RELATIONSHIPS AND AUTOPOLYPLOID EVOLUTION IN THE *MEDICAGO SATIVA* COMPLEX (ALFALFA AND ALLIES; LEGUMINOSAE)

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Tee Havananda

January 2013

RELATIONSHIPS AND AUTOPOLYPLOID EVOLUTION IN THE *MEDICAGO SATIVA* COMPLEX (ALFALFA AND ALLIES; LEGUMINOSAE)

Tee Havananda, Ph. D.

Cornell University 2013

The *Medicago sativa* complex comprises several morphologically and genetically diverse diploid and autopolyploid taxa, including autotetraploid cultivated alfalfa (*M. sativa* subsp. *sativa*). Its members can be divided into three morphological groups: *M. sativa* subsp. *caerulea* and subsp. *sativa* with blue flowers and coiled pods, subsp. *falcata* with yellow flowers with falcate pods, and subsp. *glomerata* with yellow flowers and coiled pods; each group contains both diploid and tetraploid cytotypes. Although alfalfa is well studied, closely related tetraploids and their hypothesized diploid progenitors have received much less study. Questions regarding their relationships, their controversial taxonomy, and autopolyploid evolution remain to be addressed. Genetic variation and differentiation were estimated, and phylogenetic and network relationships were constructed based on nucleotide sequences from the mitochondrial genome for the diploid members of the complex, and from chloroplast and nuclear genomes for both the diploid and tetraploid members. Independent perspectives on the species' evolutionary history were afforded because each genome has a different inheritance pattern. Mitochondrial DNA is maternally inherited, chloroplast DNA is biparentally, but largely paternally inherited, and nuclear DNA is biparentally inherited. At the diploid level, subsp. *caerulea* is genetically differentiated from diploid subsp. *falcata* for chloroplast haplotypes and nuclear alleles, although there are some shared haplotypes and alleles probably due to limited gene flow. Data from

mitochondrial haplotypes, however, show no differentiation between the two diploids, which is likely due to bidirectional introgression of the mitochondrial genome. At the tetraploid level, genetic differentiation was found between subsp. *sativa* and tetraploid subsp. *falcata* in both the chloroplast and nuclear genomes. Although chloroplast data support a simple autopolyploid origin of subsp. *sativa* from diploid subsp. *caerulea*, a contrasting history involving past introgression from closely related *M. prostrata* is suggested for tetraploid subsp. *falcata*, raising questions about its autopolyploid origin. Nuclear data, however, show that tetraploid *falcata* most likely has originated from diploid *falcata* through autopolyploidy in a similar pattern to that of subsp. *sativa*. Despite the existence of hybrids, gene flow and introgression are limited and morphologically and genetically distinctive subspecies persist.

BIOGRAPHICAL SKETCH

"We may have nothing as an inheritance for you, but we can leave you an education," a mother told her fourth child who was not really enjoying doing his homework. The young boy remembers that simple sentence and witnesses how much his parents value education providing every support for their five children so they would have an education in any fields at the highest degree they want. That fourth child is now about to graduate with a Ph.D., the highest academic degree one can accomplish.

Mr. Tee Havananda (pronounced hə-wɑː-noʊn) is the fourth child of Dr. Sriwongse, a pediatrician and surgeon, and Mrs. Chaweewan Havananda, a registered nurse. He was born in 1971 in Bangkok, Thailand. He attended grade school at Kasetsart University Laboratory School (Satit Kaset) from grade 1. Realizing the importance of extramural activities, his parents provided him with experiences of various kinds, e.g. sports, music, and camps. He went on several summer camps, including a recruiting camp for the Children's International Summer Village (CISV) program in 1982. He was recruited as one of the four delegates from Thailand to join a 4-week long CISV camp in Hong Kong in 1983, his first time abroad. After graduating high school from Satit Kaset and passing the National University Entrance Examination, he was accepted to an undergraduate program in agriculture by the Faculty of Agriculture at Kasetsart University. He chose to major in horticulture and worked on a morphological study of bird's nest fern for his Special Problem course during his senior year, under the supervision of M.L. Charuphant Thongtham, an Associate Professor in the Department of Horticulture at the university.

He received a degree of Bachelor of Science (Agriculture) in Horticulture in 1993 and then pursued further education for a master degree at the same institution. Admiring M.L.

Thongtham's guidance, Mr. Havananda continued to have him as his advisor. He briefly worked in M.L. Thongtham's breeding program for canning pineapple before deciding to return to work with his favorite plants, ferns. Introduced by his advisor, he learned about tree ferns and did his Master degree thesis on tree ferns in Thailand. His research project was funded by Queen Sirikit Botanic Garden (QSBG), The Botanical Garden Organization, Thailand. Working on his research on tree ferns for about 3 years brought him the opportunity to travel around the country with M.L. Thongtham, Dr. Piyakaset Suksathan, who had been his school mate since grade 1, and new friends, whose love of various kinds of plants is tremendous. The experience he gained during this period on several trips with these people was invaluable and unforgettable.

After graduating his Master degree in Agriculture (Horticulture) from Kasetsart University in 1998, Mr. Havananda worked as a scientist at the QSBG in Mae Rim, Chiang Mai. Later that year, the QSBG was offered two Training Fellowships from the Muséum National d'Histoire Naturelle, Paris, France, and he was fortunately granted one of them. During the three-month training at the world-renowned museum, he gained hands-on experience in seed collection and exchange, and also on herbarium specimen cataloging. He spent some time in this precious three months visiting many divisions in this huge natural history museum, botanical gardens in other parts of France, and a few other surrounding countries. This was undoubtedly another memorable and cherished period in his life.

In mid 1999, he was informed of an opening lecturer position in the Department of Horticulture, Faculty of Agriculture, Kasetsart University, his alma mater, and was convinced he should be a candidate. After hard consideration, he applied and went for an interview for the position. He was chosen over several other candidates by the recruitment committee to take the position that would be located in Kamphaeng Saen campus, Nakhon Pathom, about 80

kilometers northwest of Bangkok. With his secure job and good relationships with colleagues, it was a very difficult decision for him to resign from the QSBG to take a new job. Tee started working in the lecturer position at the end of September 1999.

In the early stage of his academic career, his responsibilities were mainly teaching and overseeing undergraduate student activities. He was a manager, co-manager, and co-lecturer in various courses such as Floriculture, Plant Materials, and Systematics of Horticultural Crops. He was also an advisor and co-advisor overseeing several research projects of mostly undergraduate students. Because of the importance of further education for this career, he was looking for options and preparing himself for Ph.D. program applications during these first few years. Being intimidated by the reputation in academic excellence of the institutions in the Ivy League such as Cornell University, he initially looked at other places with similarly interesting programs. In looking for advice from people who had gone through the process of choosing, applying for, and attending a Ph.D. program abroad, he talked to Dr. Julapark Chunwongse, a Cornell alumnus, who advised him to "aim high and do not underestimate yourself." Because Tee was interested in the field of molecular systematics, Dr. Julapark also suggested that he should learn about Dr. Jeff J. Doyle's research at Cornell. Tee took the advice and put Cornell at the top of the list of schools to which he would apply for a Ph.D. program.

In 2002, the Center for Agricultural Biotechnology (CAB) at Kasetsart University was offering various supports under a staff development program. Tee was financially supported by the center affording him a three-month visit to Dr. Doyle's lab at Cornell in that summer. During the three months, he gained hands-on experience in molecular laboratory techniques working under Jane L. Doyle's supervision and improved his English skills that were vital for graduate school admission in the US. The visit to the Doyle lab also provided him the opportunity to be

familiar with the surroundings of the institution and town, which would later turn out to be where he attended his Ph.D. program. After that valuable trip to the US, he got himself qualified for a Ph.D. program admission with only the financial requirement left to be met. Compared with US students, it was understood that it is much more competitive for international students to receive financial support from a US source. CAB recognized the financial obstacle and kindly offered Tee partial financial support, pending his acceptance to a Ph.D. program. Meeting other requirements and the almost certainty of financial support led the Field of Plant Breeding at Cornell to admit Tee to its Ph.D. program under Dr. Doyle's supervision. The admission to Cornell allowed Tee to qualify for full financial support from the Office of the Higher Education Commission, Thailand, which became available shortly before his departure from Thailand. Tee was allowed only one source of funding, so he decided to return CAB's funding with gratitude. He received the funding from the Office of the Higher Education Commission with a condition of returning to work for Kasetsart University after graduation. He started his Ph.D. program in Fall 2003, majoring in Plant Breeding with Plant Genetics as an area of concentration. He chose to minor in Plant Biology with a concentration on Systematic Botany.

His second stay at Cornell University made him grow significantly as a student, a researcher, and a person. Academically, he was taught and trained in subjects and techniques virtually new to him in a wonderful academic environment. His research, on a topic and plants on which he had never before worked, has brought two publications so far, an accomplishment he did not know he was capable of. He went through highs and lows that will only be beneficial to his career. Personally, he learnt how to live a life away from home at a place on the opposite side of the world. Living in a diverse community, he experienced different cultures, perspectives, and life styles. His married life literally started here. He married his wife, Sutani (Yuktanan)

Havananda, in Thailand less than a month before he had to leave her for Cornell. It took them five months to get things ready for her to follow him to the US and to really live together. Besides taking marriage lessons while doing a Ph.D., he also started learning parenthood lessons when his wife gave birth to a baby girl, Teerada (Teenie), in 2007. His time as a Ph.D. student has brought him much more than just an academic degree and is undoubtedly a momentous stage of his life.

Tee has now returned to work as a lecturer in the Department of Horticulture, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Kamphaeng Saen campus, Nakhon Pathom, Thailand.

I would like to dedicate this dissertation to my parents, my four siblings, and last but not least, my wife and our daughter.

ACKNOWLEDGEMENTS

This dissertation would not have been existed without the help and support of many people. They contributed one way or another throughout the research and writing processes. I would like to express my sincere gratitude to them.

First of all, I cannot be grateful enough to my advisor and Special Committee chair, Dr. Jeffrey J. Doyle, for his guidance, advice, and support from our very first communication before we even met. His intelligence, expertise, and experience were fundamental in the shaping of this research and in helping me overcome the hurdles along the way to its completion. I appreciate the patience and extra hours he applied in dealing with my language barriers, particularly in the writing of this dissertation. I am indebted to his understanding and encouragement that have never failed to boost my morale whenever I feel low. “Student first” attitude and willingness to educate are only a few of his qualities exemplifying how great a teacher he is. It is a great pleasure to learn these qualities from him for my teaching career.

I would like to express my great appreciation to other members of my Special Committee, Dr. Melissa Luckow and Dr. Donald R. Viands, as well as Dr. Stephen Kresovich, a former member. Their kind advice and constructive comments on my study plan and dissertation were very significant to the completion of my Ph.D. Special acknowledgment goes to Dr. Luckow for giving me the opportunity to expand my teaching experience as an assistant in her Taxonomy of Cultivated Plants class. Such experience will be very valuable for a similar class I will be responsible for when I return to work in Thailand and I am grateful for that.

I am very much indebted to Jane L. Doyle for all her assistance from my first visit and throughout my time as a student in the Doyle lab. I have benefited by her step-by-step instruction and advice on laboratory techniques. Her direct contribution to this research is also gratefully

acknowledged. I thank her for enlightening conversations we had over the years on topics as various as politics, religion, cooking, and gardening. I will miss working on a lab bench across from hers.

This research would have been much more difficult without contributions from Dr. E. Charles Brummer and Dr. Muhammet Şakiroğlu. I gratefully thank them for everything from plant and DNA materials, voucher information, to comments and suggestions. I owe sincere thankfulness to Dr. Ivan Maureira Butler for his guidance and assistance in the initial stage of this research during his time as a postdoc in the Doyle lab. His friendship toward me is also very appreciated.

It is a great pleasure to acknowledge other members of the Doyle lab, Dr. Sue Sherman-Broyles, Dr. Bernard Pfeil, Dr. Ashley Egan, Dr. Shannon Straub, Dr. Dan Ilut, Dr. Jeremy Coate, and Fronny Plume, for their various help, support, encouragement, and friendship. I had such a pleasant time being associated, professionally and personally, with them and hope for any future collaboration we may have. Many thanks also go to other past and present Doyle lab's members and many students of the L.H. Baily Hortorium and of the Field of Plant Breeding. Having them as friends made my time at Cornell more valuable.

I would like to express my appreciation to Dr. Elizabeth Earle for her kind help while she was Director of Graduate Study in the Field of Plant Breeding. I thank the administration staffs of the Department of Plant Breeding and Genetics, the Department of Plant Biology, and the Graduate School, Student Assistance at the International Students and Scholars Office, and staffs at both Plant Science and Guterman greenhouse facilities. Without their assistance my study and research at Cornell would not have been as smooth. My special thanks go to Mann Library and its staff for providing massive academic resources, great services, and a wonderful place for self-

study. I convey particular acknowledgment to Jeff Diver, a former Coordinator at the Stone Computing Center, Mann Library. I gained great experience working under his supervision as a Student Operator and Student Supervisor at the Stone Center from 2007 to 2009.

As an international student from Thailand, I am grateful to Cornell Thai Association and the Thai community in Ithaca. I thank fellow past and present Thai students, Dr. Adirek Rugkong, Dr. Rujira and Prasit Deewatthanawong, Dr. Wannasarn Noonsuk (and his wife, Nareerat), Dapong Boon-long, to name a few, for their friendships and support. I express my great appreciation to Apikanya and Matt McCarty, Chanita Somton, Sanit and Nujaree Chutintaranond, and Yuttanun and Amy Nakaranuruck. Their goodwill and generosity toward my family and me have been constant since we first met in Ithaca. Adirek, Apikanya, and Matt deserve special mention for their hospitality when accommodations were needed in the first and the last period of my time in Ithaca. I feel very fortunate to have known and become good friends with many of the Thai people we met here, and I cherish the good memories we shared.

My first visit to Cornell as a visiting fellow opened up the opportunity for me to become a student at the prestigious institution. Thus, I am obliged by the financial support for that visit from the Center of Agricultural Biotechnology (CAB), Kasetsart University, and the moral support and help from Dr. Pongthep Akrathanakul and Dr. Julapark Chunwongse, the director and the vice director of the center, respectively. I am greatly indebted to The Office of Higher Education Commission, the Royal Thai Government, who financially supported my Ph.D. study. I thank the Office of Educational Affairs, Royal Thai Embassy, Washington, D.C., and its staff for their kind facilitation on funding and several other matters. I would like to express my gratitude to the Department of Horticulture, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Kamphaeng Saen campus, of which I am in the service, for allowing me to

take the time I needed for my Ph.D. study, even though it was longer than anticipated. The former and present heads of the department, colleagues, and administrative staff deserve particular acknowledgment for their assistance on business matters at work during my absence. Especially for Dr. Parson Saradhuldhat, a closest friend and a colleague, I am grateful for his great help on the more personal matters needed to be done in Thailand while I was at Cornell.

I would not be what I am today without my family. The unwavering love and faith my parents have for me are immeasurable. They have instilled me with the values of education and hard work. I thank them for everything. I thank my siblings, San, Nat, Sasiwan, and Karn, and their spouses, Wandee, Euearee, Sompop, and Vatinee, respectively, for being supportive and encouraging. In particular, I am very grateful to Nat, Euearee, and their son, Punn, for their hospitality at their home in Texas in the last several weeks of our stay in the US, when I was preparing for the Dissertation Defense and short-term housing in Ithaca was not affordable. I also extend my gratitude to my mother-in-law, Boonsuk, and sisters-in-law, Waraporn and Patra Yuktanan, for warmly letting me be a part of their family and for letting her daughter/their little sister be with me on the opposite side of the world.

Finally, and most importantly, I wish to thank my wife, Sutani (Ni). Ni, words do not suffice to express my appreciation for your sacrifice and patience throughout these years of my Ph.D. pursuit. Your love, understanding, and companionship were the foundation of my life far away from home that carried me through the good times and bad. I am forever grateful to you for providing me a life-changing status, fatherhood, when you gave birth to our girl, Teerada (Teenie). Experiencing the challenge of parenthood together with you has been nothing short of wonderful. I am also thankful for you, Teenie. Thank you for being everything you are and giving me every day joy and wonder; they restored the energy I needed for my study. Living life

with both of you has made me grow as a person more than I can recognize from any other periods of my life.

This research was financially supported by various grants awarded to my advisor by the US National Science Foundation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

RELATIONSHIPS AMONG DIPLOID MEMBERS OF THE *MEDICAGO SATIVA* (FABACEAE) SPECIES COMPLEX BASED ON CHLOROPLAST AND MITOCHONDRIAL DNA SEQUENCES[*]

Abstract

The *Medicago sativa* species complex includes tetraploid cultivated alfalfa and several other diploid and tetraploid taxa that are recognized either as subspecies of *M. sativa* or as separate species. The two principal diploid taxa are "caerulea," with purple flowers and coiled pods, and "falcata" with yellow flowers and falcate pods. To understand the evolutionary relationships among taxa in the complex, sequence variation in two noncoding regions of cpDNA (*rpl20-rps12* and *trnS-trnG* spacers) and three regions of mitochondrial DNA (mtDNA: *nad4* intron, *nad*7 intron, and *rpS14-cob* spacer) were surveyed from 48 (37 for mtDNA) individuals representing these and other diploid taxa in the complex. These sequences afforded independent perspectives on the evolutionary history of the group, because mtDNA is maternally inherited in *Medicago* whereas cpDNA is biparentally inherited with strong paternal bias. Twenty and 21 haplotypes were identified for cpDNA and mtDNA, respectively. Haplotype networks were constructed and tests of differentiation were conducted. Results from cpDNA sequences supported the recognition of "caerulea" and "falcata" as differentiated taxa, despite the presence of some shared haplotypes, in agreement with morphological characters. In contrast, no significant evidence of mtDNA haplotype differentiation was observed. Incongruence between cpDNA and mtDNA is more likely explained by introgression of the mitochondrial

[*] This chapter was published in Havananda, T., E. C. Brummer, I. J. Maureira-Butler, and J. J. Doyle. 2010. Relationships among diploid members of the *Medicago sativa* (Fabaceae) species complex based on chloroplast and mitochondrial DNA sequences. *Systematic Botany* 35: 140–150. Written authorization for the use of the material in this dissertation was obtained from *Systematic Botany*.

genome than by incomplete lineage sorting of mtDNA haplotypes, given the expected smaller effective population size for uniparentally inherited mtDNA than for biparentally inherited cpDNA. Moreover, the two taxa are readily crossable, making natural hybridization possible. The long-time disagreement on whether to recognize "falcata" as a separate species or a subspecies of *M. sativa* s. l. is due to the common problem of unequal rates of differentiation for different characters during speciation.

Introduction

Polyploidy is one of the most important forces driving plant evolution and diversity; it can give rise to genomic rearrangements, including exchanges between genomes and gene loss, and changes in gene expression (e.g., Wang et al., 2006; Gaeta et al., 2007). Many recent studies have focused on various aspects of polyploidy (Soltis et al., 2004b) to understand the process of polyploidization (e.g., reviewed in Wendel, 2000) and its consequences for plant evolution (e.g., Soltis and Soltis, 2000; Adams and Wendel, 2005; Comai, 2005; Adams, 2007). However, more studies have been done in allopolyploids than in autopolyploids (e.g., Doyle et al., 2004; Soltis et al., 2004a; Dubcovsky and Dvorak, 2007). Genetic studies have revealed that autopolyploidy is much more common than generally recognized, yet autopolyploidy has often been overlooked as an important part of evolution in natural populations (Soltis et al., 2004b; Soltis et al., 2007). The *Medicago sativa* L. polyploid complex contains diploid and autotetraploid cytotypes and has potential to be a good model system for evolutionary study of autopolyploidy. Because alfalfa (*M. sativa*), an autotetraploid (Stanford, 1951; Quiros, 1982) member of the complex, is one of the world's most important forage crops, a considerable amount of genetic information and resources such as germplasm collections exist and are available for the study of this complex. In

addition, the *M. sativa* complex is congeneric with the legume genomic model species, *M. truncatula* Gaertn. Except for alfalfa, little is known about other members of this complex, particularly the diploids. To use the complex to study the evolution of autopolyploidy, it is important to understand patterns of genetic variation and evolutionary relationships among its members. Initially, we focus on diploid taxa that are hypothesized to have given rise to tetraploid members in the complex.

The *M. sativa* complex is composed of perennial, outcrossing, and morphologically differentiated but often interfertile taxa recognized at various taxonomic ranks by different authors. Lesins and Lesins (1979) recognized most members as species, whereas Quiros and Bauchan (1988) and Small and Jomphe (1989) recognized many of the same taxa as subspecies of *M. sativa*. These taxa are found in diploid ($2n = 2x = 16$) and/or tetraploid ($2n = 4x = 32$) forms with weak hybridization barriers both between taxa at the same ploidy level and between taxa from different ploidy levels (the latter when unreduced $2n$ gametes are produced from a diploid parent) (McCoy, 1982; McCoy and Bingham, 1988; Quiros and Bauchan, 1988). Based on morphological characters, the principal diploid members include *M. sativa* L. subspecies *caerulea* (Less. ex Ledeb.) Schmalh., *M. sativa* L. subsp. *falcata* (L.) Arcang. (= *M. falcata* L.), *M. sativa* L. subsp. *hemicycla* (Grossh.) C. R. Gunn (a putative natural diploid hybrid between the former two taxa; Quiros and Bauchan, 1988), and *M. sativa* L. subsp. *glomerata* (Balb.) Rouy (= *M. glomerata* Balb.) (Small and Jomphe, 1989) (Figure 1.1). Tetraploid taxa in the complex are hypothesized either to be derived directly from diploids or by hybridization between tetraploid taxa of different origins. For example, tetraploid *M. sativa* L. subsp. *sativa* is considered to be derived directly from diploid subsp. *caerulea*, and hybridization between subsp.

sativa and tetraploid subsp. *falcata* may have given rise to *M. sativa* L. subsp. *varia* (Martyn) Arcang. (Quiros and Bauchan, 1988; Small and Jomphe, 1989) (Figure 1.1).

An additional species, *M. prostrata* Jacq., is placed in subsection *Falcatae* of section *Falcago* along with the members of the *M. sativa* complex by Lesins and Lesins (1979). This species exists at both the diploid and tetraploid levels. It can be crossed to the members of the complex at both ploidy levels but with a hybridization barrier when it serves as maternal parent (Lesins, 1962, 1968) and natural hybrids of this species and *M. s.* subsp. *falcata* have been reported by Lesins and Lesins (1979).

Key morphological characters distinguishing typical plants of each member of the complex are flower color, pod shape, and pod pubescence (Table 1.1). However, each taxon shows considerable morphological variability, presumably due to the observed ability of the members to interbreed freely with each other (McCoy and Bingham, 1988). This has led to conflicting taxonomic treatments and the proliferation of names (Lesins and Lesins, 1979). The most conspicuous morphological differences in two key diploid members of the complex, *M. s.* subsp. *caerulea* ("caerulea") and subsp. *falcata* ("falcata"), are purple vs. yellow flowers and coiled vs. sickle-shaped pods, respectively. The ancestral states of these two characters were thought by Lesins and Lesins (1979) to be yellow flowers and coiled pods, the states of *M. s.* subsp. *glomerata*, which they considered to be the ancestor of "caerulea" and "falcata." These two diploid taxa also have cytogenetic differences as shown in their C-banding karyotypes (Bauchan and Hossain, 1997).

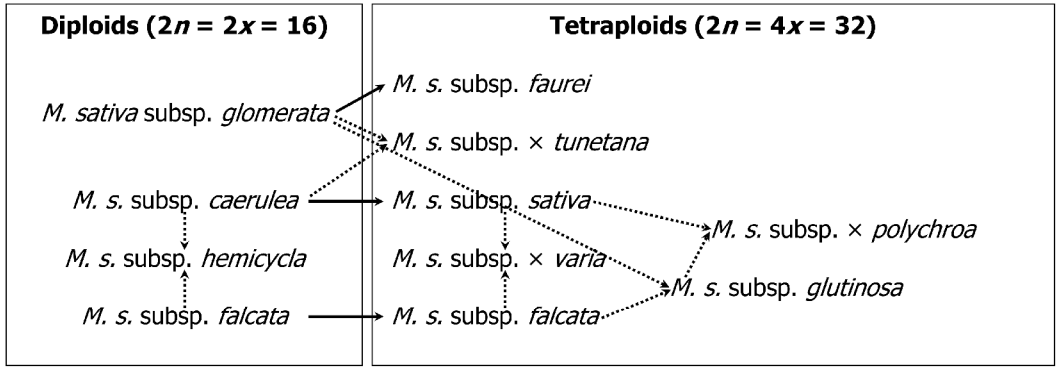


FIGURE 1.1. Hypothesized relationships among taxa in the *Medicago sativa* complex, modified from Quiros and Bauchan (1988). Arrows with solid lines denote autopolyploidy, and those with dashed lines denote hybridization.

TABLE 1.1. Principal characters used in distinguishing diploid members of the *Medicago sativa* complex (Lesins and Lesins, 1979; Quiros and Bauchan, 1988; Small and Jomphe, 1989)

Taxon	Ploidy level	Flower color	Pod shape	Pod pubescence
M. s. subsp. *caerulea*	Diploid	Violet	Coiled	Simple (glabrous)
M. s. subsp. *falcata*	Di-, tetraploid	Yellow	Straight or sickle	Simple or glandular
M. s. subsp. *glomerata*	Di-, tetraploid	Yellow	Coiled	Glandular
M. s. subsp. *hemicycla*	Diploid	Variegated	Curved or loosely coiled	Simple
M. prostrata	Di-, tetraploid	Yellow	Coiled	Glandular (simple, glabrous)

Taxonomic and phylogenetic studies of *Medicago* have been based mainly on morphology, with some use of cytological and biochemical data. More recently, molecular data have been employed and several molecular phylogenetic studies on *Medicago* have been conducted at the generic and specific levels based on nuclear ribosomal gene internal transcribed spacers (nrDNA ITS) and nrDNA external transcribed spacers (ETS) (Bena et al., 1998a–c; Downie et al., 1998; Bena, 2001), chloroplast DNA (cpDNA) (Valizadeh et al., 1996; Steele and Wojciechowski, 2003), and combined nuclear (nDNA) and mitochondrial DNA (mtDNA) (Maureira, 2004; Maureira-Butler et al., 2008). However, relationships among taxa in the *M. sativa* complex were not a main focus of these studies.

Within the complex, genetic variation within and among subspecies and populations was evaluated previously using several molecular markers: Brummer et al. (1991) and Kidwell et al. (1994) with nuclear restriction fragment length polymorphism (RFLP); Diwan et al. (1997) and Falahati-Anbaran et al. (2007) with simple sequence repeats (SSR); Ghérardi et al. (1998) with randomly amplified polymorphic DNA (RAPD); Segovia-Lerma et al. (2003) with amplified fragment length polymorphism (AFLP); Skinner (2000) with fragment length polymorphism in cpDNA hypervariable regions; Muller et al. (2001, 2003) with mtDNA RFLP; and Muller et al. (2006) with nDNA sequences. These studies gave similar results showing considerable genetic variation within and among the accessions included in each study, the majority of which were cultivated accessions of "falcata" and subsp. *sativa*. Although very little research addressing the relationships among wild populations has been carried out, various nuclear markers have shown that "falcata" is clearly separated from subsp. *sativa* in cultivated germplasms (reviewed in Brummer, 2004). A few studies included plant materials from "caerulea" and "falcata" and addressed relationships among subspecies. Brummer et al. (1991) found that accessions of

"caerulea" and subsp. *sativa* were grouped separately from those of "falcata" based on nuclear RFLP. Based on mtDNA RFLP, Muller et al. (2003) suggested that genetic differentiation has developed between "falcata" (undetermined ploidy level) and "caerulea" or subsp. *sativa* based on a unique mitochondrial haplotype (mitotype) found only in "falcata." However, they also discovered shared mitotypes between "falcata" and "caerulea" and/or subsp. *sativa* that could be explained by recent common ancestry and gene flow between them. Later, Muller et al. (2006) surveyed DNA sequence diversity at two nuclear loci on a subset of samples used in Muller et al. (2003) and found no differentiation in the *M. sativa* complex. These results leave in question the genetic boundaries of the wild members of the complex.

Because organellar DNA is generally homoplasmic in plants, it is a simpler system than nuclear DNA for a systematic study of an outcrossing, highly heterozygous taxon such as *M. sativa*. In *Medicago*, even though cpDNA can be biparentally transmitted (Smith et al., 1986), inheritance is largely paternal (Schumann and Hancock, 1989; Smith, 1989; Masoud et al., 1990) and sorting out occurs rapidly (Johnson and Palmer, 1989); thus, most individuals are homoplasmic. Mitochondrial DNA in alfalfa is maternally inherited (Schumann and Hancock, 1989; Forsthoefel et al., 1992). With two different inheritance patterns, these two genomes present independent perspectives on the evolutionary history of the group. DNA from both genomes has been shown to provide a source of variation for inferring evolutionary relationships in the complex (Skinner, 2000; Maureira, 2004; Maureira-Butler et al., 2008; Muller, 2001, 2003).

To characterize the *M. sativa* complex, we focus initially on understanding the evolutionary relationships among diploid members. We reconstruct relationships among chloroplast and mitochondrial haplotypes of diploid taxa in the complex, focusing on the two

main morphologically distinct taxa, “caerulea” and “falcata” that are putative progenitors of tetraploids subsp. *sativa* and subsp. *falcata*, respectively. The contentious taxonomic status of “falcata” as either a separate species or a subspecies of *M. sativa* sensu lato (e.g., Lesins and Lesins, 1979; Quiros and Bauchan, 1988; Small and Jomphe, 1989) is addressed and the different perspectives on the evolutionary history of the group from the two organellar genomes are discussed.

Materials and Methods

Plant Material—Forty-eight individuals of diploid *Medicago* were included in this study. They comprised 40 wild accessions of the *M. sativa* complex chosen from the USDA *M. sativa* germplasm collection to represent the taxonomic, geographic, and morphological diversity of the complex, two genotypes from a synthetic population SD201 developed from wild and cultivated diploid “falcata” genotypes (Boe et al., 1998), one genotype from the artificially diploidized *M. s.* subsp. *sativa* germplasm CADL (“cultivated alfalfa at the diploid level”; Bingham and McCoy, 1979), four *M. prostrata* diploids, and one *M. truncatula* plant as an outgroup (Appendix 1). All plants were grown in greenhouses at Iowa State Univ. and/or the Univ. of Georgia between 2003 and 2008, except for a single diploid accession of *M. s.* subsp. *glomerata* which was grown at Cornell University. Ploidy of all genotypes was confirmed by flow cytometry using methods described previously (Brummer et al., 1999). Due to low mtDNA variation found in preliminary data, only 37 individuals were used in the mtDNA dataset (Appendix 1).

PCR Amplification and DNA Sequencing—Total genomic DNA was extracted at Iowa State University or the University of Georgia using a CTAB protocol (Doyle and Doyle, 1987).

Extracted DNA was used as template for amplification of two cpDNA regions (*rpl20-rps12* spacer and *trnS-trnG* spacer) and four mtDNA regions (*nad1* intron, *nad4* intron, *nad7* intron, and *rpS14-cob* spacer). Primers used for PCR and sequencing in this study, including internal primers for the *nad4* and *nad7* introns, are shown in Table 1.2. For each of the two chloroplast regions and mitochondrial *rpS14-cob* spacer, we were able to obtain a single sequence for each individual using the PCR amplification primers. Due to the length of the mitochondrial *nad4* and *nad7* introns (>2 kb), we had difficulties in obtaining complete sequences using only the two amplification primers, hence additional internal primers of these two regions were necessary (Figure 1.2). For both *nad4* and *nad7* introns, amplification of shorter segments from a pair of primers, one from the amplification pair and one from the inner pair, gave better quality product for sequencing. The whole length of these two regions was assembled from these shorter overlapping segments.

The PCR mixture consisted of 10× PCR buffer containing 67 mM Tris pH 8.0, 2 mM $MgCl_2$, 250 μM of each dNTP, 2% DMSO, and 2 μM (for the *rpl20-rps12* spacer) or 10 μM of each primer. Temperature cycling conditions run on MJ Research (Waltham, Massachusetts, USA) or Techne (Princeton, New Jersey, USA) thermal cyclers were 5 min at 94°C; 38–40 cycles of 30 s at 94°C, 30–45 s at different annealing temperatures for each pair of primers; 1–2 min at 72°C; and 7 min at 72°C. The different annealing temperatures were 55°C for the *rpl20-rps12* spacer, 55.5–56.5°C for the *trnS-trnG* spacer, 58°C for the *nad1* and *nad4* introns, and 60°C for the *rpS14-cob* spacer. For the *nad7* intron, the cycling conditions were 5 min at 94°C; 17 cycles of 30 s at 94°C, 30 s at 62°C (−0.5°C per cycle), and 2 min at 72°C; 20 cycles of 30 s at 94°C, 30 s at 54°C, and 2 min at 72°C; and a final cycle of 7 min at 72°C. Agarose gel

TABLE 1.2. Chloroplast and mitochondrial DNA primer sequences used in this study.

Region	Name	F/R [b]	Primer sequence (5'-3')	Reference
cpDNA				
rpl20-rps12 spacer	rpl20 F	F	CCT CGT TTA ATT CTG GTC ATT G	
	5′-rps12	R	GTC GAG GAA CAT GTA CTA GG	Hamilton, 1999
trnS-trnG spacer	trnS	F	TAC AAC GGA TTA GCA ATC C	Doyle et al., 1992
	trnG-UCC 5′	R	ATA CCA CTA AAC TAT ACC C	Doyle et al., 1992
mtDNA				
nad1 intron [a]	nad1 exon B	F	GCA TTA CGA TCT GCA GCT CA	Demesure et al., 1995
	nad1 exon C	R	GGA GCT CGA TTA GTT TCT GC	Demesure et al., 1995
nad4 intron	nad4 exon 1	F	CAG TGG GTT GGT CTG GTA TG	Demesure et al., 1995
	nad4 exon 2	R	TCA TAT GGG CTA CTG AGG AG	Demesure et al., 1995
	nad4 innerF	F	TGT TTG TTC GCG AGA ATG GAT TCC	
	nad4 innerR	R	TGT AAA TCG GCG GTC CCT GTT TGG	
nad7 intron	nad7 exon4	F	AGG AGA TCG CTA TGA TCG TTA CTG T	
	nad7 exon5	R	ACG ACG GTA GGG ACG ATT GCT TC	
	nad7innerF1	F	GAA ACG CGG CGG CAT AGG AAC C	
	nad7innerR1	R	TCC CGC TTC CTC GCA TCT GCG C	
	nad7innerF2	F	CCC AAT GAT AAA CCA CTA ACA CCC	
	nad7innerR2	R	GCG AAG CGT TCT ATT GGT TTC CC	
rpS14-cob spacer	rpS14	F	CAG GGG TCG CCC TCG TTC CG	Demesure et al., 1995
	cob	R	GTG TGG AGG ATA TAG GTT GT	Demesure et al., 1995

[a] Not used in the analyses due to lack of variation detected in a subset of samples
[b] Forward or reverse primer

electrophoresis with ethidium bromide visualization was performed to check the success of the reaction and number of bands produced. Successful PCR products were separated on 0.8–1% agarose gels, excised, and then purified using QIAquick Gel Extraction Kit (Qiagen, Valencia, California, USA). Purified PCR products were used as templates for sequencing with primers used in amplifications and internal primers (for the *nad4* and *nad7* mitochondrial introns). DNA sequencing reactions were performed using BigDye® Terminator (Applied Biosystems, Foster City, California, USA) and run on an ABI 3730xl DNA sequencer (Applied Biosystems) by the Cornell Biotechnology Core Facility (CLC). Some sequencing reactions were performed using ABI BigDye® Terminator in 96-well plates and cleaned up by ethanol precipitation before being run on the DNA sequencer by the CLC.

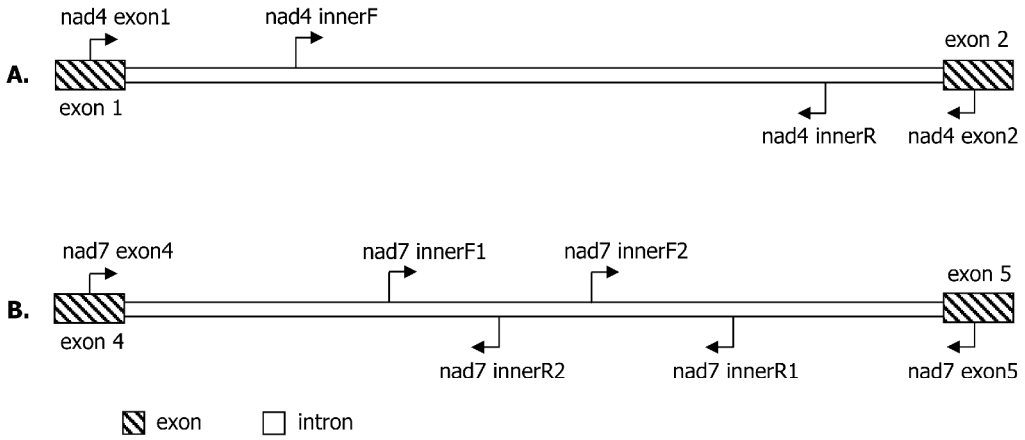


FIGURE 1.2. Primers and internal primers positions. A. *nad4* intron. B. *nad7* intron.

Data Analyses—SEQUENCE ALIGNMENT AND ANALYSES—For each individual, at least two sequences (forward and reverse, and internal sequences for the *nad4* and *nad7* mitochondrial introns) were aligned to obtain a full-length sequence for each region. Manual adjustment was performed to resolve reading errors found in any of these sequences in the contig. Due to similarity of sequences of each region across all the samples, it was not necessary to use a

multiple-sequence alignment program. Sequences were aligned using the Sequence Assembly feature on Sequencher™ versions 4.2 and 4.5 (Gene Codes, Co., Ann Arbor, Michigan, USA) and adjusted manually. Indels of variable length (one to four bp) in a mononucleotide A repeat found in the *trnS-trnG* spacer were excluded. Sequences of the two chloroplast regions and three mitochondrial regions were combined as cpDNA and mtDNA datasets, respectively, to maximize phylogenetic signal. Insertions/deletions (indels) were considered single mutation events in the alignment regardless of the length of each indel and coded as presence or absence characters as suggested in the simple indel coding method of Simmons and Ochoterena (2000). Coded indels were included in both datasets for phylogenetic and network analyses. Alignments used in this study are available from TreeBASE (study number SN4467). Haplotypes were assigned to sequences with unique combinations of polymorphisms in cpDNA and mtDNA datasets.

Phylogenetic analysis was performed using maximum parsimony (MP) implemented in PAUP* 4.0b10 (Swofford, 2003). Chloroplast DNA and mtDNA datasets were analyzed independently. Only one sequence representing each haplotype was used in the phylogenetic analysis. Heuristic searches were performed with 100 replicates of random stepwise additions holding 10 trees at each addition step with tree-bisection-reconnection (TBR) branch swapping and unlimited MaxTrees. Internal support for MP was estimated with 100 bootstrap replicates using full heuristic search with TBR and Multrees in effect.

Due to low variation at the intraspecific level, traditional phylogenetic analyses often result in poorly resolved phylogenetic trees. In addition, coexistence of a persistent ancestral haplotype and its multiple descendants results in a haplotype tree with multifurcations (Posada and Crandall, 2001). Network approaches take these population-level phenomena into account,

allowing more appropriate analysis of intraspecific data (Templeton et al., 1992). Network analysis was performed for each dataset using the statistical parsimony algorithm implemented in TCS ver. 1.21 (Clement et al., 2000). All sequences were included in the datasets to allow the calculation of haplotype frequencies.

ORGANELLAR DNA VARIATION AND GENETIC DIFFERENTIATION—Number of haplotypes (H), haplotype diversity (H_d), average number of nucleotide differences per site between two sequences (nucleotide diversity, π), and the proportion of segregating polymorphic sites per nucleotide (the Watterson estimator, θ_w) were calculated using DnaSP version 4.10.9 (Rozas et al., 2003) within each taxon and for all *M. sativa* accessions. To estimate the genetic differentiation between "caerulea" and "falcata" and between each of these two taxa and *M. prostrata*, Arlequin version 3.11 (Excoffier et al., 2005) was used to perform analysis of molecular variance (AMOVA) at two hierarchical levels (among and within taxa) for cpDNA and mtDNA datasets separately.

Results

Chloroplast Haplotypes—Amplifications of the *rpl20-rps12* and *trnS-trnG* intergenic spacers produced a single band in gel electrophoresis of the PCR product for every individual. However, DNA polymorphisms in single individuals were observed in "caerulea" accession PI 641380 at both cpDNA regions and in "falcata" accessions PI 494662 and PI 577558 at the *rpl20-rps12* region, suggesting heteroplasmy. In each case, only one position was polymorphic. Thus there were two haplotypes in each of the two heteroplasmic "falcata" individuals and up to four haplotypes in "caerulea" accession PI 641380. For the two heteroplasmic "falcata" individuals, combining sequences from the two regions was straightforward as each of them had

only one *trnS-trnG* sequence to be combined with one of the two possible *rpl20-rps12* sequences. One of the two haplotypes in each of these two individuals was found to be a singleton (Figure 1.3). Inference of haplotypes in "caerulea" PI 641380 was done using the subtraction method of Clark (1990). Four possible combined sequences in this individual were manually compared to unambiguous sequences found in homoplasmic individuals and only two of them matched sequences described from homoplasmic individuals, thus two haplotpyes were inferred in this individual (Figure 1.3). Two haplotypes from heteroplasmic individuals were named haplotype "a" and "b," e.g., Msc641380a and Msc641380b as shown in Figure 1.3. In total, 51 sequences were obtained from the 48 samples. Alignments were straightforward and required the manual addition of four gaps in the 719 bp *rpl20-rps12* alignment, and seven gaps for the 349 bp *trnS-trnG* alignment.

The combined sequence of the two regions after alignment was 1,071 bp (including seven coded indels) 15 of which were parsimony informative and 23 were autapomorphic. Thirty nucleotide substitutions and seven indels were found (14 substitutions and three indels, when *M. truncatula* was not included). With gaps coded, 20 haplotypes were identified, nine of which were singleton haplotypes, excluding *M. truncatula* (six singletons of 17 haplotypes, when indels were excluded). With indels not considered, haplotype (gene) diversity (H_d) for all *M. sativa* was 0.897 with nucleotide diversity (π) 0.00284 and $\theta_w = 0.00286$ (Table 1.3).

Maximum parsimony analysis identified 5,570 trees (length = 57; consistency index [CI] with/without autapomorphies = 0.737/0.559; retention index [RI] = 0.792). The strict consensus was mostly unresolved (trees not shown). In the statistical parsimony network analysis (Figure 1.3), haplotypes found in most "caerulea" and "falcata" fell into two separate groups. Eight "caerulea" accessions (53.3%) shared one of the two most common haplotypes and seven

accessions (30.4%) of "falcata" shared the other most common haplotypes with one accession of "caerulea" and one of subsp. *hemicycla*. *Medicago truncatula* was not connected to the network under the 95% statistical parsimony criterion of TCS. Two haplotypes of *M. prostrata*, a species considered a close ally of the complex (Lesins and Lesins, 1979), were outliers in the network. Haplotypes of subsp. *hemicycla* (the hypothesized hybrid of "caerulea" and "falcata") were identical to haplotypes found in one or the other putative parental taxon.

Analysis of molecular variance (AMOVA) of cpDNA data from "caerulea" versus "falcata" showed that 43.44% of total variation was explained by differences among taxa, indicating strongly significant genetic differentiation ($P < 0.0001$) between the two taxa (Table 1.4). Both taxa were also strongly differentiated from *M. prostrata* ($P < 0.0001$).

Haplotype differentiation within both the "caerulea" and the "falcata" haplotype clusters generally corresponds to "caerulea" and "falcata" ecogeographic subgroups based on nuclear SSR data (M. Şakiroğlu and E. C. Brummer, unpubl. data).

Mitochondrial Haplotypes—We were able to amplify all four mitochondrial regions, *nad1* intron, *nad4* intron, *nad7* intron, and *rpS14-cob* spacer. Single PCR products were obtained from the amplification of each region for each individual. Heteroplasmy was not observed in any of the mitochondrial regions.

Sequences of *nad1* were obtained from a subset of samples, including two *M. prostrata*, two "caerulea," 11 "falcata," one subsp. *hemicycla*, one subsp. *sativa* (CADL), and the outgroup, *M. truncatula*. All sequences were almost identical; only one substitution polymorphism was observed in a single accession from 1,513 bp of aligned sequences. With this low level of variation, this gene did not appear to be informative for this study and was not sampled further or included in subsequent analyses.

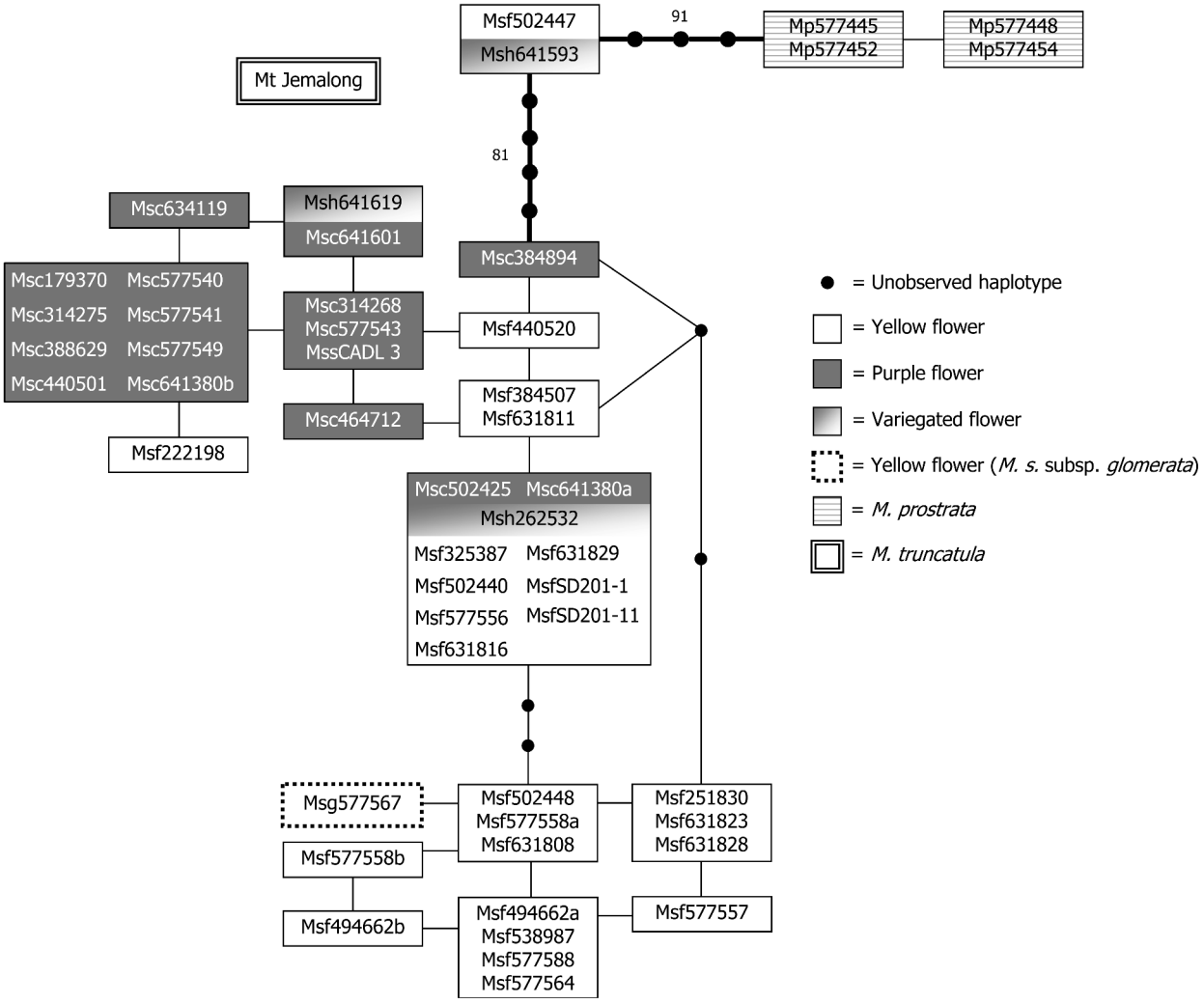


FIGURE 1.3. Chloroplast haplotype statistical parsimony network of diploid *Medicago*. Each box represents a haplotype and contains individual(s) that possess that haplotype. Taxon names are abbreviated as follows: Mp = *Medicago prostrata*; Msc = *M. sativa* subsp. *caerulea* ("caerulea"); Msf = *M. s.* subsp. *falcata* ("falcata"); Msh = *M. s.* subsp. *hemicycla*; Msg = *M. s.* subsp. *glomerata*; Mss = *M. s.* subsp. *sativa*; Mt = *M. truncatula*; names are followed by accession number (USDA Plant Introductory [PI] number) or variety name. Two haplotypes found in heteroplasmic individuals are designated with "a" or "b" at the end of the accession number. *Medicago truncatula* was not connected to the network under the 95% statistical parsimony criterion of TCS. Thick lines show branches resolved both on strict and majority-rule consensus trees from a maximum parsimony analysis. Bootstrap values (>50%) are shown above or to the left of branches.

TABLE 1.3. Summary statistics of chloroplast and mitochondrial DNA sequence variation for diploid *Medicago*. For each taxon, number of accessions and standard measures of diversity are indicated.

	Taxon	No. of accessions [a]	No. of polymorphic sites [b]	Gaps	*H*	H_d	π	θ_w
cpDNA	All *M. sativa*	43 (46)	13 (2)	20	15	0.897	0.00284	0.00286
	M. s. caerulea	15 (16)	5 (1)	18	7	0.750	0.00142	0.00146
	M. s. falcata	23 (25)	11 (6)	20	8	0.833	0.00223	0.00282
	M. s. glomerata	1 (1)	n/a	n/a	n/a	n/a	n/a	n/a
	M. s. hemicycla	3 (3)	9 (9)	18	3	1.000	0.00580	0.00580
	M. s. sativa	1 (1)	n/a	n/a	n/a	n/a	n/a	n/a
	M. prostrata	4 (4)	1 (0)	29	2	0.667	0.00065	0.00053
mtDNA	All *M. sativa*	32	23 (11)	27	18	0.861	0.00087	0.00113
	M. s. caerulea	9	16 (9)	17	7	0.917	0.00112	0.00116
	M. s. falcata	18	16 (8)	15	10	0.850	0.00083	0.00097
	M. s. glomerata	1	n/a	n/a	n/a	n/a	n/a	n/a
	M. s. hemicycla	3	5 (5)	23	3	1.000	0.00066	0.00066
	M. s. sativa	1	n/a	n/a	n/a	n/a	n/a	n/a
	M. prostrata	4	8 (1)	13	3	0.833	0.00102	0.00086

Notes: H = number of haplotypes; H_d = haplotype (gene) diversity; π = average number of nucleotide difference per site between two sequences (nucleotide diversity); θ_w = the Watterson estimator, proportion of segregating polymorphic sites per nucleotide ($\theta = 4N\mu$); n/a = not applicable.

[a] Number of sequences used in parentheses.

[b] Number of singleton variable sites in parentheses.

TABLE 1.4. Analysis of molecular variance (AMOVA) of genetic variation in the two chloroplast and three mitochondrial regions for diploid *Medicago*. Taxon names in the first column were abbreviated as follows: Mp = *Medicago prostrata*; Msc = *M. sativa* subsp. *caerulea* ("caerulea"); Msf = *M. s.* subsp. *falcata* ("falcata").

	cpDNA				mtDNA			
Source of variation	df	% of total variation	Φ_{ST}	*P*	df	% of total variation	Φ_{ST}	*P*
Msc vs Msf								
Among taxa	1	43.44	0.434	<0.0001	1	−1.67	−0.017	>0.10
Within taxa	39	56.56			25	101.67		
Msc vs Mp								
Among taxa	1	85.78	0.858	<0.001	1	−14.35	−0.143	>0.10
Within taxa	18	14.22			11	114.35		
Msf vs Mp								
Among taxa	1	69.65	0.696	<0.0001	1	−6.27	−0.063	>0.10
Within taxa	27	30.35			20	106.27		

The alignments of the three remaining mitochondrial regions were simple due to the similarity of the sequences across the samples. The lengths of the sequences of the three regions after alignment were 1,952 bp for *nad4* intron, 2,177 bp for *nad7* intron, and 958 bp for *rpS14-cob* spacer. These made the combined alignment 5,087 bp (5,101 bp when coded indels were included), of which 15 characters were parsimony informative and 25 were autapomorphic. Twenty-six nucleotide substitutions and 14 indels were found (24 substitutions and 12 indels when *M. truncatula* was not included). With gaps coded and *M. truncatula* excluded, 21 haplotypes were identified from 36 diploid individuals. Of these, 15 haplotypes were singleton haplotypes (19 haplotypes and 15 singleton haplotypes, when indels were excluded). With indels not considered, haplotype (gene) diversity (H_d) for all *M. sativa* was 0.861 with nucleotide diversity (π) 0.00087 and $\theta_w = 0.00113$ (Table 1.3).

Maximum parsimony analysis of combined mtDNA sequences of the 21 haplotypes identified 236 trees with 52 steps (CI with/without autapomorphies = 0.788/0.593; RI = 0.814). The strict consensus tree was mostly unresolved (trees not shown).

Statistical parsimony network analysis performed in TCS revealed no partitioning of haplotypes into groups corresponding to subspecies (Figure 1.4). The most common haplotype was shared by individuals from "caerulea," "falcata," and *M. prostrata*. *Medicago truncatula* was linked to the haplotype network under the 95% cut-off, reflecting the low divergences of the mitochondrial haplotypes even at the interspecific level.

AMOVA showed no significant differentiation between "caerulea" and "falcata" and between either of them and *M. prostrata* (Table 1.4).

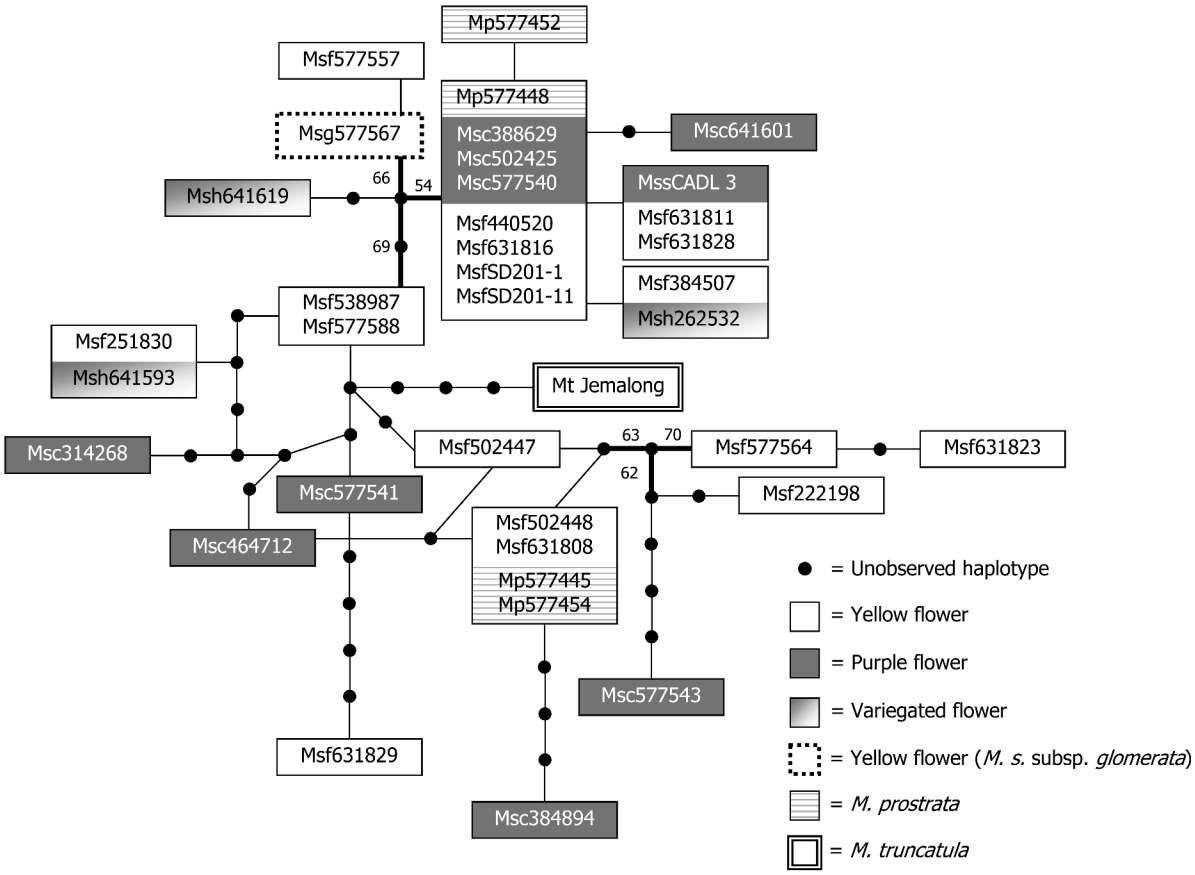


FIGURE 1.4. Mitochondrial haplotype statistical parsimony network of diploid *Medicago*. Conventions as in Figure 1.3.

Discussion

The taxonomic status of members of the *M. sativa* species complex has long been contentious (e.g., Lesins and Lesins, 1979; Quiros and Bauchan, 1988; Small and Jomphe, 1989). One of the main points of argument has been whether *M. s.* subsp. *falcata* should be recognized as a separate species or as a subspecies of *M. sativa* sensu lato. Lesins and Lesins (1979) ranked this taxon as a species, *M. falcata* L., based on its yellow flowers and sickle-shaped pods compared to the purple flowers and coiled pods of diploid and polyploid *M. sativa* (Table 1.1). Because the remaining members of the section all have coiled pods, Lesins and Lesins (1979) suggested that *M. falcata* is a "younger" species evolving in response to relatively recent

environmental changes. In contrast, Quiros and Bauchan (1988) considered the taxon as a subspecies of *M. sativa* based on cytological and genetic evidence. Small and Jomphe (1989) also treated this taxon as a subspecies of *M. sativa*. Despite the disagreement in the rank of *M. s.* subsp. *falcata*, they designated the diploid cytotype of purple flowered *M. sativa* as *M. s.* subsp. *caerulea* (= *M. s.* subsp. "*coerulea*" Schmalh. in Lesins and Lesins [1979] and Quiros and Bauchan [1988]; Small and Jomphe, 1989). We wished to determine whether the diploid form of "falcata" and diploid *M. sativa* ("caerulea") are genetically discrete taxa based on their organellar genomes, as suggested by their phenotypic differences. If they are not distinct, this would lend support to the view that they do not deserve recognition as separate species.

Previous molecular studies have disagreed concerning the distinction between "caerulea" and "falcata." Brummer et al. (1991) found that "falcata" plants formed phylogenetic clusters distinct from "caerulea" plants based on nuclear RFLP variation detected by 19 cDNA probes. In contrast, based on genetic distances determined from shared hypervariable cpDNA fragments, Skinner (2000) found that accessions of *Medicago* subspecies other than subsp. *sativa*, including "caerulea" and "falcata," did not readily cluster. However, results in Skinner (2000) may not reflect true relationships among *Medicago* taxa in the study given the tendency for excessive homoplasy in simple chloroplast repeats, even at low taxonomic levels (Doyle et al., 1998). Muller et al. (2003), found high levels of mtDNA RFLP diversity, but, similar to our results, could not detect genetic differentiation between the two taxa. Using nuclear sequence polymorphisms, Muller et al. (2006) found no differentiation between the two taxa. This is in contrast to our preliminary data based on sequences of the CNGC5 nuclear region from 34 of the diploid plants used in this study, which revealed differentiation between "caerulea" and "falcata" (T. Havananda and J. J. Doyle, unpubl. data).

Our cpDNA data represent the first intensive sampling of the diploid members of the complex at the sequence level. The cpDNA haplotype network showed that haplotypes from most accessions of "caerulea" formed a group separate from those of "falcata." The presence of cpDNA heteroplasmy in some individuals, a phenomenon found in this genus and expected given the biparental transmission of cpDNA (Johnson and Palmer, 1989), did not affect the overall picture. Statistically significant cpDNA haplotype differentiation between the two taxa also was supported by AMOVA results. A few shared haplotypes and mostly unresolved phylogenetic trees are consistent with limited gene flow due to incomplete reproductive isolation between the two taxa. Thus, the chloroplast genome appears to be tracking similar relationships as flower color and pod shape, which are the most important diagnostic morphological characters in the genus *Medicago* (Lesins and Lesins, 1979; Quiros and Bauchan, 1988; Small and Jomphe, 1989). Two nuclear loci are known to be involved in flower color in *M. sativa* and even though they both are on linkage group 1, they are not tightly linked (they are at least 39 cM apart; Kiss et al., 1993; Kalo et al., 2000). Several nuclear genes may be involved in pod shape in *Medicago*. Six non-linked genetic factors determining pod coiling were proposed from a study of hybridization between *M. hybrida* and *M. suffruticosa*, which are species in the same section with *M. sativa* (Lesins, 1969). Thoquet et al. (2002) named a gene responsible for the direction of pod coiling (clockwise vs. anticlockwise) *SPC* (Sense of Pod Coiling) and mapped it on linkage group 7 of *M. truncatula*, which is homologous to linkage group 7 of *M. sativa* (Kalo et al., 2000). Other genes may be involved in pod shape in legumes, for example, three pod length quantitative trait loci (QTLs) and three pod width QTLs were detected on chromosomes 1, 5, and 6 and chromosomes 1, 2, and 5, respectively, in *Lotus japonicus* (Gondo et al., 2007). With several genes from different linkage groups involved in floral and pod characters, it may suggest

that variation patterns of these traits are representative of many segments of the nuclear genome. The agreement between the distribution of these traits and chloroplast haplotype variation suggests that the chloroplast genome is tracking the same history as much of the nuclear genome.

In contrast, differentiation between these two taxa was not observed based on mitochondrial haplotypes identified from the same individuals, even though, by sequencing over three times as many nucleotides per individual, we detected a similar number of haplotypes as were identified from chloroplast regions (Table 1.3). In plants, incongruence is often observed between nuclear and chloroplast datasets (e.g., Rieseberg et al., 1996; van der Niet and Linder, 2008), but far fewer examples exist involving the plant mitochondrial genome (e.g., Maureira-Butler et al., 2008; Tsutsui et al., 2009). This is presumably because the mitochondrial genome generally provides much less sequence variation and is thus infrequently used in plant systematic studies (Palmer, 1992).

Either incomplete lineage sorting or hybridization could explain the lack of genetic differentiation between "caerulea" and "falcata" we found in our mtDNA haplotype data. However, lineage sorting alone cannot explain the differences between our mtDNA and cpDNA results. Because of uniparental transmission, the effective population size of mtDNA is smaller than that of cpDNA (biparentally transmitted, though largely paternally inherited in *Medicago*) or nDNA. With smaller effective population size, mitochondrial haplotypes should sort faster because their expected coalescence time is shorter than that for haplotypes randomly drawn from an effectively larger population (e.g., Moore, 1995). With strong paternal inheritance bias of cpDNA, it may be argued that the effective population size of cpDNA may not be much larger than that of mtDNA. Even if that is the case, mtDNA haplotypes should not sort out more slowly than cpDNA haplotypes. Thus, if both mtDNA and cpDNA are neutrally evolving, and if

cpDNA shows differentiation between the two taxa, then lineage sorting of mtDNA haplotypes should be more complete and a similar or greater degree of differentiation of the two subspecies should be observed in mtDNA. However, this is not observed in our mtDNA analyses.

If incomplete lineage sorting cannot fully explain the lack of mtDNA differentiation between the two subspecies, hybridization or introgression is likely to be responsible. Maureira-Butler et al. (2008) provided evidence that hybridization is a pervasive and ongoing process throughout the history of *Medicago*. In the ***M. sativa*** complex, the members are outcrossing and interfertile, resulting in natural hybrids such as the named taxon ***M. s.*** subsp. ***hemicycla*** (Lesins and Lesins, 1979; Quiros and Bauchan, 1988; Small and Jomphe, 1989). With the ability to hybridize and frequent occurrence of fertile hybrids, introgression can take place (Ellstrand et al., 1999). Even though introgression is common in plants (Rieseberg and Wendel, 1993), fewer studies have reported mtDNA introgression in plants (e.g., Senjo et al., 1999; Martinsen et al., 2001) compared to the many cases in animals (e.g., reviewed in Ballard and Whitlock, 2004; McGuire et al., 2007), again presumably because mtDNA is less commonly used in plant than in animal studies. One of the consequences of introgression is ambiguous taxonomic boundaries in populations or species involved in the introgression (Rieseberg and Wendel, 1993). If introgression of mtDNA has been taking place from "caerulea" to "falcata" or vice versa, any sign of prior divergence between the two subspecies may be confounded and no significant genetic differentiation will be detected.

Cytoplasmic genomes (chloroplast and mitochondrial) often have greater rates of introgression than nuclear DNA, due to their haploid nature and uniparental inheritance (Martinsen et al., 2001; Ballard and Whitlock, 2004). Introgression involving the chloroplast genome is pervasive enough in plants to have generated the term "chloroplast capture"

(Rieseberg and Soltis, 1991). Chan and Levin (2005) demonstrated that the rate of introgression depends on the type of reproductive isolation barriers and mode of genetic inheritance. They suggested that prezygotic isolation is a general explanation for biased introgression of maternally-inherited genomic components and that maternally inherited DNA may introgress more rapidly than paternally and biparentally inherited DNA. In *M. sativa*, cpDNA inheritance is biparental, albeit strongly paternal, whereas mtDNA inheritance is strictly maternal (Schumann and Hancock, 1989; Forsthoefel et al., 1992). Even if introgression of cpDNA has occurred in this species, it should be at a slower rate than that of mtDNA due to the different mode of inheritance. This could explain why we did not observe the effect of introgression on cpDNA differentiation between the two differentiated subspecies as we observed from mtDNA.

Hybridization between "caerulea" and "falcata" has long been hypothesized, with *M. s.* subsp. *hemicycla* thought to be their natural hybrid (Lesins and Lesins, 1979; Quiros and Bauchan, 1988). Because of the possibility of biparental transmission of cpDNA (Smith et al., 1986), *M s.* subsp. *hemicycla* might be expected to be heteroplasmic if it is an F1 hybrid. However, due to the cpDNA paternal transmission bias, heteroplasmy would be unlikely even in an F1 hybrid. We did not observe heteroplasmy in our *M. s.* subsp. *hemicycla* individuals, which were from maintained populations of wild material, and hence were not F1 hybrids. Because of the strong paternal transmission bias of cpDNA (Schumann and Hancock, 1989) it is possible, for example, that subsp. *hemicycla* PI 641619, which has a chloroplast haplotype identical to that of "caerulea" PI 641601 (Figure 1.3), may have "caerulea" as its male parent. However, a possibility of this haplotype coming from the female parent cannot be ruled out because of the biparental cpDNA inheritance. Similarly, mtDNA haplotypes from subsp. *hemicycla* samples may suggest the identity of their female parents based on the strict maternal inheritance of

mtDNA (Schumann and Hancock, 1989; Forsthoefel et al., 1992). But since there was no differentiation between mtDNA haplotypes of "caerulea" and "falcata," subsp. *hemicycla* that possesses identical haplotypes with "caerulea," for example, may not necessarily have "caerulea" as female parent if that haplotype was introgressed from "falcata."

Lesins and Lesins (1979) hypothesized diploid *M. s.* subsp. *glomerata*, characterized by yellow flowers and coiled pods, to be the ancestor of "caerulea" and "falcata." Under coalescent theory, the most frequent ancestral haplotype is most likely the oldest (Posada and Crandall, 2001). It may be expected that subsp. *glomerata* would have a haplotype that is most common in the entire complex. The cpDNA haplotype from the single diploid subsp. *glomerata* accession (PI 577567) surveyed in this study was unique to that taxon and sister to the "falcata" haplotype group (Figure 1.3), suggesting that this haplotype is not an ancestral haplotype. However, because subsp. *glomerata* can hybridize freely with "falcata" (Lesins, 1968), the haplotype from this accession may be a "falcata" haplotype obtained through hybridization.

The relationships of *M. prostrata* cpDNA and mtDNA haplotypes observed in our study suggest the possibility of a complex history for these accessions, which contain well-differentiated cpDNA haplotypes, but have mtDNA haplotypes very similar to "caerulea" and "falcata." Given that cpDNA is largely paternally inherited whereas mtDNA is strictly maternally inherited in *Medicago* (Schumann and Hancock, 1989; Forsthoefel et al., 1992) and that *M. prostrata* can cross successfully with either of these two subspecies when *M. prostrata* is a male parent (Quiros and Bauchan, 1988), these *M. prostrata* accessions could be the products of crosses in which *M. prostrata* was the male parent and either "caerulea" or "falcata" was the female parent.

Given the morphological differences and chloroplast genomic differentiation between "caerulea" and "falcata," it seems clear that they merit recognition as separate taxa. It is the rank at which they should be recognized that is contentious, which is not surprising given the ongoing debate about species concepts and recognition criteria. If species are segments of separately evolving metapopulation lineages, whose different biological properties serve as lines of evidence supporting a hypothesis of lineage separation, it is not expected that evidence of population separation will evolve simultaneously for all characters, or in the same sequence in all cases (de Queiroz, 2007). For example, reproductive isolation may evolve before sorting of polymorphisms in some groups, or only afterwards in others, with fixation of morphological characters only weakly correlated with either reproductive isolation or fixation of neutral alleles (de Queiroz, 2007). In the *M. sativa* complex, reproductively compatible taxa have distinct morphologies and it has been speculated that some of these characters have an adaptive basis. Lesins and Lesins (1979) hypothesized that the coiled pods of purple-flowered "caerulea" were adapted for dispersal by rolling with the wind on open ground in warm, dry, semi-desert conditions, in contrast to the straighter legumes of yellow-flowered "falcata," which is adapted to a steppe environment. If chloroplast and mitochondrial variation are neutrally-evolving, it is possible that these genomes might not become fixed as quickly as adaptively-driven morphological features. The free introgression of these genomes would also be possible, with uniparentally transmitted mtDNA showing greater evidence of introgression than biparentally transmitted cpDNA. None of this, however, answers the practical question of whether it is more appropriate to consider these closely related, morphologically distinct taxa that share haplotypes, potentially due to both incomplete sorting and introgression, as species or as subspecies. Because species ranking, in contrast with species grouping, is tied to multiple potentially conflicting

semisubjective criteria (Baum, 2009), we recommend continuing the more recent practice of regarding them as subspecies.

ACKNOWLEDGEMENTS—We would like to thank Muhammet Şakiroğlu for kindly providing us genomic DNA and information on the plant materials. We are grateful to Jane Doyle and others in the Doyle lab for technical help, valuable discussion, and comments. We also thank Donald Viands, Melissa Luckow, and two anonymous reviewers for their helpful comments on the manuscript. Financial support to TH by the Office of the Higher Education Commission, the Royal Thai Government is greatly appreciated. This work was supported by NSF DEB-051667 to JJD.

REFERENCES

ADAMS, K. L. 2007. Evolution of duplicate gene expression in polyploid and hybrid plants. *Journal of Heredity* 98: 136–141.

ADAMS, K. L. AND J. F. WENDEL. 2005. Novel patterns of gene expression in polyploid plants. *Trends in Genetics* 21: 539–543.

BALLARD, J. W. O. AND M. C. WHITLOCK. 2004. The incomplete natural history of mitochondria. *Molecular Ecology* 13: 729–744.

BAUCHAN, G. R. AND M. A. HOSSAIN. 1997. Karyotypic analysis of C-banded chromosomes of diploid alfalfa: *Medicago sativa* ssp. *caerulea* and ssp. *falcata* and their hybrid. *Journal of Heredity* 88: 533–537.

BAUM, D. A. 2009. Species as ranked taxa. *Systematic Biology* 58: 74–86.

BENA, G. 2001. Molecular phylogeny supports the morphologically based taxonomic transfer of the "medicoid" *Trigonella* species to the genus *Medicago* L. *Plant Systematics and Evolution* 229: 217–236.

BENA, G., M.-F. JUBIER, I. OLIVIERI, AND B. LEJEUNE. 1998a. Ribosomal external and internal transcribed spacers: Combined use in the phylogenetic analysis of *Medicago* (Leguminosae). *Journal of Molecular Evolution* 46: 299–306.

BENA, G., B. LEJEUNE, J.-M. PROSPERI, AND I. OLIVIERI. 1998b. Molecular phylogenetic approach for studying life-history evolution: The ambiguous example of the genus *Medicago* L. *Proceedings of the Royal Society*, *B*, *Biological Sciences* 265: 1141–1151.

BENA,G., J.-M. PROSPERI, B. LEJEUNE, AND I. OLIVIERI. 1998c. Evolution of annual species of the genus *Medicago*: A molecular phylogenetic approach. *Molecular Phylogenetic and Evolution* 9: 552–559.

BINGHAM, E. T. AND T. J. MCCOY. 1979. Cultivated alfalfa at the diploid level: Origin, reproductive stability, and yield of seed and forage. *Crop Science* 19: 97–100.

BOE, A., R. BORTNEM, K. F. HIGGINS, A. D. KRUSE, K. D. KEPHART, AND S. SELMAN. 1998. Breeding yellow-flowered alfalfa for combined wildlife habitat and forage purposes. South Dakota Agricultural Experiment Station Bulletin 727, South Dakota State University, Brookings, South Dakota, USA.

BRUMMER, E. C. 2004. Genomics research in alfalfa, *Medicago sativa* L. *In* R. F. Wilson, H. T. Stalker, and E. C. Brummer [eds.], Legume crop genomics, 110–142. AOCS Press, Champaign, Illinois, USA.

BRUMMER, E. C., P. CAZCARRO, AND D. LUTH. 1999. Ploidy determination of alfalfa germplasm accessions using flow cytometry. *Crop Science* 39: 1202–1207.

BRUMMER, E. C., G. KOCHERT, AND J. H. BOUTON. 1991. RFLP variation in diploid and tetraploid alfalfa. *Theoretical and Applied Genetics* 83: 89–96.

CHAN, K. M. A. AND S. A. LEVIN. 2005. Leaky prezygotic isolation and porous genomes: Rapid introgression of maternally inherited DNA. *Evolution* 59: 720–729.

CLARK, A. G. 1990. Inference of haplotypes from PCR-amplified samples of diploid populations. *Molecular Biology and Evolution* 7: 111–122.

CLEMENT M, D. POSADA, AND K. CRANDALL. 2000. TCS: A computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657–1659.

COMAI, L. 2005. The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* 6: 836–846.

DEMESURE, B., N. SODZI, AND R. J. PETIT. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* 4: 129–131.

DE QUEIROZ, K. 2007. Species concepts and species delimitation. *Systematic Biology* 56: 879–886.

DIWAN, N., A. A. BHAGWAT, G. B. BAUCHAN, AND P. B. CREGAN. 1997. Simple sequence repeat DNA markers in alfalfa and perennial and annual *Medicago* species. *Genome* 40: 887–895.

DOWNIE, S. R., D. S. KATZ-DOWNIE, E. J. ROGERS, H. L. ZUJEWSKI, AND E. SMALL. 1998. Multiple independent losses of the plastid *rpo*C1 intron in *Medicago* (Fabaceae) as inferred from phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacer sequences. *Canadian Journal of Botany* 76: 791–803.

DOYLE, J. J. AND J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

DOYLE, J. J., J. I. DAVIS, R. J. SORENG, D. GARVIN, AND M. J. ANDERSON. 1992. Chloroplast DNA inversions and the origin of the grass family (Poaceae). *Proceedings of the National Academy of Sciences, USA* 89: 7722–7726.

DOYLE, J. J., J. L. DOYLE, J. T. RAUSCHER, AND A. H. D. BROWN. 2004. Evolution of the perennial soybean polyploid complex (*Glycine* subgenus *Glycine*): A study of contrasts. *Biological Journal of the Linnean Society* 82: 583–597.

DOYLE, J. J., M. MORGANTE, S. V. TINGEY, AND W. POWELL. 1998. Size homoplasy in chloroplast microsatellites of wild perennial relatives of soybean (*Glycine* subgenus *Glycine*). *Molecular Biology and Evolution* 15: 215–218.

DUBCOVSKY, J. AND J. DVORAK. 2007. Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* 316: 1862–1866.

ELLSTRAND, N. C., H. C. PRENTICE, AND J. F. HANCOCK. 1999. Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics* 30: 539–563.

EXCOFFIER, L., G. LAVAL, AND S. SCHNEIDER. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1: 47–50.

FALAHATI-ANBARAN, M., A. A. HABASHI, M. ESFAHANY, S. A. MOHAMMADI, AND B. GHAREYAZIE. 2007. Population genetic structure based on SSR markers in alfalfa (*Medicago sativa* L.) from various regions contiguous to the centres of origin of the species. *Journal of Genetics* 86: 59–63.

FORSTHOEFEL, N. R., H. J. BOHNERT, AND S. E. SMITH. 1992. Discordant inheritance of mitochondrial and plastid DNA in diverse alfalfa genotypes. *Journal of Heredity* 83: 342–345.

GAETA, R. T., J. C. PIRES, F. INIGUEZ-LUY, E. LEON, AND T. C. OSBORN. 2007. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *Plant Cell* 19: 3403–3417.

GHÉRARDI, M., B. MANGIN, B. GOFFINET, D. BONNET, AND T. HUGUET. 1998. A method to measure genetic distance between allogamous populations of alfalfa (*Medicago sativa*) using RAPD molecular markers. *Theoretical and Applied Genetics* 96: 406–412.

GONDO, T., S. SATO, K. OKUMURA, S. TABATA, R. AKASHI, AND S. ISOBE. 2007. Quantitative trait locus analysis of multiple agronomic traits in the model legume *Lotus japonicus*. *Genome* 50: 627–637.

HAMILTON, M. B. 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* 8: 521–523.

JOHNSON, L. B. AND J. D. PALMER. 1989. Heteroplasmy of chloroplast DNA in *Medicago*. *Plant Molecular Biology* 12: 3–11.

KALÓ, P., G. ENDRE, L. ZIMÁNYI, G. CSANÁDI, AND G. B. KISS. 2000. Construction of an improved linkage map of diploid alfalfa (*Medicago sativa*). *Theoretical and Applied Genetics* 100: 641–657.

KIDWELL, K. K., D. F. AUSTIN, AND T. C. OSBORN. 1994. RFLP evaluation of nine *Medicago* accessions representing the original germplasm sources for North-American alfalfa cultivars. *Crop Science* 34: 230–236.

KISS, G. B., G. CSANÁDI, K. KÁLMÁN, P. KALÓ, AND L. ÖKRÉSZ. 1993. Construction of a basic genetic map for alfalfa using RFLP, RAPD, isozyme and morphological markers. *Molecular and General Genetics* 238: 129–137.

LESINS, K. 1962. Interspecific crosses involving alfafa. III. *Medicago sativa* L. × *M. prostrata* Jacq. *Canadian Journal of Genetics and Cytology* 4: 14–23.

LESINS, K. 1968. Interspecific crosses involving alfalfa. IV. *Medicago glomerata* × *M. sativa* with reference to *M. prostrata*. *Canadian Journal of Genetics and Cytology* 10: 536–544.

LESINS, K. 1969. Relationship of taxa in genus *Medicago* as revealed by hybridization. IV. *M. hybrida* × *M. suffruticosa*. *Canadian Journal of Genetics and Cytology* 11: 340–345.

LESINS, K. A. AND I. LESINS. 1979. Genus *Medicago* (Leguminosae): A taxogennetic study. The Hague: Dr. W. Junk, The Hague, Netherlands.

MARTINSEN, G. D., T. G. WHITHAM, R. J. TUREK, AND P. KEIM. 2001. Hybrid populations selectively filter gene introgression between species. *Evolution* 55: 1325–1335.

MASOUD, S. A., L. B. JOHNSON, AND E. L. SORENSEN. 1990. High transmission of paternal plastid DNA in alfalfa plants demonstrated by restriction fragment polymorphic analysis. *Theoretical and Applied Genetics* 79: 49–55.

MAUREIRA, I. J. 2004. Alfalfa (*Medicago sativa* subsp. *sativa*) population improvement: Genetic diversity within and among *Medicago* species and trait-associations with molecular markers. Ph.D. dissertation, University of Wisconsin, Madison, Wisconsin, USA.

MAUREIRA-BUTLER, I. J., B. E. PFEIL, A. MUANGPROM, T. C. OSBORN, AND J. J. DOYLE. 2008. The reticulate history of *Medicago* (Fabaceae). *Systematic Biology* 57: 466–482.

MCCOY, T. J. 1982. The inheritance of *2n* pollen formation in diploid alfalfa *Medicago sativa*. *Canadian Journal of Genetics and Cytology* 24: 315–323.

MCCOY, T. J. AND E. T. BINGHAM. 1988. Cytology and cytogenetics of alfalfa. *In* A. A. Hanson [ed.], D. K. Barnes and R. R. Hill Jr. [co-eds.], Alfalfa and alfalfa improvement, Agronomy monograph 29, 737–776. American Society of Agronomy, Crop Science Society of Agronomy, and Soil Science Society of Agronomy, Madison, Wisconsin, USA.

MCGUIRE, J. A., C. W. LINKEM, M. S. KOO, D. W. HUTCHISON, A. K. LAPPIN, D. I. ORANGE, J. LEMOS-ESPINAL, B. R. RIDDLE, AND J. R. JAEGER. 2007. Mitochondrial introgression and incomplete lineage sorting through space and time: Phylogenetics of crotaphytid lizards. *Evolution* 61: 2879–2897.

MOORE, W. S. 1995. Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 49: 718–726.

MULLER, M.-H., C. PONCET, J. M. PROSPERI, S. SANTONI, AND J. RONFORT. 2006. Domestication history in the *Medicago sativa* species complex: Inferences from nuclear sequence polymorphism. *Molecular Ecology* 15: 1589–1602.

MULLER, M. H., J. M. PROSPERI, S. SANTONI, AND J. RONFORT. 2001. How mitochondrial DNA diversity can help to understand the dynamics of wild-cultivated complexes. The case of *Medicago sativa* in Spain. *Molecular Ecology* 10: 2753–2763

MULLER, M. H., J. M. PROSPERI, S. SANTONI, AND J. RONFORT. 2003. Inferences from mitochondrial DNA patterns on the domestication history of alfalfa (*Medicago sativa*). *Molecular Ecology* 12: 2187–2199.

PALMER, J. D. 1992. Mitochondrial DNA in plant systematics: Applications and limitations. *In* D. Soltis, P. Soltis, and J. J. Doyle [eds.], Molecular systematics of plants, 36–49. Chapman and Hall, New York, USA.

POSADA, D. AND K. A. CRANDALL. 2001. Intraspecific gene genealogies: Trees grafting into networks. *Trends in Ecology & Evolution* 16: 37–45.

QUIROS, C. F. 1982. Tetrasomic segregation for multiple alleles in alfalfa. *Genetics* 101: 117–127.

QUIROS, C. F. AND G. R. BAUCHAN. 1988. The genus *Medicago* and the origin of the *Medicago sativa* complex. *In* A. A. Hanson [ed.], D. K. Barnes and R. R. Hill Jr. [co-eds.], Alfalfa and alfalfa improvement, Agronomy monograph 29, 93–124. American Society of Agronomy, Crop Science Society of Agronomy, and Soil Science Society of Agronomy, Madison, Wisconsin, USA.

RIESEBERG, L. H. AND D. E. SOLTIS. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* 5: 65–84.

RIESEBERG, L. H. AND J. F. WENDEL. 1993. Introgression and its consequences in plants. *In* R. G. Harrison [ed.], Hybrid zones and the evolutionary process, 70–109. Oxford University Press, New York, New York, USA.

RIESEBERG, L. H., J. WHITTON, AND C. R. LINDER. 1996. Molecular marker incongruence in plant hybrid zones and phylogenetic trees. *Acta Botanica Neerlandica* 45: 243–262.

ROZAS, J., J. C. SÁNCHEZ-DELBARRIO, X. MESSEGUER, AND R. ROZAS. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496–2497.

SCHUMANN, C. M. AND J. F. HANCOCK. 1989. Paternal inheritance of plastids in *Medicago sativa*. *Theoretical and Applied Genetics* 78: 863–866.

SEGOVIA-LERMA, A., R. G. CANTRELL, J. M. CONWAY, AND I. M. RAY. 2003. AFLP-based assessment of genetic diversity among nine alfalfa germplasms using bulk DNA templates. *Genome* 46: 51–58

SENJO, M., K. KIMURA, Y. WATANO, K. UEDA, AND T. SHIMIZU. 1999. Extensive mitochondrial introgression from *Pinus pumila* to *P. parviflora* var. *pentaphylla* (Pinaceae). *Journal of Plant Research* 112: 97–105.

SIMMONS, M. P. AND H. OCHOTERENA. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369–381.

SKINNER, D. Z. 2000. Non random chloroplast DNA hypervariablility in *Medicago sativa*. *Theoretical and Applied Genetics* 101: 1242–1249.

SMALL, E. AND M. JOMPHE. 1989. A synopsis of the genus *Medicago* (Leguminosae). *Canadian Journal of Botany* 67: 3260–3294.

SMITH, S. E. 1989. Influence of parental genotype on plastid inheritance in *Medicago sativa*. *Journal of Heredity* 80: 214–217.

SMITH, S. E., E. T. BINGHAM, R. W. FULTON. 1986. Transmission of chlorophyll deficiencies in *Medicago sativa*: Evidence for biparental inheritance of plastids. *Journal of Heredity* 77: 35–38.

SOLITS, D. E., P. S. SOLTIS, J. C. PIRES, A. KOVARIK, J. A. TATE, AND E. MAVRODIEV. 2004a. Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): Cytogenetic, genomic and genetic comparisons. *Biological Journal of the Linnean Society* 82: 485–501.

SOLTIS, D.E., P. S. SOLTIS, D. W. SCHEMSKE, J. F. HANCOCK, J. N. THOMPSON, B. C. HUSBAND, AND W. S. JUDD. 2007. Autopolyploidy in angiosperms: Have we grossly underestimated the number of species? *Taxon* 56: 13–30.

SOLTIS, D. E., P. S. SOLTIS, AND J. A. TATE. 2004b. Advances in the study of polyploidy since *Plant Speciation*. *New Phytologist* 161: 173–191.

SOLTIS, P. S. AND D. E. SOLTIS. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences*, *USA* 97: 7051–7057.

STANFORD, E. H. 1951. Tetrasomic inheritance in alfalfa. *Agronomy Journal* 43: 222–225.

STEELE, K. P. AND M. F. WOJCIECHOWSKI. 2003. Phylogenetic analyses of tribes Trifolieae and Vicieae, based on sequences of the plastid gene, *mat*K (Papilionoideae: Leguminosae). *In* B. B. Klitgaard and A. Bruneau [eds.], Advances in legume systematics, part 10, Higher level systematics, 355–370. Royal Botanic Gardens, Kew, UK.

SWOFFORD, D. L. 2003. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sinauer, Sunderland, Massachusetts, USA.

TEMPLETON, A. R., K. A. CRANDALL, AND C. F. SING. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132: 619–633.

THOQUET, P., M. GHÉRARDI, E.-P. JOURNET, A. KERESZT, J.-M. ANÉ, J.-M. PROSPERI, AND T. HUGUET. 2002. The molecular genetic linkage map of the model legume *Medicago truncatula*: an essential tool for comparative legume genomics and the isolation of agronomically important genes. *BMC Plant Biology* 2: 1.

TSUTSUI, K., A. SUWA, K. SAWADA, T. KATO, T. A. OHSAWA, AND Y. WATANO. 2009. Incongruence among mitochondrial, chloroplast, and nuclear gene trees in *Pinus* subgenus *Strobus* (Pinaceae). *Journal of Plant Research* 122: 509–521.

VALIZADEH, M., K. K. KANG, A. KANNO, AND T. KAMEYA. 1996. Analysis of genetic distance among nine *Medicago* species by using DNA polymorphisms. *Breeding Science* 46: 7–10.

VAN DER NIET, T. AND H. P. LINDER. 2008. Dealing with incongruence in the quest for the species tree: a case study from the orchid genus *Satyrium*. *Molecular Phylogenetics and Evolution* 47: 154–174.

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: 507–517.

WENDEL, J. F. 2000. Genome evolution in polyploids. *Plant Molecular Biology* 42: 225–249.

CHAPTER 2

COMPLEX PATTERNS OF AUTOPOLYPLOID EVOLUTION IN ALFALFA AND ALLIES (*MEDICAGO SATIVA*; LEGUMINOSAE)*

Abstract

• *Premise of the study:* Although there is growing evidence that autopolyploidy is a widespread and important evolutionary phenomenon, it has received less attention than allopolyploidy. *Medicago sativa* comprises several diploid and autopolyploid taxa, including autotetraploid cultivated alfalfa, and affords an opportunity to elucidate the evolutionary history of a morphologically and genetically complex autopolyploid system.

• *Methods:* Phylogenies and haplotype networks were constructed from two chloroplast noncoding regions (*rpl20-rps12* and *trnS-trnG* spacers) across seven diploid and polyploid infraspecific taxa of *M. sativa* and five additional closely related *Medicago* species, and genetic differentiation was estimated.

• *Key results:* The two most prominent *M. sativa* autopolyploids have contrasting evolutionary histories. Chloroplast data support a simple autopolyploid origin of subsp. *sativa* (alfalfa) from diploid subsp. *caerulea*, from which it is distinguishable in several quantitative characters. In contrast, morphologically identical diploid and autopolyploid cytotypes of subsp. *falcata* were found to possess very different chloroplast haplotypes, suggesting past introgression from *M. prostrata* into the polyploid. Despite the presence of hybrids between tetraploid subspecies *falcata* and *sativa*, there was little evidence of introgression of chloroplast genomes from either subspecies into the other.

* This chapter was published in Havananda, T., E. C. Brummer, and J. J. Doyle. 2011. Complex patterns of autopolyploid evolution in alfalfa and allies (*Medicago sativa*; Leguminosae). *American Journal of Botany* 98: 1633–1646. Written authorization for the use of the material in this dissertation was obtained from *American Journal of Botany*.

• *Conclusions:* Autopolyploid evolution in *M. sativa* is complicated and has followed very different paths in different subspecific taxa. The potential exists for gene flow in virtually all combinations of subspecies both within and between ploidies, yet despite the existence of hybrids, morphologically and genetically distinctive subspecies persist.

Introduction

Polyploidy is prevalent in plant evolution. Fossil data suggest that 70% of angiosperms have a polyploid ancestry (Masterson, 1994), and it is now clear that angiosperms, and indeed all seed plants, are fundamentally polyploid (Jiao et al., 2011). A recent evaluation of chromosome numbers has led to the conclusion that 15% of speciation events in angiosperms involve polyploidization (Wood et al., 2009). With so much evidence for multiple genome-wide duplications in many angiosperm genomes, Soltis et al. (2009, p. 336) suggested that "the question is no longer 'What proportion of angiosperms are polyploid?', but 'How many episodes of polyploidy characterize any given lineage?'" Generally, polyploidy is categorized into two types: allopolyploidy and autopolyploidy. A "classic" allopolyploid in both the genetic and taxonomic senses is a product of chromosome doubling that generally involves hybridization of two or more differentiated genomes from different species, forms bivalents at meiosis, and has disomic genetic segregation, whereas an autopolyploid is produced by genome doubling generally within one species, may form multivalents at meiosis, and could have polysomic segregation (e.g., Wendel and Doyle, 2005; Doyle and Egan, 2010). However, many different combinations of these characteristics exist (Ramsey and Schemske, 1998). Given the many unanswered questions concerning polyploid evolution (Soltis et al., 2010), our understanding of this important process will be enhanced by empirical studies covering the full spectrum of

genetic and taxonomic possibilities. This is especially true of autopolyploidy, which is less studied than allopolyploidy, despite being more common and having more evolutionary advantages than typically recognized (Soltis and Soltis, 2000; Soltis et al., 2004, 2007; Parisod et al., 2010). This is even reflected in nomenclature, and Soltis et al. (2007) noted that the traditional practice of lumping autopolyploids and their diploid progenitor(s) together under the same specific epithet fails to capture the biological reality that autopolyploids often achieve their own unique evolutionary trajectory.

Medicago sativa L. is an attractive system for studying autopolyploidy for several reasons. One of its tetraploid members, cultivated alfalfa, is one of the most important forage crops in the world, grown on over 80 million acres worldwide (Michaud et al., 1988; Frame et al., 1997; Russelle, 2001). A great deal is known about its genetics through decades of plant breeding studies, and it is rich in genetic resources. However, much less is known about other tetraploids in this species or about the diploids hypothesized to have given rise to them and to alfalfa, despite the use of some of these taxa in alfalfa breeding programs. Indeed, it is still an open question whether there may be taxonomic allopolyploids in this species, given the morphological diversity and controversial taxonomic history of diploid members (e.g., Havananda et al., 2010).

Medicago sativa is composed of several perennial, outcrossing, and often interfertile taxa from section *Medicago* of the genus *Medicago* (Small and Jomphe, 1989). It is a polyploid complex comprising two ploidal levels, diploid ($2n = 2x = 16$) and tetraploid ($2n = 4x = 32$), with only weak hybridization barriers at the same and across ploidal levels (the latter through unreduced $2n$ gametes produced from a diploid parent: Bingham, 1968; McCoy, 1982; McCoy and Bingham, 1988; Quiros and Bauchan, 1988). The tetraploids in this species are genetic

autotetraploids, exhibiting tetrasomic inheritance (Stanford, 1951; Quiros, 1982; McCoy and Bingham, 1988) and occasional quadrivalents at meiosis (Armstrong, 1971; Gillies, 1972; Mariani and Veronesi, 1979). Flower color, pod shape, and pod pubescence are key morphological characters distinguishing three principal diploid members of this species, with variation of these characters and ploidal level characterizing other members (Table 2.1). Taxonomic treatments of *M. sativa* have long conflicted regarding ranking and nomenclature. Lesins and Lesins (1979) ranked most taxa in *M. sativa* as different species, whereas Quiros and Bauchan (1988) and Small and Jomphe (1989) recognized many of them as subspecies. Morphological variability due to the ability to intercross among members of this species was considered a cause for the confusion and proliferation of names during the long taxonomic history of *M. sativa* (Quiros and Bauchan, 1988).

Our usage of these names is as follows. The principal diploid taxa in this species include *M. sativa* subspecies *caerulea* (Less. Ex Ledeb.) Schmalh. (hereafter referred as "caerulea"), *M. s.* subsp. *falcata* (L.) Arcang. (= *M. falcata* L.; hereafter "diploid falcata"), and *M. s.* subsp. *glomerata* (Balb.) Rouy (= *M. glomerata* Balb.; hereafter "diploid glomerata"). Polyploidization and hybridization were hypothesized as giving rise to other members of this species (Figure 2.1). *Medicago sativa* subsp. *sativa* (hereafter "sativa"), including cultivated alfalfa, is thought to be derived by autopolyploidy from caerulea, and both diploid falcata and diploid glomerata also have morphologically similar tetraploid cytotypes, "tetraploid falcata" and "tetraploid glomerata", respectively (Small, 1985; the latter is also known as *M. glutinosa* M. Bieb.). Diploid *M. s.* subsp. ×*hemicycla* (Grossh.) C. R. Gunn (hereafter "hemicycla") is a putative natural diploid hybrid between caerulea and diploid falcata, and tetraploid *M. s.* subsp. ×*varia* (Martyn) Arcang. (hereafter "varia") is a natural hybrid between sativa and tetraploid falcata

(Lesins and Lesins, 1964; Small and Brookes, 1984; Quiros and Bauchan, 1988). Tetraploid *M. s.* subsp. ×*tunetana* Murb. (hereafter "tunetana") is thought to be either a tetraploidized hybrid of caerulea and diploid glomerata or a hybrid of sativa and tetraploid glomerata (Lesins and Lesins, 1979). Small (1986a) considered tunetana and many other names, e.g., *M. s.* subsp. *faurei* Maire and *M. polychroa* Grossh., as synonyms for "*M. s.* subsp. *sativa* × subsp. *glomerata*."

Closely related to *M. sativa* is *M. prostrata* Jacq., which is found in diploid and tetraploid cytotypes (Lesins and Lesins, 1960; Quiros and Bauchan, 1988). *Medicago prostrata* can be hybridized with members of *M. sativa* successfully when it is used as the paternal parent (Lesins, 1962, 1968) and plants appearing to be natural hybrids of falcata and *M. prostrata* were observed in Italy (Lesins and Lesins, 1979). The ability to hybridize and the occurrence of natural hybrids between *M. prostrata* and members of *M. sativa* suggest that *M. prostrata* could have been involved in the evolutionary history of *M. sativa*. Although boundaries of *M. sativa* within *Medicago* seem identifiable in taxonomic studies of the genus based on morphology (e.g., Lesins and Lesins, 1979; Small and Jomphe, 1989), they often have been ambiguous in molecular systematic studies. For example, phylogenies from different molecular analyses suggest that other perennial *Medicago* species, such as *M. cretacea* M. Bieb. and *M. saxatilis* M. Bieb., may have closer relationships with members of *M. sativa* than suggested by taxonomy (Bena et al., 1998; Bena, 2001; Steele et al., 2010).

Hypothesized relationships among members of *M. sativa* presented by Lesins and Lesins (1979) and Quiros and Bauchan (1988) (Figure 2.1) were based on morphological, cytological, and biochemical evidence. Although there have been a number of studies using various kinds of molecular data to evaluate genetic variation among members in this species, most have focused on cultivated alfalfa (sativa) and included few accessions from other taxa (e.g., Kidwell et al.,

1994; Crochemore et al., 1996; Diwan et al., 1997; Segovia-Lerma et al., 2003), and few have addressed relationships involving its wild members (Brummer et al., 1991; Skinner, 2000; Muller et al., 2003, 2006). Most recently, the relationships of a wide range of wild diploid members of *M. sativa* were investigated by Havananda et al. (2010), using cpDNA and mitochondrial DNA (mtDNA), and by Şakiroğlu et al. (2010), using simple sequence repeat (SSR) markers to estimate genetic diversity and infer population structure. Both of these studies showed that blue-flowered and yellow-flowered diploids in this species (caerulea and diploid falcata) are genetically differentiated, though this had not been detected in some previous studies (e.g., Skinner, 2000; Muller et al., 2006). No molecular studies have directly addressed the origins of tetraploids in this species.

Medicago sativa is relatively unusual among flowering plants in having predominantly paternal transmission of cpDNA (Schumann and Hancock, 1989; Smith, 1989; Masoud et al., 1990). Here we follow up our finding of chloroplast haplotype differentiation among the blue- and yellow-flowered diploid members of this species (Havananda et al., 2010). We explore genetic diversity among tetraploid members of *M. sativa* using the same cpDNA regions to elucidate origins of tetraploids and to understand patterns of gene flow within and between ploidal levels. We sample a number of diploid and tetraploid individuals, including hybrids and other *Medicago* species, and reconstruct relationships among cpDNA haplotypes of these samples. We find unexpectedly different evolutionary histories of the two main autopolyploids and discuss the implications. Our chloroplast data reinforce the growing consensus that autopolyploid evolution is not as straightforward as it may seem.

TABLE 2.1. Principal characters used in distinguishing members of *Medicago sativa* (Lesins and Lesins, 1979; Quiros and Bauchan, 1988; Small and Jomphe, 1989) (Modified from Havananda et al., 2010)

Taxon	Ploidy	Flower color	Pod shape	Pod pubescence
M. sativa subsp. *caerulea*	Diploid	Violet	Coiled	Simple or glabrous
M. sativa subsp. *sativa*	Tetraploid	Violet	Coiled	Simple or glabrous
M. sativa subsp. *falcata*	Di-, tetraploid	Yellow	Straight or sickle	Simple or glandular
M. sativa subsp. *glomerata*	Di-, tetraploid	Yellow	Coiled	Glandular
M. sativa subsp. ×*hemicycla*	Diploid	Segregated or variegated	Curved or loosely coiled	Simple
M. sativa subsp. ×*varia*	Tetraploid	Segregated or variegated	Curved or loosely coiled	Simple
M. sativa subsp. ×*tunetana*	Tetraploid	Segregated or variegated	Coiled	Glandular

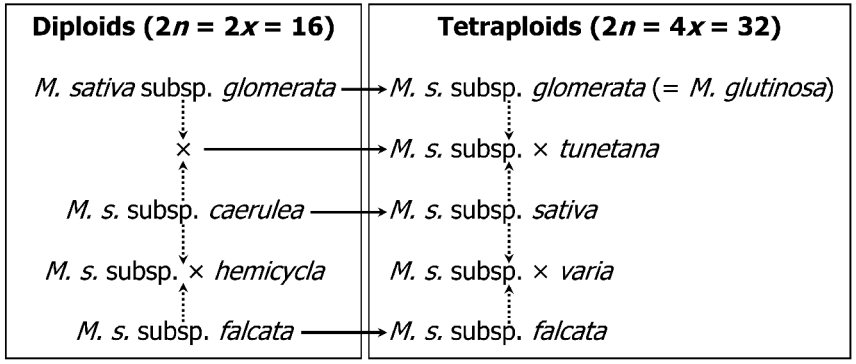


FIGURE 2.1. Hypothesized relationships among taxa in *Medicago sativa* (Lesins and Lesins, 1979; Quiros and Bauchan, 1988; Small and Jomphe, 1989), modified from Havananda et al. (2010). Arrows with solid lines denote autopolyploidy, and those with dashed lines denote hybridization.

Materials and Methods

Medicago accessions from diverse geographical areas were selected from the *Medicago* species collection maintained by the USDA National Plant Germplasm System and listed in the Germplasm Resources Information Network (GRIN; http://www.ars-grin.gov/cgi-bin/npgs/html/splist.pl?7365) and from the South Australian Research and Development Institute (SARDI), Australia (one accession of glomerata). In addition, two genotypes from a synthetic population SD201 developed from wild and cultivated diploid falcata genotypes (Boe et al., 1998), one genotype from WISFAL germplasm (PI 560333, an artificially synthesized tetraploid falcata converted from diploid through a series of 2*x*–4*x* backcrosses; Bingham, 1990, 1993), and one genotype from the artificially diploidized sativa germplasm CADL ("cultivated alfalfa at the diploid level"; Bingham and McCoy, 1979) were also included. A total of 131 individuals from these accessions, comprising 121 individuals (48 diploids and 73 tetraploids) from seven taxa in *M. sativa* and 10 individuals from other *Medicago* species, were included in this study (Appendix 2.1). One hundred and twelve individuals were from wild accessions, 13 individuals from cultivated, and six individuals from accessions with uncertain improvement status. Diploid individuals analyzed for cpDNA and mtDNA variation in Havananda et al. (2010), including 43 individuals from *M. sativa*, four *M. prostrata*, and one *M. truncatula* as an outgroup, were included in this study. All plants were grown in greenhouses at Iowa State University and/or the University of Georgia (by E. C. Brummer between 2003 and 2008) or at Cornell University (by T. Havananda between 2006 and 2010).

Ploidy of most accessions was confirmed by flow cytometry (Brummer et al., 1999; Şakiroğlu et al., 2010; Şakiroğlu and Brummer, 2011; see Appendix 2.2). The flow cytometry method used by Brummer et al. (1999) was also used to determine the ploidy of many other

accessions (E. C. Brummer, Samuel Roberts Noble Foundation, unpublished data; M. Şakiroğlu, Kafkas University, Turkey, unpublished data): young fully expanded leaves were prepared following Galbraith et al. (1983) using their "chopping buffer" (pH 7.0; 45 mmol/L magnesium chloride, 30 mmol/L sodium citrate, 20 mmol/L 4-morpholinepropane sulfonate, and Triton X-100 [1 mg/mL]), but using propidium iodine (100 µg/mL in chopping buffer) to stain nuclei. Samples were run on a flow cytometer (Cytomics FC 500; Beckman-Coulter, Fullerton, California, USA) at the University of Georgia Flow Cytometry Facility; plants with known ploidy, tetraploid sativa (ABI 408) and diploid sativa '3W,' were used as standards giving single peaks on the output DNA histogram; ploidy of the samples was determined by comparing their DNA histogram peaks with the standard peaks. Flow cytometry for all glomerata accessions (T. Havananda, unpublished data) was conducted using methods described by Straub (2010): fresh leaves were finely sliced in 1 mL of cold buffer (10 mmol/L $MgSO_4$, 50 mmol/L KCl, 5 mmol/L HEPES, 0.099% dithiothreitol, 0.247% Triton X-100); 10 µL of propidium iodine (5 mg/mL) and 5 µL RNase (10 mg/mL) were used to treat nuclei in each sample and incubate for at least 15 min; a trout erythrocyte nuclei (TEN) cytometry control (BioSure, Grass Valley, California, USA) was used as internal size standard; these samples were run on a flow cytometer (FACSCalibur, BD Biosciences, San Jose, California, USA) at the Cornell University College of Veterinary Medicine; sample 2C DNA content was calculated using the formula provided in Doležel and Bartoš (2005), and 2C values estimated for *Medicago* by Blondon et al. (1994) were used as reference for ploidy determination of each sample. Ploidy and ploidy assessment reference of each *M. sativa* accession are listed in Appendix 2.2. Individuals from accessions identified as caerulea or hemicycla in GRIN but found to be tetraploids by flow cytometry are considered here as sativa or varia, respectively (noted in Figure 2.3, Appendices 2.1 and 2.2).

Some tetraploid plants grown from seeds obtained from GRIN had flower color that did not correspond to the taxon in which they were classified; e.g., an individual from sativa accession PI 499661 had yellow flowers instead of the expected purple flowers (noted in Figure 2.3 and Appendix 2.1). These ambiguous plants, though included in phylogenetic and network analyses, were not included in DNA variation and genetic differentiation analyses.

Total genomic DNA was extracted using a CTAB protocol (Doyle and Doyle, 1987). PCR amplification and DNA sequencing of two chloroplast DNA regions (*rpl20-rps12* spacer and *trnS-trnG* spacer) were performed using primers and protocols described previously (Havananda et al., 2010). Both strands of each region were sequenced. The program Sequencher version 4.2 or later (Gene Codes, Ann Arbor, Michigan, USA) was used to edit, assemble sequences from both strands, and align sequences across all samples. Alignments were then adjusted manually. Insertions/deletions (indels) of variable length (1–4 bp) in a mononucleotide A repeat found in the *trnS-trnG* spacer were excluded. Sequences of the two chloroplast regions were combined to maximize phylogenetic signal. The simple indel coding (Simmons and Ochoterena, 2000) option in the program SeqState version 1.37 (Müller, 2005) was used to code indels in the alignment. Binary characters of coded indels were added to the alignment for phylogenetic and network analyses. Sequences with unique combinations of cpDNA polymorphisms were designated as unique haplotypes. The alignment used in this study is available in Appendices S1 and S2 (see Supplemental Data with the online version of this article[*]).

Haplotype relationships were reconstructed with maximum parsimony (MP) and maximum likelihood (ML) phylogenetic methods, as well as two network approaches,

[*] Published in Havananda, T., E. C. Brummer, and J. J. Doyle. 2011. Complex patterns of autopolyploid evolution in alfalfa and allies (*Medicago sativa*; Leguminosae). *American Journal of Botany* 98: 1633–1646.

NeighborNet (NNet) and statistical parsimony. For MP, ML, and NNet analyses, OTUs were the 34 haplotypes identified by sequencing. The program PAUP* 4.0b10 (Swofford, 2003) was used for MP analysis using heuristic searches with 100 replicates of random stepwise additions holding 10 trees at each step, tree-bisection-reconnection (TBR) branch swapping, and a maximum of 100,000 trees saved. Bootstrapping was performed with 100 replications using full heuristic search with TBR and Multrees in effect. The ML analysis was performed using the program PhyML 3.0 online web server (http://www.atgc-montpellier.fr/phyml/; Guindon and Gascuel, 2003), with the general time reversible (GTR) DNA substitution model. Branch support was estimated by bootstrap analysis with 1,000 replicates. NNet analysis was implemented in the program SplitsTree4 version 4.10 (Huson and Bryant, 2006) using UncorrectedP distance. Support for splits was estimated by 1,000 bootstrap replicates. Evidence for recombination was tested using the pairwise homoplasy index (PHI) test (Bruen et al., 2006) also implemented in SplitsTree4. Statistical parsimony network analysis was performed with the program TCS version 1.21 (Clement et al., 2000) using all sequences to allow the calculation of haplotype frequencies. To measure cpDNA variation within each taxon and for all *M. sativa* accessions, the program DnaSP version 5 (Librado and Rozas, 2009) was used to calculate number of haplotypes (H), haplotype diversity (H_d), average number of nucleotide differences per site between two sequences (nucleotide diversity, π), and the proportion of segregating polymorphic sites per nucleotide (the Watterson estimator, θ_w). Genetic differentiation between diploid progenitors and their tetraploid descendants, and among tetraploid taxa was estimated based on partitioning of cpDNA variation by the analysis of molecular variance (AMOVA) performed in the program Arlequin version 3.11 (Excoffier et al., 2005).

Results

In addition to 51 sequences from 48 diploid individuals obtained by Havananda et al. (2010), we obtained 84 combined sequences of *rpl20-rpl20* and *trnS-trnG* chloroplast intergenic spacers from 78 individuals of *M. sativa* (5 diploids and 73 tetraploids) and five individuals of other *Medicago* species (Appendix 2.1). Nine tetraploid individuals, three each from tetraploid falcata, glomerata, and sativa, were found to be taxonomically ambiguous because their flower color did not match the description of the taxon to which they had been assigned in GRIN. The number of individuals from each of these three taxa mentioned hereafter reflects the exclusion of these ambiguous individuals.

Heteroplasmy was detected in one tetraploid individual (sativa PI 577496.2) in addition to three diploid individuals previously reported by Havananda et al. (2010); no individuals from hybrid taxa (e.g., varia) showed heteroplasmy, despite the potential for biparental chloroplast transmission in *M. sativa* (Smith et al., 1986; Lee et al., 1988). Polymorphisms found in the heteroplasmic tetraploid individual were a single nucleotide substitution in the *rpl20-rps12* spacer and a 3-bp indel in a mononucleotide A repeat in the *trnS-trnG* spacer. Because indels of variable length in this mononucleotide repeat were excluded, there was effectively only one *trnS-trnG* sequence for this individual, thus two haplotypes were inferred in this individual based on the single nucleotide polymorphism (SNP) in the *rpl20-rps12* spacer. The two haplotypes from heteroplasmic individuals were designated "a" and "b" (Figure 2.3; Appendix 2.3[*]). The lengths of the two sequenced regions after alignment were 712 bp and 319 bp, including six and three added gaps, for *rpl20-rps12* and *trnS-trnG*, respectively. The combined and indel-coded

[*] This appendix was published as Supplemental Data ("Appendix S3") with the online version of Havananda, T., E. C. Brummer, and J. J. Doyle. 2011. Complex patterns of autopolyploid evolution in alfalfa and allies (*Medicago sativa*; Leguminosae). *American Journal of Botany* 98: 1633–1646. Written authorization for the use of the material in this dissertation was obtained from *American Journal of Botany*.

sequences of 1,031 bp from the two regions included 19 parsimony informative and 24 autapomophic sites. These variable sites were composed of 34 nucleotide substitutions and 9 indels (19 and 5, when *M. truncatula* was excluded). Thirty-four haplotypes were identified of which 20 were singleton haplotypes, including the outgroup, *M. truncatula*. Measures of genetic diversity are summarized in Table 2.2.

Phylogenetic and network analyses identify three haplotype groups—The 100,000 maxtrees limit was reached and the parsimony search was terminated due to limited storage capacity. These trees had a length of 66 steps, a consistency index of 0.667, and a retention index of 0.807. The strict consensus tree was mostly unresolved (tree not shown). The ML analysis identified a tree with three main clades, two of which had >50% bootstrap support (Appendix 2.3). The NNet analysis identified the same three main groups as the ML analysis (Figure 2.2). These same three major groups could also be seen in the haplotype network estimated by statistical parsimony analysis, each separated by two or three unobserved haplotypes (Figure 2.3). *Medicago truncatula* was not connected to the network under the 95% statistical parsimony criterion of TCS and thus is not included in the following discussion. Both network methods indicated character conflict in the data set (Figures 2.2, 2.3). Given the rarity of recombination between different chloroplast genomes, we employed only a single test for recombination (PHI test; Bruen et al., 2006), and the test did not find statistically significant evidence for recombination ($P = 0.319$).

TABLE 2.2. Summary statistics of chloroplast DNA sequence variation for *Medicago* taxa. For each taxon, number of accessions and standard measures of diversity (indels were not considered) are indicated. Only taxa with multiple individuals are presented.

Taxon	No. of individuals [a]	No. of polymorphic sites [b]	Gaps	H [c]	H_d	π	θ_w
M. sativa subsp. *caerulea*	15 (16)	5 (4)	20	7 (1–3, 5, 7, 9, 15)	0.750	0.00148	0.00152
M. sativa subsp. *sativa* [d, e]	20 (21)	11 (4)	33	8 (3–5, 11–13, 18, 27)	0.671	0.00193	0.00313
M. sativa subsp. *falcata* (2*x*)	23 (25)	11 (5)	22	8 (4, 7, 8, 10, 22, [28+29], [27+31], [30+32])	0.833	0.00233	0.00295
M. sativa subsp. *falcata* (4*x*) [d]	22 (22)	11 (8)	33	7 (6, [16+18], 17, 19, 22, 24,31)	0.645	0.00213	0.00309
M. sativa subsp. *glomerata* (4*x*) [d]	4 (4)	0	20	1 (5)	n/a	n/a	n/a
M. sativa subsp. ×*hemicycla*	8 (8)	11 (8)	31	4 (1, 7, 18, 22)	0.750	0.00437	0.00433
M. sativa subsp. ×*varia* [f]	15 (15)	10 (9)	31	6 (5, 7, 13, 18, [21+22], 23)	0.848	0.00435	0.00314
M. sativa subsp. ×*tunetana*	4 (4)	2 (2)	20	2 (5, 13)	0.667	0.00135	0.00110
M. prostrata	4 (4)	1 (1)	31	2 (18, 20)	0.667	0.00068	0.00056
M. papillosa	2 (2)	9 (0)	31	2 (5, 18)	1.000	0.00918	0.00918

Notes: H = number of haplotypes; H_d = haplotype (gene) diversity; π = average number of nucleotide differences per site between two sequences (nucleotide diversity); θ_w = the Watterson estimator, proportion of segregating polymorphic sites per nucleotide ($\theta = 4N\mu$); n/a = not applicable.

[a] Number of sequences used in parentheses.

[b] Number of parsimony informative sites in parentheses.

[c] Haplotype numbers in parentheses corresponding to those in Figures 2.2 and 2.3. Because indels were not considered in this analysis, some separate haplotypes in network analyses (Figures 2.2 and 2.3) were identified as a single haplotypes here and shown in brackets.

[d] Excluding ambiguous individuals with flower color that does not correspond to the description of their named taxa.

[e] Including three individuals from accessions listed as *M. s.* subsp. *caerulea* in GRIN, but found to be tetraploids by flow cytometry.

[f] Including four individuals from accessions listed as *M. s.* subsp. ×*hemicycla* in GRIN, but found to be tetraploids by flow cytometry.

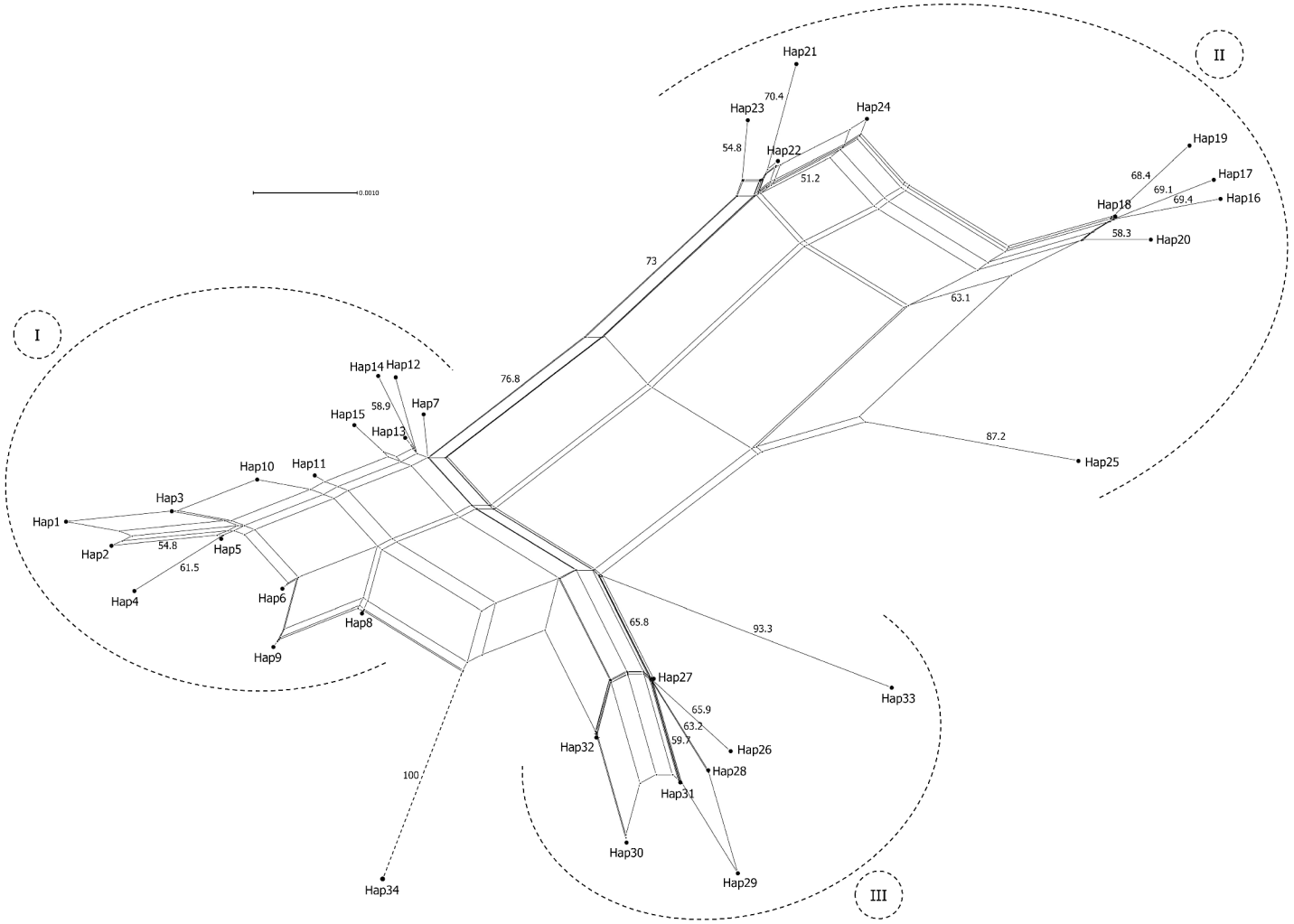
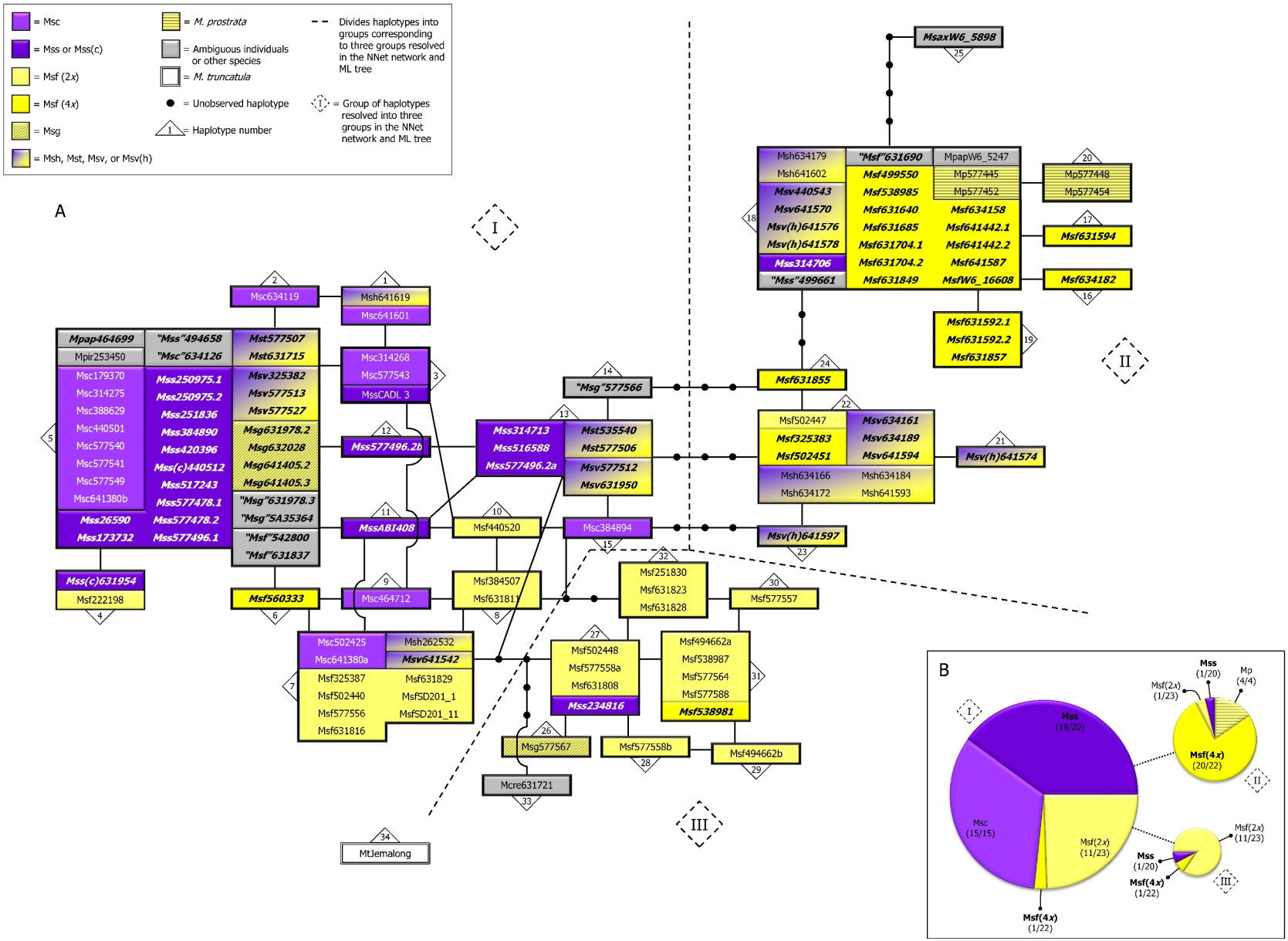


FIGURE 2.2. NeighborNet network for the 34 chloroplast haplotypes identified in diploid and polyploid taxa of *Medicago*. "Hap" followed by a number at each terminal is a haplotype number. Individual(s) possessing each haplotype is listed in Figure 2.3. Curved dashed lines and Roman numerals indicate groups of haplotypes. Numbers next to edges are bootstrap values estimated from 1,000 replicates (only values over 50% are shown). A very long edge connecting *M. truncatula* (Hap34) with the network is shortened and shown as a straight dashed line.

FIGURE 2.3. Chloroplast haplotype statistical parsimony network of diploid and polyploid taxa of *Medicago*. (A) Full network. Each box with a wide border represents a haplotype and contains one or more individuals that possess that haplotype. Taxa sharing a haplotype are separated by thin lines. Tetraploids and hexaploid (*M. saxatilis*) are in bold, italic characters. Taxonomically ambiguous individuals' names are in quotation marks. Different individuals from the same accession are denoted with decimal number; e.g., Msf631592.1 is individual number one from falcata accession PI 631592. Two haplotypes found in heteroplasmic individuals are designated with "a" or "b" at the end of the accession or individual number. The numbers in triangles are haplotype numbers corresponding to the terminals of the NeighborNet network (Figure 2.2). Dashed lines and Roman numerals indicate groups of haplotypes resolved in the NeighborNet network and ML tree (Appendix 2.3). *Medicago truncatula* was not connected to the network under the 95% statistical parsimony criterion of TCS. (B) Summary of A, showing proportions of the key taxa (caerulea, sativa, diploid falcata, tetraploid falcata, and *M. prostrata*) in each group of haplotypes. Each pie chart represents each group of haplotypes connected with dotted lines simplifying relationships among groups. Pie chart sizes are roughly proportional to numbers of individuals among the groups of haplotypes. Slices in the pie charts were calculated from number of individuals of each key taxon per total number of individuals of all key taxa in each group of haplotypes. Number of individuals possessing haplotype(s) in each group per total number of individuals of each taxon, excluding ambiguous individuals, is shown in parentheses under the abbreviated taxon name, e.g., "(20/22)" under **Msf(4*x*)** in group II pie chart means there are 20 of the total 22 tetraploid falcata individuals possessing group II haplotypes. Abbreviations: Mcre, *Medicago cretacea*; Mp, *M. prostrata*; Mpap, *M. papillosa*; Mpir, *M. pironae*; Msax, *M. saxatilis*; Msc, *M. sativa* subsp. *caerulea* (caerulea); Msf, *M. s.* subsp. *falcata* (falcata); Msg, *M. s.* subsp. *glomerata* (glomerata); Msh, *M. s.* subsp. ×*hemicycla* (hemicycla); Mss, *M. s.* subsp. *sativa* (sativa); Mss(c), *M. s.* subsp. *sativa* originally listed as *M. s.* subsp. *caerulea* in GRIN, but found to be tetraploids by flow cytometry; Msv, *M. s.* subsp. ×*varia* (varia); Msv(h), *M. s.* subsp. ×*varia* originally listed as *M. s.* subsp. ×*hemicycla* in GRIN, but found to be tetraploids by flow cytometry; Mst, *M. s.* subsp. ×*tunetana* (tunetana); Mt, *M. truncatula*; names are followed by an accession number (USDA Plant Introduction [PI] number or SARDI [SA] number), genotype, or variety name. CADL = cultivated alfalfa at the diploid level.


= Msc
= Mss or Mss(c)
= Msf (2x)
= Msf (4x)
= Msg
= Msh, Mst, Msv, or Msv(h)
= M. prostrata
= Ambiguous individuals or other species
= M. truncatula
= Unobserved haplotype
= Haplotype number
Divides haplotypes into groups corresponding to three groups resolved in the NNet network and ML tree
= Group of haplotypes resolved into three groups in the NNet network and ML tree
A
I
II
III
MsaxW6_5898
Msh634179
Msh641602
"Msf"631690
Msf499550
MpapW6_5247
Mp577445
Mp577452
Mp577448
Mp577454
Msv440543
Msv641570
Msv(h)641576
Msv(h)641578
Mss314706
"Mss"499661
Msf538985
Msf631640
Msf631685
Msf631704.1
Msf631704.2
Msf631849
Msf634158
Msf641442.1
Msf641442.2
Msf641587
MsfW6_16608
Msf631594
Msf634182
Msf631592.1
Msf631592.2
Msf631857
Msf631855
Msc634119
Msh641619
Msc641601
Mpap464699
Mpir253450
"Mss"494658
"Msc"634126
Mst577507
Mst631715
Msc179370
Msc314275
Msc388629
Msc440501
Msc577540
Msc577541
Msc577549
Msc641380b
Mss250975.1
Mss250975.2
Mss251836
Mss384890
Mss420396
Mss(c)440512
Mss517243
Mss577478.1
Mss577478.2
Mss577496.1
Msv325382
Msv577513
Msv577527
Msg631978.2
Msg632028
Msg641405.2
Msg641405.3
"Msg"631978.3
"Msg"SA35364
"Msf"542800
"Msf"631837
Mss26590
Mss173732
Mss(c)631954
Msf222198
Msf560333
Msc314268
Msc577543
MssCADL 3
Mss577496.2b
MssABI408
Msc464712
"Msg"577566
Mss314713
Mss516588
Mss577496.2a
Mst535540
Mst577506
Msv577512
Msv631950
Msf440520
Msc384894
Msf384507
Msf631811
Msf502447
Msf325383
Msf502451
Msv634161
Msv634189
Msv641594
Msv(h)641574
Msh634166
Msh634184
Msh634172
Msh641593
Msv(h)641597
Msf251830
Msf631823
Msf631828
Msf577557
Msc502425
Msc641380a
Msh262532
Msv641542
Msf325387
Msf502440
Msf577556
Msf631816
Msf631829
MsfSD201_1
MsfSD201_11
Msf502448
Msf577558a
Msf631808
Mss234816
Msf494662a
Msf538987
Msf577564
Msf577588
Msf538981
Msg577567
Msf577558b
Msf494662b
Mcre631721
MtJemalong
B
Mss
(18/20)
Msc
(15/15)
Msf(2x)
(11/23)
Msf(4x)
(1/22)
Msf(4x)
(20/22)
Mss
(1/20)
Mp
(4/4)
Msf(2x)
(1/23)
Msf(2x)
(11/23)
Mss
(1/20)
Msf(4x)
(1/22)

The three groups contained 15, 10, and 8 haplotypes, respectively; all three included haplotypes sampled from polyploids (Figure 2.3). Group I haplotypes were found predominantly in caerulea individuals at the diploid level and sativa individuals at the polyploid level. Group III was dominated by haplotypes from diploid falcata individuals; few tetraploids had group III haplotypes. Group II haplotypes were found mostly in polyploids, primarily in tetraploid falcata individuals; all of the diploid *M. prostrata* individuals had haplotypes from this group. Each of the three groups included haplotypes found in other *Medicago* species (discussed later).

Chloroplast evidence for origins of* M. sativa *polyploids—All caerulea and most sativa individuals had group I haplotypes, with most individuals of both taxa sharing haplotype 5 (Figure 2.3). Accordingly, AMOVA results indicated that there is no significant differentiation between the two taxa for chloroplast haplotypes (Table 2.3), as expected for an autopolyploid origin of sativa from caerulea.

In contrast, nearly mutually exclusive sets of haplotypes were found in diploid and tetraploid falcata, belonging to groups III and II, respectively (Figure 2.3). Only a single diploid falcata individual had a group II haplotype, and only a single tetraploid falcata individual had a group III haplotype. AMOVA results showed that a large proportion of the total variation in "falcata" (64.20%, $P < 0.0001$) was explained by differences between diploids and tetraploids (Table 2.3). The predominant nonhybrid diploid taxon with group II haplotypes was not diploid falcata, but rather *M. prostrata*, and the most common group II haplotype was shared by that diploid species and tetraploid falcata (Figure 2.3). AMOVA showed no significant differentiation between *M. prostrata* and tetraploid falcata.

The four individuals verified as representing the remaining nonhybrid tetraploid member of *M. sativa*, glomerata, were all found to have the common group I haplotype 5, in contrast to

the only diploid individual sampled, which had group III haplotype 26. Thus for glomerata, as for falcata, diploids and tetraploids are strongly differentiated (Table 2.3), though with such a small sample it is difficult to make robust conclusions.

TABLE 2.3. Analysis of molecular variance (AMOVA) of genetic variation in the two chloroplast regions for diploid and polyploid taxa of *Medicago*.

Source of variation	df	% of Total variation	Fixation index (Φ_{ST})	*P*
Caerulea vs. diploid falcata				
Among taxa	1	43.44	0.434 [a]	<0.0001
Within taxa	39	56.56		
Caerulea vs. sativa [b]				
Among taxa	1	4.61	0.046	>0.05
Within taxa	35	95.39		
Diploid falcata vs. tetraploid falcata [b]				
Among taxa	1	64.20	0.642	<0.0001
Within taxa	45	35.80		
Tetraploid falcata [b] vs. sativa [b]				
Among taxa	1	73.48	0.735	<0.0001
Within taxa	41	26.52		
M. prostrata vs. caerulea				
Among taxa	1	85.78	0.858 [a]	<0.0005
Within taxa	18	14.22		
Diploid glomerata vs. tetraploid glomerata [b]				
Among taxa	1	100.00	1.000	>0.10
Within taxa	3	0		
M. prostrata vs. sativa [b]				
Among taxa	1	80.53	0.805	<0.0005
Within taxa	23	19.47		
M. prostrata vs. diploid falcata				
Among taxa	1	69.65	0.696 [a]	<0.0005
Within taxa	27	30.35		
M. prostrata vs tetraploid falcata [b]				
Among taxa	1	5.51	0.055	>0.10
Within taxa	24	94.49		

[a] Previously reported in Havananda et al. (2010).

[b] Ambiguous individuals with flower color that does not correspond to the description of their named taxa were excluded.

Our larger sample of haplotypes confirms the finding of Havananda et al. (2010) that caerulea and diploid falcata are differentiated from one another (Table 2.3). As noted in that paper, the picture is complex, with diploid falcata being very polymorphic: most diploid falcata individuals had haplotypes belonging to group III; several had haplotypes from group I, which is dominated by caerulea, with the two taxa sharing one haplotype (group I haplotype 7; Figure 2.3); there was also one diploid falcata individual with a group II haplotype. Our single sample of the other nonhybrid diploid taxon of *M. sativa*, glomerata, had a group III haplotype related to diploid falcata. *Medicago papillosa* and *M. pironae* both had the most common group I haplotype (haplotype 5) but are morphologically distinct from caerulea. *Medicago papillosa* also shared the most common group II haplotype (haplotype 18) with *M. prostrata*. The final *Medicago* species sampled here, *M. cretacea* and *M. saxatilis* (hexaploid), had divergent haplotypes that were most closely related to group III and group II, respectively.

Hybrid taxa and implications for gene flow at the tetraploid level—Sativa and tetraploid falcata had mutually exclusive sets of haplotypes, with the exception of a single sativa individual with the common tetraploid falcata group II haplotype 18 (an individual from accession PI 560333 with group I haplotype 6 is "WISFAL," a synthetic tetraploid falcata produced through a series of 2*x*–4*x* backcrossing using falcata-sativa tetraploid hybrid as a nonrecurrrent male parent; Bingham, 1990, 1993). Thus it appears that there has been little gene flow between sativa and tetraploid falcata.

Chloroplast haplotypes from plants with intermediate morphologies were represented in groups I and II, but not group III, which was unexpected given the potentially biparental transmission of cpDNA in *M. sativa* (Smith et al., 1986; Lee et al., 1988). Haplotypes from

individuals of the diploid hybrid, hemicycla, were found primarily in group II (six individuals), with two individuals having group I haplotypes.

Tetraploid hybrids between plants with blue flowers (sativa) and yellow flowers (tetraploid falcata) are classified as varia (Figure 2.1). All such plants in our sample had either group I haplotypes, consistent with sativa as the plastid donor, or group II haplotypes, consistent with tetraploid falcata as the plastid donor. AMOVA results (Table 2.3) confirm the differentiation of sativa and tetraploid falcata.

Another tetraploid hybrid, tunetana, is hypothesized to be a tetraploidized hybrid of caerulea and diploid glomerata, or a hybrid of sativa and tetraploid glomerata (Figure 2.1). Sampled tunetana individuals shared group I haplotypes (haplotypes 5 and 13) with caerulea/sativa and tetraploid glomerata, but did not share group III haplotypes with diploid glomerata, supporting the hybrid origin of tunetana at the tetraploid level. However, with only a single sample of diploid glomerata, the alternative hypothesis cannot be ruled out.

Discussion

Different evolutionary trajectories of two closely related autopolyploids—Evolution in allopolyploids involves both genome merger and genome duplication, and there is much debate about which contributes most to the dynamic nature of allopolyploids (e.g., Doyle et al., 2008). In contrast, autopolyploidy mostly involves genome duplication; although hybridization in the plant breeding sense of combining different genotypes also may occur, heterogeneity is not immediately fixed by disomic segregation. In both the genetic and taxonomic senses of autopolyploidy, the expectation is that the doubled genome of an autopolyploid has been derived from a more similar genetic background—possibly from a single progenitor genotype—than that

of an allopolyploid, whose progenitors need not even be closely related species. In terms of the genetic differentiation of their progenitor genome(s), autopolyploids share some attributes with both allopolyploids and homoploid hybrids. Like the former, they have doubled genomes, but like the latter, their formation seems to be limited to genetically similar parents (Buggs et al., 2009), and genetic variation can be lost by segregation.

Because allopolyploids are fundamentally hybrids, it is expected that different allopolyploids can follow very different evolutionary trajectories (Mallet, 2007). Even multiple origins of the same allopolyploid might show this effect, in the absence of lineage recombination (Doyle et al., 1999), because each origin can bring together different combinations of alleles at each homoeologous locus, leading to different transgressive effects. In contrast, one might expect evolutionary outcomes of autopolyploids to track more closely their single diploid progenitor species. This should be particularly true of autopolyploids formed independently within the same species. This is not only because the genetic base is potentially narrower to begin with, but also because polysomic inheritance leads to the segregational loss of initial allelic variation. The combination of these phenomena may help explain why autopolyploids are often difficult to detect in nature (Soltis et al., 2010).

Medicago sativa provides a good system in which to test such expectations. There is morphological and genetic variation at the diploid level, which is partitioned into differentiated but interfertile taxa, notably blue-flowered caerulea and yellow-flowered diploid falcata. Each has been hypothesized to give rise to an autopolyploid: sativa (which includes cultivated alfalfa) from caerulea (Stanford, 1951; Quiros, 1982; Small and Jomphe, 1989) and tetraploid falcata from diploid falcata (Lesins and Lesins, 1979). We can ask whether these parallel autopolyploid "experiments" meet the expectation of yielding similar results.

Patterns of chloroplast genome variation suggest that the answer is "no." Sativa shows what might be called the "classic" pattern for an autopolyploid. It differs from its presumed diploid progenitor, caerulea, only in subtle characters, mostly involving size and robustness (Lesins and Lesins, 1979; Small, 1985; Quiros and Bauchan, 1988; Small and Jomphe, 1989). Our results show that the two taxa have very closely related chloroplast haplotypes, with most of their individuals share the most common haplotype, and are undifferentiated genetically for this characteristic. The presence of several loops in the group I network make it difficult to estimate how many of the remaining four haplotypes found in sativa, but not in caerulea, were derived directly from caerulea, either as independent origins or by subsequent gene flow involving unreduced caerulea gametes.

The surprise is the history of tetraploid falcata as suggested by our cpDNA results. Despite a long tradition of taxonomic splitting in *M. sativa* that led to names being given to very minor morphological variants, only a single subspecies, *M. sativa* subsp. *falcata*, has been recognized on the basis of morphology, and only by counting chromosomes was it recognized that diploid and tetraploid cytotypes existed. The expectation was that these cytotypes should be as undifferentiated genetically as they are morphologically. Yet in the tetraploid, chloroplast haplotypes predominate that are more closely related to haplotypes from a species outside *M. sativa*, *M. prostrata*, and are almost completely lacking in our relatively broad sample of diploid falcata.

The most likely origin of tetraploid falcata is presumably still autopolyploidy from diploid falcata, given that the two cytotypes are morphologically identical, and this hypothesis is supported by preliminary data from a low copy nuclear gene (T. Havananda, E. C. Brummer, and J. J. Doyle, unpublished data). The presence of prostrata plastid genomes in tetraploid falcata

could be due to lineage sorting at either the diploid or tetraploid level. Alternatively, hybridization and introgression could be responsible. These phenomena are thought to be common in *Medicago* as a whole (Maureira-Butler et al., 2008) and could also produce other unexpected results such as the pattern of divergent haplotypes seen in *M. papillosa*. Although our *M. prostrata* samples were all diploid, the species also has tetraploid cytotypes (Lesins and Lesins, 1960), so introgression at the tetraploid level between *M. prostrata* and falcata is possible. Introgression from *M. prostrata* into tetraploid falcata, one or more times, could result in the replacement of the diploid falcata chloroplast genome by that of *M. prostrata*. Introgression seems likely given: (1) the ability of *M. prostrata* as a male parent to hybridize with diploid falcata (Lesins, 1962), (2) the largely paternal inheritance of cpDNA in *Medicago* (Schumann and Hancock, 1989; Smith, 1989; Masoud et al., 1990), and (3) the existence of unreduced gametes in *M. prostrata*, making interploidal crosses possible (Sorensen et al., 1980). If introgression is responsible for this pattern, nuclear genes controlling key morphological characters apparently have not been introgressed along with the chloroplast.

There appears to have been little if any gene flow between the two falcata cytotypes subsequent to the incorporation of the *M. prostrata* chloroplast genome into tetraploid falcata, at least not with typical diploid falcata as the pollen parent. In the case of sativa and caerulea, as in many polyploid/diploid pairs, the presence of multiple haplotypes in the tetraploid that are identical or closely related to haplotypes in the diploid is consistent with either multiple origins of the polyploid or a single origin and subsequent gene flow. But in tetraploid falcata, if there were multiple origins from diploid falcata, much of the evidence has been erased for the chloroplast genome, by the replacement of diploid falcata chloroplast haplotypes with haplotypes typical of *M. prostrata*. It is possible, however, that the sharing of group II haplotype 22 by two

tetraploid falcata and one diploid falcata individuals, and the sharing of group III haplotype 31 by one tetraploid falcata and several diploid falcata individuals could suggest additional origins of tetraploid falcata from diploid falcata.

It is possible that the progenitor of tetraploid falcata was a diploid falcata genotype with a chloroplast haplotype typical of *M. prostrata*. This could be due to either introgression of the chloroplast genome of *M. prostrata* into diploid falcata, or retention of a shared chloroplast polymorphism between the two diploids. Lesins (1962) showed that *M. prostrata* as pollen parent can hybridize with diploid falcata. Our sample of diploid falcata includes no individuals with this *M. prostrata* haplotype, though one individual has a group II haplotype found in tetraploid falcata. Regardless of how this hypothesized progenitor came to have the *M. prostrata* haplotype, we arrive at the same conclusion of low subsequent interploidy gene flow, because such gene flow, at least if it involved known common genotypes of diploid falcata as pollen donors, would be expected to replace the *M. prostrata* haplotype with haplotypes typical of diploid falcata haplotypes.

Medicago papillosa, which shares a haplotype with *M. prostrata* and tetraploid falcata, could possibly be involved in whatever hybridization or lineage sorting event(s) resulted in diploid and tetraploid falcata having different haplotypes. However, this seems unlikely because, unlike *M. prostrata*, there is no evidence for natural hybridization between *M. papillosa* and members of *M. sativa* (Lesins and Lesins, 1979). Hybrid progeny from artificial crosses can only be recovered by ovule-embryo culture (McCoy and Smith, 1986; McCoy and Quarisa, 1989).

Thus, within a single recognized species, *M. sativa*, we have different outcomes involving autopolyploidy. The “classic” autopolyploid pattern, possibly involving multiple origins and/or interploidy gene flow in caerulea/sativa, has led to shared chloroplast haplotypes

and slight morphological differentiation that is recognized taxonomically, albeit at the subspecific level. This is consistent with the recommendations of Soltis et al. (2007), given the taxonomic complexities at the diploid level in *M. sativa*. Subspecific, rather than specific recognition is also consistent with the possibility that gene flow continues between caerulea and sativa, possibly unidirectionally via unreduced gametes from caerulea.

In striking contrast, the yellow-flowered diploid member of *M. sativa* has given rise to an autopolyploid that, although having a morphology indistinguishable from the diploid, has acquired a very different predominant chloroplast haplotype, possibly by hybridization at the tetraploid level with a species outside *M. sativa*, *M. prostrata*. Regardless of the origin of this haplotype, there appears to be little genetic exchange between the two cytotypes of falcata. The apparent lack of gene flow and differentiation at the chloroplast genome level would argue for formal taxonomic distinction between the two cytotypes (Soltis et al., 2007), but whether at the species or subspecies level is debatable and must await further evidence to determine to what degree the nuclear genome of tetraploid falcata is derived from diploid falcata as opposed to *M. prostrata.* Practical considerations also exist, such as determining whether the type for *M. falcata* L. is diploid or tetraploid, given the lack of morphological differentiation between the two cytotypes and their incomplete differentiation even for cpDNA haplotypes.

Evidence for lineage recombination following autopolyploidy—Multiple origins of a polyploid provide it with a means of increasing its genetic variation. But if multiple origins are not followed by gene flow among the different polyploid genotypes, the result is a set of separate polyploid populations each of which initially is genetically depauperate. Gene flow at the polyploid level leads to lineage recombination (Doyle et al., 1999) in which characters are

reassorted such that novel combinations of alleles at multiple loci are observed in different polyploids.

Given the narrower genetic base expected in autopolyploids relative to allopolyploids, it seems reasonable that the potential for reproductive incompatibilities would be more likely in the latter than in the former. Certainly, there is every reason to expect that gene flow, and hence lineage recombination, would be a common phenomenon among independently formed autopolyploids within the same species.

In *M. sativa*, three of the major groups of polyploids are (1) blue-flowered plants mostly with chloroplast group I haplotypes 5 and 13; (2) yellow-flowered plants predominantly with chloroplast group II haplotypes 18 and 19; and (3) variegated- or greenish-flowered plants with diverse group I and group II haplotypes. These three groups correspond to sativa, tetraploid falcata (yellow-flowered tetraploid glomerata was too sparsely sampled to consider here), and varia, respectively. The existence of varia plants is clear evidence of hybridization (either bidirectional or unidirectional with biparental transmission) between sativa and tetraploid falcata, which would be a prerequisite for lineage recombination at the tetraploid level in *M. sativa*.

However, despite the demonstrable ability of these taxa to hybridize both in the greenhouse and in nature, there is little evidence for late generation hybrids at the tetraploid level. In other words, there is little evidence for lineage recombination, or we would have observed numerous individuals with falcata morphologies and caerulea/sativa chloroplast haplotypes and with sativa morphologies and *M. prostrata*/tetraploid falcata haplotypes. There are only two such individuals in our sample, and sativa and tetraploid falcata are strongly differentiated from one another (Table 2.3).

Conflicting evidence for lineage recombination in *M. sativa* involves two other tetraploids, glomerata and tunetana. Our findings support the hypothesized origin of tunetana from hybridization of sativa and tetraploid glomerata and suggest gene flow at the tetraploid level between the two parental tetraploid taxa. The occurrence of several individuals with tetraploid glomerata morphologies and caerulea/sativa chloroplast haplotypes suggests lineage recombination. Here too, *M. sativa* has a complicated history.

***Conclusions*—**UNEXPECTED COMPLEXITY OF AUTOPOLYPLOIDY?—Our studies of *M. sativa* based on chloroplast sequence data suggest that its evolutionary history may be even more complicated than the histories of other well-studied polyploidy complexes, including *Tolmiea menziesii*, *Galax urceolata*, *Chamerion angustifolium*, and *Heuchera grossulariifolia* (Soltis et al., 2007 and references therein). It is always useful to be reminded that our textbook characterizations of nature are often drastically oversimplified. For example, the facile characterization of polyploids as either auto- or allo- may hide a spectrum of genetic and taxonomic possibilities (Soltis et al., 2010). It seems fair to say that autopolyploidy is considered by many a simpler process than allopolyploidy, if only because it may occur within a single species. Yet the more detailed picture of *M. sativa* provided by the chloroplast genome has revealed unexpectedly different evolutionary histories between two very closely related autopolyploid subspecies within the same taxonomic species.

The chloroplast data presented here raise many additional questions and suggest that we have only scratched the surface of the evolutionary complexity of this group. Nuclear gene data should resolve the issue of whether diploid falcata is the direct progenitor of tetraploid falcata, as morphology would suggest, or whether *M. prostrata* played a role in the origin of tetraploid falcata beyond donating a few introgressed genes. Similarly, further sampling of accessions and

nuclear genes will be required to elucidate the evolutionary history of polyploid glomerata accessions—morphologically distinctive plants with yellow flowers like falcata, yet having chloroplast haplotypes typical of blue-flowered caerulea and sativa.

What is clear is that detailed study, even of well-known groups such as alfalfa, can lead to new insights into evolution by autopolyploidy. Such studies reinforce the conclusion that autopolyploid evolution is far from simple.

ACKNOWLEDGEMENTS—The authors thank M. Şakiroğlu for providing information, ploidy, and genomic DNA of some accessions, especially diploids. They acknowledge J. L. Doyle for technical assistance, A. N. Egan for analytical help, and others in the Doyle laboratory and I. J. Maureira-Butler for helpful discussion and comments. The authors also thank D. I. Dickerman for laboratory help and J. L. Wszolek for greenhouse assistance. T.H. was financially supported by the Office of the Higher Education Commission, the Royal Thai Government. This work was supported by NSF DEB-0516673 to J.J.D. and the USDA-DOE Plant Feedstock Genomics for Bioenergy award 2006-35300-17224 to E.C.B. and J.J.D.

REFERENCES

ARMSTRONG, K. C. 1971. Chromosome associations at pachytene and metaphase in *Medicago sativa. Canadian Journal of Genetics and Cytology* 13: 697–702.

BENA, G. 2001. Molecular phylogeny supports the morphologically based taxonomic transfer of the "medicagoid" *Trigonella* species to the genus *Medicago* L. *Plant Systematics and Evolution* 229: 217–236.

BENA, G., B. LEJEUNE, J.-M. PROSPERI, AND I. OLIVIERI. 1998. Molecular phylogenetic approach for studying life-history evolution: The ambiguous example of the genus *Medicago* L. *Proceedings of the Royal Society of London, B, Biological Sciences* 265: 1141–1151.

BINGHAM, E. T. 1968. Transfer of diploid *Medicago* spp. germplasm to tetraploid *M. sativa* L. in 4x–2x crosses. *Crop Science* 8: 760–762.

BINGHAM, E. T. 1990. Backcrossing tetraploidy into diploid *Medicago falcata* L. using 2*n* eggs. *Crop Science* 30: 1353–1354.

BINGHAM, E. T. 1993. Registration of WISFAL alfalfa (*Medicago sativa* subsp. *falcata*) tetraploid germplasm derived from diploids. *Crop Science* 33: 217–218.

BINGHAM, E. T., AND T. J. MCCOY. 1979. Cultivated alfalfa at the diploid level: Origin, reproductive stability, and yield of seed and forage. *Crop Science* 19: 97–100.

BLONDON, F., D. MARIE, S. BROWN, AND A. KONDOROSI. 1994. Genome size and base composition in *Medicago sativa* and *Medicago truncatula* species. *Genome* 37: 264–270.

BOE, A., R. BORTNEM, K. F. HIGGINS, A. D. KRUSE, K. D. KEPHART, AND S. SELMAN. 1998. Breeding yellow-flowered alfalfa for combined wildlife habitat and forage purposes. South Dakota Agricultural Experiment Station Bulletin 727, South Dakota State University, Brookings, South Dakota, USA.

BRUEN, T. C., H. PHILIPPE, AND D. BRYANT. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* 172: 2665–2681.

BRUMMER, E. C., P. M. CAZCARRO, AND D. LUTH. 1999. Ploidy determination of alfalfa germplasm accessions using flow cytometry. *Crop Science* 39: 1202–1207.

BRUMMER, E. C., G. KOCHERT, AND J. H. BOUTON. 1991. RFLP variation in diploid and tetraploid alfalfa. *Theoretical and Applied Genetics* 83: 89–96.

BUGGS, R. J. A., P. S. SOLTIS, AND D. E. SOLTIS. 2009. Does hybridization between divergent progenitors drive whole-genome duplication? *Molecular Ecology* 18: 3334–3339.

CLEMENT, M., D. POSADA, AND K. A. CRANDALL. 2000. TCS: A computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657–1659.

CROCHEMORE, M. L., C. HUYGHE, M. C. KERLAN, F. DURAND, AND B. JULIER. 1996. Partitioning and distribution of RAPD variation in a set of populations of the *Medicago sativa* complex. *Agronomie* 16: 421–432.

DIWAN, N., A. A. BHAGWAT, G. B. BAUCHAN, AND P. B. CREGAN. 1997. Simple sequence repeat DNA markers in alfalfa and perennial and annual *Medicago* species. *Genome* 40: 887–895.

DOLEŽEL, J., AND J. BARTOŠ. 2005. Plant DNA flow cytometry and estimation of nuclear genome size. *Annals of Botany* 95: 99–110.

DOYLE, J. J., AND J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

DOYLE, J. J., J. L. DOYLE, AND A. H. D. BROWN. 1999. Origins, colonization, and lineage recombination in a widespread perennial soybean polyploid complex. *Proceedings of the National Academy of Sciences, USA* 96: 10741–10745.

DOYLE, J. J., AND A. N. EGAN. 2010. Dating the origins of polyploidy events. *New Phytologist* 186: 73–85.

DOYLE, J. J., L. E. FLAGEL, A. H. PATERSON, R. A. RAPP, D. E. SOLTIS, P. S. SOLTIS, AND J. F. WENDEL. 2008. Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics* 42: 443–461.

EXCOFFIER, L., G. LAVAL, AND S. SCHNEIDER. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1: 47–50.

FRAME, J., J. F. L. CHARLTON, AND A. S. LAIDLAW. 1997. Temperate forage legumes. CAB International, Wallingford, UK.

GALBRAITH, D. W., K. R. HARKINS, J. M. MADDOX, N. M. AYRES, D. P. SHARMA, AND E. FIROOZABADY. 1983. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* 220: 1049–1051.

GILLIES, C. B. 1972. Pachytene chromosomes of perennial *Medicago* species. I. Species closely related to *M. sativa. Hereditas* 72: 277–288.

GUINDON, S., AND O. GASCUEL. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* 52: 696–704.

HAVANANDA, T., E. C. BRUMMER, I. J. MAUREIRA-BUTLER, AND J. J. DOYLE. 2010. Relationships among diploid members of the *Medicago sativa* (Fabaceae) species complex based on chloroplast and mitochondrial DNA sequences. *Systematic Botany* 35: 140–150.

HUSON, D. H., AND D. BRYANT. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254–267. [Software available from http://www.splitstree.org/]

JIAO, Y., N. J. WICKETT, S. AYYAMPALAYAM, A. S. CHANDERBALI, L. LANDHERR, P. E. RALPH, L. P. TOMSHO, ET AL. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97–100.

KIDWELL, K. K., D. F. AUSTIN, AND T. C. OSBORN. 1994. RFLP evaluation of nine *Medicago* accessions representing the original germplasm sources for North-American alfalfa cultivars. *Crop Science* 34: 230–236.

LEE, D. J., T. K. BLAKE, AND S. E. SMITH. 1988. Biparental inheritance of chloroplast DNA and the existence of heteroplasmic cells in alfalfa. *Theoretical and Applied Genetics* 76: 545–549.

LESINS, K. 1962. Interspecific crosses involving alfalfa. III. *Medicago sativa* L. × *M. prostrata* Jacq. *Canadian Journal of Genetics and Cytology* 4: 14–23.

LESINS, K. 1968. Interspecific crosses involving alfalfa. IV. *Medicago glomerata* × *M. sativa* with reference to *M. prostrata. Canadian Journal of Genetics and Cytology* 10: 536–544.

LESINS, K., AND I. LESINS. 1960. Sibling species in *Medicago prostrata* Jacq. *Canadian Journal of Genetics and Cytology* 2: 416–417.

LESINS, K., AND I. LESINS. 1964. Diploid *Medicago falcata* L. *Canadian Journal of Genetics and Cytology* 6: 152–163.

LESINS, K. A., AND I. LESINS. 1979. Genus *Medicago* (Leguminosae): A taxogenetic study. Dr. W. Junk, The Hague, Netherlands.

LIBRADO, P., AND J. ROZAS. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.

MALLET, J. 2007. Hybrid speciation. *Nature* 446: 279–283.

MARIANI, A., AND F. VERONESI. 1979. Cytological and fertility relationships of different *Medicago* species and cytogenetic behavior of their hybrids. *Genetica Agraria* 33: 245–268.

MASOUD, S. A., L. B. JOHNSON, AND E. L. SORENSEN. 1990. High transmission of paternal plastid DNA in alfalfa plants demonstrated by restriction fragment polymorphic analysis. *Theoretical and Applied Genetics* 79: 49–55.

MASTERSON, J. 1994. Stomatal size in fossil plants: Evidence for polyploidy in majority of angiosperms. *Science* 264: 421–424.

MAUREIRA-BUTLER, I. J., B. E. PFEIL, A. MUANGPROM, T. C. OSBORN, AND J. J. DOYLE. 2008. The reticulate history of *Medicago* (Fabaceae). *Systematic Biology* 57: 466–482.

MCCOY, T. J. 1982. The inheritance of 2*n* pollen formation in diploid alfalfa *Medicago sativa.* *Canadian Journal of Genetics and Cytology* 24: 315–323.

MCCOY, T. J., AND E. T. BINGHAM. 1988. Cytology and cytogenetics of alfalfa. *In* A. A. Hanson [ed.], D. K. Barnes and R. R. Hill Jr. [co-eds.], Alfalfa and alfalfa improvement, Agronomy monograph 29, 737–776. American Society of Agronomy, Crop Science Society of Agronomy, and Soil Science Society of Agronomy, Madison, Wisconsin, USA.

MCCOY, T. J., AND G. L. QUARISA. 1989. Allotetraploid behavior of hybrids of *Medicago sativa* L. and *M. papillosa* Boiss. *Genome* 32: 6–11.

MCCOY, T. J., AND L. Y. SMITH. 1986. Interspecific hybridization of perennial *Medicago* species using ovule-embryo culture. *Theoretical and Applied Genetics* 71: 772–783.

MICHAUD, R., W. F. LEHMAN, AND M. D. RUMBAUGH. 1988. World distribution and historical development. *In* A. A. Hanson [ed.], D. K. Barnes and R. R. Hill Jr. [co-eds.], Alfalfa and alfalfa improvement, Agronomy monograph 29, 93–124. American Society of Agronomy, Crop Science Society of Agronomy, and Soil Science Society of Agronomy, Madison, Wisconsin, USA.

MÜLLER, K. 2005. SeqState: Primer design and sequence statistics for phylogenetic DNA datasets. *Applied Bioinformatics* 4: 65–69.

MULLER, M.-H., C. PONCET, J. M. PROSPERI, S. SANTONI, AND J. RONFORT. 2006. Domestication history in the *Medicago sativa* species complex: Inferences from nuclear sequence polymorphism. *Molecular Ecology* 15: 1589–1602.

MULLER, M. H., J. M. PROSPERI, S. SANTONI, AND J. RONFORT. 2003. Inferences from mitochondrial DNA patterns on the domestication history of alfalfa (*Medicago sativa*). *Molecular Ecology* 12: 2187–2199.

PARISOD, C., R. HOLDEREGGER, AND C. BROCHMANN. 2010. Evolutionary consequences of autopolyploidy. *New Phytologist* 186: 5–17.

QUIROS, C. F. 1982. Tetrasomic segregation for multiple alleles in alfalfa. *Genetics* 101: 117–127.

QUIROS, C. F., AND G. R. BAUCHAN. 1988. The genus *Medicago* and the origin of the *Medicago sativa* complex. *In* A. A. Hanson [ed.], D. K. Barnes and R. R. Hill Jr. [co-eds.], Alfalfa and alfalfa improvement, Agronomy monograph 29, 93–124. American Society of Agronomy, Crop Science Society of Agronomy, and Soil Science Society of Agronomy, Madison, Wisconsin, USA.

RAMSEY, J., AND D. W. SCHEMSKE. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* 29: 467–501.

RUSSELLE, M. P. 2001. Alfalfa: After an 8,000-year journey, the "Queen of Forages" stands poised to enjoy renewed popularity. *American Scientist* 89: 252–261.

ŞAKIROĞLU, M., AND E. C. BRUMMER. 2011. Clarifying the ploidy of some accessions in the USDA alfalfa germplasm collection. *Turkish Journal of Botany* 35: 509–519.

ŞAKIROĞLU, M., J. J. DOYLE, AND E. C. BRUMMER. 2010. Inferring population structure and genetic diversity of broad range of wild diploid alfalfa (*Medicago sativa* L.) accessions using SSR markers. *Theoretical and Applied Genetics* 121: 403–415.

SCHUMANN, C. M., AND J. F. HANCOCK. 1989. Paternal inheritance of plastids in *Medicago sativa. Theoretical and Applied Genetics* 78: 863–866.

SEGOVIA-LERMA, A., R. G. CANTRELL, J. M. CONWAY, AND I. M. RAY. 2003. AFLP-based assessment of genetic diversity among nine alfalfa germplasms using bulk DNA templates. *Genome* 46: 51–58.

SIMMONS, M. P., AND H. OCHOTERENA. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369–381.

SKINNER, D. Z. 2000. Non random chloroplast DNA hypervariablility in *Medicago sativa. Theoretical and Applied Genetics* 101: 1242–1249.

SMALL, E. 1985. Morphological differentiation in *Medicago sativa* s.l. in relation to ploidy. *Canadian Journal of Botany* 63: 1747–1752.

SMALL, E. 1986a. Taxonomy of glandular wild alfalfa (*Medicago sativa*). *Canadian Journal of Botany* 64: 2125–2129.

SMALL, E. 1986b. A clarification of the *Medicago papillosa* complex. *Canadian Journal of Botany* 64: 2800–2806.

SMALL, E., AND B. S. BROOKES. 1984. Taxonomic circumscription and identification in the *Medicago sativa-falcata* (alfalfa) continuum. *Economic Botany* 38: 83–96.

SMALL, E., AND M. JOMPHE. 1989. A synopsis of the genus *Medicago* (Leguminosae). *Canadian Journal of Botany* 67: 3260–3294.

SMITH, S. E. 1989. Influence of parental genotype on plastid inheritance in *Medicago sativa. Journal of Heredity* 80: 214–217.

SMITH, S. E., E. T. BINGHAM, AND R. W. FULTON. 1986. Transmission of chlorophyll deficiencies in *Medicago sativa*: Evidence for biparental inheritance of plastids. *Journal of Heredity* 77: 35–38.

SOLTIS, D. E., V. A. ALBERT, J. LEEBENS-MACK, C. D. BELL, A. H. PATERSON, C. ZHENG, D. SANKOFF, ET AL. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96: 336–348.

SOLTIS, D. E., R. J. A. BUGGS, J. J. DOYLE, AND P. S. SOLTIS. 2010. What we still don't know about polyploidy. *Taxon* 59: 1387–1403.

SOLTIS, D. E., P. S. SOLTIS, D. W. SCHEMSKE, J. F. HANCOCK, J. N. THOMPSON, B. C. HUSBAND, AND W. S. JUDD. 2007. Autopolyploidy in angiosperms: Have we grossly underestimated the number of species? *Taxon* 56: 13–30.

SOLTIS, D. E., P. S. SOLTIS, AND J. A. TATE. 2004. Advances in the study of polyploidy since *Plant Speciation. New Phytologist* 161: 173–191.

SOLTIS, P. S., AND D. E. SOLTIS. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences, USA* 97: 7051–7057.

SORENSEN, E. L., N. SANGDUEN, AND G. H. LIANG. 1980. Transfer of glandular hairs from diploid *Medicago prostrata* to tetraploid *M. sativa. In* D. K. Barnes [ed.], Report of The Twenty-seventh Alfalfa Improvement Conference, Madison, Wisconsin, 8–10 July, 53. Agricultural Research, North Central Region, USDA-SEA, Peoria, Illinois, USA.

STANFORD, E. H. 1951. Tetrasomic inheritance in alfalfa. *Agronomy Journal* 43: 222–225.

STEELE, K. P., S. M. ICKERT-BOND, S. ZARRE, AND M. F. WOJCIECHOWSKI. 2010. Phylogeny and character evolution in *Medicago* (Leguminosae): Evidence from analyses of plastid *trnK*/*matK* and nuclear *GA3ox1* sequences. *American Journal of Botany* 97: 1142–1155.

STRAUB, S. C. K. 2010. Molecular phylogentics of *Amorpha* L. (Fabaceae) and conservation genetics of rare *Amorpha* species. Ph.D. dissertation, Cornell University, Ithaca, New York, USA.

SWOFFORD, D. L. 2003. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sinauer, Sunderland, Massachusetts, USA.

WENDEL, J. F., AND J. J. DOYLE. 2005. Polyploidy and evolution in plants. *In* R. J. Henry [ed.], Plant diversity and evolution: Genotypic and phenotypic variation in higher plants, 97–117. CABI Publishing, Wallingford, UK.

WOOD, T. E., N. TAKEBAYASHI, M. S. BARKER, I. MAYROSE, P. B. GREENSPOON, AND L. H. RIESEBERG. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences, USA* 106: 13875–13879.

CHAPTER 3

THE LOW COPY NUCLEAR GENE CNGC5 PROVIDES INSIGHTS INTO RELATIONSHIPS AND AUTOPOLYPLOID EVOLUTION IN THE *MEDICAGO SATIVA* (LEGUMINOSAE) COMPLEX

Abstract

The assumption that autopolyploidy is a simpler evolutionary process than allopolyploidy is poorly supported and merits more extensive study. The *Medicago sativa* complex comprises several diploid and autopolyploid taxa, including autopolyploid cultivated alfalfa (*M. sativa* subsp. *sativa*). Because of its agricultural importance, the *M. sativa* complex is well studied and has many genetic and genomic resources, making it a potential model for investigating autopolyploid evolution. At the diploid level, there are two key morphologically and genetically differentiated groups, *M. sativa* subsp. *caerulea* with violet flowers and coiled pods and subsp. *falcata* with yellow flowers and falcate pods. Each has been hypothesized to give rise to an autopolyploid: subsp. *caerulea* to subsp. *sativa*, and diploid subsp. *falcata* to a tetraploid cytotype of the same name. Previous chloroplast DNA sequence data suggest different evolutionary trajectories of the two autopolyploids and raised questions regarding the origins of tetraploid subsp. *falcata*. In order to address these questions and to study the autopolyploid patterns in the complex, a nuclear low-copy gene, CNGC5, was used to explore genetic variation among diploid and tetraploid taxa, including hybrids and the closely related *M. prostrata*. Relationships among CNGC5 alleles were reconstructed and genetic differentiation was estimated. Our results show genetic differentiation between the two key morphological groups at both ploidal levels. Nuclear data also show that the tetraploid subsp. *falcata* most likely was

derived from its diploid cytotype in a parallel pattern to the autopolyploidy of subsp. *sativa* from subsp. *caerulea*. The nomenclature of the members in the complex is discussed.

Introduction

The recognition of polyploidy as an important and common phenomenon in plants has been accumulated through increasing evidence of its consequences on genome evolution (e.g., reviewed in Adams and Wendel, 2005; Doyle et al., 2008) and speciation (e.g., Soltis et al., 2007; Wood et al., 2009), and its frequency in the evolutionary history of seed plants (Masterson, 1994; Soltis et al., 2009; Jiao et al., 2011). Polyploidy is generally categorized into two types: allopolyploidy and autopolyploidy. The two types of polyploidy can be defined in two senses, genetic and taxonomic. Whereas a classic allopolyploid is a product of hybridization between two or more species (taxonomic) and would show bivalents at meiosis and disomic segregation (genetic), a classic autopolyploid is formed from within a single species, would have multivalents at meiosis, and would show polysomic segregation. However, a taxonomic autopolyploid can be genetically allopolyploid and vice versa (Ramsey and Schemske, 1998; Doyle and Egan, 2010, p. 74). Despite a variety of empirical studies on polyploids, many questions remain unanswered (Soltis et al., 2010), especially for autopolyploidy, which is less studied than allopolyploidy. Interest in autopolyploids has increased as the phenomenon has been recognized to be more common and have more evolutionary advantages than previously considered (Soltis and Soltis, 2000; Soltis et al., 2004, 2007; Parisod et al., 2010).

The genus *Medicago* contains several species that have multiple cytotypes, including at least four taxonomic autopolyploids with morphologically similar diploid and tetraploid cytotypes: *M. lupulina* L., *M. papillosa* Boiss., *M. prostrata* Jacq., and *M. sativa* L. (Small,

2011). Besides *M. truncatula*, a genomic model legume (Young et al., 2011), *M. sativa* is probably the most studied species of the genus due to the agricultural importance of cultivated alfalfa, *M. sativa* subsp. *sativa*. Alfalfa is considered the "Queen of Forages," grown on about 22 million acres each year in the U.S. and 80 million acres worldwide (Russelle, 2001). In addition to being a forage crop, a honey plant, and human food, more recently, alfalfa has potential as a biofuel crop (e.g., McCaslin and Miller, 2007; Martin and Jung, 2010). Its importance is also shown in a large number of studies on such topics as taxonomy, genetics, cytology, domestication, agricultural practices, crop production, and pharmaceuticals. Genetic resources for alfalfa are also abundant (Small, 2011). The knowledge and genetic resources provide a platform for the study of polyploidy in this species complex.

Medicago sativa consists of diploid ($2n = 2x = 16$) and tetraploid ($2n = 4x = 32$) cytotypes, both of which are perennial and outcrossing (Lesins and Lesins, 1979; Quiros and Bauchan, 1988; Small, 2011). Plants of the two ploidy levels can hybridize through unreduced gametes produced from a diploid parent (Bingham, 1968; McCoy, 1982; McCoy and Bingham, 1988; Quiros and Bauchan, 1988). The tetraploid members are genetic autopolyploids, exhibiting tetrasomic inheritance (Stanford, 1951; Quiros, 1982; McCoy and Bingham, 1988) and occasional quadrivalents at meiosis (Armstrong, 1971; Gillies, 1972; Mariani and Veronesi, 1971). Morphologically, three key diploid taxa can be identified based on flower color, pod shape, and pod pubescence: *Medicago sativa* subsp. *caerulea* (Less. Ex Ledeb.) Schmalh. (hereafter referred as "caerulea") has violet or blue-violet flowers and coiled pods without glandular hairs; *M. s.* subsp. *falcata* (L.) Arcang. (= *M. falcata* L.; hereafter "diploid falcata") has yellow flowers and falcate pods with or without glandular hairs; and *M. s.* subsp. *glomerata* (Balb.) Rouy (= *M. glomerata* Balb.; hereafter "diploid glomerata") has yellow flowers and

coiled pods with glandular hairs (Table 3.1). Each diploid is hypothesized to have produced autopolyploids, and hybridization at the diploid and polyploid levels is thought to have given rise to other members of the complex. Taxonomic rank and nomenclature of the members vary in different treatments. Most members were recognized as different species by Lesins and Lesins (1979), whereas they are considered subspecies of *M. sativa* by Quiros and Bauchan (1988), Small and Jomphe (1989), and Small (2011). Our usage of names is adapted from Quiros and Bauchan (1988) and Small (2011) and is given, along with key characters, in Table 3.1.

Interspecific hybridization studies for breeding purposes have shown that *M. prostrata* is the easiest species from outside the complex to hybridize with *M. sativa* (Small, 2011). Lesins (1962, 1968) showed that *M. prostrata* can be hybridized with *M. sativa* at both diploid and tetraploid levels, but only when the former is used as the paternal parent. In addition, natural hybrids of falcata and *M. prostrata* were reported by Lesins and Lesins (1979, p. 95) without indicating their ploidy level. Morphologically, *M. prostrata* can be confused with glomerata as both have yellow flowers and coiled, nonspiny pods with gland-tipped trichomes, except that glomerata does not have the long reflexed pedicels of *M. prostrata* (Small, 2011).

In the past few years, the relationships, evolutionary history, and genetic diversity of a wide range of members in the complex have been studied based on mitochondrial DNA (mtDNA: Havananda et al., 2010), cpDNA (Havananda et al., 2010, 2011) and simple sequence repeat (SSR) markers (Şakiroğlu et al., 2010). At the diploid level, both cpDNA and SSR data revealed genetic differentiation between violet-flowered caerulea and yellow-flowered diploid falcata, which had not been detected in some previous studies by, e.g., Skinner (2000) using fragment length polymorphism in cpDNA hypervariable regions, Muller et al. (2006) using nuclear DNA (nDNA) sequences, and Havananda et al. (2010) using mtDNA haplotypes. At the

tetraploid level, the same cpDNA regions used by Havananda et al. (2010) were used to elucidate genetic diversity, origins of polyploids, and patterns of gene flow within and between ploidal levels (Havananda et al., 2011). Although they found a "classic" autopolyploid pattern with a lack of chloroplast haplotype differentiation between caerulea and sativa, this was not true of diploid and tetraploid falcata. Surprisingly, tetraploid falcata as a whole was strongly differentiated from diploid falcata, with most individuals having haplotypes more closely related to those of *M. prostrata*. This finding raised the questions of whether tetraploid falcata arose from diploid falcata as traditionally hypothesized, or from *M. prostrata* as suggested by chloroplast haplotypes.

Nucleotide sequence data from different nDNA loci have also been used in phylogenetic studies of the genus *Medicago*, including a few samples from *M. sativa* complex (Maureira et al., 2008, cyclic nucleotide-gated channel [CNGC5] and putative coatomer beta subunit [*β-cop*-like]; Steele, 2010, gibberellin 3-β-hydroxylase [*GA3ox1*]). These studies reveal ambiguous taxonomic boundaries of *M. sativa* within the genus, suggesting closer relationships of other species with *M. sativa* than morphology suggests. Within the *M. sativa* complex, Muller et al. (2006) surveyed DNA sequence diversity at two nuclear loci, pectate lyase homologue (*pect*) and NADH-dependent glutamate synthase (*glu*), to estimate genetic diversity and infer domestication history of the species. They found no genetic differentiation among the taxa in the complex and suggested that gene flow and recent divergence accounted for the low differentiation; interploidal gene flow was suggested to have contributed recurrently to homogenizing genetic diversity between caerulea and autopolyploid sativa.

TABLE 3.1. Usage of names and principal characters used in distinguishing members of *Medicago sativa* and *M. prostrata* (Quiros and Bauchan, 1988; Small, 2011) (Modified from Havananda et al., 2011)

Taxon	Referred in this chapter as	Ploidy	Flower color	Pod shape	Pod pubescence
M. sativa subsp. *caerulea* (Less. Ex Ledeb.) Schmalh.	caerulea	Diploid	Violet	Coiled	Simple or glabrous
M. s. subsp. *sativa*	sativa	Tetraploid	Violet	Coiled	Simple or glabrous
M. s. subsp. *falcata* (L.) Arcang.	Diploid or tetraploid falcata	Di-, tetraploid	Yellow	Straight or sickle	Simple or glandular
M. s. subsp. *glomerata* (Balb.) Rouy	Diploid or tetraploid glomerata	Di-, tetraploid	Yellow	Coiled	Glandular
M. s. subsp. ×*hemicycla* (Grossh.) C. R. Gunn	hemicycla	Diploid	Segregated or Variegated	Curved or loosely coiled	Simple
M. s. subsp. ×*varia* (Martyn) Arcang.	varia	Tetraploid	Segregated or Variegated	Curved or loosely coiled	Simple
M. s. subsp. ×*tunetana* Murb.	tunetana	Tetraploid	Segregated or Variegated	Coiled	Glandular
M. prostrata Jacq.	*M. prostrata*	Di-, tetraploid	Yellow	Coiled	Glandular

In order to understand the complex patterns involving autopolyploidy in the *M. sativa* complex and to address the questions raised by Havananda et al. (2011), we use a low-copy nuclear gene, CNGC5 (Maureira et al., 2008), to explore genetic variation among the members of the complex. We reconstruct relationships among nDNA alleles of a number of non-hybrid and hybrid diploids and tetraploids in the complex to infer origins of the tetraploids and understand patterns of gene flow within and between ploidal levels. We estimate genetic differentiation between the two key morphological groups, violet-flowered with coiled pods and yellow flowered with falcate pods, at both ploidal levels, and elucidate the polyploid origins of the two main autopolyploids. We compare variation patterns at CNGC5 with those from cpDNA (Havananda et al., 2010, 2011) and address the nomenclature of the members of the complex.

Materials and Methods

One hundred and twenty accessions (130 individuals) of *Medicago* from diverse geographical areas were chosen from the *Medicago* species collection maintained by the USDA National Plant Germplasm System and listed in the Germplasm Resources Information Network (GRIN; http://www.ars-grin.gov/cgi-bin/npgs/html/splist.pl?7365), along with one accession of glomerata from the South Australian Research and Development Institute (SARDI). Four developed germplasms, two from synthetic population SD201 developed from wild and cultivated diploid falcata genotypes (Boe et al., 1998), one genotype from tetraploid WISFAL germplasm (PI 560333, an artificially synthesized tetraploid falcata converted from diploid WISFAL using 2*n* egg cells; Bingham, 1990, 1993), and one genotype from the artificially diploidized sativa germplasm CADL ("cultivated alfalfa at the diploid level"; Bingham and McCoy, 1979), were also included. All plants were grown in greenhouses at Iowa State

University and/or the University of Georgia (by E. C. Brummer between 2003 and 2008) or at Cornell University (by T. Havananda between 2006 and 2010). Ploidy of most accessions was confirmed by flow cytometry using various methods (described in Havananda et al., 2011).

A CTAB protocol (Doyle and Doyle, 1987) was used to extract genomic DNA of all plants. The low-copy nuclear gene CNGC5 was amplified in all samples by PCR using primers used successfully across *Medicago*, *Trigonella*, and *Trifolium* by Maureira et al. (2008). Each PCR reaction of 12.5 µL consisted of 0.75 µL template DNA, 0.5 U of *Taq* DNA polymerase (New England BioLabs, Inc., Ipswich, Massachusetts, USA), 10 µM of each primer, and 10× PCR buffer containing 67 mM Tris pH 8.0, 2 mM $MgCl_2$, 250 µM of each dNTP, 2% DMSO. Temperature cycling conditions run on MJ Research (Waltham, Massachusetts, USA) or Techne (Princeton, New Jersey, USA) thermal cyclers were 5 min at 94°C; 38–40 cycles of 30 s at 95°C, 30 s at 58–60°C; 1 min at 72°C; and 7 min at 72°C. Alternatively, touchdown PCR cyclers (95°C for 3 min followed by 15 cycles of 95°C for 30 s, 70°C for 45 s, 72°C for 75 s with a decrease in annealing temperature of 1°C per cycle, followed by 20 cycles with 56°C annealing temperature and a final extension at 72°C for 5 min) were used. The success of the PCR and number of bands produced were visualized by ethidium bromide in agarose gel electrophoresis. Successful PCR with single band product was purified using QIAquick PCR Purification Kit (Qiagen, Valencia, California, USA) according to manufacturer's protocol. Enzyme PCR purification protocol was also alternatively used. Enzyme PCR clean-up mix for 20 µL of PCR product was composed of 0.5 µL of Exonuclease I (20,000 U/mL; New England BioLabs, Inc.), 0.5 µL of Antarctic Phosphatase (5,000 U/mL; New England BioLabs, Inc.), and 3.0 µL of 1× Standard *Taq* reaction buffer (New England BioLabs, Inc.). The PCR clean-up reaction was incubated at 37°C for 45 min and then 90°C for 10 min on a thermocycler. The PCR products appeared more than one

band on the gel electrophoresis were separated on 1–2% agarose gels, excised, and then purified using QIAquick Gel Extraction Kit (Qiagen). Purified PCR products were used as templates for sequencing of both strands with primers used in amplifications and specific primers (see below). DNA sequencing reactions were performed using BigDye® Terminator (Applied Biosystems, Foster City, California, USA) and run on an ABI 3730xl DNA sequencer (Applied Biosystems) by the Cornell Biotechnology Core Facility (CLC). Some sequencing reactions were performed using ABI BigDye® Terminator in 96-well plates and cleaned up by ethanol precipitation before being run on the DNA sequencer by the CLC.

A single-strand conformation polymorphism (SSCP) analysis was performed to screen for any DNA polymorphisms in single individuals. PCR products were run on 0.7× MDE™ (Cambrex Bioscience Rockland, Rockland, Maine, USA) gels at 1–2 W at 4°C for 24–64 hrs. The gels were then stained with SYBR Gold (Invitrogen, Carlsbad, California, USA) for visualization. Bands excised from SSCP gels were mashed in water, incubated at 60°C for 5 min, and then centrifuged for 10 min. The supernatant was used as a template for CNGC5 re-amplification. Re-amplified PCR products were purified and used for sequencing. PCR re-amplification, purification, and sequencing of DNA extracted from SSCP gels were performed as described above.

From direct sequencing, several diploid and polyploid individuals were found to be heterozygous and to carry more than one polymorphic site, and phase determination was needed to reveal allelic variation in those individuals. In such cases, SSCP analysis, allele-specific primers, and/or cloning was implemented. Allele-specific primers were designed from sequence alignments of homozygous samples and heterozygous samples (Table 3.2). Because homozygous samples did not account for all alleles, heterozygous samples were included in the alignment so

that polymorphic sites that were not present across the homozygotes could be detected and used in the specific primer design. These primer sets were used in PCR amplification and/or sequencing of CNGC5 alleles from heterozygotes. A few of these primer sets were also used in PCR for some samples, when PCR from the original primers were not successful, to screen for polymorphisms from which they were designed. PCR products from heterozygous samples whose alleles could not be separated by PCR or sequencing with specific primers were cloned using the TOPO TA Cloning® Kit for Sequencing (Invitrogen) according to the manufacturer's protocol modified by doing ½ volume reaction. Bacterial colonies were picked and mixed in 15 µL TE buffer (10 mmol/L Tris; 0.1 mmol/L EDTA, pH 8.0; Integrated DNA Technologies, Inc., Coralville, Iowa, USA); 0.5 µL of the colony was used as template for 50 µL PCR reaction using the original CNGC5 primers with final ingredient concentration as described above. Prior to starting PCR cycles, the PCR reaction was incubated at 94°C for 10 min to lyse the bacterial cell. PCR thermal cycling conditions were as described above. Two to 16 clones per individuals were sequenced in both directions. Sequencing of PCR products obtained from these two techniques was performed as described above for PCR products amplified from the original primers.

For each allele, at least two sequences (forward and reverse) plus sequences generated by using specific primers or by cloning, were aligned to assemble a full length CNGC5 sequence. Full-length sequences across alleles were aligned and manually adjusted. Sequence assembly, alignment, and editing were done using the program Sequencher™ version 4.2 or later (Gene Codes, Ann Arbor, Michigan, USA). Insertions/deletions (indels) in the alignment were coded using the simple indel coding method (Simmons and Ochoterena, 2000) implemented in the program SeqState version 1.37 (Müller, 2005). Binary characters of coded indels were added to the alignment for phylogenetic and network analyses.

TABLE 3.2. Allele-specific primers designed from CNGC5 sequence alignments across homo- and heterozygous individuals and used for PCR and/or sequencing.

Primer name [a]	Type of polymorphism [c]	Direct Sequence (read in 5'-3' orientation) [b]	Specific Primer [a]
CNGC5F_43A	SNP	GTCACAAACACTTAAAATTAGTGAAACW	GTCACAAACACTTAAAATTAGTGAAACA
CNGC5F_43T			GTCACAAACACTTAAAATTAGTGAAACT
CNGC5F_74C	SNP	CGTGKGACTAAAAAGGACTTGACTTY	CGTGTGACTAAAAAGGACTTGACTTC
CNGC5F_74T			CGTGTGACTAAAAAGGACTTGACTTT
CNGC5F_90CT	Indel	TTAGTGAAACACCAAAACACTCTCT[ct]TA	TTAGTGAAACACCAAAACACTCTCTCT
CNGC5F_90TA			TTAGTGAAACACCAAAACACTCTCTTA
CNGC5F_121A	SNP	AAGGACTTGACTTTTTTCCGATAAAGR	AAGGACTTGACTTTTTTCCGATAAAGA
CNGC5F_121G			AGGACTTGACTTTTTTCCGATAAAGG
CNGC5F_128A	SNP	GACTTGACTTTTTTCCGATAAAGAACM	GACTTGACTTTTTTCCGATAAAGAACA
CNGC5F_128C			ACTTGACTTTTTTCCGATAAAGAACC
CNGC5F_146TG	Indel	GATCAGTTCCACATTGATGTTGT[tgt]GC	GATCAGTTCCACATTGATGTTGTTG
CNGC5F_149GC			GATCAGTTCCACATTGATGTTGTGC
CNGC5F_167T	Indel	TGTTGTGCAGTATATCACATCAAGT[t]A	TGTTGTGCAGTATATCACATCAAGTT
CNGC5F_168A			TGTTGTGCAGTATATCACATCAAGTA
CNGC5F_168A	SNP	GTAGAAAATGAAATGTTGTCGATCAGW	GTAGAAAATGAAATGTTGTCGATCAGA
CNGC5F_168T			GTAGAAAATGAAATGTTGTCGATCAGT
CNGC5F_192A	SNP	AGTTATCATTTAGTCCTTCTGCATGM	AGTTATCATTTAGTCCTTCTGCATGA
CNGC5F_192C			GTTATCATTTAGTCCTTCTGCATGC
CNGC5F_289C	SNP	TGATCTCTCTCTTAGATTAGTTACCY	GATCTCTCTCTTAGATTAGTTACCC
CNGC5F_289T			TGATCTCTCTCTTAGATTAGTTACCT
CNGC5F_indCC	Indel	CTCTCTTAGATTGATTTCATTGAT[ccttatttcactgat]WGA	CTCTCTTAGATTGATTTCATTGATCC
CNGC5F_indTG			CTCTCTTAGATTGATTTCATTGATTG

TABLE 3.2 (Continued)

Primer name [a]	Type of polymorphism [c]	Direct Sequence (read in 5'-3' orientation) [b]	Specific Primer [a]
CNGC5R_191A	SNP	MGTATAAAAATATAGTGTTTGAGATTTGT	ACAAATCTCAAACACTATATTTTTATACT
CNGC5R_191C			ACAAATCTCAAACACTATATTTTTATACG
CNGC5R_272C	SNP	YAACTCATAGATTTGCTTTTCAAATTCT	GAATTTGAAAAGCAAATCTATGAGTTG
CNGC5R_272T			AGAATTTGAAAAGCAAATCTATGAGTTA
CNGC5R_328A	SNP	WCAAATTCTTTTWGGTATTTAAACATGC	GCATGTTTAAATACCTAAAAGAATTTGT
CNGC5R_328T			GCATGTTTAAATACCTAAAAGAATTTGA
CNGC5R_380C	SNP	STTATCTCTCTTAGATTGATTTCATTG	CAATGAAATCAATCTAAGAGAGATAAG
CNGC5R_380G			CAATGAAATCAATCTAAGAGAGATAAC
CNGC5R_496A	SNP	WACTTATTTTCATGGATTTGTAGGTTC	GAACCTACAAATCCATGAAAATAAGTA
CNGC5R_496T			GAACCTACAAATCCATGAAAATAAGTT
CNGC5R_508C	SNP	YATTTACTTATTTTCATGGATTTGTAGG	CCTACAAATCCATGAAAATAAGTAAATG
CNGC5R_508T			CCTACAAATCCATGAAAATAAGTAAATA
CNGC5R_742C	SNP	YGAGTTCTGTGGGGAGGAGCTTT	AAAGCTCCTCCCCACAGAACTCG
CNGC5R_742T			AAAGCTCCTCCCCACAGAACTCA
CNGC5_370F	Indel	CTCTCTTAGATTGATTTCATTGAT[ccttatttcactgat]WGA	CTCTCTTAGATTGATTTCATTGATTGA
CNGC5_370R		TTGAT[ccttatttcactgat]WGATTTTCAAATTTTTAAAATTCGA	TCGAATTTTAAAAATTTGAAAATCAATCAA
CNGC5F_256C	SNP	GTTACCYGGTCSATGAACTATTTCAY	GTTACCYGGTCSATGAACTATTTCAC

[a] Specific primers with "R" in their name were designed for reverse direction. Bases in these primers read in 3'-5' orientation.
[b] Insertion/deletion bases are in lowercase in square brackets.
[c] SNP = single nucleotide polymorphism; Indel = insertion/deletion polymorphism

Allele phylogenies were reconstructed using maximum likelihood (ML), networks were estimated using two approaches, NeighborNet (NNet) and statistical parsimony. OTUs in ML and NNet analyses were the alleles identified by sequencing. The ML analysis was conducted using the program PhyML 3.0 online web server (http://www.atgc-montpellier.fr/phyml/; Guindon et al., 2010), with the general time reversible (GTR) DNA substitution model. Bootstrap analysis was performed with 1,000 replicates to estimate branch support. The output tree from the program PhyML was viewed and edited using the program FigTree version 1.3.1 (Rambaut, 2006–2009). NNet analysis was implemented in the program SplitsTree4 version 4.10 (Huson and Bryant, 2006) using uncorrectedP distance. Bootstrap analysis of 1,000 replicates was performed to estimate support for splits. The pairwise homoplasy index (PHI) test (Bruen et al., 2006) implemented in SplitsTree4 was used to test for evidence of recombination. Statistical parsimony network analysis (Templeton et al., 1992) was performed with the program TCS version 1.21 (Clement et al., 2000) with all sequences included to allow the calculation of haplotype frequencies. To measure CNGC5 DNA variation within each taxon and for all *M. sativa* accessions, the program DnaSP version 5 (Librado and Rozas, 2009) was used to calculate number of haplotypes (H), haplotype diversity (H_d), average number of nucleotide differences per site between two sequences (nucleotide diversity, π), and the proportion of segregating polymorphic sites per nucleotide (the Watterson estimator, θ_w). Genetic differentiation between the diploids, the tetraploids, diploid progenitors and their tetraploid descendants, and between each of them and *M. prostrata* was estimated by analysis of molecular variance (AMOVA) performed in the program Arlequin version 3.5 (Excoffier and Lischer, 2010).

Results

The success of CNGC5 amplification and sequencing varied among the 134 individuals sampled. For some individuals, only weak PCR products were obtained, even after considerable experimentation with different cycling conditions. When high quality sequence was obtained, heterozygosity was readily determined from the sequences. In cases where more than one polymorphic site was found in a sequence, phase determination was attempted using SSCP and/or allele-specific primer approaches, whereas cloning was used mostly for tetraploids, which could have up to four alleles per individual. Although SSCP produced multiple bands in several individuals, suggesting heterozygosity, re-amplification of CNGC5 from the DNA excised from SSCP gels was not successful and complete sequences of the alleles could not be obtained. The allele-specific primer approach, however, was much more successful, especially with heterozygous diploids. For heterozygous diploids, primers were designed based on substitution or insertion/deletion polymorphisms, and were used to amplify and directly sequence individual alleles. For the more difficult samples possessing length variation, more polymorphic sites, or more than two alleles, allele sequences were determined by cloning. For a few tetraploid individuals where more than four different sequences were revealed from cloning, sequences identical to those obtained from direct sequencing using original or specific primers for other individuals were chosen, as these should be actual alleles as opposed to polymerase error. Other sequences obtained from cloning that had a unique combination of substitutions and/or indels observed in other alleles were considered new, actual alleles.

CNGC5 sequence data were obtained for 55 individuals, 34 of which were heterozygotes whose allelic composition was successfully determined. The 55 individuals included 11 caerulea, 19 diploid falcata, seven tetraploid falcata, five hemicycla, four sativa, two varia, six *M.*

prostrata, and one *M. truncatula* (Appendix 3.1). The alignment of CNGC5 sequence was 860 bp long, including 11 coded indels, and contained 49 nucleotide substitutions (847 bp, including 5 coded indels, with 37 nucleotide substitutions when *M. truncatula* was excluded). Thirty of the polymorphisms in the dataset were parsimony-informative and the other 30 were autapomorphic. From a total of 95 sequences obtained, 39 alleles were identified, 24 of which were singleton alleles, including the single allele from *M. truncatula*. Allelic composition of each individual is given in Appendix 3.2. Measures of genetic diversity are summarized in Table 3.3.

CNGC5 allele relationships—The ML analysis identified a tree with logL = −1668.34 on which relationships among included taxa were not strongly resolved (Figure 3.1). All alleles from the sativa complex were placed in a clade separate from the three alleles sampled from prostrata individuals. All 11 alleles of the violet-flowered taxa, caerulea and sativa (allele numbers 1–7 and 10–13), were grouped together on one clade along with five of 18 alleles of the yellow-flowered taxa, diploid and tetraploid falcata (alleles 8, 9, 12, 14, 15). In contrast, the remaining alleles from falcata accessions formed a clade in the tree. NNet analysis also identified the same grouping of all caerulea and sativa alleles together and the same five alleles of diploid and tetraploid falcata (Figure 3.2). The allele from *M. truncatula*, when included, was connected to the NNet, with a very long edge and additional splits, to alleles from prostrata and tetraploid falcata (Figure 3.2). For the statistical parsimony analysis, the maximum connection at 95% confidence was 12 steps. The *M. truncatula* allele was not connected to the network at this limit, but when connection limit was set to 20 steps, the allele was connected to a tetraploid falcata allele (allele 34; Figure 3.3). Similar to ML and NNet results, all caerulea and sativa alleles were closely related to each other, whereas most diploid and tetraploid falcata alleles also appeared more closely related to each other than to caerulea/sativa alleles (Figure 3.3). Hybrid taxa,

hemicycla and varia, shared identical or closely related alleles with their parent taxa. They did not appear to form any unique groups of alleles. Only a single loop appeared in the network, which was caused by a single nucleotide substitution between the pairs of alleles in the loop (alleles 25, 26, 29). The pairwise homoplasy index (PHI) test (Bruen et al., 2006) did not find statistically significant evidence for recombination ($P = 0.093$). Alleles from *M. prostrata* were different from *M. sativa* alleles and connected to the network with more steps than most other alleles. From the network, *M. prostrata* alleles had closer relationships with alleles from diploid/tetraploid falcata than with those from caerulea/sativa. AMOVA results showed significant differentiation between *M. prostrata* and both the violet-flowered and yellow-flowered taxa at both ploidal levels (Table 3.4).

The network was dominated by two common alleles, one primarily found in caerulea individuals, and the other in falcata individuals. Caerulea had less CNGC5 DNA variability, carrying alleles that were closer to each other in the network than were alleles from diploid falcata (Table 3.3; Figure 3.3). One diploid falcata individual had caerulea allele 12, whereas another six individuals had alleles more closely related to cearulea alleles than to most other diploid falcata alleles. Despite this, diploid falcata was significantly differentiated from caerulea (Table 3.4). A similar pattern was seen at the tetraploid level: the level of DNA sequence variation, as measured by nucleotide diversity (π) and the Watterson estimator (θ_w), was lower in sativa than in tetraploid falcata (Table 3.3). Two tetraploid falcata individuals were found to have alleles more closely related to sativa alleles than to other tetraploid falcata alleles. But, as with caerulea and diploid falcata, AMOVA indicated a significant differentiation between the two tetraploid taxa (Table 3.4).

Of eight alleles found in sativa, four were also found in caerulea and the other four were closely related to caerulea alleles. AMOVA results showed that these tetraploid and diploid taxa are not significantly differentiated from each other (Table 3.4). For tetraploid falcata, although only three of 13 alleles were also found in diploid falcata, AMOVA results showed no significant differentiation between the two falcata cytotypes (Table 3.4).

TABLE 3.3. Summary statistics of DNA sequence variation for *Medicago* taxa. For each taxon, number of accessions and standard measures of diversity (indels were not considered) are indicated. Only taxa with multiple individuals were presented.

Taxon	No. of individuals [a]	No. of polymorphic sites [b]	No. of indels	H	H_d	π	θ_w
M. s. subsp. *caerulea*	11 (16)	5 (2)	4	5	0.533	0.00122	0.00180
M. s. subsp. *sativa* [c]	4 (11)	7 (4)	0	8	0.945	0.00263	0.00284
M. s. subsp. *falcata* (2*x*)	19 (30)	17 (13)	2	12	0.903	0.00481	0.00511
M. s. subsp. *falcata* (4*x*)	7 (17)	14 (11)	2	12	0.941	0.00448	0.00493
M. s. subsp. ×*hemicycla*	5 (9)	6 (5)	2	5	0.806	0.00271	0.00263
M. s. subsp. ×*varia* [d]	2 (4)	5 (0)	0	3	0.833	0.00297	0.00324
All *M. sativa*	48 (87)	30 (21)	8	30	0.920	0.00418	0.00714
M. prostrata	6 (7)	6 (6)	15	3	0.762	0.00357	0.00296

Notes: H = number of haplotypes; H_d = haplotype (gene) diversity; π = average number of nucleotide differences per site between two sequences (nucleotide diversity); θ_w = the Watterson estimator, proportion of segregating polymorphic sites per nucleotide ($\theta = 4N\mu$).

[a] Number of sequences used in parentheses.

[b] Number of parsimony informative sites in parentheses.

[c] Including two individuals from accessions listed as *M. s.* subsp. *caerulea* in GRIN, but found to be tetraploids by flow cytometry.

[d] Listed as *M. s.* subsp. ×*hemicycla* in GRIN, but found to be tetraploids by flow cytometry.

FIGURE 3.1. Maximum likelihood (ML) tree for the 39 CNGC5 alleles identified in *Medicago*. Numbers above branches are bootstrap values estimated from 1,000 replicates (only values over 50% are shown). Number at each terminal is an allele number, which is followed by an abbreviated taxon name and an accession number (USDA Plant Introduction [PI] number) or a variety name of an individual or a representative (in brackets) of the individuals possessing that allele. Abbreviations: Mp, *M. prostrata*; Msc, *M. sativa* subsp. *caerulea* (caerulea); Msf, *M. s.* subsp. *falcata* (falcata); Msh, *M. s.* subsp. ×*hemicycla* (hemicycla); Mss, *M. s.* subsp. *sativa* (sativa); Mss(c), *M. s.* subsp. *sativa* originally listed as *M. s.* subsp. *caerulea* in GRIN, but found to be tetraploids by flow cytometry; Mt, *M. truncatula*. Different individuals from the same accession are denoted with decimal number; e.g., Msf631592.1 is the individual number one from falcata accession PI 631592. Multiple alleles found in heterozygous individuals are designated with "a," "b," "c," or "d" following an underscore at the end of the accession or individual number. Individuals possessing each allele are listed in Figure 3.3.

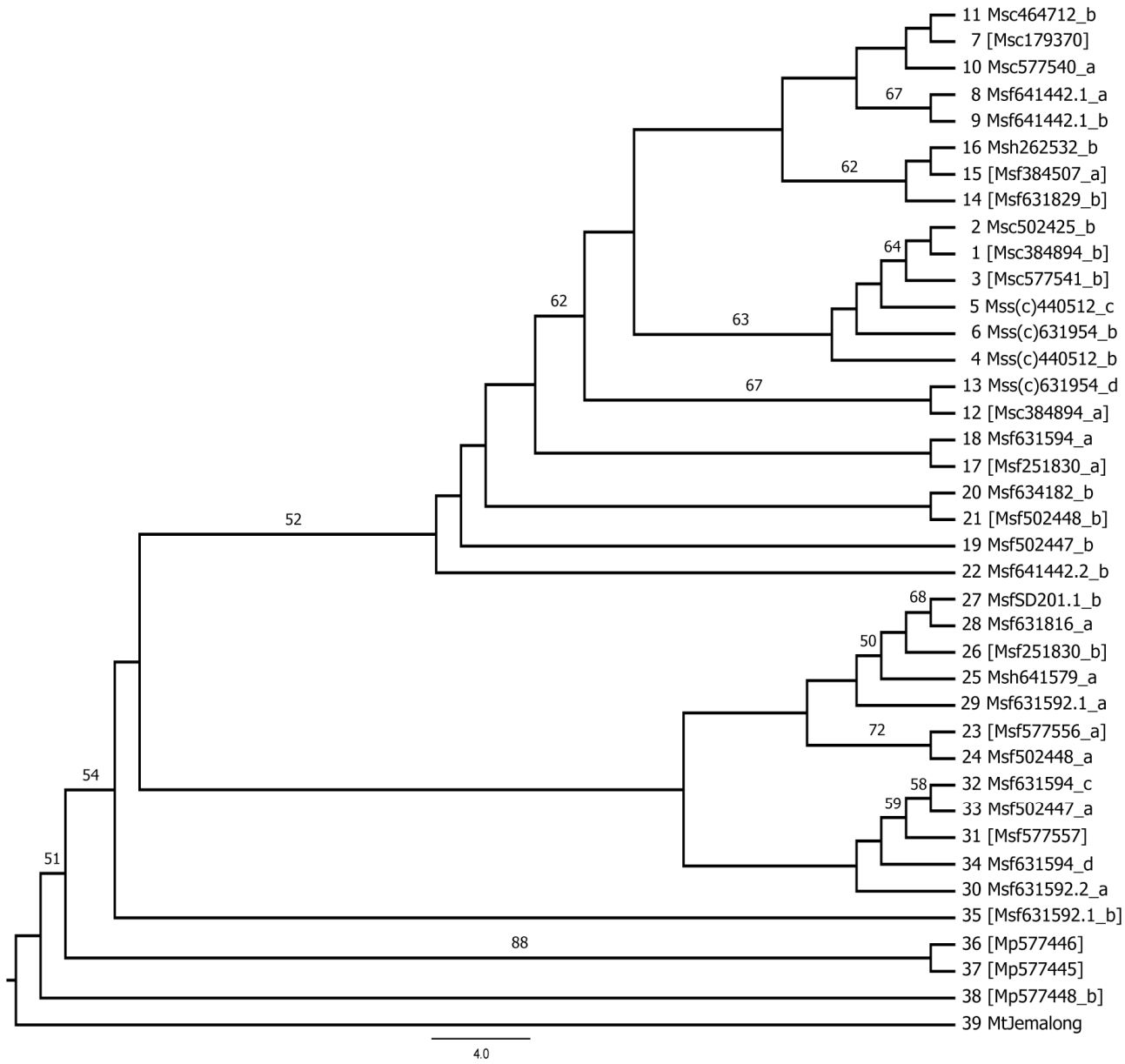

11 Msc464712_b
7 [Msc179370]
10 Msc577540_a
8 Msf641442.1_a
9 Msf641442.1_b
16 Msh262532_b
15 [Msf384507_a]
14 [Msf631829_b]
2 Msc502425_b
1 [Msc384894_b]
3 [Msc577541_b]
5 Mss(c)440512_c
6 Mss(c)631954_b
4 Mss(c)440512_b
13 Mss(c)631954_d
12 [Msc384894_a]
18 Msf631594_a
17 [Msf251830_a]
20 Msf634182_b
21 [Msf502448_b]
19 Msf502447_b
22 Msf641442.2_b
27 MsfSD201.1_b
28 Msf631816_a
26 [Msf251830_b]
25 Msh641579_a
29 Msf631592.1_a
23 [Msf577556_a]
24 Msf502448_a
32 Msf631594_c
33 Msf502447_a
31 [Msf577557]
34 Msf631594_d
30 Msf631592.2_a
35 [Msf631592.1_b]
36 [Mp577446]
37 [Mp577445]
38 [Mp577448_b]
39 MtJemalong
67
62
64
63
62
67
52
68
50
72
58
59
54
51
88
4.0

FIGURE 3.2. NeighborNet network for the 39 CNGC5 alleles identified in *Medicago*. Numbers next to edges are bootstrap values estimated from 1,000 replicates (only values over 50% are shown). Number at each terminal is an allele number (in parentheses). Dashed lines denote splits and a very long edge appeared in the network when *M. truncatula* was included in the analysis. Conventions are as in Figure 3.1.

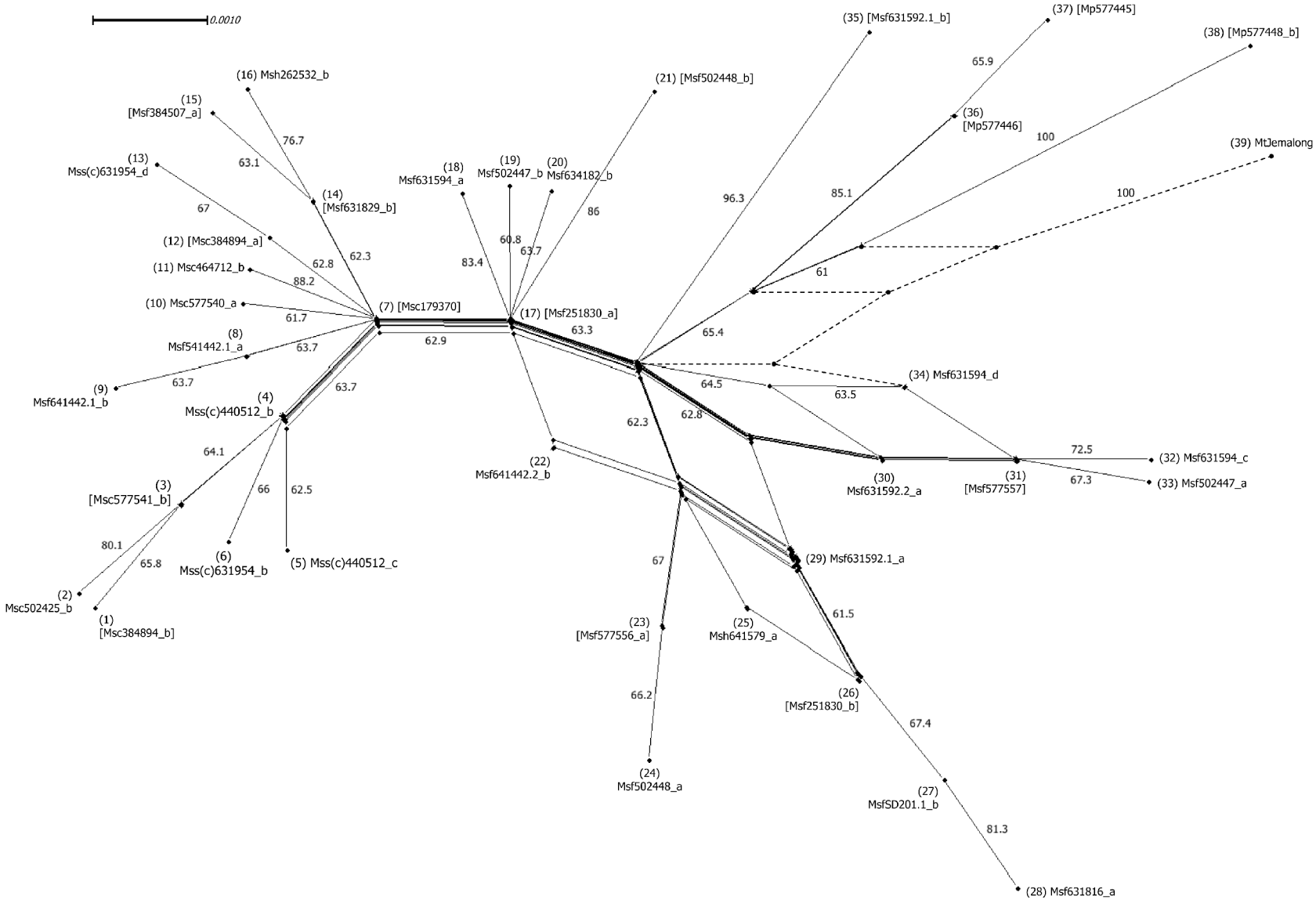

0.0010
(1) [Msc384894_b]
(2) Msc502425_b
(3) [Msc577541_b]
(4) Mss(c)440512_b
(5) Mss(c)440512_c
(6) Mss(c)631954_b
(7) [Msc179370]
(8) Msf541442.1_a
(9) Msf641442.1_b
(10) Msc577540_a
(11) Msc464712_b
(12) [Msc384894_a]
(13) Mss(c)631954_d
(14) [Msf631829_b]
(15) [Msf384507_a]
(16) Msh262532_b
(17) [Msf251830_a]
(18) Msf631594_a
(19) Msf502447_b
(20) Msf634182_b
(21) [Msf502448_b]
(22) Msf641442.2_b
(23) [Msf577556_a]
(24) Msf502448_a
(25) Msh641579_a
(26) [Msf251830_b]
(27) MsfSD201.1_b
(28) Msf631816_a
(29) Msf631592.1_a
(30) Msf631592.2_a
(31) [Msf577557]
(32) Msf631594_c
(33) Msf502447_a
(34) Msf631594_d
(35) [Msf631592.1_b]
(36) [Mp577446]
(37) [Mp577445]
(38) [Mp577448_b]
(39) MtJemalong
80.1
65.8
64.1
66
62.5
63.7
63.7
63.7
63.7
61.7
88.2
62.8
62.3
67
63.1
76.7
62.9
83.4
60.8
63.7
86
63.3
96.3
65.4
85.1
61
65.9
100
100
64.5
63.5
62.8
62.3
72.5
67.3
67
61.5
66.2
67.4
81.3

FIGURE 3.3. CNGC5 allele statistical parsimony network of *Medicago*. Each box with a wide border represents an allele and contains one or more individuals that possess that allele. Taxa sharing a haplotype are separated by thin lines. Tetraploids are in boldface, italic characters. Different individuals from the same accession and alleles found in single individuals are denoted as in Figure 3.1. The numbers in triangles are allele numbers corresponding to those at the terminals in Figures 3.1 and 3.2. *Medicago truncatula* was not connected to the network under the 95% statistical parsimony criterion of TCS. Dashed line denotes where an allele from *M. truncatula* would connect to the network when connection limit was set to 20 steps. Abbreviations are as in Figure 3.1 with the addition of Msv(h), *M. s.* subsp. ×*varia* originally listed as *M. s.* subsp. ×*hemicycla* in GRIN, but found to be tetraploids by flow cytometry; CADL = cultivated alfalfa at the diploid level.

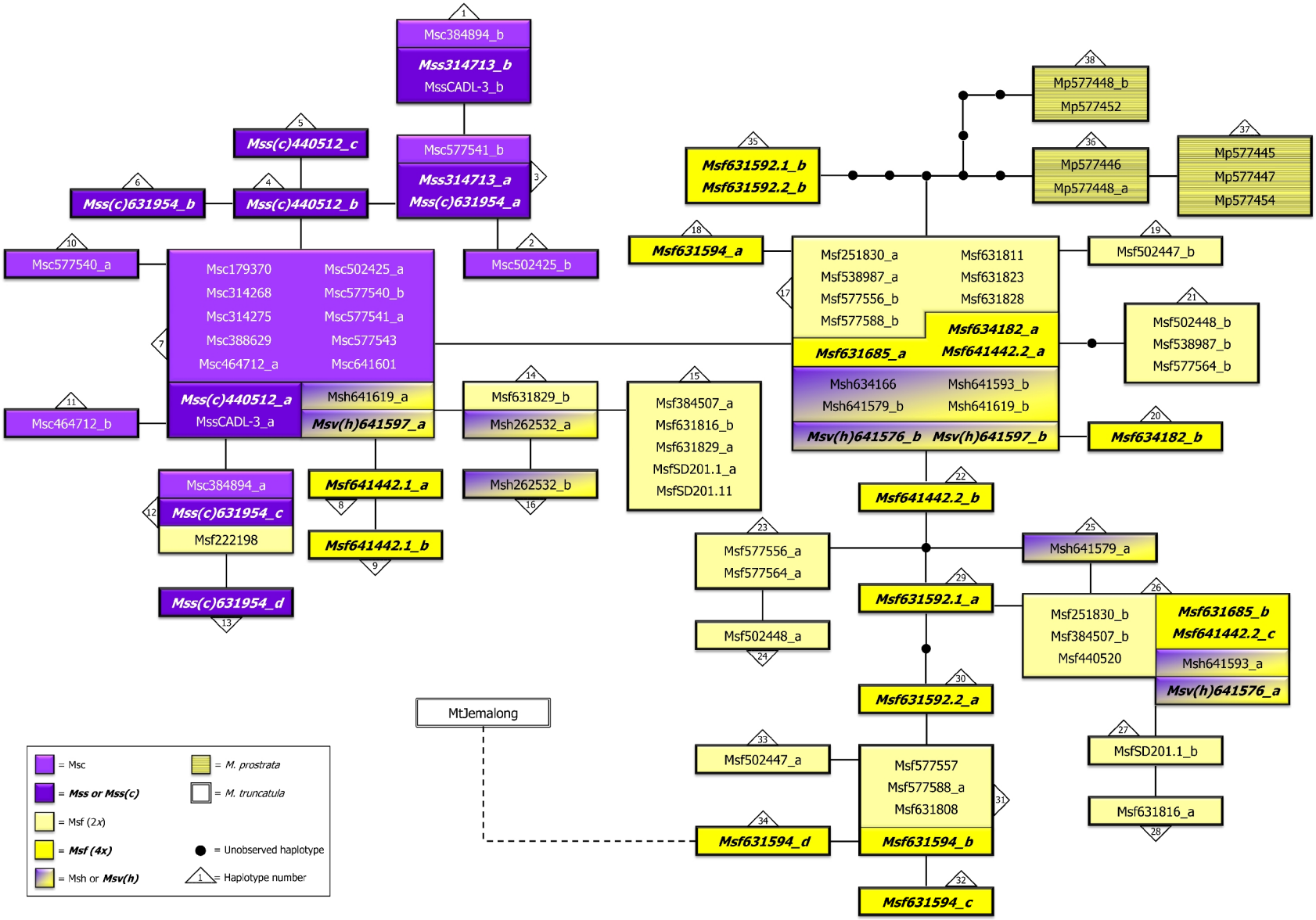

1
Msc384894_b
Mss314713_b
MssCADL-3_b
5
Mss(c)440512_c
Msc577541_b
Mss314713_a
Mss(c)631954_a
3
6
Mss(c)631954_b
4
Mss(c)440512_b
2
Msc502425_b
10
Msc577540_a
Msc179370
Msc314268
Msc314275
Msc388629
Msc464712_a
Msc502425_a
Msc577540_b
Msc577541_a
Msc577543
Msc641601
7
11
Msc464712_b
Mss(c)440512_a
MssCADL-3_a
Msh641619_a
Msv(h)641597_a
14
Msf631829_b
Msh262532_a
15
Msf384507_a
Msf631816_b
Msf631829_a
MsfSD201.1_a
MsfSD201.11
Msc384894_a
12
Mss(c)631954_c
Msf222198
Mss(c)631954_d
13
8
Msf641442.1_a
9
Msf641442.1_b
16
Msh262532_b
38
Mp577448_b
Mp577452
35
Msf631592.1_b
Msf631592.2_b
36
Mp577446
Mp577448_a
37
Mp577445
Mp577447
Mp577454
18
Msf631594_a
Msf251830_a
Msf538987_a
Msf577556_b
Msf577588_b
Msf631811
Msf631823
Msf631828
17
Msf631685_a
Msf634182_a
Msf641442.2_a
Msh634166
Msh641579_b
Msh641593_b
Msh641619_b
Msv(h)641576_b
Msv(h)641597_b
19
Msf502447_b
21
Msf502448_b
Msf538987_b
Msf577564_b
20
Msf634182_b
22
Msf641442.2_b
23
Msf577556_a
Msf577564_a
24
Msf502448_a
25
Msh641579_a
29
Msf631592.1_a
26
Msf251830_b
Msf384507_b
Msf440520
Msf631685_b
Msf641442.2_c
Msh641593_a
Msv(h)641576_a
30
Msf631592.2_a
27
MsfSD201.1_b
28
Msf631816_a
MtJemalong
33
Msf502447_a
34
Msf631594_d
Msf577557
Msf577588_a
Msf631808
31
Msf631594_b
32
Msf631594_c
= Msc
= Mss or Mss(c)
= Msf (2x)
= Msf (4x)
= Msh or Msv(h)
= M. prostrata
= M. truncatula
= Unobserved haplotype
1 = Haplotype number

TABLE 3.4. Analysis of molecular variance (AMOVA) of genetic variation in CNGC5.

Source of variation	df	% of total variation	Fixation index (Φ_{ST})	*P*
Caerulea vs. diploid falcata				
Among taxa	1	23.72	0.237	<0.0001
Within taxa	44	76.28		
Caerulea vs. hemicycla				
Among taxa	1	23.85	0.238	<0.0001
Within taxa	23	76.15		
Diploid falcata vs. hemicycla				
Among taxa	1	−2.16	−0.022	>0.50
Within taxa	37	102.16		
Caerulea vs. sativa [a]				
Among taxa	1	8.48	0.084	>0.05
Within taxa	25	91.52		
Diploid falcata vs. tetraploid falcata				
Among taxa	1	0.98	0.009	>0.10
Within taxa	45	99.02		
Tetraploid falcata vs. sativa [a]				
Among taxa	1	35.09	0.351	<0.0001
Within taxa	26	64.91		
Sativa [a] vs. varia [b]				
Among taxa	1	31.53	0.315	<0.005
Within taxa	13	68.47		
Tetraploid falcata vs. varia [b]				
Among taxa	1	−8.60	−0.086	>0.50
Within taxa	19	108.60		
Diploid falcata vs. sativa [a]				
Among taxa	1	28.47	0.285	<0.0001
Within taxa	39	71.53		
M. prostrata vs. caerulea				
Among taxa	1	57.44	0.574	<0.0001
Within taxa	21	42.56		
M. prostrata vs. sativa [a]				
Among taxa	1	52.41	0.524	<0.0001
Within taxa	16	47.59		
M. prostrata vs. diploid falcata				
Among taxa	1	43.30	0.433	<0.0001
Within taxa	35	56.70		
M. prostrata vs. tetraploid falcata				
Among taxa	1	39.46	0.395	<0.0001
Within taxa	22	60.54		

TABLE 3.4 (Continued)

[a] Listed as *M. s.* subsp. *caerulea* in GRIN, but found to be tetraploids by flow cytometry and listed here as *M. s.* subsp. *sativa*

[b] Listed as *M. s.* subsp. ×*hemicycla* in GRIN, but found to be tetraploids by flow cytometry and listed here as *M. s.* subsp. ×*varia*.

Discussion

Relationships among diploids in the* M. sativa *complex—Nuclear low copy gene data presented here confirm the pattern of relationships among diploid members of the *M. sativa* complex revealed by cpDNA data (Havananda et al., 2010, 2011) and nuclear SSR markers (Şakiroğlu et al., 2010), which also agree with the distribution of morphological characters. The nDNA allele network showed that all alleles from plants with violet flowers and coiled pods, classified as caerulea, formed a group separate from most alleles from plants with yellow flowers and sickle-shaped pods, classified as diploid falcata. Although diploid falcata had an allele (allele 12) shared with caerulea, and two other alleles (alleles 14 and 15) more similar to the most common caerulea allele (allele 7), it was genetically differentiated from caerulea (Table 3.4). Interestingly, the diploid falcata individual (Msf PI 222198) that shared allele 12 with caerulea also possesses a caerulea-like chloroplast haplotype (Havananda et al., 2010). This is in contrast with data from nuclear SSR marker showing that a different individual from the same accession possesses a falcata genome composition and is classified in the falcata highland ecotype group (Şakiroğlu et al., 2010). The presence of a caerulea chloroplast haplotype and a caerulea nuclear allele in this morphologically "falcata" diploid accession could be due to introgression, involving caerulea as a paternal parent. It should be noted that, based on the accession information in GRIN (http://www.ars-grin.gov/cgi-bin/npgs/acc/display.pl?1179727, accessed October 4, 2011), this accession is not wild material, has an uncertain improvement status, and was

identified as caerulea prior to the work of Şakiroğlu et al. (2010). Removal of this questionable accession would increase support for genetic differentiation between the two diploid taxa.

The observation that caerulea and diploid falcata share identical or closely related alleles is not unexpected, given possibility of gene flow between them due to their incomplete reproductive isolation and the observation that hybridization is a pervasive and ongoing process throughout the history of *Medicago* (Maureira et al., 2008). The finding that hemicycla, a putative natural hybrid between caerulea and diploid falcata (Lesins and Lesins, 1979; Quiros and Bauchan, 1988), shares CNGC5 alleles with both of its parents supports the hypothesis of its hybrid origin, and documents the possibility of introgression involving caerulea and diploid falcata.

Polyploid origins—Tetraploids in the *M. sativa* complex are hypothesized to be genetic autopolyploids, exhibiting tetrasomic inheritance (Stanford, 1951; Quiros, 1982; McCoy and Bingham, 1988) and occasional quadrivalents at meiosis (Armstrong, 1971; Gillies, 1972; Mariani and Veronesi, 1979). CNGC5 nuclear DNA data support the hypothesis that both sativa and tetraploid falcata are taxonomic autopolyploids: there are two major groups of alleles, one group consisting of all alleles from caerulea and sativa individuals, and a second group comprising most alleles from diploid and tetraploid falcata individuals (Figure 3.3). AMOVA results support this conclusion, indicating no significant differentiation between caerulea and sativa for CNGC5 alleles, or between diploid and tetraploid falcata (Table 3.4).

The presence of multiple identical or closely related alleles in the caerulea/sativa and diploid/tetraploid falcata pairs could be due to multiple origins of autopolyploidy, a phenomenon also known to occur in many other autopolyploid species (e.g., Soltis and Rieseberg, 1986; Ness et al., 1989; Yang et al., 2006; Frizzi et al., 2007). However, a single origin followed by

interploidal gene flow could also result in multiple shared alleles in the diploid/tetraploid pairs. Soltis et al. (2010), referring to Stebbins (1971), noted that interploidal gene flow could occur through two pathways: (1) through the existence of triploid hybrids that may produce tetraploid progeny allowing gene flow between the diploid progenitor and the tetraploid; or (2) through unreduced ($2n = 2x = 16$) gametes produced by the diploid that can unite with reduced ($1n = 2x = 16$) gametes normally produced from the tetraploid, resulting in a tetraploid hybrid. In the *M. sativa* complex, triploids are rare, suggesting a very effective triploid block in the complex (Veronesi et al., 1986; McCoy and Bingham, 1988). It is more probable that interploidal gene flow in the complex occurs through unreduced gametes, whose existence has been shown in many studies (e.g., Bingham, 1968; McCoy, 1982; McCoy and Bingham, 1988).

The finding that alleles in varia are identical to alleles in sativa or tetraploid falcata is consistent with its hybrid origin at the tetraploid level by crosses between sativa and tetraploid falcata (Quiros and Bauchan, 1988). This hybridization suggests that gene flow at the polyploid level is possible and could lead to lineage recombination (Doyle et al., 1999) in which new genotypes result from the reassortment of characters from polyploids originating independently from different diploid genotypes. Alleles 8 and 9 in tetraploid falcata plants that are closely related to sativa alleles could be the outcomes of lineage recombination. Because no tetraploid falcata nDNA alleles were observed in sativa plants, gene flow at the tetraploid level may be unidirectional from sativa into tetraploid falcata; however, our sample of sativa was small. The high level of polymorphism in diploid falcata (several alleles in typical falcata group and several individuals with alleles in the caerulea/sativa group) could suggest that the presence of alleles 8 and 9 in tetraploid falcata could represent additional origins of polyploidy. However, it is difficult to estimate the number of polyploidy events because distinguishing between multiple

origins and interploidal gene flow after polyploidization is complicated, with both having the effect of increasing genetic variation in polyploids (e.g., Soltis and Soltis, 2009; Soltis et al., 2010).

Medicago sativa *complex and* M. prostrata—*Medicago prostrata* and the *M. sativa* complex are placed in section *Medicago*, subsection *Medicago*. The key morphological characters of *M. prostrata*, perennial habit with yellow flowers and coiled, non-spiny pods that have gland-tipped trichomes, are very similar to those of glomerata, but *M. prostrata* has long reflexed pedicels that glomerata lacks (Small, 2011). Evidence that *M. prostrata* is genetically closely related to the *M. sativa* complex was provided by successful artificial crosses, but with observed interspecific barriers when *M. prostrata* was the maternal parent (Lesins, 1962, 1968; Sorensen et al., 1980). In addition, Lesins and Lesins (1979, p. 95) reported plants appearing to be natural hybrids of falcata and *M. prostrata* in Italy. The relationship of *M. prostrata* to the *M. sativa* complex has taken on increased relevance with the suggestion by Havananda et al. (2011) that *M. prostrata* may have been involved in the origin of tetraploid falcata. They found that over 90% of the tetraploid falcata individuals sampled had chloroplast haplotypes that were identical or most similar to haplotypes found in *M. prostrata*, and were more distantly related to haplotypes common in diploid falcata. Although this pattern of chloroplast haplotype variation could be due to lineage sorting or to hybridization/introgression, Havananda et al. (2011) asked "whether diploid falcata is the direct progenitor of tetraploid falcata, as morphology would suggest, or whether *M. prostrata* played a role in the origin of tetraploid falcata beyond donating a few introgressed genes." The CNGC5 data show that, at least at this nuclear locus, *M. prostrata* and tetraploid falcata (and other *M. sativa* taxa) have significantly differentiated alleles, suggesting that the nuclear genome of tetraploid falcata is not derived directly from *M. prostrata*,

but more probably from diploid falcata through autopolyploidy as previously hypothesized. The presence of *M. prostrata* chloroplast haplotypes in tetraploid falcata is most likely due to introgression, given the weakness of barriers to crossing involving these taxa specifically (see above) and in *Medicago* more generally (Maureira-Butler et al., 2008).

Autopolyploidy, speciation, and taxonomy of the* M. sativa *complex—Soltis et al. (2007) suggested that many autopolyploids do not receive taxonomic recognition as distinct species from their diploid progenitors despite fulfilling the requirements of several species concepts. They also suggested that recognition and naming of diploids and autopolyploids as distinct species should occur only after case-by-case, careful studies to determine whether they meet criteria for species recognition. Like many examples given in Soltis et al. (2007), the *M. sativa* complex contains diploid and autotetraploid cytotypes currently classified as a single species (Small and Jomphe, 1989; Small, 2011). However, taxonomic considerations concerning polyploidy are tied to the issue of how many diploid species to recognize. The taxonomy of the complex has long been controversial. There are two main morphological groups, comprising plants with violet flowers and coiled pods (caerulea-sativa) and plants with yellow flowers and falcate pods (falcata), plus intermediates between them. Both groups and their putative hybrids have diploid and tetraploid cytotypes (Figure 3.1). It has been debated whether falcata should be considered a distinct species or a subspecies of *M. sativa*. Lesins and Lesins (1979) considered falcata as a separate species, whereas Quiros and Bauchan (1988), Small and Jomphe (1989), and Small (2011) ranked it at the intraspecific level.

The two morphologically differentiated diploids, though sympatric, occupy different ecological niches: caerulea, with coiled pods, is hypothesized to be adapted for dispersal by rolling on open ground in semidesert conditions, whereas diploid falcata has straighter pods

thought to be adapted for easier seed dispersal in the more closed, denser plant community of steppes, and is adapted to colder environments (Lesins and Lesins, 1979). Genetic differentiation between the two diploids was detected based on cpDNA (Havananda et al., 2010, 2011), nuclear SSR (Şakiroğlu et al., 2010), and the nDNA sequence data presented here. However, neither their chloroplast haplotypes nor nuclear alleles form monophyletic groups. The putative hybrid of these two diploids, hemicycla, is observed in nature (Lesins and Lesins, 1979; Small and Bauchan, 1984 [as varia]) and artificial crosses between caerulea and diploid falcata yield viable seeds (Mariani and Veronesi, 1979), suggesting that these two diploids are not reproductively isolated. Thus, although they are differentiated and their recognition as separate taxa is merited, we maintain our previous suggestion of "continuing the more recent practice of regarding them as subspecies" (Havananda et al., 2010). If caerulea and diploid falcata were recognized as separate species, then it would be reasonable to ask whether either of the autopolyploids derived from each should also be recognized at the specific level.

Recognition of subspecies is typically based on morphological characters, but conflicting phylogenetic patterns from genetic variation revealed in recent studies has caused concern that such characters may not reflect underlying genetic structure and phylogenies (Haig et al., 2006). Though our data, as well as chloroplast DNA data (Havananda et al., 2011), agree with the intraspecific recognition of the group of plants with purple flowers and coiled pods separate from the group of plants with yellow flowers and falcate pods in the *M. sativa* complex, they are not always concordant with the subspecific designation of taxa within each group and of the hybrid taxa defined by morphology in relation to ploidy. For the group of plants with violet flowers and coiled pods, the two recognized subspecies, caerulea and autotetraploid sativa, are quantitatively morphologically differentiated. Yet, the present data from nuclear genome, agree with the

chloroplast data in suggesting that autotetraploid sativa and its putative diploid progenitor, caerulea, are not genetically differentiated from one another. As for the group of plants with yellow flowers and falcate pods in the complex, the two falcata cytotypes are not morphologically distinguishable, and hence have been recognized as a single subspecies (Small, 1985, 2011) despite their ploidal difference. The differentiation and the lack of gene flow at the chloroplast genome level between diploid and tetraploid falcata reported by Havananda et al. (2011) could argue for taxonomic recognition of the two cytotypes. The lack of differentiation in their nuclear genomes suggested by the data presented here, on the other hand, agrees with their lack of morphological differentiation and suggests that the presence of different chloroplast haplotypes in tetraploid falcata is likely due to introgression.

The distinct intermediate morphology of the two putative hybrid taxa, diploid hemicycla and tetraploid varia, led to their recognition as distinct subspecies (e.g., Quiros and Bauchan, 1988 and Small, 2011). Nuclear SSR data from a more extensive sampling of diploid taxa obtained by Şakiroğlu et al. (2010) revealed that although many hemicycla individuals had admixture patterns consistent with early generation hybrids between caerulea and falcata, other accessions formed a distinct group consistent with their forming a genetically distinct taxon diverged from both putative parents. However, no genetic differentiation was detected between these taxa and their presumed parents, for either nDNA (this study) or cpDNA (Havananda et al., 2011). Based on our results, hemicycla and varia appear to be early-generation hybrids, and subsequent gene flow between them and their parents seems to be limited and is not homogenizing the allelic variation between the two parental morphological groups, as suggested by genetic differentiation both between the diploid parental taxa of hemicycla (caerulea and diploid falcata: Havananda et al., 2010, 2011 [cpDNA], Şakiroğlu et al., 2010 [nuclear SSR], this

study [nDNA]) and between the putative tetraploid parents of varia (sativa and tetraploid falcata: Havananda et al., 2011 [cpDNA], this study [nDNA]). These findings make the recognition of hybrid subspecies questionable. If they are recognized, whether they should be segregated into separate subspecies according to ploidy—diploid hemicycla and tetraploid varia—is also debatable. Quiros and Bauchan (1988) considered them separate taxa based mainly on ploidy, whereas Small and Jomphe (1989) and Small (2011) considered both cytotypes as varia because they are not morphologically distinguishable from one another, a situation similar to diploid and tetraploid falcata. Because there is no nuclear allele differentiation between the two cytotypes of either of the parents, i.e. caerulea vs. sativa and diploid vs. tetraploid falcata, hybrids at both ploidal levels it is not surprising that the hybrid taxa are not differentiated, and in fact share CNGC5 alleles. The nuclear DNA data, as well as chloroplast results (Havananda et al., 2011), hence agree with the recognition of a single hybrid subspecies, varia, with two morphologically indistinguishable cytotypes suggested by Small and Jomphe (1989).

The conflict among morphological characters, ploidy, and molecular data makes suggesting subspecific boundaries among taxa in the *M. sativa* complex a difficult issue. Due to the overlapping natural distributions of caerulea and falcata (both cytotypes), and the widespread ranges of sativa and varia due to humans (Small, 2011), subspecific boundaries of these taxa are very hazy. There is yet additional complexity in the complex: It will be interesting to include glomerata—another non-hybrid taxon that has a distinctive combination of characters, having yellow flowers of falcata and coiled pod similar to caerulea/sativa, and which includes both diploid and tetraploid cytotypes. Inclusion of tunetana, a tetraploid hybrid between tetraploid glomerata and sativa, could also lead to better understanding of the history of the complex.

Conclusions—Data from a low-copy nuclear gene, CNGC5, show that the two main morphologically differentiated members of the *M. sativa* complex, caerulea/sativa and diplioid/tetraploid falcata, are genetically differentiated at both the diploid and tetraploid levels, with only limited gene flow between them at their respective ploidal levels. The nuclear DNA data are concordant with hypothesized hybrid origins of hemicycla and varia, which appear to be early-generation hybrids and probably do not act as a genetic bridge between their putative non-hybrid parents. In addition, the data also show a "classic" autopolyploid pattern, possibly involving multiple origins and/or interploidal gene flow, in both caerulea/sativa and diploid/tetraploid falcata pairs. There is no evidence based on nuclear data that *M. prostrata* was the direct progenitor of tetraploid falcata as was suggested by Havananda et al. (2011) as one explanation for the surprising differentiation of cpDNA haplotypes between diploid and tetraploid falcata. Instead, the data from CNGC5 support the alternative hypothesis of Havananda et al. (2011): autopolyploid origin of tetraploid falcata from diploid falcata with later introgression, probably at the tetraploid level, from *M. prostrata*.

Our data, including those of Havananda et al. (2010, 2011), agree with the recognition of caerulea and diploid falcata at the subspecific level as suggested by morphology. For autotetraploid and hybrid taxa of the complex, there is conflicting evidence regarding their subspecific recognition. Elucidating the evolutionary history and taxonomic treatment of the complex requires further studies with more extensive sampling of accessions, especially tetraploids and hybrids, and additional nuclear genes.

REFERENCES

ADAMS, K. L. AND J. F. WENDEL. 2005. Novel patterns of gene expression in polyploid plants. *Trends in Genetics* 21: 539–543.

ARMSTRONG, K. C. 1971. Chromosome associations at pachytene and metaphase in *Medicago sativa*. *Canadian Journal of Genetics and Cytology* 13: 697–702.

BINGHAM, E. T. 1968. Transfer of diploid *Medicago* spp. germplasm to tetraploid *M. sativa* L. in 4x–2x crosses. *Crop Science* 8: 760–762.

BINGHAM, E. T. 1990. Backcrossing tetraploidy into diploid *Medicago falcata* L. using 2*n* eggs. *Crop Science* 30: 1353–1354.

BINGHAM, E. T. 1993. Registration of WISFAL alfalfa (*Medicago sativa* subsp. *falcata*) tetraploid germplasm derived from diploids. *Crop Science* 33: 217–218.

BINGHAM, E. T., AND T. J. MCCOY. 1979. Cultivated alfalfa at the diploid level: Origin, reproductive stability, and yield of seed and forage. *Crop Science* 19: 97–100.

BOE, A., R. BORTNEM, K. F. HIGGINS, A. D. KRUSE, K. D. KEPHART, AND S. SELMAN. 1998. Breeding yellow-flowered alfalfa for combined wildlife habitat and forage purposes. South Dakota Agricultural Experiment Station Bulletin 727, South Dakota State University, Brookings, South Dakota, USA.

BRUEN, T. C., H. PHILIPPE, AND D. BRYANT. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* 172: 2665–2681.

CLEMENT, M., D. POSADA, AND K. A. CRANDALL. 2000. TCS: A computer program to estimate gene genealogies. *Molecular ecology* 9: 1657–1659.

DOYLE, J. J., AND J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

DOYLE, J. J., AND A. N. EGAN. 2010. Dating the origins of polyploidy events. *New Phytologist* 186: 73–85.

DOYLE, J. J., L. E. FLAGEL, A. H. PATERSON, R. A. RAPP, D. E. SOLTIS, P. S. SOLTIS, AND J. F. WENDEL. 2008. Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics* 42: 443–461.

ELLSTRAND, N. C., H. C. PRENTICE, AND J. F. HANCOCK. 1999. Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics* 30: 539–563.

EXCOFFIER, L., AND H. E. L. LISCHER. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.

FRIZZI, G., M. MIRANDA, C. PANTANI, AND F. TAMMARO. 2007. Allozyme differentiation in four species of the *Crocus cartwrightianus* group and in cultivated saffron (*Crocus sativus*). *Biochemical Systematics and Ecology* 35: 859–868.

GILLIES, C. B. 1972. Pachytene chromosomes of perennial *Medicago* species. I. Species closely related to *M. sativa*. *Hereditas* 72: 277–288.

GUINDON, S., J. DUFAYARD, V. LEFORT, M. ANISIMOVA, W. HORDIJK, AND O. GASCUEL. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology* 59: 307–321.

HAIG, S. M., E. A. BEEVER, S. M. CHAMBERS, H. M. DRAHEIM, B. D. DUGGER, S. DUNHAM, E. ELLIOTT-SMITH, ET AL. 2006. Taxonomic considerations in listing subspecies under the U.S. Endangered Species Act. *Conservation Biology* 20: 1584–1594.

HAVANANDA, T., E. C. BRUMMER, AND J. J. DOYLE. 2011. Complex patterns of autopolyploid evolution in alfalfa and allies (*Medicago sativa*; Leguminosae). *American Journal of Botany* 98: 1633–1646.

HAVANANDA, T., E. C. BRUMMER, I. MAUREIRA-BUTLER, AND J. J. DOYLE. 2010. Relationships among diploid members of the *Medicago sativa* (Fabaceae) species complex based on chloroplast and mitochondrial DNA sequences. *Systematic Botany* 35: 140–150.

HUSON, D. H., AND D. BRYANT. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254–267. [Software available from http://www.splitstree.org/]

JIAO, Y., N. J. WICKETT, S. AYYAMPALAYAM, A. S. CHANDERBALI, L. LANDHERR, P. E. RALPH, L. P. TOMSHO ET AL. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97–100.

LESINS, K. 1962. Interspecific crosses involving alfalfa. III. *Medicago sativa* L. × *M. prostrata* Jacq. *Canadian Journal of Genetics and Cytology* 4: 14–23.

LESINS, K. 1968. Interspecific crosses involving alfalfa. IV. *Medicago glomerata* × *M. sativa* with reference to *M. prostrata*. *Canadian Journal of Genetics and Cytology* 10: 536–544.

LESINS, K. A. AND I. LESINS. 1979. Genus *Medicago* (Leguminosae): A taxogenetic study. Dr. W. Junk, The Hague, Netherlands.

LIBRADO, P., AND ROZAS, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25**:** 1451–1452.

MARIANI, A., AND F. VERONESI. 1979. Cytological and fertility relationships of different *Medicago* species and cytogenetic behavior of their hybrids. *Genetica Agraria* 33: 245–268.

MARTIN, N. P., AND H.-J. G. JUNG. 2010. Alfalfa: Potential for new feed and biofuel–USDFRC research update. Proceedings of the Winter Seed School Conference, Las Vegas, Nevada, 17–19 January, 105–111. The Northwest Alfalfa Seed Growers Association, Kennewick, Washington, USA. Also at website http://www.ars.usda.gov/sp2UserFiles/Place/36553000/pdf%27s/Alfalfa-potentialfornewfeedfuel_final_1-10-10.pdf.

MASTERSON, J. 1994. Stomatal size in fossil plants: Evidence for polyploidy in majority of angiosperms. *Science* 264: 421–424.

MAUREIRA-BUTLER, I. J., B. E. PFEIL, A. MUANGPROM, T. C. OSBORN, AND J. J. DOYLE. 2008. The reticulate history of *Medicago* (Fabaceae). *Systematic Biology* 57: 466–482.

MCCASLIN, M., AND D. MILLER. 2007. The future of alfalfa as a biofuels feedstock. Proceedings of the 37th California Alfalfa & Forage Symposium, Monterey, California, 17–19 December. UC Cooperative Extension, Agronomy Research and Information Center, Plant Sciences Department, University of California, Davis, California, USA. Also at website http://alfalfa.ucdavis.edu/2007AlfalfaConference/2007/07-43.pdf.

MCCOY, T. J. 1982. The inheritance of 2*n* pollen formation in diploid alfalfa *Medicago sativa*. *Canadian Journal of Genetics and Cytology* 24: 315–323.

MCCOY, T. J., AND E. T. BINGHAM. 1988. Cytology and cytogenetics of alfalfa. *In* A. A. Hanson [ed.], D. K. Barnes and R. R. Hill, Jr. [co-eds.], Alfalfa and alfalfa improvement, Agronomy monograph 29, 737–776. American Society of Agronomy, Crop Science Society of Agronomy, and Soil Science Society of Agronomy, Madison, Wisconsin, USA.

MUANGPROM, A., S. G. THOMAS, T. P. SUN, AND T. C. OSBORN. 2005. A novel dwarfing mutation in a green revolution gene from *Brassica rapa*. *Plant Physiology* 137: 931–938.

MÜLLER, K. 2005. SeqState: Primer design and sequence statistics for phylogenetic DNA datasets. *Applied Bioinformatics* 4: 65–9.

MULLER, M.-H., C. PONCET, J. M. PROSPERI, S. SANTONI, AND J. RONFORT. 2006. Domestication history in the *Medicago sativa* species complex: Inferences from nuclear sequence polymorphism. *Molecular Ecology* 15: 1589–1602.

NESS, B. D., D. E. SOLTIS, AND P. S. SOLTIS. 1989. Autopolyploidy in *Heuchera micrantha* (Saxifragaceae). *American Journal of Botany* 76: 614–626.

OTTO, S. P., AND J. WHITTON. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* 34: 401–437.

QUIROS, C. F. 1982. Tetrasomic segregation for multiple alleles in alfalfa. *Genetics* 101: 117–127.

QUIROS, C. F., AND G. R. BAUCHAN. 1988. The genus *Medicago* and the origin of the *Medicago sativa* complex. *In* A. A. Hanson [ed.], D. K. Barnes and R. R. Hill, Jr. [co-eds.], Alfalfa and alfalfa improvement, Agronomy monograph 29, 93–124. American Society of Agronomy, Crop Science Society of Agronomy, and Soil Science Society of Agronomy, Madison, Wisconsin, USA.

RAMBAUT, A. 2006–2009. FigTree, version 1.3.1. Institute of Evolutionary Biology, University of Edinburgh. [Software available from http://tree.bio.ed.ac.uk/software/figtree/]

RAMSEY, J., AND D. W. SCHEMSKE. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* 29: 467–501.

RUSSELLE, M.P. 2001. Alfalfa: After an 8,000-year journey, the "Queen of Forages" stands poised to enjoy renewed popularity. *American Scientist* 89: 252–261.

ŞAKIROĞLU, M., J. J. DOYLE, AND E. C. BRUMMER. 2010. Inferring population structure and genetic diversity of broad range of wild diploid alfalfa (*Medicago sativa* L.) accessions using SSR markers. *Theoretical and Applied Genetics* 121: 403–415.

SIMMONS, M. P., AND H. OCHOTERENA. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369–381.

SKINNER, D. Z. 2000. Non random chloroplast DNA hypervariablility in *Medicago sativa*. *Theoretical and Applied Genetics* 101: 1242–1249.

SMALL, E. 1985. Morphological differentiation in *Medicago sativa* s.l. in relation to ploidy. *Canadian Journal of Botany* 63: 1747–1752.

SMALL, E. 2011. Alfalfa and relatives: Evolution and classification of *Medicago*. NRC Research Press, Ottawa, Ontario, Canada.

SMALL, E., AND G. R. BAUCHAN. 1984. Chromosome numbers of the *Medicago sativa* complex in Turkey. *Canadian Journal of Botany* 62: 749–752.

SOLTIS, D. E., V. A. ALBERT, J. LEEBENS-MACK, C. D. BELL, A. H. PATERSON, C. ZHENG, D. SANKOFF, ET AL. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96: 336–348.

SOLTIS, D. E., R. J. A. BUGGS, J. J. DOYLE, AND P. S. SOLTIS. 2010. What we still don't know about polyploidy. *Taxon* 59: 1387–1403.

SOLTIS, D. E., AND L. H. RIESEBERG. 1986. Autopolyploidy in *Tolmiea menziesii* (Saxifragaceae): Genetic insights from enzyme electrophoresis. *American Journal of Botany* 73: 310–318.

SOLTIS, D. E., P. S. SOLTIS, D. W. SCHEMSKE, J. F. HANCOCK, J. N. THOMPSON, B. C. HUSBAND, AND W. S. JUDD. 2007. Autopolyploidy in angiosperms: Have we grossly underestimated the number of species? *Taxon* 56: 13–30.

SOLTIS, P. S., AND D. E. SOLTIS. 2009. The role of hybridization in plant speciation. *Annual Review of Plant Biology* 60: 561–588.

SORENSEN, E. L., N. SANGDUEN, G. H. LIANG. 1980. Transfer of glandular hairs from diploid *Medicago prostrata* to tetraploid *M. sativa*. *In* D. K. Barnes [ed.], Report of The Twenty-seventh Alfalfa Improvement Conference, Madison, Wisconsin, 8–10 July, 53. Agricultural Research, North Central Region, USDA-SEA, Peoria, Illinois, USA.

STANFORD, E. H. 1951. Tetrasomic inheritance in alfalfa. *Agronomy Journal* 43: 222–225.

STEELE, K. P., S. M. ICKERT-BOND, S. ZARRE, AND M. F. WOJCIECHOWSKI. 2010. Phylogeny and character evolution in *Medicago* (Leguminosae): Evidence from analyses of plastid *trnK*/*matK* and nuclear *GA3ox1* sequences. *American Journal of Botany* 97: 1142–1155.

TEMPLETON, A. R., K. A. CRANDALL, AND C. F. SING. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132: 619–633.

VERONESI, F., A. MARIANI, AND E. T. BINGHAM. 1986. Unreduced gametes in diploid *Medicago* and their importance in alfalfa breeding. *Theoretical and Applied Genetics* 72: 37–41.

VORSA, N., AND E. T. BINGHAM. 1979. Cytology of 2*n* pollen formation in diploid alfalfa, *Medicago sativa*. *Canadian Journal of Genetics and Cytology* 21: 525–530.

WENDEL, J. F. 2000. Genome evolution in polyploids. *Plant Molecular Biology* 42: 225–249.

WOOD, T. E., N. TAKEBAYASHI, M. S. BARKER, I. MAYROSE, P. B. GREENSPOON, AND L. H. RIESEBERG. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences, USA* 106: 13875–13879.

YANG, W., B. J. GLOVER, G. RAO, AND J. YANG. 2006. Molecular evidence for multiple polyploidization and lineage recombination in the *Chrysanthemum indicum* polyploid complex (Asteraceae). *New Phytologist* 171: 875–886.

YOUNG, N. D., F. DEBELLE, G. E. D. OLDROYD, R. GEURTS, S. B. CANNON, M. K. UDVARDI, V. A. BENEDITO, ET AL. 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480: 520–524.

APPENDIX 1*

Information concerning plant materials and DNA sequences used in Chapter 1 is presented as follows: **taxon name accession number (USDA Plant Introduction [PI])**; original locality of collection; voucher information (if available); GenBank accession numbers: *rpl20-rps12* spacer, *trnS-trnG* spacer, *nad4* intron, *nad7* intron, *rpS14-cob* spacer (& joins two different sequences from the same chloroplast region submitted for single heteroplasmic individuals, — = sequence not obtained). Taxon names are abbreviated as follows: Mp = *Medicago prostrata* (Jacq.); Msc = *M. sativa* L. subsp. *caerulea* (Less. ex Ledeb.) Schmalh. ("caerulea"); Msf = *M. s.* subsp. *falcata* (L.) Arcang. ("falcata"); Msg = *M. s.* subsp. *glomerata* (Balb.) Rouy; Msh = *M. s.* subsp. *hemicycla* (Grossh.) C. R. Gunn; Mss = *M. s.* subsp. *sativa*; Mt = *M. truncatula* Gaertn. CADL = cultivated alfalfa at the diploid level. All plants were grown in greenhouses at Iowa State University and/or the University of Georgia, except for Msg PI 577567 which was grown at Cornell University. Voucher specimens are deposited in the University of Georgia Herbarium (GA) or Herbarium of the L. H. Bailey Hortorium (BH). Collector's abbreviations: MS, Muhammet Şakiroğlu; TH, Tee Havananda. Note: [a] heteroplasmic for chloroplast region(s); [b] listed as *M. s.* subsp. *caerulea* in the Germplasm Resources Information Network (GRIN); [c] listed as *M. s.* subsp. *sativa* in GRIN; [d] listed as *M. s.* subsp. *falcata* in GRIN.

Mp PI 577445; Italy; FJ652795, FJ695087, FJ694976, FJ695013, FJ695050. **Mp PI 577448**; Italy; FJ652796, FJ695088, FJ694977, FJ695014, FJ695051. **Mp PI 577452**; Italy; FJ652797, FJ695089, FJ694978, FJ695015, FJ695052. **Mp PI 577454**; Greece; FJ652798, FJ695090, FJ694979, FJ695016, FJ695053. **Msc PI 179370**; Turkey; *MS 1* (GA); FJ652799, FJ695091, —, —, —. **Msc PI 314268**; Uzbekistan; FJ652801, FJ695092, FJ694980, FJ695017, FJ695054. **Msc PI 314275**; Uzbekistan; *MS 37* (GA); FJ652802, FJ695093, —, —, —. **Msc PI 384894**; Iran; FJ652803, FJ695094, FJ694981, FJ695018, FJ695055. **Msc PI 388629**; Uzbekistan; *MS GH1* (GA); FJ652804, FJ695095, FJ694982, FJ695019, FJ695056. **Msc PI 440501**; Kazakhstan; *MS 53* (GA); FJ652805, FJ695096, —, —, —. **Msc PI 464712**; Turkey; *MS 72* (GA); FJ652806, FJ695097, FJ694983, FJ695020, FJ695057. **Msc PI 502425**; Russia; FJ652807, FJ695098, FJ694984, FJ695021, FJ695058. **Msc PI 577540**; Turkey; FJ652808, FJ695099, FJ694985, FJ695022, FJ695059. **Msc PI 577541**; Kazakhstan; *MS 274* (GA); FJ652809, FJ695100, FJ694986, FJ695023, FJ695060. **Msc PI 577543**; Georgia; *MS 181* (GA); FJ652810, FJ695101, FJ694987, FJ695024, FJ695061. **Msc PI 577549**; Georgia; *MS 346* (GA); FJ652811, FJ695102, —, —, —. **Msc PI 634119**; Kazakhstan; *MS 169* (GA); FJ652812, FJ695103, —, —, —. **Msc PI 641380** [a]; Russia; *MS 193* (GA); FJ652813 & FJ652814, FJ695104 & FJ695105, —, —, —. **Msc PI 641601**; Kazakhstan; *MS 197* (GA); FJ652815, FJ695106, FJ694988, FJ695025, FJ695062. **Msf PI 222198** [b]; Afghanistan; *MS 227* (GA); FJ652800, FJ695107, FJ694989, FJ695026, FJ695063. **Msf PI 251830**; Austria; *MS 384* (GA); FJ652816, FJ695108, FJ694990, FJ695027, FJ695064. **Msf PI 325387**; Russia; *MS 340* (GA); FJ652817, FJ695109, —, —, —. **Msf PI 384507**; Russia; FJ652818, FJ695110, FJ694991, FJ695028, FJ695065. **Msf PI 440520** [c]; Russia; FJ652819, FJ695111, FJ694992, FJ695029,

---

* This appendix was published in Havananda, T., E. C. Brummer, I. J. Maureira-Butler, and J. J. Doyle. 2010. Relationships among diploid members of the *Medicago sativa* (Fabaceae) species complex based on chloroplast and mitochondrial DNA sequences. *Systematic Botany* 35: 140–150. Written authorization for the use of the material in this dissertation was obtained from *Systematic Botany*.

FJ695066. **Msf PI 494662** [a]; Romania; *MS 107* (GA); FJ652820 & FJ652821, FJ695112, —, —, —. **Msf PI 502440**; Russia; FJ652822, FJ695113, —, —, —. **Msf PI 502447**; Russia; *MS 268* (GA); FJ652823, FJ695114, FJ694993, FJ695030, FJ695067. **Msf PI 502448**; Russia; *MS 113* (GA); FJ652824, FJ695115, FJ694994, FJ695031, FJ695068. **Msf PI 538987**; Russia; *MS 344* (GA); FJ652825, FJ695116, FJ694995, FJ695032, FJ695069. **Msf PI 577556**; Bulgaria; *MS 134* (GA); FJ652826, FJ695117, —, —, —. **Msf PI 577557**; Bulgaria; *MS GH2* (GA); FJ652827, FJ695118, FJ694996, FJ695033, FJ695070. **Msf PI 577558** [a]; Russia; *MS 282* (GA); FJ652828 & FJ652829, FJ695119, —, —, —. **Msf PI 577564**; Russia; *MS 349* (GA); FJ652830, FJ695120, FJ694997, FJ695034, FJ695071. **Msf PI 577588**; Bulgaria; FJ652831, FJ695121, FJ694998, FJ695035, FJ695072. **Msf PI 631808**; Russia; *MS 382* (GA); FJ652832, FJ695122, FJ694999, FJ695036, FJ695073. **Msf PI 631811**; Kazakhstan; FJ652833, FJ695123, FJ695000, FJ695037, FJ695074. **Msf PI 631816**; Russia; *MS 149* (GA); FJ652834, FJ695124, FJ695001, FJ695038, FJ695075. **Msf PI 631823**; Germany; *MS GH4* (GA); FJ652835, FJ695125, FJ695002, FJ695039, FJ695076. **Msf PI 631828**; Alaska; FJ652836, FJ695126, FJ695003, FJ695040, FJ695077. **Msf PI 631829**; Russia; *MS 295* (GA); FJ652837, FJ695127, FJ695004, FJ695041, FJ695078. **Msf SD201-1**; synthetic variety; *MS 381* (GA); FJ652838, FJ695128, FJ695005, FJ695042, FJ695079. **Msf SD201-11**; synthetic variety; *MS 201-11* (GA); FJ652839, FJ695129, FJ695006, FJ695043, FJ695080. **Msg PI 577567**; Italy; *TH 2007-2* (BH); FJ652840, FJ695130, FJ695007, FJ695044, FJ695081. **Msh PI 262532** [d]; Russia; FJ652841, FJ695131, FJ695008, FJ695045, FJ695082. **Msh PI 641593**; Kazakhstan; FJ652842, FJ695132, FJ695009, FJ695046, FJ695083. **Msh PI 641619**; Kazakhstan; *MS 377* (GA); FJ652843, FJ695133, FJ695010, FJ695047, FJ695084. **Mss CADL-3**; cultivated; *MS GH3* (GA); FJ652844, FJ695134, FJ695011, FJ695048, FJ695085. **Mt 'Jemalong'**; cultivated; FJ652845, FJ695135, FJ695012, FJ695049, FJ695086.

APPENDIX 2.1*

Taxa, plant accession numbers, voucher information of plant materials, and GenBank accession numbers of DNA sequences used in Chapter 2. Different individuals from the same accession are denoted with decimal number. CADL = cultivated alfalfa at the diploid level. Original locality of collection is according to the Germplasm Resources Information Network (GRIN). All plants were grown in greenhouses at Iowa State University and/or the University of Georgia (voucher specimens collected by Muhammet Şakiroğlu [MS]) or Cornell University (collected by Tee Havananda [TH]). Voucher specimens are deposited in the University of Georgia Herbarium (GA) or Herbarium of the L. H. Bailey Hortorium (BH).

***Taxon*—Accession number** (USDA Plant Introduction [PI] number, SARDI [SA] number, genotype, or variety name; underlined accessions were used in Havananda et al. (2010); italicized accessions are polyploids); Original locality of collection; *Voucher specimen* (if available); Herbarium; GenBank accessions: *rpl20-rps12* spacer, *trnS-trnG* spacer ("&" joins two different sequences from the same chloroplast region submitted for single heteroplasmic individuals).

***Medicago cretacea* M. Bieb.—PI 631721**; Russia; HQ198905, HQ198989.

M. papillosa* Boiss.—*PI 464699; Turkey; HQ198906, HQ198990. **W6 5247**; Turkey; HQ198907, HQ198991.

***M. pironae* Vis.—PI 253450**; Slovenia; *TH 2009-8*; BH; HQ198908, HQ198992.

***M. prostrata* Jacq.—PI 577445**; Italy; FJ652795, FJ695087. **PI 577448**; Italy; FJ652796, FJ695088. **PI 577452**; Italy; FJ652797, FJ695089. **PI 577454**; Greece; FJ652798, FJ695090.

***M. sativa* subsp. *caerulea* (Less. Ex Ledeb.) Schmalh.—PI 179370**; Turkey; *MS 1*; GA; FJ652799, FJ695091. **PI 314268**; Uzbekistan; FJ652801, FJ695092. **PI 314275**; Uzbekistan; *MS 37*; GA; FJ652802, FJ695093. **PI 384894**; Iran; FJ652803, FJ695094. **PI 388629**; Uzbekistan; *MS GH1*; GA; FJ652804, FJ695095. **PI 440501**; Kazakhstan; *MS 53*; GA; FJ652805, FJ695096. **PI 464712**; Turkey; *MS 72*; GA; FJ652806, FJ695097. **PI 502425**; Russia; FJ652807, FJ695098. **PI 577540**; Turkey; FJ652808, FJ695099. **PI 577541**; Kazakhstan; *MS 274*; GA; FJ652809, FJ695100. **PI 577543**; Georgia; *MS 181*; GA; FJ652810, FJ695101. **PI 577549**; Georgia; *MS 346*; GA; FJ652811, FJ695102. **PI 634119**; Kazakhstan; *MS 169*; GA; FJ652812, FJ695103. **PI 641380** [a]; Russia; *MS 193*; GA; FJ652813 & FJ652814, FJ695104 & FJ695105. **PI 641601**; Kazakhstan; *MS 197*; GA; FJ652815, FJ695106.

---

* This appendix was published in Havananda, T., E. C. Brummer, and J. J. Doyle. 2011. Complex patterns of autopolyploid evolution in alfalfa and allies (*Medicago sativa*; Leguminosae). *American Journal of Botany* 98: 1633–1646. Written authorization for the use of the material in this dissertation was obtained from *American Journal of Botany*.

M. sativa* subsp. *falcata* (L.) Arcang.**—**PI 222198**; Afghanistan; *MS 227*; GA; FJ652800, FJ695107. **PI 251830**; Austria; *MS 384*; GA; FJ652816, FJ695108. ***PI 325383; Ukraine; HQ198910, HQ198994. **PI 325387**; Russia; *MS 340*; GA; FJ652817, FJ695109. **PI 384507**; Russia; FJ652818, FJ695110. **PI 440520**; Russia; FJ652819, FJ695111. **PI 494662** [a]; Romania; *MS 107*; GA; FJ652820 & FJ652821, FJ695112. ***PI 499550***; China; HQ198911, HQ198995. **PI 502440**; Russia; FJ652822, FJ695113. **PI 502447**; Russia; *MS 268*; GA; FJ652823, FJ695114. **PI 502448**; Russia; *MS 113*; GA; FJ652824, FJ695115. ***PI 502451***; Russia; HQ198912, HQ198996. ***PI 538981***; Russia; HQ198913, HQ198997. ***PI 538985***; Kazakhstan; HQ198914, HQ198998. **PI 538987**; Russia; *MS 344*; GA; FJ652825, FJ695116. ***PI 542800*** [b]; Slovenia; *TH 2009-9*; BH; HQ198915, HQ198999. ***PI 560333***; synthetic variety; HQ198934, HQ199018. **PI 577556**; Bulgaria; *MS 134*; GA; FJ652826, FJ695117. **PI 577557**; Bulgaria; *MS GH2*; GA; FJ652827, FJ695118. **PI 577558** [a]; Russia; *MS 282*; GA; FJ652828 & FJ652829, FJ695119. **PI 577564**; Russia; *MS 349*; GA; FJ652830, FJ695120. **PI 577588**; Bulgaria; FJ652831, FJ695121. ***PI 631592.1***; Italy; *TH 2009-10*; BH; HQ198916, HQ199000. ***PI 631592.2***; Italy; *TH 2009-1*; BH; HQ198917, HQ199001. ***PI 631594***; Greece; *TH 2010-6*; BH; HQ198918, HQ199002. ***PI 631640***; Mongolia; *TH 2009-11*; BH; HQ198919, HQ199003. ***PI 631685***; Mongolia; *TH 2010-7*; BH; HQ198920, HQ199004. ***PI 631690*** [b]; Bulgaria; *TH 2009-12*; BH; HQ198921, HQ199005. ***PI 631704.1***; China; *TH 2009-3*; BH; HQ198922, HQ199006. ***PI 631704.2***; China; *TH 2009-3*; BH; HQ198923, HQ199007. **PI 631808**; Russia; *MS 382*; GA; FJ652832, FJ695122. **PI 631811**; Kazakhstan; FJ652833, FJ695123. **PI 631816**; Russia; *MS 149*; GA; FJ652834, FJ695124. **PI 631823**; Germany; *MS GH4*; GA; FJ652835, FJ695125. **PI 631828**; Alaska; FJ652836, FJ695126. **PI 631829**; Russia; *MS 295*; GA; FJ652837, FJ695127. ***PI 631837*** [b]; Sweden; *TH 2009-5*; BH; HQ198924, HQ199008. ***PI 631849***; Sweden; HQ198925, HQ199009. ***PI 631855***; Sweden; HQ198926, HQ199010. ***PI 631857***; Sweden; HQ198927, HQ199011. ***PI 634158***; Kazakhstan; HQ198928, HQ199012. ***PI 634182***; Kazakhstan; *TH 2009-24*; BH; HQ198929, HQ199013. ***PI 641442.1***; Russia; HQ198930, HQ199014. ***PI 641442.2***; Russia; *TH 2009-6*; BH; HQ198931, HQ199015. ***PI 641587***; Kazakhstan; HQ198932, HQ199016. **SD201-1**; synthetic variety; *MS 381*; GA; FJ652838, FJ695128. **SD201-11**; synthetic variety; *MS 201-11*; GA; FJ652839, FJ695129. ***W6 16608*** Mongolia; HQ198933, HQ199017.

M. sativa* subsp. *glomerata* (Balb.) Rouy**—PI 577566*** [b]; Russia; HQ198935, HQ199019. **PI 577567**; Italy; *TH 2007-2*; BH; FJ652840, FJ695130. ***PI 631978.2***; Georgia; HQ198936, HQ199020. ***PI 631978.3*** [b]; Georgia; HQ198937, HQ199021. ***PI 632028***; Georgia; HQ198938, HQ199022. ***PI 641405.2***; France; HQ198939, HQ199023. ***PI 641405.3***; France; HQ198940, HQ199024. ***SA 35364*** [b]; unknown; HQ198941, HQ199025.

***M. sativa* subsp. ×*hemicycla* (Grossh.) C. R. Gunn**—**PI 262532**; Russia; FJ652841, FJ695131. **PI 634166**; Kazakhstan; *TH 2009-7*; BH; HQ198942, HQ199026. **PI 634172**; Kazakhstan; *TH 2009-13*; BH; HQ198943, HQ199027. **PI 634179**; Kazakhstan; *TH 2009-25*; BH; HQ198944, HQ199028. **PI 634184**; Kazakhstan; *TH 2009-14*; BH; HQ198945, HQ199029. **PI 641593**; Kazakhstan; FJ652842, FJ695132. **PI 641602**; Kazakhstan; *TH 2009-16*; BH; HQ198946, HQ199030. **PI 641619**; Kazakhstan; *MS 377*; GA; FJ652843, FJ695133.

M. sativa* subsp. *sativa*—*ABI 408; cultivated; HQ198969, HQ199053. **CADL-3**; cultivated; *MS GH3*; GA; FJ652844, FJ695134. ***PI 26590***; Algeria; *TH 2009-32*; BH; HQ198952, HQ199036. ***PI 173732***; Turkey; HQ198947, HQ199031. ***PI 234816***; Switzerland; HQ198948, HQ199032. ***PI 250975.1***; Macedonia; *TH 2009-34*; BH; HQ198949, HQ199033. ***PI 250975.2***; Macedonia; *TH 2010-13*; BH; HQ198950, HQ199034. ***PI 251836***; Italy; HQ198951, HQ199035. ***PI 314706***; Kazakhstan; HQ198953, HQ199037. ***PI 314713***; Kazakhstan; *TH 2010-8*; BH; HQ198954, HQ199038. ***PI 384890***; Iran; HQ198955, HQ199039. ***PI 420396***; Spain; *TH 2009-33*; BH; HQ198956, HQ199040. ***PI 440512*** [c]; Kazakhstan; *TH 2009-28*; BH; HQ198957, HQ199041. ***PI 494658*** [b]; Romania; *TH 2010-1*; BH; HQ198958, HQ199042. ***PI 499661*** [b]; China; *TH 2010-2*; BH; HQ198959, HQ199043. ***PI 516588***; Morocco; HQ198960, HQ199044. ***PI 517243***; Portugal; *TH 2009-35*; BH; HQ198961, HQ199045. ***PI 577478.1***; France; *TH 2010-14*; BH; HQ198962, HQ199046. ***PI 577478.2***; France; *TH 2010-9*; BH; HQ198963, HQ199047. ***PI 577496.1***; Tunisia; *TH 2009-36*; BH; HQ198964, HQ199048. ***PI 577496.2*** [a]; Tunisia; *TH 2010-3*; BH; HQ198965 & HQ198966, HQ199049 & HQ199050. ***PI 631954*** [c]; Pakistan; *TH 2009-29*; BH; HQ198967, HQ199051. ***PI 634126*** [b, c]; Kazakhstan; *TH 2009-30*; BH; HQ198968, HQ199052.

M. sativa* subsp. ×*tunetana* Murb.**—PI 535540***; Tunisia; HQ198970, HQ199054. ***PI 577506***; Russia; *TH 2009-22*; BH; HQ198971, HQ199055. ***PI 577507***; Georgia; *TH 2009-17*; BH; HQ198972, HQ199056. ***PI 631715***; Italy; *TH 2009-26*; BH; HQ198973, HQ199057.

M. sativa* subsp. ×*varia* (Martyn) Arcang.**—PI 325382***; Ukraine; *TH 2009-18*; BH; HQ198974, HQ199058. ***PI 440543***; Kazakhstan; *TH 2009-19*; BH; HQ198975, HQ199059. ***PI 577512***; Turkey; *TH 2010-4*; BH; HQ198976, HQ199060. ***PI 577513***; Russia; *TH 2009-20*; BH; HQ198977, HQ199061. ***PI 577527***; Russia; *TH 2009-23*; BH; HQ198978, HQ199062. ***PI 631950***; China; *TH 2009-31*; BH; HQ198979, HQ199063. ***PI 634161***; Kazakhstan; *TH 2010-5*; BH; HQ198980, HQ199064. ***PI 634189***; Kazakhstan; *TH 2010-10*; BH; HQ198981, HQ199065. ***PI 641542***; Mongolia; *TH 2009-21*; BH; HQ198982, HQ199066. ***PI 641570***; Kazakhstan; *TH 2009-27*; BH; HQ198983, HQ199067. ***PI 641574*** [d]; Kazakhstan; *TH 2009-2*; BH; HQ198984, HQ199068. ***PI 641576*** [d]; Kazakhstan; HQ198985, HQ199069. ***PI 641578*** [d]; Kazakhstan; *TH 2009-15*; BH; HQ198986, HQ199070. ***PI 641594***; Kazakhstan; *TH 2010-11*; BH; HQ198987, HQ199071. ***PI 641597*** [d]; Kazakhstan; HQ198988, HQ199072.

M. saxatilis* M. Bieb.**—W6 5898***; France; *TH 2010-12*; BH; HQ198909, HQ198993.

***M. truncatula* Gaertn.**—**'Jemalong'**; cultivated; FJ652845, FJ695135.

[a] Heteroplasmic.
[b] Ambiguous individual (flower color did not correspond with the taxa in which they were classified).
[c] Listed as *M. s.* subsp. *caerulea* in GRIN, but found to be tetraploids by flow cytometry and listed here as *M. s.* subsp. *sativa*
[d] Listed as *M. s.* subsp. ×*hemicycla* in GRIN, but found to be tetraploids by flow cytometry and listed here as *M. s.* subsp. ×*varia*.

APPENDIX 2.2[*]

Ploidy and references for ploidy determination of each *Medicago* accession.

[*] This appendix was published in Havananda, T., E. C. Brummer, and J. J. Doyle. 2011. Complex patterns of autopolyploid evolution in alfalfa and allies (*Medicago sativa*; Leguminosae). *American Journal of Botany* 98: 1633–1646. Written authorization for the use of the material in this dissertation was obtained from *American Journal of Botany*.

Appendix 2.2

Taxon and accession	Origin	Ploidy	References [a]
M. cretacea			
PI 631721	Russia	2*x*	12
M. papillosa			
W6 5247	Turkey	2*x*	7
PI 464699	Turkey	4*x*	7
M. pironae			
PI 253450	Slovenia	2*x*	12
M. prostrata			
PI 577445	Italy	2*x*	5
PI 577448	Italy	2*x*	1
PI 577452	Italy	2*x*	n/a
PI 577454	Greece	2*x*	5
M. sativa subsp. *caerulea*			
PI 179370	Turkey	2*x*	2
PI 314268	Uzbekistan	2*x*	1
PI 314275	Uzbekistan	2*x*	2
PI 384894	Iran	2*x*	9
PI 388629	Uzbekistan	2*x*	5
PI 440501	Kazakhstan	2*x*	2
PI 464712	Turkey	2*x*	2
PI 502425	Russia	2*x*	2
PI 577540	Turkey	2*x*	4
PI 577541	Kazakhstan	2*x*	2
PI 577543	Georgia	2*x*	3
PI 577549	Georgia	2*x*	2
PI 634119	Kazakhstan	2*x*	3
PI 641380	Russia	2*x*	2
PI 641601	Kazakhstan	2*x*	3
M. sativa subsp. falcata (2*x*)			
PI 222198	Afghanistan	2*x*	2
PI 251830	Austria	2*x*	1, 2
PI 325387	Russia	2*x*	1, 2
PI 384507	Russia	2*x*	1
PI 440520	Russia	2*x*	5
PI 494662	Romania	2*x*	1, 2
PI 502440	Russia	2*x*	1, 3
PI 502447	Russia	2*x*	1, 2
PI 502448	Russia	2*x*	1, 2
PI 538987	Russia	2*x*	1, 2
PI 577556	Bulgaria	2*x*	1, 2
PI 577557	Bulgaria	2*x*	1
PI 577558	Russia	2*x*	1, 2
PI 577564	Russia	2*x*	1, 2
PI 577588	Bulgaria	2*x*	4
PI 631808	Russia	2*x*	1, 2
PI 631811	Kazakhstan	2*x*	1
PI 631816	Russia	2*x*	1, 2
PI 631823	Germany	2*x*	1, 5
PI 631828	Alaska	2*x*	1
PI 631829	Russia	2*x*	1, 2
SD201-1	Cultivated	2*x*	1, 2
SD201-11	Cultivated	2*x*	1, 2
M. sativa subsp. *falcata* (4*x*)			
PI 325383	Ukraine	4*x*	1
PI 499550	China	4*x*	1, 5
PI 502451	Russia	4*x*	1, 3
PI 538981	Russia	4*x*	4
PI 538985	Kazakhstan	4*x*	1
PI 542800 [b]	Slovenia	4*x*	3
PI 560333	Cultivated	4*x*	1
PI 631592	Italy	4*x*	3
PI 631594	Greece	4*x*	3
PI 631640	Mongolia	4*x*	3
PI 631685	Mongolia	4*x*	3
PI 631690 [b]	Bulgaria	4*x*	3
PI 631704	China	4*x*	3

Appendix 2.2 (Continued)

Taxon and accession	Origin	Ploidy	References [a]
PI 631837 [b]	Sweden	4*x*	3
PI 631849	Sweden	4*x*	3
PI 631855	Sweden	4*x*	3
PI 631857	Sweden	4*x*	1, 3
PI 634158	Kazakhstan	4*x*	3
PI 634182	Kazakhstan	4*x*	3
PI 641442	Russia	4*x*	3
PI 641587	Kazakhstan	4*x*	3
W6 16608	Mongolia	4*x*	4
M. sativa subsp. *glomerata*			
PI 577567	Italy	2*x*	6
PI 577566 [b]	Russia	4*x*	6
PI 631978 [b]	Georgia	4*x*	6
PI 632028	Georgia	4*x*	6
PI 641405	France	4*x*	6
SA 35364 [b]	Unknown	4*x*	6
M. sativa subsp. ×*hemicycla*			
PI 262532	Russia	2*x*	1
PI 634166	Kazakhstan	2*x*	8
PI 634172	Kazakhstan	2*x*	8
PI 634179	Kazakhstan	2*x*	8
PI 634184	Kazakhstan	2*x*	8
PI 641593	Kazakhstan	2*x*	5, 8
PI 641602	Kazakhstan	2*x*	8
PI 641619	Kazakhstan	2*x*	2
M. sativa subsp. *sativa*			
CADL-3	Cultivated	2*x*	10
ABI 408	Cultivated	4*x*	11
PI 26590	Algeria	4*x*	1
PI 173732	Turkey	4*x*	1, 5
PI 234816	Switzerland	4*x*	9
PI 250975	Macedonia	4*x*	1
PI 251836	Italy	4*x*	1
PI 314706	Kazakhstan	4*x*	1, 5
PI 314713	Kazakhstan	4*x*	1
PI 384890	Iran	4*x*	1, 5
PI 420396	Spain	4*x*	1
PI 440512 [c]	Kazakhstan	4*x*	3
PI 494658 [b]	Romania	4*x*	1
PI 499661 [b]	China	4*x*	1
PI 516588	Morocco	4*x*	1, 5
PI 517243	Portugal	4*x*	1
PI 577478	France	4*x*	1
PI 577496	Tunisia	4*x*	1
PI 631954 [c]	Pakistan	4*x*	3
PI 634126 [b, c]	Kazakhstan	4*x*	3
M. sativa subsp. ×*tunetana*			
PI 535540	Tunisia	4*x*	9
PI 577506	Russia	4*x*	9
PI 577507	Georgia	4*x*	9
PI 631715	Italy	4*x*	9
M. sativa subsp. ×*varia*			
PI 325382	Ukraine	4*x*	9
PI 440543	Kazakhstan	4*x*	9
PI 577512	Turkey	4*x*	9
PI 577513	Russia	4*x*	9
PI 577527	Russia	4*x*	9
PI 631950	China	4*x*	9
PI 634161	Kazakhstan	4*x*	8
PI 634189	Kazakhstan	4*x*	8
PI 641542	Mongolia	4*x*	9
PI 641570	Kazakhstan	4*x*	8
PI 641574 [d]	Kazakhstan	4*x*	3
PI 641576 [d]	Kazakhstan	4*x*	3
PI 641578 [d]	Kazakhstan	4*x*	3
PI 641594	Kazakhstan	4*x*	8

Appendix 2.2 (Continued)

Taxon and accession	Origin	Ploidy	References [a]
PI 641597 [d]	Kazakhstan	4*x*	3
M. saxatilis			
W6 5898	France	6*x*	12
M. truncatula			
'Jemalong'	Cultivated	2*x*	13

[a] Key to references: 1. Brummer et al. (1999), root tip chromosome counts and flow cytometry. 2. Şakiroğlu et al. (2010), flow cytometry. 3. Şakiroğlu and Brummer (2011), flow cytometry. 4. E. C. Brummer (unpublished data), flow cytometry. 5. M. Şakiroğlu (Kafkas University, Turkey, unpublished data), flow cytometry. 6. T. Havananda (unpublished data), flow cytometry. 7. Small (1986b), chromosome counts. 8. Chromosome number appears on record in GRIN, but ploidy determination technique was not indicated. 9. Ploidy was presumed based on taxon identified in GRIN. 10. Artificially diploidized from cultivated tetraploid ("cultivated alfalfa at the diploid level," Bingham and McCoy, 1979), known ploidy. 11. An elite *M. sativa* subsp. *sativa* genotype from ABI Alfalfa, Inc. (Ames, Iowa, USA), known ploidy. 12. Lesins and Lesins (1979), chromosome counts for these taxa, not accession-specific. 13. Blondon et al. (1994), flow cytometry. n/a = no reference for ploidy determination.

[b] Accession with ambiguous individual (flower color did not correspond with the taxa in which they were classified).

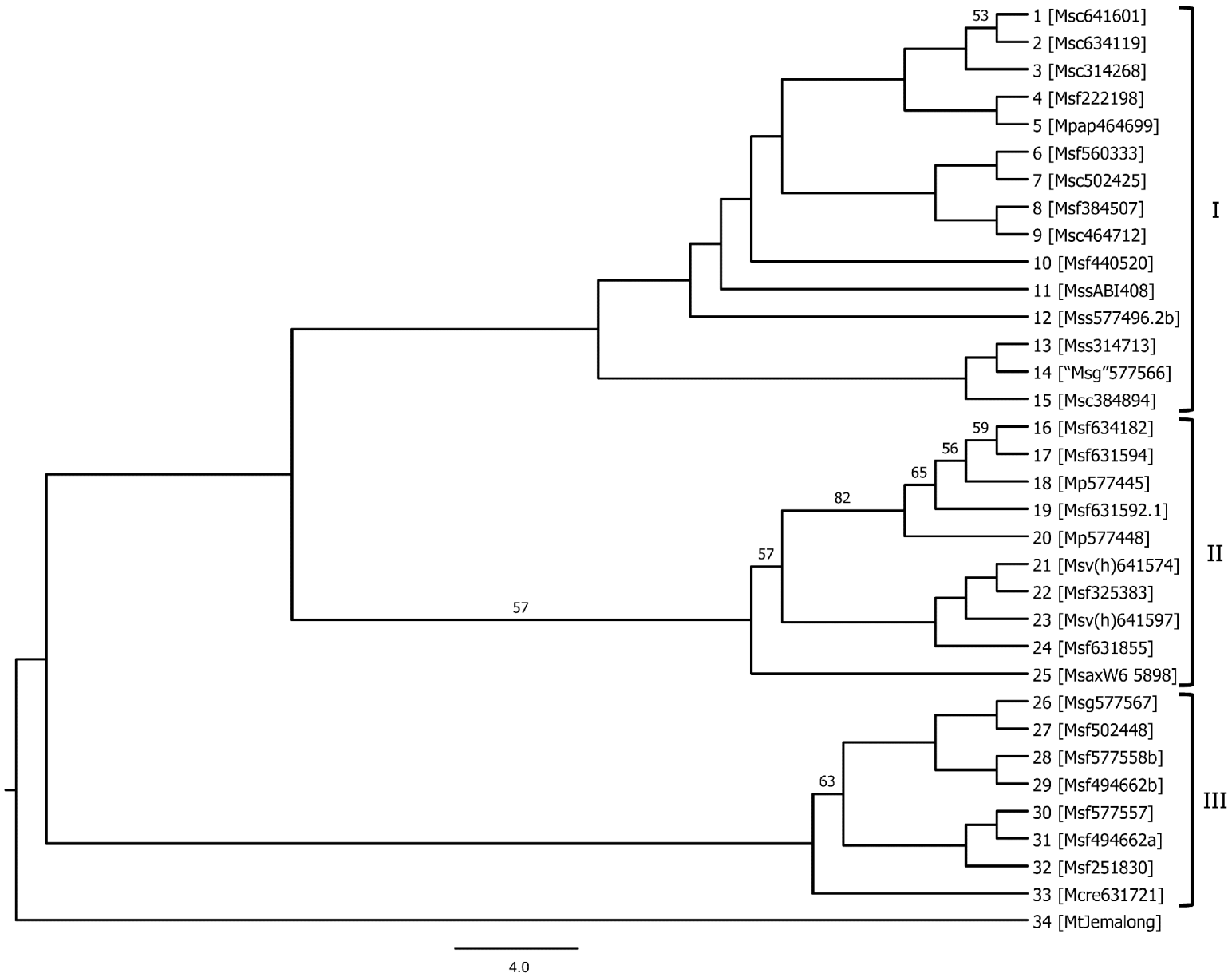
[c] Listed as *M. s.* subsp. *caerulea* in GRIN, but found to be tetraploids by flow cytometry and listed here as *M. s.* subsp. *sativa*

[d] Listed as *M. s.* subsp. ×*hemicycla* in GRIN, but found to be tetraploids by flow cytometry and listed here as *M. s.* subsp. ×*varia*.

APPENDIX 2.3*

Maximum likelihood (ML) tree for the chloroplast 34 haplotypes identified in diploid and polyploid taxa of *Medicago*. Numbers above branches are bootstrap values estimated from 1,000 replicates (only values over 50% are shown). Thick brackets with Roman numerals indicate groups of haplotypes. Number at each terminal is a haplotype number. Name following haplotype number (in brackets) is a representative of the one or more individuals that possesses that haplotype. Abbreviations: Mcre, *Medicago cretacea*; Mp, *M. prostrata*; Mpap, *M. papillosa*; Msax, *M. saxatilis*; Msc, *M. sativa* subsp. *caerulea* (caerulea); Msf, *M. s.* subsp. *falcata* (falcata); Msg, *M. s.* subsp. *glomerata* (glomerata); Mss, *M. s.* subsp. *sativa* (sativa); Msv(h), *M. s.* subsp. ×*varia* originally listed as *M. s.* subsp. ×*hemicycla* in GRIN, but found to be tetraploids by flow cytometry; Mt, *M. truncatula*; names are followed by an accession number (USDA Plant Introduction [PI] number) or variety name. Taxonomically ambiguous individuals' names are in quotation marks. Different individuals from the same accession are denoted with decimal number, e.g., Msf631592.1 is individual number one from falcata accession PI 631592. Two haplotypes found in heteroplasmic individuals are designated with "a" or "b" at the end of the accession or individual number.

* This appendix was published as Supplemental Data ("Appendix S3") with the online version of Havananda, T., E. C. Brummer, and J. J. Doyle. 2011. Complex patterns of autopolyploid evolution in alfalfa and allies (*Medicago sativa*; Leguminosae). *American Journal of Botany* 98: 1633–1646. Written authorization for the use of the material in this dissertation was obtained from *American Journal of Botany*.


53
1 [Msc641601]
2 [Msc634119]
3 [Msc314268]
4 [Msf222198]
5 [Mpap464699]
6 [Msf560333]
7 [Msc502425]
8 [Msf384507]
9 [Msc464712]
10 [Msf440520]
11 [MssABI408]
12 [Mss577496.2b]
13 [Mss314713]
14 ["Msg"577566]
15 [Msc384894]
I
59
56
65
82
57
57
16 [Msf634182]
17 [Msf631594]
18 [Mp577445]
19 [Msf631592.1]
20 [Mp577448]
21 [Msv(h)641574]
22 [Msf325383]
23 [Msv(h)641597]
24 [Msf631855]
25 [MsaxW6 5898]
II
63
26 [Msg577567]
27 [Msf502448]
28 [Msf577558b]
29 [Msf494662b]
30 [Msf577557]
31 [Msf494662a]
32 [Msf251830]
33 [Mcre631721]
III
34 [MtJemalong]
4.0

APPENDIX 3.1

Information for taxa and accessions, of which CNGC5 sequence were successfully obtained, and GenBank accession numbers of DNA sequences used in Chapter 3. Different individuals from the same accession are denoted with decimal number. Original locality of collection is according to the Germplasm Resources Information Network (GRIN). All plants were grown in greenhouses at Iowa State Univ. and/or the Univ. of Georgia (voucher specimens collected by Muhammet Şakiroğlu [MS]) or Cornell University (collected by Tee Havananda [TH]). Voucher specimens are deposited in the University of Georgia Herbarium (GA) or Herbarium of the L. H. Bailey Hortorium (BH). CADL = cultivated alfalfa at the diploid level.

***Taxon*—Accession number** (USDA Plant Introduction [PI] number, genotype, or variety name; italicized accessions are polyploids); Original locality of collection; *Voucher specimen* (if available); Herbarium; GenBank accession numbers for CNGC5 ("&" joins two to four different sequences submitted for single heterozygous individuals).

***M. prostrata* Jacq.—PI 577445**; Italy; JQ964325. **PI 577446**, Italy; *TH 2011-17*; BH; JQ964326. **PI 577447**; Italy; *TH 2011-18*; BH; JQ964327. **PI 577448** [a]; Italy; JQ964328 & JQ964329. **PI 577452**; Italy; JQ964330. **PI 577454**; Greece; JQ964331.

***M. sativa* subsp. *caerulea* (Less. Ex Ledeb.) Schmalh.—PI 179370**; Turkey; *MS 1*; GA; JQ964332. **PI 314268**; Uzbekistan; JQ964333. **PI 314275**; Uzbekistan; *MS 37*; GA; JQ964334. **PI 384894** [a]; Iran; JQ964335 & JQ964336. **PI 388629**; Uzbekistan; *MS GH1*; GA; JQ964337. **PI 464712** [a]; Turkey; *MS 72*; GA; JQ964338 & JQ964339. **PI 502425** [a]; Russia; JQ964340 & JQ964341. **PI 577540** [a]; Turkey; JQ964342 & JQ964343. **PI 577541** [a]; Kazakhstan; *MS 274*; GA; JQ964344 & JQ964345. **PI 577543** [a]; Georgia; *MS 181*; GA; JQ964346. **PI 641601**; Kazakhstan; *MS 197*; GA; JQ964347.

M. sativa* subsp. *falcata* (L.) Arcang.—PI 222198**; Afghanistan; *MS 227*; GA; JQ964348. **PI 251830** [a]; Austria; *MS 384*; GA; JQ964349 & JQ964350. **PI 384507** [a]; Russia; JQ964351 & JQ964352. **PI 440520**; Russia; JQ964353. **PI 502447** [a]; Russia; *MS 268*; GA; JQ964354 & JQ964355. **PI 502448** [a]; Russia; *MS 113*; GA; JQ964356 & JQ964357. **PI 538987** [a]; Russia; *MS 344*; GA; JQ964358 & JQ964359. **PI 577556** [a]; Bulgaria; *MS 134*; GA; JQ964360 & JQ964361. **PI 577557**; Bulgaria; *MS GH2*; GA; JQ964362. **PI 577564** [a]; Russia; *MS 349*; GA; JQ964363 & JQ964364. **PI 577588** [a]; Bulgaria; JQ964365 & JQ964366. ***PI 631592.1 [a]; Italy; *TH 2009-10*; BH; JQ964367 & JQ964368. ***PI 631592.2*** [a]; Italy; *TH 2009-1*; BH; JQ964369 & JQ964370. ***PI 631594*** [a]; Greece; *TH 2010-6*; BH; JQ964371 & JQ964372 & JQ964373 & JQ964374. ***PI 631685*** [a]; Mongolia; *TH 2010-7*; BH; JQ964375 & JQ964376. **PI 631808**; Russia; *MS 382*; GA; JQ964377. **PI 631811**; Kazakhstan; JQ964378. **PI 631816** [a]; Russia; *MS 149*; GA; JQ964379 & JQ964380. **PI 631823**; Germany; *MS GH4*; GA; JQ964381. **PI 631828**; Alaska; JQ964382. **PI 631829** [a]; Russia; *MS 295*; GA; JQ964383 & JQ964384. ***PI 634182*** [a]; Kazakhstan; *TH 2009-24*; BH; JQ964385 & JQ964386. ***PI 641442.1*** [a]; Russia; JQ964387 & JQ964388. ***PI 641442.2*** [a]; Russia; *TH 2009-6*; BH; JQ964389 & JQ964390 & JQ964391. **SD201-1** [a]; synthetic

variety; *MS 381*; GA; JQ964392 & JQ964393. **SD201-11**; synthetic variety; *MS 201-11*; GA; JQ964394.

***M. sativa* subsp. ×*hemicycla* (Grossh.) C. R. Gunn—PI 262532** [a]; Russia; JQ964395 & JQ964396. **PI 634166**; Kazakhstan; *TH 2009-7*; BH; JQ964397. **PI 641579** [a]; Kazakhstan; JQ964398 & JQ964399. **PI 641593** [a]; Kazakhstan; JQ964400 & JQ964401. **PI 641619** [a]; Kazakhstan; *MS 377*; GA; JQ964402 & JQ964403.

M. sativa* subsp. *sativa*—CADL-3** [a]; cultivated; *MS GH3*; GA; JQ964413 & JQ964414. ***PI 314713 [a]; Kazakhstan; *TH 2010-8*; BH; JQ964404 & JQ964405. ***PI 440512*** [a, c]; Kazakhstan; *TH 2009-28*; BH; JQ964406 & JQ964407 & JQ964408. ***PI 631954*** [a, c]; Pakistan; *TH 2009-29*; BH; JQ964409 & JQ964410 & JQ964411 & JQ964412.

M. sativa* subsp. ×*varia* (Martyn) Arcang.— *PI 641576 [a, d]; Kazakhstan; JQ964415 & JQ964416. ***PI 641597*** [a, d]; Kazakhstan; JQ964417 & JQ964418.

***M. truncatula* Gaertn.—'Jemalong'**; cultivated; JX014414.

[a] Heterozygous.

[b] Ambiguous individual (flower color did not correspond with the taxa in which they were classified).

[c] Listed as *M. s.* subsp. *caerulea* in GRIN, but found to be tetraploids by flow cytometry and listed here as *M. s.* subsp. *sativa*

[d] Listed as *M. s.* subsp. ×*hemicycla* in GRIN, but found to be tetraploids by flow cytometry and listed here as *M. s.* subsp. ×*varia*.

APPENDIX 3.2

Allelic composition of each individual included in Chapter 3. Base position of each polymorphism is a position in an alignment of all individuals with *M. truncatula*. Indels are shown in italic characters. Colon ":" represents base deletion. Taxon name abbreviation is as Figures 3.1 and 3.3. Tetraploids are in boldface, italic characters. Allele number corresponds to Figures 3.1–3.3.

Appendix 3.2

Base position	1	*19–20*	28	49	62	65	68	*75–76*	79	88	99	105	*116–118*	142	*144*	153	166	214	228	233	247	248	268	280	283	289	293	294	320	345	*349–363*	364	367	437	446	452	456	613	617	685	745	838	
Polymorphism	A/T	*TC/:*	G/T	C/T	A/G	A/C	A/G	*TA/:*	C/G	G/T	A/G	A/T	*TGT/:*	G/T	*T/:*	C/G A/G	A/C	G/T	C/T	C/G	C/T	C/T	A/T	A/T	C/T	A/C	C/T	G/T	C/G	C/T	*15bp/:*	A/T	C/T	C/T	C/T	A/T	C/T	C/T	C/T	C/T	A/G	Allele number
Mp 577445	A	TC	G	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	C	T	T	A	T	A	T	G	C	T	15 bp	A	T	C	T	C	T	C	C	T	T	A	37
Mp 577446	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	C	T	T	A	T	A	T	G	C	T	15 bp	A	T	C	T	C	T	C	C	T	T	A	36
Mp 577447	A	TC	G	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	C	T	T	A	T	A	T	G	C	T	15 bp	A	T	C	T	C	T	C	C	T	T	A	37
Mp 577448_a	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	C	T	T	A	T	A	T	G	C	T	15 bp	A	T	C	T	C	T	C	C	T	T	A	36
Mp 577448_b	A	TC	T	C	A	C	A	TA	G	G	G	T	TGT	G	T	G	C	T	T	C	C	C	T	A	T	A	T	G	C	T	:	T	T	C	T	C	T	C	C	C	T	A	38
Mp 577452	A	TC	T	C	A	C	A	TA	G	G	G	T	TGT	G	T	G	C	T	T	C	C	C	T	A	T	A	T	G	C	T	:	T	T	C	T	C	T	C	C	C	T	A	38
Mp 577454	A	TC	G	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	C	T	T	A	T	A	T	G	C	T	15 bp	A	T	C	T	C	T	C	C	T	T	A	37
Msc 179370	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	7
Msc 314268	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	7
Msc 314275	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	7
Msc 384894_a	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	T	C	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	12
Msc 384894_b	A	TC	T	T	G	C	A	TA	G	G	G	T	TGT	T	T	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	15 bp	T	T	C	T	C	A	C	C	T	T	A	1
Msc 388629	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	7
Msc 464712_a	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	7
Msc 464712_b	A	TC	T	T	A	C	A	TA	G	G	G	T	:	G	T	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	11
Msc 502425_a	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	7
Msc 502425_b	A	TC	T	T	G	C	A	TA	G	G	G	T	TGT	G	:	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	15 bp	T	T	C	T	C	A	C	C	T	T	A	2
Msc 577540_a	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	T	T	A	T	G	G	C	15 bp	T	T	C	T	C	T	C	C	T	T	A	10
Msc 577540_b	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	7
Msc 577541_a	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	7
Msc 577541_b	A	TC	T	T	G	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	15 bp	T	T	C	T	C	A	C	C	T	T	A	3
Msc 577543	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	7
Msc 641601	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	7
Msf 222198	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	T	C	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	12
Msf 251830_a	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	17
Msf 251830_b	T	TC	T	T	A	C	A	TA	G	G	G	A	TGT	G	T	G	A	T	T	C	T	C	T	A	T	A	T	G	C	T	15 bp	T	T	C	T	T	T	C	C	T	T	A	26
Msf 384507_a	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	15 bp	T	C	T	T	C	T	C	C	T	T	A	15
Msf 384507_b	T	TC	T	T	A	C	A	TA	G	G	G	A	TGT	G	T	G	A	T	T	C	T	C	T	A	T	A	T	G	C	T	15 bp	T	T	C	T	T	T	C	C	T	T	A	26
Msf 440520	T	TC	T	T	A	C	A	TA	G	G	G	A	TGT	G	T	G	A	T	T	C	T	C	T	A	T	A	T	G	C	T	15 bp	T	T	C	T	T	T	C	C	T	T	A	26
Msf 502447_a	A	TC	T	T	A	C	A	TA	G	T	G	T	TGT	G	T	G	A	T	C	C	T	C	T	A	T	A	T	G	C	T	15 bp	T	T	C	T	T	T	T	C	T	T	A	33
Msf 502447_b	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	A	T	C	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	19
Msf 502448_a	A	TC	T	T	A	C	G	TA	G	G	G	A	TGT	G	T	G	A	T	T	G	T	C	T	A	T	A	T	G	C	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	24
Msf 502448_b	A	TC	T	T	A	A	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	A	A	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	21
Msf 538987_a	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	17
Msf 538987_b	A	TC	T	T	A	A	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	A	A	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	21
Msf 577556_a	A	TC	T	T	A	C	G	TA	G	G	G	A	TGT	G	T	G	A	T	T	C	T	C	T	A	T	A	T	G	C	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	23
Msf 577556_b	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	17
Msf 577557	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	C	C	T	C	T	A	T	A	T	G	C	T	15 bp	T	T	C	T	T	T	T	C	T	T	A	31
Msf 577564_a	A	TC	T	T	A	C	G	TA	G	G	G	A	TGT	G	T	G	A	T	T	C	T	C	T	A	T	A	T	G	C	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	23
Msf 577564_b	A	TC	T	T	A	A	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	A	A	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	21
Msf 577588_a	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	C	C	T	C	T	A	T	A	T	G	C	T	15 bp	T	T	C	T	T	T	T	C	T	T	A	31
Msf 577588_b	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	17
Msf 631592.1_a	A	TC	T	T	A	C	A	TA	G	G	G	A	TGT	G	T	G	A	T	T	C	T	C	T	A	T	A	T	G	C	T	15 bp	T	T	C	T	T	T	C	C	T	T	A	29
Msf 631592.1_b	A	TC	T	T	A	C	A	TA	C	G	G	T	TGT	G	T	G	A	G	T	C	T	C	T	A	T	A	T	T	C	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	35
Msf 631592.2_a	A	TC	T	T	A	C	A	TA	G	G	G	T	TGT	G	T	G	A	T	C	C	T	C	T	A	T	A	T	G	C	T	15 bp	T	T	C	T	T	T	C	C	T	T	A	30
Msf 631592.2_b	A	TC	T	T	A	C	A	TA	C	G	G	T	TGT	G	T	G	A	G	T	C	T	C	T	A	T	A	T	T	C	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	35
Msf 631594_a	A	TC	T	T	A	C	A	:	G	G	G	T	TGT	G	T	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	15 bp	T	T	C	T	C	T	C	C	T	T	A	18

Appendix 3.2 (Continued)

Base position	1	*19–20*	28	49	62	65	68	*75–76*	79	88	99	105	*116–118*	142	*144*	153	166	214	228	233	247	248	268	280	283	289	293	294	320	345	*349–363*	364	367	437	446	452	456	613	617	685	745	838	
Polymorphism	A/T	*TC/:*	G/T	C/T	A/G	A/C	A/G	*TA/:*	C/G	G/T	A/G	A/T	*TGT/:*	G/T	*T/:*	C/G A/G	A/C	G/T	C/T	C/G	C/T	C/T	A/T	C/T	A/C	C/T	G/T	C/G	C/T	*15bp/:*	C/T A/T	C/T	C/T	C/T	A/T	C/T	C/T	C/T	A/G	Allele number
Msf 631594_b	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	C	C	T	C	T	A	T	A	T	G	C	T	*15 bp*	T	T	C	T	T	T	T	C	T	T	A	31
Msf 631594_c	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	C	C	T	C	T	A	T	A	C	G	C	T	*15 bp*	T	T	C	T	T	T	T	C	T	T	A	32
Msf 631594_d	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	C	C	T	C	T	A	T	A	T	G	C	T	*15 bp*	T	T	C	T	C	T	T	C	T	T	A	34
Msf 631685_a	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	17
Msf 631685_b	T	*TC*	T	T	A	C	A	*TA*	G	G	G	A	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	C	T	*15 bp*	T	T	C	T	T	T	C	C	T	T	A	26
Msf 631808	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	C	C	T	C	T	A	T	A	T	G	C	T	*15 bp*	T	T	C	T	T	T	T	C	T	T	A	31
Msf 631811	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	17
Msf 631816_a	T	*:*	T	T	A	C	A	*TA*	G	G	G	A	*TGT*	G	*T*	C	A	T	T	C	T	C	T	A	T	A	T	G	C	T	*15 bp*	T	T	C	T	T	T	C	C	T	T	A	28
Msf 631816_b	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	C	T	T	C	T	C	C	T	T	A	15
Msf 631823	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	17
Msf 631828	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	17
Msf 631829_a	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	C	T	T	C	T	C	C	T	T	A	15
Msf 631829_b	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	T	T	T	C	T	C	C	T	T	A	14
Msf 634182_a	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	17
Msf 634182_b	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	G	20
Msf 641442.1_a	A	*TC*	T	T	A	C	A	*TA*	G	G	A	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	8
Msf 641442.1_b	A	*TC*	T	T	A	C	A	*TA*	G	G	A	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	T	C	C	C	T	C	C	T	T	A	9
Msf 641442.2_a	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	17
Msf 641442.2_b	A	*TC*	T	T	A	C	A	*TA*	G	G	G	A	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	22
Msf 641442.2_c	T	*TC*	T	T	A	C	A	*TA*	G	G	G	A	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	C	T	*15 bp*	T	T	C	T	T	T	C	C	T	T	A	26
Msf SD201.1_a	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	C	T	T	C	T	C	C	T	T	A	15
Msf SD201.1_b	T	*TC*	T	T	A	C	A	*TA*	G	G	G	A	*TGT*	G	*T*	C	A	T	T	C	T	C	T	A	T	A	T	G	C	T	*15 bp*	T	T	C	T	T	T	C	C	T	T	A	27
Msf SD201.11	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	C	T	T	C	T	C	C	T	T	A	15
Msh 262532_a	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	T	T	T	C	T	C	C	T	T	A	14
Msh 262532_b	A	*:*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	T	T	T	C	T	C	C	T	T	A	16
Msh 634166	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	17
Msh 641579_a	T	*TC*	T	T	A	C	A	*TA*	G	G	G	A	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	C	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	25
Msh 641579_b	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	17
Msh 641593_a	T	*TC*	T	T	A	C	A	*TA*	G	G	G	A	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	C	T	*15 bp*	T	T	C	T	T	T	C	C	T	T	A	26
Msh 641593_b	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	17
Msh 641619_a	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	7
Msh 641619_b	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	17
Mss 314713_a	A	*TC*	T	T	G	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	A	C	C	T	T	A	3
Mss 314713_b	A	*TC*	T	T	G	C	A	*TA*	G	G	G	T	*TGT*	T	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	A	C	C	T	T	A	1
Mss 440512_a	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	7
Mss 440512_b	A	*TC*	T	T	G	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	4
Mss 440512_c	A	*TC*	T	T	G	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	A	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	5
Mss 631954_a	A	*TC*	T	T	G	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	A	C	C	T	T	A	3
Mss 631954_b	A	*TC*	T	T	G	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	C	A	6
Mss 631954_c	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	C	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	12
Mss 631954_d	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	C	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	T	T	T	A	13
Mss CADL3_a	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	7
Mss CADL3_b	A	*TC*	T	T	G	C	A	*TA*	G	G	G	T	*TGT*	T	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	A	C	C	T	T	A	1
Msv(h) 641576_a	T	*TC*	T	T	A	C	A	*TA*	G	G	G	A	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	C	T	*15 bp*	T	T	C	T	T	T	C	C	T	T	A	26
Msv(h) 641576_b	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	17
Msv(h) 641597_a	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	T	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	7
Msv(h) 641597_b	A	*TC*	T	T	A	C	A	*TA*	G	G	G	T	*TGT*	G	*T*	G	A	T	T	C	T	C	T	A	T	A	T	G	G	T	*15 bp*	T	T	C	T	C	T	C	C	T	T	A	17