Novel Genetic Resources for Plant Morphology Study

Preliminary Morphology Study of *Arabidopsis thaliana*

Populations Generated by Reverse Breeding

Mengfan Li

Novel Genetic Resources for Plant Morphology Study

Preliminary Morphology Study of *Arabidopsis thaliana* Populations Generated by Reverse Breeding

by

Mengfan Li

OCTOBER 2014

Wageningen, The Netherlands

Program: MSc Plant Sciences

Thesis MSc Plant Sciences (GEN-80436)

Specialisation: Plant Breeding and Genetic Resources

Student Registration Number: 910717515110

Supervisor: Cris Wijnen

Examiners: Joost Keurentjes, Hans de Jong

Chair Group: Laboratory of Genetics, Plant Sciences Group, Wageningen UR

Acknowledgements

Firstly I would like to thank Cris Wijnen, my thesis supervisor, whom has been guiding me a lot in scientific researching and writing with great patience during the past six months. I was encouraged by his motivation and enthusiasm.

Also I would like to thank Genetics group, especially Prof. Dr. Joost Keurentjes, for allowing me to do my MSc thesis with this topic. To me it was an exciting experience to understand this interesting new breeding technique. It also helped me to get my interest in plant genetic research and to have new plans for my future study and career.

Furthermore, I want to thank all the people who helped me in the experiments. I appreciate the supports of my family and friends in my study and life. Also I would like to thank Prof. Dr. Hans de Jong and Dr. Eric Wijnker, whom have contributed a lot to found the basis of this topic.

Preface

As a person who always loves our nature and sciences, I decided to start my discovery of the nature from plants years ago. While I was studying horticulture technology as my first step, I was amazed by the delicate biological mechanisms inside these quiet creatures. With the willing to study plants, I become an MSc Plant Sciences student in Wageningen UR, specialized in plant breeding and genetic resources.

In one of my MSc courses, *"Genetic Analysis, Tools, and Concepts"*, I learnt about reverse breeding. It's a valuable novel breeding technique. I found it very interesting. Therefore I decided to do my MSc thesis in the reverse breeding project.

Besides the knowledge, I experienced and saw researches in reality during my thesis work. I am often glad to do new things. However, during this study I had to face repetitive and non-progressing works as well. It is like anything else in life, ups and downs are unavoidable. Challenges are great chances leading to growth. Once I found myself getting stuck in, I was upset and then excited like a stop light turning green. I really appreciate the challenges I have met and the supports from people around me.

I hope that this study may form a little step to initiate the future development of the topic, just like this experience forms a step for my future study.

Abstract

Understanding leaf development, inflorescence development, and silique development are important to agricultural economy. Different genetic resources and mapping methods were developed to analyse these traits. In the present study, plant morphological traits were studied in an *Arabidopsis thaliana* double haploid (DH) population and chromosome substitution lines (CSLs) generated by reverse breeding. Also steps were made to develop new CSLs in several accessions.

Leaves from each line of the DH population were taken and analysed for multiple traits related to leaf morphology by using the software ImageJ. It was found that leaf compactness, which was a good proxy to quantify leaf serrations in previous studies, is not the best one to be used in a population that varies in its ratio of blade length : width. The relationship between perimeter and area could be further studies and employed to quantify the variation of leaf shape in one population.

The length of the main inflorescence stem and siliques of chromosome substitution library between *Col-0* and *Ler* were measured. To identify the main effects at chromosome level, student t-test was used for analysis. Chromosomes 2, 3 and 5 were found to be the major chromosomes contributing to the variation of main inflorescence stem and silique development in chromosome substitution library Col-Ler.

Following the protocol of hybrid recreation by reverse breeding, nine accessions were transformed by *RNAi:DMC1* construct to create achiamstic plants. Three achiasmatic WS-2 plants were found in previously transformed plants.

Summary

The main text of this report consists of introduction, material & methods, results, and discussions. The contents can be divided by two research goals: 1) to obtain new CSLs in different accessions by implementing reverse breeding techniques; 2) to study the morphology of reverse breeding generated DHs and CSLs. To assist reading, the works related to the first research goal are divided according to the order of reverse breeding techniques in chapter 2 and 3. The second research goal is divided into two sections, namely "Leaf morphology study of DHs" and "Morphology study of CSLs" in chapter 2, 3, and 4.

This study presents two populations in *Arabidopsis thaliana*, double haploids (DHs) and chromosome substitutions lines (CSLs), generated by reverse breeding. It demonstrates how these populations can be utilized for morphology study. Besides, a start was made to develop new populations, as several accessions were transformed with *RNAi:DMC1* construct and previously transformed plants were selected. Three new achiasmatic plants of WS-2 were found. The selection efficiency of CSLs largely decreased when seedlings were kept in the climate chamber for a longer period before selection.

In the DH population it was tried to quantify leaf morphology using the software ImageJ. Leaf compactness, which was a good proxy in previous studies, is not the best one to quantify leaf blade serrations in a population that varies in its ratio of blade length : width. The relationship between perimeters and areas of a group of similarly shaped fractal sets could be further studied and employed in DHs to quantify the variation of leaf morphology in one population.

By using CSLs, major chromosomes that contribute to the variation of morphology traits can be identified. Using a simple t-test, chromosomes 2, 3 and 5 were identified as the major contributors to variation of main inflorescence stem and silique development in the chromosome substitution library *Col-Ler*. It gives an overview of the genetic interactions on complete chromosome level.

Contents

1. Introduction

Hybrid progeny is commonly observed to have enhanced performance on many traits compared to their homozygous parental lines. However, elite hybrid lines cannot be maintained by propagation through seeds, due to chromosomal crossover and trait segregation in the F2 population. This is especially a barrier for uncharacterized elite heterozygotes lines.

Confronting the challenge of preserving heterozygous genotypes by seeds, reverse breeding was introduced as a novel plant breeding technique [1]. It was proven that reverse breeding could be used to construct homozygous complementing parental lines that can recreate the vigorous hybrid by mating [2]. Reverse breeding essentially consists of suppression of crossover recombination in the selected plant and subsequent regeneration of doubled haploids (DHs) from achiasmatic (non-crossover) gametes [3]. Various applications of reverse breeding, such as breeding on single chromosome level and reconstruction of heterozygous germplasm, entirely rely on the technical realization of these two essential steps [1].

Crossover recombination in F1 hybrid of two natural *A. thaliana* accessions can be suppressed by the RNAi construct targeting the meiotic recombinase *DISRUPTED MEIOTIC cDNA 1* (*DMC1*) [3]. To achieve intact-chromosome inheritance without crossovers, at least one parental line should contain the dominant negative *RNAi:DMC1* construct [3]. During meiosis, chiasmata formation leads to proper segregation, while the segregation is commonly unbalanced without chiasmata. Some viable gametes may also exist in the products of the unbalanced segregation at frequencies depending on the chromosome number [1].

The transgenic *cenh3 GFP-tailswap* rescue line, which is in the *Col-0* background, is crossed to the selected achiasmatic parent to generate *Arabidopsis* haploids through whole-genome elimination [4]. Altered centromeric histone 3 (CENH3) proteins cause the elimination of the genome from the embryo after fertilization, resulting in haploids only contains genome of the achiasmatic parent [4]. Chromosomes will be doubled spontaneously in the next generation [3].

Arabidopsis thaliana is the only plant species used for reverse breeding research at this moment, due to its small number of chromosome pairs (n=5) and the abundant genomic information since it became a model plant [5]. In theory the probability to get balanced gametes in *A. thaliana* is 1 in 32, while that of maize (10 chromosome pairs) is 1 in 1024 [1].

Non-recombinant meiotic segregation can be used to generate chromosome substitution lines (CSLs). Chromosome substitution libraries (CSLib) are sets of lines in which one or more chromosomes are substituted by the corresponding homolog from another genotype [3]. These immortal homozygous lines can be preserved by seeds and paired to recreate the hybrid [2].

The haploid inducer, *cenh3 GFP-tailswap* can be used to create double haploids (DHs) in merely three generations, while recombinant inbred lines (RILs), having similar features as DHs, would need eight generations [6]. This short generation time is an advantage over RILs to enable the development of DH populations from new accessions.

These newly generated populations are to serve studies of *A. thaliana* in different aspects. Understanding leaf development, inflorescence development, and silique development are important to agricultural economy. This research was to study plant morphology of DHs and CSLs.

By studying leaf morphology, this report displays method and parameters to quantify leaf morphology in DHs. As one example, leaf margin protrusions, termed serrations, is a key feature of leaf shape [7]. Parameters such as dissection index and leaf compactness were used to quantify and distinguish leaf serrations at different level in mutants and transgenic lines in previous studies [7, 8]. Through this study, it was expected to find out if the parameters used for mutant and T-DNA lines could also be used to study leaf morphology of DHs.

Development of inflorescence stems and siliques are major reproductive breeding traits. This research exhibits the new genetic resources, CSLs, for morphology study. The length of the main inflorescence stem and the length of siliques are taken as examples to demonstrate the method of utilizing this genetic resource. Using this resource it was expected to identify main effects and possible epistatic interactions of complex traits at complete chromosome level [6].

2. Methods and Materials

Transformation with *RNAi:DMC1* construct

Nine natural accessions of *Arabidopsis thaliana* (Table 1) were transformed with the *RNAi:DMC1* construct. At least ten seeds of each accession were sown on a 5-cm filter paper per Petri dish with sterilized demi water. First, seeds were stratified in dark at 4°C for 4 days and afterwards in the climate chamber for 2-3 days, with a 16:8 h photoperiod under 24°C. The seedlings were transferred to greenhouse conditions and five seedlings per accession were transplanted onto rockwool to guarantee enough available plants for transformation.

Table 1 Nine *Arabidopsis thaliana* accessions used for transformation

Name	Accession
Cvi-0	CS76116
Bur-0	CS76105
Mib-22	CS76182
Tsu-0	CS28780
Rrs-7	CS28713
Est-1	CS76127
C24	CS76106
Col-0	N60000
Col-0	186AV1B3

Transformation with the RNAi:*DMC1* construct was performed on flowers 49 days after transplanting onto rockwool (39 days for CS60000 and 186AV1B3), using floral dip [9]. *A.tumefaciens* strain *C58* harbouring *pRZ051* (Appendix A) was previously made and ready for use for infection of Arabidopsis flowers. Main stems of plants were cut off about 2-3 weeks before infection to induce side branching and thus increase the number of flowers for transformation. To obtain higher transformation efficiency, siliques were trimmed off the day of and before the infection. Detailed protocol of bacteria preparation and floral dipping can be found in Appendix B. After growth till maturation in the greenhouse seeds from transformed plants were collected, labelled, and stored for future use.

Selection of achiasmatic plants

Nine *A. thaliana* lines were selected for achiasmatic plants (Table 2). To sterilize seeds, 1 ml of HCL was added into 50 ml of bleach to produce chlorine gas in a closed desiccator inside a fume hood for 2 hours. About 50-100 sterilized seeds were sown on each squared Petri dish containing 50 mL of sterile ½ MS medium (pH 5.8) consisting of 0.22%(w:v) MS with vitamin, 1%(w:v) sucrose, Kanamycin,

dissolved in demi water and solidified with 1% (w:v) Daishin agar. To prevent the antibiotic from breaking down, 1 mL of 100 mg/mL Kanamycin stock solution was added when the medium cooled down below 60°C after autoclaving, obtaining an end concentration of 100 µg/mL in the medium.

Table 2 Previously transformed Arabidopsis thaliana accessions that were selected for achiasmatic plants

Name	Accession
WS-2	CS28827
TNZ-1	CS77369
Lov-5	CS76175
UKNW06-460	CS76279
Tamm-2	CS76244
Ler[a]	CS76164
Ler[b]	CS76164
Col-0[a]	186AV1S1
Col-0[b]	186AV1S1

[a] Transformed by strain LBA4404

[b] Transformed by strain C58

Seeds on ½ MS plates were stratified in the dark at 4°C for 3 days and then kept in a climate chamber for 3 days under 24°C, with a 16:8 h photoperiod. After one week of growth, Kanamycin-resistant seedlings were selected and transplanted on rockwool in the greenhouse. Yellowish seedlings that soon died afterwards in the plates were defined as Kanamycin susceptible.

After 15-day growth on rockwool, leaf samples were taken from the selected plants and sent to RijkZwaan for genotyping and verification of the presence of the construct. The markers used for genotyping and confirming the construct can be found in Appendix C. Genotypes of the samples were determined by comparing the results with the genotypes of the parental accessions. Plants having the parental genotype and a semi-sterile phenotype [2] were considered as achiasmatic plants, and thus harvested and stored for later use.

Identification of the location of the construct

To discover the insertion sites of the RNAi construct in the confirmed transformants, thermal asymmetric interlaced polymerase chain reaction (TAIL-PCR) was used to amplify the flanking sequences and for subsequent discovery of the location of the construct via homology searches (i.e. Basic Local Alignment Search Tool [BLAST]) [10].

Before TAIL-PCR, genomic DNA from leaves of achiasmatic plants was prepared to serve as template in the PCRs (Appendix D). Two sets of primers were used for TAIL-PCR [10]. One set of primers consists of three specific primers complementary to the insertion sequence (Table 3). The other primer set consists of arbitrary degenerate primers (AD primers), of which the sequence is shown in Table 4. AD primers were designed to hybridize to many sites randomly in the genome, including the unknown flanking sequences. According to the suggestion of previous experiments [10], the concentration of each AD primer set (12 μM) in the PCR reaction were higher than that of the insertion-specific primers (5 uM) to ensure the sufficient random priming throughout the genome.

Table 3 The specific reverse primers used for TAIL-PCR

Primer	Sequence (5'->3')	Start	Stop	Tm*	GC%	SC*	SC 3*
DMC-1	ACAATTCCACACAACATACGAGCCG	116	92	58.12	48	4	2
DMC-2	AGCCGGAAGCATAAAGTGTAAAGCC	96	72	57.86	48	4	1
DMC-3	TCCAGTCGGGAAACCTGTCG	115	96	55.99	60	5	2

***Tm=melting temperature; SC=self-complementarity; SC 3= Self 3' complementarity

Table 4 AD Primer sets that were used in TAIL-PCR

Primer	Primer sequence (5'–3')	Length	Degeneracy	Average Tm*	Average GC%
AD1	NGTCGASWGANAWGAA	16 bp	128-fold	46.6°C	43.8
AD2	TGWGNAGSANCASAGA	16 bp	128-fold	49.2°C	50
AD3	AGWGNAGWANCAWAGG	16 bp	128-fold	46.6°C	43.8

W = A or T, S = G or C, N = A or T or G or C.

*Average Tm=Average melting temperature

In order to amplify the specific template, three rounds of TAIL-PCR were carried out in the experiments using three different insertion-specific primers [10]. The location of insertion-specific primers in the *RNAi:DMC1* construct is shown in Figure 1.

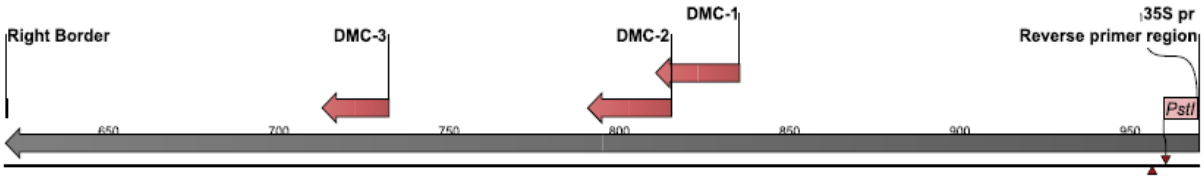

Figure 1 The insertion-specific primers (DMC-1, -2, -3) that will be used in TAIL-PCR to amplify the flanking sequences of the insert. Length of DMC-1 and DMC-2 is 25 bp and that of DMC-3 is 20bp. DMC-1 and DMC-2 have 5 bp overlap.

In the first round, a pool of three AD primer sets was used together with insertion-specific primer DMC-1. Three technical replicates were used for running the primary PCR. For the secondary TAIL-PCR, three replicates of each PCR product from the first round were made, except only one replicate was made for AD2 primer. In the second round, primer DMC-2 was used with the three separated AD primer sets. After the second round, gel electrophoresis was performed for the PCR products. According to the pattern of the agarose, the samples with the corresponding AD primers that got clear bands were to be continued for the tertiary TAIL-PCR, together with primer DMC-3. One replicate was used. Gel electrophoresis was run after the third TAIL-PCR. Visible PCR products on the agarose gels were purified and sent to GATC for sequencing together with the insertion-specific primer DMC-3.

Using BLAST, the location of the construct can discovered in the future. The PCR programs of TAIL-PCR were following the methods of Singer and Burke [10]. Concentrations of primers and chemicals used for each PCR round are listed in Appendix E. Purification of the PCR products in the bands were done by using the QIAquick® Gel Extraction Kit (50) (gel purification protocol can be found at http://sites.bio.indiana.edu/~chenlab/protocol_files/agarose_gel_extraction.pdf). The protocol of gel electrophoresis is in Appendix F.

Selection of CSLs

Formerly two achiasmatic *Col-0* lines containing *RNAi:DMC1* construct on different chromosomes (as females) were crossed to other accessions respectively to obtain achiasmatic hybrids. Afterwards the hybrids were crossed to *cenh3 GFP-tailswap* (as female) to generate haploids. Selections were performed in the offspring listed in Table 5 for desired haploid CSLs.

Table 5 Offspring of crosses between achiasmatic *Col-0* and different accesstions, which were selected for haploid CSLs

Plant Line	Name	Accession	Maternal Col_0 *
C4_fei_C3	Fei-0	CS76129	C4
C4_bor_C3	Bor-4	CS76100	C4
C4_bur-0_C3	Bur-0	CS76105	C4
C5_bor_C3	Bor-4	CS76100	C5
C5_bur-0_C3	Bur-0	CS76105	C5

*C4 and C5 stands for achiasmatic Col-0 with the *RNAi:DMC1* construct respectively on chromosome 4, and chromosome 5.

About 50 seeds were sown on each squared Petri dish containing 50 mL of sterile ½ MS medium (pH 5.8) consisting of 0.22%(w:v) MS with vitamin dissolved in demi water and solidified with 1% (w:v) Daishin agar. Stratification of seeds on plates took 3 days in the dark under 4°C. Then the plates were kept in a climate chamber for 3 days under 24°C, with a 16:8 h photoperiod. About half of the seeds from plant line C4_fei_C3 stayed in the climate chamber for 3 days, and the other half stayed in the climate chamber for 2 days.

After germination in the climate chamber, seedlings were selected phenotypically. Seedlings having very short roots and hanging cotyledons were considered to be selfings of *cenh3 GFP-tailswap* and thus discarded [3]. Only seedlings with relatively long roots and raised cotyledons were picked out and transplanted onto rockwool in the greenhouse.

Three to six weeks later, leaf samples were taken and sent to RijkZwaan for genotyping and verification of the presence of the *RNAi:DMC1* construct. As mentioned before, 39 markers were used for genotyping (Appendix C). Genotype of the samples were later determined by comparing the results to the genotype of the parental accessions. Plants with intact substituted chromosomes of the corresponding parental accessions and the semi-sterile phenotype [3] were considered as chromosome substitution haploids, and thus harvested and stored.

Leaf morphology study of DHs

Recombinant hybrids between Br-0 (cs76101) and Ty-0 (cs28786) were crossed to the haploid inducer line resulting in recombinant haploid progeny. In the next generation double haploids (DHs) were obtained through spontaneous chromosomal doubling. The DH population was firstly sown on filter paper in petri dishes with 3 replicates per line, stratified in the dark under 4°C for 3 days, and then kept in the climate chamber for 3 days under 24°C, with a 16:8 h photoperiod. After that seedlings were transplanted to greenhouse. Six weeks later, vernalization of these plants took place in a cold growing chamber (4°C) for eight weeks. Then the plants were taken back to the greenhouse for flowering and harvesting.

75 DH lines were taken for leaf morphology study. The largest rosette leaf by eye was taken from one of the three replicates of each line after flowering in the greenhouse. Fresh leaves were cut off at the very start of the leaf petiole. Wrinkles on the leaves were carefully unfolded and flattened by sandwiching the leaves between two pieces of thin transparent plastic cards. Images were taken using a

charge coupled device camera at 300 DPI as colour JPG file for each leaf with a ruler as scale.

Leaf shape parameters were acquired by using the plug-in LeafJ in software ImageJ [11]. Direct output from LeafJ are petiole length, blade length, blade width, blade area, and blade perimeter. Parameters such as leaf compactness (blade perimeter^2 : blade area), leaf L:W ([petiole length + blade length]/blade width) and blade L:W ratio (blade length/blade width) were calculated [8]. The distributions of these traits were ploted. The pairwise correlation between these traits were calculated. To plot blade area and blade perimeter, values of these two traits were log transformed.

Morphology study of CSLs

The seeds was firstly sown on filter paper in petri dishes with 10 replicates per line, stratified in the dark under 4°C for 3 days, and then kept in the climate chamber for 3 days. Subsequently, seedlings were transferred to climate chamber (16 hours of 125 µmol light under 20°C, 8 hours of darkness under 18°C, constant 70% relative humidity). For morphology study, one plant was taken randomly among the replicates that having similar phenotype in each line.

From the chromosome substitution library *Col x Ler* (CSLib, 32 lines containing substitution between chromosomes of *Col-0* and *Ler*), 32 lines were planted and 27 lines were used (Table 6), due to imperfection of this CSLib. All of the 32 lines have cytoplasm of Col-0. The used ones were confirmed to be the right genotype and phenotype. The ignored lines were CSLine 9, 17, 25, 27, and 29. The reference lines were CSLine 1 (parental *Col-0*) and CSLine 32 (having cytoplasm from *Col-0* and nuclear genomes from *Ler*).

Depending on the plant line, the first 9 to 25 rosette leaves of these plants were cut off 1 to 4 days after flowering. Images were acquired by the method described for the DH population. When most of the flowers on the plants were withered (3 to 5 weeks after flowering), the length of the main inflorescence stem (MSTL) and the length from the node of the highest vegetative side branch to the top of the main stem (B-TL) were measured with a ruler.

The 5th to 9th silique on the main stem were cut off at the start from the stem and taken images. For the lines that did not have mature siliques start from the 5th silique, siliques higher than the 9th were taken to insure that at least 2 mature siliques were obtained. Measurements of silique length were done in ImageJ. The

average length of the 5th to 9th silique (5-9SL) and silique petiole (5-9SPL), the obtained mature siliques (MSL) and silique petiole (MSPL) were estimated.

The mean differences between lines with chromosomes from *Col-0* and *Ler* were calculated. Statistical differences of means were analysed with the Student's t-test with 2 classes of significance: $p<0.05$ and $p<0.001$ (95%confidence interval, two-sided).

Table 6 Genotypes of the lines of CSLib Col-Ler. Red and blue cells represent chromosomes from *Col-0* and *Ler* respectively.

CSLine	Chr.1	Chr.2	Chr.3	Chr.4	Chr.5
Col-0	Col-0	Col-0	Col-0	Col-0	Col-0
2	Col-0	Col-0	Col-0	Col-0	Ler
3	Col-0	Col-0	Col-0	Ler	Col-0
4	Col-0	Col-0	Col-0	Ler	Ler
5	Col-0	Col-0	Ler	Col-0	Col-0
6	Col-0	Col-0	Ler	Col-0	Ler
7	Col-0	Col-0	Ler	Ler	Col-0
8	Col-0	Col-0	Ler	Ler	Ler
9*	Col-0	Ler	Col-0	Col-0	Col-0
10	Col-0	Ler	Col-0	Col-0	Ler
11	Col-0	Ler	Col-0	Ler	Col-0
12	Col-0	Ler	Col-0	Ler	Ler
13	Col-0	Ler	Ler	Col-0	Col-0
14	Col-0	Ler	Ler	Col-0	Ler
15	Col-0	Ler	Ler	Ler	Col-0
16	Col-0	Ler	Ler	Ler	Ler
17*	Ler	Col-0	Col-0	Col-0	Col-0
18	Ler	Col-0	Col-0	Col-0	Ler
19	Ler	Col-0	Col-0	Ler	Col-0
20	Ler	Col-0	Col-0	Ler	Ler
21	Ler	Col-0	Ler	Col-0	Col-0
22	Ler	Col-0	Ler	Col-0	Ler
23	Ler	Col-0	Ler	Ler	Col-0
24	Ler	Col-0	Ler	Ler	Ler
25*	Ler	Ler	Col-0	Col-0	Col-0
26	Ler	Ler	Col-0	Col-0	Ler
27*	Ler	Ler	Col-0	Ler	Col-0
28	Ler	Ler	Col-0	Ler	Ler
29*	Ler	Ler	Ler	Col-0	Col-0
30	Ler	Ler	Ler	Col-0	Ler
31	Ler	Ler	Ler	Ler	Col-0
32	Ler	Ler	Ler	Ler	Ler

*Except lines with stars (CSLine 9, 17, 25, 27, 29), the rest of the CSLib was used for phenotype.

3. Results

Selection of achiasmatic plants

At least two Kanamycin resistant seedlings of each transformed accession were found and grown in the greenhouse. However, only three out of nine resistant plants showed a semi-sterile phenotype. Genotypes of the three plants were confirmed to be WS-2 and contained the *RNAi:DMC1* construct according to data returned by Rijk-Zwaan.

Identification of the location of the construct

TAIL-PCR was performed on one of the three selected achiasmatic plants. After the secondary TAIL-PCR, obvious bands were observed on agarose gel for two samples (Figure 2), both with AD3 primers. These PCR products were diluted and amplified in the third PCR round with AD3 and DMC-3. Two bands appeared on the agarose gel after the third PCR (Figure 3). According to the figure, one band was about 1000 bp, the other one was about 600 bp. These two products were amplified again through tertiary TAIL-PCR, subsequently purified separately from agarose gel and sent for sequencing. The sequence can be analysed and used to identify the location of the contract later.

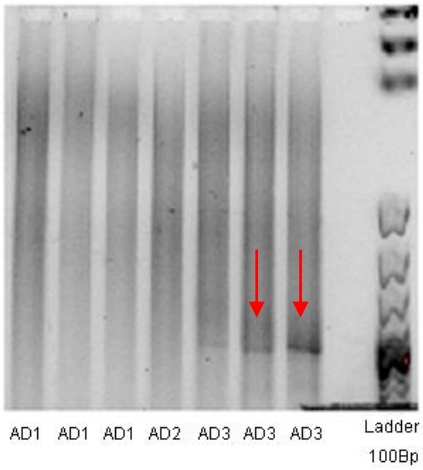

Figure 2 Agarose gel image of the secondary PCR products. The primers used were insertion-specific primer DMC-2 and AD primers labelled at the bottom of each sample. Red arrows indicate the bands. The ladder for the upper rows ran into the lower part, forming the three bands above the lower ladder.

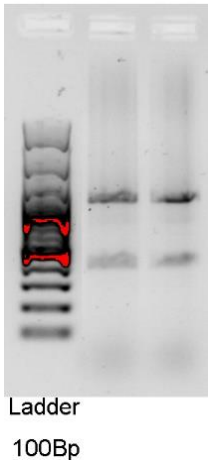

Figure 3 Agarose gel image of the tertiary PCR products. The primers used were insertion-specific primer DMC-3 and AD 3 primers.

Selection of CSLs

There were 96 seedlings of C4_fei_C3 picked out from the petri dishes. These seedlings were kept in the climate chamber for 3 days before selection. Among the survived plants, a few had the phenotype of aneuploids and the rest were semi-sterile. Leaf samples of 42 plants were collected for genotyping. One sample had the same genotype with *Fei-0* on chromosome 3 and with *Col-0* on the other four chromosomes. Thus it was confirmed to be a CSL. The rest of the samples had *Col-0* genotype and semi-sterile phenotype. The selection efficiency of CSL was 1% (= 1 CSL / 96 selected seedlings x 100%).

In the plates kept in the climate chamber for 2 days, totally 22 seedlings were selected for C4_fei_C3 and other populations listed in Table 5. According to the genotyping results, 2 plants from C4_fei_C3 were found to be CSLs, of which one had the same genotype as the CSL selected previously. Only one sample was found to be *cenh3 GFP-tailswap*. The rest plants were DHs and heterozygotes. The selection efficiency was 9%.

Leaf morphology study of DHs

The leaves taken from the DHs were serrated at different levels. The continuous distribution pattern of the petiole length, blade length, blade width, blade area, and blade perimeter, leaf compactness, blade length: blade width ratio, and leaf full length: width ratio, are shown in Figure 4.

Pairwise correlation of the traits are shown in Figure 5. The correlation between BP and BA was the highest among all the pairs. BA was highly correlated with BL and BW (r=0.85, r=0.91 respectively). BP was highly correlated with BL and BW (r=0.91, r=0.78 respectively). However the correlation between BW and BL was lower (r=0.58). LC was correlated with BL:W (r=0.57), but not with BA (0.04). There was no significant correlation between PL and any other traits, except L:W with which correlation was expected.

The log-log perimeter-area plot is given in Figure 6. The correlation between Log (BP) and Log (BA) is 0.95.

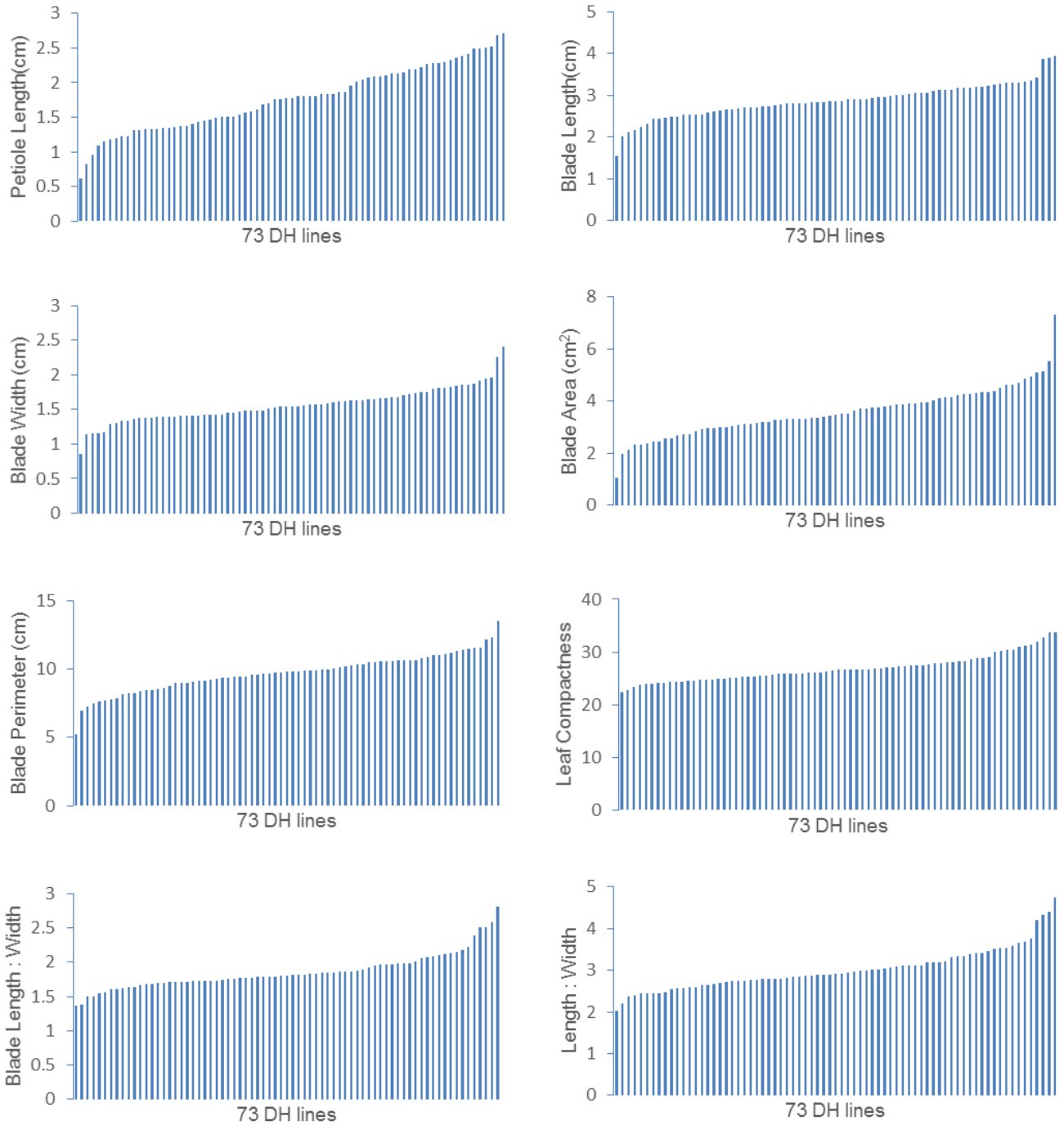

Figure 4 Distribution of petiole length, blade length, blade width, blade area, blade perimeter, leaf compactness, blade length: blade width, and leaf full length: width in the DH population.

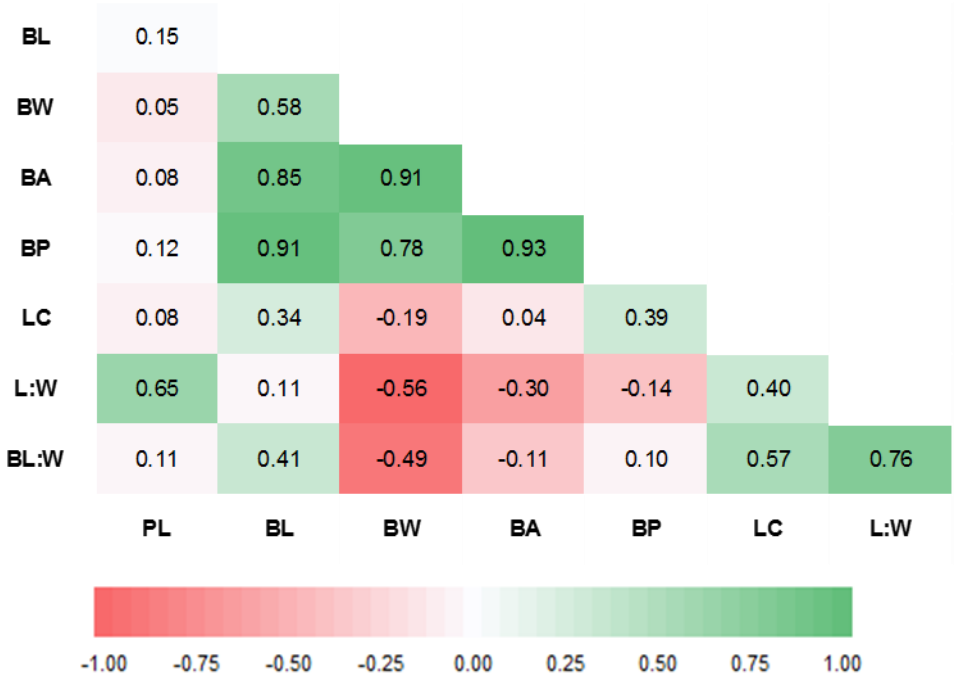

Figure 5 Pairwise correlations between the following traits: blade length (BL), blade width (BW), blade area (BA), blade perimeter (BP), leaf compactness (LC), leaf full length : width (L:W), blade length : blade width (BL:W), petiole length (PL).

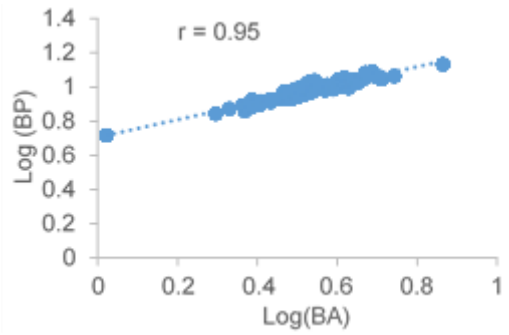

Figure 6 The plot of Log (BP) and Log (BA) of the DH population. The dotted line shows the correlation trend. r = 0.95.

Morphology study of CSLs

The distributions of the average length of the mature siliques (MSL) and silique pedicel (MSPL), and that of the 5th to 9th siliques (5-9SL) and silique pedicel (5-9SPL), are shown in Figure 7. These traits display continuous distribution. There is no significant gap between the performance of *Col-0* and CSLine 32. Among the

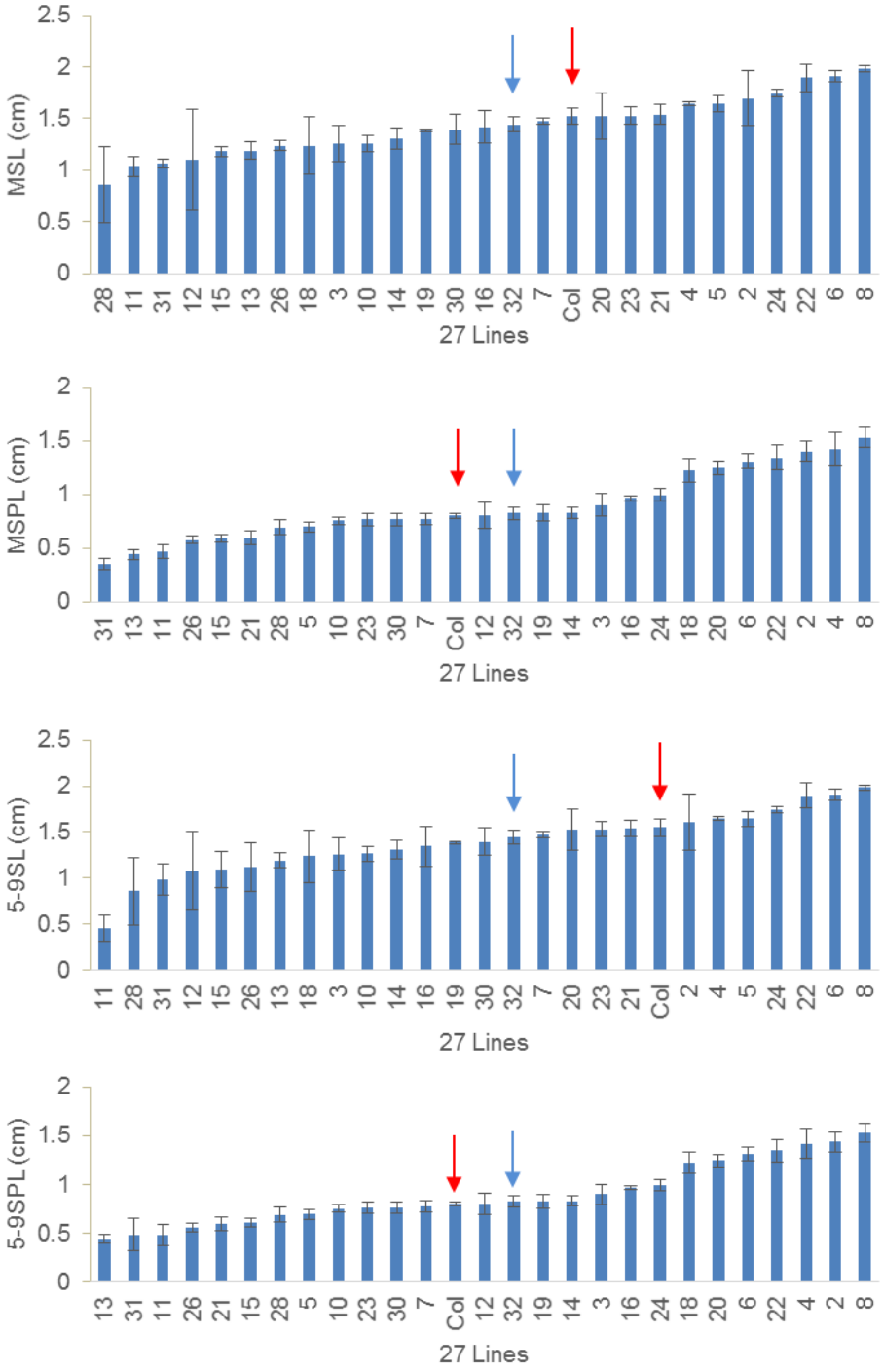

Figure 7 The distribution of MSL, MSPL, 5-9SL and 5-9SPL. The error bars represent standard deviation. The red and blue arrows point out the distribution of line Col-0 and line 32 respectively.

phenotyped 27 lines, CSLine 2, 11, 12, 15, 16, 26, and 31 did not start to produce mature siliques from the 5th silique.

The distributions of the length from the highest side branch on the main stem to the top (B-TL) and the main stem length (MSTL) are shown in Figure 8. These two traits look like discrete traits. *Col-0* is in the higher discrete groups.

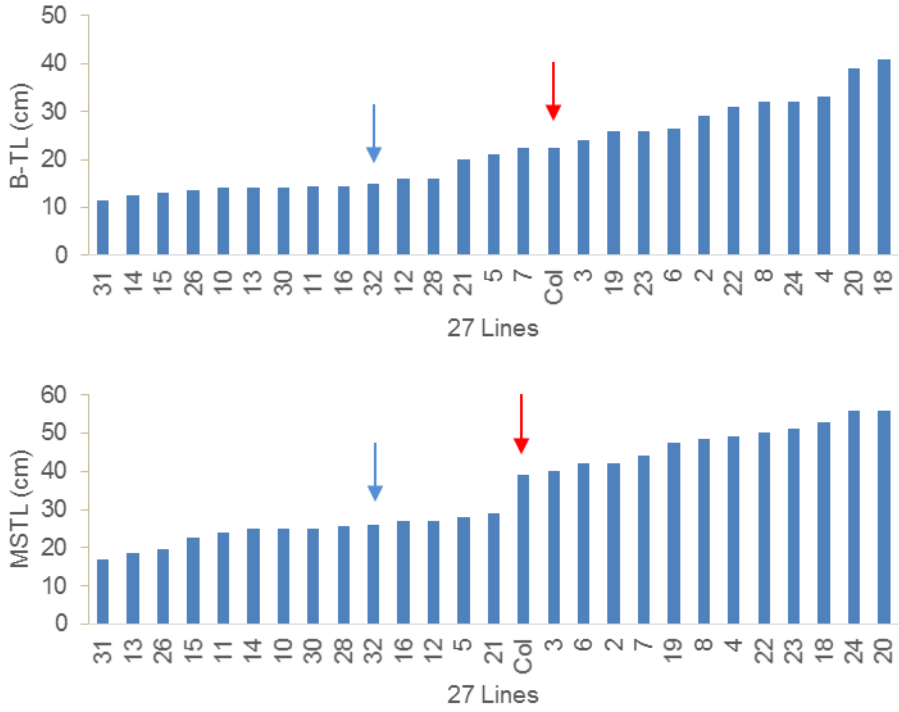

Figure 8 The distribution of B-TL and MSTL. The red and blue arrows point out the distribution of line *Col-0* and CSLine 32 respectively. There was no replicates for these two traits.

The mean differences between lines with chromosomes from *Col-0* and from *Ler* are shown in Table 7. Chromosome 1 was not significantly associated with any of the traits explained here. While chromosome 2 was significantly associated with all. The positive differences indicate that the CSLs with chromosome 2 from *Col-0* have bigger B-TL and MSTL. Chromosome 3 was significantly associated with the length of the siliques. The CSLs with chromosome 3 from *Ler* have longer siliques. The lines with chromosome 5 from *Ler* have longer siliques and pedicels. Chromosome 4 was associated with 5-9SL, while not with MSL.

Table 7 Mean differences between lines having *Col-0* chromosomes and *Ler* chromosomes. Traits: the length of the mature siliques (MSL) and silique pedicel (MSPL), the average length of the 5th to 9th siliques (5-9SL) and silique pedicel (5-9SPL), the length of the main stem (MSTL), the length between the node of the highest vegetative side branch to the top of the main stem (B-TL).

Traits	Chr. 1	Chr. 2	Chr. 3	Chr. 4	Chr. 5
MSL	0.03	0.39**	-0.22**	0.10	-0.13*
MSPL	0.05	0.38**	0.07	0.00	-0.39**
5-9SL	-0.02	0.46**	-0.26**	0.14*	-0.17*
5-9SPL	0.05	0.37**	0.06	-0.00	-0.37**
MSTL	-4.5	21.5**	3.3	-4.4	-4.5
B-TL	-3.2	14.3**	3.7	-0.8	-4.1

The difference with the levels of significance are symbolized: *p < 0.05, **p < 0.001.

4. Discussions

Reverse breeding techniques

Among previously transformed plants, three achiasmatic plants, 2 genotypes, of *WS-2* were found and thus these can be used to create new CSLib. If the *RNAi:DMC1* construct is identified to be on different chromosomes, construct-free CSLs can be made. The newly transformed accessions (Table 1) can be selected for kanamycin resistance in the future to enrich the CSLib.

The selection efficiency of CSLs largely decreased when the seedlings stayed in the climate chamber for a longer period before selection. Besides the correct CSLs, the rest of the selected seedlings were selfings of *cenh3 GFP-tailswap*, DHs, and heterozygotes. This was mainly because the phenotypes of seedlings became less discriminative. The most obvious phenotype of a *cenh3 GFP-tailswap* seedling was short roots when it just germinated. As the seedling grows, this character became much less obvious among CSL, DH, and heterozygotes.

Leaf morphology study of DHs

The DH population was phenotyped but not genotyped yet. By studying the phenotype, it displays a resource to study the explanation of the observations on cellular and molecular level in the future.

The continuous distribution pattern of phenotypes indicates a quantitative inheritance pattern for these morphological traits. A few lines were extreme (either high or low) in petiole length, blade length, blade width, blade area and blade perimeter, which may represent QTLs that have strong genetic effects.

Blade area and blade perimeter were highly correlated with blade width and blade length respectively (Figure 5). It simply shows that blade expansion on lateral and longitudinal directions are the main factors determining blade area and perimeter. Thus the growth of protrusions at leaf margin, as another factor influencing blade area and perimeter, was not as variable as blade directional development in this DH population.

The lower correlation between blade width and blade length may indicate that common genes as well as independent genes maybe controlling these two parameters. A proof of this could be that the expansion rate of blade in lateral and longitudinal directions were different at early stage, but become equal at later

developmental stage [12]. The mechanism behind it makes blade area a complex trait for morphology study.

Leaf petiole length was not obviously correlated with blade length and blade width. This result is consistent with earlier report [12] that in low light condition petiole length largely increased, while blade L:W followed the same pattern as in high light condition. Studies of photoreceptor mutants indicate that PHYB and CRY1 promote leaf blade expansion and inhibit petiole elongation in red light and blue light, respectively [13]. The contrasting reaction of petiole and blade L:W to environment may suggest that different genes are controlling the growth of petiole and blade.

The very low correlation between LC and BA (0.04) confirms that LC is scale independent [8]. However, the correlation between LC and BL:W is much higher (0.57). It suggests that a narrower leaf has a higher LC, which means the narrower leaf is less compacted. The use of a few mutant and T-DNA *A. thaliana* lines with representative blade shape in the earlier studies was successful [7, 8]. Leaf compactness could describe leaf blade serration in a population with constant length : width ratio. But it may become biased to describe leaf serrations when the leaves have successive and variable blade L:W in one population.

The perimeters and areas of a group of similarly shaped fractal sets are related to one another by power-law relationships. And that means the log perimeter and log area are correlated to one another linearly in these sets. It has been employed in many fields such as physics and geology studies [14]. Here in the present study, this relationship was utilized to quantify and to indicate whether there is variation of leaf shape in the DH population. The correlation between Log(BP) and Log(BA) deviated from 1. It means that there is variation of blade shape in this population. However the deviation was not very big, and thus the population was not highly variable in leaf shape. The variation may consist the variation of leaf L:W and levels of serrations mainly. The application of the relationship between perimeter and area could be further studied in order to employ it in leaf morphology study.

Previously the methods to describe leaf serrations and the variation in leaf area characteristics within collections of plants were studied by using software LAMINA [15]. But it could not run on computer operation system Windows 7, as the software was updated no longer. Applicable methods and up-to-date software for description of the leaf serrations and the variation of leaf shape within one population are in need of development in the future.

The major advantage of generating DHs by reverse breeding is that it merely takes three generations to reach 100% homozygosity, while recombinant inbred lines (RILs) take 8 generations. Besides, current RIL populations of *A. thaliana* are made between early flowering plants. Also one of the parental lines are mostly *Col-0*. By using reverse breeding, the short construction time enables building DH populations from late flowering accessions that require extra eight weeks of vernalization. New populations will increase the variation for genetic studies. For future study, one can take leaves in a late flowering DH population before vernalization in order to analyse vegetative traits without being affected by reproductive growth.

Morphology study of CSLs

From CSLs, it is possible to obtain hybrids without *RNAi:DMC1* construct even though the parental lines are transgenic [3]. On one hand the presence of the construct disturbs chromosomal segregation, resulting in low number of viable seeds which causes difficulty to preserve these CSLs, and it influences the phenotype of the transgenic plants. On the other hand transgene-free product through reverse breeding may be more acceptable than transgenic crops, since there are many concerns around transgenic crops on different aspects, for example, environment, human health, and socio-economic impact [16]. Nevertheless, at this moment it is not clear how to position transgene-free products resulting from a transgenic breeding approach.

CSLs have been shown to provide an excellent resource for mapping of reproductive traits. The development of the 5^{th} to 9^{th} siliques among the CSLs was not identical. Seven lines did not start to produce mature siliques from the 5^{th} silique, while most of the lines did. The immature siliques influenced the data, resulting in the association between 5-9SL and chromosome 4.

The continuous distribution of MSL, MSPL, 5-9SL, and 5-9SPL suggests that these traits may have at least two chromosomes as main factors interacting with one another (Figure 7). This was confirmed with a t-test between lines with *Col-0* chromosomes and *Ler* chromosomes (Table 7). Lines having chromosome 2 from *Col-0*, chromosome 3 and 5 from *Ler* would have the longest average mature siliques in this CSLib. And those were CSLine 24, 22, 6, and 8.

However, yield is highly sensitive to environment and controlled by multiple genes and yield traits were found to have relatively high variation among replicates in previous study [17]. Therefore the mature siliques from the bottom of the main

inflorescence stems may represent the silique length better than the 5th to 9th siliques.

As with inflorescence stem length, the discrete distributions of B-TL and MSTL were explained by the results of t-test that only chromosome 2 was significantly involved. Both the distribution and the t-test for main effects suggest this is a simple trait, influenced by a single chromosome. It indicates that the difference of main inflorescence stem length between *Col-0* and *Ler* were mainly due to the difference on chromosome 2. Consequently lines inherit chromosome 2 from *Col-0* were higher than that of *Ler*.

The reason why chromosome 2 was associated with all the traits above may be that the *erecta* gene is present on chromosome 2 of *Ler*, which is a mutation and shows an altered organ shape such as compact inflorescence and short and blunt siliques [18].

The ignored CSLines were found to contain the construct and therefore produce aneuploids. These lines can be included in morphology study after correction. Genetic interactions were observed in this CSLib. Futher crosses between CSLines could be made upon specific chromosome(s). The lines with specific recombinant chromosome(s) can be used to study genetic interactions. Compare to conventional RILs, it provides a less noisy genome background. Also to build a CSLib takes less generations than a RIL and that provides a quick overview of the interactions for further studies.

5. References

1 Dirks, R.*, et al.* (2009) Reverse breeding: A novel breeding approach based on engineered meiosis. *Plant Biotechnology Journal* 7, 837-845

2 Wijnker, E.*, et al.* (2012) Reverse breeding in Arabidopsis thaliana generates homozygous parental lines from a heterozygous plant. *Nature Genetics* 44, 467-470

3 Wijnker, E.*, et al.* (2014) Hybrid recreation by reverse breeding in Arabidopsis thaliana. *Nature Protocols* 9, 761-772

4 Ravi, M. and Chan, S.W.L. (2010) Haploid plants produced by centromere-mediated genome elimination. *Nature* 464, 615-618

5 Koornneef, M. and Meinke, D. (2010) The development of Arabidopsis as a model plant. *Plant Journal* 61, 909-921

6 Wijnen, C.L. and Keurentjes, J.J.B. (2014) Genetic resources for quantitative trait analysis: Novelty and efficiency in design from an Arabidopsis perspective. *Current Opinion in Plant Biology* 18, 103-109

7 Bilsborough, G.D.*, et al.* (2011) Model for the regulation of Arabidopsis thaliana leaf margin development. *Proceedings of the National Academy of Sciences of the United States of America* 108, 3424-3429

8 Backhaus, A.*, et al.* (2010) Leafprocessor: A new leaf phenotyping tool using contour bending energy and shape cluster analysis. *New Phytologist* 187, 251-261

9 Clough, S.J. and Bent, A.F. (1998) Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant Journal* 16, 735-743

10 Singer, T. and Burke, E. (2003) High-throughput TAIL-PCR as a tool to identify DNA flanking insertions. *Methods in molecular biology (Clifton, N.J.)* 236, 241-272

11 Maloof, J.N.*, et al.* (2013) LeafJ: an ImageJ plugin for semi-automated leaf shape measurement. *Journal of visualized experiments : JoVE*

12 Kalve, S.*, et al.* (2014) Three-dimensional patterns of cell division and expansion throughout the development of Arabidopsis thaliana leaves. *Journal of experimental botany*

13 Kozuka, T.*, et al.* (2005) The different growth responses of the Arabidopsis thaliana leaf blade and the petiole during shade avoidance are regulated by photoreceptors and sugar. *Plant and Cell Physiology* 46, 213-223

14 Cheng, Q. (1995) The perimeter-area fractal model and its application to geology. *Mathematical Geology* 27, 69-82

15 Bylesjö, M.*, et al.* (2008) LAMINA: A tool for rapid quantification of leaf size and shape parameters. *BMC Plant Biology* 8

16 Robinson, J. (1999) Ethics and transgenic crops: A review. *Electronic Journal of Biotechnology* 2, 71-81

17 Moore, S. and Lukens, L. (2011) An evaluation of arabidopsis thaliana hybrid traits and their genetic control. *G3: Genes, Genomes, Genetics* 1, 571-579

18 Torii, K.U.*, et al.* (1996) The arabidopsis ERECTA gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. *Plant Cell* 8, 735-746

Appendix

Appendix A

RNAi:DMC1 construct

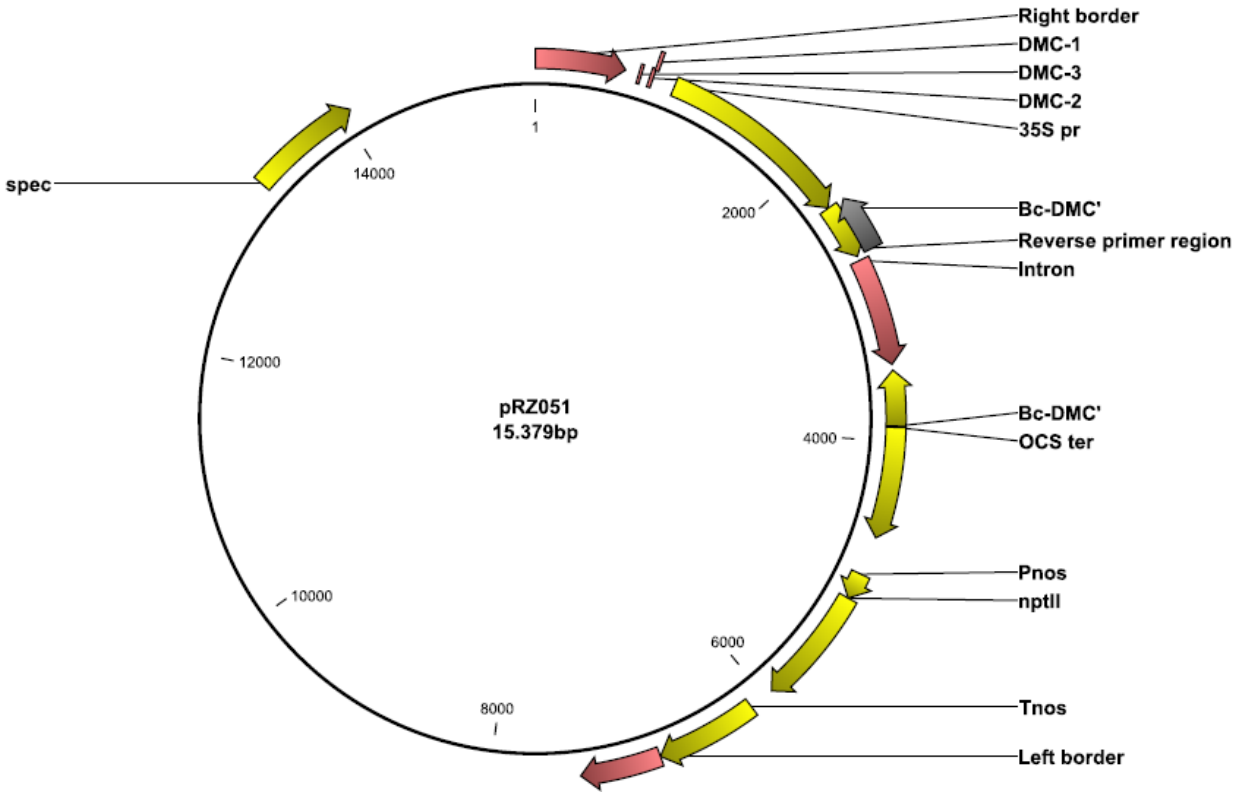

Figure S1 Structural representation of the *RNAi:DMC1* construct. T-DNA is in between of right border and left border. It comprises the reversed sequence of the last three exons of gene *DMC1*, a constitutive promoter 35S, *nptII* gene with the *pnos* promoter and *ocs* terminator.

Appendix B

Transformation protocol

Agrobacterium tumefaciens mediated transformation through flowerdip

YEBS media preparation

Bacteria are grown in liquid culture at 28°C, ~250 rpm, in YEBS liquid media

- 1 g/L yeast extract
- 5 g/L beef extract
- 5 g/L sucrose
- 5 g/L bacto-peptone
- 0.5 g/L magnesium sulphate; adjusted pH 7.

Agrobacterium preparation

1. Start 50 ml culture in YEBS with a bacterial streak. Use appropriate antibiotics (Table S1) for bacterial selection of the Ti-plasmid and the introduced transformation vector.

Table S1 concentration of antibiotics for bacterial selection of the T—plasmid and the introduce transformation vector

Antibiotics	Concentration
Streptomycine	100 mg/L
Spectinomycine	300 mg/L
Rifampicine	40 mg/L
Gentamycine	25 mg/L

2. Grow at 28°C until cell density is saturated (typically 2-3 days)
3. Pour the 50 ml culture into 450 ml of YEBS.
4. Grow at 28°C for 8 hours (start growing in the morning, dip in the afternoon)
5. Add and mix 150 µl Silwett L77 to the culture. (Amount of Silwett is plant genotype dependent advised is an amount between 100 and 200 µl).
6. The solution is ready for dipping (Direct Dip Solution), lines that have bolted and where flowers are visible are ready to be dipped. NOTE: siliques in slightly overgrown plants should be removed, as they will not be transformed.

Transformation through flower dip

1. All flowers are dipped in the Direct Dip solution.
2. 2 plants of the same accession are placed in the pot.
3. After dipping the plants are placed on its side and covered by plastic bags to ensure a high humidity, thus increasing transformation efficiency.
4. The covered plants are placed in trays inside a special cabinet in order to separate from non-transformed plants.
5. After 4 days, the bags are removed, and the plants put upright. From this point on, the plants are watered again with Hyponex growing solution.
6. All new shoots formed after these 4 days, are removed, increasing the transformation efficiency.
7. When the siliques show signs of ripening, a special paper bag is placed over the plants and the whole plant is bent downwards, ensuring all seeds are collected inside the paper bag. The plants will not be watered again from this point onwards.
8. The plants are left to dry and ripen. After several days, the stems are cut and all transformed seeds are collected inside the paper bag.
9. The collected seeds are dry sterilized and then grown aseptically for transformant selection.

GMO handling / disposal

- The *Agrobacterium* strain(s) are grown in a MLII lab. And transported to the greenhouse (PKMII) in a double sealed container.
- After the flower dip, the *Agrobacterium* strain(s) are disposed of in a designated GMO bin (located in the MLII lab).
- The Rockwool blocks of the left over plants are disposed in the designated GMO disposal bin in the greenhouse (a red bag inside a autoclave bag).
- The special trays are collected in a red bag, and then placed inside an autoclave bag, to be sterilized and stored for future use.
- The seeds are dry sterilized before future use.

Appendix C

Markers for genotyping

	Chromosome 1										Chromosome 2								Chromosome 3							Chromosome 4							Chromosome 5							Construct	
Markers	592939	3504562	wak1	M235	AIGI	GAPB_2	19996564	M305	ADH	29393153	68457	172469	3994520	7994335	11991100	15493536	18753024	19694139	580137	4000301	8000279	13491841	15990167	20428680	23443472	11691	641363	4040382	8001301	11999130	14177230	18488307	342420	RCL1B	8000694	11982076	FM4_2	23115566	26479586	NptII	promotor 35S

Appendix D

DNA extraction with CTAB

1. Add a single ball bearing to each 1.2 ml screw-lid tube containing tissue. Put samples into liquid nitrogen
2. Disrupt tissue by shaking with paint shaker for 2 min. Repeat if clumps of tissue remain.
3. Centrifuge briefly to bring down tissue dust.
4. Add 300 µl CTAB, close the lid loosely heat at 65°C in water bath for 30 min.
5. Let cool to room temperature.
6. Centrifuge briefly and add 300 µl chloroform (done in hood). Use a TIP BOX lid for chloroform. Close tightly with tube lid and vortex vigorously for 10-20 seconds.
7. Centrifuge at 3250 rpm for 15 min.
8. During centrifugation, prepare new tubes by adding 200 µl of very cold (-20°C) isopropyl alcohol to each well.
9. Transfer 200 µl of the chloroform-extracted supernatant to the new tubes. Be very careful not to transfer any of the goop from the interface, or any of the organic layer.
10. Close the tubes with lids and centrifuge at 3250 rpm for 15 min.
11. Pour off the liquid into the sink; the pellet of DNA should stay behind.
12. Wash with 200 µl of 70% ethanol, re-cover the tubes and centrifuge for 7-10 min. at 3250 rpm.
13. Dump off the liquid.
14. Repeat step 12.
15. Dump off the liquid again and blot the tubes dry on paper towels. Let them air dry for >3 hours (or until no liquid can be seen inside and the tubes don't smell like ethanol)
16. Resuspend the pellets in 50 µl of sterilized demi-water and let sit for >3 hours at 4°C. The DNA can then be stored.
17. Use 1 µl in a 10 µl PCR reaction

Table S2 Ingredients of CTAB

Chemicals	Amount for 1 L 2X CTAB	Final concentration
CTAB*	20g	2%
NaCl	81.82g	1.4M
1M Tris (pH 8)	100 ml	100 mM
0.5M EDTA	40 ml	20 mM

Appendix E

Concentrations of primers and chemicals used for TAIL-PCR

Table S3 Concentrations of primers and chemicals used for 1 x reaction volume for each TAIL-PCR round. 4x AD pool consists of 12 µM AD1, 12 mM AD2 and 12 mM AD3.

PCR round	Volume	Chemicals/ primers
	3.0 µl	MQ
	2.0 µl	5x PCR buffer
	0.4 µl	5mM dNTP's
	0.6 µl	25mM MgCl2
Primary PCR	0.4 µl	5uM prm DMC-1
	2.5 µl	4x AD pool
	0.1 µl	Go-taq
	1 µl	DNA template
		10 µl total volume
	4.0 µl	MQ
	2.0 µl	5x PCR buffer
	0.4 µl	5mM dNTP's
	0.6 µl	25mM MgCl2
Secondary PCR	0.4 µl	5uM prm DMC-2
	1.5 µl	4x AD1/ AD2/ AD3 [a]
	0.1 µl	Go-taq
	1 µl	DNA (1:50) [b]
		10 µl total volume
	7.0 µl	MQ
	4.0 µl	5x PCR buffer
	0.8 µl	5mM dNTP's
	1.2 µl	25mM MgCl2
Tertiary PCR	0.8 µl	5uM prm DMC-3
	5.0 µl	4x AD primers [c]
	0.2 µl	Go-taq
	1 µl	DNA (1:50) [d]
		20 ul total volume

[a] In the secondary PCR, the three AD primers were used separately.

[b, d] DNA samples used for the secondary and tertiary PCR were 50 times diluted PCR product from previous PCR round.

[c] According to the pattern of the agarose gel after secondary PCR, the samples with the corresponding AD primers that got clear bands were to be continued with for the tertiary TAIL-PCR.

Appendix F

Gel electrophoresis

Material needed: Agarose (1% by weight), TBE Buffer, 6X PCR sample loading buffer, DNA ladder, Electrophoresis chamber, Gel casting tray and combs, electric power, GelRed stain (5 µl for 100 ml TBE buffer), Staining tray, Gloves, Pipette and tips.

Recipe of 1L 5X TBE Buffer stock solution

- 54 g of Tris base
- 27.5 g of boric acid
- 20 ml of 0.5 M EDTA (pH 8.0)

PCR products are loaded to the solidified agarose gel and applied with the voltage of 75W for 45 min to 1 hour.