An Integrated Genetic Linkage Map of the Soybean Genome

P. B. Cregan,* T. Jarvik, A. L. Bush, R. C. Shoemaker, K. G. Lark, A. L. Kahler, N. Kaya, T. T. VanToai, D. G. Lohnes, J. Chung, and J. E. Specht

ABSTRACT

A number of molecular genetic maps of the soybean [*Glycine max* (L.) Merr.] have been developed over the past 10 yr. These maps are primarily based on restriction fragment length polymorphism (RFLP) markers. Parental surveys have shown that most RFLP loci have only two known alleles. However, because the soybean is an ancient polyploid, RFLP probes typically hybridize and map to more than one position in the genome. Thus, the polymorphic potential of an RFLP probe is primarily a function of the frequency of the two alleles at each locus the probe detects. In contrast, simple sequence repeat (SSR) markers are single locus markers with multiple alleles. The polymorphic potential of an SSR marker is dependent on the number of alleles and their frequencies. Single locus markers provide an unambiguous means of defining linkage group homology across mapping populations. The objective of the work reported here was to develop and map a large set of SSR markers. A total of 606 SSR loci were mapped in one or more of three populations: the USDA/Iowa State *G. max* × *G. soja* F_2 population, the Univ. of Utah Minsoy × Noir 1 recombinant inbred population, and the Univ. of Nebraska Clark × Harosoy F_2 population. Each SSR mapped to a single locus in the genome, with a map order that was essentially identical in all three populations. Many SSR loci were segregating in two or all three populations. Thus, it was relatively simple to align the 20+ linkage groups derived from each of the three populations into a consensus set of 20 homologous linkage groups presumed to correspond to the 20 pairs of soybean chromosomes. On the basis of in situ segregation or linkage reports in the literature all but one of the classical linkage groups can now be assigned to a corresponding molecular linkage group.

Genetic linkage maps serve the plant geneticist in a number of ways, from marker assisted selection in plant improvement to map-based cloning in molecular genetic research. Thus, in a widely studied and economically important species such as soybean, a well developed and broadly useful linkage map would be a valuable resource. Ideally, such a map should include many classical markers with discernible phenotypic effects, isozyme markers, as well as a large number of highly informative DNA markers evenly spaced throughout the genome.

Using a mapping population derived from an interspecific *G. max* × *G. soja* cross, Shoemaker and Olson (1993) developed a molecular genetic linkage map that consisted of 25 linkage groups with about 365 RFLP, 11 RAPD (random amplified polymorphic DNA), three classical markers, and four isozyme loci. The current soybean classical marker map consists of 68 loci dispersed among 20 small linkage groups with a few loci each (Palmer and Shoemaker, 1998). A partial integration of the various marker types into a common linkage map was recently achieved by Shoemaker and Specht (1995). These authors used a soybean mapping population derived from a mating of near-isogenic lines of the cultivars Clark and Harosoy to create a linkage map that included 13 classical and 7 isozyme loci along with 110 RFLP and 8 RAPD loci. A set of anchoring RFLP loci that segregated in both the Clark × Harosoy and the *G. max* × *G. soja* mapping populations was used to identify linkage group homologies between the molecular and classical marker maps.

Two characteristics of RFLP markers in soybean tend to complicate the task of consolidating linkage maps from different mapping populations. First, only rarely have more than two alleles been identified at RFLP loci in soybean. Because these two alleles generally have asymmetric frequencies, e.g., $p > 0.9$, $q < 0.1$ (Keim et al., 1989; Keim et al., 1992), the likelihood that any two genotypes will be polymorphic at a particular RFLP locus is relatively low. This is particularly true when both parents of the mapping population come from adapted soybean germplasm pools (Apuya et al., 1988; Lark et al., 1993). For example, Muehlbauer et al. (1991) also observed that only one-third of the available RFLP probes tested on donor parent, near-isogenic line, or recurrent parent triplets were actually polymorphic between the two parents. Similarly, Shoemaker and Specht (1995) reported that only 118 of 365 RFLP markers polymorphic in the *G. max* × *G. soja* population were segregating in the Clark × Harosoy mapping population. Thus, a polymorphic fragment mapped in one population may not be segregating in another. A second factor that complicates the use of RFLP markers in soybean is the detection of multiple DNA fragments (i.e., multiple loci) with most probes. This may be the result of the tetraploid origin of soybean (Hymowitz

P.B. Cregan, USDA-ARS, Soybean and Alfalfa Research Lab., Beltsville, MD 20705; T. Jarvik and K.G. Lark, Dep. of Biology, Univ. of Utah, Salt Lake City, UT 84112; A.L. Bush and R.C. Shoemaker, USDA-ARS-CICG, Dep. of Agronomy, Iowa State Univ., Ames, IA 50011; A.L. Kahler, Biogenetic Services Inc., 801 32 Ave., Brookings, SD 57006; N. Kaya, Yuzuncu Yil Univ., College of Agriculture, 65080-Van, Turkey; T.T. vanToai, USDA-ARS, Soil Drainage Unit, 590 Woody Hayes Drive, Columbus, OH 43210; D.G. Lohnes, Dep. of Horticulture and Crop Science, The Ohio State Univ. and the Ohio Agric. Res. and Dev. Center, Wooster, OH 44691; J. Chung, Dep. of Agronomy, Gyeong Sang National Univ., Chinju, Korea 660-701; and J.E. Specht, Dep. of Agronomy, Univ. Nebraska, Lincoln, NE 68583-0915. The mention of trade or manufacturer names is made for information only and does not imply an endorsement, recommendation, or exclusion by USDA-ARS. Received 12 Jan. 1998. *Corresponding author (pcregan@nal.usda.gov).

Published in Crop Sci. 39:1464–1490 (1999).

Abbreviations: AFLP, amplified fragment length polymorphism; BAC, bacterial artificial chromosome; CLG, classical linkage group; cM, centimorgan; MLG, molecular linkage group; NIL, near isogenic line; PCR, polymerase chain reaction; QTL, quantitative trait loci; RAPD, random amplified polymorphic DNA; RIL, recombinant inbred line; RFLP, restriction fragment length polymorphism; SSR, simple sequence repeat.

and Singh, 1987). Because one fragment in a multiple banding pattern may segregate in one population and a different or an additional fragment in another, one must define a RFLP locus not only by the probe and restriction enzyme being used, but also by the molecular weight of the segregating fragments(s). Up to 19 independent loci have been mapped by specific RFLP probes (Mansur et al., 1996). The multiplicity of RFLP loci can make RFLP linkage maps ambiguous with respect to RFLP locus identity, and often precludes the use of such loci for the evaluation of linkage group homology among different maps.

One possible solution to this complexity is the use of single locus DNA markers with multiple alleles. In soybean, the highly polymorphic nature (i.e., multi-allelism) of simple sequence repeat (SSR) or microsatellite DNA markers is quite clear as shown by initial work of Akkaya et al. (1992) and Morgante and Olivieri (1993). Subsequent reports (Rongwen et al., 1995; Maughan et al., 1995; Powell et al., 1996; Diwan and Cregan, 1997) have described highly polymorphic microsatellite loci with as many as 26 alleles. A high level of allelic diversity makes it likely that a particular SSR locus will be polymorphic in many of the two-parent populations derived from the hybridization of adapted soybean genotypes. Multiple allele molecular markers are much more useful than dimorphic markers when tracking the fate of genomic segments in multi-parent pedigrees and in multi-parent mated populations. Another virtue of SSR markers is their simplicity. In the development of these markers, care is taken to select polymerase chain reaction (PCR) primers that produce one amplification product in an inbred soybean genotype (Cregan et al., 1994). Primer sets producing more than one product are discarded. Thus, the difficulty of the genetic interpretation of multiple banding patterns is eliminated. In addition, as reported in humans and other mammalian species, microsatellite loci in soybean seem to distribute fairly randomly throughout the genome, with minimal evidence of clustering (Akkaya et al., 1995).

While extensive SSR or microsatellite DNA based maps are now available and used by human and other mammalian geneticists (Hudson et al., 1995; Dietrich et al., 1994; Archibald et al., 1995), relatively smaller numbers of SSR markers have been developed and integrated into existing plant linkage maps. Bell and Ecker (1994) reported the assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Akkaya et al. (1995) integrated 40 SSR markers into a soybean linkage map and Mansur et al. (1996) added another 22 loci shortly thereafter. Senior et al. (1996) mapped 42 distinct GenBank-derived SSR loci in maize (*Zea mays* L.). More recently, Sharon et al. (1997) mapped 25 SSR loci on a genetic linkage map of avocado (*persea americana* Mill.). Development and mapping of wheat (*Triticum aestivum* L.) microsatellite loci on a similar scale has been reported (Röder et al., 1995; and Bryan et al., 1997).

The availability of a molecular genetic map saturated with highly informative, PCR-based, single-locus, multiple-allele molecular markers would be of great utility to a wide range of soybean researchers. Therefore, the first objective of the work reported here was to develop a set of SSR markers that could be used under a standard set of amplification conditions. The second objective was to map those SSR loci in three existing mapping populations to attempt the alignment of homologous linkage groups and the identification of 20 consensus linkage groups corresponding to the 20 pairs of soybean chromosomes.

MATERIALS AND METHODS

Development of Simple Sequence Repeat Markers

The selection of SSR-containing sequences from GenBank and the basic procedures of cloning, identification, and sequencing of microsatellite-containing 500- to 700-bp genomic clones of 'Williams' soybean DNA were described previously (Cregan et al., 1994; Akkaya et al., 1995). One important difference in the development of the SSR loci reported here was the use of OLIGO (National Biolabs, St. Paul, MN) software for PCR primer selection. Primers were selected by a number of criteria that included (i) annealing temperature of 47 ± 0.5°C, (ii) minimal 3′ end homologies, (iii) low 3′ end stability to prevent false priming, and (iv) the presence of a GC-clamp, if possible, near the 5′ end of each primer.

Each selected primer pair was initially tested in two different PCR amplification reactions. The first used as template the plasmid containing the Williams soybean genomic insert from which the sequence data for the selection of the primer set was selected. The second reaction used genomic DNA of Williams soybean as template. When both amplification reactions yielded a single product of predicted size, the primer set was further tested on a set of 10 soybean genotypes in order to obtain an estimate of the level of SSR length polymorphism associated with each locus. The genotypes included the cultivars Clark (Maturity Group [MG] IV), Harosoy (MG II), Jackson (MG VII), Williams (MG III), Amsoy (MG II), Archer (MG I), Fiskeby V (MG 000), Minsoy (MG 0), Noir 1 (MG 0), and Tokyo (MG VII). Primer sets that produced multiple products in any of the 10 genotypes were discarded. The size of the alleles (i.e., base pair number in the PCR products) produced by each genotype was determined with DNA sequencing gels with a modified sequencing gel formulation (6% [w/v] acrylamide:bis-acrylamide [19:1], 5.6 *M* ultrapure urea, and 30% [v/v] formamide in TBE buffer) as described by Cregan and Quigley (1997). These allele size estimates were used to calculate the gene diversity or informativeness of each SSR locus. This calculation was described by Anderson et al. (1993).

Soybean Mapping Populations and Genetic Marker Data

USDA/Iowa State University Population. This is an F_2-derived mapping population from the interspecific cross of the *G. max* Breeding Line A81-356022 and *G. soja* (wild soybean) PI 468.916. This population currently consists of 59 F_2 plant derivatives and has been described in detail (Shoemaker and Specht, 1995; Shoemaker and Olson, 1993). The extraction of DNA and the mapping of RFLP loci in this population was previously described by Keim et al. (1988). DNA isolation from the *G. max* × *G. soja* population and RFLP probe hybridization procedures were the same as previously described (Diers et al., 1992a, c; Keim et al., 1990).

University of Utah Recombinant Inbred Line (RIL) Population. Originating from a cross of Minsoy × Noir I, this

population consists of 240 F_7-derived RIL and has been described previously (Mansur et al., 1996). The DNA isolation and RFLP analysis were described by Lark et al. (1993) and Mansur et al. (1996).

University of Nebraska Population. This is an F_2 -derived population from the cross of near isogenic lines (NILs) of the important cultivars Clark and Harosoy. Each of the Clark and Harosoy NILs used as the parents carries a number of pigmentation and/or morphological mutants thereby allowing the mapping of these classical genetic loci along with molecular loci. The population consists of derivatives of 57 F_2 plants and previously was described by Shoemaker and Specht (1995). The isolation of DNA and the RFLP analysis was described earlier (Shoemaker and Specht, 1995).

The segregation of alleles at each SSR locus in each of the above populations was determined by amplifying template DNA from each RIL or F_2 derivative followed by electrophoretic separation of the resulting products on DNA sequencing gels. These procedures are described in Cregan and Quigley (1997).

Genetic Mapping

MAPMAKER 3.0b (Lander et al., 1987; Lincoln and Lander, 1993) was used to group and order genetic loci within each of the three mapping populations. Marker loci were first grouped at LOD 5.0 and then ordered by repetitive use of the Ripple command of MAPMAKER with a window size of 6. The Kosambi centimorgan function was used with error detection on. The error detection probability level was set at 5% in the case of the USDA/Iowa State *G. max* × *G. soja* and University of Nebraska Clark × Harosoy populations and at 1% in the analysis of the Univ. of Utah Minsoy × Noir 1 population. The absence of heterozygous genotypes in the latter population made genotypic classification less prone to error.

RESULTS AND DISCUSSION

A total of 606 SSR loci were mapped in one or more of the three mapping populations. This number includes 544 new loci that were not previously reported by either Akkaya et al. (1995) or Mansur et al. (1996). One important criterion used during the development of these loci was the requirement that each primer set produce only a single PCR product with each of 10 soybean genotypes as described above. Thus, each of the SSR loci reported here maps to a single locus. This avoids the ambiguity that sometimes results with soybean RFLP probes that hybridize to two or more positions in the soybean genome.

The total numbers of markers (SSR, RFLP, RAPD, AFLP, isozyme, and classical) mapped in any one of the three populations ranged from 523 to 1004 (Table 1). The total number of unique loci in the three maps combined totaled 1423, which included 606 SSR, 689 RFLP, 79 RAPD, 11 AFLP, 10 isozyme, and 26 classical loci. Before the inclusion of the SSRs in the *G. max* × *G. soja* map, the last reported MAPMAKER analysis indicated a total of 25 linkage groups (Shoemaker and Specht, 1995). This number was reduced to 23 by the addition of the SSR loci. The addition of well over 300 new SSR loci more than doubled the number of markers on the Univ. of Utah map and reduced the number of linkage groups from 36 to 22. By aligning the linkage groups in these two maps with the linkage groups of the Clark × Harosoy population, on the basis of the presence of common SSR loci, a total of 20 linkage groups are now readily discernable (Fig. 1). For example, the two separate *G. max* × *G. soja* linkage groups D1a and Q are very likely one linkage group given SSR loci they have in common with D1a+Q-U08 of the Minsoy × Noir 1 map (Fig. 1, Panel MLG D1a+Q). Indeed if the LOD threshold is reduced to 2.7, MAPMAKER joins D1a and Q. On the basis of a similar rationale, *G. max* × *G. soja* groups W and D1b (Fig. 1, Panel MLG D1b+W) were combined on the basis of alignment with the homologous Clark × Harosoy linkage groups. Similarly, Utah linkage groups U18 and U02 were joined (Fig. 1, Panel MLG E), as were U13a and U13b (Fig. 1, Panel MLG F) on the basis of alignments with linkage groups E-ISU and F-ISU, respectively, of the *G. max* × *G. soja* map. The alignment of linkage groups based upon the presence of common SSR loci across the three maps resulted in the establishment of 20 consensus linkage groups (Fig. 1). In the case of both the *G. max* × *G. soja* and the Clark × Harosoy maps, only one (Y-ISU) and two (CH54 and CH24) small linkage groups, respectively, could not be aligned with any of the 20 consensus groups. The fact that these remaining small groups contain no SSR loci (Fig. 1, Panel MLG Y + Unlinked) made their alignment particularly difficult.

As the development of the *G. max* × *G. soja* and the Minsoy × Noir 1 maps progressed over the past few years, consolidations and occasional subdivisions of linkage groups have occurred. In addition, the names of linkage groups have been changed. As a result, it may be difficult to relate linkage groups reported in older literature with one of the 20 consensus linkage groups shown in Fig. 1. Table 2 provides a historical summary of how linkage groups in the *G. max* × *G. soja* and the Minsoy × Noir 1 maps have been combined and/or renamed to yield the 20 consensus groups presented in Fig. 1. We have assigned new linkage group

Table 1. Numbers of SSR, RFLP, RAPD, isozyme, and classical genetic markers mapped in the USDA/Iowa State Univ., the Univ. of Utah, and the Univ. of Nebraska mapping populations.

Mapping population	Marker total	Marker type					
		SSR	RFLP	RAPD	AFLP	Isozyme	Classical
	no.						
USDA/Iowa State Univ. (A81-356022 × G. *soja* PI 468.916)	1004	486	501	10	0	4	3
Univ. of Utah (Minsoy × Noir 1)	633	412	209	0	0	2	10
Univ. of Nebraska (Clark × Harosoy)	523	339	95	57	11	7	14

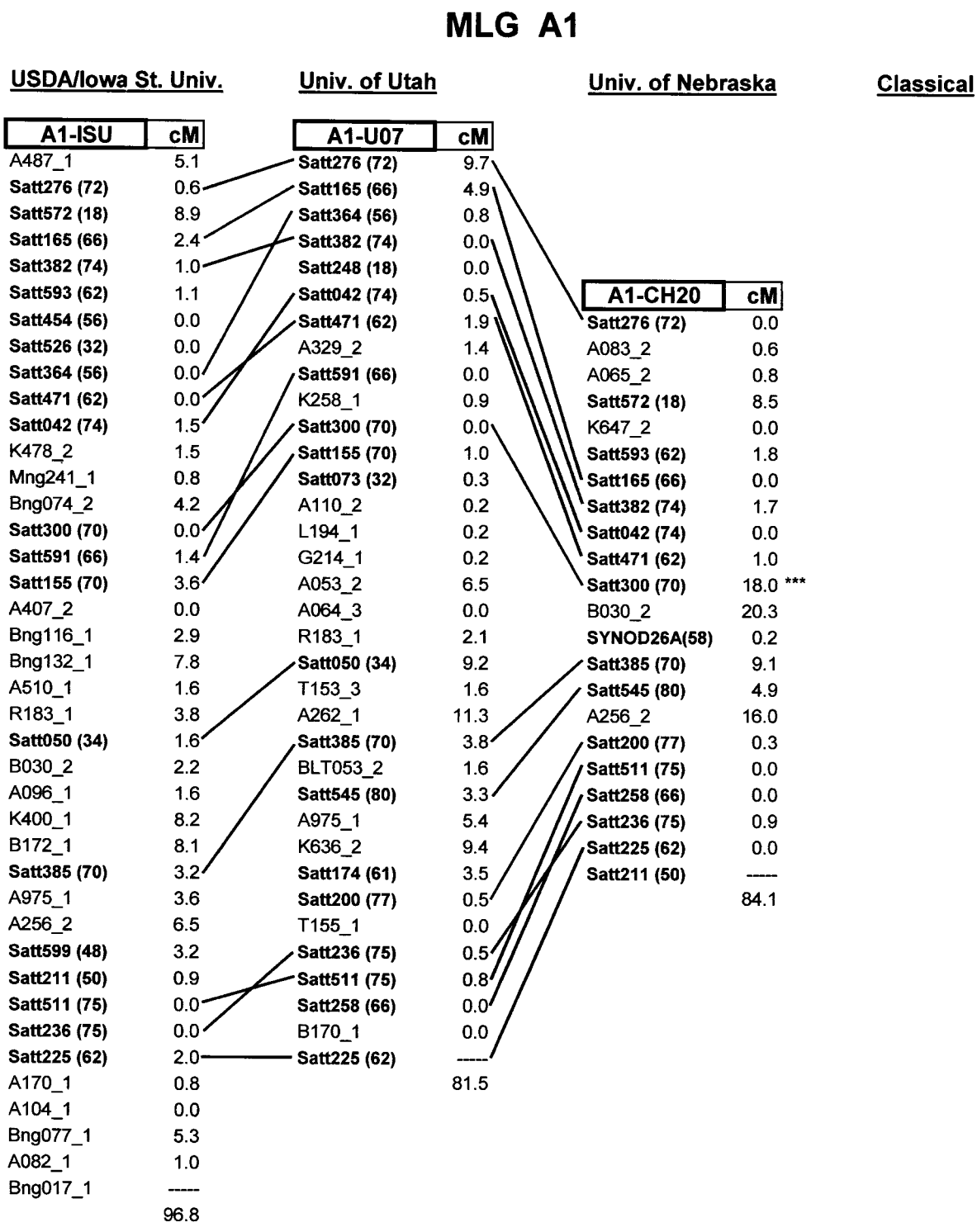
MLG A1
USDA/Iowa St. Univ.
Univ. of Utah
Univ. of Nebraska
Classical
A1-ISU
cM
A487_1 5.1
Satt276 (72) 0.6
Satt572 (18) 8.9
Satt165 (66) 2.4
Satt382 (74) 1.0
Satt593 (62) 1.1
Satt454 (56) 0.0
Satt526 (32) 0.0
Satt364 (56) 0.0
Satt471 (62) 0.0
Satt042 (74) 1.5
K478_2 1.5
Mng241_1 0.8
Bng074_2 4.2
Satt300 (70) 0.0
Satt591 (66) 1.4
Satt155 (70) 3.6
A407_2 0.0
Bng116_1 2.9
Bng132_1 7.8
A510_1 1.6
R183_1 3.8
Satt050 (34) 1.6
B030_2 2.2
A096_1 1.6
K400_1 8.2
B172_1 8.1
Satt385 (70) 3.2
A975_1 3.6
A256_2 6.5
Satt599 (48) 3.2
Satt211 (50) 0.9
Satt511 (75) 0.0
Satt236 (75) 0.0
Satt225 (62) 2.0
A170_1 0.8
A104_1 0.0
Bng077_1 5.3
A082_1 1.0
Bng017_1 -----
96.8
A1-U07
cM
Satt276 (72) 9.7
Satt165 (66) 4.9
Satt364 (56) 0.8
Satt382 (74) 0.0
Satt248 (18) 0.0
Satt042 (74) 0.5
Satt471 (62) 1.9
A329_2 1.4
Satt591 (66) 0.0
K258_1 0.9
Satt300 (70) 0.0
Satt155 (70) 1.0
Satt073 (32) 0.3
A110_2 0.2
L194_1 0.2
G214_1 0.2
A053_2 6.5
A064_3 0.0
R183_1 2.1
Satt050 (34) 9.2
T153_3 1.6
A262_1 11.3
Satt385 (70) 3.8
BLT053_2 1.6
Satt545 (80) 3.3
A975_1 5.4
K636_2 9.4
Satt174 (61) 3.5
Satt200 (77) 0.5
T155_1 0.0
Satt236 (75) 0.5
Satt511 (75) 0.8
Satt258 (66) 0.0
B170_1 0.0
Satt225 (62) -----
81.5
A1-CH20
cM
Satt276 (72) 0.0
A083_2 0.6
A065_2 0.8
Satt572 (18) 8.5
K647_2 0.0
Satt593 (62) 1.8
Satt165 (66) 0.0
Satt382 (74) 1.7
Satt042 (74) 0.0
Satt471 (62) 1.0
Satt300 (70) 18.0 ***
B030_2 20.3
SYNOD26A(58) 0.2
Satt385 (70) 9.1
Satt545 (80) 4.9
A256_2 16.0
Satt200 (77) 0.3
Satt511 (75) 0.0
Satt258 (66) 0.0
Satt236 (75) 0.9
Satt225 (62) 0.0
Satt211 (50) -----
84.1

Fig. 1.

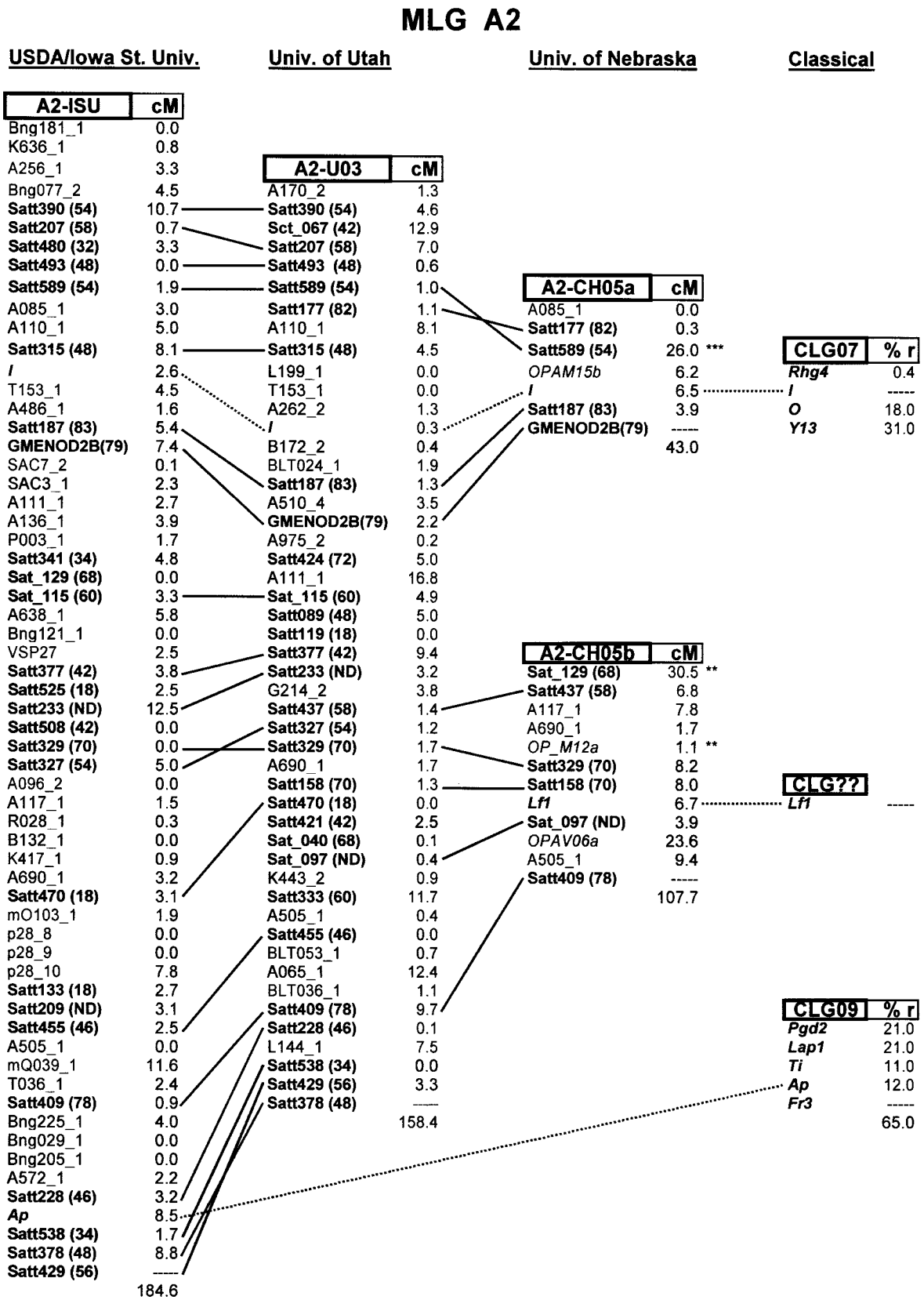

Fig. 1. continued

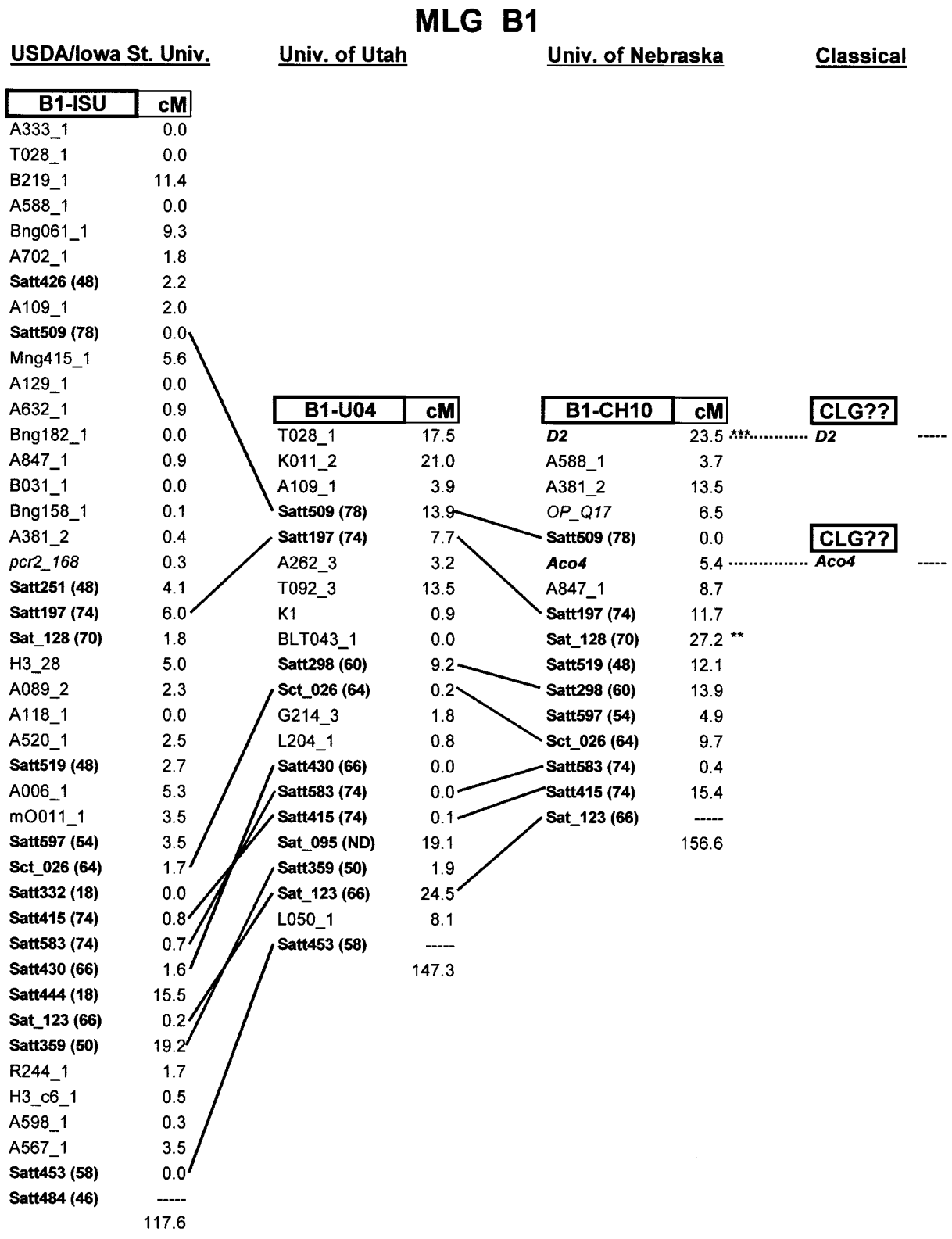

Fig. 1. continued

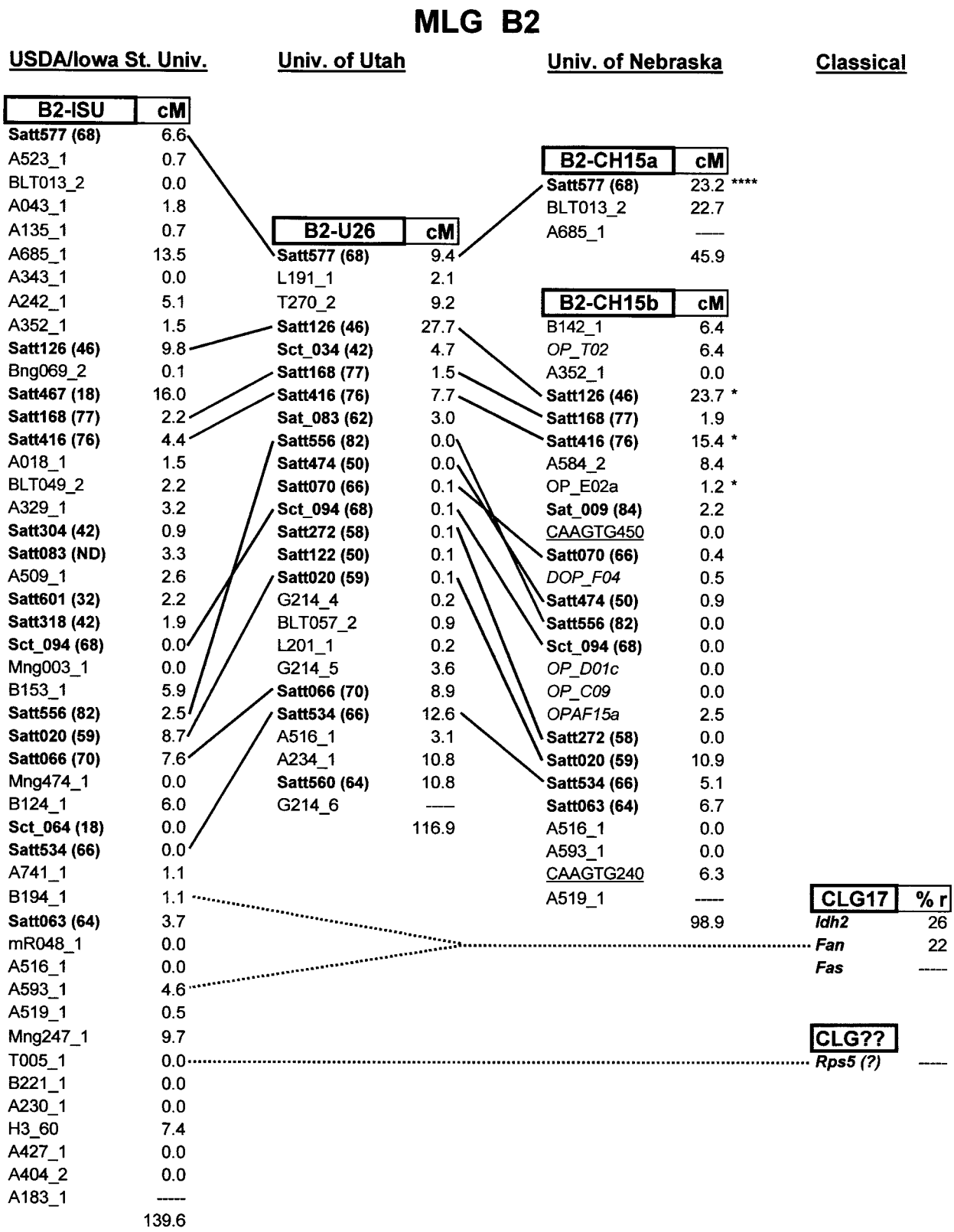

Fig. 1. continued

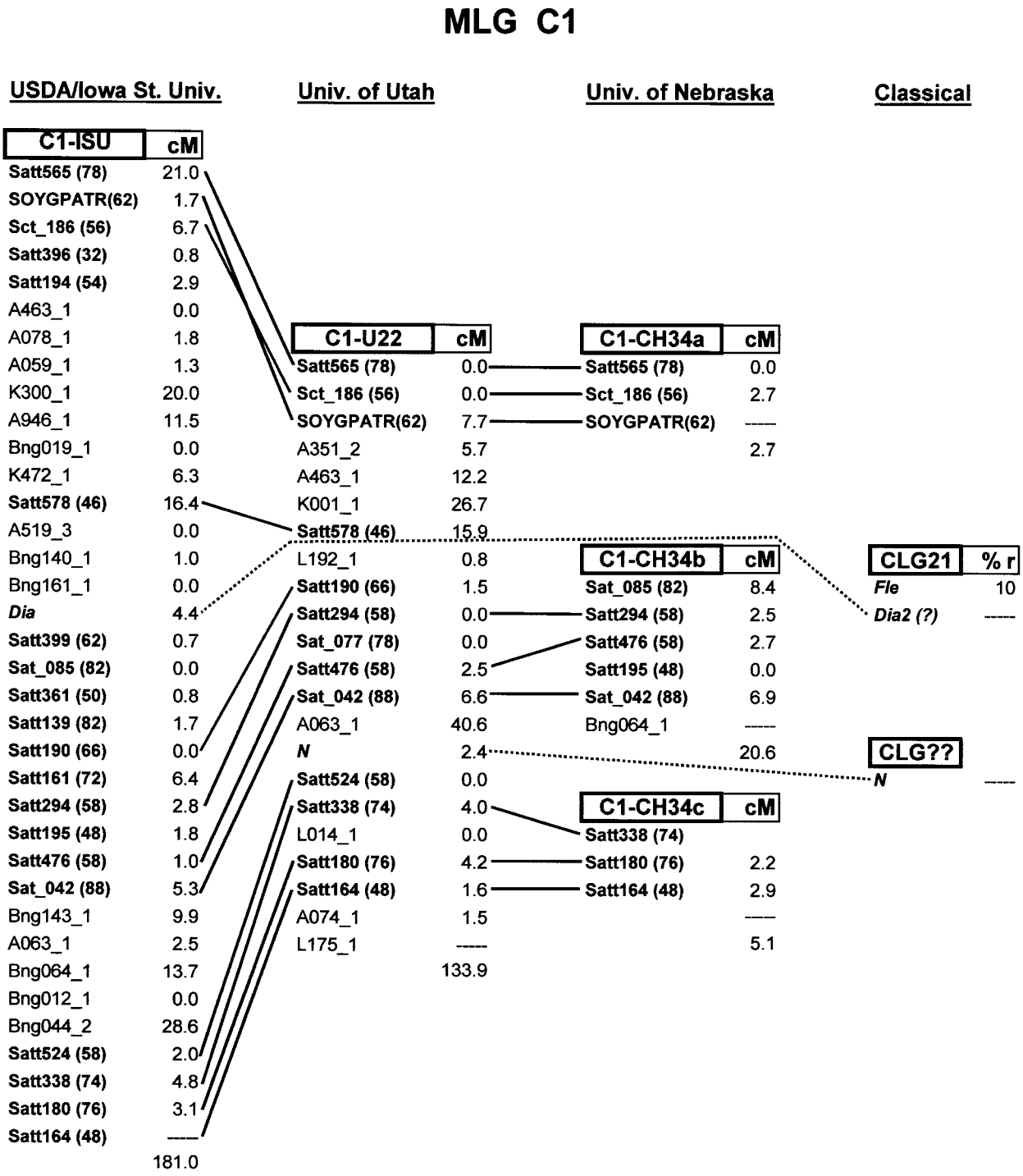

Fig. 1. continued

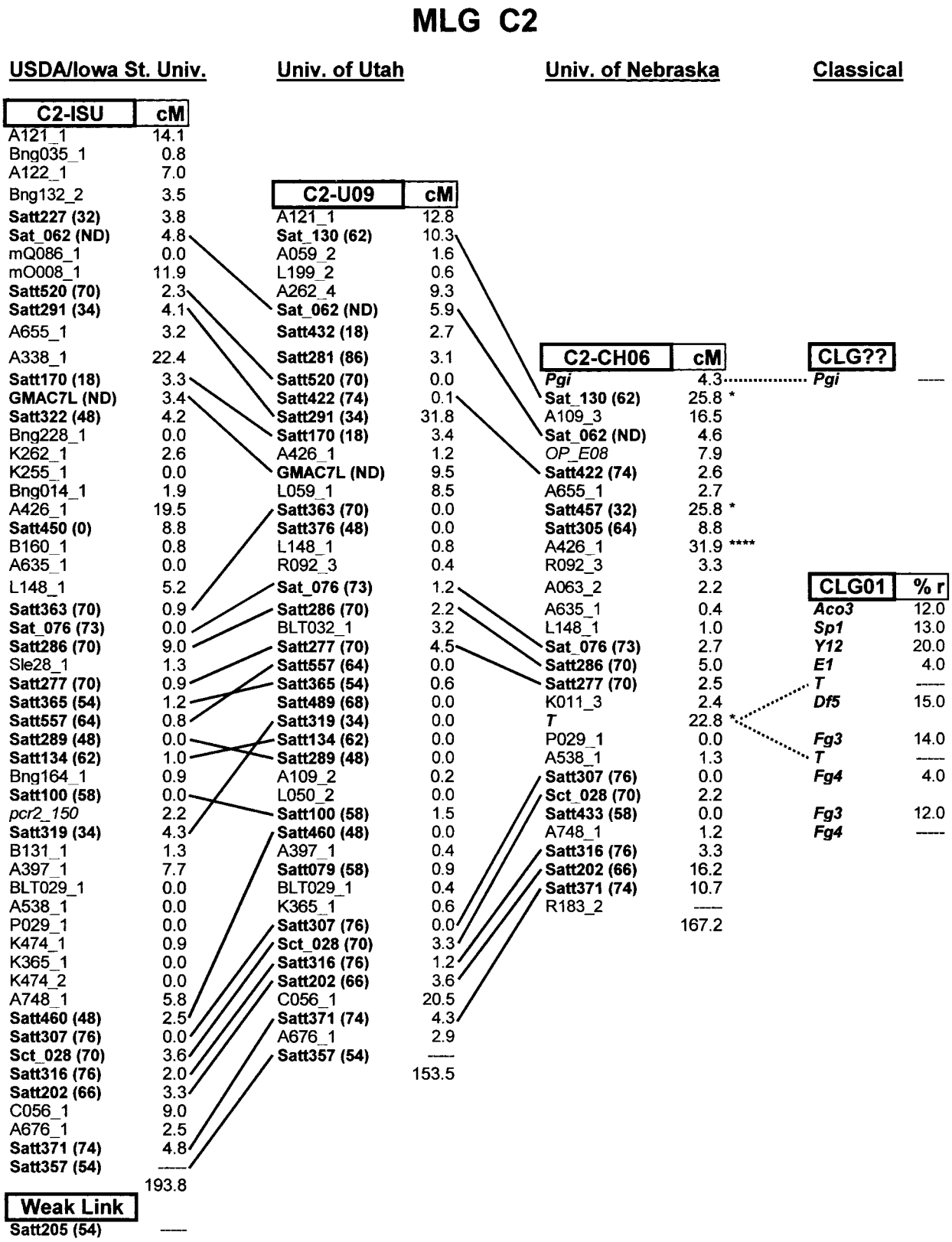

Fig. 1. continued

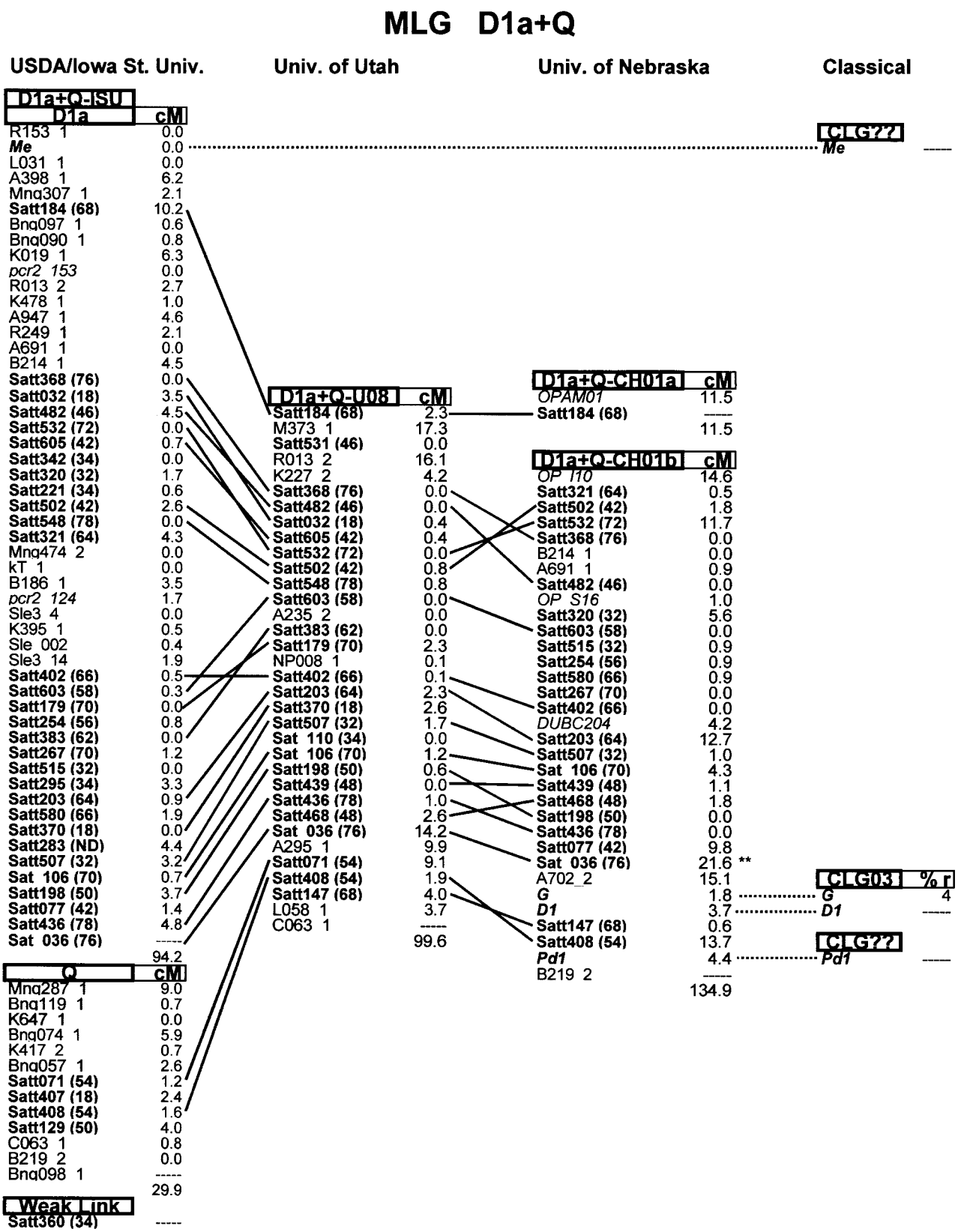

Fig. 1. continued

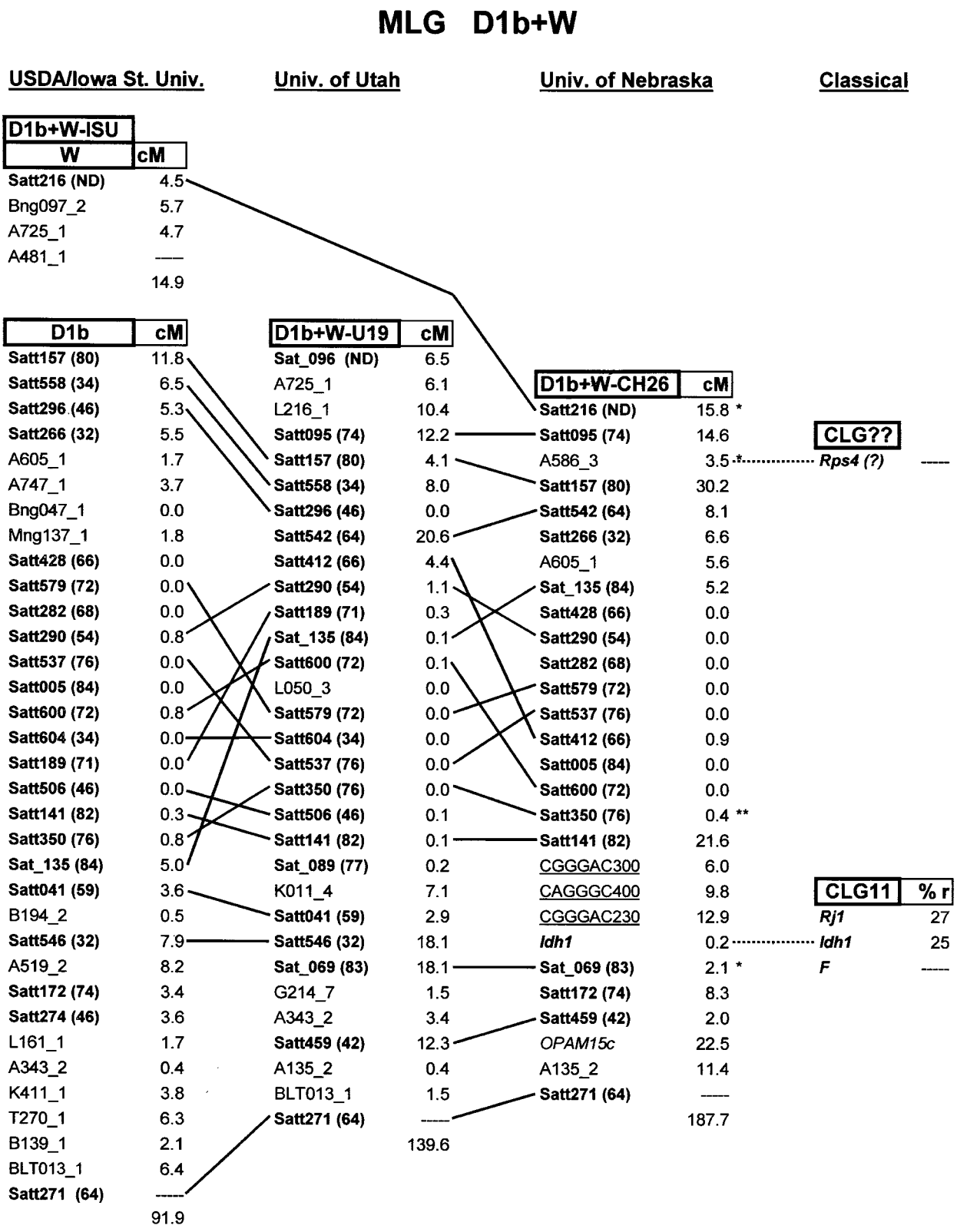

Fig. 1. continued

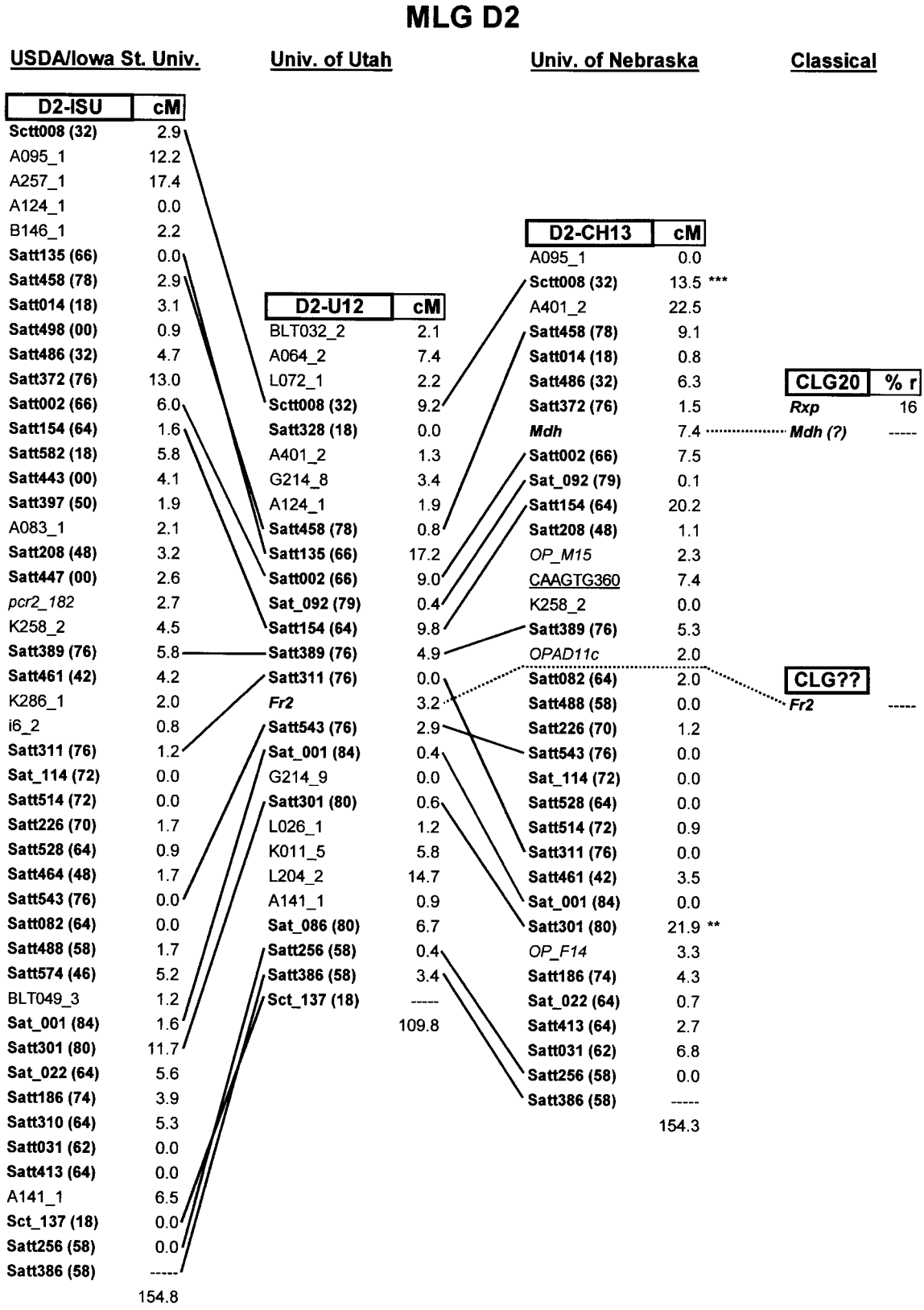

Fig. 1. continued

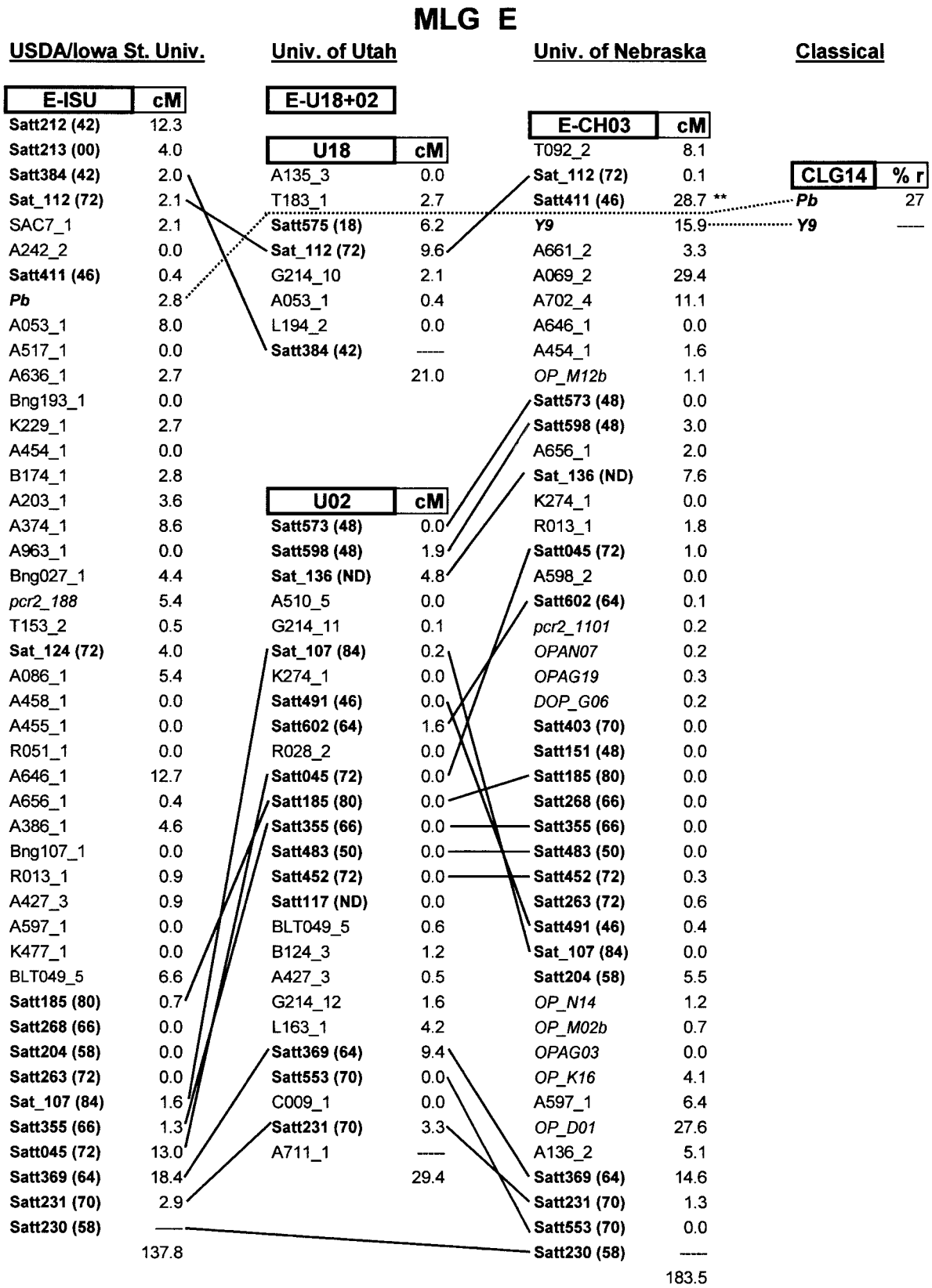

Fig. 1. continued

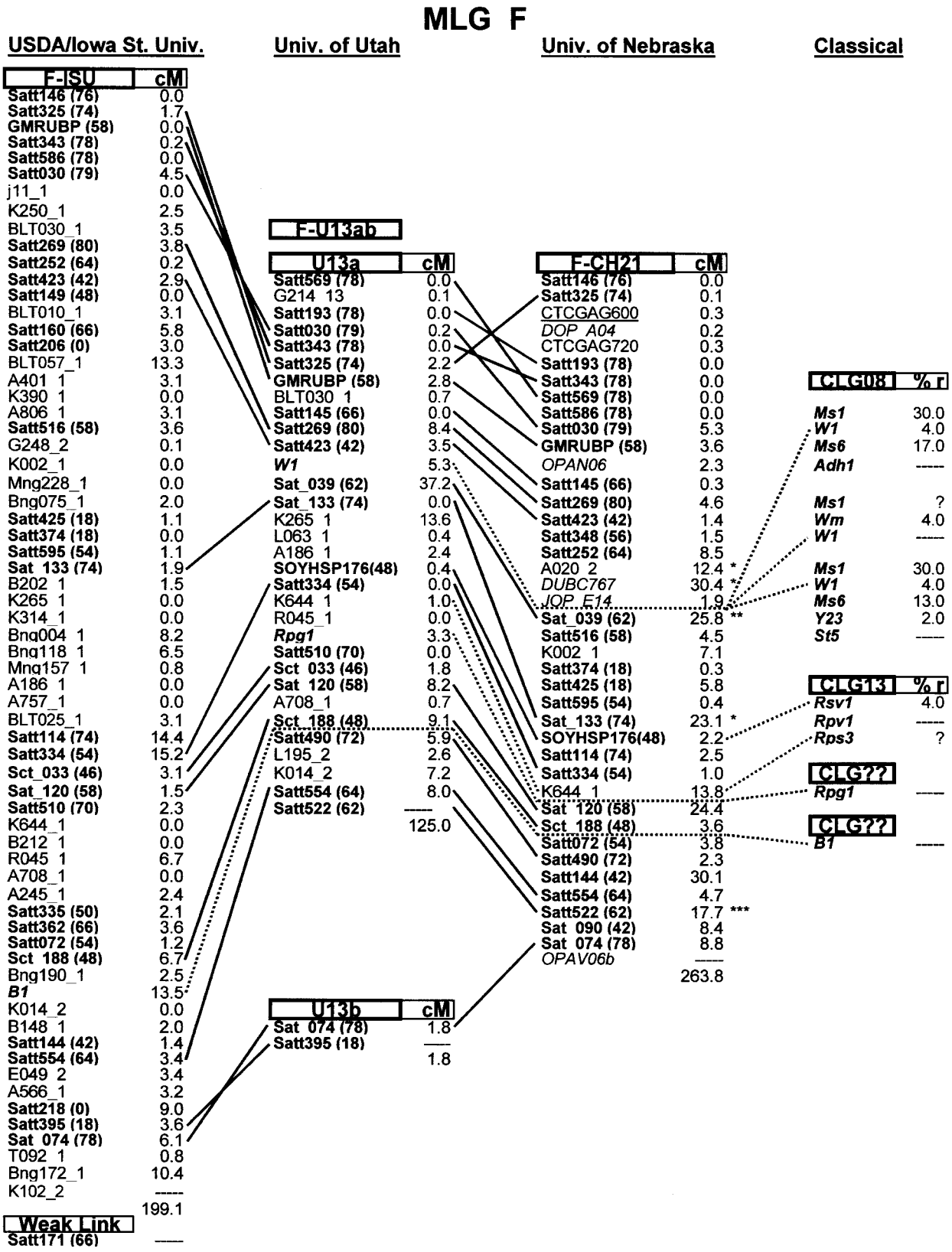

Fig. 1. continued

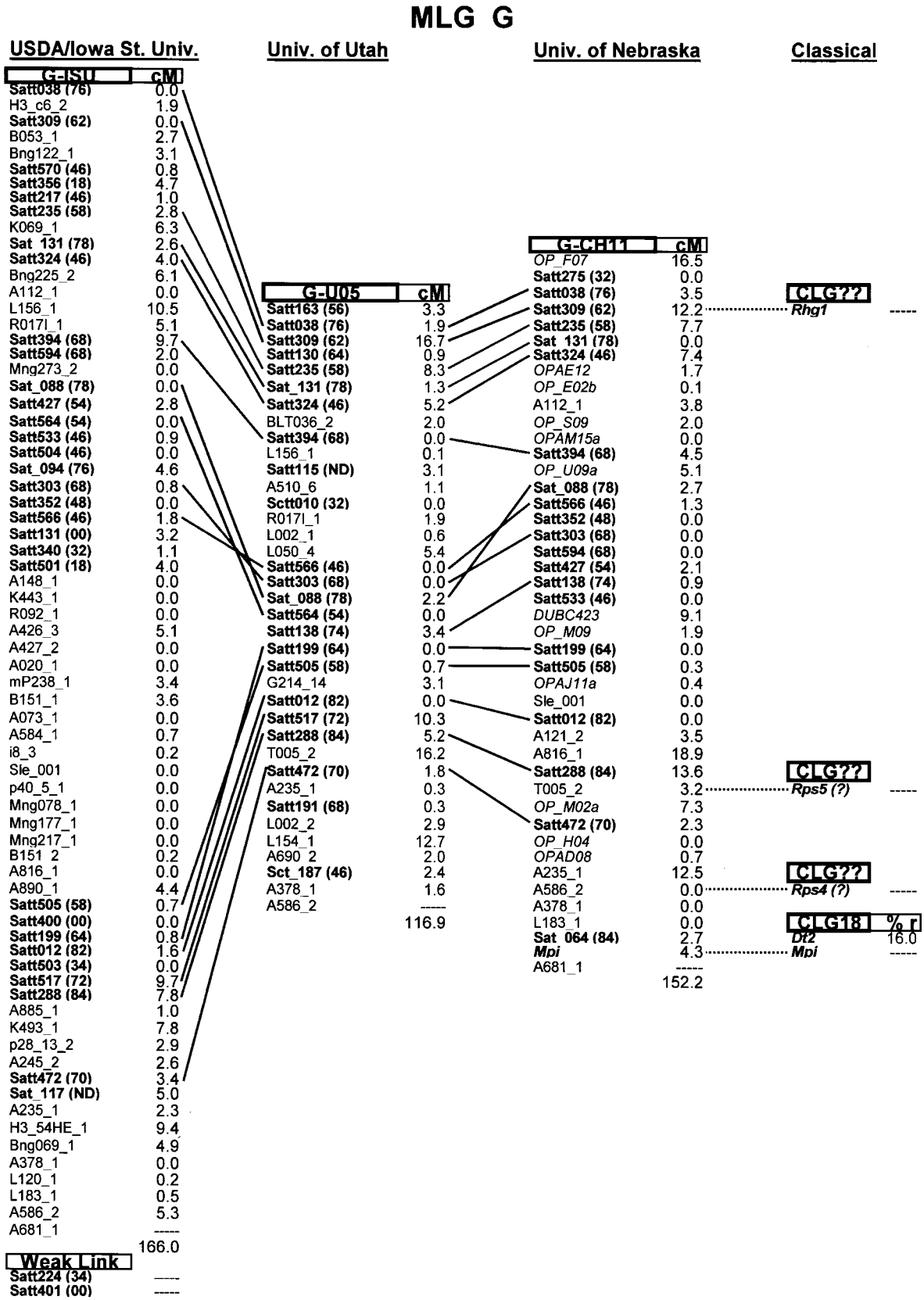

Fig. 1. continued

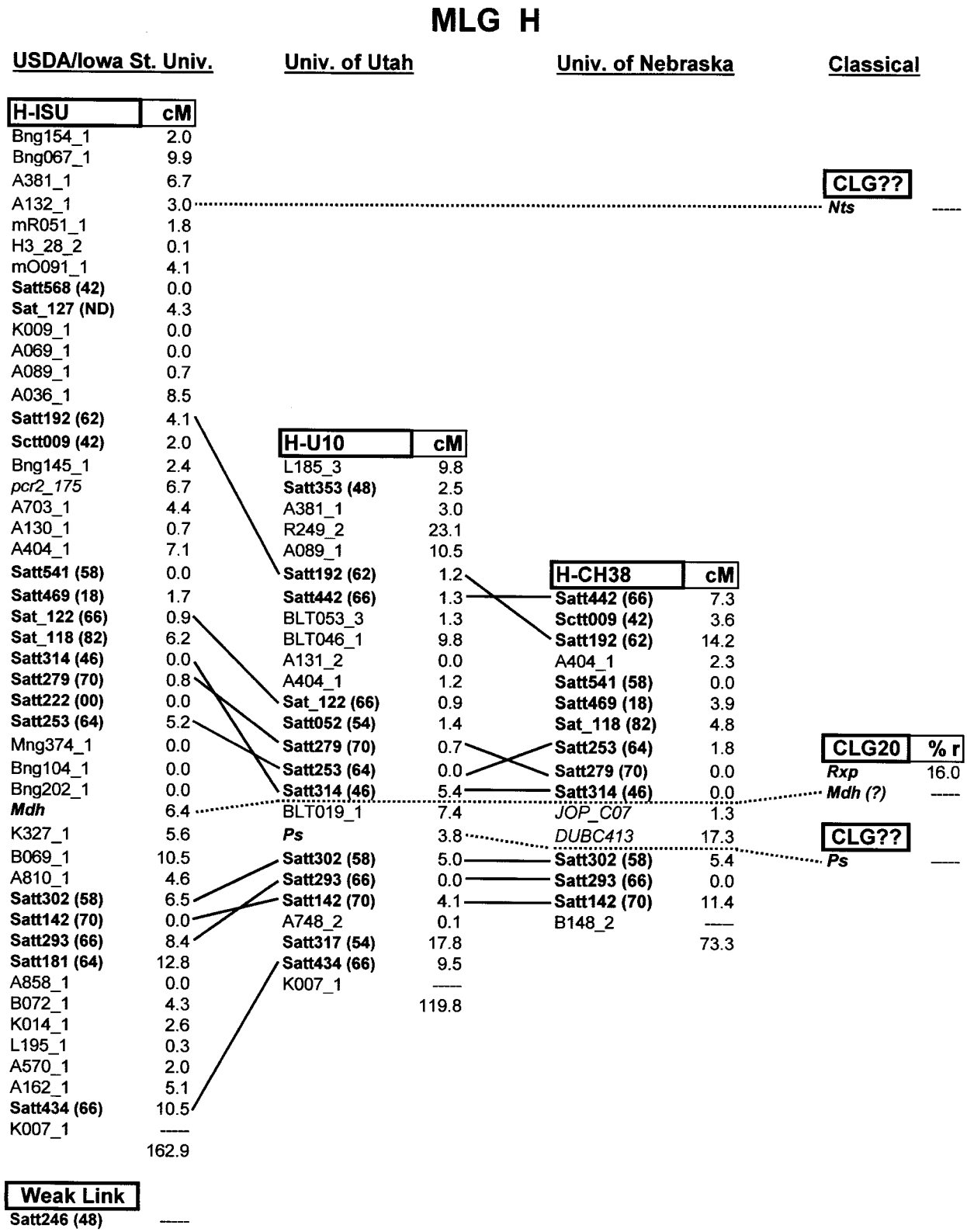

Fig. 1. continued

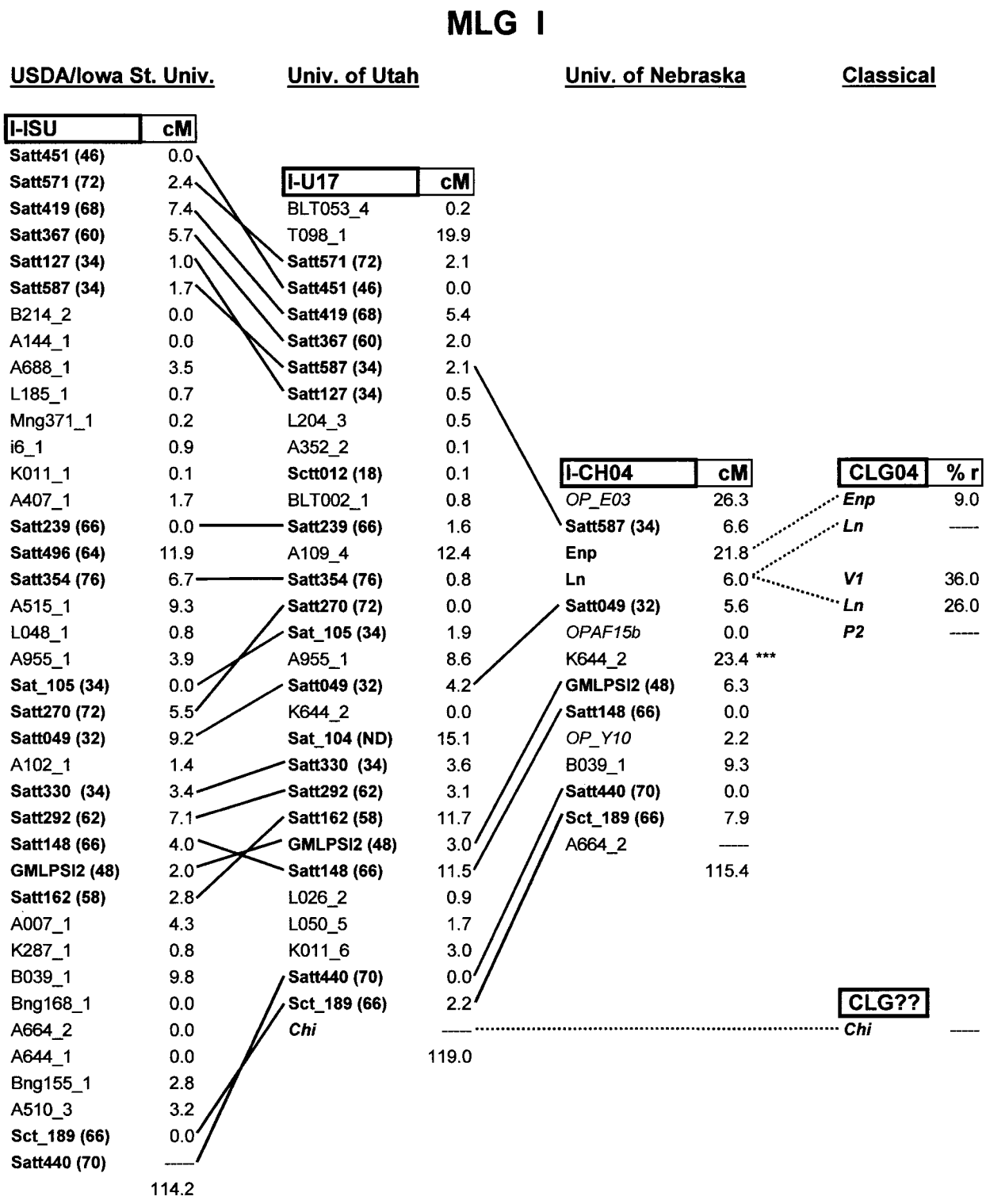

Fig. 1. continued

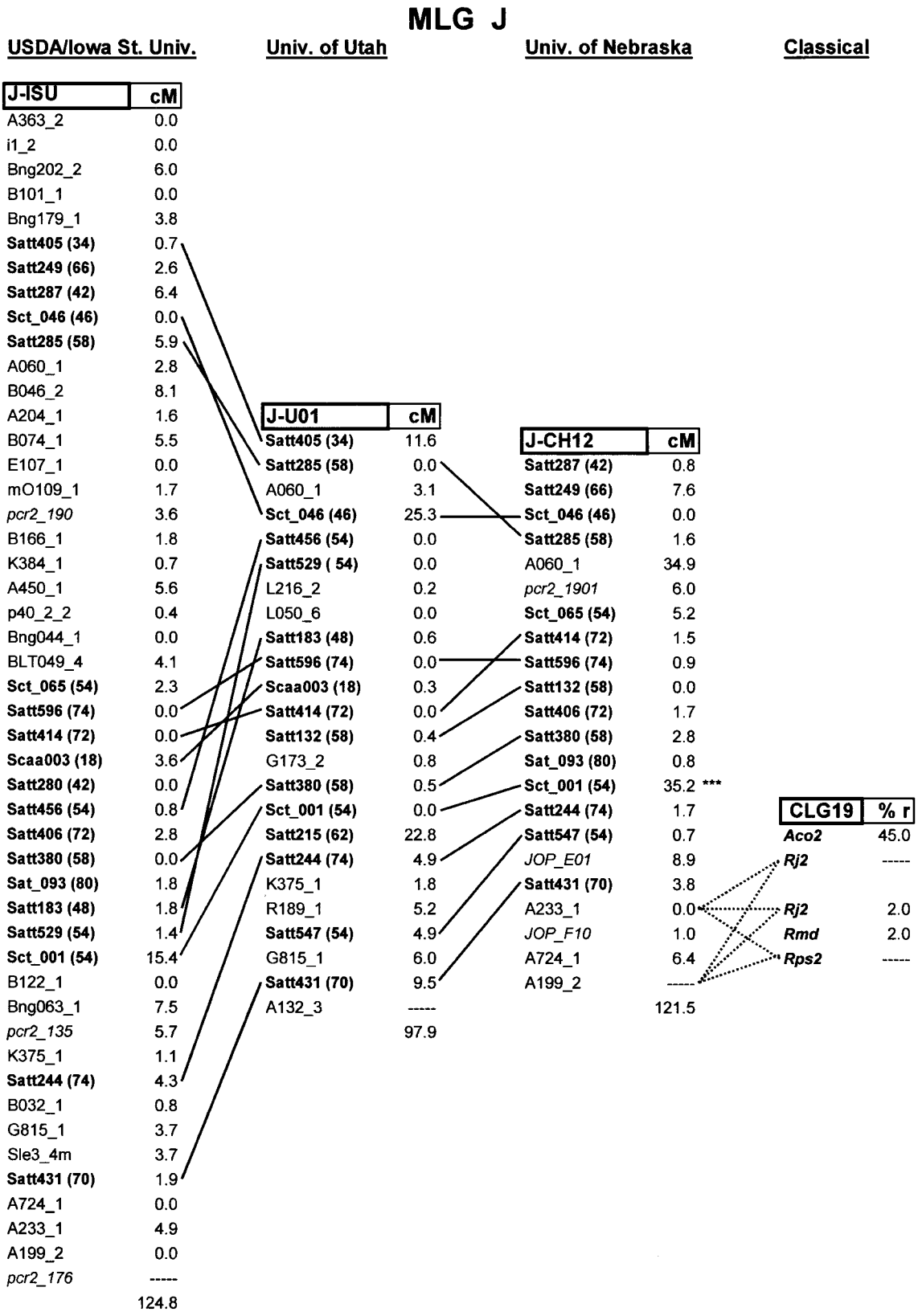

Fig. 1. continued

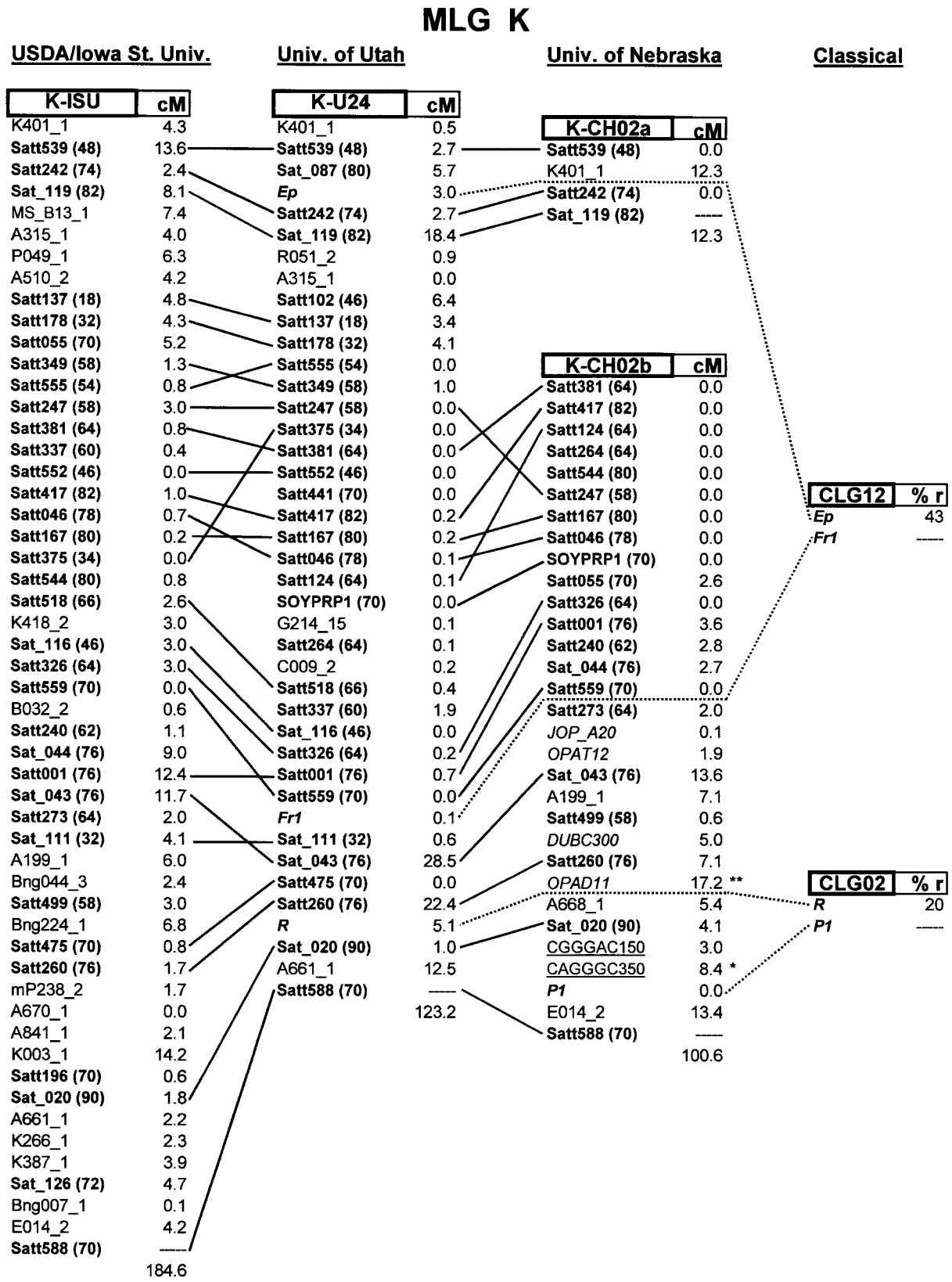

Fig. 1. continued

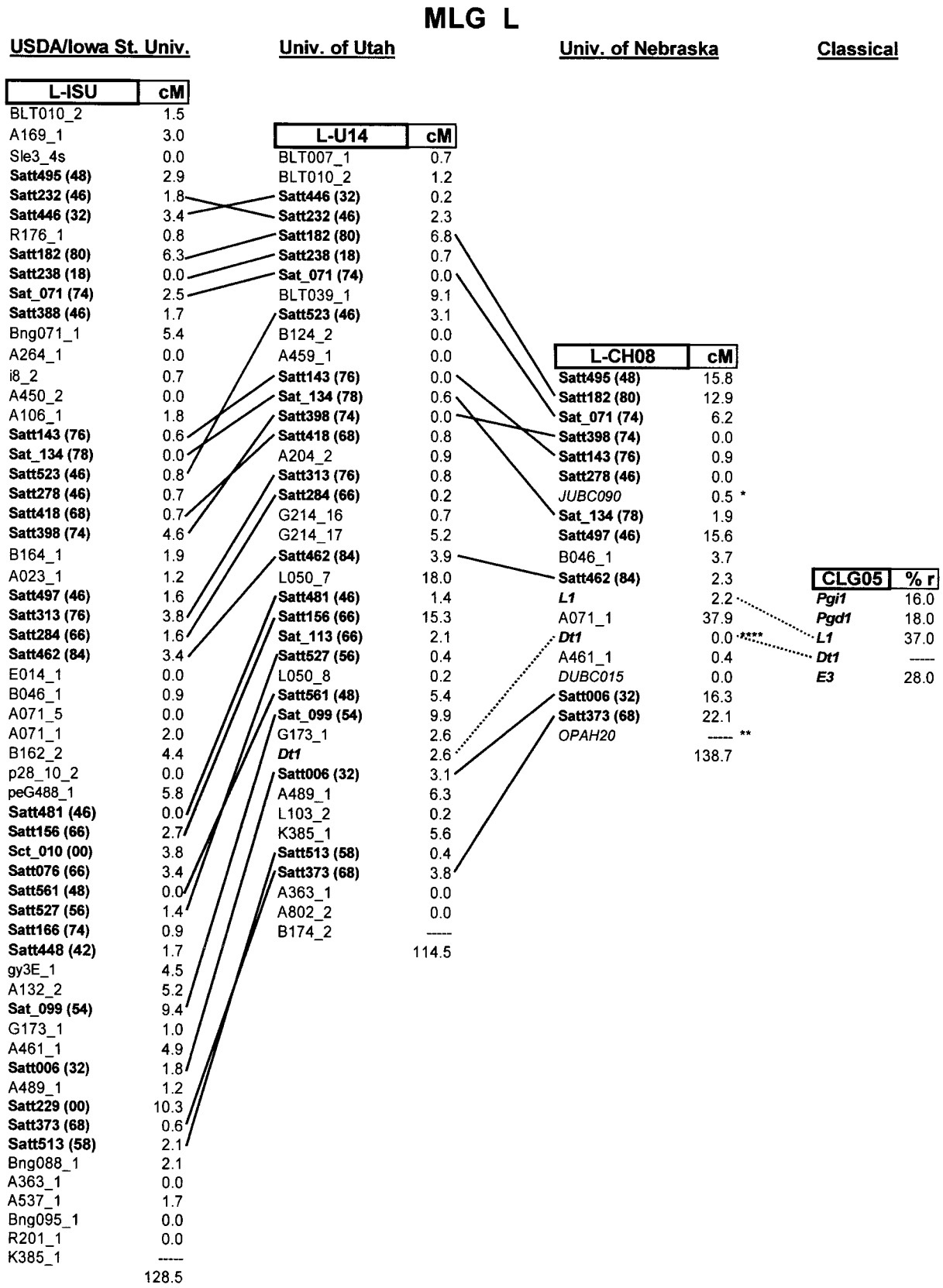

Fig. 1. continued

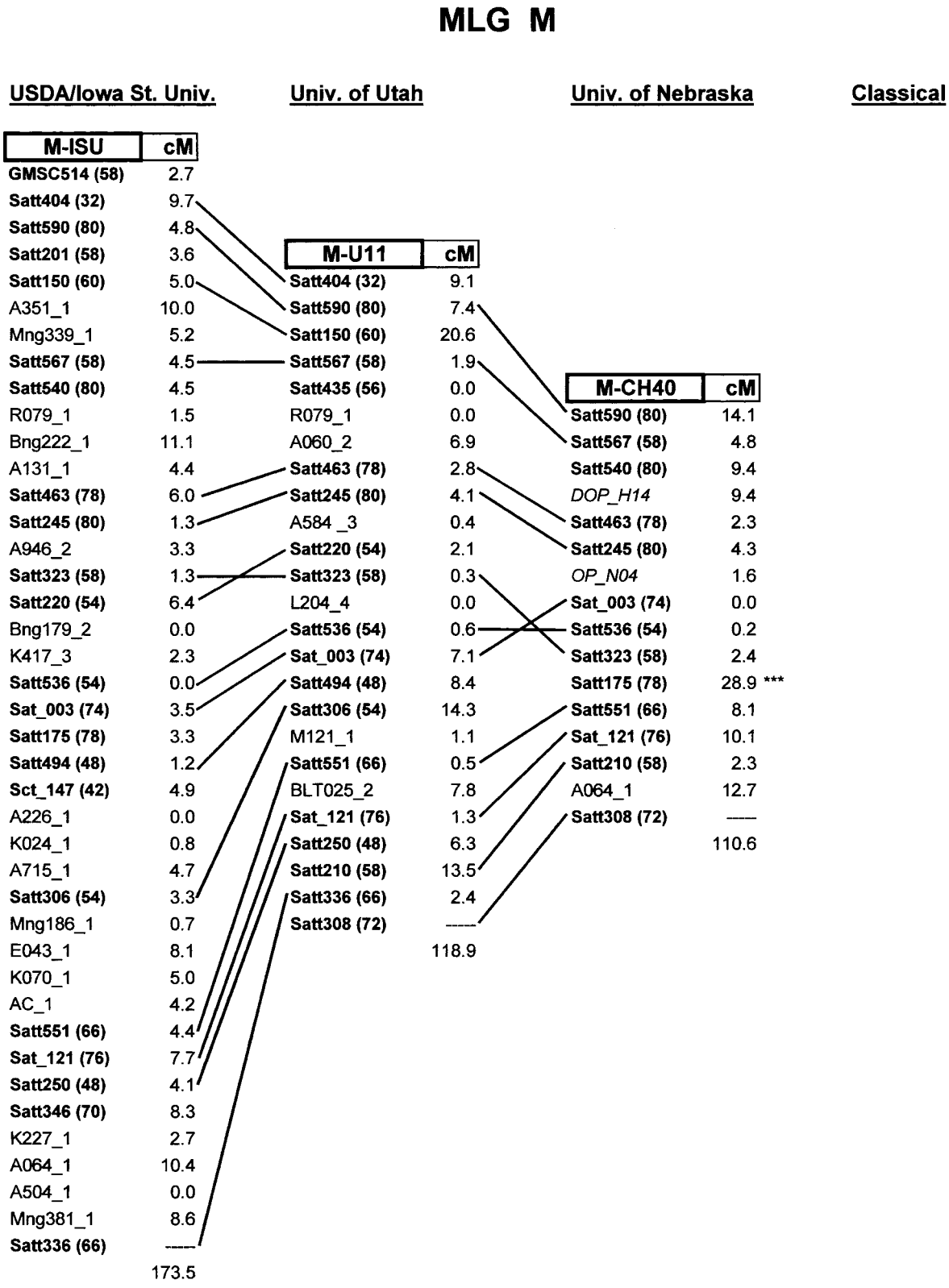

Fig. 1. continued

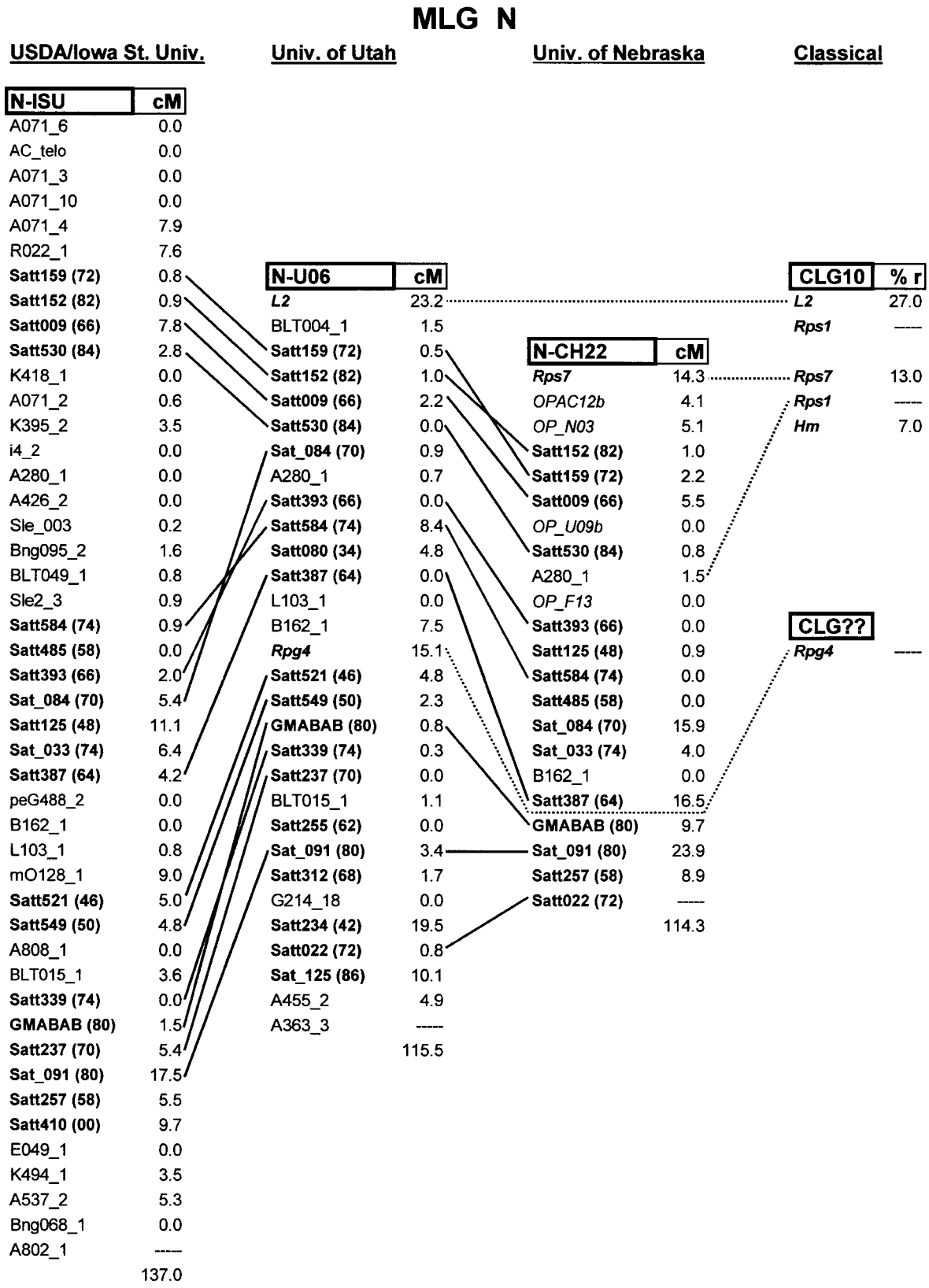

Fig. 1. continued

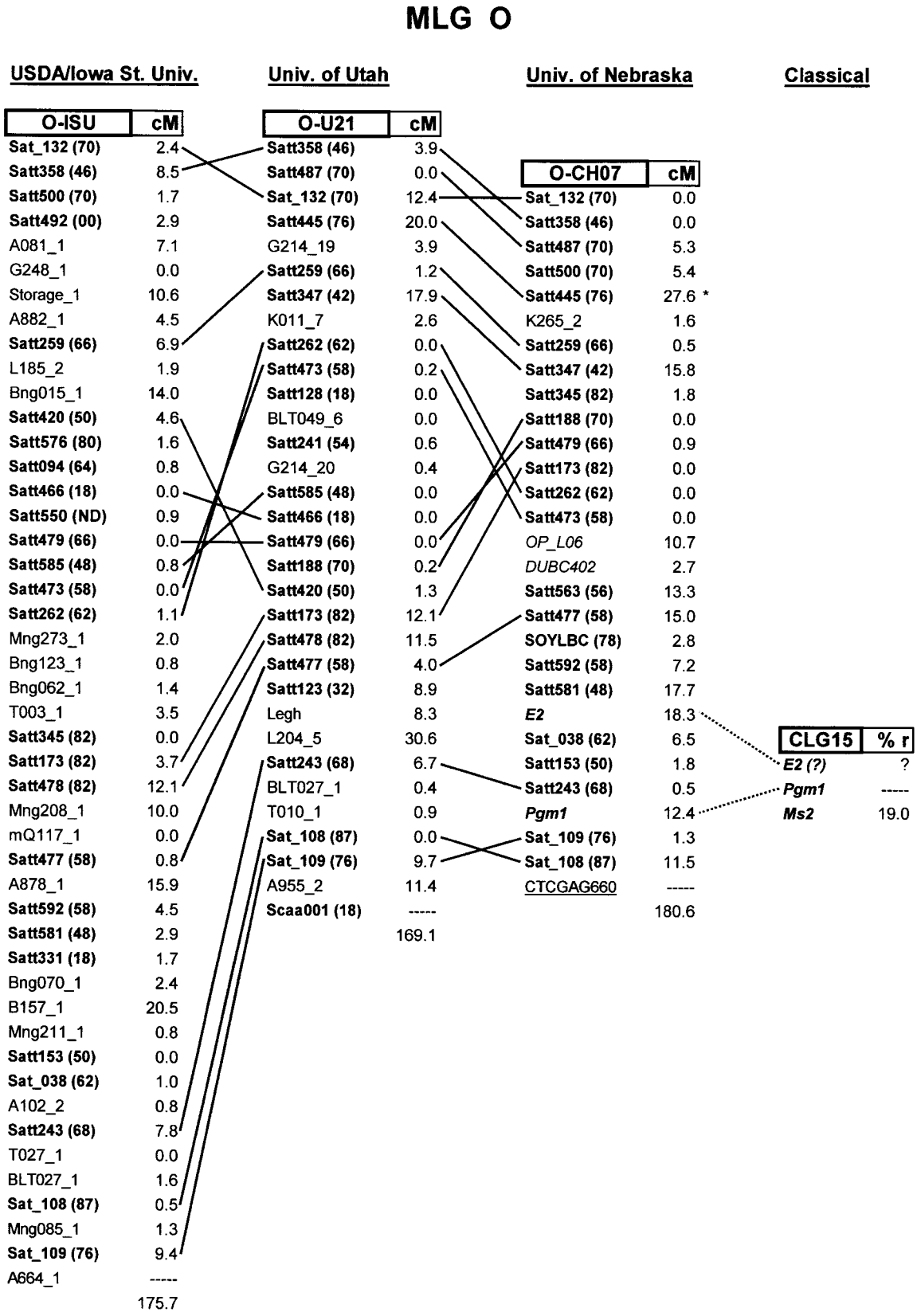

Fig. 1. continued

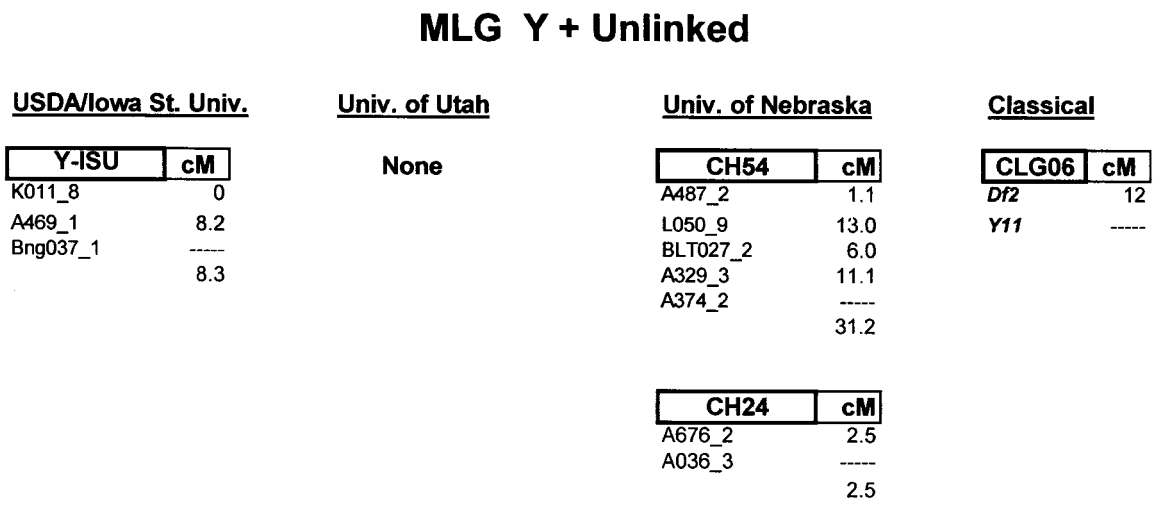

Fig. 1. Genetic maps of 20 consensus soybean linkage groups defined using three mapping populations: The USDA/Iowa State Univ., *G. max* × *G. soja* F_2 population consisting of 59 F_2 plants; the Univ. of Utah, Minsoy × Noir 1 RIL population of 240 lines; and the Univ. of Nebraska, Clark × Harosoy F_2 population consisting of 57 plants; and corresponding classical linkage groups. The first 20 panels correspond to the linkage 20 consensus linkage groups in Table 2. The last panel (Panel Y-ISU+Unlinked) are those linkage groups for which there is as yet no corresponding consensus group. Loci referred to as "Weak Link" do not coalesce with their homologous group at LOD 5.0 but do coalesce at a lower LOD. On the University of Nebraska map, one asterisk indicates the distance is supported by a LOD score of less than 5.0 but greater than 4.0, two asterisks indicates the distance is supported by a LOD score of less than 4.0 but greater than 3.0, three asterisks indicates the distance is supported by a LOD score of less than 3.0 but greater than 2.0, four asterisks indicates the distance is supported by a LOD score of less than 2.0 but greater than 1.5.Centimorgan (cM) distances between adjacent loci are indicated. The number in parentheses following each SSR locus is the gene diversity × 100. An ND in parenthesis following an SSR locus indicates that the gene diversity was not determined. RFLP loci are denoted using the Arial Normal font, SSR loci are in Arial Bold font, RAPD loci are in *Arial Italic font*, AFLP loci are in *Arial Normal font and underlined*, and classical loci are in Arial Bold Italic font.

names to the 20 consensus groups that correspond as closely as possible to the USDA/Iowa St. Univ. names used in the past. In the near future, we anticipate the availability of all 20 primary trisomics of soybean (Xu et al., 1997). Given the availability of a large set of highly polymorphic SSR markers with unambiguous linkage group assignments, it should be a relatively simple task to associate each linkage group with its corresponding chromosome. When these associations are made, linkage groups can be assigned numbers equivalent to those assigned to the 20 soybean chromosomes.

Positioning of SSR Markers within Linkage Groups

In previous work, Akkaya et al. (1995) indicated that the first 40 SSR loci mapped in soybean appeared to

Table 2. The 20 consensus soybean linkage groups based upon alignments of the USDA/Iowa State Univ. *G. max* × *G. soja*, the Univ. of Utah Minsoy × Noir 1, and the Univ. of Nebraska Clark × Harosoy molecular genetic maps and a summary of the corresponding linkage group name(s) that has been used to refer to each in the scientific literature and corresponding classical linkage groups.

	USDA/Iowa State Univ. *G. max* × *G. soja* map					Univ. of Utah Minsoy × Noir 1 map		Classical linkage groups
Consensus linkage groups (Fig. 1)	Current linkage groups (Fig. 1)	Shoemaker and Specht, 1995 and Shoemaker et al., 1996	Shoemaker and Olson, 1993 and Shoemaker, 1994	Diers et al., 1992b	Keim et al., 1990	Current linkage groups (Fig. 1)	Mansur et al., 1996	Palmer and Shoemaker, 1998
A1	A1-ISU	A1	A	B, O	A, Z	A1-U07	U07	
A2	A2-ISU	A2	A	B, W, Z4	A, S	A2-U03	U03	CL07, CL09
B1	B1-ISU	B1, S	B, S	H, Y	I	B1-U04	U04	
B2	B2-ISU	B2, P	B, P	J, M	–	B2-U26	U1c TOP, U26, U24	CL17
C1	C1-ISU	C1	C	E, V	T	C1-U22	U22, U10b, U28	CL21(?)
C2	C2-ISU	C2	C, U	E, Z3	M, V, Y	C2-U09	U09, U20	CL01
D1a+Q	D1a+Q-ISU	D1, Q	D, Q	I, Z1	H	D1a+Q-U08	U08	CL03
D1b+W	D1b+W-ISU	D1	D, W	Z2, Z6	–	D1b+W-U19	U19, U25	CL11
D2	D2-ISU	D2, R	D, R	U, X	N, X	D2-U12	U12, U16	CL20(?)
E	E-ISU	E	E	A	B	E-U18+U02	U18, U2a	CL14
F	F-ISU	F	F, X	C	E, W	F-U13a+U13b	U13, U15	CL08, CL13
G	G-ISU	G	G	D	C	G-U05	U05	CL18
H	H-ISU	H	H	F	F, P	H-U10	U10a, U23	CL20(?)
I	I-ISU	I	I	K	J	I-U17	U2c, U17, U29	CL04
J	J-ISU	J	J	Q, R	L, O	J-U01	U1a	CL19
K	K-ISU	K	K	G	D	K-U24	U1cBOTTOM, U24	CL12, CL2
L	L-ISU	L	L	P, S	R	L-U14	U2b, U14	CL05
M	M-ISU	M	M	L	G	M-U11	U11	
N	N-ISU	N	N	T, Z5	K	N-U06	U06	CL10
O	O-ISU Y-ISU	O	O, V	N	–	0-U21	U1b, U21	CL15

Table 3. Intervals of greater than 20 cM in the Univ. of Utah (Minsoy × Noir 1) genetic linkage map that do not contain simple sequence repeat (SSR) markers, the SSR loci or linkage group end flanking each interval, and the estimated distance between the flanking SSR loci.

Linkage group (Univ. of Utah)	Flanking SSR loci or linkage group end	Distance between flanking SSR loci
		cM
A1-U07	Satt050-Satt385	22.1
A1-U03	Satt424-Sat_115	21.8
B1-U04	Top of linkage group-Satt509	42.2
	Satt197-Satt298	25.3
	Sat_123-Satt453	32.6
B2-U26	Satt577-Satt126	20.7
	Satt126-Sct_034	27.7†
	Satt534-Satt560	26.5
C1-U22	SOYGPATR-Satt578	52.3
	Sat_042-Satt524	49.6
C2-U09	Sat_130-Sat_062	21.8
	Satt291-Satt170	31.8†
	Satt202-Satt371	24.1
D1a+Q-U08	Satt531-Satt368	20.3
	Sat_036-Satt071	24.1
D1b+W-U19	Sat_096-Satt095	23.0
	Satt542-Satt412	20.6†
	Sat_069-Satt459	23.0
D2-U12	Satt301-Sat_086	23.2
E-U18+U02	Satt384-Satt598	50†
F-U13ab	Satt522-Sat_074	50†
G-U05	Satt288-Satt472	21.4
H-U10	Satt353-Satt192	39.1
I-U17	Top of linkage group-Satt571	20.1
J-U01	Sct_046-Satt456	25.3†
	Satt215-Satt244	22.8†
K-U24	Sat_043-Satt475	28.5†
	Satt260-Sat_020	27.5†
L-U14	Satt462-Satt481	21.9
M-U11	Satt150-Satt567	20.6†
N-U06	Top of linkage group-Satt159	24.7
	Satt387-Satt521	22.6
O-U21	Satt445-Satt259	23.9
	Satt347-Satt262	20.5
	Satt123-Satt243	47.8
	Sat_109-Scaa001	21.1

† No RFLP locus is present in the interval defined by the two flanking SSR loci.

distribute randomly throughout the genome. However, these authors also observed one cluster of five SSR loci that would have a low probability of occurring by random chance if SSR loci were truly distributed randomly in the soybean genome. With the mapping of as many as 500 SSR loci in a single mapping population, it is still not clear if there is substantial clustering of SSRs. Portions of many linkage groups contain groups of SSR loci and similar groups of RFLP loci. Linkage group D1a (Fig. 1, Panel MLG D1a+Q) demonstrates this clustering especially in the case of the *G. max* × *G. soja* map. At the top of the linkage group is a group of 16 markers, 13 of which are RFLPs. Adjacent to these are groups of uninterrupted stretches of 11 and 18 SSR markers separated by a group of eight loci, seven of which are RFLPs. Linkage group E (Fig. 1, Panel MLG E) in the *G. max* × *G. soja* map has a very long stretch of mostly RFLP loci in which there is only one SSR (Sat_124). Flanking this region are two clusters of markers that are mostly SSRs. A similar cluster of RFLP loci occurs in linkage group G (Fig. 1, Panel MLG G) of the *G. max* × *G. soja* map between SSR loci Satt501 and Satt505. While this type of clustering does not occur in every linkage group, it is a fairly frequent occurrence.

As indicated above, it is not clear from observing the maps presented in Fig. 1 that clustering of markers is the result of the clustering of RFLP loci or of SSR loci or both. In the development of libraries from which SSR-containing genomic clones were selected, many different restriction enzymes and combinations of restriction enzymes were used to create genomic fragments in the 500 to 700 bp range (Cregan et al., 1994; Cregan et al., 1999). This was done to avoid duplicate clones and to sample different portions of the soybean genome. In contrast, one of the techniques in the development of the RFLP probes used here was the development and isolation of probes from *Pst*I libraries (Keim and Shoemaker, 1988). The rationale for this approach is that methylation-sensitive enzymes such as *Pst*I preferentially cut non-methylated regions which are presumed to contain less repetitive DNA. However, this approach may have produced sets of restriction fragments that were not randomly distributed throughout the genome. Thus, it is possible that clustering of markers we have noted here is more closely associated with RFLP than with SSR loci.

Regardless of the reason for the possible clustering of loci, the application of the set of SSR loci described here will be detrimentally affected by large intervals or gaps in which no markers are present. For example, it is obvious that quantitative trait loci (QTL) in genomic regions lacking markers will be undetectable. A total of 36 intervals of greater than 20 cM that do not contain an SSR locus are present in the Minsoy × Noir 1 map (Table 3). Each of the 20 consensus linkage groups contains at least one such gap suggesting that such intervals are not confined to a specific subset of chromosomes. In many instances such as those in linkage group E-U18+02 and G-U05, clusters of RFLP loci are present in the intervals lacking SSRs. In an effort to place SSR loci in regions with only RFLP markers, we are attempting to use bacterial artificial chromosome (BAC) clones for targeted SSR marker development as described by Cregan et al. (1999). These BAC clones are being selected via hybridization with the RFLP probes that map to regions devoid of SSR loci. However, in 10 of the 36 intervals of greater than 20 cM listed in Table 3, no RFLP loci are present in the gaps flanked by the SSR loci. It may be difficult to develop markers that map to these intervals.

Informativeness of SSR Loci

Each SSR locus shown in Fig. 1 is followed by a two-digit value (in parentheses) which is the gene diversity of the locus multiplied by 100. Gene diversity was determined on a group of 10 *G. max* genotypes representing a range of diversity within the cultivated species. Adapted cultivars such as Williams, Clark, Amsoy, Harosoy, Jackson, and Archer are included along with the more exotic genotypes Fiskeby V, Minsoy, Noir 1, and Tokyo. The gene diversity scores are provided to allow the soybean breeder or geneticist to choose markers with the greatest probability of detecting polymorphism. For example, for QTL discovery research, one can use Fig.

1 to select a set of maximally informative loci distributed throughout the 20 soybean linkage groups in order to obtain a maximum coverage of the genome. In general, loci with AT core motifs (designated Sat_xxx) are most informative, followed by those with ATT cores (designated Sattxxx). A small number of SSRs with CT or CTT cores are also included. These generally have low gene diversity values.

The Integration of Classical Genetic Loci into the Soybean Map

Palmer and Shoemaker (1998) defined 20 classical linkage groups (CLG) that contain a total of 68 of the more than 250 classical pigmentation, morphological, isozyme, disease resistance, etc. genes that have been named. The inclusion of the Univ. of Nebraska population derived from the cross of NILs of Clark × Harosoy in this study is important because it segregated for 14 classical loci (Shoemaker and Specht, 1995). These classical loci and other classical loci linked to them are now integrated with molecular markers to form a comprehensive genetic map. Some of the linkage data were reported earlier (Shoemaker and Specht, 1995), but it is reproduced here for completeness. As shown in Fig. 1 and listed in Table 2, all but one CLG (CLG06, Fig. 1, Panel MLG Y+Unlinked) can now be associated with a molecular linkage group.

The integration of the classical, RFLP, and SSR markers into one comprehensive linkage map is a powerful tool for the advancement of soybean genetics. For those studies that have been based upon RFLP marker technology it provides an opportunity to convert to a PCR-based marker system. Because SSR markers are so commonly used by human geneticists, technologies for their use are readily available. For example, automated allele sizing is being applied to plant genetic studies (Kresovich et al., 1995; Diwan and Cregan, 1997). In the future, high throughput systems will provide the opportunity for genetic analysis that is appropriate to the needs of large scale plant improvement programs that routinely analyze thousands or hundreds of thousands of segregating progeny.

ACKNOWLEDGMENTS

The authors wish to thank Mr. Edward Fickus, Mr. Charles Quigley, Mr. Patrick Elia, Ms. Sarah Hyatt, Ms. Carol Kosman, Mr. Jason Kenworthy (Beltsville, MD), and Mr. Mike Livingston (Univ. of Nebraska) for their excellent technical assistance during the course of the work reported here. This research was supported in part by funds from the USDA, Cooperative State Research Service, National Research Initiative, Competitive Grants Program, Grant No. 9300874 to PBC, JES, and RCS for which the authors are very grateful. In addition, the authors wish to extend their thanks to the United Soybean Board (USB Grant #6016) for support of the development and mapping of soybean simple sequence repeat markers.

REFERENCES

Akkaya, M.S., A.A. Bhagwat, and P.B. Cregan. 1992. Length polymorphism of simple sequence repeat DNA in soybean. Genetics 132: 1131–1139.

Akkaya, M.S., R.C. Shoemaker, J.E. Specht, A.A. Bhagwat, and P.B. Cregan. 1995. Integration of simple sequence repeat DNA markers into a soybean linkage map. Crop Sci. 35:1439–1445.

Anderson, J.A., G.A. Churchill, J.E. Autrique, S.D. Tanksley, and M.E. Sorrells. 1993. Optimizing parental selection for genetic linkage maps. Genome 36:181–186.

Apuya, N.R., B.L. Frazier, P. Keim, E.J. Roth, and K.G. Lark. 1988. Restriction fragment length polymorphisms as genetic markers in soybean, *Glycine max* (L.) Merrill. Theor. Appl. Genet. 75:889–901.

Archibald, A.L., C.S. Haley, J.F. Brown, S. Couperwhite, H.A. McQueen, D. Nicholson, W. Coppieters, A. Van de Weghe, A. Stratil, A-K. Wintero, M. Fredholm, N.J. Larson, V.H. Nielsen, D. Milan, N. Woloszyn, A. Robic, M. Dalens, J. Riquet, J. Gellin, J-C. Caritez, G. Burgaud, L. Ollivier, J-P. Bidanel, M. Vaiman, C. Renard, H. Geldermann, R. Davoli, D. Ruyter, E.J.M. Verstege, M.A.M. Groenen, W. Davies, B. Hoyheim, A. Keiserud, L. Andersson, H. Ellegren, M. Johansson, L. Marklund, J.R. Miller, D.V. Anderson Dear, E. Signer, A.J. Jeffreys, C. Moran, P. Le Tissier, M.F. Muladno Rothschild, C.K. Tuggle, D. Vaske, J. Helm, H-C. Liu, A. Rahman, T.P. Yu, R.G. Larson, and C.B. Schmitz. 1995. The PiGMap Consortium linkage map of the pig (*Sus scrofa*). Mamm. Genome 6:157–175.

Bell, C.J., and J.R. Ecker. 1994. Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics 19:137–144.

Bryan, G.J., A.J. Collins, P. Stephenson, A. Orry, J.B. Smith, and M.D. Gale. 1997. Isolation and characterisation of microsatellites from hexaploid bread wheat. Theor. Appl. Genet. 94:557–563.

Cregan, P.B., A.A. Bhagwat, M.S. Akkaya, and Jiang Rongwen. 1994. Microsatellite fingerprinting and mapping of soybean. Meth. Mol. Cell. Biol. 5:49–61.

Cregan, P.B., and C.V. Quigley. 1997. Simple sequence repeat DNA marker analysis. p. 173–185. *In* G. Caetano-Anolles and P.M. Gresshoff (ed.) DNA markers: Protocols, applications and overviews. John Wiley & Sons, New York.

Cregan, P.B., J. Mudge, E.W. Fickus, L.F. Marek, D. Danesh, R. Denny, R.C. Shoemaker, B.F. Matthews, T. Jarvik, and N.D. Young. 1999. Targeted isolation of simple sequence repeat markers through the use of bacterial artificial chromosomes. Theor. Appl. Gent. (In press).

Diers, B.W., W. Fehr. P. Keim, and R.C. Shoemaker. 1992a. RFLP analysis of soybean seed protein and oil content. Theor. Appl. Gent. 83:608–612.

Diers, B.W., P. Keim, W.R. Fehr, and R.C. Shoemaker. 1992b. RFLP analysis of soybean seed protein and oil content. Theor. Appl. Genet. 83:608–612.

Diers, B.W., L. Mansur, J. Imsande, and R.C. Shoemaker. 1992c. Mapping of *Phytophthora* resistance loci in soybean with restriction fragment length polymorphism markers. Crop Sci. 32:377–383.

Dietrich, W.F., J.C. Miller, R.G. Steen, M. Merchant, D. Damron, R. Nahf, A. Gross, D.C. Joyce, M. Wessel, R.D. Dredge, A. Marquis, L.D. Stein, N. Goodman, D.C. Page, and E.S. Lander. 1994. A genetic map of the mouse with 4006 simple sequence length polymorphism. Nature Genet. 7:220–225.

Diwan, N., and P.B. Cregan. 1997. Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. Theor. Appl. Genet. 95:723–733.

Hymowitz, T., and R.J. Singh. 1987. Taxomony and speciation. p. 23–48. *In* R.J. Wilcox (ed.) Soybeans: Improvement, production, and uses. 2nd ed. Agron. Monogr. 16. ASA, CSSA, SSSA, Madison, WI.

Hudson, T.J., L.D. Stein, S.S. Gerety, J. Ma, A.B. Castle, J. Silva, D.K. Slonim, R. Baptista, L. Kruglyak, S.H. Xu, X. Hu, A.M.E. Colbert, C. Rosenberg, M.P. Reeve-Daly, S. Rozen, L. Hui, X. Wu, C. Vestergaard, K.M. Wilson, J.S. Bae, S. Marita, S. Ganiatsas, C.A. Evans, M.M. DeAngelis, K.A. Ingalls, R.W. Nahf, L.T. Horton, M.O. Anderson, A.J. Collymore, W. Ye, V. Kouyoumjian, I.S. Zemsteva, J. Tam, R. Devine, D.F Courtney, M.T. Renaud, H. Nguyen, T.J. O'Connor, C. Fizames, S. Faure, G. Gyapay, C. Dib, J. Morissette, J.B. Orlin, B.W. Birren, N. Goodman, J. Weissenbach, T.L. Hawkins, S. Foote, D.C. Page, and E.S. Lander. 1995. An STS-based map of the human genome. Science 270:1945–1954.

Keim, P., W. Beavis, J. Schupp, and R. Freestone. 1992. Evaluation of soybean RFLP marker diversity in adapted germplasm. Theor. Appl. Genet. 85:205–212.

Keim, P., B.W. Diers, T.C. Olson, and R.C. Shoemaker. 1990. RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics 126:735–742.

Keim, P., T. Olson, and R. Shoemaker. 1988. A rapid protocol for isolating soybean DNA. Soybean Genet. Newslet. 15:150–152.

Keim, P., and R.C. Shoemaker. 1988. Construction of a random recombinant DNA library that is primarily single copy sequence. Soybean Genet. Newslet. 15:147–148.

Keim, P., R.C. Shoemaker, and R.G. Palmer. 1989. Restriction fragment length polymorphism diversity in soybean. Theor. Appl. Genet. 77:786–792.

Kresovich, S., A.K. Szewc-McFadden, and S.M. Bliek. 1995. Abundance and characterization of simple-sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* L. (Rapeseed). Theor. Appl. Genet. 91:206–211.

Lander, E.S., P. Green, J. Abrahamson, A. Barlow, M.J. Daly, S.E. Lincoln, and L. Newburg. 1987. Mapmaker: an interactive computer package for constructing genetic linkage maps of experimental and natural populations. Genomics 1:174–181.

Lark, K.G., J.M. Weisemann, B.F. Matthews, R. Palmer, K. Chase, and T. Macalma. 1993. A genetic map of soybean (*Glycine max* L.) using an intraspecific cross of two cultivars: 'Minsoy' and 'Noir 1'. Theor. Appl. Genet. 86:901–906.

Lincoln, S.E., and S L. Lander. 1993. Mapmaker/exp 3.0 and Mapmaker/QTL 1.1 Whitehead Inst. of Med. Res. Tech Report. Cambridge, MA.

Mansur, L.M., J.H. Orf, K. Chase, T. Jarvik, P.B. Cregan, and K.G. Lark. 1996. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Sci. 36:1327–1336.

Maughan, P.J., M.A. Saghai Maroof, and G.R. Buss. 1995. Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. Genome 38:715–723.

Morgante, M., and A.M. Olivieri. 1993. PCR-amplified microsatellites as markers in plant genetics. Plant J. 3:175–182.

Muehlbauer, G.J., P.E. Staswick, J.E. Specht, G.L. Graef, R.C. Shoemaker, and P. Keim. 1991. RFLP mapping using near-isogenic lines in the soybean [*Glycine max* (L.) Merr.]. Theor. Appl. Genet. 81: 189–198.

Palmer, R.G., and R.C. Shoemaker. 1998. Soybean genetics. p. 45–82. *In* M. Hrustic et al. (ed.) Soybean Institute of Field and Vegetable Crops. Novi Sad, Yugoslavia.

Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalski. 1996. The comparison of RFLP, RAPD, AFLP, and SSR (microsatellite) markers for germplasm analysis. Mol. Breeding 2:225–238.

Röder, M.S. J. Plashke, S.U. König, A. Börner, M.E. Sorrells, S.D. Tanksley and M.W. Ganal. 1995. Abundance, variability and chromosomal location of microsatellites in wheat. Mol. Gen. Genet. 246:327–333.

Rongwen, Jiang, M.S. Akkaya, A.A. Bhagwat, U. Lavi, and P.B. Cregan. 1995. The use of microsatellite DNA markers for soybean genotype identification. Theor. Appl. Genet. 90:43–48.

Senior, M.L., E.C.L. Chin, M. Lee, C.W. Stuber, and J.S.C. Smith. 1996. Simple sequence repeat markers developed from maize sequences found in the Genbank database: map construction. Crop Sci. 36:1676–1683.

Sharon, D., P.B. Cregan, S. Mhameed, M. Kusharska, J. Hillel, E. Lahav, and U. Lavi. 1997. An integrated genetic linkage map of avocado. Theor. Appl. Genet. 95:911–921.

Shoemaker, R.C. 1994. RFLP Map of soybean. p. 299–309. *In* R.L. Phillips and I.K. Vasil (ed.) DNA-based markers in plants. Kluwer Academic Press, Dordrecht, the Netherlands.

Shoemaker, R.C., and T.C. Olson. 1993. Molecular linkage map of soybean (*Glycine max* L. Merr.). p. 6.131–6.138. *In* S.J. O'Brien (ed.) Genetic maps: Locus maps of complex genomes. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Shoemaker, R.C., K.M. Polzin, L.L. Lorenzen, and J.E. Specht. 1996. Molecular Genetic Mapping of Soybean. p. 37–56. *In* D.P.S. Verma and R.C. Shoemaker (ed.) Soybean: Genetics, molecular biology and biotechnology. CIAB International, Wallingford, UK.

Shoemaker, R.C., and J.E. Specht. 1995. Integration of the soybean molecular and classical genetic linkage groups. Crop Sci. 35: 436–446.

Xu, S.J., R.J. Singh, and T. Hymowitz. 1997. Establishment of a cytogenetic map of soybean: Progress and Prospective. Soybean Genet. Newslet. 24:121–122.