GENETIC INVESTIGATIONS OF HETEROSIS AND INBREEDING DEPRESSION IN DIPLOID AND TETRAPLOID ALFALFA

by

XUEHUI LI

(Under the Direction of E. Charles Brummer)

ABSTRACT

Alfalfa (*Medicago sativa*) is one of the most important forage crops in the world; however, yield increases have lagged behind many other crops. Hybrids between *Medicago sativa* subsp. *sativa* and *M. sativa* subsp. *falcata* often express heterosis for yield. Understanding the genetic mechanisms causing heterosis could assist alfalfa breeders to improve yield.

In the current studies, we investigated the genetic basis of heterosis and inbreeding depression from several perspectives. First, using classical quantitative genetics, we assessed the importance of dominance, over-dominance, epistasis, and multiple allelic interactions on biomass yield across generations. Our results suggested that the complementary interactions of favorable alleles/linkats from both *M. sativa* subsp. *sativa* and *M. sativa* subsp. *falcata* may play an important role in heterosis for yield. Second, we used Affymetrix gene arrays to compare gene expression profiles between alfalfa hybrids that expressed heterosis for biomass yield and a hybrid that did not. More nonadditive expression and expression outside the parental range was observed in the heterotic hybrids compared to the non-heterotic hybrid. Based on our results, we hypothesized that nonadditive expression at different developmental stages and in different tissues may contribute to biomass yield heterosis. Third, we developed an F_1 population derived

from a diploid *M. sativa* subsp. caerulea x subsp. *falcata* cross and an F_2 population from one F_1 individual. The two populations were germinated and grown under the same environmental conditions, and assessed for segregation distortion patterns using molecular markers. In the F_2 population, 58 of 80 SSR markers showed segregation distortion (SD), mostly with heterozygote excess. The clustering of the markers suggested that multiple segregation distortion loci (SDL) are located on different chromosomes of alfalfa. Over-dominance or pseudo-overdominance of SDL may explain the heterozygote excess of distorted markers. The effects of SD on linkage mapping and QTL mapping were discussed.

INDEX WORDS: Alfalfa, *Medicago sativa* subsp *sativa*, *M. sativa* subsp. *falcata*, Hybrid, Heterosis, Inbreeding depression, Biomass, Gene expression, Nonadditive expression, Segregation distortion (SD), Segregation distortion loci (SDL)

GENETIC INVESTIGATIONS OF HETEROSIS AND INBREEDING DEPRESSION IN DIPLOID AND TETRAPLOID ALFALFA

by

XUEHUI LI

B. S., China Agricultural University, China, 1998

M. S., China Agricultural University, China, 2001

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2009

© 2009

Xuehui Li

All Rights Reserved

GENETIC INVESTIGATIONS OF HETEROSIS AND INBREEDING DEPRESSION IN DIPLOID AND TETRAPLOID ALFALFA

by

XUEHUI LI

Major Professor:	E. Charles Brummer
Committee:	Michael L. Arnold H. Roger Boerma Steven J. Knapp Wayne A. Parrott

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2009

DEDICATION

I dedicate my dissertation to my loving and supporting wife Yanling Wei, and to my wonderful kids Amanda and Braden.

ACKNOWLEDGEMENTS

This dissertation could not be completed without so many people's help and support.

First, I would like to give my special thanks to Dr. E. Charles Brummer, my major professor, for his guidance in my study and research, and his patience and encouragement in my career development.

Second, I would like to thank my current committee members, Dr. Michael L. Arnold, Dr. H. Roger Boerma, Dr. Steven J. Knapp, Dr. Wayne A.Parrott, and my previous committee members at Iowa State University, Dr. Ken Moore, Dr. Jode Edwards, Dr. Ted Bailey, Dr. Jean-Luc Jannink, for their willingness to serve on my comittee and their critical advices on my research projects.

Third, I would like to thank all of our previous and current lab members for their friendship and support. Baldo, Heathcliffe, Joe, Valentin, Mindy, Julia, Muhammet, Rafael, Ananta, Hammad, and Xiaojuan provided valuable discussions on my research; Mark, Diane, Donald, Jonathan, and Wesley helped with field management and data collection; our current lab manager, Yanling, who's also my wife, provided technical support for my research.

Most valuably, my wife's continuous love and support has accompanied me through all of the difficult as well as fun times during my study and research. At the mean time, I'd like to also give my appreciation to my kids, Amanda and Braden for their lovely smiling and happiness brought to me.

Finally, I'd like to give my thanks to all of those that their names are not mentioned here but who helped and encouraged me through my study.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

ALFALFA BIOLOGY, GERMPLASM, AND BREEDING

Alfalfa (*M. sativa*) is one of the most important forage crops in the world. It is a perennial, outcrossing species, suffers from extreme inbreeding depression, and exhibits a weak self-incompatibility system. Cultivated alfalfa is a tetrasomic tetraploid. Alfalfa belongs to a cluster of taxa termed the *M. sativa-falcata* complex which includes both diploid and tetraploid members. The diploids (2n=2x=16) include *M. sativa* subsp. *caerulea*, *M. sativa* subsp. *falcata* and subsp. x *hemicycla*; tetraploid (2n=4x=32) taxa include *M. sativa* subsp. *sativa*, subsp. *falcata*, and subsp. x *varia*. The most closely related species are *M. glutinosa* (2x) and *M. glomerata* (4x) (both of which are sometimes given subspecific status) and the somewhat more clearly distinct species, *M. prostrata*. The species within the complex can be intercrossed to each other and share the same karyotype (Quiros and Bauchan, 1988). Historically, nine germplasms of alfalfa (*M. falcata*, *M. varia*, Ladak, Turkistan, Flemish, Chilean, Peruvian, Indian and African) have been introduced into USA (Barnes et al., 1977).

Most current alfalfa cultivars are synthetic populations developed by recurrent phenotypic selection. Since the 1950s, alfalfa breeders have generally focused on improving insect and disease resistances, thereby indirectly improving biomass, and large parental populations (>100 individuals) have been used to avoid inbreeding depression. The increase of biomass before 1980 probably was due to the incorporation of multiple resistances and possibly

the greater heterozygosity resulting from the inclusion of many different germplasm sources into the breeding programs. However, biomass has increased little since the 1980s (USDA, 2004). One likely cause is that selection has not been specifically focused on biomass yield so that alleles controlling yield have not been consolidated in the extant breeding materials (Barnes et al., 1977).

HETEROSIS ON BIOMASS IN ALFALFA

One effective way to improve yield in the future is to capitalize on non-additive gene action by harnessing heterosis. Non-additive genetic variances for yield were found to be large in alfalfa (Riday and Brummer, 2005), suggesting that breeding methods that used this variation could lead to yield improvements. Alfalfa breeders have attempted to create inbred lines to produce hybrids since the 1930s, but because of severe inbreeding depression, no true inbred lines have been developed (the genetic load that an autotetraploid can carry is huge compared to that of a diploid, like maize, which helps explain some of the failure to derive inbreds). Brummer (1999) proposed a semi-hybrid strategy to capture partial heterosis and also proposed three hypothetical heterotic groups in alfalfa, including *M. sativa* subsp. *falcata*, dormant or moderately dormant *M. sativa* subsp *sativa* and nondormant *M. sativa* subsp *sativa*. Hybrids between subspecies *sativa* and *falcata* showed heterosis for biomass production (Bhandari et al., 2007; Riday and Brummer, 2002; Riday and Brummer, 2005; Sriwatanapongse and Wilsie, 1968), but hybrids between the dormant and nondormant sativa germplasm groups did not (Sakiroglu and Brummer, 2007). While these field-based observations demonstrate the potential of heterosis expression in alfalfa with some germplasm, a fuller understanding of the genetic mechanisms causing heterosis could assist breeders in reliably creating high-yielding hybrids.

GENETIC BASIS OF HETEROSIS, INBREEDING DEPRESSION, AND OUTBREEDING DEPRESSION

In classic quantitative genetics, three main hypotheses (dominance, overdominance, and epistasis hypothesis) have been proposed to explain heterosis (Falconer and Mackay, 1996). All three hypotheses postulate that physical allelic variation between parents results in novel allelic interactions in F_1 hybrids, which in turn causes heterosis. Although not always explicitly stated, all three mechanisms concurrently may play a role in heterosis.

Most genetic knowledge about heterosis in alfalfa has been inferred from inbreeding studies. Severe inbreeding depression results from the high level of deleterious recessive alleles carried by alfalfa plants (Jones and Bingham, 1995). Busbice and Wilsie (1966) found that the effect of level of inbreeding depression on biomass yield in tetraploid alfalfa was more severe than predicted by decreased heterozygosity alone, leading them to propose that severe inbreeding depression in tetraploids might be due to the rapid loss of multiple-allelic interactions within a locus. Inbreeding studies with two-allele tetraploid alfalfa suggested that heterozygosity *per se* (overdominance) probably is not the main cause of heterosis, but instead, that complementary genetic factors, caused by dominance and/or epistasis, is likely the predominant contributor to heterosis (Bingham et al., 1994; Kimbeng and Bingham, 1998; Woodfield and Bingham, 1995).

Crosses between different species or distantly related populations frequently lead to substantial loss of vigor, a phenomenon called outbreeding depression (Lynch and Walsh, 1998). In some situations, outbreeding depression is observed not in the F_1, which may show heterosis, but in the F_2 or later generations, in which the vigor is significantly less than the mean of original parents (Lynch and Walsh, 1998). Heterosis and inbreeding depression can be generally explained by dominance effects between alleles within loci, while outbreeding depression is

probably due to the loss of favorable epistatic interactions (co-adapted gene complexes) present in the parental lines (Lynch and Walsh, 1998; Templeton, 1986).

The molecular basis of heterosis is obscure. Alleles at a given locus can have different levels of gene expression (Brem et al., 2002; Kliebenstein et al., 2006), and heterosis may be explained at the molecular level by the combined allelic expression levels in F_1 hybrids, and in particular, by nonadditive expression at each locus involved in a trait (Birchler et al., 2003). Variable levels of nonadditive gene expression have been found in previous studies of hybrids compared to their parents (Cui et al., 2006; Stupar and Springer, 2006; Swanson-Wagner et al., 2006; Vuylsteke et al., 2005; Wang et al., 2006). However, the different degrees and types of nonadditive expression observed in these studies could be due to biological, technical, and/or statistical analysis differences, so generalizations about nonadditive gene expression in hybrids across studies and species are difficult to make. Unfortunately, none of these experiments assessed gene expression in hybrids that do not show a heterotic response for the trait of interest, making conclusions that nonadditive expression is related to heterosis difficult to support.

SEGREGATION DISTORTION IN ALFALFA

The molecular mapping of biomass can identify QTLs and further facilitate breeding, and also facilitate the genetic investigations of biomass and heterosis at the same time. It is documented that distorted markers may bias the estimation of genetic distance between markers and can affect orders among markers (Lorieux et al., 1995a; Lorieux et al., 1995b). And it will further affect the genetic mapping of phenotypic traits (Vogl and Xu, 2000; Xu, 2008). Several genetic maps have been constructed for diploid alfalfa using partially inbred mapping populations (F_2 or backcross) derived from intra-subspecies and inter-subspeices crosses

(Brummer et al. 1993; Echt et al. 1994; Kaló et al. 2000; Kiss et al. 1993). Serious segregation distortion (18-54%) was found in those studies. Interestingly, distorted markers were usually found clustered and segregation distortion with heterozygote excess was commonly observed. A biological explanation for this extreme distortion is that alfalfa carries a high level of deleterious recessive alleles due to its outcrossing and self-incompatible nature, and that gametic and/or (perhaps especially) zygotic selection against deleterious recessive alleles during inbreeding causes skewed ratios of marker genotypes. If the deleterious recessive alleles at different loci are linked in repulsion phase, then excessive heterozygotes at linked marker loci would be evident because either homozygote would be linked to a different deleterious recessive allele.

In a non-inbred diploid alfalfa mapping population (F_1 population), only 9% of mapped marker loci showed significant segregation distortion (SD), which supports the above explanation (Tavoletti et al., 1996). Other supporting evidence is that relatively low segregation distortion (4-9% of mapped alleles) was observed in a tetraploid backcross mapping population, because of "the greater buffering capacity of autotetraploids against the effects of deleterious recessive alleles" (Brouwer and Osborn, 1999). However, other F_1 tetraploid alfalfa mapping populations showed substantially more distortion; in one, 35% of the AFLP and 25% SSR markers were significantly distorted (Julier et al., 2003), and in another, about 32% markers exhibited SD (Robins et al., 2007). Different locations and numbers of segregation distortion loci (SDLs) in different populations could be one possible explanation for the variable SD levels observed. Different SDLs have been reported in different populations, such as in maize (Lu et al., 2002) and rice (Xu et al., 1997). The genetic distance between parents also has an effect on SD level among different mapping populations (Chetelat et al., 2000). In addition, non-biological factors such as environment (especially conditions during seed production, or interaction

between environment and viability genes), sample size, genotyping errors, statistical testing methodology, etc. affect estimation of SD. Thus, analysis of segregation distortion in comparable mapping populations in one single study grown under similar environmental conditions, such as F_1 populations that share a common parent or F_1 and F_2 populations derived from same parent pair, may be helpful for understanding the mechanisms of segregation distortion.

The consistency with which segregation distortion is observed in the alfalfa genome and the potential relationship of SD to inbreeding depression and heterosis suggest that investigation into the mechanism of segregation distortion is needed. The early genetic maps developed in alfalfa were all in diploid populations, for which mapping is more tractable than tetraploids, but the maps were primarily constructed with RFLP and RAPD markers, and have not been integrated into a single, reference map. Newer maps in tetraploid populations have used SSR markers, and SNP markers are now under development. Alfalfa suffers from not having a single framework map on which most available genetic markers are mapped that can serve as a reference for other mapping projects.

STRUCTURE OF DISSERTATION RESEARCH

My dissertation will investigate the genetic basis of heterosis and inbreeding depression from several perspectives. First, using classical quantitative genetics, I assessed the importance of dominance, over-dominance, epistasis, and multiple allelic interactions on biomass yield across generations (Chapter 2). Second, I used Affymetrix gene arrays to compare gene expression profiles between heterotic and nonheterotic hybrids (Chapter 3). Third, I developed one F_1 and one F_2 population, which are derived from same two parents, to assess segregation

distortion patterns and investigate the effect of SD on mapping (Chapter 4). Finally, I offer brief conclusions of all three experiments (Chapter 5).

REFERENCES

Barnes, D.K., E.T. Bingham, R.P. Murphy, O.J. Hunt, D.F. Beard, W.h. Skrdla, and L.R. Teuber 1977. Alfalfa germplasm in the United States: Genetic vulnerability, use, improvement, and maintenance. 1571. U.S. Gov. Print. Office, Washington, D.C.

Bhandari, H.S., C.A. Pierce, L.W. Murray, and I.M. Ray. 2007. Combining Abilities and Heterosis for Forage Yield among High-Yielding Accessions of the Alfalfa Core Collection Crop Sci. 47:665-671.

Bingham, E.T., R.W. Groose, D.R. Woodfield, and K.K. Kidwell. 1994. Complementary gene interaction in alfalfa is greater in autotetraploids than diploid. Crop Sci. 34:823-829.

Birchler, J.A., R.K. Dawe, and J.F. Doebley. 2003. Marcus rhoades, preferential segregation and meiotic drive. Genetics 164:835-41.

Brem, R.B., G. Yvert, R. Clinton, and L. Kruglyak. 2002. Genetic dissection of transcriptional regulation in budding yeast. Science 296:752-5.

Brouwer, D.J., and T.C. Osborn. 1999. A molecular marker linkage map of tetraploid alfalfa (Medicago sativa L.) Theor Appl Genet 99:1194:1200.

Chetelat, R.T., V. Meglic, and P. Cisneros. 2000. A genetic map of tomato based on BC(1) Lycopersicon esculentum x Solanum lycopersicoides reveals overall synteny but suppressed recombination between these homeologous genomes. Genetics 154:857-67.

Cui, X., J. Affourtit, K.R. Shockley, Y. Woo, and G.A. Churchill. 2006. Inheritance patterns of transcript levels in F1 hybrid mice. Genetics 174:627-37.

Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to Quantitative Genetics. 4 ed., Longman, Essex, England.

Jones, J.S., and E.T. Bingham. 1995. Inbreeding depression in alfalfa and cross-pollinated crops. John Wiley & Sons, New York.

Julier, B., S. Flajoulot, P. Barre, G. Cardinet, S. Santoni, T. Huguet, and C. Huyghe. 2003. Construction of two genetic linkage maps in cultivated tetraploid alfalfa (Medicago sativa) using microsatellite and AFLP markers. BMC Plant Biol 3:9.

Kimbeng, C.A., and E.T. Bingham. 1998. Population improvement in alfalfa: fertility and S1 forage yield performance in original and improved populations. Crop Sci. 37:1509-1513.

Kliebenstein, D.J., M.A. West, H. van Leeuwen, K. Kim, R.W. Doerge, R.W. Michelmore, and D.A. St Clair. 2006. Genomic survey of gene expression diversity in Arabidopsis thaliana. Genetics 172:1179-89.

Lorieux, M., X. Perrier, B. Goffinet, C. Lanaud, and D.G.d. León. 1995a. Maximum-likelihood models for mapping genetic markers showing segregation distortion. 1. Backcross populations Theor Appl Genet 90:73-80.

Lorieux, M., X. Perrier, B. Goffinet, C. Lanaud, and D.G.d. León. 1995b. Maximum-likelihood models for mapping genetic markers showing segregation distortion. 2. F2 populations Theor Appl Genet 90:81-89.

Lu, H., J. Romero-Severson, and R. Bernardo. 2002. Chromosomal regions associated with segregation distortion in maize. Theor Appl Genet 105:622-628.

Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA. Chapter 9:222-226.

Quiros, C.F., and G.R. Bauchan. 1988. The Genus *Medicago* and the origin of the *Medicago sativa* complex, *In* A. A. Hanson, ed. Alfalfa and alfalfa improvement. American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc., Madison, Wisconsin, USA.

Riday, H., and E.C. Brummer. 2002. Forage Yield Heterosis in Alfalfa. Crop Sci 42:716-723.

Riday, H., and E.C. Brummer. 2005. Heterosis in a Broad Range of Alfalfa Germplasm. Crop Science 45:8-17.

Robins, J.G., D. Luth, T.A. Campbell, G.R. Bauchan, C. He, D.R. Viands, J.L. Hansen, and E.C. Brummer. 2007. Genetic Mapping of Biomass Production in Tetraploid Alfalfa Crop Sciences 47:1-10.

Sakiroglu, M., and E.C. Brummer. 2007. Little Heterosis between Alfalfa Populations Derived from the Midwestern and Southwestern United States Crop Sci 47:2364-2371.

Sriwatanapongse, S., and C.P. Wilsie. 1968. Intra- and intervariety crosses of Medicag sativa L. and Medicago falcata L. . Crop Sci. 8:465-466.

Stupar, R.M., and N.M. Springer. 2006. Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid. Genetics 173:2199-210.

Swanson-Wagner, R.A., Y. Jia, R. DeCook, L.A. Borsuk, D. Nettleton, and P.S. Schnable. 2006. All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. Proc Natl Acad Sci U S A 103:6805-10.

Tavoletti, S., F. Veronesi, and T.C. Osborn. 1996. RFLP Linkage Map of an Alfalfa Meiotic Mutant Based on an F1 Population Journal of Heredity 87:167:170.

Templeton, A.R. 1986. Coadaptation and outbreeding depression. In M. E. Soule (ed.), Conservation biology: the science of scarcity and diversity, pp. 105-116. Sinauer Assoc., Sunderland, MA.

Vogl, C., and S. Xu. 2000. Multipoint mapping of viability and segregation distorting loci using molecular markers. Genetics 155:1439-47.

Vuylsteke, M., F. van Eeuwijk, P. Van Hummelen, M. Kuiper, and M. Zabeau. 2005. Genetic analysis of variation in gene expression in Arabidopsis thaliana. Genetics 171:1267-75.

Wang, J., L. Tian, H.S. Lee, N.E. Wei, H. Jiang, B. Watson, A. Madlung, T.C. Osborn, R.W. Doerge, L. Comai, and Z.J. Chen. 2006. Genomewide nonadditive gene regulation in Arabidopsis allotetraploids. Genetics 172:507-17.

Woodfield, D.R., and E.T. Bingham. 1995. Improvement in two-allele autotetraploid populations of alfalfa explained by accumulation of favorable alleles. Crop Sci. 35:988-994.

Xu, S. 2008. Quantitative trait locus mapping can benefit from segregation distortion. Genetics 180:2201-8.

Xu, Y., L. Zhu, J. Xiao, N. Huang, and S.R. McCouch. 1997. Chromosomal regions associated with segregation distortion of molecular markers in F2, backcross, doubled haploid, and recombinant inbred populations in rice (Oryza sativa L.). Mol Gen Genet 253:535-45.

CHAPTER 2

INBREEDING DEPRESSION FOR FERTILITY AND BIOMASS IN ADVANCED GENERATIONS OF INTER- AND INTRA-SUBSPECIES HYBRIDS OF TETRAPLOID ALFALFA[1]

[1] Xuehui Li and E. Charles Brummer. 2009. *Crop Science*. 49:13-19.
Reprinted here with permission of publisher.

ABSTRACT

Hybrids between *Medicago sativa* subsp. *sativa* and *M. sativa* subsp. *falcata* often express heterosis for yield. Yield generally declines when hybrids are inbred due to the loss of desirable dominant alleles or epistatic combinations. We hypothesized that inbreeding sativa-falcata hybrids may show more extensive yield loss than intra-subspecific crosses due to the large genetic difference between subspecies. Our objective was to compare fertility in the greenhouse and biomass production in the field of F1 hybrids and their F2 and S1 generations in both intra- and inter-subspecies crosses. Field experiments including F1, F2, S1, and double cross populations were planted in two locations in Iowa in 2003 and evaluated for yield in two years. Sativa x falcata crosses (SFC) showed greater F1 fertility than intra-subspecies crosses suggesting that the two subspecies have similar chromosome structure. The SFC did not show high-parent F1 heterosis probably due to an intensive harvest schedule favoring sativa x sativa hybrids. However, SFC showed greater inbreeding depression in both F2 and S1 generations than intra-subspecific crosses, possibly reflecting a greater loss of favorable complementary gene interactions. These results suggest that selection within each parental population followed by intercrossing to produce population hybrids may be a better way to improve biomass with this germplasm than advancing hybrids into a recurrent selection program.

INTRODUCTION

Alfalfa (*Medicago sativa*) is one of the most important forage crops in the world; however, yield increases have lagged behind many other crops (Riday and Brummer, 2002; Lamb et al., 2006). One cause of the limited yield improvement is possibly that alfalfa breeders have focused selection efforts more on insect and disease resistance selection than on yield *per*

se (Barnes et al., 1977; Hill, 1983; Lamb et al., 2006). A potentially effective way to improve yield in the future is to capitalize on non-additive gene action by harnessing heterosis. Alfalfa breeders and researchers have attempted to develop inbred lines to use as parents of hybrids since the 1930s, but because of severe inbreeding depression, inbred lines are difficult or impossible to produce. A semi-hybrid breeding strategy based on population crosses would avoid the need for inbred lines and may capture some heterosis (Brummer, 1999). Population hybrids between *M. sativa* subsp. *sativa* and subsp. *falcata* show heterosis for biomass production (Riday and Brummer, 2002; Riday and Brummer, 2005; Sriwatanapongse and Wilsie, 1968), suggesting possible heterotic groups within alfalfa germplasm.

Most genetic knowledge about heterosis in alfalfa has been inferred from inbreeding studies. Generally, inbreeding depression is considered as the converse of heterosis, although the genetic mechanism could be somewhat different (Ritland, 1996). Inbreeding depression in tetraploid alfalfa is more severe than predicted based solely on the decrease in heterozygosity, suggesting that it is due to the loss of multiple-allele interactions (Busbice and Wilsie, 1966) or more likely, the loss of favorable dominant alleles linked in repulsion phase and/or of desirable epistatic combinations of alleles (Bingham et al., 1994; Kimbeng and Bingham, 1998; Woodfield and Bingham, 1995). The latter explanation essentially considers the genome to consist of blocks of linked loci, called linkats (Demarly, 1979). For convenience in this paper, we will use ‘allele’ interchangeably with ‘linkat’.

The striking differences in morphology, geographical distribution, and phenology between sativa and falcata suggest that more favorable complementary gene interactions (Bingham et al., 1994) would accumulate in their F1 hybrid than in hybrids derived from intra-subspecies crosses, generating heterosis, but also resulting in larger inbreeding depression in F2

or later generations if the hybrid were advanced, causing greater biomass yield loss. In particular, we expected that favorable epistatic allele combinations likely exist in each heterotic group. While no deleterious effect would be observed in the F1, additional generations would lead to disruptions of these coadapted gene complexes that could produce greater yield loss than similar advanced generations in intra-subspecies crosses. If inter-subspecies crosses have greater depression in advanced generations of F1 hybrids, then advancing hybrid populations into a traditional recurrent selection program would be unadvisable. A better approach for germplasm maintenance and for repeated capitalization on heterosis would be to keep the parental germplasms separate and only produce hybrids between them as the final step in the breeding process, as suggested by Brummer (1999).

Progressive heterosis occurs if double cross hybrids have higher yield than single cross hybrids. Because tetraploids can accumulate more favorable alleles at one locus and have a greater ability to accumulate at least one favorable allele at each of several different loci compared to diploids, maximum heterosis may not be reached in a single cross hybrid, as it is in diploids, but may be progressive, where the maximum might not occur until the double cross (or even later) generation, depending on the inbreeding levels of the parents (Bingham et al., 1994).

The primary objectives of this experiment were to test the hypotheses (1) that advanced generations of inter-subspecies alfalfa crosses show greater depression for fertility and biomass than intra-subspecies hybrids, and (2) that multiple allelic interactions and/or epistasis contribute to biomass production, and (3) that progressive heterosis is observed in alfalfa germplasm.

MATERIALS AND METHODS

Plant materials

We sampled 4 elite sativa genotypes (ABI314, ABI408, C96-513 and RP-93-377) and 4 wild or semi-improved falcata genotypes (WISFAL4, WISFAL6, C25-6 and PI502453-1) as parents. These genotypes are the same as those used by Riday and Brummer (2002) and complete details of their origins can be found in that paper.

We made eight intra-subspecies crosses (four sativa×sativa crosses [SSC] and four falcata×falcata crosses [FFC]) and four inter-subspecies sativa×falcata crosses (SFC) (Table 1). About 20 plants from each F1 hybrid population were sib-pollinated (~100 florets per plant) to obtain the F2 generation and each plant was also self-pollinated (~250 florets per plant) to obtain the S1 generation. Only florets located on the main stems were pollinated. All sib- and self-pollinations were done under the same environmental conditions (25/18 °C, day/night temperature) in the Iowa State University agronomy greenhouse in 2003. Self and sib fertility data (pods per floret tripped, seeds per pod, seeds per floret tripped, and weight per seed) were collected on each F1 plant. The 12 F1 hybrid populations were reciprocally intercrossed (with 20 plants per F1 hybrid population) to produce 11 double-cross (DC) hybrids: 4 sativa×sativa double crosses, 3 falcata×falcata double crosses and 4 sativa×falcata double crosses (Table 1). Each parent was also self-pollinated to produce the PS1 generation. About 20 plants from each PS1 population were sib-pollinated to obtain the PF2 generation. The falcata parents exhibited poor fertility, so only the four sativa parents produced enough seeds for the field experiment. The seeds of each generation of each cross were bulked equally across genotypes for the field experiment.

Field experiment

The F1, DC, F2, and S1 generations for all 12 crosses, 2 check cultivars ('5454' and 'Vernal'), 4 PF2 populations, and some additional F1 hybrid populations (Table 1) for a total of 64 entries, were planted in the field at the Agronomy and Agricultural Engineering Research Farm west of Ames, IA, in a Nicollet loam soil (fine-loamy, mixed, super-active, mesic Aquic Hapludolls) on 18 Aug. 2003 and at the Northeast Research Farm south of Nashua, IA, in a Readlyn loam (fine-loamy, mixed, mesic Aquic Hapludolls) on 23 Aug. 2003. The experimental design was an 8×8 lattice with 3 replications at each location. Each entry was planted in one-row plots, 3 m long and seeded at a rate of 75 seeds per plot. Plots were spaced 60 cm apart within rows, with rows spaced 75 cm apart.

Harvests for biomass were taken 4 times at both locations in 2004 and 2005 (May 31, June 29, July 28, and September 3, 2004 and May 20, June 20, July 25, and August 25, 2005 at Ames; June 4, July 7, August 2, and September 5, 2004 and May 31, June 30, August 1, and August 29, 2005 at Nashua). In 2004, each plot was hand-harvested. The forage was dried for 5 days at 60° C in a forced-air dryer and then weighed to get the whole plot dry biomass. In 2005, plots were machine-harvested at each harvest and weighed wet. Six subsamples were randomly taken, weighted wet, dried for 5 days at 60° C in a forced-air dryer and then weighed dry. An average dry matter percentage was calculated and used to derive the whole plot dry biomass. The yearly dry biomass of each plot is the sum of the whole plot dry biomass of the four harvests each year. Stand was scored during the first week after each harvest by estimating the percentage of each row that was occupied by plants. The mean stand score for each plot was computed as the average of the individual stand scores after each harvest each year. To correct

for variations in stand, the yearly dry biomass per plot was calculated by dividing the yearly dry biomass of each plot by its mean stand score.

Statistical analysis

Fertility data were analyzed using the GLM procedure of the SAS software package (SAS, 2000). Least square means were computed for cross types, and compared using Fisher's Least Significant Difference (LSD). Cross type and fertility type (sib- or self-fertility) were considered to be fixed effects, and entry nested within cross type was considered to be a random effect.

Yearly dry matter biomass (adjusted for stand as described above) was analyzed using the MIXED model from SAS (SAS, 2000). Year, location, year × location interaction, the interaction of all three with entry, replication, and block were considered as random effects; entry was considered a fixed effect. Least square means were estimated for entry in each year – location combination separately due to significant year × entry and location × entry interactions. The estimated least square means for the 12 crosses for each year – location combination were fitted to the following model:

$$y_{ijk} = \mu + T_i + C_{j(i)} + G_k + TG_{ik} + \varepsilon_{ijk}$$

where y_{ijk} is the estimated least square mean of biomass for a given entry, μ is the overall mean, T_i is the cross type (FFC, SSC, or SFC) effect, $C_{j(i)}$ is the entry (i.e., an individual cross between two parental genotypes) effect within each cross type, G_k is the generation (F1, DC, F2 and S1) effect, TG_{ik} is the interaction effect between cross type and generation, and ε_{ijk} is the random error associated with *i*th cross type, *j*th cross and *k*th generation. Only entry was considered to be a random effect. Least square means were estimated for cross type at each

generation. F2 depression was measured as the deviation of the F2 value from the F1 hybrid value; S1 depression was measured analogously as the difference between the S1 and F1 values. F2 and S1 depression were measured by making contrasts between generations, implemented with the ESTIMATE command in SAS (SAS, 2000). Double-cross progressive heterosis was measured as the deviation of the double-cross generation from the F1 parental mean value, and statistical significance determined using contrasts as for depression. Differences among cross types for yield at each generation were evaluated when F-tests were significant by making pairwise contrasts using the ESTIMATE command of SAS. Differences in yield between generations within a given cross type were evaluated by contrasts using the ESTIMATE command of SAS (SAS, 2000). The comparison among cross types was possible because the mean values of the different entries within each cross types served as replications. For the double crosses, we did not multiple entries within each cross type, and the mean separation was not conducted.

The four SSC entries had data from four generations (F1, F2, S1, and PF2), so their yearly biomass means were fit to a set of linear models to determine the effect of multiple allele interactions on biomass production (The falcata parents did not produce enough viable plants to provide sufficient seed for PF2 testing.) Because of missing values for some entries at Nashua, only data from Ames were used for this analysis. The equations are as follows:

$$\text{Biomass} = \text{Year} + R \qquad \text{(M1)}$$

$$\text{Biomass} = \text{Year} + R + K \qquad \text{(M2)}$$

$$\text{Biomass} = \text{Year} + \text{Generation} \qquad \text{(M3)}$$

where year was a random effect and the other variables were fixed. R represents the proportion of two-allele interactions present within plants in each generation, assuming that the

parents were non-inbred and unrelated. Under this assumption, the maximum number of two-allele interactions would be present in the F1 populations, and plants in the succeeding inbred generations would have fewer interactions as they approached homozygosity. Similarly, K is proportion of possible three-allele interactions possible at each generation. The values of R and K were calculated using formulas developed by Busbice and Wilsie (1966) and Busbice (1969). For the four generations evaluated in this experiment, these values are R=1.0 and K=1.0 for F1; R=0.917 and K=0.750 for F2; R=0.833 and K=0.500 for S1, and R=0.779 and K=0.417 for PF2. F-tests were conducted sequentially to test for the presence of an effect due to two- or three-allele interactions on biomass yield. A lack-of-fit test between models M2 and M3 was used to evaluate the presence of a combined effect due to four-allele interaction and epistasis.

Unless otherwise indicated, statistical significance of results was assessed at the 5% probability level.

RESULTS AND DISCUSSION

Fertility in advanced generations

In the overall analysis, the interaction of cross type with fertility type was present for all the traits (data not shown), so further analyses were conducted separately for sib- and self-fertility. Among cross types, both sib and self fertility of SFC was either higher than or equal to the higher intra-subspecies hybrid for pods per floret, seeds per pod, and seeds per floret (Table 2). However, SSC showed the greatest value for weight per seed, with FFC the lowest and SFC intermediate based on both sib- and self-pollination (Table 2). Thus, these results clearly show that fertility of inter-subspecies hybrids do not manifest any serious fertility barriers.

Differences in the content of constitutive heterochromatin identified by C-banding has been noted between diploid *M. sativa* subsp. *caerulea* (the putative ancestor of cultivated alfalfa, which has similar C-banding patterns to subsp. *caerulea* [Bauchan and Hossain, 2001]) and *M. sativa* subsp. *falcata* (Bauchan and Hossain, 1997). Thus, while these two results are limited in their scope, they suggested that chromosomal differences could exist between subspecies and that these could affect the progeny of SFC. If these chromosome variations or rearrangements interfered with normal meiosis of the F1 hybrid of SFC, they could cause the fertility and/or biomass of later generations to break down because chromosomal aberrations would be likely. However, in our experiment, the fertility of SFC hybrids did not break down for either self- or sib-pollination, and on the contrary, SFC showed greater fertility than SSC or FFC. A previous analysis of diploid inter-species hybrids between sativa and falcata showed that they displayed cytologically normal meiosis and fertility data of parents and hybrids suggested that the three species belonged to one polymorphic species (Sprague, 1959). Our fertility data on tetraploid sativa x falcata hybrids gave the similar results. Collectively, our results suggest the absence of agronomically undesirable chromosomal variation between subspecies.

Biomass yield depression in advanced generations

Because of interactions among entry, year, and location (data not shown), we analyzed the biomass data separately for each year and location combination, which we term 'environments' for convenience. We had previously shown high parent heterosis of SFC in an experiment that included the same germplasm we used here (Riday and Brummer, 2002). In this experiment, the yearly F1 biomass production of SFC was equal to SSC in both years at Nashua, but lower than SSC in both years at Ames (Table 3). The contrast between our experiment and

that of Riday and Brummer (2002) could be due to the different harvest managements in the two experiments. The former experiment harvested plants three times per year, whereas we harvested four times per year. SFC hybrids exhibited a faster growth rate at later growth stages (Riday and Brummer, 2004), which makes a less frequent harvest schedule advantageous for the display of heterosis in SFC. Indeed, results of individual harvests in the current experiment showed that SFC performed as well as or better than SSC for first harvest yield, but underperformed in successive harvests (data not shown). Like the previous experiment, the FFC yield was either the lowest or among the lowest (Table 3).

For the two inbred generations (F2 and S1), SSC yields were among the highest and FFC among the lowest in all environments (Table 3). The yearly biomass of the F2 generation of SFC was less than SSC except at Nashua in 2004, where the SFC F2 generation yielded similarly to SSC. The SFC F2 biomass was greater than FFC except at Nashua in 2004, where the SFC F2 biomass was equal to FFC (Table 3). In the S1 generation, SSC had the highest yields, SFC were intermediate, and FFC had the lowest yields in Ames, but at Nashua in both years, SFC and FFC had lower yields than SSC (Table 3).

The yield depression of advanced generations showed somewhat different results than the yield *per se* of each generation. The depression between the F1 and F2 generations was similar among cross types in 2004. However, in 2005, the SFC decline was greater than either of the other two types, which were equal, at Ames, and depression at Nashua was greater in SFC than in SSC (Table 4). For S1 yield depression, SFC always had among the largest declines, except at Nashua in 2005, where no difference among crosses were noted (Table 4). Thus, when taken collectively, the data suggest that SFC hybrids showed greater yield loss when advanced to successive generations than either SSC or FFC.

Serious inbreeding depression for biomass has been repeatedly documented in alfalfa (Busbice and Wilsie, 1966; Gallais, 1984; Jones and Bingham, 1995; Kimbeng and Bingham, 1998). In this experiment, we showed similar results, with the S1 generation of all three cross types and the F2 generation of SFC showing biomass depression across all environments. However, the F2 generation of SSC and FFC showed biomass depression only in some environments. Of the three cross types, the depression level for SFC was always the greatest or among the greatest. The greater depression of SFC could be due to the greater loss of allelic (or linkat) interactions (Bingham et al., 1994). Even though SSC had higher yields in some situations than SFC, the depression observed following inbreeding was less than that seen for SFC.

Progressive heterosis

Double crosses of SSC resulted in higher yields than SFC or FFC in all environments except Nashua 2005 (Table 3). SFC double crosses were generally superior to FFC but inferior to SSC. In terms of yield differences from the single cross (F1) yield, the SSC double crosses showed yield improvement, or progressive heterosis, at Ames in both years (Table 4). No progressive heterosis was seen for the other cross types, whose DC yields were not different from the F1 yields, although all were numerically trending lower.

Tetrasomic tetraploids may exhibit progressive heterosis, where the maximum performance is not observed until the double cross or later generations, but expression depends on the parents being largely inbred (Dudley, 1964). Progressive heterosis has been observed in both unimproved and improved two-allele tetraploid sativa populations (Groose et al., 1989), and our results suggest it can be generated even among elite tetraploid alfalfa individuals. This result

suggests that the sativa genotypes used here, all of which were derived from commercial breeding programs, must be partially inbred (Dudley, 1964). By contrast, the falcata genotypes, which did not derive from highly selected, commercially viable germplasm pools, appear to be highly heterozygous, preventing the expression of progressive heterosis in either FFC or SFC double crosses. We did not find progressive heterosis in double crosses between SFC in this experiment. Therefore, double cross hybrids may be worth further investigation in sativa germplasm, but do not appear sensible for sativa x falcata hybrids.

Multiple allelic interactions and epistasis

We fit the yearly biomass data from the four SSC crosses, their respective advanced generations, and the F2 generations of their parents grown at Ames to a set of linear models designed to determine the effect of the allelic interactions on biomass yield based on methods outlined by Busbice and Wilsie (1966). F-tests were constructed between full and reduced models to test for the presence of multiple allelic interactions and/or epistatic effects. The results indicated that an effect due to two-allele interactions was significant for all crosses (Table 5). This suggests that dominance is an important factor in biomass yield. However, only two of the four crosses showed an effect of three allele interactions, and just one showed a significant lack-of-fit between the model consisting of two- and three-allele interactions and one consisting solely of generations, denoting an effect on yield due to four allele interactions and/or epistasis in sativa x sativa hybrid populations. This result reinforces our suggestion above that alfalfa genotypes are partially inbred and perhaps, related.

Considerations for using falcata germplasm

Crosses between different species or distantly related populations frequently leads to substantial loss of vigor, a phenomenon called outbreeding depression (Lynch and Walsh, 1998). In some situations, outbreeding depression is not observed in the F1, which may show heterosis, but in the F2 or later generations, in which the vigor is significantly less than the mean of original parents (Lynch and Walsh, 1998). Generally, dominance is considered the predominant cause of heterosis and inbreeding depression, but outbreeding depression is probably due to the loss of favorable epistatic effects (co-adapted gene complexes) present in the parental lines (Lynch and Walsh, 1998). The higher yield depression observed in SFC compared to SSC or FFC suggests that outbreeding depression may be present, but since we could not include the parental genotypes in the experiment, we could not test this hypothesis directly.

The potential value of falcata germplasm to improve biomass yield in alfalfa has been demonstrated in crosses with elite sativa genotypes (Riday and Brummer, 2002; Riday and Brummer, 2005; Segovia-Lerma et al., 2004). Further, genetic mapping has identified quantitative trait locus (QTL) alleles for yield from falcata that are complementary to sativa (Robins et al., 2007). The results of the present experiment offer some ideas on how to best use falcata in germplasm improvement. Sativa and falcata germplasm may have different suites of favorable alleles, or coadapted gene complexes, which perform well in hybrids between the germplasms but which are disrupted upon further generation advance. Because of this, the direct introduction of falcata germplasm into a primarily sativa-based program may not be the best approach to using this germplasm.

Instead, falcata germplasm could be used in two ways to improve alfalfa. First, marker-assisted introgression would help incorporate individual falcata alleles into cultivated germplasm

without bringing in large genomic regions that could be disruptive to positive epistatic combinations within sativa germplasm, which could largely remain intact. Second, the easy loss of the complementary favorable alleles from falcata germplasm that could occur during a wholesale introgression and selection program (Simmonds, 1993) or the disruption of desirable epistatic combinations of alleles within each of the subspecies' genomes could be avoided by improving sativa and falcata separately, and using them to produce hybrid or semihybrid cultivars but not advancing the hybrids further in the breeding program (Brummer, 1999).

However, although we indicate the presence of a larger yield penalty from in advanced generations of SFC than in SSC, the extra yield depression may not be severe enough to unduly penalize a breeder who wished to advance sativa x falcata hybrids further. Practical experience with a given germplasm pool will indicate the best approach.

REFERENCES

Barnes, D.K., E.T. Bingham, R.P. Murphy, O.J. Hunt, D.F. Beard, W.H. Skrdla, and L.R. Teuber. 1977. Alfalfa germplasm in the United States: Genetic vulnerability, use, improvement, and maintenance. 1571. U.S. Gov. Print. Office, Washington, D.C.

Bauchan, G.R., and M.A. Hossain. 1997. Karyotypic analysis of C-banded chromosomes of diploid alfalfa: *Medicago sativa* ssp. *caerulea* and spp. *falcata* and their hybrid. J. Hered. 88:533-537.

Bauchan, G.R. and M.A. Hossain. 2001. Distribution and characterization of heterochromatic DNA in the tetraploid African population alfalfa genome. Crop Sci 41:1921-1926.

Bingham, E.T., R.W. Groose, D.R. Woodfield, and K.K. Kidwell. 1994. Complementary gene interaction in alfalfa is greater in autotetraploids than diploids. Crop Sci. 34:823-829.

Brummer, E.C. 1999. Capturing heterosis in forage crop cultivar development. Crop Sci. 39:943-954.

Busbice, T.H. 1969. Inbreeding in synthetic varieties. Crop Sci. 9:601-604.

Busbice, T.H. and C.P. Wilsie. 1966. Inbreeding depression and heterosis in autotetraploids with application to *Medicago sativa* L. Euphytica 15:52-67.

Demarly, Y. 1979. The concept of linkat. p. 257-265. *In* A.C. Zeven and A.M. van Harten (ed.) Proc. Conference Broadening Genetic Base of Crops. Centre for Agricultural Publishing and Documentation, Wageningen, Netherlands. Pudoc, Wageningen.

Dudley, J.W. 1964. A genetic evaluation of methods of utilizing heterozygosis and dominance in autotetraploids. Crop Sci 4:410-413.

Gallais, A. 1984. An analysis of heterosis vs. inbreeding effects with an autotetraploid cross-fertilized plant: *Medicago sativa* L. Genetics 106:123-137.

Groose, R.W., L.E. Talbert, W.P. Kojis, and E.T. Bingham. 1989. Progressive heterosis in autotetraploid alfalfa: studies using two types of inbreds. Crop Sci 29:1173-1177.

Hill, R.R., Jr. 1983. Heterosis in population crosses of alfalfa. Crop Sci. 23:48:50.

Jones, J.S. and E.T. Bingham. 1995. Inbreeding depression in alfalfa and cross-pollinated crops. Plant Breed. Rev. 13:209-233.

Kimbeng, C.A., and E.T. Bingham. 1998. Population improvement in alfalfa: fertility and S1 forage yield performance in original and improved populations. Crop Sci 38:1509-1513.

Lamb, J.F.S., C.C. Sheaffer, L.H. Rhodes, R.M. Sulc, D.J. Undersander, and E.C. Brummer. 2006. Five decades of alfalfa cultivar improvement: impact on forage yield, persistence, and nutritive value. Crop Sci. 46:902:909.

Lynch, M. and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA.

Riday, H. and E.C. Brummer. 2002. Forage yield heterosis in alfalfa. Crop Sci 42:716-723.

Riday, H. and E.C. Brummer. 2004. Morphological variation of *Medicago sativa* subsp. *falcata* genotypes and their hybrid progeny. Euphytica 138:1-12.

Riday, H. and E.C. Brummer. 2005. Heterosis in a broad range of alfalfa germplasm. Crop Sci 45:8-17.

Ritland, K. 1996. Inferring the genetic basis of inbreeding depression in plants. Genome 39:1-8.

Robins, J.G., D. Luth, T.A. Campbell, G.R. Bauchan, C. He, D.R. Viands, J.L. Hansen, and E.C. Brummer. 2007. Genetic mapping of biomass production in tetraploid alfalfa. Crop Sci 47:1-10.

SAS Institute. 2000. SAS language and procedure: Usage. Release 8.0, 1st ed. SAS institute, Cury, NC.

Segovia-Lerma, A., L.W. Murray, M.S. Townsend, and I.M. Ray. 2004. Population-based diallel analyses among nine historically recognized alfalfa germplasms. Theor. Appl. Genet. 109:1568-75.

Simmonds, N.W. 1993. Introgression and incorporation. Strategies for the use of crop genetic resources. Biological Reviews 68:539-562.

Sprague, E.W. 1959. Cytological and fertility relationships of *Medicago sativa*, *Medicago falcata*, and *Medicago gaetula*. Agron. J. 51:249-252.

Sriwatanapongse, S., and C.P. Wilsie. 1968. Intra- and intervariety crosses of *Medicago sativa* L. and *Medicago falcata* L. Crop Sci. 8:465-466.

Woodfield, D.R., and E.T. Bingham. 1995. Improvement in two-allele autotetraploid populations of alfalfa explained by accumulation of favorable alleles. Crop Sci. 35:988-994.

Table 2.1. Cross types, names, and pedigrees of materials used in the experiment.

Cross Type	Cross Name	Pedigree
FFC	HRMS10	WISFAL-4×WISFAL-6
FFC	HRMS11	WISFAL-4×C25-6
FFC	HRMS12	C25-6×PI502453-1
FFC	HRMS04	WISFAL-6×PI502453-1
SFC	HRMS13	C96-513×C25-6
SFC	HRMS14	ABI408×WISFAL-6
SFC	HRMS15	ABI314×WISFAL-4
SFC	HRMS03	RP-93-377×PI502453-1
SSC	HRMS06	C96-513×RP-93-377
SSC	HRMS07	ABI408×RP-93-377
SSC	HRMS08	ABI314×C96-513
SSC	HRMS09	ABI314×ABI408
FFDC	HRMS10×12	(WISFAL-4×WISFAL-6) × (C25-6×PI502453-1)
FFDC	HRMS12×10	(C25-6×PI502453-1)× (WISFAL-4×WISFAL-6)
FFDC	HRMS11×04	(WISFAL-4×C25-6) × (WISFAL-6×PI502453-1)
SFDC	HRMS13×14	(C96-513×C25-6) × (ABI408×WISFAL-6)
SFDC	HRMS14×13	(ABI408×WISFAL-6) × (C96-513×C25-6)
SFDC	HRMS03×15	(RP-93-377×PI502453-1) × (ABI314×WISFAL-4)
SFDC	HRMS15×03	(ABI314×WISFAL-4) × (RP-93-377×PI502453-1)
SSDC	HRMS06×09	(C96-513×RP-93-377) × (ABI314×ABI408)
SSDC	HRMS09×06	(ABI314×ABI408) × (C96-513×RP-93-377)
SSDC	HRMS07×08	(ABI408×RP-93-377) × (ABI314×C96-513)
SSDC	HRMS08×07	(ABI314×C96-513) × (ABI408×RP-93-377)
FFC	HRMS17	WISFAL-6×C25-6
SFC	HRMS01	C96-513×WISFAL-6
SSC	HRMS20	ABI408×C96-513

† SSC = sativa x sativa crosses; SFC = sativa x falcata crosses; FFC = falcata x falcata crosses; DC=Double cross.

Table 2.2. Self- and sib-fertility for three groups of alfalfa hybrids based on hand pollinations in the greenhouse at Ames, IA.

Cross type†	Pods per floret		Seeds per pod		Seeds per floret		Weight per seed	
	Self	Sib	Self	Sib	Self	Sib	Self	Sib
	no.		no.		no.		mg	
FFC	0.30 a‡	0.42 c	1.28 b	1.92 b	0.43 ab	0.85 c	1.96 c	2.29 c
SFC	0.31 a	0.56 a	1.57 a	2.89 a	0.55 a	1.79 a	2.72 b	2.58 b
SSC	0.24 b	0.50 b	1.37 b	2.97 a	0.38 b	1.58 b	3.36 a	3.02 a

† SSC = sativa x sativa crosses; SFC = sativa x falcata crosses; FFC = falcata x falcata crosses

‡Values within columns followed by different letters are significantly different at α=0.05.

Table 2.3. Biomass yield of F1, F2, S1 and DC generations for the three alfalfa cross types grown at two Iowa locations for two years.

Location	Cross type†	2004				2005			
		F1	F2	S1	DC	F1	F2	S1	DC
		g plot^{-1}							
Ames	SSC	1801 a‡	1581 a	1378 a	2062 a	4056 a	3822 a	2918 a	4492 a
	SFC	1616 b	1335 b	1044 b	1540 b	3353 b	2675 b	2221 b	3100 b
	FFC	982 c	830 c	726 c	836 c	2099 c	1813 c	1551 c	1973 c
Nashua	SSC	824 a	722 a	618 a	834 a	2771 a	2559 a	2085 a	2534 a
	SFC	781 a	665 a	539 b	749 b	2598 ab	2030 b	1615 b	2581 a
	FFC	631 b	558 b	508 b	590 c	2322 b	1924 b	1608 b	2038 b

† SSC = sativa x sativa crosses; SFC = sativa x falcata crosses; FFC = falcata x falcata crosses

‡Mean values within generations and locations followed by different letters are significantly different at α=0.05.

Table 2.4. Biomass yield depression between the F1 and S1 or F2 generations and progressive heterosis for biomass yield in double crosses relative to single crosses for three cross types at two Iowa locations and in two years.

		2004			2005		
Location	Cross type	DC v. MF1†	F2 v. F1	S1 v. F1	DC v. MF1	F2 v. F1	S1 v. F1
		----------g plot^{-1}----------					
Ames	SSC	261**	-220a*‡	-422ab**	436*	-234b	-1138a**
	SFC	-76	-281a**	-572a**	-254	-678a**	-1132a**
	FFC	-146	-152a	-256b**	-126	-286b	-542b**
Nashua	SSC	10	-102a**	-206ab**	-237	-212b	-686a**
	SFC	-32	-116a**	-242a**	-17	-568a**	-984a**
	FFC	-40	-73a	-123b**	-284	-398ab*	-714a**

*, ** indicate that the mean depression value is significantly different from zero at $\alpha = 0.05$ or 0.01, respectively.

† DC = Double Cross; MF1 = mean of two F1 parental populations that were crossed to get double cross.

‡Mean values within generations and locations followed by different letters are significantly different at $\alpha=0.05$.

Table 2.5. F-statistics for testing the significance of the effect of different allele interactions in explaining the yield variation among four generations of alfalfa plants derived from four sativa x sativa hybrid populations and grown in the field at Ames, IA for two years.

Entry	Two allele interaction	Three allele interaction	Four allele interaction and epistasis
HRMS06	22.69***	0.68	2.47
HRMS07	21.02***	3.66*	0.03
HRMS08	28.29***	0.21	4.56**
HRMS09	38.21***	7.76**	0.23

*, **, *** represent a significant effect of the particular interaction at $\alpha = 0.1$, 0.05, or 0.01, respectively.

CHAPTER 3

COMPARATIVE GENE EXPRESSION PROFILES BETWEEN HETEROTIC AND NON-HETEROTIC HYBRIDS OF TETRAPLOID *MEDICAGO SATIVA*[2]

[2] Xuehui Li, Yanling Wei, Dan Nettleton, and E. Charles Brummer. Submitted to *BMC Plant Biology*. 2/25/2009.

ABSTRACT

Background

Heterosis, the superior performance of hybrids relative to parents, has clear agricultural value, but its genetic control is unknown. Our objective was to test the hypotheses that hybrids expressing heterosis for biomass yield would show more gene expression levels that were different from midparental values and outside the range of parental values than hybrids that do not exhibit heterosis.

Results

We tested these hypotheses in three *Medicago sativa* (alfalfa) genotypes and their three hybrids, two of which expressed heterosis for biomass yield and a third that did not, using Affymetrix *M. truncatula* GeneChip arrays. Alfalfa hybridized to approximately 47% of the *M. truncatula* probe sets. Probe set signal intensities were analyzed using MicroArray Suite v.5.0 (MAS) and robust multi-array average (RMA) algorithms. Based on MAS analysis, the two heterotic hybrids performed similarly, with about 27% of genes showing differential expression among the parents and their hybrid compared to 12.5% for the non-heterotic hybrid. At a false discovery rate of 0.15, 4.7% of differentially expressed genes in hybrids (~300 genes) showed nonadditive expression compared to only 0.5% (16 genes) in the non-heterotic hybrid. Of the nonadditively expressed genes, approximately 50% showed expression levels that fell outside the parental range in heterotic hybrids, but only one of 16 showed a similar profile in the non-heterotic hybrid. Results were similar with the RMA analysis. Genes whose expression differed in the parents were three times more likely to show nonadditive expression than genes whose parental transcript levels were equal.

Conclusions

We conclude that nonadditive expression of transcript levels may contribute to heterosis for biomass yield in alfalfa.

BACKGROUND

Heterosis is a phenomenon in which offspring show increased fitness relative to their parents (Shull, 1908). In classic quantitative genetics, three main hypotheses have been proposed to explain heterosis (Falconer and Mackay, 1996). One is the dominance hypothesis, which suggests heterosis results from the complementation of favorable alleles of different loci in F_1 hybrids. Under the dominance hypothesis, each heterozygous locus in F_1 hybrids contributes to a trait value within the range of the two homozygous parents, but summing locus effects across the genome gives the hybrid its advantage over its parents. The second is the over-dominance hypothesis, which states that a heterozygous locus in an F_1 hybrid will perform better than either homozygous locus in parents; therefore, heterozygosity *per se* causes heterosis. Finally, the third hypothesis suggests that epistasis plays the predominant role in heterosis expression, and recent evidence in *Arabidopsis* shows that it plays a role in heterosis of biomass (Kusterer et al., 2007). All three hypotheses postulate that physical allelic variation between parents results in novel allelic interactions at given loci in F_1 hybrids, which in turn causes heterosis. Although not always explicitly stated, all three mechanisms concurrently may play a role in heterosis.

The underlying genetic causes of heterosis are not understood. Alleles at a given locus may be expressed at different levels (Brem et al., 2002; Kliebenstein et al., 2006), and heterosis

may be explained at the molecular level by the combined allelic expression in F_1 hybrids, and in particular, by nonadditive expression, at each locus involved in a trait (Birchler et al., 2003). Nonadditive expression in transcript levels could be classified in two ways. First, the hybrid expression level could be different from the midparental value but within the range of the parental values. Second, the hybrid expression could be outside of the parental expression level, such that the hybrid's expression is significantly above the high parent or below the low parent.

Nonadditive expression in F_1 hybrids has been documented in several cases. In maize, Auger et al (Auger et al., 2005) used northern blot assays to analyze 30 transcripts in two maize inbred lines and their two reciprocal hybrids and found that 19 and 20 transcripts showed nonadditive expression. Of the 24 genes showing nonadditive expression in at least one hybrid, 16 showed hybrid patterns that fell outside the parental range of expression. More recent microarray experiments conducted on the same maize hybrid family (B73 x Mo17) have shown ~20% of genes show nonadditive expression (Stupar and Springer, 2006; Swanson-Wagner et al., 2006). However, these two experiments differed in the number of genes whose expression was higher or lower than the parental values, ranging from about 14% of genes [9] to nearly none (Stupar and Springer, 2006). Similar experiments have been conducted in Arabidopsis, Drosophila, and rice (Gibson et al., 2004; Huang et al., 2006; Vuylsteke et al., 2005; Wang et al., 2006), all of which show substantial nonadditive gene expression, but the level of expression outside the parental range is variable. However, the different degrees and types of nonadditive expression observed in these studies could be due to biological, technical, and/or statistical analysis differences, so generalizations about nonadditive gene expression in hybrids across studies and species are difficult. Unfortunately, none of these experiments assessed gene expression in hybrids that do not show a heterotic response for the trait of interest, making

conclusions that nonadditive expression is related to heterosis difficult to support. More recently, an analysis of six hybrids expressing varying levels of high parent heterosis for different seedling traits found similar expression patterns among the hybrids (Stupar et al., 2008). The authors suggest that differences in transcriptional diversity among parents, rather than expression patterns *per se* in hybrids, may be involved with heterosis expression.

Cultivated *Medicago sativa* (alfalfa) is a tetrasomic tetraploid consisting of two major subspecies, *M. sativa* subsp. *sativa* and subsp. *falcata*. Hybrids between these groups often express heterosis for biomass yield and other quantitative traits (Bhandari et al., 2007; Riday and Brummer, 2002a; Riday and Brummer, 2002b; Riday and Brummer, 2005; Riday et al., 2002). This finding may help breeders improve the yield of this important forage crop, which has recently seen a productivity plateau (Lamb et al., 2006; Riday and Brummer, 2002a). While these field-based observations demonstrate the potential for heterosis expression in alfalfa, a fuller understanding of the molecular genetic mechanisms causing heterosis could assist breeders in reliably creating high-yielding hybrids.

In this experiment, we grew three tetraploid alfalfa hybrids, two of which expressed heterosis for biomass yield in field experiments and a third that did not (Riday and Brummer, 2002a), and assessed global gene expression using Affymetrix *Medicago* GeneChip arrays. With these data, we tested the hypotheses that (i) more genes with nonadditive expression patterns would be identified in heterotic than in non-heterotic hybrids when hybrids were compared to their respective parents, (ii) more genes would show expression patterns that were higher than the high parent or lower than the low parent in heterotic than in non-heterotic hybrids, and (iii) the two heterotic hybrids would have similar expression profiles.

METHODS

Plant Growth, Experiment Design and Sampling

We focused on three genotypes and their hybrids. The parents consisted of one genotype from a semi-improved germplasm of subsp. *falcata*, WISFAL-6 (P1), and two elite genotypes from commercial alfalfa breeding germplasm of subsp. *sativa*, ABI408 (P2) and C96-513 (P3). These three genotypes and their hybrids (H12, H13 and H23) have been extensively evaluated for biomass yield, nutritive value, and agronomic traits in a series of previous papers (Riday and Brummer, 2002a; Riday and Brummer, 2002b; Riday et al., 2002). The two sativa×falcata hybrids had previously exhibited heterosis for biomass yield and the sativa×sativa hybrid did not when evaluated in a field experiment (Riday and Brummer, 2002a). For convenience in the following narrative, we refer to the three parents and their three hybrid populations as the six entries evaluated in the study. Also, we will refer to the hybrids expressing heterosis for biomass as "heterotic hybrids" and the hybrid which did not as a "non-heterotic hybrid."

The experimental design was a randomized complete block design (RCBD) with four replications. Each replication included 2 clones for each parent and a single clone for each of 10 genotypes in each hybrid family, for a total of 36 plants. Because the parents were not inbred lines, a cross between them results in a segregating F_1 population. Thus, the 10 F_1 individuals per family represented the hybrid population for the array experiment. Plants were grown in growth chambers (two replications in each of two chambers) under controlled conditions of 25ºC and a 16 hr photoperiod. After being placed into the chambers, plants were maintained for 30 days at which point all biomass was clipped to a 5cm height above soil. Twenty-three days following clipping, the upper fully expanded leaf on a given stem was sampled for RNA isolation and microarray analysis. We sampled five trifoliate leaves from each of the two clones

for each parent, and one trifoliate leaf from each of 10 genotypes for each hybrid. The leaves for each parent or hybrid were pooled prior to RNA extraction. Leaves were harvested, quickly frozen in liquid nitrogen, and stored at -80 ºC until RNA isolation. After sampling leaves, the whole plants were cut and dried at 60 ºC for four days to measure the dry weight. Mid-parent heterosis for yield was calculated on a dry weight basis as the difference between the mean value of an F_1 population and the mean of the parents.

RNA isolation and hybridization

The total RNA for array hybridizations was extracted from frozen leaf tissue with Trizol reagent using standard procedures (Puissant and Houdebine, 1990). Gene expression was assayed using *Medicago* Affymetrix GeneChips, which include 61,278 genes identified from EST collections and genome sequencing data in ***M. truncatula***, ***Sinorhizobium meliloti*** and ***M. sativa***, together with hybridization controls, housekeeping controls, and Poly-A controls. For the experiment, four biological replications of the six entries resulted in 24 GeneChip hybridizations.

First strand cDNA synthesis, GeneChip hybridization, and array staining were conducted at the Iowa State University GeneChip Facility (http://www.biotech.iastate.edu/facilities/genechip/Genechip.htm). Arrays were scanned with a GeneChip Scanner 3000 7G. The gene expression of each probe set on the array was determined from the scanned signal intensities using the Affymetrix® MicroArray Suite v.5.0 (MAS) software and the robust multi-array average (RMA) software (Irizarry et al., 2003). The data resulting from both methods have been uploaded to the MIAMExpress public database (http://www.ebi.ac.uk/miamexpress/, accession number: E-MEXP-1579).

Statistical analysis of microarray data

MAS determines the actual expression intensity of each probe set and provides a detection call indicating whether the estimated expression level is reliable by classifying each probe set on each chip as present (P), marginal (M), or absent (A). Thus, using MAS, we first compared genotypes based on detection calls, and second based on the actual expression intensities of each probe set, filtered by detection call as suggested by previous studies (McClintick and Edenberg, 2006; Pepper et al., 2007). With RMA, we compared genotypes based on expression intensities of each probe set, the only result RMA provides.

Comparisons based on detection calls: Each chip contains 61,278 probe sets. Because our experiment included four replications (corresponding to four separate chips for each entry), each entry received four signal calls for each probe set. For a given entry, a probe set that was PPPP, PPPM, PPPA, or PPMM across the four replications was designated as present, a probe set that was MAAA or AAAA was designated as absent, and the remaining probe sets were designated as marginal.

Comparisons based on expression level differences: Expression intensity data from MAS were log transformed and normalized by median centering prior to analysis. Using the transformed and normalized MAS data and the RMA expression intensity data, we fit the following mixed linear model to each probe set:

$$Y_{ij} = \mu + G_i + r_j + e_{ij}$$

where μ is the overall probe set mean, G_i (i=1,…,6) is the effect of the ith entry, r_j (j=1,…,4) is the effect of the jth replication, and e_{ij} is the random error associated with the ith

entry in the *j*th replication; r_j and e_{ij} were modeled as independent normal random effects, and the others were modeled as fixed effects.

Differential expression was evaluated (i) among the three parental entries, (ii) between the two parents of a given hybrid, and (iii) between the two parents and their hybrid by testing the null hypothesis that the entries had equal expression levels. To control for multiple testing errors, the false discovery rate (FDR) of Benjamini and Hochberg (Benjamini and Hochberg, 1995) was employed at a significance level of α=0.15, as has been used in other studies of this type (Swanson-Wagner et al., 2006). For MAS data, only probe sets that were identified as being present in at least one of the entries being compared were evaluated.

For each hybrid family (i.e., the two parents and their hybrid), probe sets with nonadditive expression were identified within the differentially expressed probe sets by contrasting the expression levels of the hybrid with the mean of the two parents. We were interested in whether the numbers of genes with nonadditive expression differed between heterotic and non-heterotic hybrid families. Therefore, we assessed four different significance level thresholds to determine the stability of the relationship between hybrid types, including *p*-values of 0.05 and 0.01 and FDR levels of 0.20 and 0.15. In order to test whether nonadditive expression in the hybrid tended to occur for probe sets that were differentially expressed between parents, we calculated an odds ratio (OR) to compare the number of nonadditively expressed probe sets that showed differential expression between parents and those that did not as follows:

$$\mathrm{OR} = \frac{m1}{n1 - m1} \Big/ \frac{m2}{n2 - m2}$$

where, *m1* is the number of probe sets with nonadditive expression patterns that also showed different expression levels between parents, *n1* is the total number of probe sets whose expression was significantly different between parents, *m2* is the number of probe sets with

nonadditive expression patterns whose expression was not significantly different between parents, and *n2* is the total number of probe sets whose expression was not significantly different between parents. The 95% confidence limits of the odds ratio were calculated using the EXACT statement and OR option in the SAS procedure FREQ (SAS, 2000).

The probe sets that showed nonadditive expression were classified as being (1) outside the parental range of expression (i.e., higher than the high parent or lower than the low parent at a *p*-value of 0.05) or (2) within the parental range of expression (i.e., equal to or less than the higher parent but greater than the midparental value or equal to or greater than the lower parent but less than the midparental value at a *p*-value of 0.05).

For MAS data, we also identified probe sets that were only expressed in the hybrid in each hybrid family (i.e., the detection call was 'present' in the hybrid and 'absent' in both parents and the actual expression level was different between the hybrid and either parent at FDR<0.15) and those expressed only in both parents and not the hybrid, using the same parameters.

Validation of gene expression via quantitative Real-Time PCR (qRT-PCR)

The qRT-PCR analysis was performed on first strand cDNA synthesized from the same RNA samples used for the microarray experiment. A poly dT primer and SuperScript II RNase H Reverse Transcriptase (Cat. No. 18064-014, Invitrogen, CA) were used to synthesize first strand cDNA. Amplification primers (see Additional file 4) were designed using Primer 3 (Rozen and Skaletsky, 2000) for nine probe sets having contrasting expression patterns among the 6 entries based on MAS data. The qRT-PCR was conducted using first strand cDNA diluted 60 times on a LightCycler 480 SYBR Green I Master (Roche Cat. No. 04-707-516-001)

following the manufacturer's protocol. The qRT-PCR data were initially analyzed with the LightCycler 480 analysis software to obtain crossing point (Cp) values for each probe set.

RESULTS

Heterosis expression

The hybrids H12 and H13 showed significant mid-parent heterosis for biomass, while hybrid H23 did not (Table 1). The entries we used in this experiment were grown in the growth chamber, but the biomass production we measured in this experiment showed the same relative patterns of heterosis as observed previously in field experiments (Riday and Brummer, 2002a). The low yield of WISFAL-6 is attributable to its slower regrowth compared to the two sativa parents.

Probe set hybridization patterns based on MAS detection calls

Of the total 61,278 probe sets on the *Medicago* chip, 25,604 (41.8%) were 'present' in at least one of the six entries in this experiment. Of these probe sets, 71.0% were present in all entries, 20.8% were present in two to five entries, and 8.2% were unique to one entry. The 61,278 probe sets were designed from 3 species: *M. sativa*, *M. truncatula*, and *S. meliloti*. About 90.6% (1,711 of 1,888) of the probe sets derived from *M. sativa* but only 46.6% (23,700 of 50,905) of those from *M. truncatula* and 1.2% (99 of 8,305) of those from *S. meliloti* were scored as present in at least one of the six entries. Of these probe sets, 90.4%, 69.7% and 1.0%, respectively, were present in all entries and 2.0%, 8.4% and 71.7%, respectively, were present only in one single entry. Because our experimental material was *M. sativa*, the observed hybridization percentages are not surprising. The 10% of *M. sativa* genes that were not present

in any individual may represent genes that were not expressed in leaves at this developmental stage and under these environmental conditions, or that were expressed at a level too low to be detected.

Comparisons between parents

MAS results: Of the 24,356 probe sets that were present in at least one of the three parents, 18,796 were present in all parents and 2,975 were only present in a single parent (Figure 1). The number of probe sets present in only one parent did not differ substantially among the three parents, and P1 (WISFAL-6), which derived from *M. sativa* subsp. *falcata*, is not obviously different from the two subsp. *sativa* parents in terms of hybridization efficiency.

Of the probe sets present in at least one parent, 10,130 showed different expression levels among the three parents. For the non-heterotic parent pair P2-P3, 4,222 of 23,341 probe sets (18.1%) were found to be differentially expressed between parents, while for the heterotic parent pairs, 7,062 of 23,522 (30.0%) were differentially expressed between P1 and P2, and 7,227 of 23,230 (31.1%) between P1 and P3 (Table 2). Despite the variation among parent pairs in the number of differentially expressed genes, each parent in each pair had higher expression for about half of the probe sets (Table 2).

The probe sets with significantly different expression between each pair of parents had between 1.16 and 1141 fold change, with an overall median fold change of 1.93; all three parent pairs showed similar median fold change (Table 2). Considering only those probe sets having at least a 2-fold difference in expression, 1,960 probe sets displayed different expression for the non-heterotic parent pair P2-P3, compared to 3,196 and 3,385 for the heterotic parent pairs P1-P2 and P1-P3, respectively (Table 2). Of the probe sets that had different expression between

parents, only about 6-8% were present in one parent and absent in the other (Table 2). This indicated that transcriptional diversity among genotypes was mainly due to transcript abundance rather than the presence or absence of expression.

RMA results: A total of 17,387 probe sets showed different expression levels among the three parents when analyzed with RMA. The RMA results showed patterns similar to the MAS results. Heterotic parent pairs had more differentially expressed genes than the non-heterotic parent pair and each parent of a particular cross contributed about 50% of the genes with higher expression (Table 2). The RMA analysis identified more differentially expressed probe sets but fewer probe sets that showed fold changes greater than two when compared to MAS (Table 2). Interestingly, however, only a fraction of the probe sets identified as differentially expressed by MAS for a given parental pair were also identified by RMA as being differentially expressed for that same parental pair (P1-P2 = 23%; P1-P3 = 24%; P2-P3 = 17%).

Comparisons between parents and their hybrid

MAS results: We further analyzed each hybrid family separately to determine the proportion of probe sets showing nonadditive expression and the prevalence of hybrid expression values outside the parental range of expression. Using a cutoff of FDR<0.15, 12.5% of probe sets displayed different expression levels among the three entries in the non-heterotic hybrid family H23, but in the heterotic hybrid families, 26.3% in H12 and 27.6% in H13 showed differences (Table 3). For each hybrid family, the probe sets with different expression can be divided into those in which the hybrid exhibits additivity of expression relative to its parents and those exhibiting nonadditive expression. We evaluated the number of probe sets with nonadditive expression using four significance thresholds ($p<0.05$, $p<0.01$, FDR<0.20, and FDR<0.15). The

numbers varied dramatically among the four cutoff levels as expected, but importantly, in all cases, the heterotic hybrids (H12 and H13) showed substantially more nonadditively expressed probe sets than the non-heterotic hybrid (Figure 2).

We calculated the numbers of probe sets showing nonadditive expression that also had different expression levels between the parents. In all three hybrid families, a higher proportion of nonadditively expressed genes were identified in the subset of probe sets that were differentially expressed between parents than in those not differentially expressed between parents. The lower limit of the 95% confidence interval for the odds ratio under all four cutoffs was approximately three or greater (Table 4), which indicated that probe sets whose expression differed between the parents had odds of nonadditive expression that were at least three times greater than the odds of nonadditive expression for probe sets whose expression did not differ between parents. Thus, heterotic hybrids showed more nonadditive expression, and the proportion of differentially expressed probe sets in heterotic parent pairs was higher than for the non-heterotic pair.

The probe sets with nonadditive expression were divided into two categories: (i) those in which the hybrid expression level fell within the parental range of expression and (ii) those in which the hybrid expression value fell outside the parental range of expression. Greater proportions of probe sets were found to fall outside the parental range of expression in heterotic hybrids than in the non-heterotic hybrid based on FDR<0.15 (Table 3) and also under the other three statistical thresholds (data not shown). Approximately 300 probe sets displayed nonadditive expression in each of the heterotic hybrids, and about half of these had expression levels that were higher than the higher parent or lower than the lower parent (Table 3). Of the 69 probe sets with non-additive expression that were in common between the two heterotic hybrids,

65 did not display nonadditive expression in the non-heterotic hybrid H23 (see Additional file 1). In the non-heterotic H23 hybrid family, no probe set was expressed only in the hybrid or only in both parents. In contrast, one probe set in H12 and 10 in H13 were expressed only in the hybrid (see Additional file 2).

RMA results: The RMA results were similar to the MAS results in that more probe sets with non-additive expression and with expression outside of the parental range were found in heterotic hybrid families than in non-heterotic hybrid families (Table 3 and Figure 2). However, only two and four probe sets showing non-additive expression overlapped between analysis methods for the H12 and H13 hybrid families, respectively, and no probe sets overlapped for the H23 hybrid family, using a cutoff of FDR<0.15. A total of 124 probe sets showed non-additive expression in both heterotic hybrids, 119 of which did not show non-additive expression in the non-heterotic hybrid H23 (see Additional file 3).

Validation of gene expression via quantitative Real Time PCR (qRT-PCR)

Quantitative RT-PCR was applied to 9 probe sets to verify the microarray data. Two of the probe sets, Mtr3074 and Mtr43518, did not differ among the six entries and all others showed differences in expression between at least two of the six entries based on the MAS data. In general, the qRT-PCR results produced relative expression patterns similar to those observed from the MAS analysis (Figure 3). However, some differences were evident. For Mtr34420, several entries had different expression patterns than those observed from the MAS analysis, and one entry with a different pattern than the MAS analysis was observed for Mtr241. A total of 135 pairwise comparisons for expression patterns are possible among the six entries across all nine probe sets (i.e., 15 pairwise comparisons for each probe set). Of these 135, 90 (67%) were

validated by qRT-PCR. Out of 15 comparisons, only 4 and 5 were validated for probe set Mtr34420 and Mtr241, respectively, while 9 to 14 comparisons were validated for other probe sets. When compared to the RMA data, 77 (57%) of 135 pairs of comparison were validated by qRT-PCR. These results suggest that overall, the broad pattern of our microarray results is an accurate depiction of the gene expression levels among these entries.

DISCUSSION

A number of algorithms are available for calculating the expression intensities on Affymetrix microarrays. Among them, MAS and RMA are two of the most widely used. Comparative studies using spike-in or dilution controls have suggested that RMA algorithms are more accurate than MAS (Irizarry et al., 2003; Zhou and Rocke, 2005), but other experiments suggest that detection calls effectively filtered MAS data, removing the vast majority of false positive results, and that the filtered-MAS data yielded better results than RMA (McClintick and Edenberg, 2006; Pepper et al., 2007; Seo et al., 2004). The contrasting results could be due to the different datasets, assessments, and assessment statistics used in different studies.

In this study, more differentially expressed genes between parental pairs were identified by RMA than by MAS, but smaller proportions of them showed fold changes greater than two. This supports the hypothesis proposed by previous studies (Irizarry et al., 2003; Seo et al., 2004) that the RMA algorithm is more sensitive, particularly at low expression levels, but this may increase the proportion of false positive results, thereby increasing noise in the data (Seo et al., 2004). Given the conflicting results of previous experiments, we analyzed our data using both methods – MAS and RMA – to determine if the results we obtained were consistent across analysis methods.

The entries used in this study were previously tested in a field experiment (Riday and Brummer, 2002a; Riday and Brummer, 2002b), which showed that the heterotic hybrids exhibited high parent heterosis for biomass yield and that these heterotic hybrids showed greater heterosis as the period of regrowth increased. Our growth chamber results indicated that the heterotic hybrids exhibited mid-parent heterosis, probably due to the shorter length of regrowth at harvest, which we limited to three weeks to avoid possible changes in gene expression due to flowering time differences, and/or to the very different environmental conditions in the chamber compared to the field. Mid-parent heterosis for biomass may not be useful for breeding applications, but it is meaningful for the genetic study of heterosis because the difference between the hybrid and the parental mean is the response variable to be related to nonadditive expression, not their absolute phenotypic performance.

We compared two hybrids expressing heterosis for biomass yield with a third hybrid that did not express heterosis. The heterotic hybrid families had a higher number and a higher proportion of genes exhibiting differential expression and nonadditive expression than did the non-heterotic family using either analysis method (RMA or MAS). Higher proportions of probe sets with expression outside of the parental range were also found in heterotic hybrids compared to a non-heterotic hybrid. At an FDR < 0.15, we found about 300 nonadditively expressed genes in heterotic hybrids based on MAS, about half of which had expression outside the parental range, compared to 16 in the non-heterotic hybrid. Similar patterns were seen with RMA. Our data suggest that genes that have non-additive expression patterns in the hybrid and, more importantly, that have expression levels higher than the high parent or lower than the low parent could play a role in heterosis for biomass yield.

Although the two analysis methods produced broadly similar results, different numbers of probe sets were identified as differentially expressed by the two methods and only a small proportion of these probe sets overlapped. This is not surprising, because the algorithms use different background correction, normalization, and summarization methods (Millenaar et al., 2006). Further investigation is needed to determine if one algorithm more accurately identified important genes in this experiment, although based on congruence with the RT-PCR results, MAS appeared to hold a slight advantage.

Our results stand in contrast to Stupar and Springer (Stupar and Springer, 2006) who found very little evidence for hybrid gene expression patterns that were nonadditive or that exceeded parental levels, and different from Uzarowska et al (Uzarowska et al., 2007) who found a large proportion of genes showing nonadditive expression (90%) and expression outside the parental range (51%) in maize. Our results are broadly similar to those of Swanson-Wagner et al. (2006). However, comparisons among experiments for the percentage of nonadditively expressed genes need to be made cautiously for a number of reasons, including the use of different statistical methods and thresholds. Recently, a few studies compared the expression profiles of a set of hybrids simultaneously. Stupar et al (Stupar et al., 2008) investigated the gene expression profile of six maize inbred-hybrid combinations with varying levels of better parent heterosis on five traits, and found a strong correlation between the number of differentially expressed genes and the level of genetic distance between inbred parents, while the proportions of nonadditive expression among the differentially expressed genes were similar among the hybrids. Interestingly, the hybrid with the smallest genetic distance – and the least high-parent heterosis for seedling traits – exhibited the greatest proportion of nonadditive expression. The authors proposed that nonadditive expression is not correlated with heterosis

levels. Guo et al (Guo et al., 2006) found that heterosis was correlated with the proportion of additively expressed genes but not with the proportion of genes with expression levels outside of the parental range in a set of 16 maize hybrids.

Our study only analyzed three hybrids, limiting our ability to generalize these results to other hybrids. Perhaps more importantly, our results need to be interpreted cautiously given that we used non-inbred parents. Unfortunately, alfalfa suffers severe inbreeding depression, and true inbred lines are not available. To account for the heterogeneity of F_1 hybrid indivduals, we pooled ten individuals for each hybrid. This can potentially lead to erroneous results, if alleles from the heterozygous parents are not present in the progeny in equal frequencies. In this case, the hybrid expression relative to the parental mean may be skewed – for example, if the progeny only received a highly expressing allele from one parent, then the overall hybrid expression level may be equal to or exceed the higher parent, even though the hybrid expression level should be additive. Without evaluating allele-specific expression patterns, this concern is difficult to allay. We examined the heterozygosity of the parents using 41 EST-SSR markers. WISFAL-6 (P1) had 1.92 alleles/marker, ABI408 (P2) had 1.95, and C96-513 had 2.15. Assuming that the SSR allele diversity mirrors the diversity of alleles producing different expression patterns, these results suggest that the three parents would have a similar chance to generate false expression results due to preferential allele inheritance. Therefore, we suggest that our comparisons among the three hybrids regarding the about the number and proportion of genes showing nonadditive expression are valid.

Although higher proportions of the nonadditive expression and expression higher or lower than either parent were found in heterotic hybrids compared to a non-heterotic hybrid in our study, the majority of genes showed additive expression in all hybrid families. We may have

underestimated the numbers of genes with nonadditive expression due to limitations in our statistical power for this experiment. However, in maize, although the F_1 hybrid between Mo17 and B73 showed significant high parent heterosis for seedling growth, only 22% of differentially expressed genes had nonadditive expression patterns and only a small proportion of them showed expression outside of the parental range, similar to our results (Swanson-Wagner et al., 2006). Springer and Stupar (Springer and Stupar, 2007) proposed that heterosis could result from the additive expression of multiple genes, whereby particularly low or high expression values that are generally detrimental to the plant are modulated in the hybrid, which expresses an average expression level in a moderate, but more biologically functional range. While this may be true in some cases, the clear differences in expression patterns between hybrid types in our experiment suggests that nonadditive expression may also be important for heterosis expression.

What is heterosis? Heterosis simply represents the manifestation of a phenotype in a hybrid that is different from the expectation of a parental average value for that phenotype, be it yield, plant height, or any other trait. The manifestation of the phenotype – particularly of quantitatively inherited traits like yield – results from the complex actions of many components, including the timing of the expression of various genes, the magnitude and location of their expression, and the interaction of their gene products. The genetic hypothesis for the cause of heterosis that has the most empirical support at the current time is that each parent contains a set of dominant alleles at loci controlling the trait and that at some loci, the other parent has recessive alleles at those loci; thus, hybridization brings these dominant alleles together, with the parents complementing each other and giving the hybrid a larger set of dominant (and desirable) alleles than either parent. Complementary expression patterns – each parent contributing alleles that show higher expression than those at the relevant loci in the other parent – could have the

same effect. Under this model, hybrids expressing heterosis should have more nonadditive expression patterns, as we have shown in our alfalfa example. Given that control of complex traits likely involves many genes and given that the expression patterns of most genes is additive, this model does not exclude the possibility that additivity also plays a role in heterosis, under the model suggested by Springer and Stupar (Springer and Stupar, 2007).

Conceivably, only a subset of genes may need to deviate from additivity of expression in order to produce a heterotic phenotype. The extent of nonadditive expression at different development stages and different tissues may vary and across the life cycle of the plant, the expression patterns cumulatively produce the observed heterotic response. *Arabidopsis* allotetraploids had little overlap between the set of genes exhibiting nonadditive expression in leaves and that in flowers, suggesting a role of developmental stages and tissue types on nonadditive gene regulation (Wang et al., 2006). If nonadditively expressed genes truly do underlie heterosis, this result suggests that different genes contribute to heterosis in different tissues and at different developmental stages. Thus, for integrative phenotypes like yield, the cumulative effect of these different genes acting at different places and times could result in heterosis. If this is the case, then the nonadditive expression observed at a single timepoint and in a single tissue, as we assayed here, would only give a small part of the overall picture of how gene expression may affect the ultimate expression of the yield phenotype. Finally, genetic divergence between the parental lines appears to result in more differential expression between parents. Both in our study and in that in *Arabidopsis* by (Wang et al., 2006), a higher proportion of nonadditive expression occurred in hybrids whose parents showed divergent expression levels than in hybrids whose parents had similar expression levels. This suggests that there could be more nonadditive expression in the crosses between more distantly related parents, exactly the

type of situation in which agronomically useful heterosis levels are also commonly observed. However, recent results in maize suggest that this may not be the case (Stupar et al., 2008).

The expression patterns of individual genes are themselves controlled by other genes, acting in *cis* or *trans* (Stupar and Springer, 2006; West et al., 2007). Thus, heterosis for an ultimate phenotype, in this case, biomass yield, must be controlled by multiple genes exhibiting some level of dominance, with some residing in each parental genome (Falconer and Mackay, 1996). The genes themselves may also be controlled by a number of other genes, and this control can result in expression levels ranging from additivity to some level of non-additivity. Genes controlling transcript levels have been inferred from experiments mapping eQTL, that is, quantitative trait loci that control the expression of a transcript (Brem and Kruglyak, 2005; Brem et al., 2002; West et al., 2007). Interestingly, no eQTL could be mapped for some genes with highly heritable transcript levels in yeast, suggesting that many loci of small effect and/or epistasis among loci controls their expression (Brem and Kruglyak, 2005).

We know that biomass yield, like many other agronomically important traits, is quantitatively inherited, suggesting that it is controlled by many loci (and possibly by multiple interactions among them), and infer that directional dominance plays a role in its control, at least in the certain hybrids that express heterosis. As a means of understanding the nature of the genetic mechanisms underlying biomass yield and yield heterosis, we identified a suite of genes whose expression in hybrids is phenotypically nonadditive, in some cases falling outside of the parental range, and a subset of which only show that expression pattern in heterotic hybrids. But expression of each individual gene is itself the result of a number of gene interactions, and hence, the regulation of expression of any single gene may also have a complex genetic basis. This complexity shows that the genetic control of quantitative traits is difficult to untangle because

many levels of interactions, from genes to gene expression profiles to proteins and metabolites, occur to produce the ultimate phenotype.

AUTHORS' CONTRIBUTIONS

XL participated in experimental design, conducted the bulk of the experimental work, analyzed the microarray and qRT-PCR data, and drafted the manuscript. YL performed qRT-PCR experiment and gene ontology analysis. DN provided advice on statistical analysis and data explanation. CB conceived, designed and supervised the study. All authors have read and approved the final manuscript.

ACKNOWLEDGEMENTS

This research was funded by a grant from the Plant Sciences Institute at Iowa State University (to ECB). Support was also provided by the Raymond F. Baker Center for Plant Breeding at Iowa State Univ. and by the University of Georgia.

REFERENCES

Auger, D.L., A.D. Gray, T.S. Ream, A. Kato, E.H. Coe, Jr., and J.A. Birchler. 2005. Nonadditive gene expression in diploid and triploid hybrids of maize. Genetics 169:389-97.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methoodol. 57:289-300.

Bhandari, H.S., C.A. Pierce, L.W. Murray, and I.M. Ray. 2007. Combining Abilities and Heterosis for Forage Yield among High-Yielding Accessions of the Alfalfa Core Collection. Crop Science 47:665-673.

Birchler, J.A., D.L. Auger, and N.C. Riddle. 2003. In search of the molecular basis of heterosis. Plant Cell 15:2236-9.

Brem, R.B., and L. Kruglyak. 2005. The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc Natl Acad Sci U S A 102:1572-7.

Brem, R.B., G. Yvert, R. Clinton, and L. Kruglyak. 2002. Genetic dissection of transcriptional regulation in budding yeast. Science 296:752-5.

Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to Quantitative Genetics. 4 ed., Longman, Essex, England.

Gibson, G., R. Riley-Berger, L. Harshman, A. Kopp, S. Vacha, S. Nuzhdin, and M. Wayne. 2004. Extensive sex-specific nonadditivity of gene expression in Drosophila melanogaster. Genetics 167:1791-9.

Guo, M., M.A. Rupe, X. Yang, O. Crasta, C. Zinselmeier, O.S. Smith, and B. Bowen. 2006. Genome-wide transcript analysis of maize hybrids: allelic additive gene expression and yield heterosis. Theor Appl Genet 113:831-45.

Huang, Y., L. Zhang, J. Zhang, D. Yuan, C. Xu, X. Li, D. Zhou, S. Wang, and Q. Zhang. 2006. Heterosis and polymorphisms of gene expression in an elite rice hybrid as revealed by a microarray analysis of 9198 unique ESTs. Plant Mol Biol 62:579-91.

Irizarry, R.A., B. Hobbs, F. Collin, Y.D. Beazer-Barclay, K.J. Antonellis, U. Scherf, and T.P. Speed. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4:249-64.

Kliebenstein, D.J., M.A. West, H. van Leeuwen, K. Kim, R.W. Doerge, R.W. Michelmore, and D.A. St Clair. 2006. Genomic survey of gene expression diversity in Arabidopsis thaliana. Genetics 172:1179-89.

Kusterer, B., J. Muminovic, H.F. Utz, H.P. Piepho, S. Barth, M. Heckenberger, R.C. Meyer, T. Altmann, and A.E. Melchinger. 2007. Analysis of a triple testcross design with recombinant inbred lines reveals a significant role of epistasis in heterosis for biomass-related traits in Arabidopsis. Genetics 175:2009-17.

Lamb, J.F.S., C.C. Sheaffer, L.H. Rhodes, R.M. Sulc, D.J. Undersander, and E.C. Brummer. 2006. Five decades of alfalfa cultivar improvement: Impact on forage yield, persistence, and nutritive value. Crop Science 46:902-909.

McClintick, J.N., and H.J. Edenberg. 2006. Effects of filtering by Present call on analysis of microarray experiments. BMC Bioinformatics 7:49.

Millenaar, F.F., J. Okyere, S.T. May, M. van Zanten, L.A. Voesenek, and A.J. Peeters. 2006. How to decide? Different methods of calculating gene expression from short oligonucleotide array data will give different results. BMC Bioinformatics 7:137.

Pepper, S.D., E.K. Saunders, L.E. Edwards, C.L. Wilson, and C.J. Miller. 2007. The utility of MAS5 expression summary and detection call algorithms. BMC Bioinformatics 8:273.

Puissant, C., and L.M. Houdebine. 1990. An improvement of the single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Biotechniques 8:148-9.

Riday, H., and E.C. Brummer. 2002a. Forage Yield Heterosis in Alfalfa. Crop Science 42:716-723.

Riday, H., and E.C. Brummer. 2002b. Heterosis of Agronomic Traits in Alfalfa Crop Science 42:1081-1087.

Riday, H., and E.C. Brummer. 2005. Heterosis in a Broad Range of Alfalfa Germplasm. Crop Science 45:8-17.

Riday, H., E.C. Brummer, and K.J. Moore. 2002. Heterosis of Forage Quality in Alfalfa. Crop Science 42:1088-1093.

Rozen, S., and H. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365-86.

SAS Institute. 2000. SAS language and procedure: Usage. Release Version 8. SAS Institute., Cary, NC.

Seo, J., M. Bakay, Y.W. Chen, S. Hilmer, B. Shneiderman, and E.P. Hoffman. 2004. Interactively optimizing signal-to-noise ratios in expression profiling: project-specific algorithm selection and detection p-value weighting in Affymetrix microarrays. Bioinformatics 20:2534-44.

Shull, G.H. 1908. The composition of a field of maize. American Breeders Assoc Rep 4:296-301.

Springer, N.M., and R.M. Stupar. 2007. Allelic variation and heterosis in maize: how do two halves make more than a whole? Genome Res 17:264-75.

Stupar, R.M., and N.M. Springer. 2006. Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid. Genetics 173:2199-210.

Stupar, R.M., J.M. Gardiner, A.G. Oldre, W.J. Haun, V.L. Chandler, and N.M. Springer. 2008. Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis. BMC Plant Biol 8:33.

Swanson-Wagner, R.A., Y. Jia, R. DeCook, L.A. Borsuk, D. Nettleton, and P.S. Schnable. 2006. All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. Proc Natl Acad Sci U S A 103:6805-10.

Uzarowska, A., B. Keller, H.P. Piepho, G. Schwarz, C. Ingvardsen, G. Wenzel, and T. Lubberstedt. 2007. Comparative expression profiling in meristems of inbred-hybrid triplets of maize based on morphological investigations of heterosis for plant height. Plant Mol Biol 63:21-34.

Vuylsteke, M., F. van Eeuwijk, P. Van Hummelen, M. Kuiper, and M. Zabeau. 2005. Genetic analysis of variation in gene expression in Arabidopsis thaliana. Genetics 171:1267-75.

Wang, J., L. Tian, H.S. Lee, N.E. Wei, H. Jiang, B. Watson, A. Madlung, T.C. Osborn, R.W. Doerge, L. Comai, and Z.J. Chen. 2006. Genomewide nonadditive gene regulation in Arabidopsis allotetraploids. Genetics 172:507-17.

West, M.A., K. Kim, D.J. Kliebenstein, H. van Leeuwen, R.W. Michelmore, R.W. Doerge, and D.A. St Clair. 2007. Global eQTL Mapping Reveals the Complex Genetic Architecture of Transcript-Level Variation in Arabidopsis. Genetics 175:1441-50.

Zhou, L., and D.M. Rocke. 2005. An expression index for Affymetrix GeneChips based on the generalized logarithm. Bioinformatics 21:3983-9

.

Table 3.1. Dry weight for three parental alfalfa genotypes and their hybrids and the mid-parental heterosis values of the hybrids.

Entry	Dry weight	Mid-Parent Heterosis	Hybrid vs. Midparent
	g/plant		*p*-value
P1 (WISFAL-6)	0.56	--	
P2 (ABI408)	2.11	--	
P3 (C96-513)	2.57	--	
H12 (WISFAL-6 × ABI408)	2.05	0.71	0.0029
H13 (WISFAL-6 × C96-513)	2.35	0.79	0.0011
H23 (ABI408 × C96-513)	2.70	0.36	0.1295

Table 3.2. The numbers and proportions of probe sets with significantly different expression levels between parental pairs, fold change in expression levels between parents at a false discovery rate of 0.15, and numbers of genes expressed only in one genotype of each parent pair.

Method	Parental comparison	Differentially expressed genes	Genes with higher expression in first parent of pair listed in second column		Fold change of all differentially expressed genes			Genes with >2 fold change	Genes present in one parent and absent in the other	
		no.	no.	%	minimum	median	maximum	no.	no.	%
MAS	P1 vs P2	7062	3814	54.0	1.17	1.92	711	3196	420	5.9
	P1 vs P3	7227	3608	49.9	1.16	1.95	1141	3385	480	6.6
	P2 vs P3	4222	2009	47.6	1.18	1.92	324	1960	329	7.8
RMA	P1 vs P2	12627	6752	53.5	1.04	1.41	312.3	1890	--	--
	P1 vs P3	12821	6538	51.0	1.05	1.41	180.8	2039	--	--
	P2 vs P3	8147	4028	49.4	1.03	1.40	175.6	1179	--	--

Table 3.3. The numbers and proportions of probe sets exhibiting nonadditive expression and expression levels outside the parental range in each hybrid family at a false discovery rate of 0.15. The total number of probe sets on the GeneChip is 61,278.

	MAS						RMA					
Probe set classification	Heterotic hybrids				Non-heterotic hybrid		Heterotic hybrids				Non-heterotic hybrid	
	H12		H13		H23		H12		H13		H23	
	no.	%	no.	%	no.	%	no.	%	no.	%	no.	%
Present in at least one parent or hybrid	24174	39.4	24296	39.6	23963	39.1						
Present and differentially expressed (MAS) or differentially expressed (RMA)	6346	26.3	6696	27.6	2986	12.5	11942		12015		6209	
Differentially expressed with nonadditive expression	279	4.4	334	5.0	16	0.5	591	4.9	922	7.7	34	0.5
Non-additive expression as above or below the parental range	128	45.9	156	46.7	1	6.2	329	55.7	428	46.4	14	41.2

Table 3.4. Confidence limits (95%) for the ratio of the odds of nonadditivity for probe sets that are differentially expressed between parents to the odds of nonadditivity for probe sets that are not differentially expressed between parents.

Family	$p < 0.05$	$p < 0.01$	FDR < 0.20	FDR < 0.15
H12	(5.3, 6.4)	(2.9, 3.7)	(3.7, 5.5)	(4.2, 7.2)
H13	(6.5, 7.9)	(3.7, 4.8)	(4.1, 6.0)	(4.8, 7.9)
H23	(18.2, 27.5)	(7.7, 12.9)	(9.7, 170.5)	(22.6, ~)

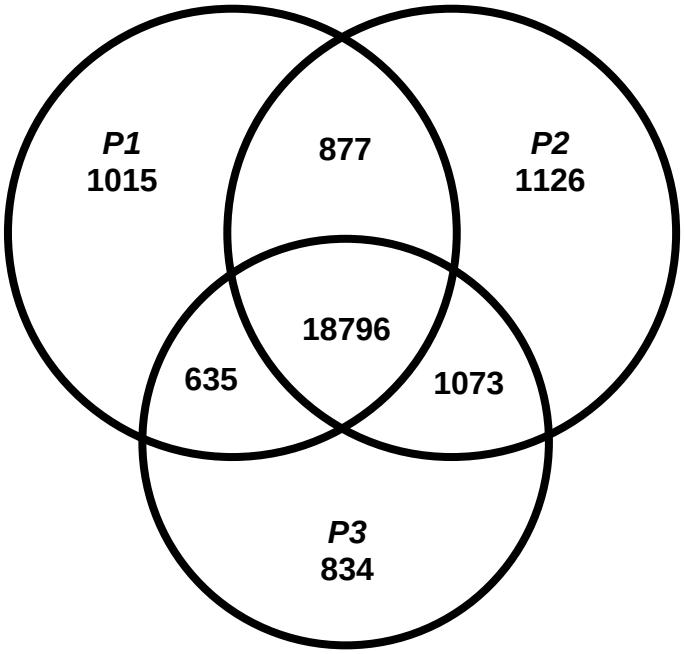
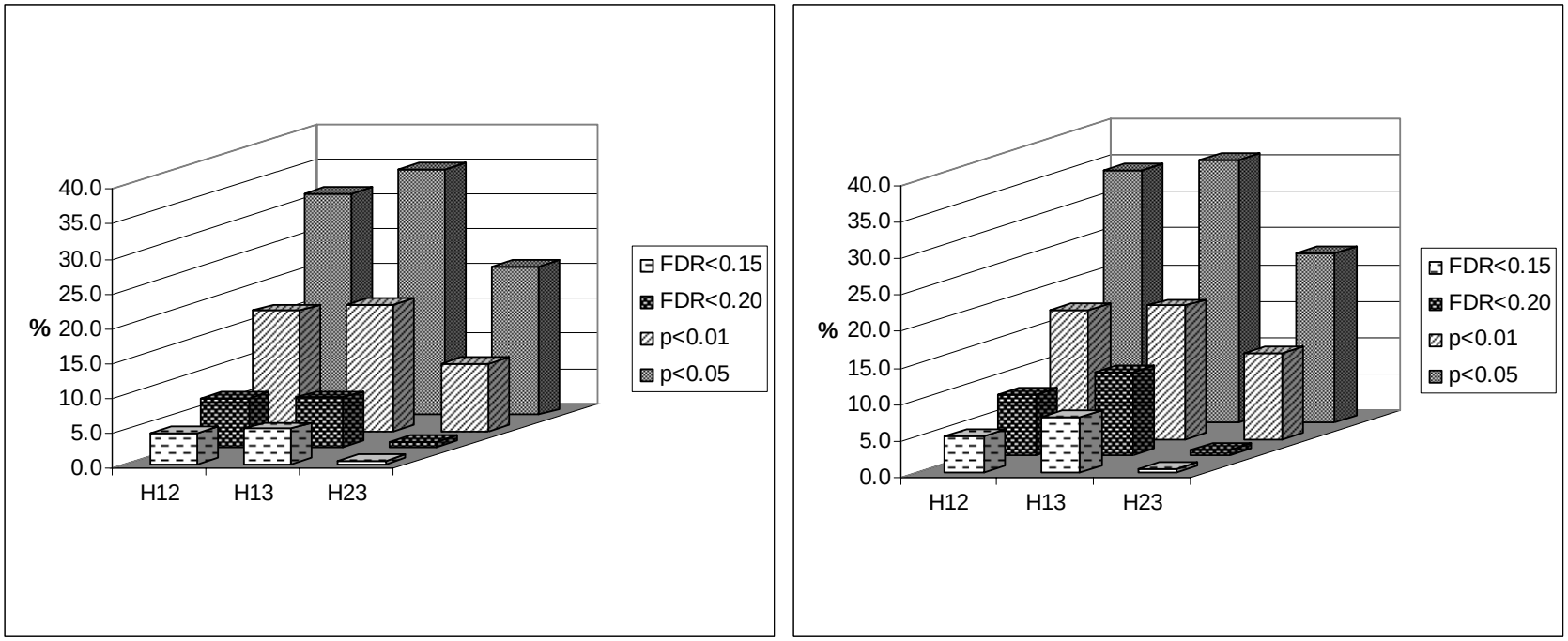


Figure 3.1. The numbers of probe sets present in one, two, or three parental genotypes.



MAS5.0 RMA

Figure 3.2. The proportion of genes showing nonadditive expression at four statistical threshold levels for the three hybrids.

FDR is the false discovery rate.

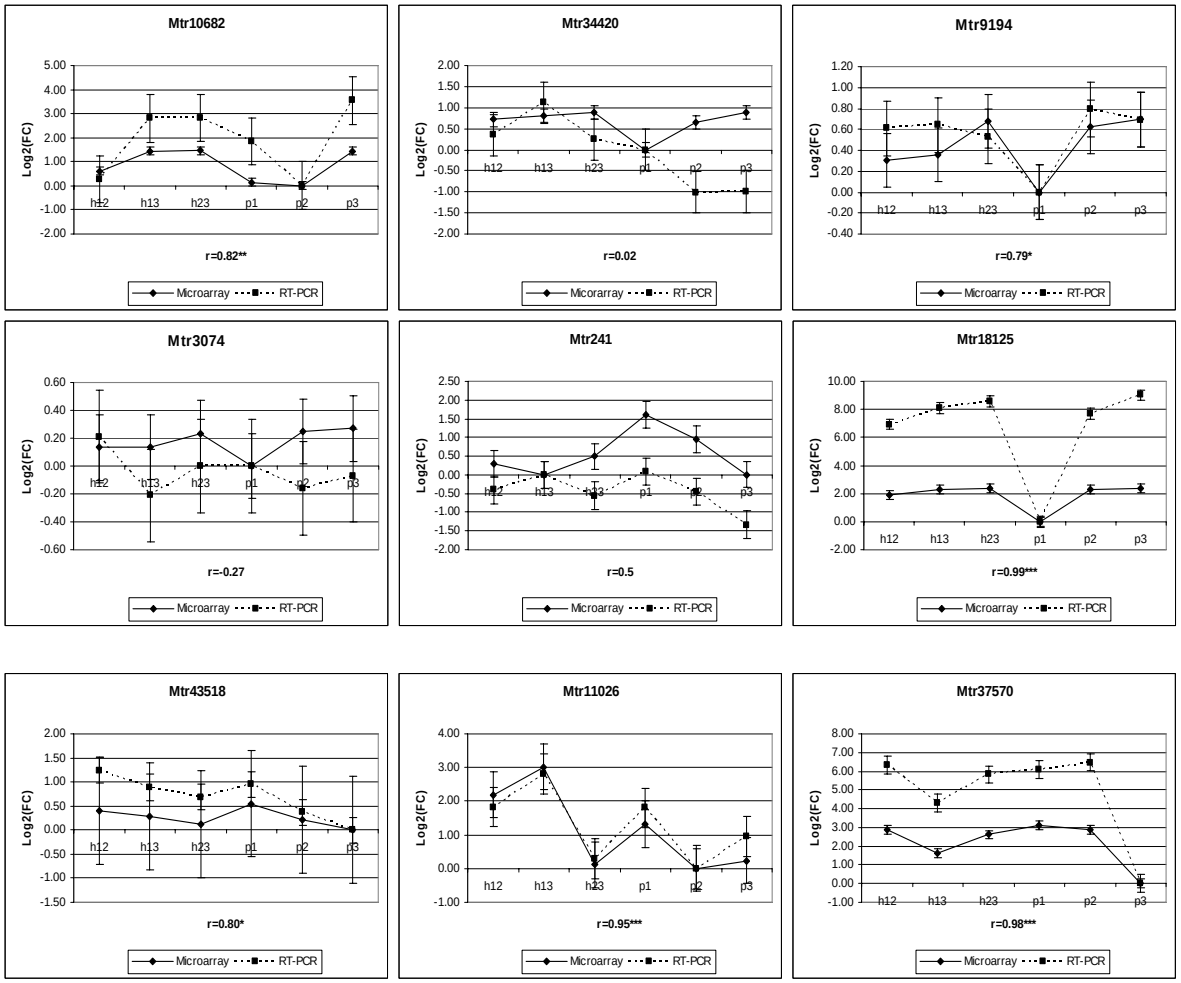


Figure 3.3. Validation of nine probe sets using quantitative Real-Time PCR (qRT-PCR). The $\log_2$-fold change of each entry relative to the entry with the minimum expression on the microarray for each probe set is plotted for both the microarray and the qRT-PCR results. Correlations between them are shown as "r". *, ** and *** represent significance level of 0.1, 0.05 and 0.01, respectively. The standard errors are represented by the vertical bars. Note that the y-axis scale differs for each gene.

CHAPTER 4

SEGREGATION DISTORTION AND GENETIC MAP CONSTRUCTION IN DIPLOID ALFALFA[3]

[3] Xuehui Li and E. Charles Brummer. To be submitted to *Theoretical and Applied Genetics*.

ABSTRACT

Segregation distortion (SD) is often observed in plant populations and affects genetic map construction and QTL mapping. To investigate the prevalence of SD in diploid alfalfa (*Medicago sativa* subsp. *caerulea* and subsp. *falcata*), we developed a segregating F_1 population and an F_2 population from one of the F_1 individuals, and genotyped both populations with SSR markers. For the F_2 population, 50 of 70 mapped markers (71.4%) were distorted. Most distorted markers were clustered into groups located on all 8 chromosomes. We identified 7 segregation distortion regions (SDRs) based on the skew direction of alleles and genotypes and the allelic and zygotic SD tests. A relatively low incidence of allelic SD compared to genotypic SD and zygotic SD suggested that zygotic selection at segregation distortion loci (SDL) is a more common cause of SD than gametic selection. Except for SDR1.1 and SDR8.1, distorted markers in all SDRs showed heterozygote excess. This could be explained by over-dominance or pseudo-overdominance of SDLs. Compared to SD in the F_2 population, a smaller percentage of loci exhibited both genotypic SD and allelic SD in the F_1 mapping population. This could be due to the low inbreeding level in the F_1 population or to non-fully informative markers. The severe SD in the F_2 population likely biased the estimated genetic distances among the distorted markers and possibly the ordering of the markers in some SDRs. However, all markers except for markers on chromosome 5 were correctly assigned to each linkage group/chromosome. The markers on chromosome 5 were assigned to two different linkage groups in the F_2 population, but one linkage group in the F_1 population. The implication for QTL mapping and breeding in alfalfa is discussed.

INTRODUCTION

Mendelian inheritance is a fundamental tenet of classical genetics, enabling geneticists to predict the expression of simple traits. Deviations from the expected Mendelian segregation ratio for various traits have been reported and investigated since the early 20th century (Mangelsdorf and Jones, 1926; Rhoades, 1942; Sandler et al., 1959), and the phenomenon was defined as segregation distortion (Sandler et al., 1959). The causes of SD can be gametic selection, zygotic selection, or both (Falconer and Mackay, 1996). Segregation distortion of molecular markers has been observed in most genetic mapping studies across all the species, including maize (Lu et al., 2002), rice (Xu et al., 1997), tomato (Paterson et al., 1988), alfalfa (Brummer et al., 1993), and populus (Yin et al., 2004). Distorted markers may bias the estimation of genetic distance between markers and may affect orders among markers (Lorieux et al., 1995a; Lorieux et al., 1995b); consequently, it will affect the genetic mapping of phenotypic traits (Vogl and Xu, 2000; Xu, 2008).

Molecular markers can facilitate the investigation of segregation distortion. If a gene contributes to gamete or zygote fitness, then it will cause markers linked to it to deviate from the expected Mendelian segregation ratios (Zamir and Tadmor, 1986). Thus, observing a cluster of markers showing SD in one or more populations suggests that the chromosomal region may have one or more genes causing segregation distortion. These regions of distorted markers are termed segregation distortion regions (SDR, Lu et al., 2002) and the gene(s) segregation distortion loci (SDL, Vogl and Xu, 2000). Based on marker clustering and skew direction, SDR have been identified in maize (Lu et al., 2002), rice (Xu et al., 1997), grape (Riaz et al., 2008), and other species (Faris et al., 1998; Lu et al., 2002; McDaniel et al., 2007). A few studies have attempted to map SDL (Hall and Willis, 2005; Wang et al., 2005).

Diploid alfalfa includes two main subspecies, *M. sativa* subsp. *caerulea* and *M. sativa* subsp. *falcata*. Several genetic maps have been constructed for diploid alfalfa using partially inbred mapping populations (F_2 or backcross) derived from intra-subspecies and inter-subspeices crosses (Brummer et al., 1993; Echt et al., 1994; Kaló et al., 2000; Kiss et al., 1993). Serious SD (18-54%) was found in those studies. Distorted markers were usually found clustered, as is often the case due to linkage, but interestingly, the segregation distortion was most commonly heterozygote excesses. A biological explanation is that alfalfa carries a high level of deleterious recessive alleles due to its outcrossing and self-incompatible nature, and that gametic and/or (perhaps especially) zygotic selection against deleterious recessive alleles during inbreeding caused skewed ratios of marker genotypes. If the deleterious recessive alleles at different loci are linked in repulsion phase, then heterozygotes at linked marker loci would be evident because either homozygote would be linked to a different deleterious recessive allele. Heterozygote excess has also been identified in outbreeding conifer populations, although the heterozygote excess is less extreme than has been observed in alfalfa (Williams et al., 2001; Williams et al., 2003).

In a non-inbred diploid alfalfa mapping population (F_1 population), only 9% of mapped marker loci showed significant segregation distortion, which supports the above explanation (Tavoletti et al., 1996). Another supporting evidence is that a relatively low segregation distortion (4-9% of mapped alleles) was observed in a tetraploid backcross mapping population, because of "the greater buffering capacity of autotetraploids against the effects of deleterious recessive alleles" (Brouwer and Osborn, 1999). However, in a non-inbred autotetraploid alfalfa mapping population (F_1 population), 35% of the AFLP and 25% SSR markers were significantly distorted (Julier et al., 2003). In another F_1 antotetraploid alfalfa mapping population, about 32%

markers exhibited segregation distortion (Robins et al., 2007). Different SDLs have been reported in different populations in maize (Lu et al., 2002) and rice (Xu et al., 1997). The genetic distance between parents also has effect on SD level among different mapping populations (Chetelat et al., 2000). Non-biological factors such as environment (especially conditions during gamete development or seed production), sample size, genotype error, and statistical methodolgy may affect estimation of SD. Thus, analysis of segregation distortion in multiple mapping populations grown under similar environmental conditions may be helpful to understand the mechanisms of segregation distortion.

To investigate the prevalence of segregation distortion, we developed an F_1 population and used one F_1 genotype to produce an F_2 population. We produced additional F_1 individuals at the same time as the F_2 population, so that we could compare both populations being produced under same environmental conditions.

The objectives of this study were to test the hypotheses that (1) greater SD will be observed in the more inbred F_2 population compared to the less inbred F_1 population, (2) the distorted markers will be clustered and predominantly characterized by heterozygote excess in the F_2 population, and (3) the estimation of genetic distances, ordering, and grouping among markers will be biased by SD.

MATERIALS AND METHODS

Generation of mapping populations

We obtained two wild, diploid alfalfa plant introductions, PI464712 (*M. sativa* subsp. *caerulea* from Turkey) and PI631817 (*M. sativa* subsp. *falcata* from Russia), from the USDA-ARS National Plant Germplasm System (NPGS). We selected a single genotype from each

accession for hybridization to produce an F_1 population. The maternal parent was PI464712-4, and the paternal parent was PI631817-1. The two accessions were collected approximately 2400 miles apart, based on passport information in the NPGS-Germplasm Resources Information Network (GRIN). One single plant from the F_1 population ([PI464712-4 × PI631817-1]-5) was self-pollinated to generate an F_2 population.

For the F_1 population, 191 of 240 seeds germinated, and 183 survived at least long enough for DNA extraction. For the F_2 population, 161 of 237 seeds germinated and DNA was extracted from 152 F_2 plants, some of which died subsequent to DNA extraction.

SSR genotyping

Tissue from young leaves of greenhouse grown plants was freeze-dried, ground, and used for DNA extraction with the CTAB method (Doyle and Doyle, 1990). SSR markers used in previous alfalfa genetic mapping projects (Julier et al., 2003; Robins et al., 2007; Sledge et al., 2005), as well as additional markers obtained from the Noble Foundation, Ardmore, OK were used for mapping. Primer sequences are available in Supplementary Table 4.1. Primers were synthesized by Integrated DNA Technologies (IDT, http://www.idtdna.com), adding 18 nucleotides of M13 universal primer sequence onto the 5' end of the forward primer (Schuelke, 2000). The M13 universal primer sequence labeled with blue (6-FAM), green (HEX), or yellow (NED) fluorescent tags were synthesized by Applied Biosystems (http://www.appliedbiosystems.com). PCR recipes and ingredients were exactly the same as Sledge et al. (2005). The PCR program used was either 30 cycles of 94 °C / 30 sec - 55 °C / 1 min - 72 °C / 1 min (Julier et al., 2003) if the Tm is listed in Supplementary Table 4.1 as "55", or as follows: 30 cycles of 95 °C / 30 sec - 60 °C / 45 sec - 72 °C / 45 sec, followed by 9 cycles of

95 °C / 30 sec - 53 °C / 45 sec - 72 °C / 45 sec if the Tm is listed as "60" (Supplementary Table 4.1). PCR products from four to eight SSR markers were diluted ten times and pooled for each individual, mixed with 0.2ul of GeneScan-500 ROX size standard (ABI catalogue # 401734), and analyzed on an ABI 3730 DNA analyzer. The data files from the sequencer were analyzed using Genemarker software (http://www.softgenetics.com) verified by visual inspection. All markers were scored as codominant loci.

Map construction

Eighty SSR markers were used to construct the genetic linkage map for the F_1 population and 80 markers were used for the F_2 population. Seventy-six markers were common in both populations. Linkage maps for the two mapping populations were constructed using JOINMAP 4.0. Markers were grouped with a LOD $\geq$ 3.0 and map distances were calculated with the Haldane function. Primer sequences, along with the EST sequences from which they derived were used to search the *M. truncatula* pseudomolecule developed from the euchromatic gene space sequence with BLAST at the website http://medicago.org/genome/cvit_blast.php. If both forward and reverse primers hit the same BAC location with a predicted amplicon size similar to our observed fragment size, we used the position with the smaller number on the alignment between the BAC sequence and either forward or reverse primer as the physical location of the corresponding marker. If only one primer got a good hit, we only claimed a physical location if the corresponding EST also hit the same region on the same BAC.

Segregation distortion analysis

F_1 populations

The parents were not inbred because alfalfa suffers from severe inbreeding depression and inbred lines cannot be developed. As a consequence, each parent can contain up to two alleles per locus and the alleles in each parent can also be different from one another. Therefore, seven genotypic segregation patterns are possible for an F_1 population (Table 4.1). For each marker, a chi-square test was used to test the genotypic deviation from the expected Mendelian genotypic ratio (Table 4.1). A separate chi-square test was used to test the allelic deviation from the expected allelic ratio 1:1 for all segregation patterns for each parent (Table 4.1). This test is the same as the two-way pseudotestcross strategy used in previous studies for testing SD in F_1 full-sib progenies (Grattapaglia and Sederoff, 1994; Tavoletti et al., 1996).

F_2 population

A chi-square test was used to identify the markers with significant genotypic segregation distortion (SD). In the F_2 population, the expected segregation ratio of A_1A_1:A_1A_2:A_2A_2 is 1:2:1. We defined significant SD at $p<0.01$ because linked markers are not independent. A modified method of two successive chi-square tests (Lorieux et al., 1995a) was used to test selection on the gametic or zygotic level for each marker. Assuming a locus with two alleles, "A_1" and "A_2," in the F_2 mapping population, p represents the frequency of allele "A_1" and q represents the frequency of allele "A_2." The maximum likelihood estimator (MLE) of p is $\hat{p} = n_{A_1A_1} + n_{A_1A_2} / 2$ and of q is $\hat{q} = n_{A_2A_2} + n_{A_1A_2} / 2$. Allelic SD was tested as a deviation from the expected allelic ratio for A_1:A_2 of 1:1. Zygotic SD was tested as deviation from the expected genotypic ratio given the estimated the allele frequency A_1A_1:A_1A_2:$A_2A_2 = n\hat{p}^2 : 2n\hat{p}\hat{q} : n\hat{q}^2$.

The chi-square test for allelic segregation distortion is as follows:

$$\chi_1^2 = \frac{(n_{A_1A_1} - n\hat{p}^2)^2}{n\hat{p}^2} + \frac{(n_{A_1A_2} - 2n\hat{p}\hat{q})^2}{2n\hat{p}\hat{q}} + \frac{(n_{A_2A_2} - n\hat{q}^2)^2}{n\hat{q}^2} \qquad (1)$$

The chi-square test for zygotic segregation distortion is as follows:

$$\chi_2^2 = \frac{(2n\hat{p} - n)^2 + (2n\hat{q} - n)^2}{n} \qquad (2)$$

When three or more markers with significant genotypic SD were clustered together on the F_2 map, we considered the group of markers as a candidate SDR if the markers in the group showed distortion in the same direction for allele and genotype and the same pattern for allelic SD and zygotic SD test. Because markers should be more distorted if closer to the SDL, we selected the most distorted marker in each SDR as the putative SDL. Based on the relative genotypic frequency at that marker, we estimated the relative viability of each genotype. For a locus with two alleles "A_1" and "A_2", the relative viability of A_1A_1, A_1A_2, and A_2A_2 are 1, 1-hs, and 1-s, where s is the selection coefficient and h is the degree of dominance. The estimation of h and s was done as described by Luo et al., (2005). Then, the additive genetic effect (a) and dominant genetic effect (d) were estimated from the relative viabilities, where a = s/2 and h = s/2-hs. A dominance ratio (d/a) of 0 suggests additive effects at the SDL; between -1 and 1 but not equal to 0 suggests dominance, and greater than 1 suggests over-dominance.

Epistasis

To identify loci with epistatic interactions for viability, we conducted a contingency test for each pair of markers using Fisher's exact test. Pairs of markers that were linked on one chromosome generally showed non-independence because of linkage disequilibrium. Therefore, we focused on marker pairs from different chromosomes. When one or more contiguous

markers from one chromosome showed significant non-independent segregation with multiple contiguous markers from another chromosome at $p<0.01$, the two loci were considered to be involved in an epistatic interaction.

RESULTS

Construction of linkage maps

In the F_1 population, 72 of 80 SSR markers were assigned to one of ten linkage groups (LG), with the other eight markers unlinked. The ten LGs were associated with the eight alfalfa chromosomes based on the markers' physical locations and genetic locations in previous genetic maps (Figure 4.1). The F_1 linkage map covered 445 cM with an average distance between markers of 6.2 cM. In the F_2 mapping population, 70 of 80 SSR markers were assigned to twelve LGs that were associated with the eight chromosomes based on previous marker locations (Figure 4.1); ten markers were unlinked. The F_2 linkage map covered 345 cM with an average distance between markers of 4.9 cM.

Segregation distortion

In the F_1 population, 21 of 80 markers (26.3%) showed significant genotypic SD. And 10 of 54 informative markers for maternal parent (18.5%) and 7 of 58 informative markers for paternal parent (12.1%) showed significant allelic SD.

For the F_2 population, 58 of 80 markers (72.5%) showed significant genotypic SD and 50 of 70 mapped markers (71.4%) were distorted. With the exception of LG4a, LG6b, and most of LG2, essentially the entirety of the alfalfa genome expressed segregation distortion. Two of the distorted markers were on opposite ends on LG 5b; the remaining 48 markers were clustered into

9 groups located on all 8 chromosomes. Of the 9 groups, the three groups on linkage groups 4b, 5a, and 8b only contained two markers each, and hence were not considered as candidate SDRs. For other 6 groups, we identified 7 SDRs based on the skew direction of alleles and genotypes and the allelic and zygotic SD tests (Figure 4.2).

The markers in SDR1.2, SDR2.1, SDR3.1, and SDR7.1 showed zygotic SD but not allelic SD, and all the markers had excess heterozygotes (Figure 4.2 and Supplementary Table 4.2), which suggests over-dominant zygotic selection. The dominance ratios (d/a value) among these putative SDL varied from 10.75 to 79.67 (Table 4.2). The markers in SDR6.1 and SDR8.1 showed both allelic SD and zygotic SD (Figure 4.2 and Supplementary Table 4.2), which suggested a certain level of zygotic selection. The dominance ratio of putative SDL on SDR6.1 is 5.33 (Table 4.2), which suggested an over-dominant zygotic selection, but the over-dominant effect was much lower than other over-dominant SDL. The dominance ratio of putative SDL on SDR8.1 was -0.30 (Table 4.2), which suggested partial dominance zygotic selection. Only markers within SDR1.1 showed allelic SD but not zygotic SD (Figure 4.2 and Supplementary Table 4.2). Assuming zygotic selection, the dominance ratio of the putative SDL was 0.37 (Table 4.2), which suggested partial dominant selection. The markers with allelic SD had different skew directions. The markers in SDR1.1 and 6.1 skewed to the paternal parent allele. The markers in SDR8.1 skewed to maternal parent allele.

Epistatic interaction contributed to SD

Through the testing for the genotypic and allelic segregation of single markers, we have identified viability loci or SDLs and estimated their genetic effect based on a single locus model explaining additive, dominance, and over-dominance effects. To investigate if epistatic

interactions contributed to viability, we conducted a contingency test for each pair of markers using Fisher's exact test. Only one pair of regions involved in espistatic interactions was identified at $p<0.01$. The pair of regions includes six markers on LG 3 (between 0 and 5.5 cM) and three markers on LG 7 (between 30.5 and 31.5 cM). Those markers were within SDR3.1 and SDR7.1 respectively, and all of them were distorted with an excess of heterozygotes and a deficit of paternal parent homozygotes. Interestingly, the male homozygous genotype for the six markers on chromosome 3 was only observed when the three markers on chromosome 7 were heterozygous, which suggested a possible deleterious/lethal epistatic interaction.

DISCUSSION

Segregation distortion

Similar to previous studies on diploid inbred alfalfa mapping populations (Brummer et al., 1993; Kaló et al., 2000), we found a large percentage of markers, essentially the whole genome, showing SD in our F_2 population, but also many markers in the F_1 population showed SD as well. Every chromosome except for chromosome 4 had at least one cluster of distorted markers, which indicated that SDLs were located on most of the chromosomes. A relatively low incidence of allelic SD compared to genotypic SD and zygotic SD suggested that zygotic selection is more common than gametic selection for SDLs in the F_2 mapping population.

Only markers in SDR1.1 showed allelic SD but not zygotic SD, which suggested gametic selection as defined by two successive chi-square tests (Lorieux et al., 1995a). However, because we collected genotypic data from germinated plants, selection before fertilization (gametic selection) could not be differentiated from selection after fertilization but before germination (zygotic selection). Therefore, the genotypic SD in SDR1.1 also could be due to

zygotic selection with a partial dominant effect. Gametic selection is not equal to dominant zygotic selection, but they could generate the same allelic and genotypic segregation ratios. Except for SDR1.1 and SDR8.1, distorted markers in all SDRs showed heterozygote excess. Segregation distortion with heterozygote excess was common in diploid alfalfa; 30 of 51 distorted markers were skewed to heterozygotes in a diploid *sativa* segregating population (Brummer et al., 1993), and distorted marker regions on all 8 chromosomes except for chromosome 4 had heterozygote excess in a *falcata* x *caerulea* population (Kaló et al., 2000). SD with heterozygote excess has also been observed in other species (Faris et al., 1998; Fishman et al., 2001; Lu et al., 2002); however, none of them show extreme distortion like alfalfa, in which SD with heterozygote excess predominates. Only 5 of 55 (9.1%) distorted markers were skewed toward heterozygotes in maize (Lu et al., 2002) and 6 of 57 (10.5%) distorted markers s in *Aegilops tauschii* (Faris et al., 1998). A few distorted markers with heterozygote excess were also observed in lettuce (Kesseli et al., 1994; Landry et al., 1987), pepper (Prince et al., 1993), and *Mimulus guttatus* (Fishman et al., 2001).

The common distortion toward heterozygote excess in alfalfa suggests that over-dominant zygotic selection is predominant. Fitness is a quantitative trait, controlled by multiple genes (Falconer and Mackay, 1996; Luo and Xu, 2003; Luo et al., 2005). Therefore, pseudo-overdominance may also explain the heterozygote excess SD in alfalfa, and is the more likely explanation. Two or more repulsion linked SDL with a dominant effect at each locus will generate markers with observed heterozyote excess. Dominance appears to be the major cause of biomass heterosis in alfalfa (Bingham et al., 1994). The SD with heterozygote excess was commonly observed in conifer populations, which was most likely explained by pseudo-overdominance of multiple lethal factors, although gametic selection between maternal and

paternal gametes is also possible (Williams et al., 2001; Williams et al., 2003). Self-incompatibility loci can be comprised tightly linked male- and female-specific genes, which are inherited as one single functional unit (Takayama and Isogai, 2005). Thus, the SD of markers with heterozygote excess caused by linked self-incompatibility loci could be considered as an example of pseudo-overdominance.

Dobzhansky-Muller (DM) incompatibilities are epistatic interactions between unlinked markers. Lethal DM incompatibilites could contribute to SD and could be identified by testing pair-wise associations of marker genotypes (McDaniel et al., 2007; Payseur and Hoekstra, 2005). For testing possible deleterious or lethal epistatic interactions, we conducted a contingency test for each pair of markers using Fisher's exact test. We found one pair of regions (SDR7.1 and SDR3.1) involved in epistatic interactions. The homozygous paternal parent genotype at markers from SDR3.1was found only when markers from SDR7.1 were heterozygous. The same phenomenon was found for all 18 marker-pairs between the two SDRs. The mechanism behind this interaction needs to be investigated further.

Based on the physical location of SDR8.1, SD with excess of one homozygote was also observed in *Medicago truncatula* (Choi et al., 2004).

Comparison of SD pattern between F_1 and F_2 population

Compared to SD in the F_2 population, a smaller percentage of loci exhibited both genotypic SD and allelic SD in the F_1 mapping population. We found two possible explanations through comparing the SD patterns of loci in SDRs between F_1 and F_2 populations. The first reason is that some markers that showed SD in the F_2 population were not fully informative in the F_1, and hence, SD may have occurred but could not be observed. Markers in SDR1.1, SDR1.2 and SDR8.1 (Figure 4.1 and Supplementary Table 4.2) fall into this class; linked loci in

those SDRs that were fully informative showed genotypic SD in the F_1 population and allelic SD for one parent, but the non-fully informative markers did not show SD even though they were closely linked to the distorted markers. This could partly explain the low level of SD found in diploid F_1 full-sib progenies in alfalfa (Tavoletti et al., 1996). The second reason is that the inbreeding level is different between the F_1 and F_2 populations. The SD may not have occurred in the F_1 population because of high heterozygosity (low inbreeding level), even if the genotypes of markers around the SDLs are fully informative.

The out-crossing nature of alfalfa causes individuals to carry many recessive deleterious alleles. The inbreeding level could play important role for SD of marker genotypes/gene genotypes. The SD of markers in SDR2.1, SDR3.1, and SDR7.1 could not be found in F_1 in this study, likely due to the low inbreeding level in F_1 population.

SD effect on genetic distances and orders among markers

The estimation of genetic distance between a distorted marker and a non-distorted marker should not be biased, but the distance between two distorted markers may be biased (Bailey, 1949; Lorieux et al., 1995a; Lorieux et al., 1995b). The bias level depends on the distortion level and the real genetic distance between the markers (Lorieux et al., 1995b). One SDL generally generates a cluster of distorted markers with the same skew direction, which will bias the estimated genetic distances. From theoretical and simulation studies, when a group of markers is distorted in the same direction, the genetic distance between pairs of markers will be underestimated; when distorted in different directions, the genetic distance between the pair of markers will be overestimated (Lorieux et al., 1995b; Zhu et al., 2007a; Zhu et al., 2007b). Therefore, the estimated genetic distances among the markers within each SDR in this study is

likely underestimated. In LG1, the two SDRs included SDL with different genetic effects (dominant and overdominant).

The precise ording and grouping of markers depends on the precise estimation of pairwise genetic distances. When two of three markers are severely distorted with the same or different skew directions in a backcross population, the order among the 3 markers could be biased, but not necessarily (Lorieux et al., 1995b). The SD pattern and the severity of SD in an F_2 population needed to cause a biased order among markers is unknown. From our study, the markers in some SDRs may have a biased order based on comparisons of the F_1, F_2, and physical maps. Although most markers were distorted in the F_2 population, all markers except for markers on chromosome 5 were correctly assigned to each linkage group/chromosome. The markers on chromosome 5 were assigned to two different linkage groups in the F_2 population, but one linkage group in the F_1 population. Although several approaches were investigated to adjust genetic distance between distorted markers considering SDL effect (Mitchell-Olds, 1995; Vogl and Xu, 2000; Zhu et al., 2007a), none of them was integrated into a standard mapping program.

Implication for QTL mapping and breeding in alfalfa

The F_1 population had a lower level of SD, which is good for precise map construction and QTL mapping. However, if markers are not full informative, QTLs might be underestimated. For example, a favorable QTL allele linked to a homozygous marker genotype could not be identified efficiently. In this study, only 39 out of 80 SSR markers were fully informative for the F_1 population. Adding more fully informative markers would create a denser map.

Partially inbred F_2 populations have high levels of SD, which bias the estimation of genetic distance and possibly order among markers, further affected QTL mapping. However, unlike F_1 populations, all markers in an F_2 population are fully informative. After identifying SDLs and estimating their effects, reconstruction of a corrected linkage map with distorted markers should be meaningful for both F_1 and inbred populations. Reduction of SD can be effected by developing advanced generation populations, in which recombination can break the linkage between markers and SDLs. Then precise mapping can be conducted to more specifically pinpoint the actual location of SDLs.

Evolutionary geneticists are interested in the genes controlling viability. Plant breeders are interested in the genes controlling agronomic traits like yield. Theoretically, it is possible that the entire chromosome could be distorted by just a few SDL (Xu, 2008). The SDRs were identified on most of 8 chromosomes in alfalfa and multiple SDRs could be located on one chromosome from this study. This indicates that segregation of a substantial number of genes could be affected, including genes controlling important agronomic traits.

The markers with allelic SD had different skew directions in this study, which indicated genetic divergence of SDL between *M. sativa* subsp. *caerulea* and *M. sativa* subsp. *falcata*. The genes controlling biomass in a previous study also came from both *M. sativa* subsp. *sativa* and *M. sativa* subsp. *falcata* (Robins et al., 2007). Alfalfa breeders generally make selections after crossing germplasms, including *M. sativa* ssp. *sativa* and *M. sativa* ssp. *falcata*. If the SDL and QTL alleles for agronomic traits are linked in repulsion, then the selection or introgression on agronomic traits could be affected by SDL. For example, if subsp. *sativa* has a favorable allele at an SDL that is linked to an unfavorable allele at a yield locus, then a beneficial allele for yield from *M. sativa* subsp. *falcata* could not be integrated into subsp. *sativa* and fixed efficiently.

Most SD found in this study was heterozygote excess, which could be due to pseudo-overdominance of two linked SDL, in which case, desirable yield genes could be in repulsion-phase linkages. Therefore, it is meaningful to map SDL and QTL for agronomic traits in same population, which can facilitate their selection and introgression.

REFERENCES

Bailey, N.T.J. 1949. The estimation of linkage with differential viability, II and III. Heredity 3:220-228.

Bingham, E.T., R.W. Groose, D.R. Woodfield, and K.K. Kidwell. 1994. Complementary gene interaction in alfalfa is greater in autotetraploids than diploid. Crop Sci. 34:823-829.

Brouwer, D.J., and T.C. Osborn. 1999. A molecular marker linkage map of tetraploid alfalfa (*Medicago sativa* L.) Theor Appl Genet 99:1194:1200.

Brummer, E.C., J.H. Bouton, and G. Kochert. 1993. Development of an RFLP map in diploid alfalfa Theor Appl Genet 86:329:332.

Chetelat, R.T., V. Meglic, and P. Cisneros. 2000. A genetic map of tomato based on BC(1) *Lycopersicon esculentum* x *Solanum lycopersicoides* reveals overall synteny but suppressed recombination between these homeologous genomes. Genetics 154:857-67.

Choi, H.K., D. Kim, T. Uhm, E. Limpens, H. Lim, J.H. Mun, P. Kalo, R.V. Penmetsa, A. Seres, O. Kulikova, B.A. Roe, T. Bisseling, G.B. Kiss, and D.R. Cook. 2004. A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. Genetics 166:1463-502.

Doyle, J.J., and J.I. Doyle. 1990. Isolation of plant DNA from fresh tissue. BRL Life Technologies: Focus 12:13–15.

Echt, C.S., K.K. Kidwell, S.J. Knapp, T.C. Osborn, and T.J. McCoy. 1994. Linkage mapping in diploid alfalfa (*Medicago sativa*). Genome 37:61-71.

Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to Quantitative Genetics. 4 ed., Longman, Essex, England.

Faris, J.D., B. Laddomada, and B.S. Gill. 1998. Molecular mapping of segregation distortion loci in Aegilops tauschii. Genetics 149:319-27.

Fishman, L., A.J. Kelly, E. Morgan, and J.H. Willis. 2001. A genetic map in the Mimulus guttatus species complex reveals transmission ratio distortion due to heterospecific interactions. Genetics 159:1701-16.

Grattapaglia, D., and R. Sederoff. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137:1121-37.

Julier, B., S. Flajoulot, P. Barre, G. Cardinet, S. Santoni, T. Huguet, and C. Huyghe. 2003. Construction of two genetic linkage maps in cultivated tetraploid alfalfa (*Medicago sativa*) using microsatellite and AFLP markers. BMC Plant Biol 3:9.

Kaló, P., G. Endre, L. Zimányi, G. Csanádi, and G.B. Kiss. 2000. Construction of an improved linkage map of diploid alfalfa (*Medicago sativa*) Theor Appl Genet 100:641:657.

Kesseli, R.V., I. Paran, and R.W. Michelmore. 1994. Analysis of a detailed genetic linkage map of *Lactuca sativa* (lettuce) constructed from RFLP and RAPD markers. Genetics 136:1435-46.

Kiss, G.B., G. Csanadi, K. Kalman, P. Kalo, and L. Okresz. 1993. Construction of a basic genetic map for alfalfa using RFLP, RAPD, isozyme and morphological markers. Mol Gen Genet 238:129-37.

Landry, B.S., R.V. Kesseli, B. Farrara, and R.W. Michelmore. 1987. A Genetic Map of Lettuce (*Lactuca sativa* L.) with Restriction Fragment Length Polymorphism, Isozyme, Disease Resistance and Morphological Markers. Genetics 116:331-337.

Lorieux, M., X. Perrier, B. Goffinet, C. Lanaud, and D.G.d. León. 1995a. Maximum-likelihood models for mapping genetic markers showing segregation distortion. 2. F2 populations Theor Appl Genet 90:81-89.

Lorieux, M., X. Perrier, B. Goffinet, C. Lanaud, and D.G.d. León. 1995b. Maximum-likelihood models for mapping genetic markers showing segregation distortion. 1. Backcross populations Theor Appl Genet 90:73-80.

Lu, H., J. Romero-Severson, and R. Bernardo. 2002. Chromosomal regions associated with segregation distortion in maize. Theor Appl Genet 105:622-628.

Luo, L., and S. Xu. 2003. Mapping viability loci using molecular markers. Heredity 90:459-67.

Luo, L., Y.M. Zhang, and S. Xu. 2005. A quantitative genetics model for viability selection. Heredity 94:347-55.

Mangelsdorf, P.C., and D.F. Jones. 1926. The Expression of Mendelian Factors in the Gametophyte of Maize. Genetics 11:423-55.

McDaniel, S.F., J.H. Willis, and A.J. Shaw. 2007. A linkage map reveals a complex basis for segregation distortion in an interpopulation cross in the moss *Ceratodon purpureus*. Genetics 176:2489-500.

Mitchell-Olds, T. 1995. Interval mapping of viability loci causing heterosis in Arabidopsis. Genetics 140:1105-9.

Paterson, A.H., E.S. Lander, J.D. Hewitt, S. Peterson, S.E. Lincoln, and S.D. Tanksley. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-6.

Payseur, B.A., and H.E. Hoekstra. 2005. Signatures of reproductive isolation in patterns of single nucleotide diversity across inbred strains of mice. Genetics 171:1905-16.

Prince, J.P., E. Pochard, and S.D. Tanksley. 1993. Construction of a molecular linkage map of pepper and a comparison of synteny with tomato. Genome 36:404-17.

Rhoades, M.M. 1942. Preferential Segregation in Maize. Genetics 27:395-407.

Riaz, S., A.C. Tenscher, J. Rubin, R. Graziani, S.S. Pao, and M.A. Walker. 2008. Fine-scale genetic mapping of two Pierce's disease resistance loci and a major segregation distortion region on chromosome 14 of grape. Theor Appl Genet 117:671-81.

Robins, J.G., D. Luth, T.A. Campbell, G.R. Bauchan, C. He, D.R. Viands, J.L. Hansen, and E.C. Brummer. 2007. Genetic Mapping of Biomass Production in Tetraploid Alfalfa. Crop Sciences 47:1-10.

Sandler, L., Y. Hiraizumi, and I. Sandler. 1959. Meiotic Drive in Natural Populations of Drosophila Melanogaster. I. the Cytogenetic Basis of Segregation-Distortion. Genetics 44:233-50.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol 18:233-4.

Sledge, M.K., I.M. Ray, and G. Jiang. 2005. An expressed sequence tag SSR map of tetraploid alfalfa (*Medicago sativa* L.). Theor Appl Genet 111:980-92.

Takayama, S., and A. Isogai. 2005. Self-incompatibility in plants. Annu Rev Plant Biol 56:467-89.

Tavoletti, S., F. Veronesi, and T.C. Osborn. 1996. RFLP Linkage Map of an Alfalfa Meiotic Mutant Based on an F1 Population Journal of Heredity 87:167:170.

Vogl, C., and S. Xu. 2000. Multipoint mapping of viability and segregation distorting loci using molecular markers. Genetics 155:1439-47.

Williams, C.G., Y. Zhou, and S.E. Hall. 2001. A chromosomal region promoting outcrossing in a conifer. Genetics 159:1283-9.

Williams, C.G., L.D. Auckland, M.M. Reynolds, and K.A. Leach. 2003. Overdominant lethals as part of the conifer embryo lethal system. Heredity 91:584-92.

Xu, S. 2008. Quantitative trait locus mapping can benefit from segregation distortion. Genetics 180:2201-8.

Xu, Y., L. Zhu, J. Xiao, N. Huang, and S.R. McCouch. 1997. Chromosomal regions associated with segregation distortion of molecular markers in F2, backcross, doubled haploid, and recombinant inbred populations in rice (*Oryza sativa* L.). Mol Gen Genet 253:535-45.

Yin, T.M., S.P. DiFazio, L.E. Gunter, D. Riemenschneider, and G.A. Tuskan. 2004. Large-scale heterospecific segregation distortion in Populus revealed by a dense genetic map. Theor Appl Genet 109:451-63.

Zamir, D., and Y. Tadmor. 1986. Unequal segregation of nuclear genes in plants. Bot. Gaz. 147:355:358.

Zhu, C., C. Wang, and Y.M. Zhang. 2007a. Modeling segregation distortion for viability selection. I. Reconstruction of linkage maps with distorted markers. Theor Appl Genet 114:295-305.

Zhu, C.S., F.H. Wang, J.F. Wang, G.J. Li, H.S. Zhang, and Y.M. Zhang. 2007b. Reconstruction of linkage maps in the distorted segregation populations of backcross, doubled haploid and recombinant inbred lines. Chinese Science Bulletin 52:1648-1653.

Table 4.1. Expected allelic and genotypic segregation patterns in an F_1 population derived from non-inbred parents.

Allele No.	Segregation type	Expected allelic ratio		Expected genotypic ratio
		maternal parent	paternal parent	
2	$A_1A_2 \times A_1A_1$	A_1: A_2 = 1:1		A_1A_2: A_1A_1 = 1:1
2	$A_1A_1 \times A_1A_2$		A_1: A_2 = 1:1	A_1A_1: A_1A_2 = 1:1
2	$A_1A_2 \times A_1A_2$			A_1A_1: A_1A_2: A_2A_2 = 1:2:1
3	$A_1A_2\times A_3A_3$	A_1: A_2 = 1:1		A_1A_3: A_2A_3 = 1:1
3	$A_1A_1 \times A_2A_3$		A_2: A_3 = 1:1	A_1A_2: A_1A_3 = 1:1
3	$A_1A_2 \times A_1A_3$	A_1: A_2 = 1:1	A_1: A_3 = 1:1	A_1A_1: A_1A_3: A_1A_2: A_2A_3 = 1:1:1:1
4	$A_1A_2 \times A_3A_4$	A_1: A_2 = 1:1	A_3: A_4 = 1:1	A_1A_3: A_1A_4: A_2A_3: A_2A_4 = 1:1:1:1

Table 4.2. Segregation distortion loci (SDL) in segregation distortion regions (SDR) in a diploid alfalfa F_2 population, their additive and dominance effects, and the dominance/additive ratio.

Marker	SDR	a	d	d/a
bg208	SDR1.1	0.36	0.13	0.37
bg249	SDR1.2	0.06	0.90	15.00
bf644149	SDR2.1	0.22	2.88	13.14
mtic237	SDR3.1	0.28	2.97	10.75
bf69	SDR6.1	0.47	2.53	5.33
aw212	SDR7.1	0.19	14.94	79.67
aw685868	SDR8.1	0.50	-0.15	-0.30

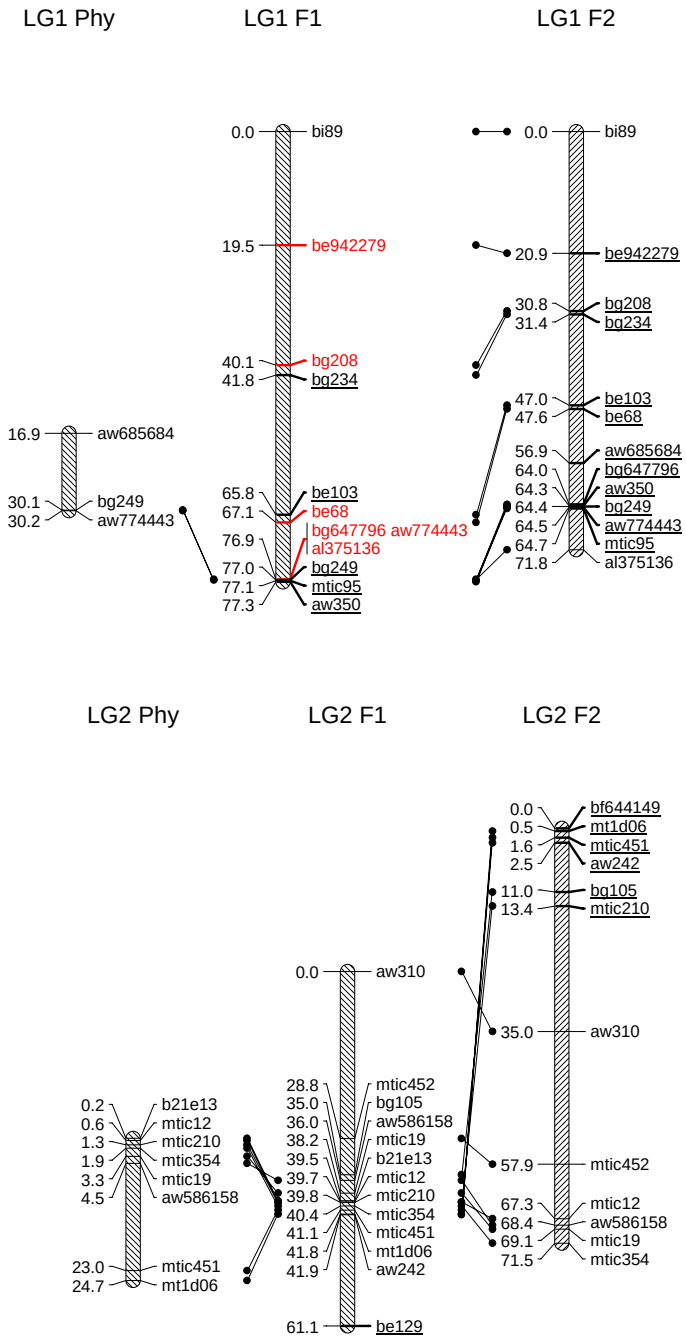


Figure 4.1. Physical map and genetic linkage maps for F_1 and F_2 populations. The physical locations indicated on the maps are all in the scale of $\times 10^{\times 6}$ base pairs. Underlined markers were significantly distorted at $p<0.01$. Markers that showed SD in the F_2 population but did not show SD in the F_1 population possibly due to the marker not being fully informative are labeled in red.

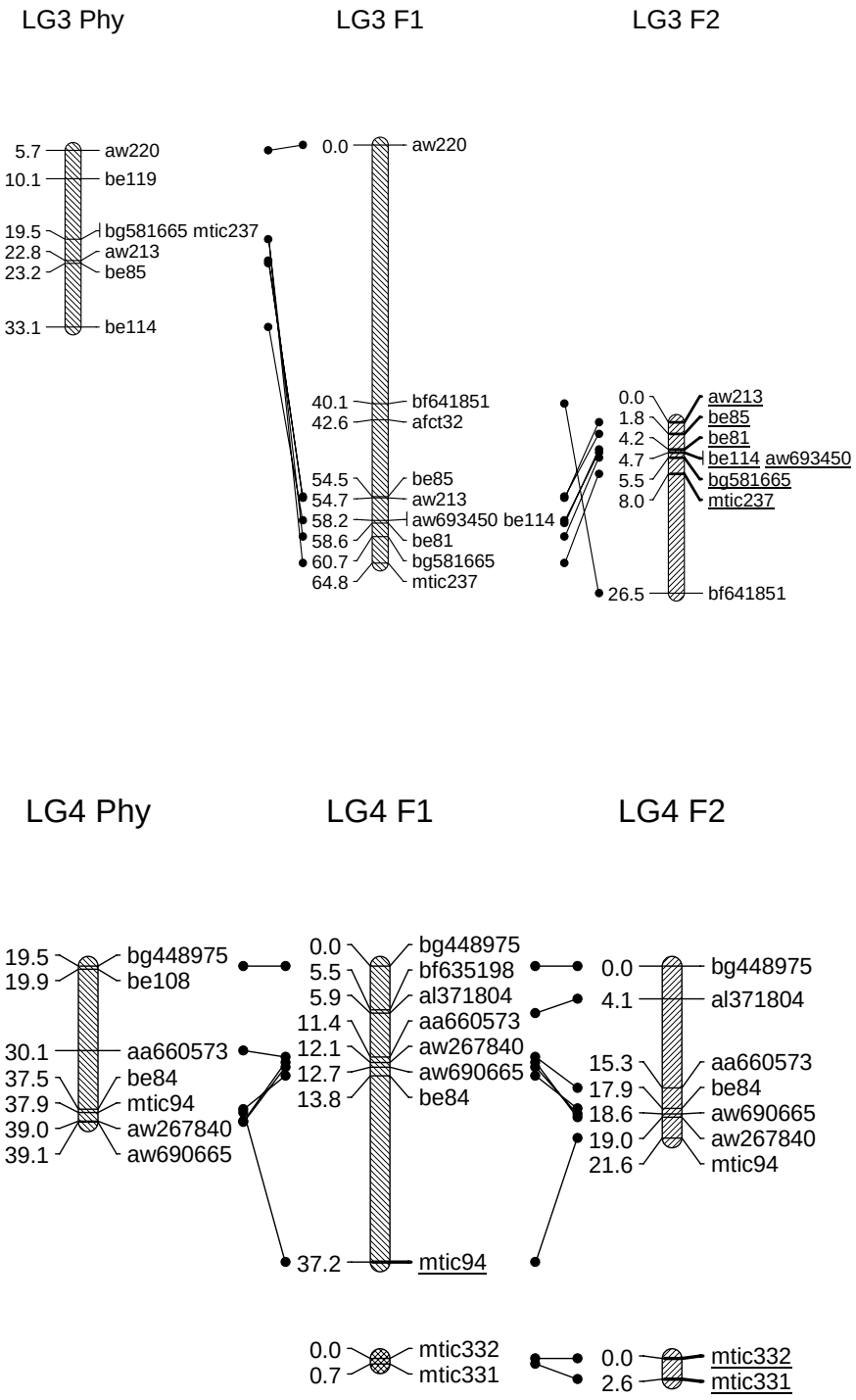


Figure 4.1. Continued.

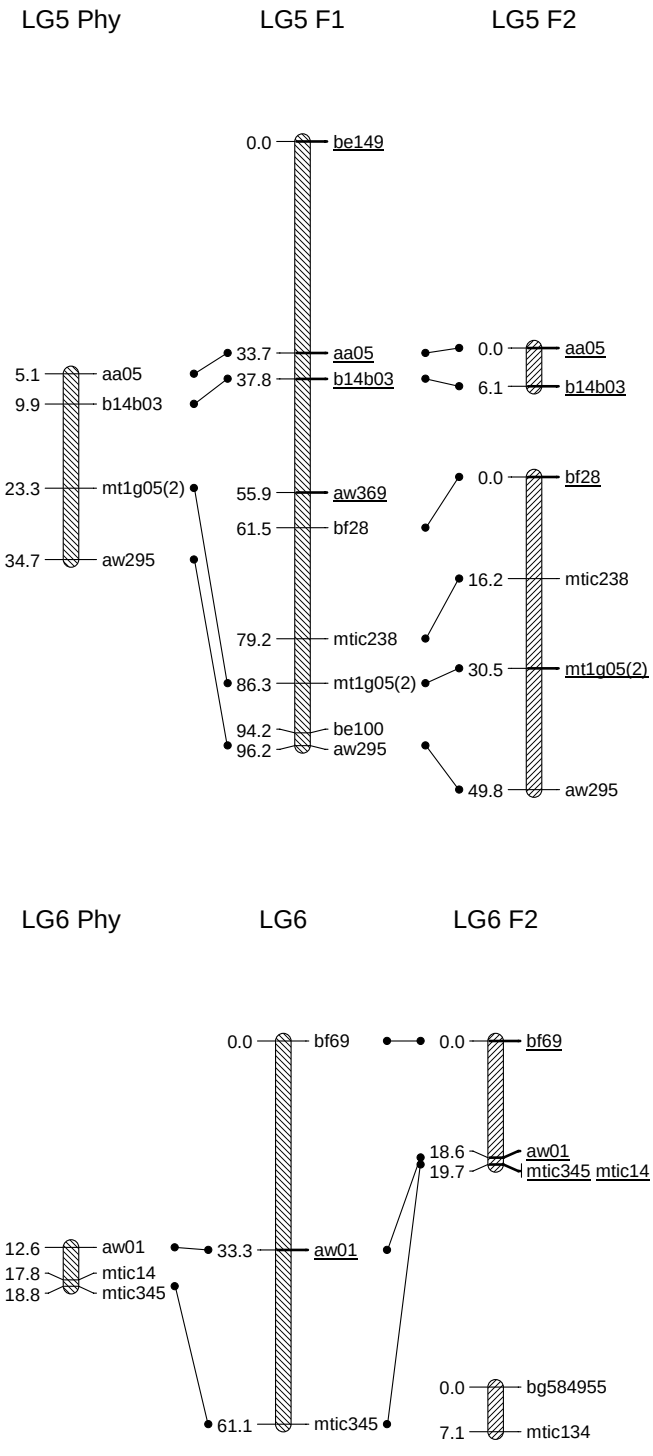


Figure 4.1. Continued.

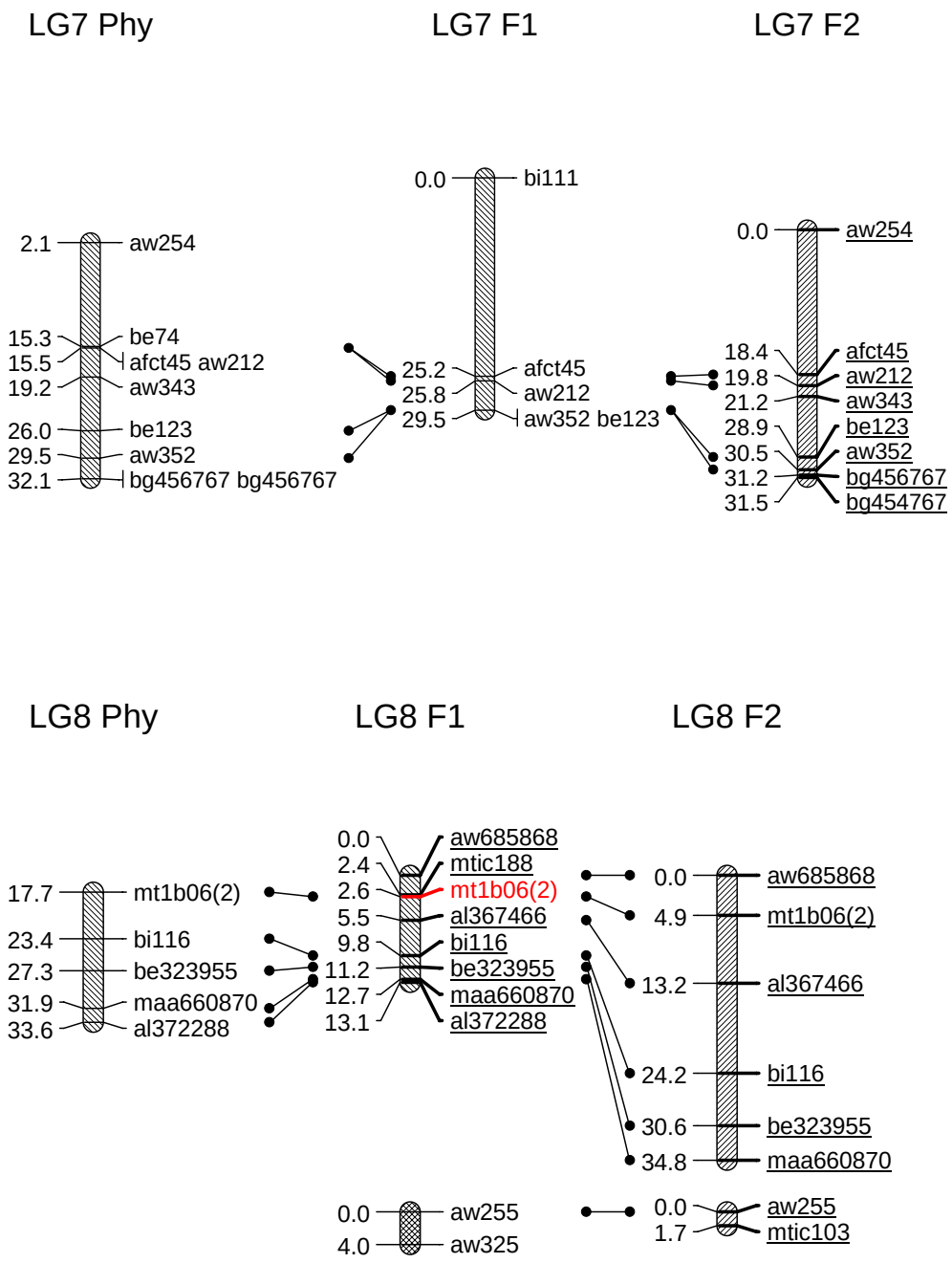


Figure 4.1. Continued.

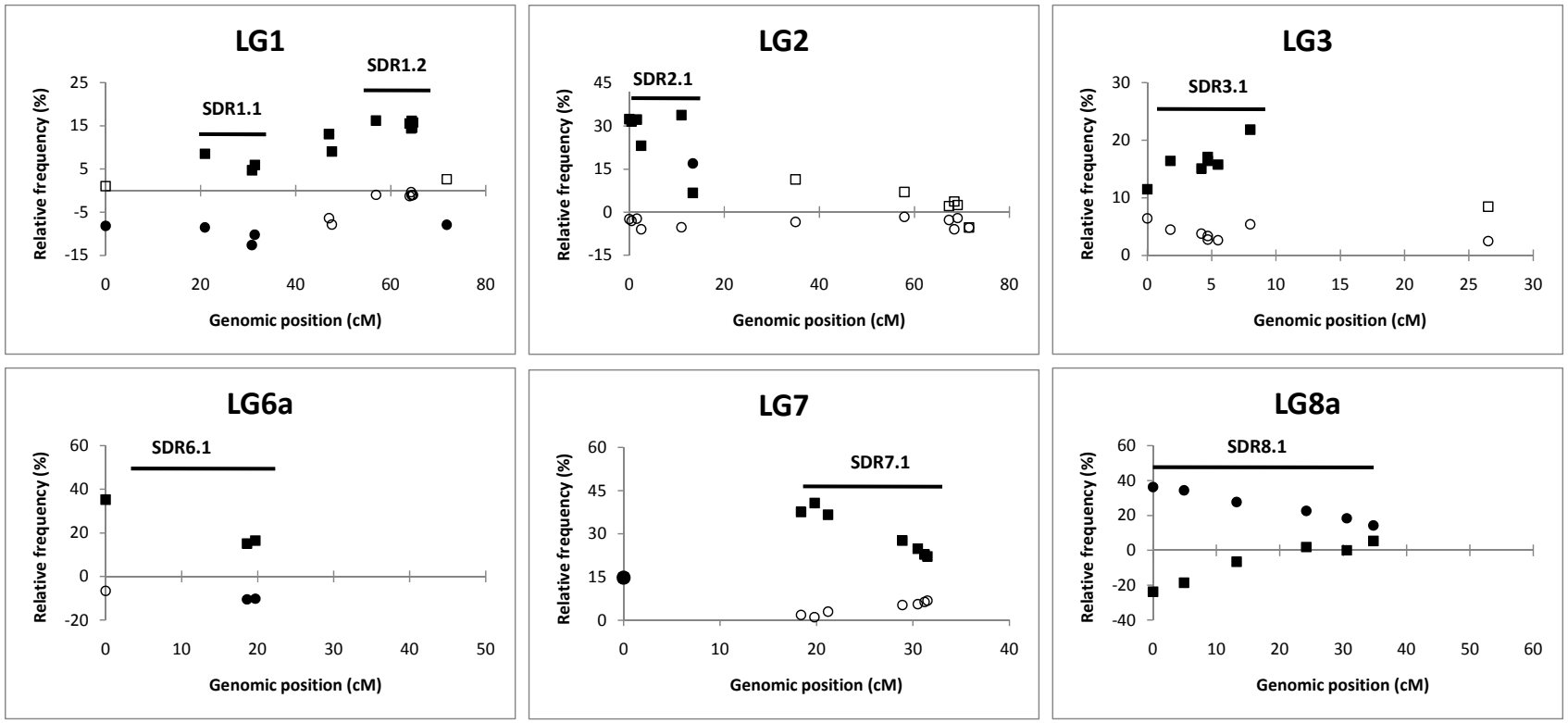


Figure 4.2. Relative frequencies of heterozygotes and maternal alleles at SSR markers across six alfalfa chromosomes in an F_2 population. Squares represent the deviation of heterozygote frequency at each locus from the expected frequency of 0.5; solid squares indicate that the marker showed genotypic SD. Circles represent the deviation of the maternal parent allele from its expected frequency of 0.5; solid circles indicate the marker showed allelic SD. The black bars represent the location of segregation distortion regions (SDRs).

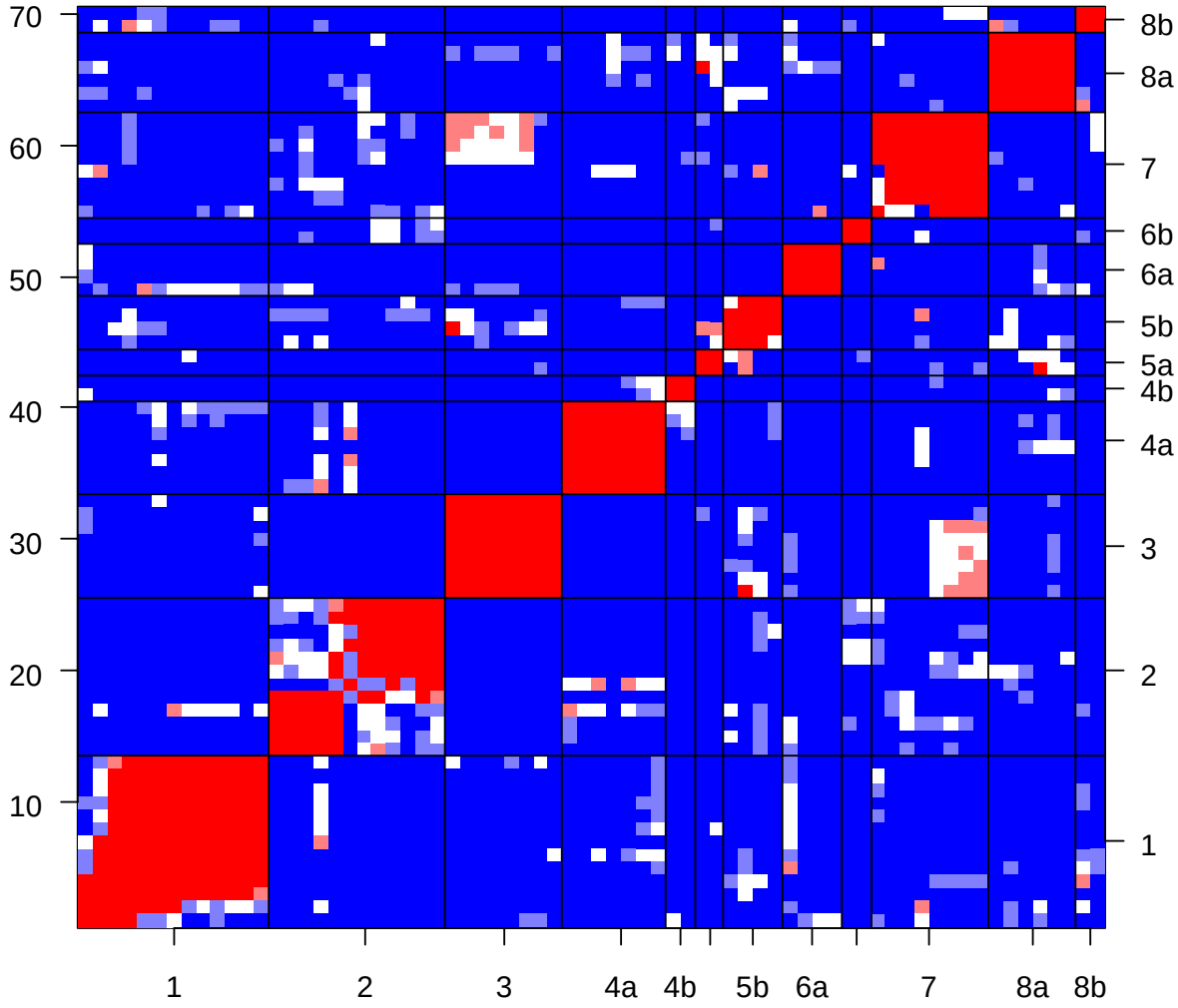


Figure 4.3. Genomewide map of pairwise genotype associations, measured by p value. The ten linkage groups we identified are shown along the x- and y-axes. The p value for the significance of the test for pairwise interaction for each pair of loci is color coded as follows: red = $p<0.001$; light red = $p<0.01$; white = $p<0.05$; light blue = $p<0.1$; and blue = $p\leq1$.

Supplementary Table 4.1. Attributes of genetic markers used to create the F_1 and F_2 genetic maps in diploid alfalfa.

Marker	Tm (°C)	Forward	Reverse
bg456767	55	TCCCTTAAATCCGTGGCTCT	TTCCCATGCAGAAGAAATCC
be323955	55	CACACTCTCTCTTCTCCGGTTC	TGTCGTCAGTGGTGGTTGTTA
al367466	55	CCTCCACACTCTCTCTCTTGC	CCGAACAATTCTCCGATGAT
aw685868	55	AAGCAAGTTCTGTTGATGGAGA	TTGTGAAAGCCAAAACACCA
bf518447	55	ATGCTACCAAGGCTGCTGAT	CACATACTTGAAGGGATGACAAG
bg581665	55	AACGGTGGTGTGTTTATTGCT	TTCCCATATGCAACAGACCTT
afct45	55	TAAAAAACGGAAAGAGTTGGTTAG	GCCATCTTTTCTTTTGCTTC
maa660456	55	GGGTTTTTGATCCAGATCTT	AAGGTGGTCATACGAGCTCC
maa660870	55	GTACATCAACAACTTTCTCCT	ATCAACAAAATTCATCGAAC
b14b03	55	GCTTGTTCTTCTTCAAGCTC	ACCTGACTTGTGTTTTATGC
b21e13	55	GCCGATGGTACTAATGTAGG	AAATCTTGCTTGCTTCTCAG
mtic19	55	TCTAGAAAAAGCAATGATGTGAGA	TGCAACAGAAGAAGCAAAACA
mtic93	55	AGCAGGATTTGGGACAGTTG	TACCGTAGCTCCCTTTTCCA
mtic94	55	GCTACAACAGCGCTACATCG	CAGGGTCAGAGCAACAATCA
mtic95	55	AAAGGTGTTGGGTTTTGTGG	AGGAAGGAGAGGGACGAAAG
mtic103	55	TGGGTTGTCCTTCTTTTTGG	GGGTGCAGAAGTTTGACCA
mtic134	55	GCAGTTCGCTGAGGACTTG	CAATTAGAGTCTACAGCCAAAAACT
mtic183	55	AAATGGAAGAAAGTGTCACG	TTCTCTTCAAGTGGGAGGTA
mtic188	55	GGCGGTGAAGAAGTAAACGA	AATCGGAGAAACACGAGCAC
mtic210	55	CCAAACTGGCTGTGTTCAAA	GCGGTAAGCCTTGCTGTATG

Supplementary Table 4.1. Continued.

Marker	Tm (°C)	Forward	Reverse
mtic237	55	CCCATATGCAACAGACCTTA	TGGTGAAGATTCTGTTGTTG
mtic238	55	TTCTTCTTCTAGGAATTTGGAG	CCTTAGCCAAGCAAGTAAAA
mtic331	55	CCCTCTTCTACCTCCTTTCCA	GGAAGAGAAGATGGGGGTGT
mtic332	55	CCCTGGGTTTTTGATCCAG	GGTCATACGAGCTCCTCCAT
mtic345	55	TCCGATCTTGCGTCCTAACT	CCATTGCGGTGGCTACTCT
mtic354	55	AAGTGCCAAAGAACAGGGTTT	AACCTACGCTAGGGTTGCAG
mtic451	55	GGACAAAATTGGAAGAAAAA	AATTACGTTTGTTTGGATGC
mtic452	55	CTAGTGCCAACACAAAAACA	TCACAAAAACTGCATAAAGC
aw220	55	GCCACAATTTTCTCATCATCAC	TGCTGCTGTGCCGTAGTAGATA
aw242	55	CGTCACACCAACTTTATCACCA	TTCTCCAAGAGGAGCTTAACCA
aw254	55	TATATGCTTGTTGAGGCCACTG	CACATCTTCGTCATCATCTTCA
aw256	55	ACCACTACTGCGTTTGTTTGTG	TAAGGAGTTTGGAATGGGAAGA
aw258	55	AATTGGAACCTATCGTTGTCGT	GAGTATCGGAAGAGGGTTGTTG
aw295	55	CAACATTCTTCCATTTCCTTCC	TCTTCATCTTCGTCGTCTTCAA
aw343	55	GGTTCGTGTATTTGTTCGATCC	AATCTCCAAGGTTCCATCTTCA
aw350	55	GAACCTCTTCTTCTCTCACACAAA	AGCATCGAACGGATGTAATTG
aw369	55	GCGCTCATCATCTTCATCTAAA	AGAATTGAGACATGGCAGAGG
aw373	55	TATCATCCTGGTTCGTTCCTCT	GGTTGAGCTTGAGAAAATCTGA
be100	55	GCATTAGCACCCTCATTCATATC	TGCAGAGACTTTTGAACACCTT
be103	55	AATGGCGAACACTTTCACTCTT	GATGGTTTCTTCGAGACGAGAG

Supplementary Table 4.1. Continued.

Marker	Tm (°C)	Forward	Reverse
be108	55	CTCCTTCATCCGTTTCTCAAAA	CGATGTTTGCCCAGAATGA
be114	55	CCACCTCATCACTCCGTAAAA	ATGAAGCTGTTGTTGTTGCAGT
be119	55	GCTAGTTCTGCTCTCACTCTCATC	CATTGTCTTTGTTGTGGAGGTG
be123	55	ATCACAAGCCTCAACAGCCATA	TTGATGGGTAAAGGAGAAGGTG
be129	55	GAAGTTGCGTCAGAGAGATCAG	AACAAACACAGGCTTCACCATA
be131	55	GCAACTCTTTCTCACTCACCA	GTTGAGTGGTGGCATTTGAAC
be149	55	GTGTTTGGGAGATTTTGAGGAG	GCATGATAGCAAGTGGAACCATA
be68	55	TCTGTTTACCACACGCAACTTC	CAGAAGCCATTAGCCTGAACAT
be74	55	TACTGTCCCAATCTTCACAACG	GCACAAGCAGCCATATTGATAG
be81	55	CGCTCTTGTTGATTCTGCTATG	TTACTCTTCTCCTTGGCAGCTT
be84	55	TCCGAACCCTACTTCCAAATTA	TGGGATACTGATTTTCTGCTTC
be85	55	TTTCCTCTTATTATTCTTTCATACCC	CTGATTCGAGATTGGGATTGAT
al372288	55	AAATATCCATTCGATACAATGACC	CAGGCCTCAAGCTAAGAACTGT
aa660573	55	TTCCGCCCATAGTCTTTGAC	TAAATGTGTCCTGCGTCTGG
afct32	55	TTTTTGTCCCACCTCATTAG	TTGGTTAGATTCAAAGGGTTAC
al371804	55	TCATGTTGCAGTTGGAAGGA	TGGTTCTTAATTTTATCCATCATCA
al375136	55	CATGTTTTTGTTTTGTTGGGG	TTCAGCTTAGCAACAAGTCTCAA
aw267840	55	GGGATTCGACGCTTATTTCA	GAAGACGGTGTTGGTAGGGA
aw586158	55	GATCAATTCGTGCAGAAGCA	ATTCATCCTTGCTCGTTTCG
aw685684	55	ATCGTCCCCACTGTGTCTTC	GTGGGGTTGGTGAGAGTGTT

Supplementary Table 4.1. Continued.

Marker	Tm (°C)	Forward	Reverse
aw690665	55	GGTTTTGGAGACATGACGGT	GTGAAGACTTTGCGGTGGAT
aw693450	55	TTTTTCACAGCACTGAAGAGG	CCATGGCGTCTACCCATTAT
aw774443	55	ATTCGCAGTGAGCTGATCCT	GACATTTGCAGACCACCATT
be323955	55	CACACTCTCTCTTCTCCGGTTC	TGTCGTCAGTGGTGGTTGTTA
be942279	55	GGTGTGTGCGGTTTAGGACT	CCTCCACATAGCTGGTCGAT
bf635198	55	ACCCCTGCTGAAACAGCATA	CTCTCCCCTAGCCTCAAAGC
bf641851	55	GAAGACACAATGAGTAGCAGAGC	TGGCCCATGTTTCTCAGACT
bf644149	55	CTTTTGCGAAATTTCCTTGC	CCTTGGTGCTTCTCATGTCA
bg448975	55	TCGGATCTGACACGATTTTG	TTGGTTAAAAGATGAAGATGAACG
bg454767	55	ACTCCCTTAAATCCGTGGCT	CACTGGAACCACGAACCTTT
bg647796	55	GCAAGAAAGCATAGGCTGAGA	GTGAAGCTGCACGAATTTCA
bg648700	55	GCTTTTCACACCTCCACTCC	ACGGGAAAGACTCCCACTCT
mtic12	55	TTCCTCTTTTGACTCATCACCA	CAACAACATGTTCATGCTTAGAAAC
mtic14	55	CAAACAAACAACACAAACATGG	CCCATTGATTGGTCAAGGTT
aa05	60	CCTTCTGCCATTCATTTCACTT	CTTCAAAGGGTCATCAAATCAC
aw01	60	ACCTGTTCTAAGGGAGATTTCG	CAGGGGAAGCATACAAAACC
aw212	60	GTCGAAATGGTTGCTTCTCTTT	GGTTAGGGTTTTGGGTTTGAA
aw213	60	ACCCTTGTGGGTTCTTCTTCTT	CATGTACGGGGATTGTTGTTTT
aw255	60	TCTCTCCATCATCACCATCATC	TGCTTGAACTTTGAGTCTTGGA
aw310	60	CCACTCAACCTCATCTCTCTACC	CAATGCAAGAAACCCTAAAAGC

Supplementary Table 4.1. Continued.

Marker	Tm (°C)	Forward	Reverse
aw325	60	TCTGTAAGAGGGTCACTGCGTA	GCTTGTTGTTGTTGTTGATGCT
aw352	60	ACGTTCCTCCTTCATCTCGTAA	ATCTCCTCGTGTATTCCTTCCA
bf28	60	TTTCAATCTTCTCCTTTGATTGC	GGCAGCCATGATACAAGTGA
bf69	60	CTCTCACCAAACCCACTTCC	TTGAAGTTGGTGGAACAGCA
bg105	60	CCAATCTCCCCTTTTTCTCC	CATTGCTGTTGGAATTGCTG
bg208	60	ACACCTCGAACAAGATTCATCC	AGTAACCGCGAACCAAAGAGTA
bg234	60	GCTGGAATACACCAAGCATGA	GCAACATACCATCCCCTAAAAG
bg249	60	GGATACAAAATCCACAAGCACA	ACATAAGCGACTGGAACAAACC
bi111	60	GCCTTTAGTGGGATGAGTTCTG	TTTTGCTGAGGTGATGATATGG
bi116	60	CACACTTTCTCGTTTGCTCTCT	TCAACCCTTCAGATTTTCTTCC
bi89	60	TGGTTACTATTCCCACCATCATC	GTTTTGTCGTTGTGGAGTTTCA
mt1b06(2)	60	GTTTCCACGTGAAAGCCAGT	CATGGGCTGATACAACACACA
mt1d06	60	CCATGGCTCTTTCCTACCAA	GAAGGTTTTGGGTGGTGATG
mt1g05(2)	60	ATCATTACCGCAGCAAATCC	TTTTTGGAGTTTTGTTGGTGG

Supplementary Table 4.2. The segregation patterns of SSR markers in the F_1 and F_2 populations of diploid alfalfa and deviations from expectations. A_1 represents the allele from maternal parent; A_2 represents the allele from paternal parent.

				F_1 population			F_2 population							
				Genotypic	Allelic		Genotypic				Allelic			Zygotic
	LG on F_2 map	Segregation Pattern	Position on F_2 map		Maternal	Paternal	Genotype Frequency (%)				Allele Frequency (%)			
				P value	P value	P value	P value	A_1A_1	A_2A_2	A_1A_2	P value	A_1	A_2	P value
bi89	1	$A_1A_2xA_3A_4$	0.0	0.016	0.496	0.002	0.019	16.3	32.7	51.0	0.005	41.8	58.2	0.842
be942279	1	$A_1A_2xA_3A_3$	20.9	0.180	0.180	.	0.002	12.2	29.3	58.5	0.004	41.5	58.5	0.046
bg208	1	$A_1A_2xA_1A_1$	30.8	0.286	0.286	.	0.000	10.1	35.3	54.7	0.000	37.4	62.6	0.142
bg234	1	$A_1A_2xA_1A_3$	31.4	0.000	0.177	0.000	0.001	11.8	32.2	55.9	0.000	39.8	60.2	0.120
be103	1	$A_1A_2xA_3A_4$	47.0	0.000	0.065	0.000	0.001	12.1	24.8	63.1	0.028	43.6	56.4	0.003
be68	1	$A_1A_2xA_3A_3$	47.6	0.188	0.188	.	0.005	12.6	28.3	59.1	0.012	42.1	57.9	0.059
aw685684	1	.	56.9	.	.	.	0.000	15.9	17.9	66.2	0.730	49.0	51.0	0.000
bg647796	1	$A_1A_2xA_3A_3$	64.0	0.146	0.146	.	0.003	16.0	18.5	65.5	0.697	48.7	51.3	0.003
aw350	1	$A_1A_2xA_3A_4$	64.3	0.000	0.328	0.000	0.002	17.4	18.1	64.4	0.908	49.7	50.3	0.002
bg249	1	$A_1A_2xA_3A_4$	64.4	0.000	0.117	0.000	0.001	15.8	18.0	66.2	0.719	48.9	51.1	0.001
aw774443	1	$A_1A_2xA_3A_3$	64.5	0.149	0.149	.	0.002	16.7	18.8	64.6	0.724	49.0	51.0	0.002
mtic95	1	$A_1A_2xA_1A_2$	64.7	0.000	.	.	0.001	16.1	18.1	65.8	0.728	49.0	51.0	0.001

Supplementary Table 4.2. Continued.

				F_1 population			F_2 population							
				Genotypic	Allelic		Genotypic				Allelic			Zygotic
	LG on F_2 map	Segregation Pattern	Position on F_2 map		Maternal	Paternal	Genotype Frequency (%)				Allele Frequency (%)			
				P value	P value	P value	P value	A_1A_1	A_2A_2	A_1A_2	P value	A_1	A_2	P value
al375136	1	$A_1A_2xA_3A_3$	71.8	0.204	0.204	.	0.018	15.8	31.6	52.6	0.006	42.1	57.9	0.618
bf644149	2	$A_1A_1xA_2A_3$	0.0	0.000	.	0.000	0.000	6.3	11.3	82.4	0.406	47.5	52.5	0.000
mt1d06	2	$A_1A_2xA_3A_4$	0.5	0.881	0.704	0.820	0.000	6.2	12.3	81.5	0.292	46.9	53.1	0.000
mtic451	2	$A_1A_2xA_3A_4$	1.6	0.839	0.655	0.766	0.000	6.6	11.2	82.2	0.422	47.7	52.3	0.000
aw242	2	$A_1A_2xA_3A_4$	2.5	0.968	0.817	0.817	0.000	7.5	19.4	73.1	0.167	44.0	56.0	0.000
bg105	2	$A_1A_2xA_3A_3$	11.0	0.496	0.496	.	0.000	2.8	13.4	83.8	0.075	44.7	55.3	0.000
mtic210	2	$A_1A_1xA_2A_3$	13.4	0.511	.	0.511	0.000	38.6	4.7	56.7	0.000	66.9	33.1	0.007
aw310	2	$A_1A_2xA_3A_4$	35.0	0.011	0.136	0.017	0.012	15.9	22.8	61.4	0.240	46.6	53.4	0.019
mtic452	2	$A_1A_2xA_3A_4$	57.9	0.513	0.411	0.940	0.197	19.9	23.2	57.0	0.565	48.3	51.7	0.226
mtic12	2	$A_1A_2xA_1A_3$	67.3	0.612	0.243	0.586	0.570	21.2	26.7	52.1	0.349	47.3	52.7	0.867
aw586158	2	$A_1A_2xA_1A_2$	68.4	0.406	.	.	0.170	16.8	29.5	53.7	0.082	44.0	56.0	0.000
mtic19	2	$A_1A_2xA_3A_4$	69.1	0.277	0.151	0.705	0.655	21.7	25.9	52.4	0.478	47.9	52.1	0.832

Supplementary Table 4.2. Continued.

				F_1 population			F_2 population							
				Genotypic	Allelic		Genotypic				Allelic			Zygotic
	LG on F_2 map	Segregation Pattern	Position on F_2 map		Maternal	Paternal	Genotype Frequency (%)				Allele Frequency (%)			
				P value	P value	P value	P value	A_1A_1	A_2A_2	A_1A_2	P value	A_1	A_2	P value
mtic354	2	$A_1A_2xA_1A_3$	71.5	0.711	0.297	0.766	0.075	22.3	33.1	44.6	0.063	44.6	55.4	0.494
be129	.	$A_1A_2xA_1A_2$	.	0.000	.	.	0.000	0.0	0.0	100.0	1.000	50.0	50.0	0.000
b21e13	.	$A_1A_2xA_3A_4$	.	0.362	0.087	0.941	0.021	28.6	14.3	57.1	0.023	57.1	42.9	0.174
aw213	3	$A_1A_2xA_1A_3$	0.0	0.523	0.157	0.637	0.002	25.7	12.8	61.5	0.027	56.4	43.6	0.010
be85	3	$A_1A_1xA_2A_3$	1.8	0.657	.	0.657	0.000	21.2	12.3	66.4	0.128	54.5	45.5	0.000
be81	3	$A_1A_1xA_2A_3$	4.2	0.334	.	0.334	0.001	21.2	13.7	65.1	0.198	53.8	46.2	0.001
be114	3	$A_1A_2xA_3A_4$	4.7	0.528	0.299	0.374	0.000	19.2	13.7	67.1	0.349	52.7	47.3	0.000
aw693450	3	$A_1A_2xA_3A_4$	4.7	0.485	0.265	0.334	0.000	20.1	13.4	66.4	0.247	53.4	46.6	0.000
bg581665	3	$A_1A_1xA_1A_2$	5.5	0.352	.	0.352	0.000	19.7	14.5	65.8	0.359	52.6	47.4	0.000
mtic237	3	$A_1A_1xA_1A_2$	8.0	0.406	.	0.406	0.000	19.5	8.7	71.8	0.064	55.4	44.6	0.000
bf641851	3	$A_1A_2xA_3A_4$	26.5	0.417	0.647	0.222	0.093	23.2	18.3	58.5	0.406	52.5	47.5	0.123
aw220	.	$A_1A_2xA_1A_2$	.	0.207	.	.	0.018	21.3	16.9	61.8	0.467	52.2	47.8	0.021

Supplementary Table 4.2. Continued.

				F_1 population			F_2 population							
				Genotypic	Allelic		Genotypic				Allelic			Zygotic
	LG on F_2 map	Segregation Pattern	Position on F_2 map		Maternal	Paternal	Genotype Frequency (%)				Allele Frequency (%)			
				P value	P value	P value	P value	A_1A_1	A_2A_2	A_1A_2	P value	A_1	A_2	P value
afct32	.	$A_1A_1xA_2A_3$	.	0.084	.	0.084	.	.	.	.		.	.	
be119	.	$A_1A_2xA_1A_3$	.				0.000	85.3	13.3	1.4	0.000	86.0	14.0	0.000
bg448975	4a	$A_1A_1xA_2A_3$	**0.0**	0.137	.	0.137	0.076	19.3	21.4	59.3	0.725	49.0	51.0	0.080
al371804	4a	$A_1A_2xA_1A_2$	**4.1**	0.519	.	.	0.070	22.5	18.5	58.9	0.490	52.0	48.0	0.085
aa660573	4a	$A_1A_1xA_2A_3$	**15.3**	0.411	.	0.411	0.276	26.7	19.3	54.0	0.204	53.7	46.3	0.575
be84	4a	$A_1A_1xA_2A_3$	**17.9**	0.307	.	0.307	0.086	35.4	19.5	45.1	0.042	57.9	42.1	0.797
aw690665	4a	$A_1A_2xA_3A_4$	**18.6**	0.783	0.657	0.374	0.113	28.9	17.4	53.7	0.049	55.7	44.3	0.562
aw267840	4a	$A_1A_2xA_3A_4$	**19.0**	0.412	0.363	0.225	0.126	27.5	17.6	54.9	0.097	54.9	45.1	0.428
mtic94	4a	$A_1A_2xA_1A_2$	**21.6**	0.000	.	.	0.136	17.3	27.6	55.1	0.103	44.9	55.1	0.438
bf635198	.	$A_1A_2xA_1A_3$	**.**	0.327	1.000	0.103	0.001	11.5	32.4	56.1	0.000	39.5	60.5	0.109
mtic332	4b	$A_1A_2xA_3A_4$	0.0	0.194	0.156	0.411	0.000	13.1	37.7	49.2	0.000	37.7	62.3	0.860
mtic331	4b	$A_1A_2xA_3A_3$	2.6	0.231	0.231	.	0.000	10.1	36.2	53.6	0.000	37.0	63.0	0.208

Supplementary Table 4.2. Continued.

				F_1 population			F_2 population							
				Genotypic	Allelic		Genotypic				Allelic			Zygotic
	LG on F_2 map	Segregation Pattern	Position on F_2 map		Maternal	Paternal	Genotype Frequency (%)				Allele Frequency (%)			
				P value	P value	P value	P value	A_1A_1	A_2A_2	A_1A_2	P value	A_1	A_2	P value
aa05	5a	$A_1A_2xA_1A_3$	0.0	0.000	0.000	0.489	0.000	16.3	13.6	70.1	0.641	51.4	48.6	0.000
b14b03	5a	$A_1A_2xA_3A_4$	6.2	0.000	0.000	0.549	0.000	16.0	10.7	73.3	0.356	52.7	47.3	0.000
bf28	5b	$A_1A_2xA_3A_4$	0.0	0.236	0.042	0.755	0.000	8.4	14.5	77.1	0.438	47.0	53.0	0.674
mtic238	5b	$A_1A_2xA_3A_4$	16.2	0.376	0.086	0.709	0.507	23.3	22.0	54.7	0.817	50.7	49.3	0.519
mt1g05(2)	5b	$A_1A_2xA_3A_4$	30.5	0.392	0.153	0.408	0.000	11.9	23.8	64.3	0.044	44.1	55.9	0.001
aw295	5b	$A_1A_1xA_2A_3$	49.8	0.599	.	0.599	0.130	18.0	28.7	53.3	0.065	44.7	55.3	0.627
be100	.	$A_1A_1xA_2A_3$	.	0.881	.	0.881	.	.	.	.		.	.	
aw369	.	$A_1A_2xA_3A_4$	.	0.002	0.001	0.435	.	.	.	.		.	.	
be149	.	$A_1A_2xA_1A_2$	.	0.000	.	.	0.000	11.7	13.3	75.0	0.796	49.2	50.8	0.000
bf69	6a	$A_1A_2xA_3A_3$	0.0	0.820	0.820	.	0.000	0.7	14.0	85.3	0.029	43.4	56.6	0.000
aw01	6a	$A_1A_2xA_3A_4$	18.6	0.007	0.025	0.454	0.000	7.0	28.0	65.0	0.000	39.5	60.5	0.000
mtic345	6a	$A_1A_2xA_3A_3$	19.7	0.459	0.459	.	0.000	6.6	27.0	66.4	0.000	39.8	60.2	0.000

Supplementary Table 4.2. Continued.

				F_1 population			F_2 population							
				Genotypic	Allelic		Genotypic				Allelic			Zygotic
	LG on F_2 map	Segregation Pattern	Position on F_2 map		Maternal	Paternal	Genotype Frequency (%)				Allele Frequency (%)			
				P value	P value	P value	P value	A_1A_1	A_2A_2	A_1A_2	P value	A_1	A_2	P value
mtic14	6a	$A_1A_1xA_2A_3$	19.7	0.204	.	0.204	0.000	6.6	27.0	66.4	0.000	39.8	60.2	0.000
mtic93	.	.	.				0.000	0.0	0.0	100.0	1.000	50.0	50.0	0.000
bg584955	6b	.	0.0				0.776	22.7	26.7	50.7	0.488	48.0	52.0	0.983
mtic134	6b	$A_1A_2xA_3A_4$	7.1	0.105	0.233	0.551	0.409	24.7	20.7	54.7	0.488	52.0	48.0	0.508
bg648700	.	$A_1A_1xA_2A_3$	.	0.645	.	0.645	.	.	.	.		.	.	
aw254	7	$A_1A_1xA_1A_2$	0.0	0.000	.	0.000	0.000	32.4	2.8	64.8	0.000	64.8	35.2	0.000
afct45	7	$A_1A_2xA_3A_3$	18.4	0.820	0.820	.	0.000	8.0	4.4	87.6	0.546	51.8	48.2	0.000
aw212	7	$A_1A_2xA_3A_4$	19.8	0.229	0.705	0.041	0.000	5.8	3.6	90.6	0.719	51.1	48.9	0.000
aw343	7	$A_1A_1xA_2A_3$	21.2	0.031	.	0.031	0.000	9.8	3.7	86.6	0.435	53.0	47.0	0.000
be123	7	$A_1A_2xA_3A_3$	28.9	0.823	0.823	.	0.000	16.4	5.9	77.6	0.066	55.3	44.7	0.000
aw352	7	$A_1A_2xA_3A_3$	30.5	0.941	0.941	.	0.000	18.2	7.0	74.8	0.058	55.6	44.4	0.000
bg456767	7	$A_1A_1xA_2A_3$	31.2	0.045	.	0.045	0.000	19.9	7.3	72.8	0.029	56.3	43.7	0.000

Supplementary Table 4.2. Continued.

				F_1 population			F_2 population							
				Genotypic	Allelic		Genotypic				Allelic			Zygotic
	LG on F_2 map	Segregation Pattern	Position on F_2 map		Maternal	Paternal	Genotype Frequency (%)				Allele Frequency (%)			
				P value	P value	P value	P value	A_1A_1	A_2A_2	A_1A_2	P value	A_1	A_2	P value
bg454767	7	$A_1A_1xA_2A_3$	31.5	0.088	.	0.088	0.000	20.7	7.1	72.1	0.023	56.8	43.2	0.000
bi111	.	$A_1A_2xA_1A_1$	.	0.484	0.484	.	0.000	8.3	16.7	75.0	0.176	45.8	54.2	0.000
be74	.	.	.				0.000	30.1	1.4	68.5	0.000	64.4	35.6	0.000
aw685868	8a	$A_1A_2xA_1A_3$	0.0	0.000	0.000	0.653	0.000	73.2	0.7	26.2	0.000	86.2	13.8	0.454
mt1b06(2)	8a	$A_1A_1xA_2A_3$	4.9	0.707	.	0.707	0.000	68.6	0.0	31.4	0.000	84.3	15.7	0.088
al367466	8a	$A_1A_2xA_3A_4$	13.2	0.000	0.000	0.551	0.000	55.9	0.7	43.4	0.000	77.6	22.4	0.009
bi116	8a	$A_1A_2xA_3A_4$	24.2	0.000	0.000	0.446	0.000	46.6	1.5	51.9	0.000	72.5	27.5	0.003
be323955	8a	$A_1A_2xA_1A_3$	30.6	0.000	0.000	0.655	0.000	43.3	6.7	50.0	0.000	68.3	31.7	0.338
maa660870	8a	$A_1A_1xA_2A_3$	34.8	0.000	0.000	.	0.000	36.5	8.1	55.4	0.000	64.2	35.8	0.044
mtic188	.	$A_1A_2xA_3A_4$	.	0.000	0.000	0.764	.	.	.	.	.	.	.	.
aw255	8b	$A_1A_2xA_1A_3$	0.0	0.463	0.240	0.309	0.000	61.1	2.7	36.2	0.000	79.2	20.8	0.476
mtic103	8b	.	1.7				0.000	60.9	1.1	38.0	0.000	79.9	20.1	0.211

Supplementary Table 4.2. Continued.

				F_1 population			F_2 population							
				Genotypic	Allelic		Genotypic				Allelic			Zygotic
	LG on F_2 map	Segregation Pattern	Position on F_2 map		Maternal	Paternal	Genotype Frequency (%)				Allele Frequency (%)			
				P value	P value	P value	P value	A_1A_1	A_2A_2	A_1A_2	P value	A_1	A_2	P value
aw325	.	$A_1A_2xA_1A_3$	.	0.841	0.542	0.760	0.000	0.7	3.3	96.0	0.644	48.7	51.3	0.000
al372288	.	$A_1A_2xA_3A_4$		0.000	0.000	0.599	.	.	.	.	.	.	.	.

CHAPTER 5

CONCLUSIONS

Hybrids between *Medicago sativa* subsp. *sativa* (sativa) and *M. sativa* subsp. *falcata* (falcata) often express heterosis for yield. We investigated the genetic basis of heterosis and inbreeding depression from several perspectives including classic quantitative genetics, gene expression profiling, and genetic mapping of segregation distortion (SD).

First, we assessed the fertility and biomass in advanced generations of inter- and intra-subspecies hybrids of tetraploid alfalfa. We found that the fertility of sativa×falcata crosses (SFC) did not break down for either self- or sib-pollination, and on the contrary, SFC showed greater fertility than either sativa×sativa crosses (SSC) or falcata×falcata crosses (FFC). However, SFC did show higher yield depression compared to SSC and FFC. F-tests were constructed between full and reduced models to test for the presence of multiple allelic interactions and/or epistatic effects. The results indicated that the two-allele interactions (dominance) plays a more important role in biomass yield compared to multiple allelic interactions and epistatic interaction.

Second, to investigate the genetic mechanism of heterosis at the gene expression level, we compared gene expression profiles of three hybrids, two of them which showed heterosis for yield and a third which did not. We found that the heterotic hybrid families had a higher number and a higher proportion of genes exhibiting nonadditive expression and expression outside the

parental range than did the non-heterotic family. This indicated that nonadditive expression and expression higher or lower than either parent might contribute to heterosis for biomass.

Third, we found a large percentage of markers (almost the whole genome) showed significant segregation distortion (SD) in a diploid *M. sativa* subsp. *caerulea* x *falcata* F_2 population. Most distorted markers in the F_2 population showed genotypic SD and zygotic SD, but not allelic SD, which suggested that zygotic selection was more frequent than gametic selection. Most of distorted markers were clustered and exhibited heterozygote excess. Either overdominance or pseudo-overdominance could explain these results at multiple segregation distortion loci (SDLs).

Collectively, our results suggest that sativa and facata contain favorable alleles or linkats (co-adapted gene complexes), and the complementary action of them (dominance or pseudo-overdominance) in inter-subspecies cross hybrids causes heterosis on yield. The outcrossing nature of alfalfa causes it to carry a high genetic load of deleterious recessive alleles for fitness (i.e., biomass yield); the exposure of deleterious recessive alleles in homozygotes and the loss of the complementary gene actions in inbred alfalfa populations cause the serious inbreeding depression. The same reason explains the severe SD in F_2 population. Although SD with heterozygote excess was common, it is more likely due to the pseudo-overdominance of segregation distortion loci (SDLs).

Although higher proportions of the nonadditive expression and expression levels higher or lower than either parent were found in heterotic hybrids compared to a non-heterotic hybrid in our study, the majority of genes showed additive expression in all hybrid families. However, the extent of nonadditive expression at different development stages and different tissues may vary and across the life cycle of the plant, the expression patterns cumulatively produce the observed

heterotic response. Therefore, for integrative phenotypes like yield, the cumulative effect of different genes acting at different places and times could result in heterosis.

Quantitative trait locus (QTL) mapping could help investigate the genetic basis of yield heterosis. Precise QTL mapping depends on precise genetic maps, which could be biased by segregation distortion (SD). Compared to SD in an F_2 population, a smaller percentage of genotypic and allelic SD were found in an F_1 mapping population. By comparing the constructed maps of the F_1 and F_2 populations derived from same pair parents, we inferred that SD mostly affected the estimation of genetic distance among markers, had a little effect on ordering, depending on the distorted level, and did not affect the grouping of markers.

Alfalfa breeders generally make selection after crossing germplasms including *M. sativa* subsp. *sativa* and *M. sativa* subsp. *falcata*. If the SDL and QTL alleles for agronomic traits are linked in repulsion, then the selection or introgression on agronomic traits could be affected by SDL. Therefore, it is meaningful to map SDL and QTL for agronomic traits with same population, which can facilitate their selection and introgression. In addition, advanced generations could be obrained by random mating, and the linkages between markers (and/or QTLs) and SDLs could be breaked. The advanced generations with less SD may facilitate map construction and QTL mapping.