FINAL REPORT OF THE RESEARCH PROJECT
"IDENTIFICATION AND ANALYSIS OF GENETIC LOCI THAT CAUSE HETEROSIS IN CROP PLANTS"
(OTKA K-105170)

Heterosis or as it is often referred hybrid vigor is the increased performance of progeny relative to their parents. Plant breeders have successfully exploited this phenomenon to increase the performance of cultivated species although the genetic determinants of heterosis was hardly understood. Several studies aimed to reveal the genetic basis of heterosis background and based on preliminary results and theoretical considerations, two main concepts were used to explain the phenomenon of heterosis. (i) One concept was the "dominance" model which explained heterosis with the complementation of recessive detrimental alleles in the hybrid. The (ii) "overdominance" model suggested that heterosis was the result of superior effect of heterozygous allele combinations.

The aim of this project was to identify and analyze genetic loci that produce heterosis in crop plants. The two concepts of heterosis was going to be tested using different plant species. The single-gene (overdominance) concept of heterosis was supported by the findings that mutations in a single tomato gene resulted in heterosis of fruit yield. It has been demonstrated that heterozygous configuration of the loss-of-function and wild-type allele of the tomato *SINGLE FLOWER TRUSS* (*SFT*) gene resulted in increased yields of F1 hybrids (Krieger et al. 2010). *SFT* encodes a transmissible polypeptide regulating the transition of vegetative growth into development of flowers and hence fruits. According to the accepted hypothesis the heterozygous state of *SFT* provides the exact dose of the gene product causing the highest yield of fruits (heterosis). This discovery could be used to exploit the natural variations of the SFT gene to evaluate its genetic diversity. This could extremely useful since the original sft mutant was generated by mutagenesis using EMS that induce unknown and unwanted mutations in the genome. In addition the identification of loss-of-function alleles of SFT in a natural collection could have the advantaged that these varieties are already locally adapted and can be directly used for breeding purposes. Based on these conception, we planned to identify tomato and pea varieties carrying variants in the tomato *SFT* gene or in the pea *FT* gene, a homolog of *SFT* (**Task 1**) which can be used in genetic crosses to produce hybrids displaying heterosis.

The analysis of the heterozygosity in alfalfa (*Medicago sativa*) was planned (**Task 2**) based on concept of the "dominance" model of heterosis. We aimed to determine the chromosomal regions which heterozygous configuration is necessary for heterosis in alfalfa. In order to study the heterosis in alfalfa we planned to generate homozygous alfalfa lines using a

novel approach by producing double haploid (DH) lines of alfalfa and generate hybrid plants with these DH lines.

Task1

(A) Analysis of genetic diversity of tomato accessions

Tomato plants heterozygous for loss of function alleles of the SINGLE FLOWER TRUSS (SFT) gene, which were located to the coding sequence of *SP3D*, was shown to display heterosis in fruit yield producing about 60% more fruit than the wild type (Krieger et al. 2010). One aim of our current project is to discover natural variations in this gene.

To achieve our goal, 299 tomato varieties with diverse phenotypes in flowering and fruit production (early and late flowering and ripening, determined and semi determined, large, many or few flowers) were selected from the tomato collection of the Plant Diversity Centre, Tápiószele, since these phenotypes might be associated with altered sequence of *SP3D*. We successfully isolated DNA from pooled leaves of 291 varieties. Our original strategy was to amplify and sequence the four exons of the gene separately in order to discover natural variations. This, however, was changed and we decided to sequence the entire gene, both exons and introns, and some portion of the untranslated 5' and 3' regions, too. Primers were designed using the 6,819 bp long genome sequence of the *SP3D* gene (NCBI entry AY186735). Finally, a right size PCR product was obtained from 262 samples which were then sequenced by a service company using the Illumina HiScanSQ platform. The 1×50 bp deep sequencing of the *SP3D* 5082 bp amplicon from 262 tomato varieties was carried out. The 55.908.308 reads were aligned to the reference sequence and computer software were used to identify SNP in the SFT gene. By this way, we have identified 101, 311 and 84 SNPs, respectively, of which 118 were detected by at least two methods. Of these 118 SNPs, four were exonic, while 114 were found in the introns and the promoter region of the *SP3D* gene. All four exonic SNPs were synonymous, thus not resulting in any amino acid change or loss of function mutation. According to published data, the quality of the nucleotides in the -62 and -61 positions upstream of the ATG start codon of SP3D are linked with the so called sympodial index, spi (i.e. the number of vegetative branches between inflorescences) in tomato. Spi2 and Spi3 plants produce 2 and 3 branches, respectively, and Spi2 plants give a higher yield because they produce more fruit due to the relatively more number of inflorescence per plant. Spi3 is associated with either a CT or a TA dinucleotide and Spi2 is associated with a CA dinucleotide. In the 651 bp long promoter region of the *SP3D* gene covered by the sequenced amplicon, we have detected polymorphism in the -62 and -61 positions. While the reference sequence contained C and T in

these positions, respectively, we detected a T at around 1 % frequency in the -62 position and an A at around 3 % frequency in the -61 position. Determining the dinucleotide frequencies revealed that the very large majority (96.91%) of the reads covering these positions had CT, while CA, TA and CC were detected at 1.96, 1.08 and 0.04 % frequencies, respectively. We have then genotyped the 262 varieties for these two position using the qPCR based KASP technology. A CT dinucleotide was found in 251 entries, seven had C/TA, one had CT/A and three had TA. The manuscript describing the genetic diversity among the germplasm items was prepared and published in the journal of Plant Gene (Marincs et al. 2017) (https://ac.els-cdn.com/S2352407316300336/1-s2.0-S2352407316300336-main.pdf?_tid=66c48621-46cb-4d23-bee9-502a20ef25cd&acdnat=1528225310_55972f616c3d0934d99cef4925f81054) and in the proceedings of Hungarian Agricultural Research (Endre et al. 2016).

Our earlier results suggested that the tomato SP3D gene, responsible for fruit yield heterosis, has no naturally occurring loss of function mutations (Marincs et al. 2017). Genome editing is a possible method to generate mutations in a given gene, and for this the required molecular components are introduced into the plant cells by transformation. In tomato, an in planta fruit transformation method was published. We have tested this method in our laboratory, because the published high transformation frequency would make this method suitable for genome editing. Unfortunately, we could not detect any transgenic plants, despite that after fruit injection transient GFP expression was observed in a number of varieties in the developing seeds. We are going to perform the traditional transformation protocol of tomato with the Agrobacterium strain carrying construct designed for the genome editing of SPD3 in tomato.

(B) Results on the genetic diversity analysis of pea varieties

To perform the same genotyping of a pea *FT* gene, an ortholog of *SP3D*, leaf samples were collected from 208 pea varieties available at the Tapiószele genebank. Two primers were designed based on the pea *FT* genomic region and a 970 bp fragment covering the entire coding region and about 500 bp of the promoter was amplified by PCR resulting in a single, appropriately sized fragment in 207 DNA samples. Our institute is participating in the evaluation of a new generation sequencing device (MinION), therefore we have sequenced the amplicon pool of the PCR fragments using that device. Surprisingly, the obtained reads did not match to the pea reference sequence at all. We, therefore, sequenced the DNA pool using traditional Sanger sequencing, which revealed 100% match between the pool and the reference. Thus we concluded that MinION sequencing had a major technical fault and we plan an Illumina deep sequencing of the pea amplicon, similar to that, which was performed in tomato.

Using the same primer pair, PCR was performed with the 208 DNA samples of pea accessions to generate fragments for sequencing which was carried out by a service provider using Illumina's NGS technology. Compared to the reference sequence, SNPs were identified in the about 70 millions 50 bp long reads obtained by new generation sequencing of the pea FT orthologue gene using bioinformatic methods. Altogether, 639 SNPs were identified, of which six and 19 were in the 5' and 3' untranslated regions, and 273 and 341 were in the introns and exons, respectively. Regarding protein function, the exonic SNPs are the most important, thus we analyzed those exonic SNPs further, which displayed at least 1% frequency, since in theory at 1% frequency that particular SNP can be homozygous or heterozygous in four varieties, respectively. Of such 42 exonic SNPs, 37 and 5 were transversion and transition, respectively. By another evaluation, of the 42 SNPs ten were synonymous, 29 were non synonymous, missense and three were non synonymous, missense i.e. stop gain. Stop-gain SNPs can result in loss-of-function of the protein. The confirmation of the these three SNPs in the varieties by using KASP genotyping, which was used before for the tomato SP3D gene in this project, is in progress to determine the varieties carrying that type of SNPs.

Task2

Analysis of heterozigosity of diploid alfalfa (*Medicago sativa*) plants

The classical way to produce heterosis is the crossing of inbred lines and the performance of the generated hybrids. In order to analyze the highest degree of heterosis in diploid alfalfa, the cross of homozygous plants was planned. To obtain homozygous diploid alfalfa lines we test the feasibility of generation doubled haploids (DHs) of alfalfa plants. To generate DHs, anther cultures of diploid M. sativa CADL (Cultivated Alfalfa at Diploid Level) and M. sativa ssp. falcata plants were induced using the methods published previously by Zagorska et al. (1997; In vitro Cell Devel. Biol.-Plant. 33:107-110). Flower buds were collected form diploid alfalfa plants and following surface sterilization, anthers were isolated from buds and cultured on callus induction medium. After ~25 days of callus induction the cultures were transferred to shoot induction medium and grown under 16/8 h photoperiod. Subsequently, plants were regenerated from calli and planted into Jiffy pots for acclimatization before transferring to greenhouse. Altogether 5030 anthers were isolated from 33 CADL lines and 103 plants deriving 12 independent lines of 5 CADL plants could be regenerated (~ 2% efficiency of regeneration).

Molecular markers were used to test the homozygosity of the 12 independent lines and to prove DH genomic background of the DH candidate plants. Genomic DNA of the 12 independent lines was purified and PCR amplifications of simple sequence repeat (SSR)

markers distributed all over the eight chromosomes of alfalfa were carried out. We selected 18 SSR markers which showed high rate of polymorphism among the parental lines. The analysis of the allele distribution of the SSR markers in the 12 independent lines revealed that the lines showed polymorphism for several SSR markers indicating the heterozygous nature of their genome. The heterozygosity of the tested SSR markers proved that these lines were not DHs. We suppose that the regenerated plants derived from the diploid cells of the cell wall of anthers and not from microspores. In order to avoid calli induction from the diploid cells of the anthers we isolated microspores from the anthers and attempted to generate DH alfalfa plants using a slightly modified DH generation method. Unfortunately we could not regenerate plants from isolated microspores. We concluded that published protocol to generate DH plants from anthers or microspore has not worked under our conditions.

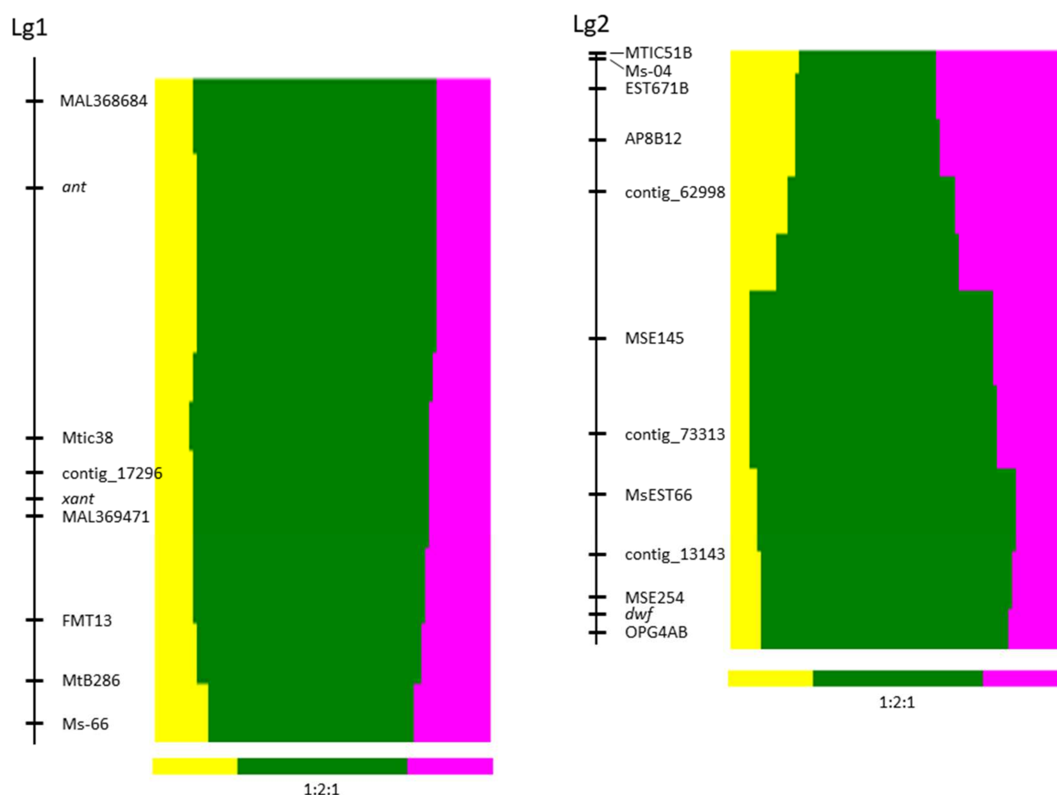
Because of the inefficiency of generation of DH plants we needed to apply an alternative way to produce hybrids for heterosis analysis. CADL plants can be regarded as partlially inbred lines. SSR markers were used to analyse the heterozygosity in large number of CADL plants and 12 independent lines were selected showing the lowest heterozigosity for the tested SSR markers. These 12 plants were selected and crossed to each other and the yellow flowered diploid *M. sativa* ssp. *quasifalcata* plants to generate F1 plants. We carried out more than 1600 crosses between the purple and white flowered CADL plants in more than 100 combinations. These crosses resulted in 74 pods of 29 crossing combinations. About 800 crosses between 47 CADL and *M. sativa* ssp. *quasifalcata* plants produced 268 pods of 38 crossing combinations. Seeds from the pods were germinated and seedlings were planted into green house. The F1 hybrid plants were grown and vegetative cuttings were prepared from the hybrid and parental plants and the propagated plant material were transferred to Kisvárda where field experiments were carried out. Altogether 814 vegetative propagation of 105 parental and hybrid plants were planted and tested for their performance in field experiment. The green yield, height, stem thickness and length and wideness of leaves of these plants were measured. Based on the yield data several F1 plants were pre-selected and self-mated in the greenhouse of ABC at Gödöllő to generate F2 plants to establish segregation populations. We have selected those crosses where the better performance of the hybrids compare to the parental lines (heterosis) was detected measuring green and dried mass and yield. We have selected the crosses of DH66xDH90 and DH83xDH66, and hybrids of DH89, DH83 and Mcw2 with *M. quasifalcata* (Mq) as these crosses produced the highest absolute values in green and dried mass. The selected F1 hybrids were self-mated to produce high number seeds to establish F2 segregating population to analyze the correlation between heterozygous configuration of chromosomal segments and the genetic

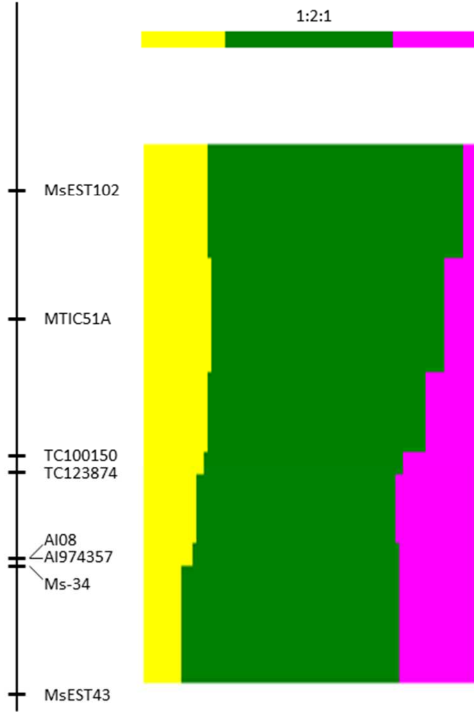
loci affecting the main characters (green yield, height, stem thickness and length, leaf sizes, etc.) supposed to be important for heterosis. Unfortunately we had difficulties in producing sufficient number of plant materials for the next step of experiments. On the one hand, a massive infection of mites occurred in our greenhouse resulted in very low efficiency of seed production of the selected F1 plants. Nevertheless we were able to generate 40-60 F2 plants/hybrids of four F1 (DH89xMq, DH83xMq, Mcw2xMq and DH83xDH66) plants to establish F2 mapping populations. The survival of the individuals in these F2 populations was very low and finally except one population (Mcw2xMq) less than ~ 25-30 plants remained per F2 populations. In addition the efficiency of production of vegetative cuttings of these left over plants was very low and the number of plants was not sufficient to start field experiments. Due to the problem in generating plant material for the next step of the project, we applied for the prolongation of the project. Additional F2 plants and vegetative cuttings were generated in the next year and the propagated plant material was transferred to Kisvárda for field experiments.

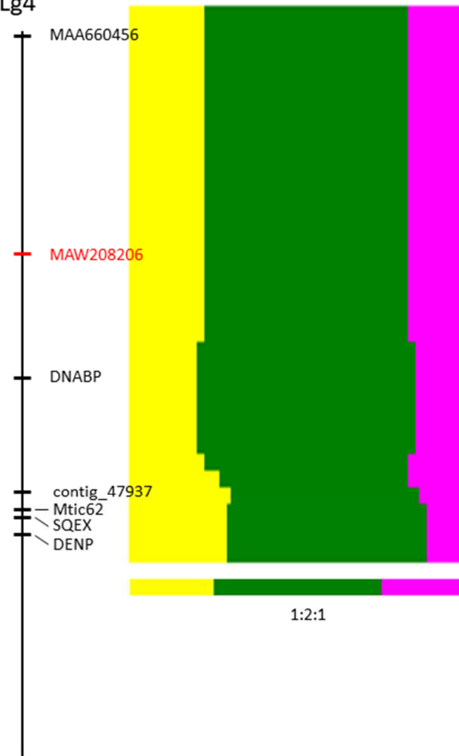
Genomic DNA was purified from the F2 individuals of three segregation populations (DH89xMq, DH83xMq, Mcw2xMq) having sufficient number of plants for genetic mapping. The genotypes of the F2 plants were determined for DNA based molecular markers developed based on the genomic resources of the model legume *M. truncatula* and alfalfa. The genetic maps of these mapping populations have been constructed (Figure 1). The height, fresh and dry mass, shoot and leaf weight, ratio of shoot/leaves of the F2 individuals of the three mapping populations was measured. We used the values of those F2 plants for further genetic analysis only when the values of the characters were generated from at least three vegetative propagated sister plants of the F2 individuals, respectively. In order to determine the chromosomal regions possessing excess of heterozygous genotypes, the segregation ratio of the genotypes for each genetic markers was calculated and presented by colors along the linkage groups of diploid alfalfa (Figure 1). Markers showing extreme distorted segregation showing the excess of heterozygous genotypes was observed in the middle and the bottom part of linkage group 2 (LG2), in the top of LG3, in the middle of LG7 and all along the linkage groups 6 and 8 (Figure 1). The aim of our study was to test the concept if there is correlation between these heterozygous chromosomal regions and quantitative traits, mainly the yield, of diploid alfalfa. In order to analyze the correlation, we had to determine the QTLs (quantitative trait loci) correlating with the variance of the traits. Therefore we attempted to determine intervals of genomic segments correlate with the variation of the measure traits using the JoinMap/MapQTL software. The genetic analysis of the quantitative traits has not detected statistically significant linkage to the markers mapped in the three genetic maps. The reason of the failure to detect

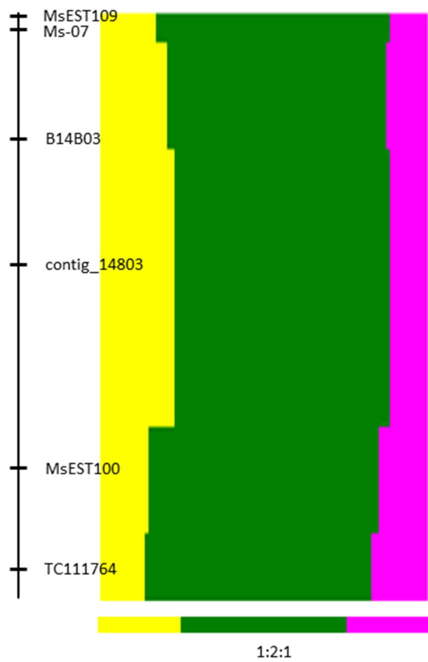
QTLs could be the (i) incomplete data of the F2 segregating populations (we had several F2 plants in all three populations without sufficient number of vegetative cuttings), (ii) the limited size of the mapping population and (iii) the high variance of the traits determined only in one season (due to the problem of generating the plant material we were not able to score the plants more than one season for the quantitative traits.

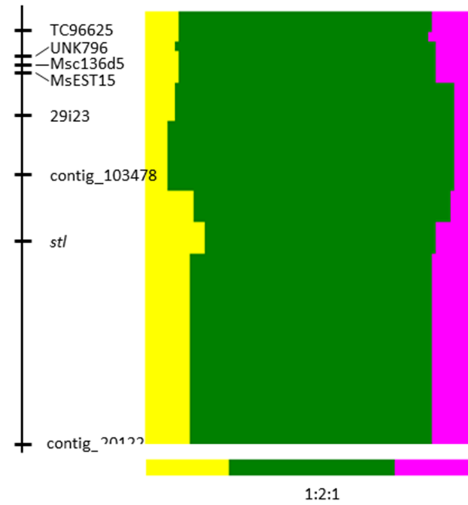
In order to exploit the results of our analysis and have sufficient and reliable data for publication, we plan to continue the QTL mapping. We are going to extend the mapping populations and generate additional vegetative cuttings of the F2 plants to have reliable yield data determined in more than one season.




Lg3
1:2:1
MsEST102
MTIC51A
TC100150
TC123874
AI08
AI974357
Ms-34
MsEST43


Lg4
MAA660456
MAW208206
DNABP
contig_47937
Mtic62
SQEX
DENP
1:2:1


Lg5
MsEST109
Ms-07
B14B03
contig_14803
MsEST100
TC111764
1:2:1


Lg6
TC96625
UNK796
Msc136d5
MsEST15
29i23
contig_103478
stl
1:2:1

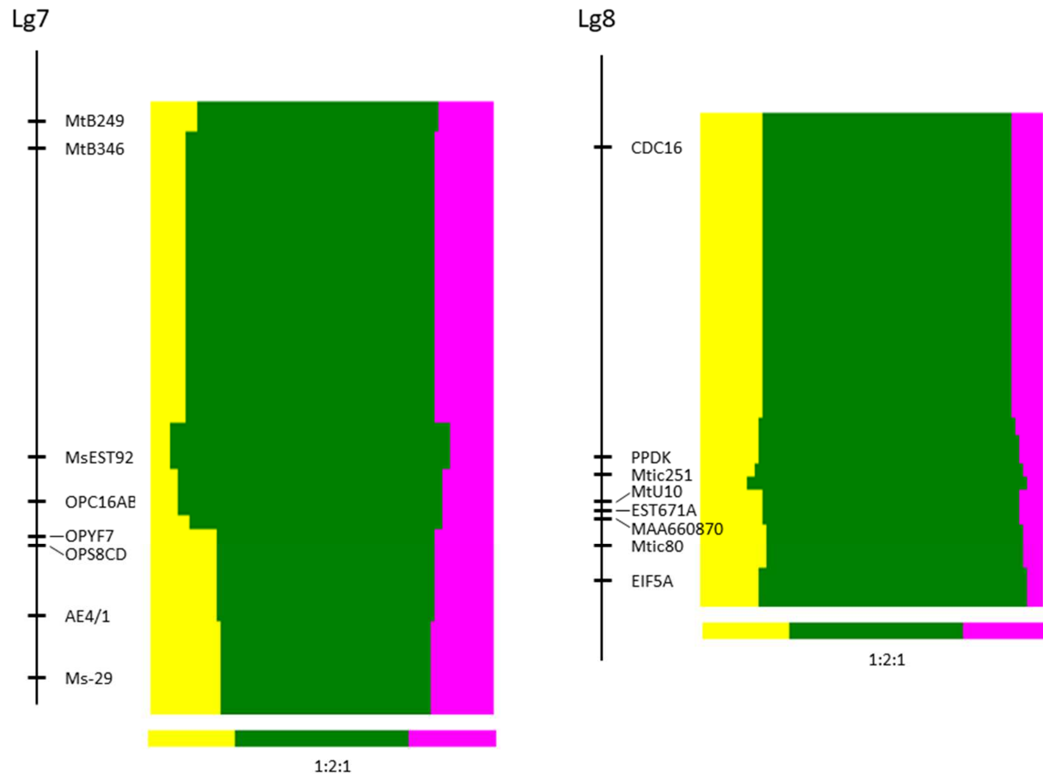


Figure 1. The core genetic markers on the eight linkage groups of the diploid alfalfa mapping populations developed in this study. Beside the linkage groups, the segregation ratio of the genetic markers are shown with colors. The yellow, green, and magenta colors represent the maternal homozygous, heterozygous, and paternal homozygous genotypes, respectively. The theoretical 1:2.1 segregation ratio of the co-dominant markers are shown below the color panels.

(B) The analysis of the translocation between chromosome 4 and 8 in the *Medicago truncatula*

In order to generate PCR-based genetic markers on diploid alfalfa linkage map, we used the genomic resources of the closely related model legume, *Medicago truncatula*. When we carried out the genetic mapping of markers predicted to be positioned on chromosome 4 of *M. truncatula*, we found that few markers showed wierd linkage relationships, they were both linked to markers on lower arm of chromosomes 4 and 8 of *M. truncatula*. The basic genetic map of the model legume *M. trucatula* was developed using the F2 individuals of the cross a *M. truncatula* genotypes A17 and A20 (Choi et al. 2004). A more detailed and focused genetic mapping was performed on the F2 population of the *M. truncatula* A17 x A20 genotypes that revealed a cruciform linkage between markers located on the lower arm of linkage group 4 and 8 (Figure 2).

Parallel to this study, we carried out the genetic mapping of an ineffective symbiotic locus of mutant 13U of *M. trucatula*. The mutant 13U was generated in the Jemalong genetic background. Using the Jemalong-based 13U plant, linear genetic linkage was observed between the genetic markers in the mapping population of Mt A20 and Mt 13U (Figure 3). A previous study suggested that the Mt A17 genome contains aberrant chromosomes which might cover a translocation between the affected chromosomes (Kamphuis et al., 2007). Therefore we suspected that the translocation between chromosomes 4 and 8 of A17 generated the unexpected linkage phenomenon and this event did not occur in the closely related Jemalong accession. To confirm and highlight the molecular events which led to the aberrant chromosomal behavior, we decided to identify the translocation breakpoint. The main point of this study was to determine to what extent the reference genome of the model legume *M. trucatula* could be used in the genetic analysis of alfalfa. In order to define the translocation breakpoints, DNA segments involved in the chromosomal rearrangements were searched for. DNA sequences between the break points pinpointed by high resolution genetic mapping were compared between the chromosomal segments of Mt A17, Jemalong and Mt A20. This analysis led to the discovery of a DNA segment where sister chromatid break and reunion occurred between the two chromosomes involved in the translocation event. We tested more than 100 ecotypes of *M. truncatula* and it is interesting to note that only the MtA17 carried this breakpoint among the studied Mt accessions. The identification of the translocation breakpoint defined precisely the genomic regions where the co-linearity between the reference model legume genome and other *Medicago* accessions and species discontinue. The manuscript presenting the details of the abnormal chromosomal configuration of the reference genotype *M. truncatula* A17 is going to be resubmitted in the next weeks (Szabó, Z., Balogh, M. Miró, K, Ellis, NTH, Cook, D., Debelle, F., Kiss, G.B. and Kaló, P. DNA sequence of the translocation breakpoint between chromosome 4 and 8 in the genome of *Medicago truncatula* A17).

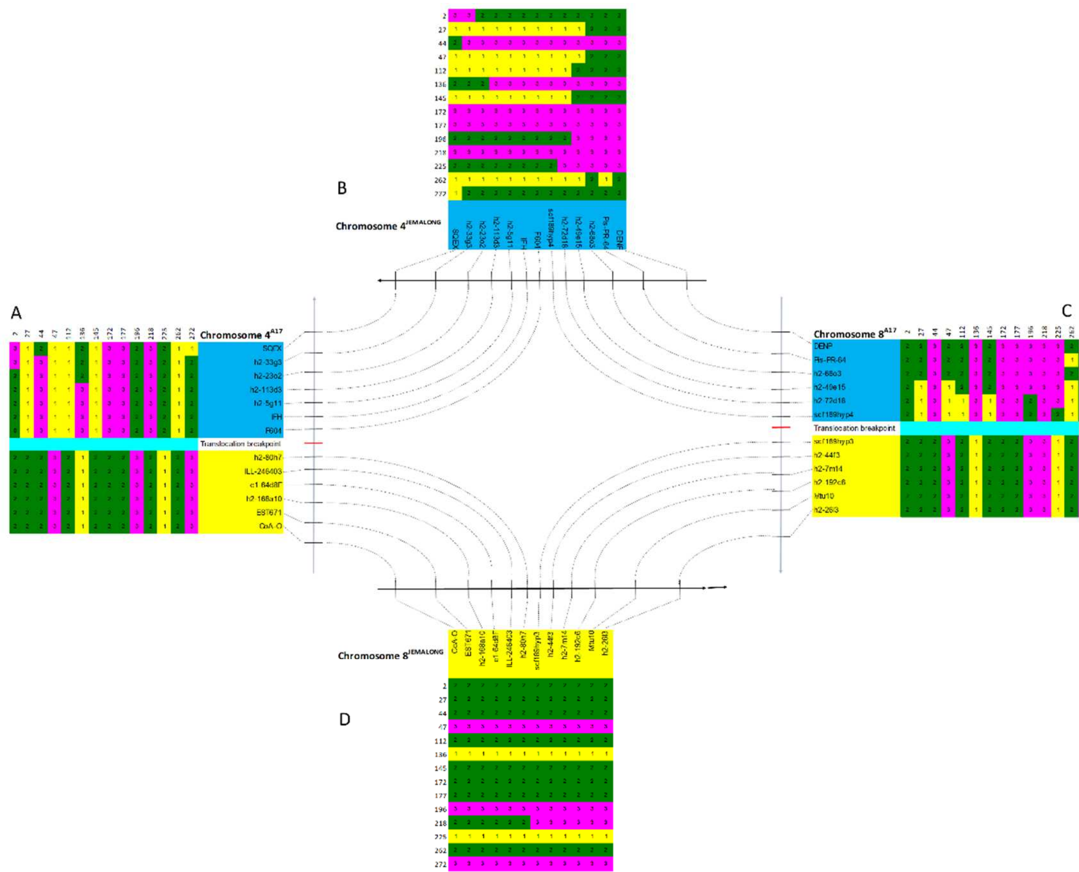


Figure 2. Color coded genotypes of F2 individuals of chromosome 4 and 8 from cross Mt 13U (Jemalong) x Mt A20 (panels B and D) and the reference *M. truncatula* genetic map developed from the cross of Mt A17 x MtA20 (panels A and C). The predicted translocation breakpoints between the chromosomes 4 and 8 of the A17 genotype are presented by turquoise bars in panel A and C.

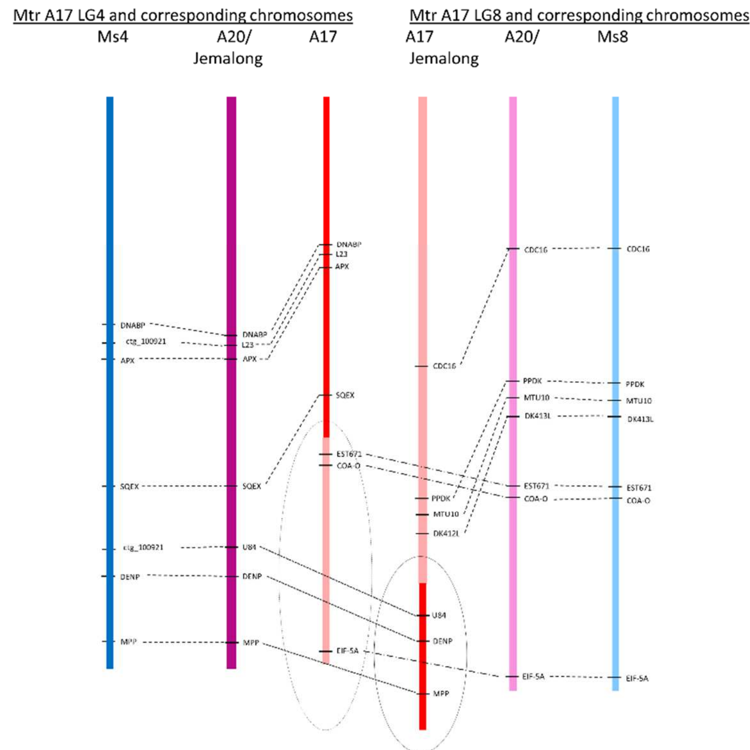


Figure 3. The position of genetic markers affected in the translocation event in A17 genotype reveals. All the *M. truncatula* accessions and ecotypes, presented by A20 and Jemalong in this figure, and alfalfa (*M. sativa*) have co-linear marker order indicating that the translocation between chromosome 4 and 8 is unique to genotype A17 (circled with solid and dashed lines).

References

Choi et al. 2004. A Sequence-Based Genetic Map of *Medicago truncatula* and Comparison of Marker Colinearity with *M. sativa*. Genetics 166: 1463–1502

Endre Barta, Zsófia Bánfalvi, Zoltán Havelda, László Hiripi, Zsigmond Jeney, János Kiss, Balázs Kolics, Ferenc Marincs, Dániel Silhavy, Viktor Stéger, Éva Várallyay (2016) Agricultural genomics: an overview of the Next Generation Sequencing projects at the NARIC-Agricultural Biotechnology Institute in Gödöllő, Hungarian Agricultural Research 25:10-21

Kamphuis et al. 2007. The *Medicago truncatula* reference accession A17 has an aberrant chromosomal configuration. New Phytologist 174: 299 –303.

Krieger U, Lippman ZB, Zamir D (2010) The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. Nature Genetics 42: 459–463.

Marincs, F, Nagy, T, Miró, K, Kollár, Z, Barta, E, Kaló, P (2017) Large-scale amplicon sequencing of the SP3D gene responsible for fruit-yield heterosis in tomato. Plant Gene 9: 45–49

Zagorska, N, Dimitrov, B, Gadeva, P and Robeva, P (1997) Regeneration and characterization of plants obtained from anther cultures in Medicago sativa L. In Vitro Cellular & Developmental Biology - Plant 33: 107-110.