Published September, 2011

RESEARCH

Variation in Biomass Yield, Cell Wall Components, and Agronomic Traits in a Broad Range of Diploid Alfalfa Accessions

Muhammet Sakiroglu,* Kenneth J. Moore, and E. Charles Brummer*

ABSTRACT

Alfalfa (*Medicago sativa* L.), one of the most important forage legumes throughout the world, has potential as a biofuel crop. Breeding to improve biomass yield and stem composition will enhance the value of alfalfa. In this experiment, we evaluated 374 individual genotypes from 120 accessions of wild diploid alfalfa collected throughout the Northern Hemisphere and representing the three diploid subspecies, *M. sativa* subsp. *falcata*, *M. sativa* subsp. *caerulea,* and *M. sativa* subsp. *hemicycla*. We measured total biomass yield, stem cell wall constituents (neutral detergent fiber, acid detergent fiber, acid detergent lignin, glucose, xylose, and arabinose), and important agronomic traits in field trials at two Georgia locations in 2007 and 2008. A large amount of phenotypic variation exists in the diploid gene pool for all traits. Yield variation among accessions spanned more than an order of magnitude. High-yielding accessions tended to have higher levels of cell wall constituents. Clustering of genotypes or accessions on the basis of phenotypic traits differed from the population structure pattern observed previously with molecular markers. The results show that phenotypic variation is very large in diploids, offering a potentially useful pool of alleles for future breeding efforts.

M. Sakiroglu, Dep. of Biology, Kafkas Univ., 36100 Kars, Turkey; K.J. Moore, Dep. of Agronomy, Iowa State Univ., Ames, IA 50011; E.C. Brummer, Samuel Roberts Noble Foundation, 2510 Sam Noble Pkwy., Ardmore, OK 73401. Received 21 Nov. 2010. *Corresponding authors (msakiroglu@kafkas.edu.tr; ecbrummer@noble.org).

Abbreviations: ADF, acid detergent fiber; ADL, acid detergent lignin; NDF, neutral detergent fiber; PCA, principal component analysis; QTL, quantitative trait loci; SSR, simple sequence repeat; TNC, total nonstructural carbohydrates.

Alfalfa (*Medicago sativa* L.), the oldest plant that has been exclusively grown for forage, is the most important forage legume in the world (Quiros and Bauchan, 1988; Michaud et al., 1988). In addition to its forage use, alfalfa can also be used as a biofeedstock for cellulosic ethanol production (Delong et al., 1995). Alfalfa stems and leaves can be separated, and the leaves could be used as a high-protein animal feed while the stems are used to produce energy (McCaslin and Miller, 2007; Lamb et al., 2007). Because alfalfa significantly reduces the need for fossil-fuel-based synthetic nitrogen fertilizers, it decreases production costs and limits environmental problems (Patzek, 2004; Crews and Peoples, 2004).

High yield and conversion efficiency are the two main qualifications needed for any crop to be considered for bioenergy production (Ragauskas et al., 2006). Alfalfa has the potential to produce high yield, but genetic improvement for yield has not been as high as that realized in the major grain crops (Hill et al., 1988). The most obvious way to increase yield is through population improvement via recurrent selection, but limited selection for yield per se has been conducted in most breeding programs. A second way to increase yield is through the use of semihybrid

Published in Crop Sci. 51:1956–1964 (2011).
doi: 10.2135/cropsci2010.11.0658
Freely available online through the author-supported open-access option.
Published online 6 July 2011.

cultivars, produced by hybridizing populations that express heterosis for biomass yield (Brummer 1999). Hybrids between semidormant *M. sativa* subsp. *sativa* and subsp. *falcata* show yield heterosis (e.g., Riday and Brummer, 2002), but more agronomically useful nondormant germplasm shows limited heterosis with semidormant cultivars (Sakiroglu and Brummer, 2007), except for the Peruvian germplasm (Segovia-Lerma et al., 2003; Maureira et al., 2004). A third method to improve yield is to identify quantitative trait loci (QTL) for yield and subsequently introgress them into modern cultivars. These "yield alleles" could be derived from any germplasm, from elite tetraploid cultivars to wild diploids. A combination of QTL introgression, recurrent selection, and heterosis capture could greatly improve alfalfa yields.

Conversion efficiency—the amount of energy that can be derived from a given amount of biomass—is the second characteristic important for a potential biofuel. It can be altered by modifying cell wall components; for example, bioethanol production is facilitated if lignin is minimized and cellulose maximized (Farrokhi et al., 2006). In alfalfa, cell wall components have been modified through selection and transgenic approaches to increase forage quality and digestibility (Buxton and Russell, 1988; Albrecht et al., 1987; Reddy et al., 2005), and improving conversion efficiency will require similar changes to the cell wall. Cellulose and hemicellulose are embedded in a matrix of lignin that needs to be disrupted by pretreatment of biomass to increase microbial access to the sugars; pretreatment is one of the most expensive procedures during cellulosic bioethanol production (Dien et al., 2006). A reduction of lignin could increase saccarification at the pretreatment stage by increasing access to cellulose and hemicellulose (Chen and Dixon, 2007; Jackson et al., 2008).

Modifying cell walls can be done by recurrent selection, by marker-assisted QTL introgression, or through transgenic approaches. Selection has been effective at modifying alfalfa for forage nutritive value (Coors et al., 1986; Demment et al., 1986; and Shenk and Elliott, 1971), essentially the same changes that are needed for bioethanol production. Genetic mapping to identify QTL affecting cell wall composition, and thereby conversion efficiency of biomass into ethanol, can be used to identify natural allelic variation that could be incorporated into breeding populations via marker-assisted selection. Down-regulation of enzymes in the lignin biosynthetic pathway by transgenic methods have significantly lowered lignin content and increased digestibility (Baucher et al., 1999; Reddy et al., 2005; Guo et al., 2001a; Nakashima et al., 2008).

Virtually all cultivated alfalfa is a tetrasomic tetraploid ($2n = 2x = 32$), and almost all germplasm evaluation and manipulation has been done on tetraploids. However, diploid ($2n = 2x = 16$) germplasm may offer a simpler means of identifying and manipulating desirable QTL alleles for important traits. Because hybridization is possible across ploidy levels through unreduced eggs (McCoy and Bingham, 1988), diploid alleles can be incorporated into tetraploid breeding populations. Three naturally occurring diploid taxa are part of the larger *Medicago sativa* species complex: *M. sativa* subsp. *falcata*, which has yellow flowers and sickle-shaped pods; *M. sativa* subsp. *caerulea*, with purple flowers and coiled pods; and their natural hybrid, *M. sativa* subsp. *hemicycla*, which possesses variegated flowers and pods with approximately one coil (Quiros and Bauchan, 1988). Previously, we investigated the population structure of a wide range of diploid alfalfa germplasm using simple sequence repeat (SSR) markers (Sakiroglu et al., 2010). The results indicated that the three subspecies of diploid germplasm were clearly differentiated by SSR markers. Additionally, two distinct *caerulea* groups corresponding to northern and southern regions and two distinct *falcata* groups differentiated by ecogeography as either lowland or upland ecotypes were identified (Sakiroglu et al., 2010).

Knowledge of the biofeedstock potential of the five germplasm groups will help breeders more effectively use diploid germplasm to develop improved cultivars. Therefore, our objectives were (i) to measure the variation for biomass yield and composition in the diploid gene pool of cultivated alfalfa as a prelude to genetic mapping in diploid populations, and (ii) to investigate the population structure of wild alfalfa on the basis of phenotype compared with that detected previously with genotypic data.

MATERIALS AND METHODS

Plant Materials and Experimental Design

We selected 374 individuals from 120 diploid accessions of *M. sativa* from the USDA collections. The accessions were sampled from throughout the Northern Hemisphere to represent the natural distribution of diploid accessions from the three subspecies. Each accession was represented by one to four individual genotypes. All genotypes were grown and clonally propagated at the University of Georgia Crop & Soil Sciences Greenhouse before transplanting to the field. We conducted flow cytometry to confirm the ploidy level of each genotype and recorded morphological characters, including flower color and pod shape, that are used for taxonomic classification. Further details are presented in a companion paper (Sakiroglu et al., 2010).

Field experiments were established on 10 May 2007 at the University of Georgia Plant Sciences Farm near Watkinsville, GA (33°52′ N; 83°32′ W), in a Cecil sandy loam (fine, kaolinitic, thermic typic kanhapludults) and on 18 June 2007 at the UGA Central Georgia Research and Education Center near Eatonton, GA (33°24′ N; 83°29′ W), in a Davidson loam (fine, kaolinitic, thermic rhodic kandiudults). At the Plant Sciences Farm, a triple α-lattice with each complete block (replication) consisting of 24 incomplete blocks of 16 entries was used for the experimental design. Ten check entries were planted to complete the design. Each plot consisted of four clones planted 15-cm apart in a single row. Plots were separated side-by-side and end-by-end by 75 cm. The trial at Eatonton included 115

entries planted in two replications as well as an additional 90 entries included in only a single replication because of limited cuttings. Plots in each replication consisted of three clones spaced 15-cm apart with plots separated in rows by 45 cm and rows spaced 75-cm apart. At both locations, plants were clipped approximately 1 mo after establishment and then allowed to regrow for the remainder of the season.

Biomass yield was measured on 17–18 Sept. 2007 at Watkinsville and 24 Sept. 2007 at Eatonton by clipping the entire plot at a uniform height of 5 cm, with the herbage placed in a paper bag, dried for 5 d at 60°C, and weighed. The stems and leaves were separated to determine stem weight of each plot, and the stem proportion of the total biomass was computed. The number of surviving plants was recorded at the time of harvest and was used to adjust yield to a four-plant plot basis. Plant regrowth was scored visually 2 wk after harvest using a scale of 1 = no regrowth to 9 = vigorous regrowth. The second year harvest was taken between 5 and 9 June 2008 at Watkinsville and 11 June 2008 at Eatonton in a similar manner. Plant survival was noted and regrowth ability was scored as described above. Before harvest, plant height of each plot in the Watkinsville trial was measured on each individual plant and a plot average computed. Stem thickness for each individual plot was measured in millimeters directly above the second node on each of four dried stems randomly selected from each plot; diameters were averaged for each plot before analysis. Spring regrowth of each plot in the Watkinsville trial was measured on 29 Mar. 2008. Stem weight and stem proportion were not recorded in 2008.

Composition Analyses

Composition analysis was conducted on stems from each of the four environments (Watkinsville 2007, Watkinsville 2008, Eatonton 2007, and Eatonton 2008). Stems were ground to pass a 1-mm mesh screen (Cyclone Mill, UDY Mfg., Ft. Collins, CO). Each stem sample was scanned by near-infrared reflectance spectroscopy (Windham et al., 1989). We used a NIRSystems 6500 scanning monochromator (NIRSystems, Silver Spring, MD) for collection of the reflectance measurements (log 1/*R*) between 1100 and 2500 nm, recorded at 4-nm intervals. A subset of 97 samples was selected for calibration of spectroscopy using chemical analyses. The 97 calibration samples were used to determine neutral detergent fiber (NDF), acid detergent fiber (ADF), and acid detergent lignin (ADL). We determined NDF and ADF for the calibration set with fiber bag technology using the ANKOM Fiber Analyzer and the Daisy incubator (ANKOM Technology, Macedon, NY) following Vogel et al. (1999). We determined ADL and ash using the method of Van Soest et al. (1991). The ANKOM bags containing the residual of the ADF procedure were placed in a 3-L Daisy^{II} incubator jar and covered with 72% H_2SO_4. The samples were rotated in the incubator for 3 h, washed in hot water for 15 min and by acetone for 10 min, dried in a 100°C oven overnight, and weighed after cooling to room temperature. Finally, the entire sample bag with its remaining material was ashed at 550°C for 4 h, and the ash was weighed. Ash weights were calculated after accounting for the sample bag material. The composition of the neutral sugars glucose, arabinose, and xylose was measured using the two-step hydrolysis procedure in the Uppsala Method (Theander et al., 1995). Ground plant samples weighing 0.1 g were placed in 50-mL test tubes and incubated with 1.25 mL of 12 M sulfuric acid for 1 h at 30°C. The acid was diluted to 0.4 M by adding 35 mL of water to the tubes, which were then autoclaved at 125°C for 1 h. After cooling, 1-mL aliquots of the hydrolysates were diluted to a volume of 100 mL in deionized water and filtered using 0.45-m syringe filters. Filtrates were analyzed for arabinose, xylose, and glucose by high-performance liquid chromatography with an electrochemical detector (Dionex ED 50 Electrochemical Detector, Dionex, Sunnyvale, CA) using a Carbopac PA200 column (Dionex, Sunnyvale, CA). Separations were made using 2 mM NaOH as the mobile phase at a flow rate of 0.48 mL min^{-1}, a column temperature of 30°C, and an injection volume of 25 μL.

Calibration equations for fibers and sugars were developed using modified partial least squares regression (Shenk and Westerhaus, 1991). Coefficients of determination (R^2) and standard errors of the calibration and cross validation were; 0.98, 0.88, and 1.18 for NDF; 0.98, 0.80, and 1.07 for ADF; 0.96, 0.34, and 0.48 for ADL; 0.97, 0.52, and 0.67 for total nonstructural carbohydrates (TNC); 0.75, 1.68, and 1.87 for glucose; 0.90, 0.21, and 0.27 for arabinose; and 0.83, 0.55, and 0.61 for xylose (Windham et al., 1989).

Data Analyses

We conducted analyses of variance for each trait using all data over the 2 yr and two locations to investigate interactions among genotypes, years, and locations. We fit a mixed linear model including replications and blocks as random effects and years, locations, genotypes and the interactions among them as fixed effects. We subsequently conducted analyses of variance by year over locations and by locations over years. We estimated least-squares means of all traits for each genotype, for each accession (across genotypes of that accession), and for five genetically distinct groups (southern *caerulea*, northern *caerulea*, hemicycla, lowland *falcata*, and upland *falcata*) we identified previously with genetic markers (Sakiroglu et al., 2010). Mean separations were conducted using Fisher's protected least significant difference (LSD). We used Spearman's rank correlation test for traits measured in both years in the Watkinsville trial to investigate ranking changes among entries. Pearson's correlations were calculated between selected traits using mean values from the Watkinsville trial. Eatonton data were not included because only a subset of genotypes was present and not all traits were assessed in that trial. All analyses were conducted using SAS 9.2 (SAS Institute, 2004). Statistical significance was assessed at the 5% probability level unless otherwise indicated.

We used 23 traits measured in 2007 and 2008 in the Watkinsville trial to compute phenotypic distances among the 374 individual genotypes. The 23 traits included biomass yield, regrowth after harvest, NDF, ADF, ADL, TNC, glucose, arabinose, and xylose in each of 2 yr; stem weight and proportion of total biomass in the stem fraction, measured in 2007; and spring regrowth, plant height, and stem thickness measured in 2008. Phenotypic data for each trait were standardized and then used to calculate a Euclidian distance between all pairs of genotypes. A principal component analysis (PCA) was conducted to reduce the dimensionality of the data, enabling visualization of the relationships among genotypes on the basis of phenotypic data. We also estimated phenotypic distance and PCA of the 120 accessions in a similar manner. These analyses were conducted using NTSYS-pc (Rohlf, 1994).

Table 1. Mean, standard deviation, and range of agronomic traits and cell wall components of 373 wild diploid alfalfa (*Medicago sativa* L.) genotypes in 2 yr (2007 and 2008) averaged across locations.

	2007			2008		
		Range			Range	
Trait†	Mean ± SD	Individuals	Accessions	Mean ± SD	Individuals	Accessions
Yield	55 ± 34	4–209	6–194	384 ± 209	9–2193	106–2185
Regrowth	2.8 ± 0.7	1.2–4.9	1.9–4.3	3.4 ± 1.2	0.1–6.1	0.9–5.0
Stem Ratio	0.4 ± 0.1	0.1–0.7	0.14–0.49	NA	NA	NA
Stem Weight	23 ± 16	0.8–104	1–94	NA	NA	NA
Spring Regrowth	NA	NA	NA	2.3 ± 0.7	0.3–6.4	1.0–6.4
Plant Height	NA	NA	NA	81 ± 14	26–133	49–117
Stem Thickness	NA	NA	NA	2.3 ± 0.6	0.4–4.1	0.83–3.82
NDF	532 ± 35	375–641	411–604	655 ± 39	511–764	519–742
ADF	384 ± 33	237–478	269–445	497 ± 37	363–581	371–572
ADL	93 ± 9	58–122	66–115	120 ± 8	88–139	94–137
TNC	162 ± 19	100–285	110–261	106 ± 12	86–169	90–161
Arabinose	33 ± 4	22–45	26–41	24 ± 3	14–39	15–39
Glucose	276 ± 17	212–318	238–302	315 ± 16	262–359	279–344
Xylose	81 ± 6	50–100	59–93	97 ± 7	67–122	77–114

†NDF = neutral detergent fiber; ADF = acid detergent fiber; ADL = acid detergent lignin; TNC = total nonstructural carbohydrates. Total biomass yield and stem weight were reported as grams plot^{-1}, stem thickness is measured in millimeters and cell wall components (NDF, ADF, ADL, TNC, arabinose, glucose, and xylose) were measured as grams kilogram^{-1}. The regrowth score of 0 indicates no growth, whereas 9 indicates abundant growth.

We previously estimated genetic distances among genotypes and among accessions on the basis of 89 SSR loci (Sakiroglu et al., 2010). We computed a correlation between the genetic and phenotypic similarity matrices for individual genotypes and for accessions using Mantel's test (Mantel, 1967) with the software program GenAlEx (Peakall and Smouse, 2001). Neighbor-joining clustering represented in the form of a dendogram was created from the phenotypic distance matrix (NTSYS-pc; Rohlf, 1994) of 120 accessions and presented using Dendroscope (Huson et al., 2007).

RESULTS AND DISCUSSION

Effects of Environments and Locations

Nine traits (NDF, ADF, ADL, TNC, arabinose, glucose, xylose, dry matter yield, and regrowth following harvest) were measured in each of 2 yr at two locations. For these traits, two- and three-way interactions among genotypes, years, and locations were present for all traits, with the exception of year × location interactions for NDF, ADF, and yield (data not shown). Because of the abundant interactions, we subsequently analyzed the data by location and by year. When analyzed by year, location effects and genotype × location effects were present for all traits except yield in both years, NDF and ADF in 2007, and NDF and stem thickness in 2008. When data were analyzed by location, year effects and year × genotype interactions were present for all traits in both the Watkinsville and Eatonton trials, with the exception of regrowth and TNC at Eatonton. Spearman rank correlations between the 2 yr were 0.58 for regrowth, 0.65 for yield, 0.39 for NDF, 0.33 for ADF, 0.40 for ADL, and 0.36 for TNC. Although these rank correlations are statistically significant, they reveal weak relationships between years, especially for composition traits.

In most previous analyses, relatively little genotype × environment interaction has been observed for composition traits in alfalfa (Sheaffer et al., 2000; Sheaffer et al., 1998), even in an analysis of a large germplasm collection (Jung et al., 1997). Further, harvest timing × genotype interactions are also typically absent (Sheaffer et al., 2000; Sheaffer et al., 1998). In our experiment, the two harvests, one in each year, were taken at considerably different times of the year and at different developmental stages of the plant—autumn of the establishment year (2007) and early summer of the first full production year (2008). The effect of plant maturity on forage quality has been documented before (Kalu and Fick, 1981; Fick and Janson, 1990). In 2008 most genotypes had fully flowered at the time we harvested in early June, but in 2007, since harvest was in autumn, genotypes expressed a wide variation in maturity. This could have accounted for some of the genotype × year interaction in this study. The second possible reason for our widespread genotype × year interactions was the age of plants. Plants in the establishment year (2007) may have different cell wall composition than in the first full production year (2008), as was observed by Sheaffer et al. (1998). Thus, the presence of genotype × year interactions in this experiment could be expected. Because of the extensive interactions, we analyzed the data separately for each year–location combination, which we denote as "environments" for convenience.

Trait Means and Correlations among Traits

Total biomass yield was much higher in 2008 than in 2007, reflecting the plants' development (Table 1). Mean regrowth in 2007 was lower than 2008, probably resulting from dormancy-induced growth cessation in at least some germplasms. Mean values for cell wall structural components were higher

and TNC was lower in 2008 than in 2007 at both locations (Table 1). Phenotypic correlations showed that biomass yield between years was correlated ($r = 0.64$; Table 2), similar to the rank correlation mentioned above. Regrowth likewise showed a strong correlation between years. The correlations among cell wall constituents in the same year were moderate ($r \leq 0.50$). Yield in either year was not strongly correlated with any cell wall composition traits in 2007 but moderate to strong correlations were observed in 2008. However, a positive relationship between biomass yield and glucose concentration in both years suggests that higher yields could result in both higher glucose per area harvested and also higher concentrations of glucose. Of course, these are phenotypic, not genetic, correlations, but nevertheless, they are encouraging in terms of biofuel production.

Phenotypic Variation among Alfalfa Accessions

Wide variation among genotypes exists for all the traits evaluated (Table 1), offering considerable potential for selection and for identification of desirable alleles potentially useful for introgression. The wide range in regrowth values following autumn harvest in 2007 suggests that diploid alfalfa germplasm spans a range of dormancy classes. Lignin limits the effective saccarification of cell wall sugars (Chen and Dixon, 2007), and higher lignin concentrations also lower forage nutritive value (Buxton and Russell, 1988; Albrecht et al., 1987; Reddy et al., 2005). Decreasing lignin is therefore desirable for alfalfa forage or for bioethanol production. We found that the lignin level (ADL) in the unimproved diploid alfalfa stems ranged between 58 and 139 g kg^{-1} depending on the genotype and the environment, or 53 to 129% of the mean. This range is comparable to the reported reduction resulting from transgenic downregulation of enzymes involved in lignin biosynthesis (Guo et al., 2001b; Chen and Dixon, 2007). A similar extent of variation exists for other cell wall components and agronomic traits as well. The tetraploid alfalfa core collection had a range of variation in NDF of 13% of the mean; ADF of 15%; and ADL of 26% (Jung et al., 1997). On the basis of our data from Watkinsville in 2008, the location that included all genotypes and fully developed plants, we found the range of variation in NDF was 40% of mean, ADF was 48%, and ADL was 50%. The variation detected in diploid germplasm was two- to four-fold higher than that reported in core collection, although we evaluated individual genotypes rather than comparing overall accessions. The variation in these traits will enable mapping the genetic control of forage quality, biofuel potential, and yield in diploid populations, thereby avoiding the complexity of tetrasomic inheritance. Our analysis of individual genotypes enables us to accurately select parents for future population development, leading to QTL discovery and manipulation.

Table 2. Pearson's correlations coefficients among genotypic means of selected cell wall constituents and yield on plants grown in 2007 and 2008.

Traits†	Year	Yield 2007‡	Yield 2008
NDF	2007	NS	0.16
ADF	2007	NS	0.14
ADL	2007	–0.16	NS
TNC	2007	NS	NS
Glucose	2007	0.29	0.19
Regrowth	2007	0.56	0.42
NDF	2008	0.29	0.54
ADF	2008	0.32	0.55
ADL	2008	0.14	0.37
TNC	2008	–0.18	–0.36
Glucose	2008	0.33	0.54
Plant height	2008	0.34	0.59
Stem thickness	2008	0.38	0.66
Regrowth	2008	0.43	0.48
Yield	2008	0.64	1.00

†NDF = neutral detergent fiber; ADF = acid detergent fiber; ADL = acid detergent lignin; TNC = total nonstructural carbohydrates.

‡Correlations shown are significant at the 1% probability level and NS indicates no significance.

Direct comparison of performance between tetraploid and diploid germplasm is difficult, not only because the germplasm sources may differ, but also because ploidy may affect cell wall composition (McCoy and Bingham, 1988). No examination of the direct effect of polyploidization has been made on alfalfa cell wall composition, but in vitro digestible dry matter was found to be equivalent between diploid and isogenic tetraploid alfalfa lines (Arbi et al., 1978). Tetraploidization increases yield up to twofold, along with an increase of 5 to 22% in TNC (Dunbier et al., 1975). Thus, lignin and cellulose content can be expected to differ across ploidy levels, but an allele contributing to a reduction of lignin or elevation of cellulose may have the same directional effect at both ploidies.

Genotype and Accession Clustering

Our previous molecular marker analysis indicated that diploid germplasm can be differentiated into five main groups, corresponding to northern and southern *M. sativa* subsp. *caerulea*, upland and lowland *M. sativa* subsp. *falcata*, and *M. hemicycla* (Sakiroglu et al., 2010). Therefore, we evaluated differences in phenotypic traits among these five groups. The groups differed for all traits (Table 3). Glucose was the most abundant monosaccharide followed by xylose (Table 3). Since higher glucose and xylose concentrations are desirable for efficient biofuel production, genotypes with high monosaccharide concentration and successful performance in the field in terms of agronomic traits are best for improving alfalfa as a bioenergy crop. Therefore, from a bioenergy perspective, the northern caerulea group and upland falcata group had most value. The northern caerulea group produced the most biomass yield and was

Table 3. Mean of cell wall components and agronomic traits of the five main populations of diploid alfalfa (*Medicago sativa* L.) in the Watkinsville trial over 2 yr (2007 and 2008).

Trait†	Units	Population				
		Southern (A) Caerulea	Northern (B) Caerulea	Hemicycla	Lowland (A) Falcata	Upland (B) Falcata
			2007			
NDF	g kg^{-1}	548a‡	538b	539b	506d	530c
ADF	g kg^{-1}	401a	388b	387b	359d	376c
ADL	g kg^{-1}	103a	96c	100b	89e	93d
TNC	g kg^{-1}	137d	153b	147c	175a	172a
Arabinose	g kg^{-1}	29c	33b	32b	36a	36a
Glucose	g kg^{-1}	260c	271b	258c	271b	276a
Xylose	g kg^{-1}	80b	81a,b	82a	75c	82a
Biomass yield	g plot^{-1}	36c	65a	29d	49b	56b
Stem proportion		0.38a	0.40a	0.38a	0.29c	0.36b
Regrowth	score	2.7c	3.1a,b	2.5d	3.3a	3.0b
			2008			
NDF	g kg^{-1}	637c	668a,b	661b	623d	669a
ADF	g kg^{-1}	479c	506a	498b	466d	507a
ADL	g kg^{-1}	117c	121a,b	122a	112d	120b
TNC	g kg^{-1}	116b	109d	108d	121a	112c
Arabinose	g kg^{-1}	25b	23d	23d	28a	24c
Glucose	g kg^{-1}	310b	322a	311b	304c	321a
Xylose	g kg^{-1}	95b	101a	100a	92c	100a
Biomass yield	g plot^{-1}	332c	605a	266d	286d	441b
Spring growth	score	2.2c	2.8a	1.9d	1.7d	2.6b
Regrowth	score	3.0b	3.1a,b	1.9c	3.2a	3.2a
Plant height	cm	81b	92a	76c	68d	83b
Stem thickness	mm	2.09c	2.76a	2.13c	1.77d	2.46b

†NDF = neutral detergent fiber; ADF = acid detergent fiber; ADL = acid detergent lignin; TNC = total nonstructural carbohydrates.

‡Values within rows followed by different letters are significantly different at $P < 0.05$.

among the highest for regrowth after harvest and xylose concentrations in 2007 and was the highest or was among the highest performing groups for glucose and xylose concentrations, biomass yield, plant height, spring recovery, and regrowth after harvest in 2008 (Table 3). Upland falcata accessions also showed promise by having high glucose, xylose, and yield, although in general, this group has slightly lower overall value than caerulea. The yield of southern caerulea, lowland falcata, and hemicycla were considerably lower than the others, limiting their value for biofuel uses. The northern caerulea accessions PI 577545 and PI 577551 and upland falcata accessions PI 631817

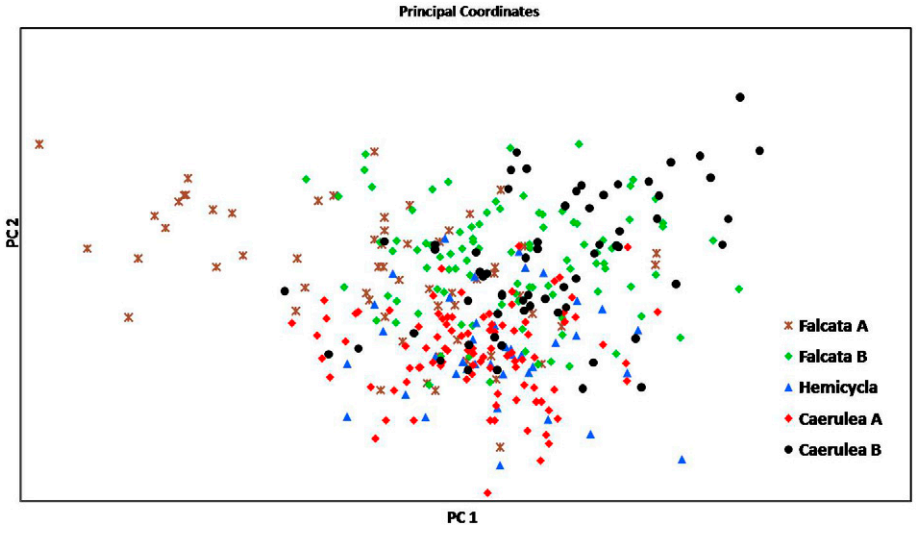

Figure 1. Plot of first two principal components from an analysis of 17 phenotypic traits on 374 diploid alfalfa (*Medicago sativa* L.) genotypes. Individual genotypes are coded as belonging to one of five populations on the basis of molecular-marker clustering conducted previously (Sakiroglu et al., 2010).

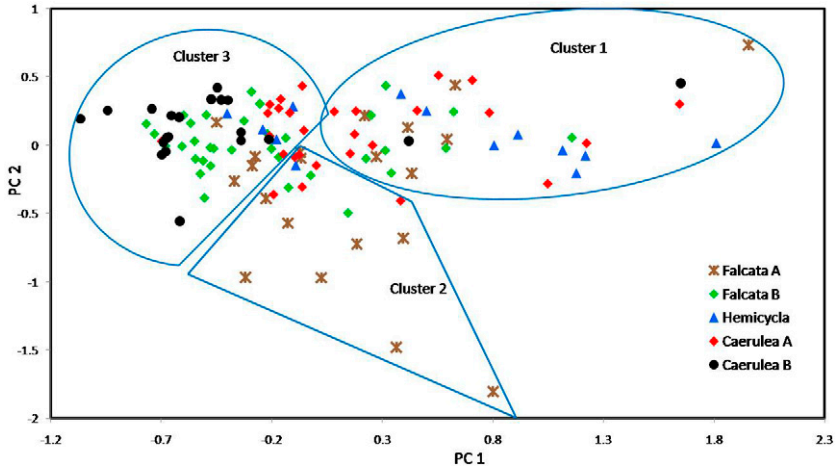

Figure 2. Plot of first two principal components from an analysis of 17 phenotypic traits on 120 diploid alfalfa (*Medicago sativa* L.) accessions. Accession classification based on Sakiroglu et al. (2010). The three clusters indicated in Fig. 3 are shown.

and SD201 had high NDF and yield (see Supplementary Table 1). Owing to the high correlation among cell wall components, high-yielding accessions with higher NDF or ADF values were also high in terms of ADL.

A principal components analysis indicated that the first two principal components accounted for 61% of the variation among genotypes (42% for PC1 and 19% for PC2). A plot of PC1 vs. PC2 did not show any clear clustering of genotypes (Fig. 1). In particular, we were interested to determining if phenotypic differences would reflect the differentiation we had observed using genetic markers. Our evaluation of population structure within diploid *M. sativa* germplasm clearly separated falcata from caerulea, leaving hemicycla in between, and provided strong evidence for further hierarchical separation of both falcata and caerulea into two groups (Sakiroglu et al., 2010). The visual clustering of the phenotypic data did not reflect that separation, although the Mantel test showed a very weak but statistically significant correlation between the phenotypic and genotypic matrices ($R = 0.15$).

We then conducted a similar analysis evaluating accessions, rather than the individual genotypes. The first principal component accounted for 42% and the second principal component accounted for 21% of the variation. Plotting the first two principal components derived from an analysis of the phenotypic data of the 120 accessions also did not cluster

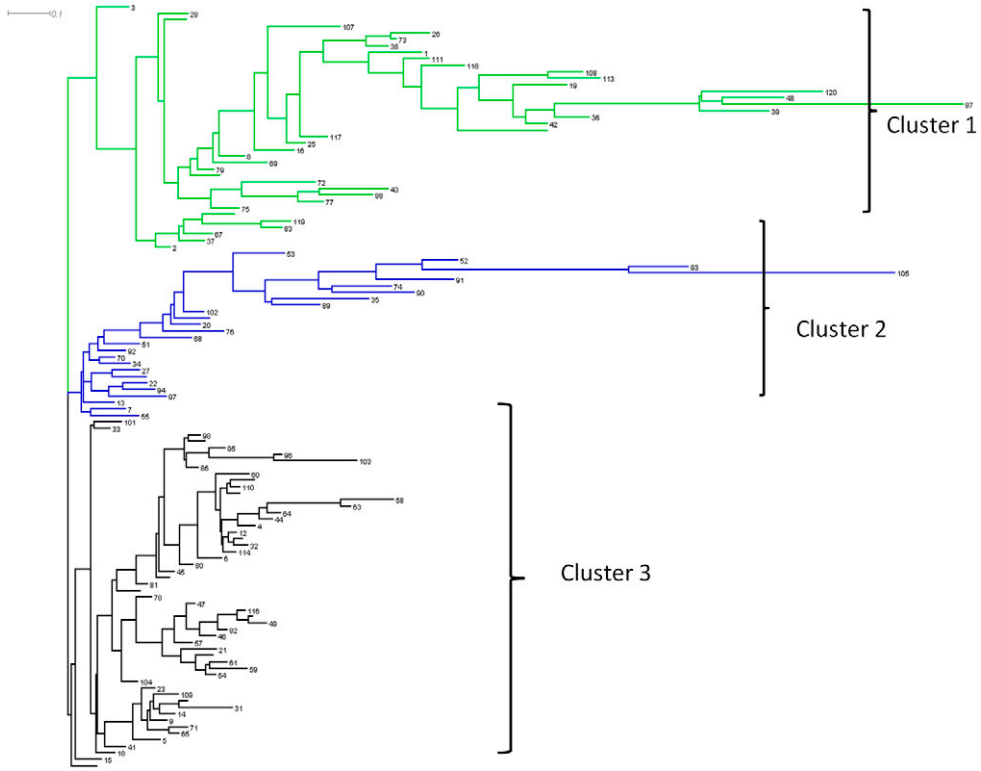

Figure 3. Neighbor-joining cladogram of 120 accessions based on 17 phenotypic traits indicating three distinct clusters of genotypes. Numbers at branch tips correspond to accession numbers in Supplementary Table 1.

accessions into distinct subgroups (Fig. 2), in accordance with the phenotypic analysis of the individual genotypes.

The distance matrix computed on the basis of phenotypic data was used to develop a neighbor-joining dendrogram showing relationships among accessions. Three clusters of accessions were evident (Fig. 3) but these did not correspond to those found with the molecular marker analysis. Each of the three clusters included accessions from all five molecular-marker defined groups (Fig. 2). When we compared the three clusters on the basis of agronomic performance and cell wall composition we found that cluster 3 had the highest performance of total biomass yield and regrowth in both years, whereas cluster 1 performed poorest in those two agronomic traits (data not shown). The other traits did not show a consistent pattern among clusters.

Practical Implications

From a breeding standpoint, the relationship between phenotypic and genetic distance can help to effectively identify valuable germplasm. The lack of a relationship between the molecular-marker-based clustering and phenotypic clustering may imply that desirable alleles for improvement of yield and cell wall traits (either in the context of forage quality or biofuel potential) are dispersed across all subspecies. Alternatively, the markers we tested may not be related to the quantitatively inherited agronomic traits we evaluated, so that genetic distances estimated from them are not reliable for predicting distributions of trait alleles, as suggested previously (Burstin and Charcosset, 1997). Certain germplasm sources, such as the northern caerulea accessions, for example, offer the best starting point for accessing new genes for cultivated alfalfa because they generally have more agronomically desirable trait values than other sources. Nevertheless, other germplasm, particularly falcatas, should be carefully evaluated for the traits of interest as these may contain novel alleles, as suggested by their molecular marker distinctiveness. The next step in using this material is to develop populations, select at the diploid level for agronomic utility, and then scale to the tetraploid level using unreduced gametes, following an analytic breeding scheme (Chase, 1963), or directly tetraploidize individual plants and use them as donors of specific alleles on the basis of future mapping efforts. The key finding in this experiment is that diploid germplasm has wide variation and could serve as a useful reservoir of beneficial alleles for cultivated tetraploid alfalfa.

Acknowledgments

The authors thank Donald Wood, Jonathan Markham, and Wesley Dean for field plot assistance; the Turkish government for funding MS; and the USDA-DOE Plant Feedstock Genomics for Bioenergy award 2006-35300-17224 (to ECB and KJM).

References

Albrecht, K.A., W.F. Wedin, and D.R. Buxton. 1987. Cell-wall composition and digestibility of alfalfa stems and leaves. Crop Sci. 27:735–741. doi:10.2135/cropsci1987.0011183X002700040027x

Arbi, N., D. Smith, E.T. Bingham, and R.M. Soberalske. 1978. Herbage Yields and Levels of N and IVDDM from Five Alfalfa Strains of Different Ploidy Levels. Agron. J. 70:873–875. doi:10.2134/agronj1978.00021962007000050039x

Baucher, M., M.A. BernardVailhe, B. Chabbert, J.M. Besle, C. Opsomer, M. VanMontagu, and J. Botterman. 1999. Down-regulation of cinnamyl alcohol dehydrogenase in transgenic alfalfa (*Medicago sativa* L.) and the effect on lignin composition and digestibility. Plant Mol. Biol. 39:437–447. doi:10.1023/A:1006182925584

Brummer, E.C. 1999. Capturing heterosis in forage crop cultivar development. Crop Sci. 39:943–954. doi:10.2135/cropsci1999.0011183X003900040001x

Burstin, J., and A. Charcosset. 1997. Relationship between phenotypic and marker distances: Theoretical and experimental investigations. Heredity 79:477–483. doi:10.1038/hdy.1997.187

Buxton, D.R., and J.R. Russell. 1988. Lignin constituents and cell-wall digestibility of grass and legume stems. Crop Sci. 28:553–558. doi:10.2135/cropsci1988.0011183X002800030026x

Chase, S.S. 1963. Analytic breeding in *Solanum tuberosum* L.: A scheme utilizing parthenotes and other diploid stocks. Can. J. Genet. Cytol. 5:359–363.

Chen, F., and R.A. Dixon. 2007. Lignin modification improves fermentable sugar yields for biofuel production. Nat. Biotechnol. 25:759–761. doi:10.1038/nbt1316

Coors, J.G., C.C. Lowe, and R.P. Murphy. 1986. Selection for improved nutritional quality of alfalfa forage. Crop Sci. 26:843–848. doi:10.2135/cropsci1986.0011183X002600050001x

Crews, T.E., and M.B. Peoples. 2004. Legume versus fertilizer sources of nitrogen: Ecological tradeoffs and human needs. Agric. Ecosyst. Environ. 102:279–297. doi:10.1016/j.agee.2003.09.018

Delong, M.M., D.R. Swanberg, E.A. Oelke, C. Hanson, M. Onischak, M.R. Schmid, and B.C. Wiant. 1995. Sustainable biomass energy production and rural economic development using alfalfa as a feedstock. p. 1582–1591. *In* D.L. Klass (ed.) Second Biomass Conf. of the Americas: Energy, Environment, Agriculture, and Industry, 21–24 Aug. 1995, Portland, OR.

Demment, M.W., L.R. Teuber, D.P. Bourque, and D.A. Phillips. 1986. Changes in forage quality of improved alfalfa populations. Crop Sci. 26:1137–1143. doi:10.2135/cropsci1986.0011183X002600060011x

Dien, B.S., H.-J.G. Jung, K.P. Vogel, et al. 2006. Chemical composition and response to dilute-acid pretreatment and enzymatic saccharification of alfalfa, reed canarygrass, and switchgrass. Biomass Bioenergy 30:880–891. doi:10.1016/j.biombioe.2006.02.004

Dunbier, M.W., D.L. Eskew, E.T. Bingham, and L.E. Schrader. 1975. Performance of genetically comparable diploid and tetraploid alfalfa: Agronomic and physiological parameters. Crop Sci. 15:211–214. doi:10.2135/cropsci1975.0011183X001500020021x

Farrokhi, N., R.A. Burton, L. Brownfield, M. Hrmova, S.M. Wilson, A. Bacic, and G.B. Fincher. 2006. Plant cell wall biosynthesis: Genetic, biochemical and functional genomics approaches to the identification of key genes. Plant Biotechnol. J. 4:145–167. doi:10.1111/j.1467-7652.2005.00169.x

Fick, G.W., and C.G. Janson. 1990. Testing mean stage as a predictor of alfalfa forage quality with growth chamber trials. Crop Sci.

30:678–682. doi:10.2135/cropsci1990.0011183X003000030040x

Guo, D., F. Chen, K. Inoue, et al. 2001a. Down-regulation of Caffeic Acid 3-*O*-methyltransferase and Caffeoyl CoA 3-*O*-methyltransferase in transgenic alfalfa (*Medicago sativa* L.): Impacts on lignin structure and implications for the biosynthesis of G and S lignin. Plant Cell 13:73–88. doi:10.1105/tpc.13.1.73

Guo, D., F. Chen, J. Wheeler, et al. 2001b. Improvement of in-rumen digestibility of alfalfa forage by genetic manipulation of lignin *O*-methyltransferases. Transgenic Res. 10:457–464. doi:10.1023/A:1012278106147

Hill, R.R., Jr., J.S. Shenk, and R.F. Barnes. 1988. Breeding for yield and quality. p. 809–825. *In* A.A. Hanson et al. (ed.) Alfalfa and alfalfa improvement. ASA, CSSA, SSSA, Madison, WI.

Huson, D.H., D.C. Richter, C. Rausch, T. Dezuian, M. Franz, and R. Rupp. 2007. Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinformatics 8:460. doi:10.1186/1471-2105-8-460

Jackson, L.A., G.L. Shadle, R. Zhou, J. Nakashima, F. Chen, and R.A. Dixon. 2008. Improving saccharification efficiency of alfalfa stems through modification of the terminal stages of monolignol biosynthesis. Bioenergy Res. 1:180–192. doi:10.1007/s12155-008-9020-z

Jung, H.G., C.C. Sheaffer, D.K. Barnes, and J.L. Halgerson. 1997. Forage quality variation in the U.S. alfalfa core collection. Crop Sci. 37:1361–1366. doi:10.2135/cropsci1997.0011183X003700040054x

Kalu, B.A., and G.W. Fick. 1981. Quantifying morphological development of alfalfa for studies of herbage quality. Crop Sci. 21:267–271. doi:10.2135/cropsci1981.0011183X002100020016x

Lamb, J.F.S., H.G. Jung, C.C. Sheaffer, and D.A. Samac. 2007. Alfalfa leaf protein and stem cell wall polysaccharide yields under hay and biomass management systems. Crop Sci. 47:1407–1415. doi:10.2135/cropsci2006.10.0665

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27:209–220.

Maureira, I.J., F. Ortega, H. Campos, and T.C. Osborn. 2004. Population structure and combining ability of diverse *Medicago sativa* germplasms. Theor. Appl. Genet. 109:775–782. doi:10.1007/s00122-004-1677-x

McCaslin, M., and D. Miller. 2007. The future of alfalfa as a biofuels feedstock. 37th California Alfalfa & Forage Symposium. 17–19 Dec. 2007, Monterey, CA.

McCoy, T.J., and E.T. Bingham. 1988. Cytology and cytogenetics of alfalfa. p. 739–776. *In* A.A. Hanson et al. (ed.) Alfalfa and alfalfa improvement. ASA, CSSA, SSSA, Madison, WI.

Michaud, R., W.F. Lehman, and M.D. Rumbaugh. 1988. World distribution and historical development p. 26–82. *In* A.A. Hanson et al. (ed.) Alfalfa and alfalfa improvement. ASA, CSSA, SSSA, Madison, WI.

Nakashima, J., F. Chen, L. Jackson, et al. 2008. Multi-site genetic modification of monolignol biosynthesis in alfalfa (*Medicago sativa* L.): Effects on lignin composition in specific cell types. New Phytol. 179:738–750. doi:10.1111/j.1469-8137.2008.02502.x

Patzek, T. 2004. Thermodynamics of the corn-ethanol biofuel cycle. Crit. Rev. Plant Sci. 23:519–567. doi:10.1080/07352680490886905

Peakall, R., and P.E. Smouse. 2001. GenAlEx V5; Genetic analysis in excel. Population genetic software for teaching and research. Australian National Univ., Canberra, Australia. http://www.anu.edu.au/BoZo/GenAlEx (verified 14 June 2011).

Quiros, C.F., and G.R. Bauchan. 1988. The genus *Medicago* and the origin of the *Medicago sativa* complex. p. 93–124. *In* A.A. Hanson et al. (ed.) Alfalfa and alfalfa improvement. ASA, CSSA, SSSA, Madison, WI.

Ragauskas, A.J., C.K. Williams, B.H. Davison, G. Britovsek, J. Cairney, C.A. Eckert, W.J. Frederick, Jr., J.P. Hallett, D.J. Leak, C.L. Liotta, J.R. Mielenz, R. Murphy, R. Templer, and T. Tschaplinski. 2006. The path forward for biofuels and biomaterials. Science 311:484–489. doi:10.1126/science.1114736

Reddy, M.S.S., F. Chen, G. Shadle, L. Jackson, H. Aljoe, and R.A. Dixon. 2005. Targeted down-regulation of cytochrome P450 enzymes for forage quality improvement in alfalfa (*Medicago sativa* L.). Proc. Natl. Acad. Sci. USA 102:16573–16578. doi:10.1073/pnas.0505749102

Riday, H., and E.C. Brummer. 2002. Forage yield heterosis in alfalfa. Crop Sci. 42:716–723. doi:10.2135/cropsci2002.0716

Rohlf, F.J. 1994. NTSYS-pc: Numerical taxonomy and multivariate analysis system, Ver. 1.80. Exeter Software, Setauket, NY 42:1081–1087.

Sakiroglu, M., and E.C. Brummer. 2007. Little heterosis between alfalfa populations derived from the midwestern and southwestern United States. Crop Sci. 47:2364–2371. doi:10.2135/cropsci2006.12.0803

Sakiroglu, 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. Theor. Appl. Genet. 121:403–415. doi:10.1007/s00122-010-1319-4

SAS Institute. 2004. SAS 9.1.2. SAS Institute, Cary, NC.

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. doi:10.1139/g02-100

Sheaffer, C.C., D. Cash, N.J. Ehlke, J.C. Henning, J.G. Jewett, K.D. Johnson, M.A. Peterson, M. Smith, J.L. Hansen, and D.R. Viands. 1998. Entry × environment interaction for alfalfa forage quality. Agron. J. 90:774–780. doi:10.2134/agronj1998.00021962009000060010x

Sheaffer, C.C., N.P. Martin, J.F.S. Lamb, G.R. Cuomo, J.G. Jewett, and S.R. Quering. 2000. Leaf and stem properties of alfalfa entries. Agron. J. 92:733–739. doi:10.2134/agronj2000.924733x

Shenk, J.S., and F.C. Elliott. 1971. Plant compositional changes resulting from two cycles of directional selection for nutritive value in alfalfa. Crop Sci. 11:521–524. doi:10.2135/cropsci1971.0011183X001100040017x

Shenk, J.S., and M.O. Westerhaus. 1991. Population structuring of near infrared spectra and modified least squares regression. Crop Sci. 31:1548–1555. doi:10.2135/cropsci1991.0011183X003100060034x

Theander, O., P. Aman, E. Westerlund, R. Andersson, and D. Pettersson. 1995. Total dietary fiber determined as neutral sugar residues, uronic acid residues, and klason lignin (The Uppsala Method): Collaborative study. J. AOAC Int. 78:1030–1044.

Van Soest, P.J., J.B. Robertson, and B.A. Lewis. 1991. Symposium: Carbohydrate methodology, metabolism, and nutritional implications in dairy cattle. J. Dairy Sci. 74:3583–3597. doi:10.3168/jds.S0022-0302(91)78551-2

Vogel, K.P., J.F. Pedersen, S.D. Masterson, and J.J. Troy. 1999. Evaluation of a filter bag system for NDF, ADF, and IVDMD forage analysis. Crop Sci. 39:276–279. doi:10.2135/cropsci1999.0011183X003900010042x

Windham, W.R., D.R. Mertens, and F.E. Barton, II. 1989. Protocol for NIRS calibration: Sample selection and equation development and validation. p. 96–103. *In* G.C. Marten et al. (ed.) Near infrared reflectance spectroscopy (NIRS): Analysis of forage quality. USDA Agric. Handbook 643.