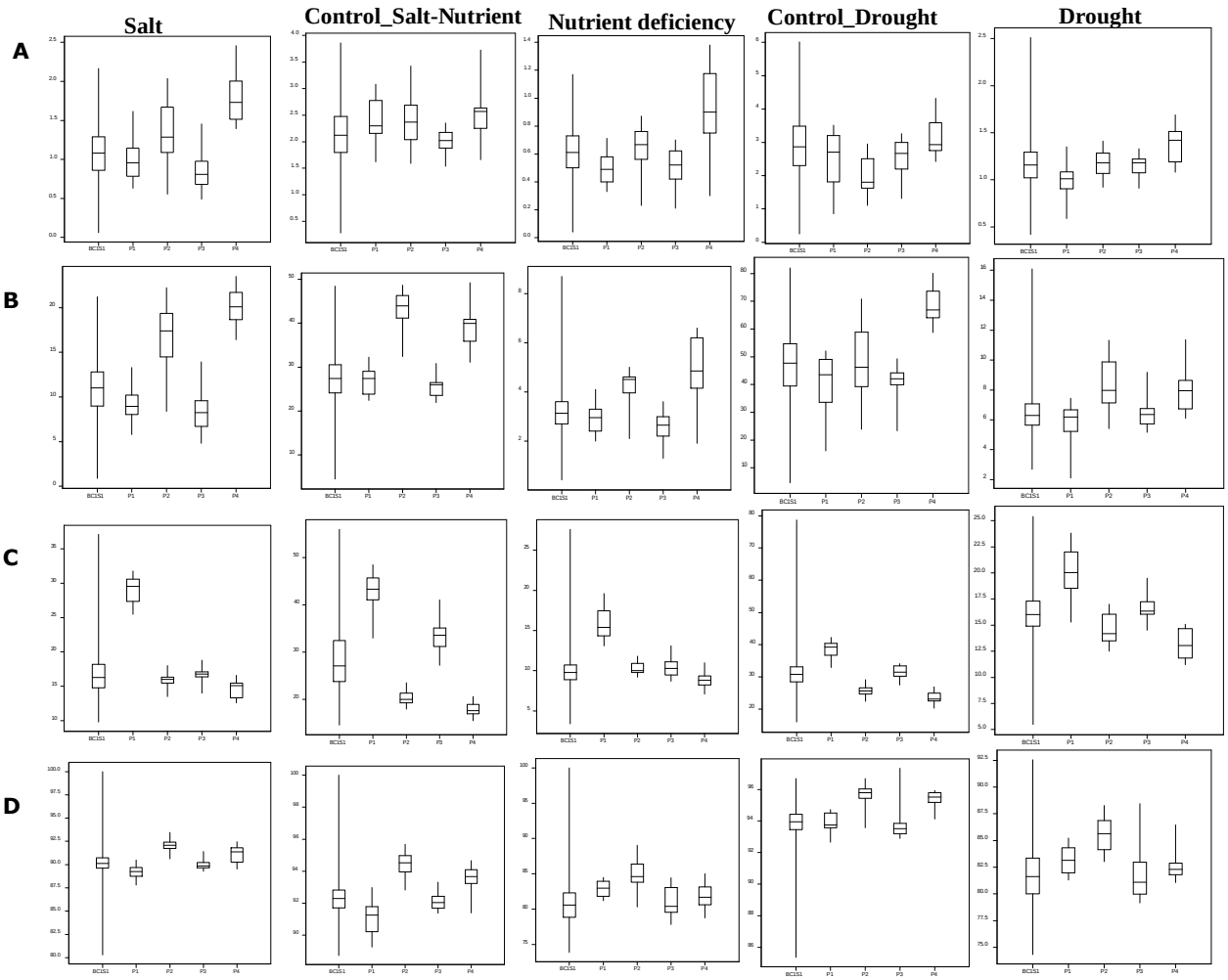
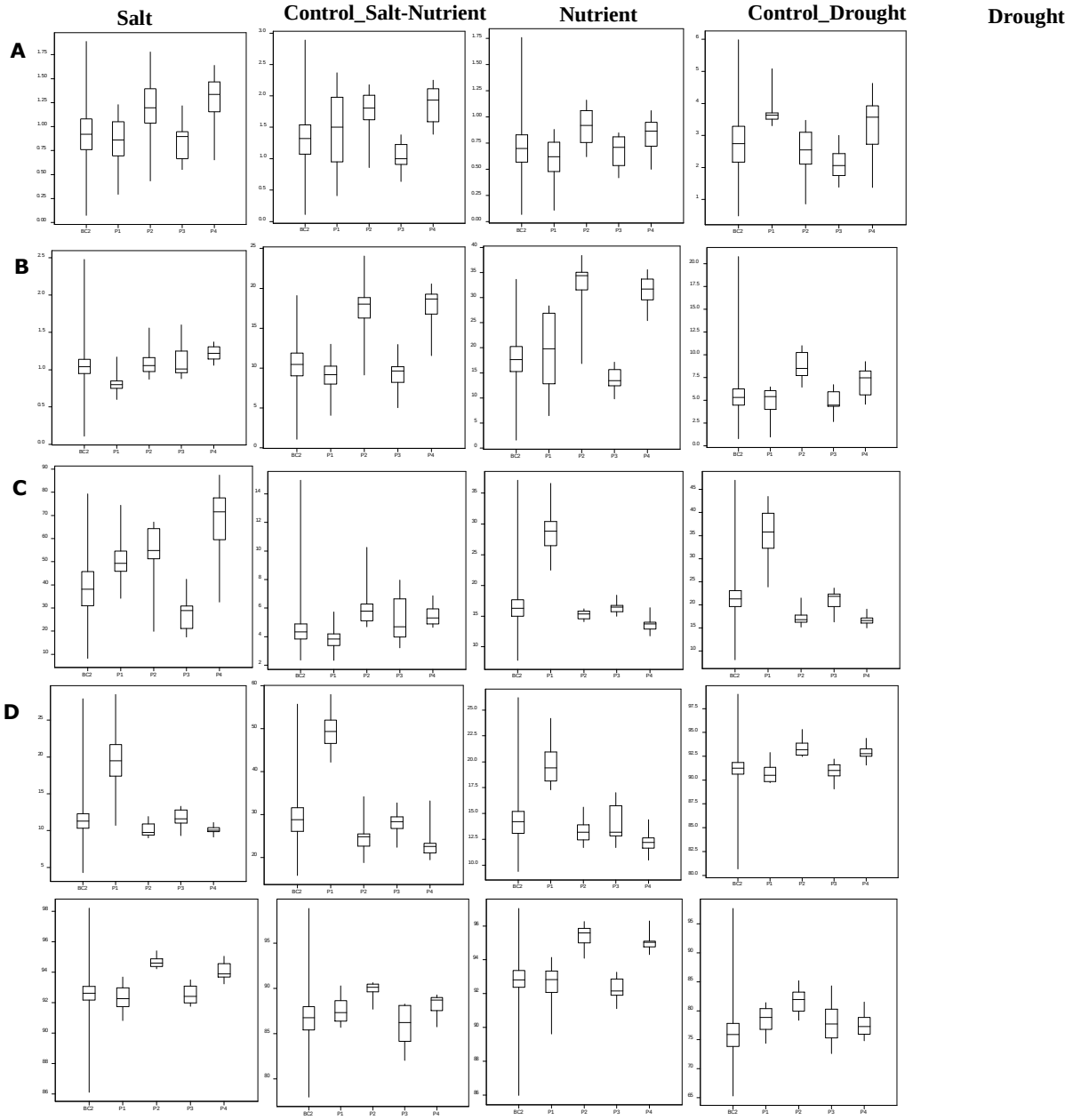


Figure S5 Boxplots representing the phenotypic variation among BC_2S_1 plants (position 1) relative to *L. serriola* acc. UC96US23 (position 2), *L. sativa* cv. Salinas (position 3), *L. serriola*/Eys (position 4) and *L. sativa* cv. Dynamite (position 5) for vigour traits dry weight (A), fresh weight (B), plant height (C) and relative moisture content (D) under the five treatments.



Chapter 5

Table S2 Analysis of variance and deviance of the phenotypic data and the significance of GxE on the traits: all the traits measured at the two sites show significant GxE except Survival

Trait	Source of variation	DF[1]	MS or MD[2]	P-value
Germination	Genotype x site x block	2447	4.8	<.001
Dry weight	genotype	101	1.03	<.001
	Site	1	88.29	<.001
	Genotype x site	101	0.38	0.007
	residual	2231	0.27	
Number of branches	genotype	100	185.12	<.001
	site	1	312.71	0.006
	Genotype x site	99	80.23	<.001
	residual	1607	41.21	
Number of basal shoots	genotype	100	64.24	<.001
	site	1	4129.04	<.001
	Genotype x site	99	32.84	<.001
	residual	1614	11.72	
Days to flowering	genotype	101	1440.2	<.001
	site	1	105506.5	<.001
	Genotype x site	101	354	0.001
	Residuals	1903	235.3	
Number of seeds per capitulum	genotype	100	65.9	<.001
	site	1	9060.06	<.001
	Genotype x site	99	23.68	<.001
	residual	1605	11.78	
Total number of capitula	genotype	100	2.35×10^{6}	<.001
	site	1	1.43×10^{8}	<.001
	Genotype x site	99	1.23×10^{6}	<.001
	residual	1613	4.63×10^{5}	
Total number of seeds	genotype	100	1.03×10^{9}	<.001
	site	1	1.54×10^{10}	<.001
	Genotype x site	99	6.15×10^{8}	<.001
	residual	1602	2.30×10^{8}	
Survival	genotype	101	5.80	<.001
	block	11	3.20	<.001
	site	1	25.8763	<.001
	Genotype x site	101	1.00	0.484
	Genotype x block	1111	1.14	0.001
	Block x .site	11	3.51	<.001
	Genotype x site. x lock	1111	0.63	1
	residual	1111	0.63	
Plant height	genotype	99	1847.2	<.001
	residual	1000	648.4	

1 DF: degrees of freedom; 2 MS: mean square, MD: mean deviance

Acknowledgements

Living in Wageningen for the last six years has been the most gratifying time of my life. I grew up scientifically, morally and spiritually, thanks to the many people who made my stay wonderful and contributed to the starting, running and finishing of the PhD program during the last four years. It is time now to thank those people who made this period so memorable.

My deepest thanks go to **my parents** who taught me that the greatest knowledge and wisdom is to fear the Lord. For every sacrifice you made to make my dreams come true I thank you. To my brothers **Jean-Louis** and **Diogène**, thanks for introducing me to sciences at an early age, if it was not for you I would never have realised my full potential and this thesis would have never seen the day. To my sisters **Umulisa**, **Cathy** and **Adelphine**, I thank you for all the encouragement and support you've given me. Your courage, strength and determination have taught me to strive to be the best I can be.

I am grateful to **Prof. Dr. Richard Visser** for giving me the opportunity to accomplish my childhood dream of doing a PhD. Working in Plant Breeding was like being part of a great international scientific family, help and expertise was always readily available. It was an honour to be a part of this family, and I thank you for giving me such a great opportunity.

Clemens van de Wiel and **René Smulders**, thank you for being my mentors, supervisors and advisors. I feel privileged for the last four years I worked under your guidance. You have taught me, not only how to become an independent scientist, but also to have a well-balanced life. I have learnt and experienced what it is like to really work under great people. It was a humbling experience working with senior scientists that were always willing to listen to my opinion, offer guidance to my ideas and quick to respond to my work. If there was an award for the best supervisors, you would have my vote.

Herman van Eck and **Björn D'hoop**, thank you for encouraging me to undertake the PhD challenge. **Peter van Tienderen**, **Yorike Hartman** and **Danny Hootfman**, thank you for the insight, experience and collaboration shared in our twin projects. **Richard Michelmore** and **Leah McHale,** thank you for your wonderful work on lettuce genotyping.

Hans Jansen, thank you for sparing time for me every time I requested an appointment with you. Every time I walked out of your office I felt statistically smarter and everything seemed easy. **Gerard van der Linden**, **Oene Dolstra** and **Eric van de Weg**, and the **abiotic stress group,** thank you for the inspiring discussions we had. **Paula Hurtado Lopez**, thank you for introducing me to QTL mapping using GenStat. **Martin Boer**, thank you for rushing to my

rescue when GenStat was giving me unrealistic explained variances. **Marieke Jeuken** and **Erik den Boer**, thank you for sharing many tips on working with lettuce. **Annie**, **Mariame**, **Janneke** and **Letty**, it was just a pleasure working with you. You are the tires of this big truck called "Wageningen UR Plant Breeding" and because of you, everything runs so smoothly. Thank you for such a great job.

Working in the greenhouse can be daunting at times so I greatly appreciate all the people that helped out with my work. **Rinie** and **your team in Nergena**, thank you for such a great help with running those huge experiments. **Zheng**, **Song** and **Bai**, it was an immense experience to be your internship supervisor; you taught me a lot, you introduced me to the fascinating Chinese culture and you brought so much help with the greenhouse work. **Marcel**, **Anitha**, **Eugène**, **Hanneke**, **Wendy**, **Raphael**, **Colette**, **Marleen**, **Myriam, Innocent, Gaspard** and **Jean de Dieu,** thanks for sparing your precious time to help me with data collection. To the UNIFARM field experiments team led by **John** and **Gerard**, my deepest gratitude goes out to you. It would have been impossible to set up my experiments and collect all the data without your help.

Living away from home can be stressful. **Jean-Louis**, **Maman Fabrice**, **Fabrice**, **Ariane** and **Bruce**, thanks for being my home away from home for the last six years. I never felt homesick. Home was just at a few-hour train journey. Ce fut mon plus grand plaisir de faire partie de votre famille pendant ces six dernières années et de voir mes potes grandir. Vous allez me manquer énormément.

Colette, **Marleen**, **Floor, Andres, Myriam**, **Evelyn** and **Hulya**, you were great office mates. I thank you for all the support you gave me, the interesting conversations we had and the dinners we shared. I hope that this is not the end of our friendship and we will keep in touch and hopefully work together in the future. You were an integral part of my amazing Wageningen experience. Floor and Colette, if my sown seeds produce their harvest (who knows?), make sure to invite me; wherever I will be I will attend.

The Rwandese community in Wageningen throughout the last six years, I enjoyed being your coordinator for a while and sharing dinners with you at the end of every period. I looked forward for the times we could get together and enjoy the Rwandese food and talk in Kinyarwanda.

Virginia, I'm sure when you left Kenya for Wageningen you had no idea your major assignment was to make Brigitte's life as easy as possible. You have been my housemate, my colleague, my friend and my sister. We have shared so much in the last four years. You

understood me better than most people did. You nursed me when I was sick, cooked for me, prayed for me, listened to me and encouraged me. Thank you for everything. They say that "friends meet to depart and they depart to meet"; I hope that the saying will be true for us.

To **Mariame**, **Anoma**, **Soifia**, **Shital**, **Edwige** and **Virginia**, I will never forget all the insightful girl talks we had. Thank you for your support and advice especially the "legs" one. It has definitely worked for many of us. **Farai**, **Busi**, **Elton**, **Edna** and **Adesua**, I thank God for bringing you into my life. Thank you for continuously offering both spiritual and emotional support throughout my time here. In the years we have been together we have seen and met many people who have come and gone but you have been constantly here. My little friends **Heather** and **Anotida**, I will miss our walks and movie times. My friend **Gaudiose,** having you around was like having a sister around. We have seen each other grow from our single Bornsesteeg days to now having homes and families of our own. I am grateful for your friendship and support.

My last and biggest thanks go to my dear husband **Jean de Dieu Havugimana**. Honey, thank you for your undivided and unwavering support. You have been my continuous source of strength. You encouraged me when I was overwhelmed, you made me laugh when I was sad, you gave me perspective in the days when the challenges seemed too great. The distance and the time difference between us did not make things easy but you did everything you could to make it work. I thank you for the long hours on the phone every single day. Thank you for helping me in the greenhouse and with my data when you were supposed to be on holidays. You are just my best half. I thank God for bringing you into my life. Je t' aime tellement.

With love,
Brigitte Uwimana
Wageningen, December 2nd, 2011

About the author

Brigitte Uwimana was born in Nyakogo, Rwanda on August 20th, 1977. After high school, she did her BSc. in Agricultural Sciences at Allahabad Agricultural Institute-Deemed University, Allahabad, Uttar Pradesh State, in India. Upon graduation in 2002, she returned to Rwanda and worked at the "Institut des Sciences Agronomiques du Rwanda" first as a focal point technician in charge of extension in the Eastern zone, then as a research assistant in the Horticulture program. In 2005 she got admitted at Wageningen University, the Netherlands to do her MSc. in Plant Sciences with a specialization in Plant Breeding and Genetic Resources. Upon graduation in 2007, she started her PhD in the department of Plant Breeding of Wageningen University, the Netherlands under the supervision of Prof. dr. Richard Visser, Dr. Clemens van de Wiel and Dr. René Smulders. She graduated on December 2nd, 2011.

Education Statement of the Graduate School

Experimental Plant Sciences

The Graduate School EXPERIMENTAL PLANT SCIENCES

Issued to: **Brigitte Uwimana**
Date: **2 December 2011**
Group: **Plant Breeding, Wageningen University & Research centre**

1) Start-up phase	*date*
► **First presentation of your project**	
Genetic analysis of the introgression process of wild (*L. serriola*) and cultivated lettuce (L. sativa): background selection and hitchhiking - Introduction of the project	Mar 04, 2008
► **Writing or rewriting a project proposal**	
► **Writing a review or book chapter**	
► **MSc courses**	
Population and Quantitative Genetics (GEN-30806)	2007/2008
► **Laboratory use of isotopes**	
Isotope course: "How to work safely with radioactive substances"	Jan 22-24 & Feb 04, 2008
Subtotal Start-up Phase	*9,0 credits**

2) Scientific Exposure	*date*
► **EPS PhD student days**	
EPS PhD day 2010, Utrecht University	Jun 01, 2010
EPS career day	Nov 18, 2011
► **EPS theme symposia**	
EPS theme 4 : 'Genome Plasticity', Leiden University	Dec 07, 2007
EPS theme 3 : 'Metabolism and adaptation', Leiden University	Feb 19, 2010
► **NWO Lunteren days and other National Platforms**	
Lunteren days 2008	Apr 07-08, 2008
Lunteren days 2010	Apr 19-20, 2010
Lunteren days 2011	Apr 04-05, 2011
► **Seminars (series), workshops and symposia**	
Plant Breeding research day 2008	Jun 06, 2008
Plant Breeding research day 2009	Mar 03, 2009
Plant Breeding research day 2010	Feb 08, 2010
Plant Breeding research day 2011	Mar 07, 2011
Symposium: "Photosynthesis: from femto to Peta and from nano to Global"	Nov 05, 2009
Seminar: "Unraveling abiotic stress tolerance mechanisms in Citrullus and Citrus" by Prof. Fenny Dane	Sep 22, 2009
Plant Sciences Seminars	Dec 2009-Sep 2011
NWO-ERGO workshop	Jun 30, 2009
NWO-ERGO conference	Feb 25-26, 2010
► **Seminar plus**	
► **International symposia and congresses**	
11th Intern. Symposium 'On the biosafety of genetically modified organisms', Buenos Aires, Argentina	Nov 15-20, 2010
Eucarpia Leafy Vegetables conference 2011, Lille, France	Aug 24-26, 2011
► **Presentations**	
Poster: "Genetic analysis of the introgression process from cultivated lettuce (Lactuca sativa) into wild lettuce (L. serriola): genetic hitchhiking and background selection tools for assessing the likelihood of the establishment of transgenes in wild relatives" (Lunteren, The Netherlands)	Apr 07-08, 2009
Presentation: "A genetic analysis of the introgression process from cultivated (Lactuca sativa) to wild lettuce (L. serriola): Crop-wild hybrids receive QTLs for stress tolerance from the crop" NOW-ERGO conference	Feb 25, 2010
Presentation: "BC1 and F1S1: The crop confers to the crop-wild hybrids tolerance against abiotic stresses", 11th ISBGMO, Buenos Aires, Argentina	Nov 15-20, 2010
Presentation: "QTLs for plant vigour under non-stress, drought, salt and nutrient deficiency conditions", Lunteren days, 2011	Apr 05, 2011
Presentation: "QTLs for plant vigour under non-stress, drought, salt and nutrient deficiency conditions", TTI Green Genetics Meeting, Utrecht	Apr 13, 2011
Presentation: "QTLs for plant vigour under non-stress, drought, salt and nutrient deficiency conditions and their effecton crop-wild introgression", Eucarpia Leafy vegetables 2011	Aug 24, 2011
► **IAB interview**	Dec 04, 2009
► **Excursions**	
Subtotal Scientific Exposure	*15,1 credits**

3) In-Depth Studies	*date*
► **EPS courses or other PhD courses**	
Principles of Ecological Genomics	Feb 23-27, 2009
Quantitative genetics of selection response	Jun 07-11, 2010
► **Journal club**	
Literature discussions, Plant Breeding	Oct 2007-Sep 2011
► **Individual research training**	
Training on QTL mapping and analysis, Kyazma	Apr 27-29, 2009
Subtotal In-Depth Studies	*8,4 credits**

4) Personal development	*date*
► **Skill training courses**	
PhD Competence Assessments	Feb 19 & Mar 12, 2008
Interpersonal communication for PhD students	Oct 2008
Techniques for writing and presenting a scientific paper	Jul 01-04, 2008
Information literacy including introduction to Endnote	May 27-28, 2008
Workshop on scientific publishing	Nov 05, 2008
Dutch I and II CENTA -WUR	Sep 2009-Jun 2010
► **Organisation of PhD students day, course or conference**	
► **Membership of Board, Committee or PhD council**	
Subtotal Personal Development	*6,0 credits**

TOTAL NUMBER OF CREDIT POINTS*	**38.5**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits

* *A credit represents a normative study load of 28 hours of study.*

This study was financially supported by the Netherlands Organization for Scientific Research (NWO) through their research programme “Ecology Regarding Genetically modified Organisms” (ERGO).

Cover picture: A branch of prickly lettuce showing ripe seeds and capitula; by Lubomir Hlasek
Cover design: Brigitte Uwimana
Printing: Wöhrmann Print Service, Zutphen, the Netherlands