



## UvA-DARE (Digital Academic Repository)

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

Hartman, Y.

**Publication date**  
2012

[Link to publication](#)

#### **Citation for published version (APA):**

Hartman, Y. (2012). *Genomic regions under selection in crop-wild hybrids of lettuce: implications for crop breeding and environmental risk assessment*. [Thesis, fully internal, Universiteit van Amsterdam].

#### **General rights**

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

#### **Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

# 2

## **QTL analysis reveals the domestication architecture of Crisphead Lettuce**

Yorike Hartman<sup>1</sup>, Danny A.P. Hooftman<sup>1,2</sup>, M. Eric Schranz<sup>1</sup>, Peter H. van Tienderen<sup>1</sup>

<sup>1</sup> Institute for Biodiversity and Ecosystem Dynamics, Universiteit van Amsterdam, The Netherlands

<sup>2</sup> Centre for Ecology and Hydrology, Wallingford, United Kingdom

### Abstract

The domestication architecture of crops is generally characterized by three trends: relatively few genomic regions with major QTL effects are involved, QTL are often clustered, and alleles derived from the crop do not always contribute to the crop phenotype. We have investigated the domestication architecture of lettuce using a Recombinant Inbred Line population from a cross between a crop *Lactuca sativa* (cv. Salinas) and its wild relative *L. serriola*. Few genomic regions with major QTL, plus various intermediate QTL, largely control the transition from wild to cultivated Crisphead lettuce. Allelic effects of all major QTL were in the expected direction, but there were intermediate QTL where the crop contributed to the wild phenotype and vice versa. We found two main regions with clusters of QTL, one on linkage group 3, where the crop allele induced lower seed output, another on linkage group 7, where the crop allele caused a delay in flowering time. Knowledge of genetic changes due to the domestication is relevant for the chance that a transgene inserted in a crop genome will spread to wild relatives due to hitchhiking effects. If a transgene would be inserted in one of these regions, background selection on the crop alleles may lead to a reduced fitness of hybrids with the transgene. QTL research on the effects of domestication genes can thus indicate regions in the crop genome that are less likely to introgress, although these results still need to be verified under field conditions.

### Introduction

The study of crop domestication has received much interest from crop breeders and evolutionary biologists alike (Burger et al. 2008; Hancock 2005). Crops serve as a model for evolution and adaptation, because of the knowledge of crop history and the selective pressures that accompanied domestication (Purugganan and Fuller 2009). Quantitative Trait Locus (QTL) studies of domestication-related traits using crosses between crops and their wild relatives have given insights into the origin of crops and speed of domestication (Gross and Olsen 2010). Furthermore, those crop–wild crosses provide potential source material for crop improvement (Isemura et al. 2010; Vaughan et al. 2007), insight on hybridization and introgression between crops and their wild relatives (Hancock 2005; Hooftman et al. 2009), and show genomic regions under selection in crop–wild hybrids (Baack et al. 2008; Dechaine et al. 2009).

Three important trends in genetic domestication architecture in crops are often highlighted and are the focus of this study: effect size of QTL, clustering of QTL, and directionality of allelic effects. First, the transition from wild-type to cultivated crop is often controlled by relatively few genomic regions with major QTL effects (Burger et al. 2008; Gross and Olsen 2010; Ross-Ibarra 2005). Such regions contain very low variation among cultivars, indicating a rapid and uniform fixation caused by continuous artificial selection over many years (Burke et al. 2007). In maize (Doebley and Stec 1993), rice bean (Isemura et al. 2010), eggplant (Doganlar et al. 2002), tomato (Frary et al. 2000), and rice (Cai and Morishima 2002), a small number of QTL cause large phenotypic changes in seed shattering, seed and fruit size, and branching patterns. In contrast, in sunflower domestication was driven by a relative high number of QTL, many with minor or intermediate effect (Burke et al. 2002, 2007). Second, domestication-related QTL are not uniformly distributed over the genome, but are often clustered (Burger et al. 2008; Ross-Ibarra 2005). For example, clustering has been found in sunflower (Burke et al. 2002; Dechaine et al. 2009), azuki bean (Kaga et al. 2008), rice bean (Isemura et al. 2010), rice (Cai and Morishima 2002), eggplant (Doganlar et al. 2002), pearl millet (Poncet et al. 2000), and tomato (Prudent et al. 2009). Third, the majority of QTL usually show phenotypic effects in the expected direction; i.e. crop alleles contribute to a crop phenotype. However, the opposite

also occurs, with crop alleles contributing to the wild phenotype and wild alleles to the crop phenotype (Ross-Ibarra 2005).

The underlying cause of these three trends is not yet well understood. It has been suggested that the relative importance of major QTL and the presence of QTL clusters reflects a rapid domestication driven by strong directional selection (Rieseberg et al. 2002; Ross-Ibarra 2005), whereas a majority of intermediate and minor QTL might reflect a gradual process with multiple domestication events (Burke et al. 2002). The clustering seen in many crop species might be caused by the fact that species with beneficial genes in clusters are more easily domesticated than species with beneficial genes more dispersed over the genome (Burger et al. 2008), or alternatively, such clustering might be the result of single genes with pleiotropic effects over several traits. In any case, more detailed analyses are necessary to draw more definitive conclusions.

Domestication research plays an important role in the debate surrounding genetically modified (GM) crops. With the introduction of GM crops concerns have arisen about hybridization and the chances for transgene escape from crops to their wild relatives. A particular concern is possible negative ecological effects, such as increased invasiveness of the wild relative (Chapman and Burke 2006; Stewart et al. 2003). In order to minimize chances of transgene escape, mitigation strategies have been proposed where the transgene is in linkage with an allele that is selected against in the wild and therefore is more likely to be purged from the wild population (Gressel 1999; Kwit et al. 2011; Stewart et al. 2003). Locating domestication-related QTL might be a way to pinpoint genomic areas where the crop allele confers a fitness disadvantage to hybrid individuals. Many crops share a number of traits, termed the ‘domestication syndrome’, which were selected for in early stages of domestication and made crops easier to cultivate (Harlan 1992). Although many crops are reported to hybridize with their wild relatives (Ellstrand 2003), these domestication traits were until recently seen as unable to introgress into the wild and were generally believed to impose negative fitness effects on hybrid individuals in the wild (Hails and Morley 2005). However, Ellstrand et al. (2011) showed 13 examples of suggested evolution of weediness from domesticated ancestors of which seven are via hybridization with wild relatives.

Due to their economic importance, the majority of domestication research to date has focused on seed-propagated annual crop species (Burger et al. 2008; Gross and Olsen 2010), in particular cereal crops such as wheat, maize, and rice. Whether or not the three general patterns described earlier extend to other types of crop remains to be seen (Gross and Olsen 2010). Within the Compositae there are two major crops, namely sunflower and lettuce (Dempewolf et al. 2008). We study domestication traits in lettuce (*Lactuca sativa*), an annual vegetable crop grown for its leafy head rather than for its seeds (achenes). The recent availability of genomic sequences, genetic maps, and genotyped Recombinant Inbred Lines (RILs) makes extensive analyses of domestication-related traits in lettuce possible (<http://compgenomics.ucdavis.edu/>).

*Lactuca serriola* is fully cross compatible and interfertile with cultivated lettuce (de Vries 1990). It occurs in Europe, North America, South Africa, Asia and Australia, mainly in pioneer, and/or disturbed ruderal habitats, but also as a weed in agricultural fields (D’Andrea et al. 2009; Lebeda et al. 2001). In the past 50–60 years, it has expanded its range dramatically in Western Europe (D’Andrea et al. 2009; Hooftman et al. 2006). This invasiveness is probably closely linked to human activities (D’Andrea et al. 2009; Lebeda et al. 2001) although effects of introgressed crop genes to the wild population cannot be excluded (Hooftman et al. 2006). Lettuce is predominantly selfing, but also insect pollinated with a reported outcrossing rate of approximately 1–5% (D’Andrea et al. 2008; Giannino et al. 2008). In a field experiment,

## Chapter 2

Hooftman et al. (2005, 2007) showed that artificial hybrids of *L. serriola* and *L. sativa* had an increased rather than a decreased fitness compared to the wild-type. This suggests that hybridization can lead to novel genetic combinations that can be more vigorous than the original parent lines and that crop genes in lettuce are not necessarily negatively selected for.

It has already been shown that RILs of the cross between cultivated *L. sativa* and wild *L. serriola* used in this study have differentiated root architecture and water use patterns (Johnson et al. 2000), seed and seedling traits (Argyris et al. 2005), and leaf biophysical and developmental traits (Zhang et al. 2007). With regard to morphological domestication traits, lettuce shows a reduced natural seed (achene) dispersal and increased seed size compared to the wild-type, similar to cereal crops (de Vries 1997). More typical for a leafy vegetable, domestication in lettuce has been targeted at delay of bolting, increased head formation, loss of spines, decrease in latex content and bitter taste, and a change from long serrate leaves to broad almost circular leaves (de Vries 1997). The cultivar used in this study, Crisphead or Iceberg lettuce (*L. sativa* cv. Salinas), is mainly cultivated in the United States and was developed to form very tight dense heads (de Vries 1997).

In this study, we will focus on the identification of domestication- and fitness-related QTL for various life stages and compare the results with the genetic architecture of domestication found in previous lettuce research and in other crops. Our focus is on developmental, leaf shape, and seed output traits. We will discuss the results in the view of the development of new transgenic crops, and the likelihood of introgression of crop genes to wild relatives.

## Material & methods

### Plant material and growing conditions

We used an existing Recombinant Inbred Line (RIL) population, selfed for nine generations, from a cross between the crop *Lactuca sativa* cv. Salinas and its wild relative *L. serriola* f. *serriola* (UC96US23). The RILs are from the same set as used in Argyris et al. (2005), Johnson et al. (2000), and Zhang et al. (2007). In total, 114 RILs were grown in the greenhouse under 6h dark and 18h light, a minimum of 18°C, under 600 W SON T-Agro lamps generating on average 160  $\mu\text{mol}/\text{m}^2/\text{sec}$  at plant level. One plant per line was grown from the end of November 2007 and followed through the entire lifecycle until the last plants produced seed in August 2008.

In addition, we performed a germination experiment in a germination cabinet under 16h of light at 20°C and 8h dark with 15°C. The experiment lasted 9 days in total until the majority of lines reached 90–100% germination. Achenes (further referred to as seeds) were placed in Petri dishes on filter paper and watered with sterilized water to induce germination. We added a small amount of TMTD (tetramethyl-thiuram-disulfide) powder to prevent the formation of fungi on the seeds. We phenotyped the 114 RIL and parent lines; all lines were replicated twice. Each Petri dish contained three to four lines with each ten seeds. Lines were randomized and, to prevent any position effects, trays and Petri dishes were shuffled around three times every day.

### Trait measurements

At several life stages, we measured different domestication- and fitness-related traits focusing on developmental, leaf shape, flowering phenotype, spines, seed phenotype, and seed output traits (Table 1).

**Table 1. Traits examined in a *Lactuca sativa* cv. Salinas × *Lactuca serriola* recombinant inbred lines population.**

Plant stage	Trait	Abbreviation	Evaluation method	Type
Seedling	Germination time (days)	GT	No. of days when 50% of seeds have germinated	Development
30 days after sowing	Rosette leaf length (cm)	RLL	Average distance from the base to leaf tip of four biggest leaves	Leaf shape
	Rosette leaf width (cm)	RLW	Average maximum width of four biggest leaves	Leaf shape
	Rosette leaf surface (cm <sup>2</sup> )	RLSU	Surface area of the biggest rosette leaf determined by scanning the leaf	Leaf shape
	Rosette leaf serration	RLSE	No serration, little serration, much serration	Leaf shape
	Rosette leaf circularity	RLC	Amount of serration of the biggest rosette leaf determined by scanning the leaf	Leaf shape
	Plant height (cm)	PH30	Length of main stem at 30 days after sowing, values log-transformed	Development
	Proportion of stem leaves (%)	PSL30	No. of stem leaves divided by the total no. of rosette and stem leaves at 30 days, values angular-transformed	Development
60 days after sowing	Number of stem leaves (count)	SL60	No. of stem leaves at 60 days after sowing	Development
	Plant height (cm)	PH60	Length of main stem at 60 days after sowing, values log-transformed	Development
Flowering	Days to first flower (day)	FLD	No. of days from sowing to flowering of first flower, values log-transformed	Development
	Flower production	FLP	No flowering or flowering	Development
	Stem leaf length (cm)	SLL	Distance from the base to leaf tip of two leaves halfway the main stem	Leaf shape
	Stem leaf width (cm)	SLW	Maximum width of two leaves halfway the main stem	Leaf shape
	Stem leaf surface (cm <sup>2</sup> )	SLSU	Surface area of a stem leaf halfway the main stem determined by scanning the leaf	Leaf shape
	Stem leaf serration	SLSE	No serration, little serration, much serration	Leaf shape
	Stem leaf circularity	SLC	Amount of serration of a stem leaf halfway the main stem determined by scanning the leaf	Leaf shape
Seed set	Days to first seed (day)	SSD	No. of days from sowing to seed set of first seed, values log-transformed	Development
	No. of stem leaves (count)	SLSS	Total no. of stem leaves	Development
	Plant height (cm)	PHSS	Length of main stem	Development
	No. of reproductive basal shoots (count)	SHN	No. of basal side shoots which have flower buds, flowers and/or seed heads	Seed output
	No. of branches main inflorescence (count)	BRN	No. of branches counted from the base of the main inflorescence to the top	Seed output
	Inflorescence shape	IS	Spike or pyramid	Flowering phenotype
	Inflorescence length (cm)	IL	Length of main inflorescence from the first node at the base until the top	Flowering phenotype
	Involucral bracts	IB	Erect, erect/reflexed, or reflexed	Flowering phenotype
	Stem thickness (cm)	STT	Diameter of the main stem 10 cm above the soil	Development
	Stem leaf spines	SLSP	No spines, few spines, or many spines	Spines
	Stem base spines	SBSP	No spines, few spines, or many spines	Spines

## Chapter 2

**Table 1.** Continued.

Plant stage	Trait	Abbreviation	Evaluation method	Type
After seed set	Seed length (mm)	SDL	Maximum distance from top to bottom of the seed	Seed phenotype
	Seed width (mm)	SDW	Maximum distance from side to side of the seed	Seed phenotype
	Seed weight (mg)	SDWT	Average weight of 50 seeds	Seed phenotype
	Pappus length (mm)	PL	Length of the stalk	Seed phenotype
	Pappus width (mm)	PW	Maximum distance between the outer tips of the radiating hairs	Seed phenotype
	No. of seeds per capitulum (count)	SDC	Average no. of seeds per capitulum based on 15 collected capitula	Seed output
	Total no. capitula	TC	Total no. of capitula developed, calculation following Hooftman et al. (2005)	Seed output
	Seed output	SDO	Total no. of seeds produced, calculation following Hooftman et al. (2005); values square-root-transformed	Seed output
	Survival until reproduction	SUR	No seeds formed or seed formation	Seed output

*Development* – Germination was counted daily for 9 days; seeds were considered as germinated if the two cotyledons were fully emerged from the seeds and roots were visible. Since most lines reached 90–100% germination, we calculated the time point at which 50% of seeds had germinated using an existing R-code (version 2.12.1; <http://lukemiller.org/index.php/2010/02/calculating-lt50-median-lethal-temperature-aka-ld50-quickly-in-r/>, March 11, 2011).

At 30 and 60 days after sowing, we measured several plant characteristics for one plant per RIL. Plant height was measured and the number of rosette and stem leaves was counted. From this we calculated the proportion of stem leaves by dividing the number of stem leaves by the total number of rosette and stem leaves. At 60 days, the majority of lines had bolted and produced shoots, therefore we only used the number of stem leaves in the final analysis. Flowering and seed set was measured as the number of days from sowing to the production of the first flower or seeds, respectively. At seed set, we also measured plant height, stem thickness of the main stem at 10 cm above the soil, and we counted the total number of stem leaves.

*Leaf shape* – Leaf shape was determined quantitatively as well as qualitatively. At 30 days after sowing, we measured the length and maximum width of the four biggest rosette leaves. At the flowering stage, we also measured the length and maximum width of two leaves halfway the main stem. In both cases, the biggest leaf was scanned and images were analyzed with ImageJ v1.41 (Abramoff et al. 2004) to determine leaf surface and circularity as a measure of the amount of serration of the leaves. Circularity was calculated as: circularity index =  $4\pi \times (\text{leaf surface}/\text{perimeter}^2)$ . In addition, we scored rosette and stem leaf serration categorically (no serration, little serration, much serration).

*Flowering phenotype and spines* – At the seed set stage, we measured the length of the main flowering stem from the first node at the base of the inflorescence until the top. In addition, we scored the inflorescence shape (spike-like: most capitula in a horizontal plane, or pyramid-like: capitula of main stem higher than those of axillary branches), the position of the involucre bracts of capitula that had set seed (erect, erect/reflexed, or completely reflexed), and the presence of spines on the stem base as well as on the stem leaves (no spines, few spines, or many spines).

*Seed phenotype and seed output* – We collected 15 capitula for seed (or the achene fruit to be precise) morphology and seed output measurements. We measured seed length and width, and also measured pappus length and width of 5 randomly selected seeds. The total number of seeds was counted for the 15 capitula to calculate the average number of seeds per capitulum. In order to calculate the total seed output, we counted the number of reproductive basal side shoots and the number of branches from the base of the main flowering stem to the top. Subsequently, we estimated the total number of capitula and the total seed output following Hooftman et al. (2005, eqn 1). Subsequently, the total seed output was calculated by multiplying the number of capitula and the average number of seeds per capitulum.

### Statistical analysis

All statistical analyses were performed in PASW Statistics 17.0 (SPSS Inc. 2009), with the exceptions mentioned earlier. Several trait values were transformed to normalize the distribution and improve QTL analyses. The fraction number of stem leaves at 30 days (relative to the total number of rosette and stem leaves) was arcsine-square-root-transformed. Plant height, days until first flowering, and days until seed set were log-transformed. Estimated seed output values ranged from 242 to 62,847 seeds; these values were square-root-transformed. For all other traits the data was normally distributed and hence, we used the trait values for the QTL analysis. We performed a Principal Component Analysis (PCA) using a Varimax rotation method to summarize the variation among RILs; missing values were replaced by the mean. The loading values of the RILs for the first two axes were saved and also used in the QTL analysis.

### Quantitative trait loci analysis

Genetic map and marker data used in the QTL analysis were obtained from The Compositae Genome Project website (<http://compgenomics.ucdavis.edu>). The genetic map consisted of 1513 markers distributed over the nine chromosomal linkage groups (<http://cgpdb.ucdavis.edu/GeneticMapView/display/>; map version: RIL\_MAR\_2007\_ratio). All QTL analyses were performed with Composite Interval Mapping (CIM) in QTL Cartographer version 2.5.008 (Wang et al. 2010). Tests for the presence of a QTL were performed at 2 cM intervals using a 10 cM window and five background cofactors, which were selected via a forward and backward stepwise regression method. Statistical significance threshold values ( $\alpha = 0.05$ ) for declaring the presence of a QTL were estimated from 1000 permutations (Churchill and Doerge 1994; Doerge and Churchill 1996). One-LOD support intervals and additive effects were calculated from the CIM results. The linkage map and QTL were drawn with MapChart 2.2 (Voorrips 2002).

## Results

We identified 71 QTL for 37 traits and an additional five QTL when summarizing these traits in PCA axes (Table 2; Figure 1). The range of Phenotypic Variation Explained (PVE) per QTL varied from 7.9% to 69.6%. The majority of QTL was of intermediate effect (PVE between 10–25%) as defined by Burke et al. (2002). Only seven QTL were of minor effect (PVE < 10%) and fourteen QTL had a major effect (PVE > 25%).

QTL were distributed over all nine linkage groups. For each trait one to six QTL were detected (mean = 1.9), except for total inflorescence length and number of seeds per capitulum, for which no QTL were detected. The 1-LOD support intervals ranged from 0.3 cM to 5.9 cM (mean = 2.6 cM). Although QTL were distributed over all linkage groups, the majority of QTL clustered with at least one other QTL. Only 17 (out of 71) QTL were uniquely located.

## Chapter 2

**Table 2. QTL detected by composite interval mapping in a *Lactuca sativa* cv. Salinas × *Lactuca serriola* recombinant inbred lines population and PCA loadings for the first two axes.** For abbreviations, see Table 1; nd = no QTL detected; PVE = Percentage Variation Explained. Positive additive effects indicate that the crop-type (*L. sativa*) allele increases trait values and negative additive effects indicate that the wild-type (*L. serriola*) allele increases trait values.

Trait	LG	Position (cM)	LOD	1-LOD interval	Additive effect	PVE (%)	Threshold 0.05	PCA-1	PCA-2
GT	7	65.9	3.9	65.4–67.7	0.39	12.1	3.2	-0.04	-0.00
RLL	7	17.4	5.1	15.5–18.5	-1.65	17.1	3.3	0.32	0.10
RLW	1	89.0	3.7	88.3–91.0	0.75	10.0	3.3	0.12	0.15
	3	37.2	5.5	36.8–40.9	-0.99	15.2			
	3	47.0	4.8	46.3–47.8	-0.90	13.4			
	7	97.8	7.1	97.1–99.0	1.03	18.5			
	7	114.6	5.1	112.8–114.7	0.89	13.9			
	8	132.0	3.4	130.9–132.9	0.71	7.9			
RLSU	1	1.7	3.8	1.6–1.9	-13.37	8.7	3.2	0.31	0.13
	1	89.0	7.8	87.9–89.7	20.53	20.4			
	3	38.8	5.7	37.9–41.0	-18.06	15.0			
RLSE	3	42.9	14.3	41.6–43.5	-0.48	34.7	3.5	-0.02	0.54
	7	97.8	7.6	96.6–100.0	0.31	16.4			
RLC	1	72.4	10.0	71.9–72.4	0.05	29.6	3.4	-0.03	-0.27
	1	76.6	12.7	76.4–77.3	0.05	35.9			
PH30	7	17.4	8.0	15.5–21.0	-0.22	23.7	4.2	0.63	0.03
PSL30	6	129.9	3.9	129.7–132.3	-0.20	8.8	3.5	0.74	0.20
	7	18.5	15.7	18.4–20.3	-0.45	43.9			
	8	33.5	3.8	32.2–34.3	-0.21	9.0			
SL60	7	19.2	5.5	18.7–22.4	-28.55	16.7	3.4	0.40	0.60
PH60	2	111.3	4.1	110.5–116.4	-0.18	8.7	3.5	0.73	0.32
	6	130.8	7.3	129.9–132.8	-0.26	17.4			
	7	21.9	15.3	19.6–22.4	-0.39	41.5			
	8	23.4	4.1	22.1–24.7	-0.19	8.9			
FLD	2	106.8	4.1	105.5–107.5	0.05	12.4	3.4	-0.82	-0.21
	7	19.2	11.0	18.5–19.9	0.08	35.0			
FLP	9	1.0	4.8	0.0–1.4	-0.18	17.0	3.4	0.10	0.05
SLL	2	17.8	3.7	15.9–18.1	1.62	11.3	3.4	0.31	-0.13
	7	109.4	4.8	108.8–109.8	-1.81	14.1			
	9	0.2	3.8	0.0–2.1	1.78	12.9			
SLW	3	37.2	6.7	36.7–39.0	-1.12	23.6	3.3	0.22	0.05
	7	7.2	3.4	6.2–7.4	-0.77	11.6			
SLSU	6	119.9	4.0	117.5–120.5	-12.44	19.3	3.4	0.08	-0.35
SLSE	3	42.9	16.2	41.6–45.8	-0.55	52.9	3.5	0.02	0.60
SLC	1	38.9	4.6	37.6–40.2	0.04	12.4	3.8	0.12	-0.37
	2	136.3	6.2	134.8–138.0	0.05	15.9			
	3	45.4	11.6	45.3–46.6	0.07	37.9			
SSD	1	30.8	3.9	30.7–31.6	0.04	13.9	3.3	-0.88	-0.25
	2	106.8	5.1	105.5–107.1	0.05	15.6			
	7	19.2	7.0	17.9–19.9	0.07	27.6			
SLSS	9	55.9	4.3	55.0–56.3	-31.92	16.5	3.5	0.18	0.58
PHSS	1	79.2	3.4	78.2–80.8	-0.04	13.0	3.4	-0.45	0.11
	6	119.9	3.5	118.2–124.1	-0.04	11.0			
	7	12.9	6.4	11.3–15.2	0.05	26.9			
SHN	3	9.2	3.5	8.1–11.5	1.18	12.2	3.4	0.43	0.74

**Table 2.** Continued

Trait	LG	Position (cm)	LOD	1-LOD interval	Additive effect	PVE (%)	Threshold 0.05	PCA-1	PCA-2
SHN	3	38.8	3.9	37.2–40.9	-1.25	15.1			
	7	18.4	4.9	16.4–19.2	-1.34	17.4			
BRN	7	25.1	5.5	23.3–25.9	2.11	16.6	3.3	-0.41	0.74
IS	4	154.9	3.9	154.3–157.0	-0.18	13.9	3.9	0.03	-0.23
IL	nd							0.16	0.54
IB	1	77.6	5.4	77.1–78.8	0.33	11.5	3.6	-0.06	0.16
	6	5.6	21.3	4.3–7.7	-0.75	61.3			
STT	5	91.0	5.3	90.5–91.6	-0.15	17.6	3.4	-0.06	-0.34
	7	111.9	3.6	110.9–113.5	-0.12	11.6			
SLSP	5	146.8	17.4	146.1–148.1	-0.60	49.7	3.9	-0.01	-0.30
SBSP	5	146.8	23.4	146.1–149.7	-0.81	69.6	3.5	-0.07	-0.46
SDL	1	69.3	4.3	67.9–73.4	-0.15	16.2	3.3	0.63	0.07
	7	10.2	6.3	8.7–12.1	-0.18	23.2			
SDW	8	19.5	5.8	17.7–21.1	-0.07	21.7	3.4	0.52	-0.00
	9	72.3	3.8	71.0–76.3	0.06	14.2			
SDWT	1	26.1	4.1	24.7–26.5	-5.16	17.0	3.4	0.68	-0.17
PL	4	0.4	3.9	0.0–1.9	-0.20	13.0	3.4	-0.35	-0.01
PW	2	134.8	5.9	132.6–137.3	-0.34	21.8	3.5	0.09	-0.01
	4	7.6	4.9	7.4–10.3	-0.30	17.0			
	4	138.5	3.8	137.5–139.2	+0.24	10.9			
SDC	nd							-0.67	0.32
TC	3	9.2	4.8	8.1–9.7	260.21	20.7	3.4	0.26	0.81
	3	38.8	7.6	37.2–40.0	-287.75	27.5			
SDO	3	38.8	4.5	37.2–40.9	-18.98	16.7	3.3	-0.44	0.63
	4	143.3	3.3	141.4–144.2	-15.48	11.4			
	7	105.2	3.6	103.5–106.7	-15.43	11.8			
SUR	6	23.9	4.6	23.3–25.3	-0.15	13.1	3.6	0.07	0.08
	7	77.0	3.4	75.3–78.0	-0.13	9.9			
<b>PCA-QTL</b>								<b>Total variance explained (%)</b>	
PCA1	1	30.8	5.7	29.8–31.6	-0.37	13.4	3.3	18.04	
	5	117.0	3.4	112.3–121.2	0.36	12.7			
	7	18.4	13.8	17.4–19.2	-0.60	35.2			
PCA2	3	38.8	10.6	37.2–41.0	-0.57	29.7	3.4	10.67	
	5	142.3	3.5	140.1–144.5	0.29	8.3			

This high level of clustering occurred mainly in two regions on linkage group (LG) 3 and LG7 (Fig. 1), where nine out of the fourteen major QTL are localized. Additional major QTL were found on LG1 for rosette leaf circularity and on LG6 for involucre bracts. The cluster on LG3 is mainly determined by leaf shape and seed output traits. At this location, four major QTL co-localize, including rosette and stem leaf serration, stem leaf circularity, and total number of capitula. This region is also supported by QTL of intermediate effect for shoot number and total seed output. These same traits have the highest loading scores for the second PCA-axis that explained 10% of the variation (Table 2). A major PCA-2 QTL (PVE = 29.7%) indeed co-localizes with the clustering on LG3, confirming this region as a major region for leaf shape and seed output traits. Another trait with a high loading score for the second PCA axis is stem base spines. Not surprisingly, another PCA-2 QTL co-localizes with two major spine-QTL found on LG5.

## Chapter 2

In contrast, QTL and PCA results show that the clustering on LG7 is the most important region for growth traits, where peak values of five major QTL co-localize within 9 cM. This includes proportion of stem leaves after 30 days, plant height after 60 days, days to first flower, days to first seed, and plant height at seed set (Table 2). These traits, except plant height at seed set, also showed the highest loading scores for the first PCA axis that explained 18% of the variation. A major PCA-1 QTL (PVE = 35.2%) confirms this clustering. Therefore, we consider this the major region for growth related and earliness traits.

For 21 traits, we identified more than one QTL. Of these, QTL effects of seven traits were always increased by the wild parent allele. These included proportion of stem leaves after 30 days, plant height at 60 days, stem leaf width, stem thickness, and seed output traits such as seed length, seed production, and seed output. In contrast, four traits (rosette leaf circularity, stem leaf circularity, days to first flower, and days to first seed) were always increased by the crop parent allele. This shows that the wild parent alleles cause a faster development with an earlier bolting and flowering time, and an increased seed output.

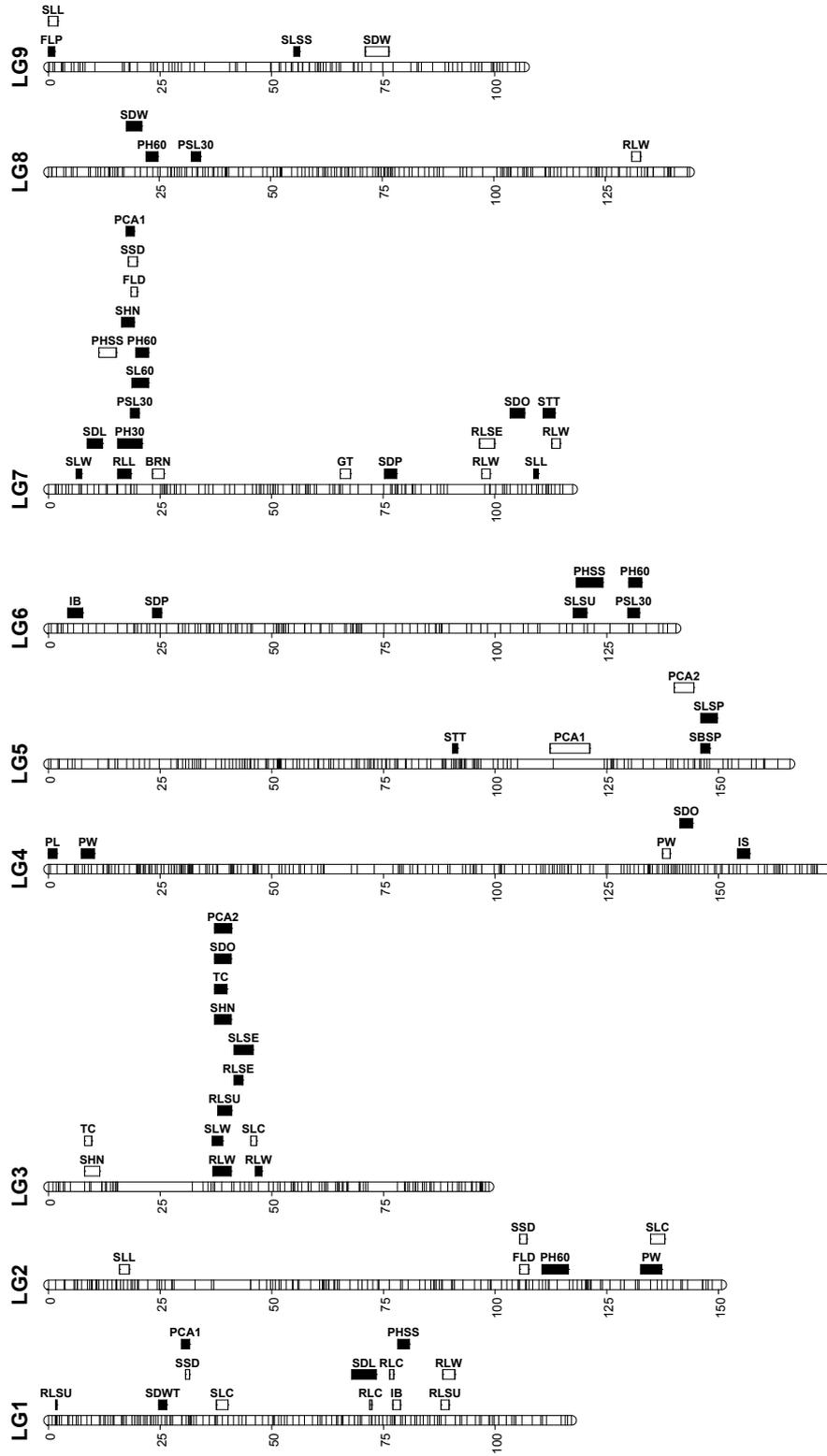
Ten traits showed opposing effects with some QTL where crop alleles resulted in higher values for a trait, but one or more others where they led to lower values. These were all QTL of intermediate effects, including a number of leaf shape traits, but also seed output traits such as seed weight, shoot number, and total number of capitula. Apparently, for each of these traits the crop allele contributed to an increase in trait values at one location, but at another location the increase came from the wild parental allele. Whether there are selective advantages of the direction of the effects could of course not be determined under greenhouse conditions.

## Discussion

We identified 71 domestication- and fitness-related QTL for this lettuce cross of which 14 were of major effect (PVE > 25% following Burke et al. 2002). Of the remaining QTL, the majority was of intermediate effect (50) and only seven were of minor effect (PVE < 10%). Cultivated Crisphead lettuce appears to have been selected for no spines, erect involucre bracts, round almost circular leaves instead of serrate leaves, and a delay in bolting beneficial for vegetative biomass production (de Vries 1997). Several QTL of major effect could be linked directly to these assumed selection regimes. For instance, two closely linked major QTL for rosette leaf circularity on linkage group (LG) 1 are directly related to the cultivated allele inducing rounder leaves. On LG5, we found two major QTL for stem base spines and stem leaf spines and as expected, the wild allele induced spine production. Similarly, for involucre bracts we found one major QTL (LG6). Here, the cultivated parental allele induced erect involucre bracts that are important to prevent shattering of the seeds (achenes).

Two regions were identified – on LG3 and LG7 – that have significant clustering of major as well as intermediate QTL. Our data show that the region on LG3 is involved in leaf shape and seed output traits, whereas the clustering on LG7 is mainly involved in developmental traits, such as height after 30 days and days to first flower, supported by some intermediate leaf shape and seed output QTL. PCA-QTL co-localize and confirm these clustering regions. Based on marker comparisons between different linkage maps, it appears that regions overlap with those found by Zhang et al. (2007) for leaf biophysical and developmental traits, such as leaf area (LG3) and absolute growth rate (LG7) although the exact map alignment is difficult due to limited marker overlap. Such clustering is consistent with either many linked genes and/or of few genes with major pleiotropic effects.

Most traits we measured were morphological or phenological (growth habit, leaf shape, flowering date) and in general such traits show little environmental variability under controlled conditions in the greenhouse. This means that they have relatively high heritability



**Figure 1. Genomic locations of quantitative trait loci detected in composite interval mapping.** The map consisted of 1513 markers indicated by horizontal lines on the bars. Map distances (cM) are shown on the left side of the bars. Bars to the right represent one LOD confidence intervals; for abbreviations, see Table 1. A white bar indicates that the crop-type (*Lactuca sativa* cv. Salinas) allele increases the trait values, whereas a black bar indicates that the wild-type (*Lactuca serriola*) allele increases the trait values.

## Chapter 2

values compared to, for example, fitness-related traits or field data (Visscher et al. 2008), also explaining the high lod-scores that could even be further improved by using more RILs and replicates. Moreover, the QTL clusters found are consistent in location with those in other experiments (Hartman et al. 2012). Therefore, we are confident in the robustness of the QTL, and the randomization of the lines in both greenhouse and germination tests makes it unlikely that environmental factors confounded the results.

### **Trends in genetic domestication architecture**

#### *Strength of QTL regulation*

Regarding the strength distribution across the QTL found, lettuce seems to take an intermediate position among other crops. For most crops, the transition from wild-type to cultivated crop seems to have been regulated by a small number of QTL that are all of large effect (Burger et al. 2008; Gross and Olsen 2010; Ross-Ibarra 2005). For example, one QTL of large effect included one single gene that was responsible for the transition from small to large fruit size in tomato (Frery et al. 2000). As already mentioned in the introduction, sunflower seems to be an exception, with many QTL of intermediate and minor effects (Burke et al. 2002, 2007). For lettuce, we found only very few minor QTL and 14 major QTL, which represented the main phenotypic transitions from wild-type to crop. This is comparable with the pattern found in the majority of crops in which phenotypic differentiation is mainly caused by a few genomic regions of major effect (Burger et al. 2008; Hancock 2005). However, the majority of QTL in this study was of intermediate effect, which is more similar to the patterns found in sunflower (Burke et al. 2002; Wills and Burke 2007).

A possible statistical explanation might be that minor QTL went undetected as a result of small sample size (Beavis 1998), biasing the results towards QTL of large effect. However, the large amount of located intermediate QTL suggests that this distribution of QTL is a real feature of the domestication architecture of lettuce. The main phenotypic differentiation between wild and cultivated lettuce seems to be the result of a few genomic regions with major QTL, plus many QTL of intermediate effect. Whether this is the result of the additive effects of individual genes or through epistatic interactions is to be studied (Uwimana 2011).

#### *Clustered domestication QTL*

A major trend prevalent across many crop species is that domestication traits are not distributed randomly throughout the genome, but are often found clustered together in specific genomic regions (Burger et al. 2008; Ross-Ibarra 2005). Indeed we identified two such clustering regions. Similarly, Zhang et al. (2007) found several clusters of QTL for shelf life, leaf biophysical, and developmental traits for lettuce. The two potential explanations are that these clustering patterns result from pleiotropic effects of a single gene affecting different traits at once or from physical linkage among multiple genes that were individually selected, or have been integrated in the crop genome due to linkage drag (Burger et al. 2008; Ross-Ibarra 2005).

A QTL spanning 10 cM may contain hundreds of genes (Collard et al. 2005; Mauricio 2001), making distinguishing between both possibilities very difficult and as yet there is no information on the identity of (candidate) genes in these regions. For the different traits in the cluster at LG3 a pleiotropic explanation is not obvious. However, for the second cluster (LG7) the majority of traits seem involved in speed of development, such as height at 30 days and days to first flower. We consider it plausible that at LG7 the clustering of these traits, together with those for leaf shape, branching, and seed output, is caused by pleiotropic effects from a common major gene for earliness of flowering. Similar results have been found in dry bean (Koinange et al. 1996), where a gene for earliness also affected branching patterns, number of

days from flowering to fruiting, and pod number. As a more recent example, in *Arabidopsis* a major flowering gene was found to be involved in germination as well (Chiang et al. 2009).

### *Direction of effects*

QTL research on crop species has revealed that in many instances alleles show opposing effects, meaning a crop allele contributing to a wild phenotype or visa versa (Ross-Ibarra 2005). Our results show that all allelic effects of major QTL were in the expected direction. However, our results show for many QTL with intermediate effects traits were increased by crop and wild alleles alike; such opposing effects have been found in previous lettuce research as well (Argyris et al. 2005; Zhang et al. 2007). We found that crop alleles, for example, induced higher germination rates, and more branches and hence, capitula that might be beneficial in the wild habitat at least under some environmental circumstances (Baack et al. 2008; Kaga et al. 2008).

The genetic architecture of lettuce thus concurs mainly with the main domestication trends observed in other species. The presence of many QTL of intermediate effect in combination with the appearance of opposing effects suggests that domestication of lettuce went readily and could reflect multiple domestication events, as suggested for sunflower (Burke et al. 2002). However, the existence of alleles with opposing effects seems to be a general pattern even for crops that are believed to be domesticated through single selection events (Ross-Ibarra 2005). Gross and Olsen (2010) showed that there are multiple pathways to make domesticated plants and that the inferences made from genetic evidence and one cross is not always straightforward. First, QTL results can be very cultivar specific (Mercer et al. 2006). The formation of a tight dense head and delay in bolting are the result of modern breeding efforts (de Vries 1997), and *L. sativa* cv. Salinas is one of the latest bolting and flowering ones (Simonne et al. 2002). Second, these results do need to be verified in the field, as results for controlled greenhouse and environmentally variable field conditions may differ due to Genotype by Environment interactions (Hails and Morley 2005; Weinig et al. 2002). Moreover, hybrid fitness and selective forces may vary across different environments and in different seasons (Chapman and Burke 2006). Further research on the domestication of lettuce is needed to be able to further disentangle the genetic basis and historic selection pressures.

### **Implications for crop breeding**

The enormous increase in genetic resources, such as genetic maps and markers, is only expected to continue as more and more species are being sequenced (Collard and Mackill 2008). With the introduction of new molecular techniques and analysis tools, crop breeding has entered a new phase and rapid progress is being made in the development of new transgenic cultivars (Vaughan et al. 2007) that have higher yields, increased resistance to diseases and herbicides, or increased tolerance to abiotic stresses, such as salt or drought (Campos et al. 2004; Cuartero et al. 2006; Warwick et al. 2008). Current lettuce breeding is strongly focused on utilization and exploitation of wild lettuce relatives (Lebeda et al. 2009); however, transgenic research is also ongoing (Park et al. 2005). To our knowledge, there are no genetically modified lettuce cultivars in production at the moment. Our results indicate that when they are produced the likelihood of escape will depend on the location of insertion (Stewart et al. 2003).

In a potential mitigation strategy, a transgene is placed in close linkage with a gene or genomic block that causes a lower fitness in the wild habitat and is therefore more likely to be purged from the wild populations (Gressel 1999; Stewart et al. 2003). Techniques for targeted insertion of transgenes to specific regions in the genome are currently being developed (Lombardo et al. 2011; Nandy and Srivastava 2011; Shukla et al. 2009). We tentatively indicate

## Chapter 2

the clustering regions on LG3 and LG7 as possible insertion sites. At these sites, crop alleles invariably contributed to e.g. lower amount of basal side shoots and lower seed output, leading to lower hybrid fitness and chances of introgression. Moreover, if the clustering at LG7 is in fact the result of a major gene for earliness, crop alleles will induce a delay in bolting which might be deleterious in the wild habitat. QTL research, such as ours, will prove as a valuable tool to already indicate regions in the crop genome less likely to introgress before any transgene is inserted (Kwit et al. 2011; Stewart et al. 2003).

### **Acknowledgements**

We are grateful to the laboratory of R. W. Michelmore (UC Davis) for graciously providing seeds from their RIL collection to us and for discussions. We like to thank the assistants and technical staff involved especially Rob Bregman, Louis Lie, Ludek Tikovsly, Thijs Hendrix, and Harold Lemereis. This study is funded by the Netherlands Organization for Scientific Research (NWO) as part of the ERGO program (838.06.042).