2009

Genetic linkage map of LCP 85-384, genetic diversity of a S. spontaneum collection and the contribution of S. spontaneum to Louisiana commerical germplasm

Suman Andru
Louisiana State University and Agricultural and Mechanical College, sumangene@yahoo.com

GENETIC LINKAGE MAP OF LCP 85-384, GENETIC DIVERSITY OF A *S. SPONTANEUM* COLLECTION AND THE CONTRIBUTION OF *S. SPONTANEUM* TO LOUISIANA COMMERCIAL GERMPLASM

A Dissertation

Submitted to the Graduate faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirement for the degree of
Doctor of Philosophy

in

The School of Plant, Environmental and Soil Sciences

By

Suman Andru
B.Sc (Agri.), Acharya NG Ranga Agricultural University, 2000
M.Sc. (Genetics and Plant Breeding), Acharya NG Ranga Agricultural University, 2003
Master of Applied Statistics, Louisiana State University, 2009

December, 2009

Dedicated to

My **Parents**

Dear wife **SWATHI**

Lovely daughter **NAYANA**

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my supervisor, Dr. Collins A Kimbeng who was abundantly helpful and offered invaluable assistance, support and guidance whilst allowing me the space to work in my own way. He was always accessible to listen and give advice. The encouragement he gave me during the course of the study is commendable and without him this thesis would not have been completed. One simply could not wish for a better or friendlier supervisor. I thank my committee members, Dr. Yong Bao Pan, Dr. Gerald Myers, Dr. Charlie Johnson and Dr. Carrie Knott, for their constructive criticism while preparing this dissertation. I would like to acknowledge the help received from Dr. Yong Bao Pan who provided the mapping population and DNA samples used for the linkage mapping study. My special thanks go to Dr. Kenneth Gravois for his critical input on my thesis.

I am indebted to all my friends, especially Jie Arro, Dr. Niranjan Baisakh, Ashok, Hanama Reddy, Thirupathi, and Prakash for sharing their knowledge during countless intellectual discussions and for the fun and wonderful moments we shared together. My special thanks to Ashok Chanda, Tom Barnett, and Gouri Shankar for their encouragement and matured suggestions during the stressful times. I also thank my student worker, Anudeep Paturi, for his help and diligence while scoring the marker profiles.

My deepest gratitude goes to my family for their unflagging love and support throughout my life without whose loving support I would not be what I am today. I thank Amma (Mrs. Saraswathi), Bapu (Mr. Madanaiah), Akka (Geetha) and Thammudu (Sujan) for their understanding and endless love through the duration of my studies. My wife, Swathi, and daughter, Nayana, the only two people that have really suffered as much as I did in getting the Ph.D deserve special accolades.

The financial support from USDA-NRI, American Sugarcane League, and the School of Plant, Environmental and Soil Sciences is gratefully acknowledged.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Sequence related amplification polymorphism (SRAP) marker technique was used to assess genetic relationships and diversity among genotypes of *Saccharum* and allied genera. In the SRAP technique, the primers were arbitrarily designed with an AT- and GC-rich motif to anneal introns and exons, respectively. The level of polymorphism observed proved that the SRAP system was robust and amplified markers across species and genera and established evolutionary history interconnecting members of the *Saccharum* complex. The resolving power of the SRAP markers coupled with the fact that some of the amplicons could be amplifying gene-rich regions from diverse loci of the genome, was indicative of its potential usefulness for linkage and quantitative trait loci (QTLs) mapping in sugarcane.

S. spontaneuam has been the most important source of wild germplasm for sugarcane cultivar development in Louisiana. Genetic diversity and structure of 51 *S. spontaneum* genotypes in the local collection (USDA, Houma, LA) was assessed using amplified fragment length polymorphism (AFLP) markers. Fifty-one genotypes grouped largely according to their geographical origins namely Central and East zones. The contribution of alleles from the *S. spontaneum* collection in the modern cultivars was low and about equal. This study also allowed us to realize that *S. spontaneum* germplasm representing the west zone was not present in the collection.

A framework genetic linkage map of 'LCP 85-384' was constructed using 300 selfed progeny based on 773 single-dose (SD) markers generated by 64 AFLP, 12 TRAP and 19 SSR markers. Out of 773 SD markers, 717 markers were assigned onto 108 co-segregation groups (CGs) with a cumulative map length of 5,384 cM. With the estimated genome size of 12,720 cM, the map covered an estimated 42% of the genome. Of the 108 CGs, 31 CGs were assigned into 12 homo(eo)logous groups (HGs) based on the SSRs and information from the parental maps.

Repulsion phase linkages studied suggested the preponderance of disomic segregation between CGs within the homo(eo)logus chromosomes. The framework map established in this study will provide an important background for mapping QTLs associated with sugar related traits and thus, information will be useful for crossing and selection of clones in the breeding program.

CHAPTER 1: GENERAL INTRODUCTION

Sugarcane is a member of the *Poaceae* family like rice and *Andropogoneae* tribe like maize and sorghum. The genus *Saccharum* contains six species, namely *S. officinarum* Linnaeus (2*n* = 80), *S. spontaneum* Linnaeus (2*n* = 40-128), *S. barberi* Jeswiet (2*n* = 81-124), *S. sinense* Roxb. (2*n* = 111-120), *S. robustum* Brandes and Jeswiet ex Grassl (2*n* = 60-80), and *S. edule* Hassk. (2*n* = 60, 70, 80) (Brandes, 1958). These *Saccharum* species (excluding *S. edule*) together with other closely related interbreeding genera [e.g. *Erianthus* (sect. Ripidium), *Narenga, Sclerostachya,* and *Miscanthus* (sect. Diandra Keng)] have been designated as the *Saccharum* complex (Mukherjee, 1957; Daniels et al., 1975). The *Saccharum* complex is postulated to have been derived from a series of polyploidization and hybridization events and represents the shared gene pool from which modern sugarcane is derived (Daniels and Roach, 1987; Sobral et al., 1994).

Until the end of 19th century, the *S. officinarum* was the only cultivated species among all the *Saccharum* species because of its characteristic high sucrose content in the stalks. It is widely believed that *S. officinarum* was domesticated from the species S. *robustum* in New Guinea (Brandes, 1958; Berding and Roach, 1987). In contrast, the *S. spontaneum* species features low sucrose and high fiber content, and resistance to various biotic and abitoic stresses. Early in the 20th century, hybridization attempts between *S. officinarum* (2n=80) and its wild relative *S. spontaneum* (2n=40-128) in Java and India, and then backcrossing of hybrids to *S. officinarum* resulted in high sugar yields and disease resistance (Roach, 1972). This entire process of developing high yielding sugarcane hybrids is popularly known as 'nobilization' (Sreenivasan et al., 1987). However, an unequal transmission of chromosome number ('2n' from *S. officinarum* and 'n' from *S. spontaneum*) had taken place during the initial hybridization, and subsequent backcrossing events (Bremer, 1961; Bhat and Gill, 1985). As a result, modern

sugarcane cultivars have chromosome numbers ranging between 2n = 100-130 with a strong prevalence of aneuploidy.

Nobilization was the major breakthrough in sugarcane improvement. However, only a few clones were involved in the original 'nobilization' event, and modern sugarcane cultivars are mostly multi-generational descendants of the original backcross populations, thus making the genetic base of cultivated sugarcane very narrow (Arceneaux, 1967; Berding and Roach, 1987). Thus, continued exploitation of the wild relatives of cultivated sugarcane is essential to tackle current challenges of further improving sucrose content and general adaptability and to take advantage of new opportunities (e.g., use of sugarcane as a feedstock for renewable energy). Evaluation of available germplasm is the first step in this process. Phenotypic trait evaluation is strongly influenced by environmental conditions. In contrast, molecular marker profiles from RFLP (D'Hont et al., 1994; Coto et al., 2002), AFLP (Besse et al., 1998; Lima et al., 2002), SSR (Piperidis et al., 2001; Cordeiro et al., 2003), and TRAP (Alwala et al., 2006) were found to be repeatable irrespective of environmental conditions and were more suited to study genetic relationships and diversity among wild and cultivated germplasm of the sugarcane. The SRAP marker technique (Li and Quiros, 2001) which amplifies the DNA at AT- and GC- rich regions has proved to be a reliable tool for studying genetic diversity and phylogeny in other crops (Riaz et al., 2001; Ferriol et al., 2003; Budak et al., 2004). However, SRAP markers have not been used to study genetic diversity and phylogeny among *Saccharum* species. Therefore, the first objective of this study was to evaluate the potential of SRAP markers for assessing genetic relationship and diversity in sugarcane germplasm collections.

Among the *Saccharum* species, *S. spontaneum* is the most important source of wild germplasm for commercial sugarcane improvement in Louisiana. *S. spontaneum* clones exhibit diversity both in their habitat and chromosome number (2n=40-128). *Saccharum spontaneum*

clones are better adapted to the temperate climates of Louisiana than *S. officinarum,* which is of tropical origin (Artschwager and Brandes, 1958). Additionally, *S. spontaneum* represents an important source of genes for vigor, ratooning ability, cold tolerance, and host plant resistance to some common diseases of cultivated sugarcane in Louisiana (Dunckelman and Breaux, 1969). Recent evidence using molecular markers has shown that *S. spontaneum* has genes that could contribute positively to sucrose accumulation in sugarcane (Ming et al., 1998; Reffay et al., 2005; Alwala et al, 2008). Despite their richness in useful genes, very few *S. spontaneum* clones have so far been successfully used in the development of sugarcane cultivars in Louisiana when compared to the number of *S. spontaneum* clones in the collection. In fact, about 85% of the cultivars presently under cultivation in Louisiana can trace their ancestry to only one *S. spontaneum* clone, US56-15-8, out of a collection of about 55 *S. spontaneum* clones. Therefore, the second objective of this study was to evaluate the genetic diversity and structure of this diversity among a local collection of *S. spontaneum* germplasm and to survey diversity of S. spontaneum alleles in Louisiana commercial genotypes.

Sugar and ethanol are two of the major economic products of sugarcane. Despite the economic importance of sugarcane, the complexity of its genome limited classical genetic research when other genetically simple crops made remarkable gains (Barnes and Bester, 2000). Despite the difficulties faced in using molecular marker technologies in polyploids, efforts to unravel the sugarcane genome remain promising with the development of theoretical aspects of genetic mapping in polyploids by Wu et al. (1992) using single dose restriction fragments (SDRF). Earlier efforts in developing linkage maps were successful in the ancestral species as well as in the commercial cultivars using the full-sib (F_1) individuals (pseudo-test cross strategy) based on RAPD, RFLP, AFLP, SSR, SRAP, TRAP, and EST-SSR markers (Da Silva et al., 1993; Al Janabi et al., 1993; Guimares et al., 1999; Atienza et al., 2002; Ming et al., 2002;

Aitken et al., 2005; Raboin et al., 2006; Edme` et al., 2006; Aitken et al., 2007; Olievera et al., 2007; Alwala et al., 2008).

LCP 85-384 was a successful sugarcane cultivar for the Louisiana sugar industryupon its release in 1993. The sugar yields of LCP 85-384 were superior over the sugar yields of previously grown hybrids by about 25 % (Gravois and Bischoff, 2008). It was commercially successful and occupied 91% of the Louisiana sugarcane acreage in 2004 because of its superior agronomic characters [good cane yield (tonnes of cane per hectare), ratooning ability and planting ratio], and resistance to various biotic and abiotic stresses [leaf scald (*Xanthomonas albilineans*), mosaic viral disease (Carla virus group) and post-freeze recovery]. For this reason, LCP 85-384 has been frequently used as a parent in the Louisiana breeding programs. A genetic linkage map of the pseudo F_2 population of LCP 85-384 is therefore, considered useful to understand the coexistence of genomic components derived from its parents and the genetic basis of the heterosis observed in the F_1 generation. Finally, the third objective of this study was to construct a genetic linkage map of LCP 85-384 using AFLP, SSR, and TRAP markers based on a selfed (S_1) progeny of LCP 85-384.

In summary, the objectives of the research were:

1. To evaluate the potential of SRAP markers for assessing genetic relationship and diversity in sugarcane germplasm collections
2. a. To evaluate the genetic diversity and structure of this diversity among a local collection of *S. spontaneum* germplasm
   b. To survey the extent to which this diversity has permeated the Louisiana commercial breeding program.
3. To construct a genetic linkage map of LCP 85-384 using AFLP, SSR, and TRAP markers based on a selfed (S_1) progeny of LCP 85-384.

1.1 References

Aitken KS, Jackson PA, and McIntyre CL (2005) A combination of AFLP and SSR markers provide extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. *Theoretical and Applied Genetics* 110: 789-801.

Aitken KS, Jackson PA, and McIntyre CL (2007) Construction of genetic linkage map for *Saccharum officinarum* incorporating both simplex and duplex markers to increase genome coverage. *Genome* 50: 742-756.

Al-Janabi SM, Honeycutt RJ, McClelland M, and Sobral BWS (1993) A genetic linkage map of *Saccharum spontaneum* L. 'SES 208'. *Genetics* 134: 1249-1260.

Alwala S, Kimbeng CA, Veremis JC, and Gravois KA (2008) Linkage mapping and genome analysis in *Saccharum* interspecific cross using AFLP, SRAP and TRAP markers. *Euphytica* 164: 37–51.

Alwala S, Suman A, Arro JA, Veremis JC and Kimbeng CA (2006) Target region amplification polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections. *Crop Science* 46: 448-455.

Arceneaux G (1967) Cultivated sugarcanes of the world and their botanical derivation. *Proceedings International Society of Sugarcane Technologists* 12:844-854.

Artschwager E and Brandes EW (1958) 'Sugarcane (*Saccharum officinarum*), origin, characteristics and descriptions of representative clones'. *US Dept Agriculture Handbook* no. 122 (USDA, Washington, DC).

Atienza SG, Satovic Z, Peterson KK, Dolstra O, and Martin A (2002) Preliminary genetic linkage map of *Miscanthus sinensis* with RAPD markers. *Theoretical and Applied Genetics* 105: 946-952.

Berding N, Roach BT (1987) Germplasm collection, maintenance, and use. In: Heinz DJ (ed.) *Sugarcane Improvement Through Breeding*. Amsterdam: Elsevier, pp. 143–210.

Besse P, Taylor G, Carroll B, Berding N, Burner D, and Mc Intyre CL (1998) Assessing genetic diversity in a sugarcane germpalsm collection using an automated AFLP analysis. *Genetica* 104:143-153.

Bhat S and Gill SS (1985) The implication of 2n egg gametes in nobilization and breeding of sugarcane. *Euphytica* 34:377-384.

Brandes EW (1958) Origin, classification, and characteristics. In: Artschwager E and Brandes EW (eds.) *Sugarcane* (*Saccharum officinarum* L.). *US Dept of Agriculture Handbook* vol 122, pp. 1-35.

Bremer G (1961) Problems in breeding and cytology of sugarcane. *Euphytica* 10:59–78.

Budak H, Shearman RC, Parmaksiz, Gaussoin RE, riordan TP, Dweikat I (2004) Molecular characterization of Buffalograss germplasm using sequence related amplified polymorphism markers. *Theoretical and Applied Genetics* 108:328-334.

Cordeiro GM, Pan YB, and Henry RJ (2003) Sugarcane microsatellites for the assessment of genetic diversity in sugarcane germplasm. *Plant Science* 165:181–189.

Coto O, Cornide MT, Calvo D, Canales E, D'Hont A, and de Prada F (2002) Genetic diversity among wild sugarcane germplasm from Laos revealed with markers. *Euphytica* 123(1):121-130.

Da Silva JAG, Sorrells ME, Burnquist W, and Tanksley SD (1993) RFLP linkage map and genome analysis of *Saccharum spontaneum. Genome* 36: 782–791.

Daniels J and Roach BT (1987) Taxonomy and evolution. In: Heinz D.J (ed.) *Sugarcane Improvement Through Breeding*. Amsterdam: Elsevier, pp. 143–210.

Daniels J, Smith P, Paton N, Williams CA (1975) The origin of the genus *Saccharum*. *Sugarcane Breeding Newsletter* 36:24-39.

D'Hont A, Lu YH, de León DG, Grivet L, Feldmann P, Lanaud C, and Glaszmann JC (1994) A molecular approach to unraveling the genetics of sugarcane, a complex polyploid of the Andropogoneae tribe. *Genome* 37(2): 222- 230.

Dunckelman PH and Breaux RD (1969) Agronomic characteristics of *Saccharum spontaneum* in culture in Houma, Louisiana. *International Sugar Journal* 71:333-334.

Edmé SJ, Glynn NG, and Comstock JC (2006) Genetic segregation of microsatellite markers in *Saccharum officinarum* and *S. spontaneum. Heredity* 97: 366-375.

Ferriol M, Pico B and Nuez F (2003) Genetic diversity of a germplasm collection of *Cucurbita pepo* using SRAP and AFLP markers. *Theoretical and Applied Genetics* 107:271-282.

Gravois KA and Bischoff KP (2008) New sugarcane varieties to the rescue. *Louisiana Agriculture* 51(2):14-16.

Guimaráes CT, Honeycutt RJ, Sills GR, and Sobral BWS (1999) Genetic linkage maps of *Saccharum officinarum* L. and *Saccharum robustum* Brandes & Jew. Ex Grassl. *Genetics and Molecular Biology* 22: 125-132.

Li G and Quiros CF (2001) Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in *Brassica*. *Theoretical and Applied Genetics* 103:455-461.

Lima MLA, Garcia AAF, Oliveira KM, Matsuoka S, Arizono H, de Souza CL Jr., and de Souza AP (2002) Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugarcane (*Saccharum* spp). *Theoretical and Applied Genetics* 104:30–38.

Ming R, Liu S-C, Bowers JE, Moore PH, Irvine JE, and Paterson AH (2002) Construction of *Saccharum* consensus genetic map from two interspecific crosses. *Crop Science* 42: 570-583.

Ming R, Liu SC, Lin YR, Da Silva JAG, Wilson W, Braga D, van Devnze A, Wenslaff F, Wu KK, Moore PH, Burnquist W, Sorrells ME, Irvine JE, Paterson AH (1998) Detailed alignment of *Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and polyploid genomes. *Genetics* 150:1663–1682.

Mukherjee SK (1957). Origin and distribution of *Saccharum*. *Botanical Gazette* 199:55–61.

Oliveira KM, Pinto LR, Marconi TG, Margarido GRA, Pastina MM, Teixeira LHM, Figueira AV, Ulian EC, Garcia AAF, and Souza AP (2007) Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross. *Molecular Breeding* 20: 189-208.

Piperidis G and D'Hont A (2001) Chromosome composition analysis of various *Saccharum* interspecific hybrids by genomic *in situ* hybridization (GISH). *Proceedings of International Society of Sugar Cane Technologists* 11:565-566.

Raboin LM, Oliveira KM, Lecunff L, Telismart H, Roques D, Butterfield M, Hoarau JY, and D'Hont A (2006) genetic mapping in sugarcane, a high polyploidy, using bi-parental progeny: identification of a gene controlling stalk colour and a new rust resistance gene. *Theoretical and Applied Genetics* 112: 1382-1391.

Reffay N, Jackson P, Aitken K, Hoarau J, D'Hont A, Besse P, and McIntyre CL (2005) Characterization of genome regions incorporated from an important wild relative into Australian sugarcane. *Molecular Breeding* 15:367–381.

Riaz A, Li G, Quresh Z, Swatt MS and Quiros CF (2001) Genetic diversity of oilseed *Brassica* inbred lines based on sequence-related amplified polymorphism and its relation to hybrid performance. *Plant Breeding* 120:411–415.

Roach B (1972) Nobilisation of sugarcane. *Proceedings of International Society of Sugar Cane Technologists* 14: 206-216.

Sobral BWS, Braga DPV, LaHood ES and Keim P (1994) Phylogenetic analysis of chloroplast restriction enzyme site mutations in the Saccharinae Griseb subtribe of the Andropogoneae Dumort. Tribe. *Theoretical and Applied Genetics* 87:843-853.

Sreenivasan TV, Ahloowalia BS and Heinz DJ (1987) Cytogenetics. In: Heinz DJ (ed.) *Sugarcane Improvement Through Breeding*. Amsterdam: Elsevier, pp. 211-253.

Wu KK, Burnquist W, Sorrells ME, Tew TL, Moore PH, and Tanksley SD (1992) The detection and estimation of linkage in polyploids using single-dose restriction fragments. *Theoretical and Applied Genetics* 83: 294-300.

CHAPTER 2: SEQUENCE-RELATED AMPLIFIED POLYMORPHISM (SRAP) MARKERS FOR ASSESSING GENETIC RELATIONSHIPS AND DIVERSITY IN SUGARCANE GERMPLASM COLLECTIONS

2.1 Introduction

Sugarcane is a member of the grass family *Poaceae*, tribe *Andropogoneae,* and genus *Saccharum*. The genus *Saccharum* contains six species, namely *S. officinarum* Linnaeus (2*n* = 80), *S. spontaneum* Linnaeus (2*n* = 40-128), *S. barberi* Jeswiet (2*n* = 81-124), *S. sinense* Roxb. (2*n* = 111-120), *S. robustum* Brandes and Jeswiet ex Grassl (2*n* = 60-80), and *S. edule* Hassk. (2*n* = 60, 70, 80) (Brandes, 1958). This classification, however, has been adjusted several times. According to Irvine (1999), only the two wild species (*S. spontaneum* and *S. robustum*) deserve species status, whereas the cultivated species (*S. officinarum*, *S. barberi, S. sinense,* and *S. edule*) should be designated as horticultural classes. *Saccharum officinarum* is thought to have been domesticated from the 2*n* = 80 form of *S. robustum* (Artschwager and Brandes, 1958); *S. barberi* and *S. sinense* are believed to have originated from natural hybridization events between *S. officinarum* and *S. spontaneum* (Daniels and Roach, 1987); *S. edule*, characterized by its abortive flowers, is thought to have arisen from intergeneric crosses between *S. officinarum* or *S. robustum* and a related genus (e.g. *Mischantus*), or derived from *S. robustum* (Williams *et al.,* 1974; Daniels and Roach, 1987; Irvine, 1999; Amalraj and Balasundaram, 2006). These *Saccharum* species (excluding *S. edule*) together with other closely related interbreeding genera [e.g. *Erianthus* (sect. Ripidium), *Narenga, Sclerostachya,* and *Miscanthus* (sect. Diandra Keng)] have been designated as the *Saccharum* complex (Mukherjee, 1957; Daniels *et al.,* 1975). The *Saccharum* complex is postulated to have been derived from a series of polyploidization and hybridization events and represents the shared gene pool from which modern sugarcane is derived (Daniels and Roach, 1987; Sobral *et al.,* 1994).

Modern sugarcane cultivars are interspecific hybrids derived by crossing the previously

cultivated *S. officinarum* with the wild *S. spontaneum* species (Price, 1963; Stevenson, 1965) to respond to diseases that affected sugar production in commercial fields. In a process coined 'nobilization', genes for stress and ratooning ability were introgressed from *S. spontaneum* into the cultivated background followed by a few backcrosses to recover the sucrose genes from the female *S. officinarum* parent (Price, 1965; Roach, 1986; Sreenivasan *et al.,* 1987). Only a few clones were involved in the original 'nobilization' events, and modern cultivars are mostly multi-generational descendants of the original backcross populations, which makes the genetic base of cultivated sugarcane narrow (Arceneaux, 1967; Berding and Roach, 1987).

Continued exploitation of the wild relatives of cultivated sugarcane is essential to meet current challenges of further improving sucrose content, general adaptability and to take advantage of new opportunities (e.g. use of sugarcane as a feedstock for renewable energy). Many important agronomic traits being influenced by environmental conditions, molecular markers are more suited to study genetic relationships and diversity among wild and cultivated germplasm and to monitor and ascertain the presence or absence of specific alleles linked to traits of interest during introgression.

Sequence-related amplification polymorphism (SRAP) is a molecular marker technique that has been employed in genetic diversity and phylogenetic studies of crop species including *Brassica napus* L. (Riaz *et al.,* 2001), *Cucurbita maxima* Duchesne (Ferriol *et al.,* 2003a), *C. pepo* L. (Ferriol *et al.,* 2003b), *C. moschata* (Ferriol *et al.,* 2004), and buffalograss (*Buchloë dactyloides* Nutt.; Budak *et al.,* 2004) but not sugarcane. SRAP markers have shown a great affinity to amplify gene-rich regions of the *Brassica* genome (Li and Quiros, 2001). The objective of this study was to evaluate the potential of SRAP markers for assessing genetic relationships and diversity in sugarcane germplasm collections.

2.2 Materials and Methods

2.2.1Plant Materials and DNA Extraction

Genotypes representing five *Saccharum* species (namely *S. officinarum, S. barberi*, *S. sinense, S. spontaneum,* and *S. robustum*) as well as cultivars, cultivar-derived mutants, and F_1 interspecific hybrids were used in this study (Table 2.1). Two clones (each of *Miscanthus* and *Erianthus*) were used as outgroups. As a baseline study, the genotypes were chosen to include some wild clones (SES 147B, Coimbatore, and LA Purple) and very early hybrids (e.g. POJ 2878) used as progenitors of US sugarcane cultivars, and different BC generation cultivars (the CP, LCP, and HoCP). The genotypes Dwarf1 and Dwarf2 are genetic mutants derived from spontaneous mutations in the cultivar LCP 81-137 (Burner, 1999). The 16-Low and 40-High are F_1 interspecific hybrids from a cross between LA Striped (S. *officinarum*) x SES 147B (*S. spontaneum*) with low and high sucrose content, respectively. The genotypes Dwarf1, Dwarf2, 16 Low and 40 High were included in the study as checks. These 30 genotypes make up part of the sugarcane working germplasm collection maintained at the USDA Sugarcane Research Unit at Houma, Louisiana and have been used in various genotyping studies (Alwala et al., 2006; Arro *et al.,* 2006).

Young leaves were collected from each genotype, placed immediately in ice, and stored at -80^{o} C. The leaves were later ground to a powder in liquid nitrogen. Genomic DNA was extracted using the Plant DNeasy Mini Kit (Qiagen, Valencia, CA) following the manufacturer's protocol. Concentrations of extracted DNA were estimated in 1.5% agarose gel, in comparison with known concentration of lambda DNA. Working DNA samples were prepared at 50 to 80 ng/µl for PCR amplification.

2.2.2 SRAP Analysis

Thirty-one SRAP primer combinations based on four forward and eight reverse primers

were used to amplify the 30 genotypes (Table 2.2). The reverse primers were 5'-end labeled with IR-Dye 700 and 800 (MWG Biotech AG, Germany). The PCR reactions were performed as described by Ferriol *et al.* (2003b) with some modifications. Briefly, PCR was performed in 10-µl reaction volume containing 0.75 µl of 1 µM each of forward primer and reverse primer, 1 µl

Table 2.1: Description of 30 genotypes of the *Saccharum* complex (made up of five *Saccharum* species and related genera) used in a SRAP marker analysis.

Serial Number	Genotype name	Genera or species[a]	Code
1	Kalingpong	*Erianthus*	Er
2	Dwarf1	*Saccharum* species hybrid (mutant)	DW1
3	Dwarf2	*Saccharum* species hybrid (mutant)	DW2
4	16 Low†	*Saccharum* species hybrid (F_1)	Hy1
5	40 High†	*Saccharum* species hybrid (F_1)	Hy2
6	POJ2878	*Saccharum* species hybrid (cultivar)	Cu1
7	LCP 85-384	*Saccharum* species hybrid (cultivar)	Cu2
8	CP 77-310	*Saccharum* species hybrid (cultivar)	Cu3
9	CP 77-407	*Saccharum* species hybrid (cultivar)	Cu4
10	LCP 85-845	*Saccharum* species hybrid (cultivar)	Cu5
11	Miscanthus	*Miscanthus*	Mi
12	Ganapathy	*S. barberi*	Sb1
13	Chin	*S. barberi*	Sb2
14	LA Stripe	*S. officinarum*	So1
15	LA Purple	*S. officinarum*	So2
16	Cuba	*S. officinarum*	So3
17	IN 84-064A	*S. officinarum*	So3
18	NG 57-54	*S. robustum*	Sr1
19	NG 57-159	*S. robustum*	Sr2
20	Molokai 5573	*S. robustum*	Sr3
21	IMP72-232	*S. robustum*	Sr4
22	NG 77-218	*S. robustum*	Sr5
23	Chukche	*S. sinense*	Ssi
24	SES 147B	*S. spontaneum*	Ssp1
25	Coimbatore	*S. spontaneum*	Ssp2
26	MPTH 97-213	*S. spontaneum*	Ssp3
27	MPTH 97-200	*S. spontaneum*	Ssp4
28	MPTH 97-107	*S. spontaneum*	Ssp5
29	PIN 84-B	*S. spontaneum*	Ssp6
30	Molokai1032B	*S. spontaneum*	Ssp7

[a]Original sugarcane cultivars (e.g. POJ 2878) were derived from crossing mainly between *S. officinarum* and *S. spontaneum* followed by several generations of backcrosses to *S. officinarum*. Present-day cultivars are selections derived from cultivar x cultivar crosses.
†*Saccharum* species hybrid (F_1 between LA Stripe (*S. officinarum*) and SES147B (*S. spontaneum*)

of 25mM $MgCl_2$, 1 µl of 10X PCR buffer, 1 µl of 2.5mM dNTPs (Promega, Madison, WI), 0.2 µl of 5U *Taq* DNA polymerase (Promega, Madison, WI), and 1-1.5 µl of 50-80 ng/µl of genomic DNA. The thermal cycler profile for PCR amplification was set on an *i-cycler* (BioRad Labs, Hercules, CA) as follows: denaturation at 94^{o}C for 4 min, followed by five cycles of denaturing at 94^{0}C for 1 min, annealing temperature at 35^{o}C for 1 min, and elongation at 72^{o}C for 1 min. In the remaining 30 cycles, the annealing temperature was increased to 50^{o}C for 1 min with a final elongation step at 72^{0}C for 7 min (Ferriol *et al.,* 2003b). The amplified fragments were separated on 6.5% polyacrylamide gels using the Li-Cor 4300 Global DNA sequencer (Li-Cor, Lincoln, NE). Digital images of the gel were saved onto a computer and scored manually.

Table 2.2: The forward and reverse SRAP primer sequences used to genotype the *Saccharum* germplasm.

Forward	
SRAP Primer	**Sequence (5'-3')**
SF1	TGAGTCCAAACCGGATA
SF2	TGAGTCCAAACCGGAGC
SF3	TGAGTCCAAACCGGAAT
SF4	TGAGTCCAAACCGGACC
Reverse	
T1	GACTGCGTACGAATTAAT
T2	GACTGCGTACGAATTTGC
T3	GACTGCGTACGAATTGAC
T4	GACTGCGTACGAATTTGA
T5	GACTGCGTACGAATTAAC
T6	GACTGCGTACGAATTGCA
T7	GACTGCGTACGAATTCAA
T8	GACTGCGTACGAATTCAC

2.2.3 Data Analysis

Digital images were scored as '1' for presence and '0' for absence of clear and unambiguous DNA fragments. The polymorphism information content (PIC) for each primer combination was determined by averaging the allele frequency over all loci using the formula:

PIC=1-$\sum f^2_i$, where f_i is the frequency of the i^{th} allele (Weir, 1990). Genetic similarity (GS_{ij}) was calculated for each pair of genotypes using Nei and Li's (Dice) similarity index (Nei and Li, 1979). This index ignores 0-0 matches in the pairwise comparisons. The GS_{ij} values were used to compute genetic distances (D_{ij}) based on the formula $D_{ij} = 1-GS_{ij}$. The genetic distance matrix was subjected to cluster analysis using the unweighted pair group method with arithmetic mean (UPGMA) in NTSYS-pc v2.1 (Rohlf, 2000). Goodness of fit between clusters in the dendrogram and the similarity index were tested by computing the cophenetic values using the COPH and MXCOMP procedures in NTSYS-pc v2.1. For comparison and to verify the robustness of the clusters, a bootstrap analysis with 10,000 replications was performed using PAUP v4.0b (Swofford, 1998). The genetic distance matrix was also subjected to a non-metric multidimensional scaling (NMDS) analysis using PROC MDS in SAS v8.2 (SAS Institute, 2004) with the ORDINAL option to highlight the resolving power of the ordination. A Shepard's plot was generated to assess goodness of fit of the NMDS plot to the distance matrix.

2.2.4 Sequencing of SRAP-derived DNA Fragments

Some of the SRAP-derived DNA fragments of a *S. officinarum* (LA Striped) and *S. spontaneum* (SES 147B) genotype amplified with SF1/T3, SF2/T3 and SF3/T3 primer combinations were excised from a silver stained PAGE (polyacrylamide gel electrophoresis) gel. The DNA fragments were re-amplified with the corresponding primer sequences and both monomorphic and polymorphic fragments were sequenced directly without cloning. The sequences obtained were compared against EST (Expressed Sequence Tag) sequences available in TIGR (The Institute for Genomic Research) database website (http://www.tigr.org/) using the BLASTn search algorithm. Monomorphic fragments were compared for homology using the ClustalW2 algorithm found at the European Bioinformatic Institute website (http://www.ebi.ac.uk/).

2.3 Results

2.3.1 SRAP Marker Profile, Polymorphism and PIC Values

Distinct DNA profiles were produced on all 30 genotypes by each of the 31 SRAP primer combinations with fragments ranging in size from 50 to 700 bp. A total of 1364 such DNA fragments were produced, with individual primer combinations amplifying from 18 (SF4/T8) to 92 (SF4/T2) for an average of 44 fragments (Table 2.3). A total of 1135 fragments (83%) were polymorphic, with 17 (SF4/T8) to 84 (SF4/T2) polymorphic fragments produced per primer combination for an average of 37 polymorphic fragments. The overall percent polymorphism was generally high and comparable to that reported from AFLP analyses of sugarcane germplasm (Besse *et al.,* 1998; Arro *et al.,* 2006; Selvi *et al.,* 2006). However, fewer fragments (44) were amplified with SRAP than with AFLP (about 110 fragments) on average.

The polymorphism information content (PIC) value is often used to measure the informativeness of a genetic marker system (Vuylsteke *et al.,* 2000) and the theoretical maximum PIC value for a dominant marker is 0.5. In this study, PIC values varied among SRAP primer combinations, ranging from 0.16 (SF1/T7) to 0.32 (SF2/T5) with an average of 0.22. These values are comparable to those previously reported in a related study using TRAP markers (Alwala *et al.,* 2006).

2.3.2 Genetic Diversity and Relationships among Genotypes

Cluster analysis and non-metric multi-dimensional scaling (NMDS) were used to assess the genetic diversity and relationships among the genotypes used in this study. Two major clusters, supported by high boostrap values (> 96%), were identified and included 27 of the 30 genotypes (Fig. 2.1). The other three genotypes (SES 147B (*S. spontaneum*), *Miscanthus* and Kalingpong (*Erianthus*)) made three single-clone clusters that joined the two major clusters at GS levels of 0.78, 0.72, and 0.56, respectively, with branches supported by high bootstrap

Table 2.3: Polymorphism and PIC values for 31 SRAP primer combinations used in genotyping 30 genotypes representing the *Saccharum* complex (made up of five *Saccharum* species and related genera).

Primer combinations	Total no. of amplified bands	Polymorphic bands		PIC[a]
		no.	%	
SF1/T1	47	36	76.60	0.19
SF1/T2	89	68	76.40	0.18
SF1/T3	43	35	81.40	0.28
SF1/T4	67	46	68.66	0.19
SF1/T5	50	35	70.00	0.17
SF1/T6	30	29	96.67	0.25
SF1/T7	39	28	71.79	0.16
SF1/T8	32	18	56.25	0.25
SF2/T1	48	43	89.58	0.22
SF2/T2	40	39	97.50	0.16
SF2/T3	43	34	79.07	0.17
SF2/T4	44	39	88.64	0.18
SF2/T5	47	38	80.85	0.32
SF2/T6	36	35	97.22	0.19
SF2/T8	24	19	79.17	0.23
SF3/T1	36	33	91.67	0.26
SF3/T2	89	77	86.52	0.18
SF3/T3	67	56	83.58	0.25
SF3/T4	42	31	73.81	0.26
SF3/T5	40	38	95.00	0.26
SF3/T6	38	34	89.47	0.27
SF3/T7	31	29	93.55	0.19
SF3/T8	32	28	87.50	0.21
SF4/T1	43	36	83.72	0.23
SF4/T2	92	84	91.30	0.19
SF4/T3	32	24	75.00	0.21
SF4/T4	45	34	75.56	0.20
SF4/T5	37	34	91.89	0.25
SF4/T6	21	19	90.48	0.32
SF4/T7	22	19	86.36	0.26
SF4/T8	18	17	94.44	0.25
Total	1364	1135		
Average	44	36.61	83.21	0.22

[a]PIC = polymorphism information content.

values. The *S. officinarum, S. sinense, S. barberi,* and *S. robustum* clones, along with the cultivar- derived mutants and hybrids, were included in Cluster I. All the *S. spontaneum*

genotypes grouped in Cluster II, except SES 147B (Ssp1) which was found to be distinct. At the 0.84 similarity level, cluster I was divided into five subgroups. LA Striped clustered closer to a group of *S. robustum* clones having GS values ranging from 0.85 to 0.88. A mixed subgroup of *S. officinarum*, *S. sinense*, *S. barberi*, and POJ 2878 was formed at GS values ranging from 0.88 to 0.93. The two dwarf genotypes shared the closest relationship (GS = 0.96) in the study and were joined to a tight subgroup formed by cultivars and LA Purple with GS values ranging from 0.92 to 0.94. The two F_1 hybrids (16-Low and 40-High) between LA Striped and SES 147B with a GS=0.88 were closer to each other than to any of the studied genotypes and were quite distant from their parents. In general, the high (>0.80) cophenetic correlation value of 0.92 and the strong bootstrap support for branches in the dendrogram indicate that the UPGMA clustering in this study represented a good fit to the distance matrix.

2.3.3 Non-metric Multi-dimensional Scaling (NMDS)

The genetic distance matrix was also analysed using the non-metric multi-dimensional scaling (NMDS) method based on three dimensions. The stress value for the three axes was 0.09, which explained 91% of the variation among the genotypes. In the three- dimensional plot (Fig. 2.2) generated, the position of the genotypes was found to be consistent with the grouping pattern of the UPGMA clustering. The first and third axes separated the *Erianthus* (Kalingpong) and *S. spontaneum* clone (SES 147B) from the other genotypes, respectively; the second axis separated *Miscanthus* and the *S. spontaneum* clones from the group of hybrids, cultivars, *S. sinense*, *S. barberi*, *S. robustum* and *S. officinarum* clones. The Shephard's plot (Fig. 2.3) indicated that the NMDS plot represented an excellent fit to the distance matrix.

2.3.4 Mean Genetic Distance between *Saccharum* spp.

Based on the genotypes included in this study, *S. officinarum*, *S. spontaneum*, and *S. robustum* had the same level of intra-species genetic similarity, with GS values around 0.74,

0.70, and 0.73, respectively (Table 2.4). The highest genetic similarity (83%) was observed within cultivars. As groups, the *S. officinarum*, *S. robustum*, and cultivars were more similar among themselves than to the *S. spontaneum* group. On the other hand, *S. spontaneum* shared the least genetic similarity with *S. robustum* (0.62) and with *S. officinarum* (0.64).

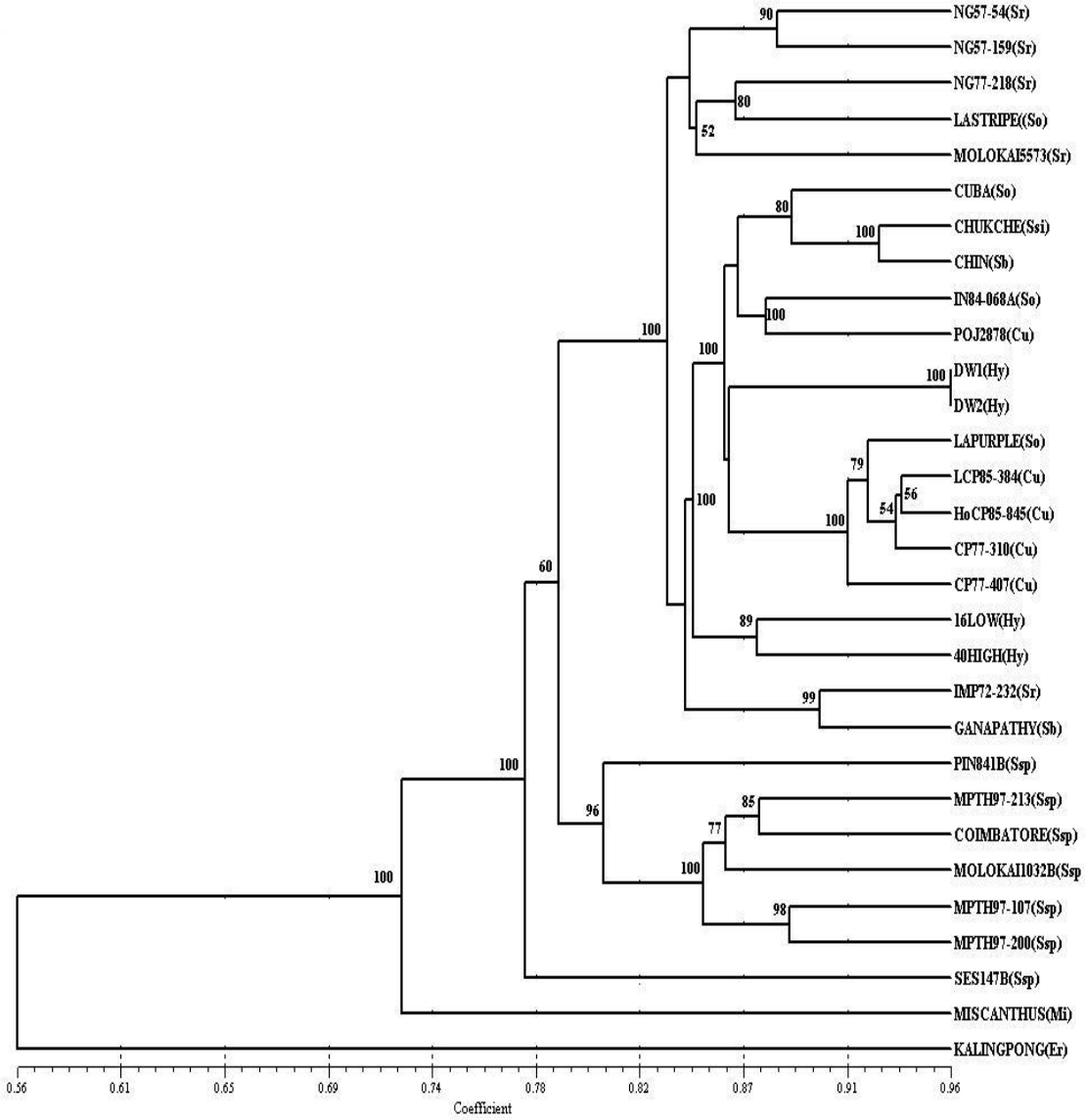


Fig. 2.1: UPGMA dendrogram showing relationships among 30 genotypes of the *Saccharum* complex, represented by *Erianthus*, *Miscanthus*, five *Saccharum* species, and cultivars. So= *S. officinarum*; Sr = *S. robustum*; Ssp = *S. spontaneum*; Sb = *S. barberi*; Ssi = *S. sinense*; Cu = cultivars; DW = dwarf genetic mutants derived from the cultivar LCP 81-137; Hy = low and high-sucrose hybrids derived from a cross between La Stripe (*S. officinarum*) x SES 147B (*S. spontaneum*).

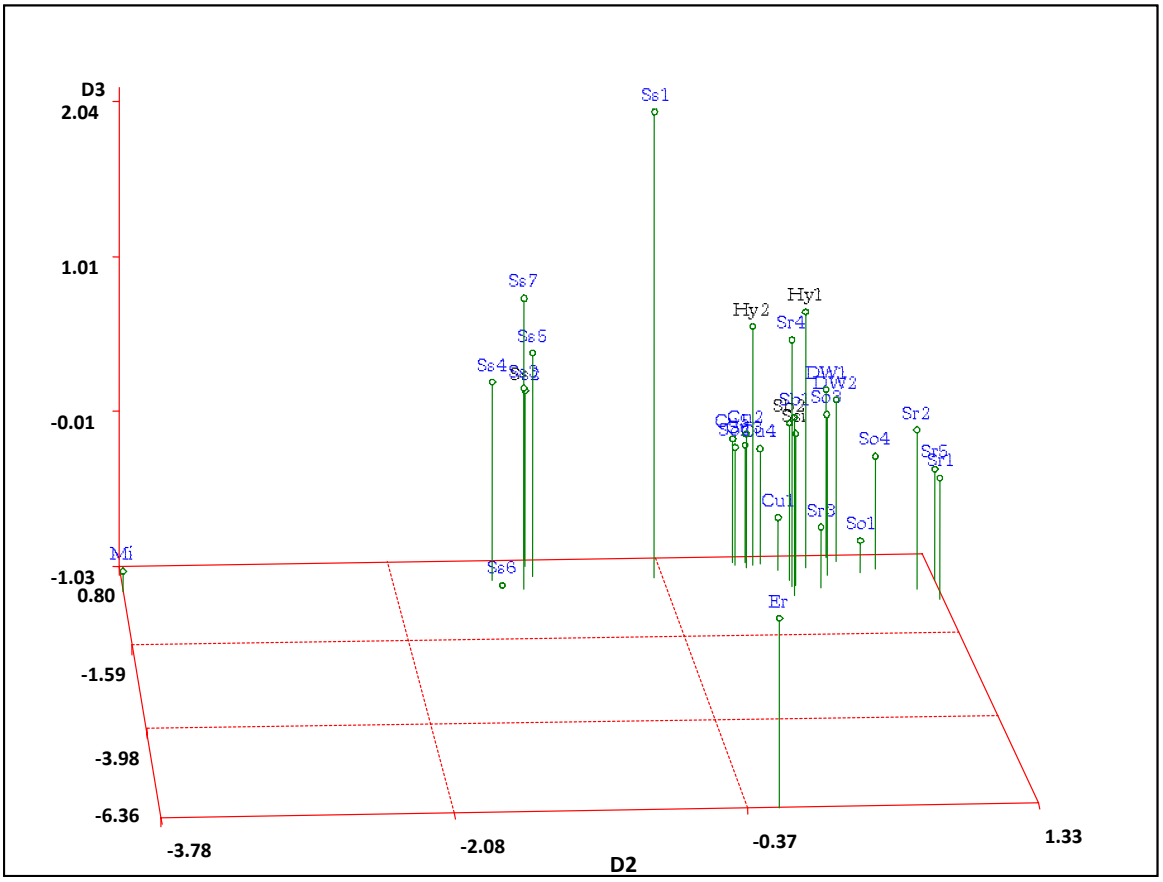


Fig 2.2: A non-metric multi-dimensional scaling plot for 30 genotypes of the *Saccharum* complex based on the SRAP markers. The coding of the genotypes is given in Table 2.1.

2.3.5 Species-specific Markers

One hundred nineteen (8.7%) of the 1364 amplified SRAP fragments were either genus- or species-specific when considering the four *S. officinarum,* five *S. robustum*, and seven *S. spontaneum* clones included in this study (Table 2.5). Markers that were present in at least two genotypes of a species and completely absent in other species were regarded as species-specific (Jannoo *et al.,* 1999). Only three primer combinations (SF2/T4, SF3/T7, and SF4/T8) did not amplify bands unique to any of the *Saccharum* species. Fifteen bands were found to be specific to *S. officinarum* or to *S. robustum*, whereas 89 markers differentiated the *S. spontaneum* clones from the other species. The primer combination SF3/T1 yielded the highest number of unique bands across all three species within a size range of 300 to 500bp. The three primer

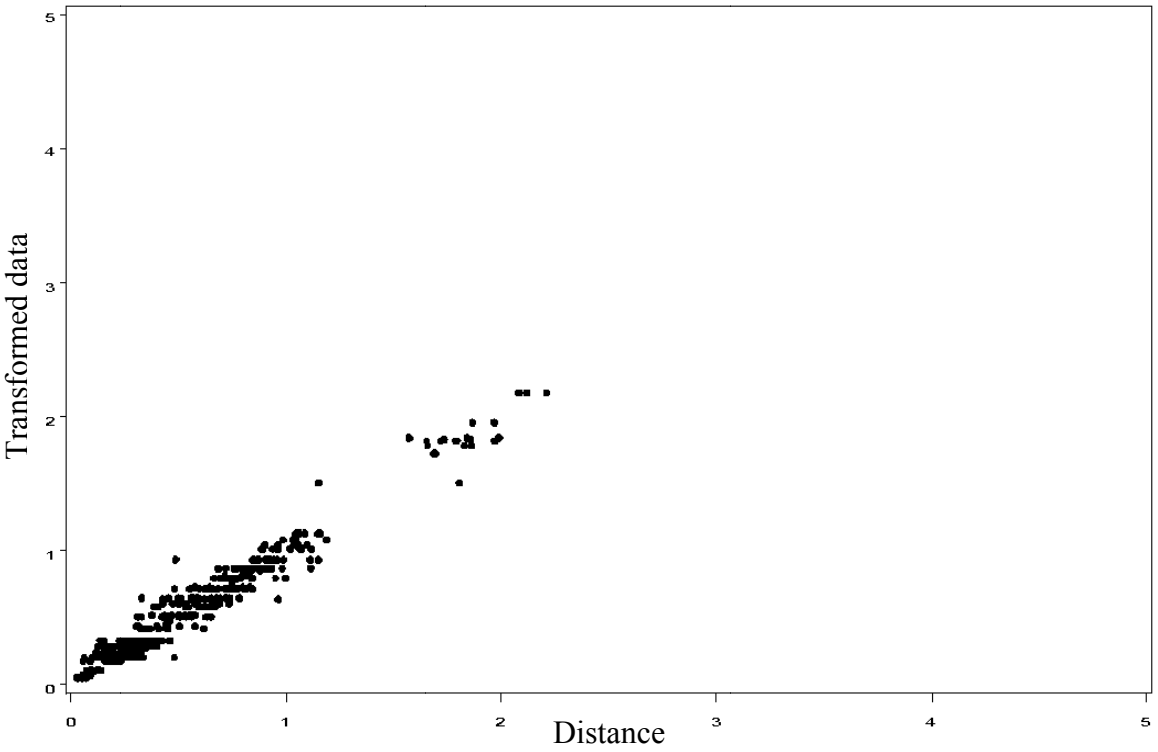


Fig. 2.3: Shephard's plot showing the overall fit of 3D-NMDS plot to the distance matrix.

Table 2.4: Mean genetic similarity (GS) estimates within and between *Saccharum* species and cultivars based on SRAP marker analysis.

	S. officinarum	***S. spontaneum***	***S. robustum***	**Cultivars**
S. officinarum	0.74[a]			
S. spontaneum	0.64	0.70		
S. robustum	0.72	0.62	0.73	
Cultivars	0.77	0.67	0.71	0.83

[a] Estimated as per the Dice coefficient (Nei and Li, 1979).

combinations (viz. SF1/T3, SF2/T1, and SF4/T2) amplified from eight to 10 unique bands in the *S. spontaneum* clones. On the other hand, *S. robustum* was identified with a total of 15 species-specific markers with band sizes ranging from 300 to 400 bp.

The SRAP fingerprinting data of cultivars were compared with those of *S. officinarum* and *S. spontaneum* to determine the genomic contribution of these progenitor species to modern cultivars. Species-specific markers in *S. officinarum* and *S. spontaneum* were traced for their

presence and absence in the modern cultivars. Among the markers traced in the cultivars, 98 (71.5%) were inherited from *S. officinarum* and 39 (28.5%) from *S. spontaneum.*

2.3.6 Sequencing of SRAP-derived Fragments

Sequences were obtained for seven monomorphic and four polymorphic fragments with sizes ranging from 99 bp to 184 bp (Table 2.6). The GC content of the sequenced SRAP fragments was high ranging from 41 to 58%. Percent homology among the monomorphic fragments following alignment with ClustalW2 ranged from 76 to 83 %. Although not ideal, this is high considering the large genome size of sugarcane and the ability of the dominant SRAP markers to produce co-migrating fragments from different regions of the genome. Additional steps in purifying and cloning before sequencing would be necessary to achieve sequences with a high level of fidelity.

For monomorphic fragments, the Blastn search was conducted using only one of the two sequences taken from a portion displaying the most homology following alignment. Blastn search of the TIGR EST database (http://www.tigr.org/) revealed homology with EST sequences of rice *(Oryza sativa*), maize (*Zea mays*), and *S. officinarum.* A monomorphic fragment of 163 bp amplified by SF3+T3 showed high homology with *S. officinarum* (84%, E = 8.6e-18) and *Oryza sativa* (68%, E=1.2) ESTs in TIGR. The *Oryza sativa* EST was said to be similar to a plastid division protein (FtsZ) of *Arabidopsis thaliana* and a Blastx search of the NCBI database (http://www.ncbi.nlm.nih.gov/) using the *Oryza sativa* EST sequence revealed high homology (65%, E=2e-19) with the protein. Similarly, a polymorphic fragment of 148 bp amplified by SF1+T3 in *S. spontaneum* showed high homology with *S. officinarum* (93%, E = 4.1e-22) and *Zea mays* (61%, E=5.1) ESTs in TIGR. The *Zea mays* EST has been tentatively annotated to a response regulator receiver which is a transcriptional regulatory protein.

Table 2.5: Number of species-specific markers amplified by 31 SRAP primer combinations among three *Saccharum* species.

Primer combinations	*S. officinarum*	*S. spontaneum*	*S. robustum*	Across species
SF1/T1	0	3	0	3
SF1/T2	0	1	0	1
SF1/T3	0	8	0	8
SF1/T4	0	6	0	6
SF1/T5	0	3	1	4
SF1/T6	0	2	0	2
SF1/T7	0	5	0	5
SF1/T8	0	0	2	2
SF2/T1	0	8	0	8
SF2/T2	2	2	1	5
SF2/T3	1	1	0	2
SF2/T4	0	0	0	0
SF2/T5	0	2	2	4
SF2/T6	0	2	1	3
SF2/T8	0	1	0	1
SF3/T1	3	6	2	11
SF3/T2	2	3	2	7
SF3/T3	2	6	0	8
SF3/T4	0	3	0	3
SF3/T5	2	3	0	5
SF3/T6	0	4	0	4
SF3/T7	0	0	0	0
SF3/T8	0	0	2	2
SF4/T1	1	1	1	3
SF4/T2	0	10	0	10
SF4/T3	0	1	0	1
SF4/T4	2	1	0	3
SF4/T5	0	1	1	2
SF4/T6	0	2	0	2
SF4/T7	0	4	0	4
SF4/T8	0	0	0	0
Total	15	89	15	119

2.4 Discussion

While various molecular marker techniques have been used to characterize sugarcane germplasm, this study was the first one to evaluate the potential of SRAP markers at inferring

genetic diversity within and among *Saccharum* and related genera. The SRAP technique, by amplifying both intronic and exonic regions of the genome, provides valuable markers for use in plant breeding. High levels of polymorphism were detected with an average PIC value of 0.22 and an average number of 37 polymorphic fragments per primer pair. This generated sufficient polymorphism (1135 out of 1364) to discriminate each of the 30 genotypes under study (Dudley, 1994) and makes SRAP comparable to the AFLP technique at amplifying the *Saccharum* genome.

The levels of polymorphism revealed a relatively low to moderate amount of intra and inter-genetic variability among this group of 30 genotypes. Very close relationships exist among them with GS values ranging from 0.60 to 0.96. The two major clusters illustrated in the dendrogram were connected at a similarity level of 0.79 with GS values ranging from 0.80 to 0.96. As expected, the closest relationships were detected among the group of cultivars, F_1 nterspecific hybrids, and LA Purple. These cultivars represent different generations of a recurrent selection program and LA Purple is one *S. officinarum* clone that was used repeatedly in the parentage of cultivars released in both the Florida and Louisiana industries (Deren, 1995).

Despite such close relationships, the SRAP system was effective at discriminating the genotypes according to the accepted lineages among members of the *Saccharum* complex. The *Erianthus* and *Miscanthus* clones appeared as two out groups in the dendrogram, sharing GS values of 0.40 and 0.57, respectively, with the two major clusters. This result supports the classification of *Erianthus* and *Miscanthus* as separate genera, but would indicate some evolutionary relationship to the *Sacharum* species (Daniels *et al.,* 1975). The *S. spontaneum* clones formed a very distinct and more diverse cluster, which is supported by previous research using isozyme analysis (Glaszmann *et al.,* 1989), RAPD markers (Nair *et al.,* 1999), SSR markers (Selvi *et al.,* 2003), comparative chloroplast genome analysis (Takahashi *et al.,* 2005),

Table 2.6. Sequence analysis of SRAP DNA fragments amplified from a *S. officinarum* (La Striped) and *S. spontaneum* (SES 147B) genotype.

Sequence code	Primer pair	Fragment source	Fragment size (bp)	GC content (%)	% homology[a]	Blastn Score (% identity)	Blastn Score (E- value)	TIGR[b] accession number	Source of accession
				Monomorphic fragments					
1	SF1+T3	*S. spontaneum*	101	45					
2	SF1+T3	*S. officinarum*	99	47	81	69	4.3	CR286450	Rice
3	SF1+T3	*S. spontaneum*	184	58					
4	SF1+T3	*S. officinarum*	173	44	76	70	0.20	TC368808	Maize
7	SF2+T3	*S. spontaneum*	140	46					
8	SF2+T3	*S. officinarum*	140	41	83	75	0.0054	TC63158	*S. officinarum*
11	SF3+T3	*S.spontaneum*	163	44	-	84	8.6e-18	TC71562	*S. officinarum*
				Polymorphic fragments					
5	SF1+T3	*S. officinarum*	145	44	-	65	0.044	CA214874	*S. officinarum*
6	SF1+T3	*S. spontaneum*	148	52	-	93	4.1e-22	CA210227	S. *officinarum*
9	SF2+T3	*S. officinarum*	100	44	-	66	1.4	TC3400008	Rice
10	SF2+T3	*S. spontaneum*	97	41	-	71	1.3	CX118790	Rice

[a] Monomorphic fragments pairs were 1,2; 3,4; and 7,8. The corresponding pair for 11 failed to amplify.

[b] For monomorphic fragments, the segment displaying the most homology was used for a Blastn search of The Institute for Genomic Research (TIGR) database.

and TRAP markers (Alwala *et al.,* 2006).*S. spontaneum* is a progenitor of modern sugarcane and it is characterized by a large intra-specific diversity in terms of morphology, species distribution, and chromosome number (Guimaraes and Sobral 1998).

The SRAP fingerprinting differentiated *S. robustum* from the other genotypes; however, a tight relationship seems to exist between some *S. officinarum* and *S. robustum* genotypes. *Saccharum officinarum* is believed to be a cultivated form of *S. robustum*, and morphological, cytological, and molecular studies have revealed considerable similarities between *S. robustum* and *S. officinarum*, in spite of differences in sugar and fiber content (Nair *et al.,* 1998; Irvine, 1999; Selvi *et al.,* 2003; Takahashi *et al.,* 2005; Alwala *et al.,* 2006). In this study, the *S. robustum* clones clustered closer to *S. officinarum* than to *S. spontaneum*. The dendrogram revealed also that there were differences within the *S. officinarum*, *S. robustum*, and *S. barberi* clones. Aitken *et al.* (2006), in an AFLP review of the *S. officinarum* germplasm, found great diversity within this species. *S. sinense* and *S. barberi* appeared together in the group, suggesting a significantly close relationship (0.93) between these species. *S. barberi* and *S. sinense* are thought to be interspecific hybrids between *S. officinarum* and *S. spontaneum* (Daniels and Roach, 1987) and this has been substantiated in sugarcane using evidence from studies based on chromosome number (Price, 1965), RFLP markers (Lu *et al.,* 1994), RAPD markers (Nair *et al.,* 1999), maize-derived microsatellite markers (Selvi *et al.,* 2003), comparative chloroplast genome analysis (Takahashi *et al.,* 2005), and TRAP markers (Alwala *et al.,* 2006).

Modern sugarcane cultivars originated from crossing the *S. officinarum* 'noble' clones with *S. spontaneum*, followed by a few backcrosses to *S. officinarum*. During this 'nobilization', the 2n somatic chromosome number of *S. officinarum* was transferred to the progeny (Bremer, 1961; Bhat and Gill, 1985; Sreenivasan *et al.,* 1987; d' Hont *et al.,* 1996). Because of this, cultivars share a greater portion of their genome with *S. officinarum* than with *S. spontaneum*.

This explains why cultivars clustered closer to and shared more unique bands with *S. officinarum* than with *S. spontaneum* in this and other studies.

Generation-wise, POJ 2878, being a less advanced cultivar, was in a different subgroup than the more modern cultivars. The latter, however, share a closer relationship with LA Purple, which is an *S. officinarum* clone used extensively in their pedigree (Deren, 1995). Furthermore, the leading cultivar in Louisiana, LCP 85-384 (CP 77-310 x CP 77-407), shared a closer relationship (0.94) with HOCP 85-845 compared to either of its parents. This relationship is supported by the fact that their grandparents were full siblings. The SRAP technique seems robust enough to describe the subtle relationship that exists among these four cultivars, some of which may have resulted from breeding and directional selection and which is generally not accounted for by pedigree data.

An appealing aspect of the SRAP system in this study is its ability to amplify species-specific markers across the *Saccharum* species, with 75% of those markers scored in th *S. spontaneum* genome. Pending an assessment of their breeding values, these markers can be useful in introgression breeding and in broadening the genetic base of sugarcane cultivars. Most of the SRAP alleles amplified in *Brassica* were evenly distributed across the genome (Li and Quiros, 2001). A similar propensity to amplify markers across the *Saccharum* genome would be a valuable addition to genetic mapping projects that particularly employ interspecific populations.

Another appealing aspect of the SRAP system is its ability to amplify exonic regions of the genome (Li and Quiros, 2001; Ferriol *et al.*, 2003a, b). All the sequenced fragments in this study showed homology with EST sequences of *S. officinarum* and several other related species [(wheat (*Triticum aestivum* L., rice and maize). Additional searches with the rice and maize EST, to benefit from the large (relative to sugarcane) bioinformatics resources available to these

crops, found high homology with known or putative protein sequences. Furthermore, the high GC content (> 41%) of the sequenced SRAP fragments in this study is indicative of their affinity to amplify exons. A GC content of over 35% from sequences of the *Arabidopsis* genome has often been associated with exonic regions (Li and Quiros, 2001).

2.5 Conclusion

This study was the first report on the utility of the SRAP marker technique to assess genetic relationships and diversity among genotypes of *Saccharum* and allied genera and should be regarded as a baseline since only a small representative number of clones were included from each species and genus. The level of polymorphism observed proved that the SRAP system was robust at amplifying markers across species and genera and did so according to the evolutionary history interconnecting members of the *Saccharum* complex. The ability to amplify species and genus-specific markers would prove to be a valuable asset during efforts to introgress useful genes and, at the same time, broaden the genetic base of modern sugarcane. The resolving power of the SRAP markers, even for the narrow genetic structure of modern cultivars, coupled with the fact that some of the amplicons could be amplifying gene-rich regions from diverse loci of the genome, is indicative of its potential usefulness for linkage and quantitative trait loci (QTL) mapping in sugarcane.

2.6 References

Aitken KS, Li JC, Jackson P, Piperidis G and McIntyre CL (2006) AFLP analysis of genetic diversity within *Saccharum officinarum* and comparison with sugarcane cultivars. *Australian Journal of Agricultural Research* 57:1167-1184.

Alwala S, Suman A, Arro JA, Veremis JC and Kimbeng CA (2006) Target region amplification polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections. *Crop Science* 46: 448-455.

Amalraj VA and Balasundaram N (2006) On the taxonomy of the members of *Saccharum* complex. *Genetic Resources and Crop Evolution* 53: 35-41.

Arceneaux G (1967) Cultivated sugarcanes of the world and their botanical derivation.

Proceedings International Society of Sugarcane Technologists 12:844-854

Arro JA, Veremis JC, Kimbeng CA and Botanga C (2006) Genetic diversity and relationships revealed by AFLP among a collection of *Saccharum spontaneum* and related species and genera. *Journal of the American Society of Sugar Cane Technologists* 26: 101-115.

Artschwager E and Brandes EW (1958) 'Sugarcane (*Saccharum officinarum*), origin, characteristics and descriptions of representative clones'. *US Dept Agriculture Handbook* no. 122 (USDA, Washington, DC).

Berding N, Roach BT (1987) Germplasm collection, maintenance, and use. In: Heinz DJ (ed.) *Sugarcane Improvement Through Breeding*. Amsterdam: Elsevier, pp. 143–210.

Besse P, Taylor G, Carroll B, Berding N, Burner D and McIntyre CL (1998) Assessing genetic diversity in a sugarcane germplasm collection using an automated AFLP analysis. *Genetica* 104: 143-153.

Bhat S and Gill SS (1985) The implication of 2n egg gametes in nobilization and breeding of sugarcane. *Euphytica* 34:377-384.

Brandes EW (1958) Origin, classification, and characteristics. In: Artschwager E and Brandes EW (eds.) *Sugarcane* (*Saccharum officinarum* L.). *US Dept of Agriculture Handbook* vol 122, pp. 1-35.

Budak H, Shearman RC, Parmaksiz, Gaussoin RE, riordan TP, Dweikat I (2004) Molecular characterization of Buffalograss germplasm using sequence related amplified polymorphism markers. *Theoretical and Applied Genetics* 108:328-334.

Burner DM (1999) Registration of Dwarf1 sugarcane genetic stock. *Crop Science* 39: 309.

Daniels J and Roach BT (1987) Taxonomy and evolution. In: Heinz D.J (ed.) *Sugarcane Improvement Through Breeding*. Amsterdam: Elsevier, pp. 143–210.

Daniels J, Smith P, Paton N, Williams CA (1975) The origin of the genus *Saccharum*. *Sugarcane Breeding Newsletter* 36:24-39.

Deren CW (1995) Genetic base of U.S. mainland sugarcane. *Crop Science* 35:1195-1199.

Dudley JW (1994) Comparison of genetic distance estimators using molecular marker data. *Joint Plant Breeding Symposia Series*. ASHS/CASSA5, 6 August 1994, pp. 3-7.

Ferriol M, Pico B and Nuez F (2003a) Genetic diversity of a germplasm collection of *Cucurbita pepo* using SRAP and AFLP markers. *Theoretical and Applied Genetics* 107:271-282.

Ferriol, M., B. Picó, and F. Nuez. (2003b). Genetic diversity of some accessions of *Cucurbita maxima* from Spain using RAPD and SRAP markers. *Genetic Resources and Crop Evolution* 50:227–238.

Ferriol M, Picó B, Córdova PF and Nuez F (2004) Molecular diversity of a germplasm collection of squash (*Cucurbita moschata*) determined by SRAP and AFLP markers. *Crop Science* 44:653-664.

Glaszmann JC, Noyer JL, Fautret A, Feldmann P and Lanaud C (1989) Biochemical genetic markers in sugarcane. *Theoretical and Applied Genetics* 789:537-543.

Guimaraes CT, Sobral WS (1998) The *Saccharum* complex: relation to other Andropogoneae. In: Janick J (ed) *Plant Breeding Reviews*, vol 16. New York: Wiley & Sons, pp. 269-287.

Irvine JE (1999) *Saccharum* species as horticultural classes. *Theoretical and Applied Genetics* 98:186–194.

Jannoo N, Grivet L, Seguin M, Paulet F, Domaingue R, Rao PS, Dookun A, d'Hont A and Glaszmann JC (1999) Molecular investigation of the genetic base of sugarcane cultivars. *Theoretical and Applied Genetics* 99:171-184.

Li G and Quiros CF (2001) Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in *Brassica*. *Theoretical and Applied Genetics* 103:455-461.

Lu YH, d'Hont A, Walker DIT and Rao PS (1994) Relationships among ancestral species of sugarcane revealed with RFLP using single copy maize nuclear probes. *Euphytica* 78:7-18.

Mukherjee SK (1957) Origin and distribution of *Saccharum*. *Botanical Gazette.* 119:55–61.

Nair NV, Nair S, Sreenivasan TV and Mohan M (1999) Analysis of genetic diversity and phylogeny in *Saccharum* and related genera using RAPD markers. *Genetic Resources and Crop Evolution* 46:73–79.

Nei M and Li WH (1979) Mathematical modelling for studying genetic variation in terms of restriction endonuclease. *Proceedings National Academy of Sciences* 76:5268-5273.

Price S (1965) Cytology of *Saccharum robustum* and related sympatric species and natural hybrids. *USDA-ARS Technical Bulletin 1337*. U.S. Government Print Office, Washington, D.C.

Price S (1963) Cytogenetics of modern sugarcane. *Economic Botany* 17:97-106.

Riaz A, Li G, Quresh Z, Swatt MS and Quiros CF (2001) Genetic diversity of oilseed *Brassica* inbred lines based on sequence-related amplified polymorphism and its relation to hybrid performance. *Plant Breeding* 120:411–415.

Roach B (1986) Evaluation and use of sugarcane germplasm. *Proceedings International Society of Sugarcane Technologists* 1:492-503.

Rohlf FJ (2000) *NTSYS-PC: numerical taxonomy and multivariate analysis system. Version 2.1*. New York: Applied Biostatistics.

SAS Institute (2004) *The SAS System for Windows Version 8.02*. Cary, NC: SAS Institue.

Selvi A, Nair N V, Noyer J L, Singh N K, Balasundaram N, Bansal K C, Koundal K R and Mohapatra T (2006). AFLP analysis of the phonetic organization and genetic diversity in the sugarcane complex. *Genetic Resources and Crop Evolution* 53: 831-842.

Selvi A, Nair NV, Balasundaram N, Mohapatra T (2003) Evaluation of maize microsatellite markers for genetic diversity analysis and fingerprinting in sugar cane *Genome* 46:394–403.

Sobral BWS, Braga DPV, LaHood ES and Keim P (1994) Phylogenetic analysis of chloroplast restriction enzyme site mutations in the Saccharinae Griseb subtribe of the Andropogoneae Dumort. Tribe. *Theoretical and Applied Genetics* 87:843-853.

Sreenivasan TV, Ahloowalia BS and Heinz DJ (1987) Cytogenetics. In: Heinz DJ (ed.) *Sugarcane Improvement Through Breeding*. Amsterdam: Elsevier, pp. 211-253.

Stevenson GC (1965) *Genetics and Breeding of Sugarcane*. London: Longmans.

Swofford DL (2002). *PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4.0b*. Sunderland, Mass: Sinauer Associates.

Takahashi S, T. Furukawa T, Asano T, Terajima Y, Shimada H, Sugimoto A, Kadowaki K (2005) Very close relationship of the chloroplast genomes among *Saccharum* species. *Theoretial and Applied Genetics* 110:1523-1529.

Vuylsteke M, Mank R, Brugmans B, Stam P and Kuiper M (2000) Further characterization of AFLP data as a tool in genetic diversity assessments among maize (*Zea mays* L.) inbred lines. *Molecular Breeding* 6: 265–276.

Weir B S (1990) *Genetic Data Analysis*. Sunderland, Mass: Sinauer Associates.

Williams CA, Harborne JB and Smith P (1974) The taxonomic significance of leaf flavonoids in *Saccharum* related genera. *Phytochemistry* 13:1141-1149.

CHAPTER 3: GENETIC DIVERSITY AMONG A LOCAL COLLECTION OF *SACCHARUM SPONTANEUM* GENOTYPES AND ITS CONTRIBUTION TO THE COMMERCIAL BREEDING GENE POOL AS REVEALED USING AFLP MARKERS

3.1 Introduction

Sugarcane (*Saccharum* spp. hybrid) is one of the crops for which interspecific hybridization has provided a major breakthrough in its improvement (Ramdoyal and Badaloo, 2002). Modern sugarcane cultivars are interspecific hybrids derived by crossing the previously cultivated *S. officinarum* with the wild relative, *S. spontaneum* (Price, 1963; Stevenson, 1965) to respond to diseases that affected sugarcane production in commercial fields. In a process coined 'nobilization', genes for stress and ratooning ability were introgressed from *S. spontaneum* into the cultivated background followed by a few backcrosses to recover the sucrose genes from the female *S. officinarum* parent (Price, 1965; Roach, 1986; Sreenivasan *et al.,* 1987).

Only a few clones were involved in the original 'nobilization' event, and modern sugarcane cultivars are mostly multi-generational descendants of the original backcross populations, thus making the genetic base of cultivated sugarcane very narrow (Arceneaux, 1967; Berding and Roach, 1987). In the US, for example, only two ancestors were found to have contributed germplasm to 90% or more of the cultivars surveyed from Louisiana (Deren, 1995). Mindful of the narrowness of the genetic base of sugarcane, the fact that the rate of genetic advance has slowed and the need to transfer disease resistance to commercial hybrids, interest in creating more diverse interspecific crosses has grown steadily since the 1960s (Berding and Roach, 1987). A basic breeding program was established in 1964 by the USDA-ARS at Houma, Louisiana with two main objectives: 1) to broaden the genetic base of cultivated sugarcane and, 2) to identify and introgress useful genes from the wild relatives into the cultivated background. A much larger collection of *Saccharum* germplasm is held at one of two world collections located in Miami, Florida.

Saccharum spontaneum continues to remain the most important source of wild germplasm for sugarcane improvement in Louisiana. *S. spontaneum* genotypes have a wide range of adaptability ranging from 8^{o} S to 40^{o} N latitude with the natural ecosystem extending from Japan and New Guinea through the Indian subcontinent to the Mediterranean and Africa (Daniels and Roach, 1987). This diversity is also exhibited by its wide polyploid range of chromosome numbers from 2n = 40 to 128, with the most frequent counts being 2n = 48, 64, 80, and 96 (Irvine, 1999). *Saccharum spontaneum* clones are considered by Louisiana breeders to be better adapted to the temperate climates of Louisiana than *S. officinarum,* which is of tropical origin (Artschwager and Brandes, 1958). Additionally, *S. spontaneum* represents an important source of genes for vigour, ratooning ability, cold tolerance, and host plant resistance to some common diseases of cultivated sugarcane in Louisiana (Dunckelman and Breaux, 1969). Indeed, resistance to mosaic virus was successfully transferred to BC progenies in cultivar x *S. spontaneum* crosses (Dunckelman and Breaux, 1972) which culminated to the commercial release of LCP 85-384 (Milligan et al., 1994) and HOCP 85-845 (Legendre et al., 1994). Recent evidence using molecular markers has shown that *S. spontaneum* could also harbor genes that contribute positively to sucrose accumulation in sugarcane (Ming et al.,1998; Reffey et al., 2005; Alwala et al., 2008).

Very few *S. spontaneum* clones have so far been successfully used in the development of sugarcane cultivars in Louisiana when compared to the number of *S. spontaneum* clones in the collection. In fact, about 85% of the cultivars presently under cultivation in Louisiana can trace their ancestry to only one *S. spontaneum* clone, US 56-15-8, from collection of about 55 *S. spontaneum* clones. The objective of this study was 1) to evaluate the genetic diversity and structure of this diversity among a local collection of *S. spontaneum* germplasm; and 2) to survey the extent to which this diversity has permeated the Louisiana commercial breeding gene pool.

This study will help answer pertinent questions relating to local germplasm enhancement efforts. For example, has most of the diversity within the collection already been tapped in the commercial breeding program? In which case, it may be necessary to explore additional sources of germplasm. Does most of the already tapped diversity come from a few clusters or is the tapped diversity uniformly dispersed across clusters? Answers to these questions will be instructive in germplasm utilization efforts with respect to strategies to employ while exploiting genotypes from this germplasm collection for cultivar development.

3.2 Material and Methods

3.2.1 Plant Material

The plant material used in this study comprised of 51 *S. spontaneum* (Table 3.1) and 66 commercial hybrid (parental) genotypes (Table 3.2). The *S. spontaneum* and commercial hybrid parental genotypes are part of the working collection of germplasm used in both basic and commercial breeding programs, respectively. This study includes S. *spontaneum* genotypes from India (16), Thailand (12), Philippines (6), Taiwan (5), China (3), Indonesia (3), Iran (2), Malaysia (2) and two genotypes of unknown origin. Genotypes selected for the study was based on availability in the collection. The commercial hybrid parental genotypes used in the study included several early hybrids that were used as progenitors of US sugarcane cultivars (Black Cheribon, POJ 2878, POJ 2725, NCO 310, Q 160 and R 570) and several generations of cultivars or advanced selections that were bred and selected in the US (Table 3.2).

3.2.2 DNA Extraction

Young leaves were collected from each genotype in greenhouse, placed immediately in ice and stored at -80^0 C. The leaves were later ground to a powder in liquid nitrogen. Total genomic DNA was extracted using the Plant DNeasy Mini Kit (Qiagen, Valencia, CA) following the manufacturer's protocol. Concentrations of extracted DNA were estimated by Nanodrop

1000 spectrophotometer (Nanodrop, Bethesda, MD) at 260 nm of UV wavelength and the DNA was stored at -20^{0}C until used.

Table 3.1: List of *S. spontaneum* genotypes that were included in the study from a local collection maintained at USDA-ARS, Sugarcane Research Unit, Houma, LA, classified according to their country of origin.

Country of origin	Genotypes*	Zone†	Region within zone†
China	GUANGXI-8605 (CHN1), GUANGXI-8721 (CHN2), GUANGXI8722 (CHN3).	East zone	II
India	COIMBATORE (IND1),IND 81161(IND2), IND 81165 (IND3),IND 8180 (IND4), IND 82275A(IND5), IND 82311(IND6), SES 006 (IND7), SES 114 (IND8), SES 147B (IND9), SES 189 (IND10), SES 205A(IND11), SES231(IND12), SES 323A(IND13), SES 8458(IND14), SH 249 (IND15), IND 81144 (IND16).	Central zone	IV
Philippines	PCANOR 842A (PHIP1), PCAV 8412A (PHIP2), PCAV 8412B (PHIP3), PCAV 8412C (PHIP4), PIN 841B (PHIP5), PQ 843 (PHIP6).	East Zone	I
Taiwan	S 66121A (TWN1), S 6684A(TWN2), S 6684B(TWN3), SPONT 24 (TWN4), TAININ (TWN5).	East Zone	II
Thailand	MPTH 97003 (THAI1), MPTH 97107 (THAI2), MPTH 97200 (THAI3), MPTH 97204 (THAI4), MPTH 97209 (THAI5), MPTH 97213 (THAI6), MPTH 97216 (THAI7), MPTH 97218 (THAI8), MPTH 97233 (THAI9), MPTH 98388 (THAI10), US 56137 (THAI11), US 56158 (THAI12).	East Zone	III
Indonesia	DIJATIROTO (INDO1), IMP 9068(IN84-21) (INDO2), IMP 9089 (IN 84-42) (INDO3)	East Zone	I
Malaysia	SES 234A (MAL1), SES 234B (MAL2)	East Zone	III
Iran	SONT 17 (IRAN1), SPONT 37(IRAN2)	Central zone	V
Unknown	MOL 1032A (UK1), MOL 1032B (UK2)	-	-

**S. spontaneum* genotypes along with their coded names that was used in the UPGMA and PCoA analysis
†Zone and regions are as classified by Panje and Babu, 1960; Tai and Miller, 2001

Table 3.2: List of the sugarcane cultivars used in the study with their parentage.

Parental Genotype†	Female parent	Male parent	Description
Black Cheribon	-	-	Old/legendary cultivar/ noble cane
Co 1148	P 4383	Co 321	Old/legendary cultivar
Co 421	POJ 2878	Co 285	Old/legendary cultivar
CP 44-154	Co 281	US 1694	Experimental clone
CP 48-103	CP 29-320	Co 290	Commercial cultivar
CP 52-68	CP 29-320	CP 38-034	Commercial cultivar
CP 57-614	CL 47-123	CP 53-017	Commercial cultivar (Rice et al., 1969)
CP 62-258	CP 53-18	CP 33-224	Experimental clone
CP 65-357	CP 52-68	CP 53-017	Commercial cultivar: Breaux et al. (1974)
CP 70-1133	CP 56-63	CP 67 poly 06	Commercial cultivar (Rice et al., 1978)
CP 70-321	CP 61-39	CP 57-614	Commercial cultivar (Fanguy et al., 1979a)
CP 70-330	CP 61-39	CP 57-614	Commercial cultivar (Fanguy et al., 1979b)
CP 72-356	CP 63-361	CP 62-258	Commercial cultivar (Breaux et al., 1981)
CP 72-370	CP 61-37	CP 52-68	Commercial cultivar (Fanguy et al., 1981)
CP 73-351	CP 65-357	L 65-69	Commercial cultivar (Breaux et al., 1982)
CP 74-383	CP 65-357	L 65-69	Commercial cultivar (Fanguy et al., 1983)
CP 76-331	CP 65-357	L 65-69	Commercial cultivar (Garrison et al., 1985)
CP 77-310	CP 52-68	L 65-69	Experimental clone
CP 77-405	CP 52-68	CP 71-424	Experimental clone
CP 77-407	CP 71-421	CP 66-315	Experimental clone
CP 79-318	CP 65-357	L 65-69	Commercial cultivar (Fanguy et al., 1989)
CP 85-830	CP 74-387	CP 77-407	Experimental clone
CP 86-916	CP 72-356	L 65-69	Experimental clone
Ho 95-988	CP 86-941	US 89-12	Commercial cultivar (Tew et al., 2005a)
HoCP 00-927	CP 89-831	LCP 85-384	Experimental clone
HoCP 00-930	CP 89-831	LCP 85-384	Experimental clone
HoCP 00-950	HoCP 93-750	HoCP 92-676	Commercial cultivar (Tew et al., 2009)
HoCP 01-544	LCP 85-384	LCP 86-454	Experimental clone
HoCP 01-553	LCP 85-384	LCP 86-454	Experimental clone
HoCP 03-741	LCP 85-384	HoCP 92-631	Experimental clone
HoCP 03-760	LCP 86-454	HoCP 92-631	Experimental clone
HoCP 85-845	CP 72-370	CP 77-403	Commercial cultivar (Legendre et al., 1994)
HoCP 89-846	CP 81-325	CP 78-304	Experimental clone
HoCP 91-552	LCP 81-10	CP 72-356	Released in 2007 as an energy cane (USDA,Houma)
HoCP 91-555	CP 83-644	LCP 82-094	Commercial cultivar (Legendre et al., 2000)
HoCP 92-631	CP 81-325	CP 71-1038	Experimental clone
HoCP 92-678	HoCP 85-845	CP 83-657	Experimental clone
HoCP 96-540	LCP 86-454	LCP 85-384	Commercial cultivar (Tew et al., 2005b)
HoCP 98-776	LCP 85-384	CP 70-1133	Experimental clone
L 00-266	HoCP 89-846	L 93-386	Experimental clone
L 01-281	LCP 86-429	LCP 85-384	Experimental clone
L 01-283	L 93-365	LCP 85-384	Commercial cultivar
L 01-292	CP 65-357	LCP 85-384	Experimental clone
L 01-296	CP 65-357	LCP 85-384	Experimental clone
L 01-299	L 93-365	LCP 85-384	Commercial cultivar
L 65-69	CP 52-1	CP 48-103	Commercial cultivar (Anzalone et al., 1974)
L 93-365	CP 78-304	CP 72-2086	Experimental clone
L 93-386	CP 79-332	L 84-290	Experimental clone
L 97-128	LCP 81-10	LCP 85-384	Commercial cultivar (Gravois et al., 2008)
L 98-209	LCP 86-454	LCP 85-384	Experimental clone
Louisiana Purple	-	-	Old/legendary cultivar/noble cane

TABLE 3.2 Continued

LCP 81-10	CP 74-328	CP 70-1133	Experimental clone
LCP 82-89	CP 52-68	CP 72-370	Commercial cultivar (Martin et al., 1992)
LCP 85-384	CP 77-310	CP 77-407	Commercial cultivar (Milligan et al., 1994)
LCP 86-429	CP 74-2013	CP 77-418	Experimental clone
LCP 86-454	CP 77-310	CP 69-380	Commercial cultivar (Martin et al., 1996)
NCo 310	Co 421	Co 312	Old/legendary cultivar
POJ 2725	POJ 2364	EK 28	Old/legendary cultivar
POJ 2878	POJ 2364	EK 28	Old/legendary cultivar
Q 160	-	-	Old/legendary cultivar
R 570	H 328560	R 397	Old/legendary cultivar
TucCP 77-42	CP 71-321	US 72-19	Bred in Louisiana and cultivated in Argentina (Mariotti et al., 1991)
US 01-39	HoCP 92-678	US 93-15	Sugarcane borer resistant clone
US 01-40	HoCP 93-775	US 93-16	Sugarcane borer resistant clone
US 93-15	CP 85-861	CP 85-834	Sugarcane borer resistant clone (White et al., 1998)
US 93-16	LCP 84-222	CP 85-843	Sugarcane borer resistant clone

† Progenitors of US sugarcane cultivars and several generations of cultivars or advanced selections that were bred and selected in the USA.

3.2.3 AFLP Marker Analysis

AFLP marker analysis was performed on the DNA of all genotypes according to the protocol of Vos et al. (1995) with some modifications. The DNA (~ 200 ng /µL) of each genotype was double-digested with *Eco*RI and *Mse*I restriction enzymes. The restricted DNA fragments were ligated to adapters specific for the *Eco*RI and *Mse*I restriction sites. A pre-selective amplification was carried out with *Eco*RI+A and *Mse*I+C primers. The resultant PCR product was then 10-fold diluted and used as template for the selective amplifications. Six IR-dye labeled (700 and 800: MWG, Germany) AFLP primer pairs (E-ACC/M-CAA, E-AAC/M-CAA, E-ACC/M-CAC, E-AAC/M-CAC, E-AGA/M-CAG, and E-AGG/M-CAG) with three selective nucleotides were used to fingerprint all the genotypes. The selective amplifications were performed with a final volume of 10 µL containing 1.5 µL of the diluted pre-amplification product, 1 µL of 1 µ*M* IR-labeled *EcoRI* primer, 1 µL of 1 µ*M MseI* primer, 0.25 µL of 5U/µLTaq DNA polymerase (Promega, Madison, WI), 1.5 µL of 2.5m*M* dNTPs, 2.0 µL of 5x PCR buffer (supplied with Taq), 1.2 µL of 25m*M* $MgCl_2$, and 1.55 µL of nano pure water. The selective amplification products were mixed with 5 µL of 5x Bromo-phenol Blue loading dye.

The mixture was denatured at 95^0C for 5 minutes and 0.75 µL was loaded on a 6.5% polyacrylamide denaturing gel in LiCor 4300 DNA Analyzer (LiCor Inc., Lincol, NE). The digital images of AFLP profiles were saved into a computer hard drive after electrophoresis. The images were manually scored as '1' for presence and '0' for absence of clear and unambiguous AFLP fragments.

3.2.4 Data Analysis

3.2.4.1 Diversity Analysis of 51 *S. spontaneum* Genotypes

The polymorphism information content (PIC) for each AFLP primer combination was computed using the following formula: $PIC=1-\Sigma f_i^2$, where f_i is the frequency of i^{th} allele (Weir, 1990). The bootstrap procedure implemented in the Dboot software (A.Coelho, personal communication) computed the minimum number of AFLP markers needed to differentiate among the genotypes with a certain level of precision using one thousand samples with replacement bootstrap iterations. The mean, variance and coefficient of variation (CV) were computed for each new marker sample. Different marker sample sizes were plotted against their corresponding CV values in a dispersion plot to identify the number of markers that would precisely estimate the genetic distances between genotypes. The AFLP binary data was also used to compute a genetic similarity (GS) matrix among the genotypes using Nei and Li's (Dice: 1979) coefficient. This coefficient disregards negative matches (0-0) in the pair-wise comparison. The similarity matrix was subjected to cluster analysis (CA) using the unweighted pair-group method with arithmetic mean (UPGMA) in NTSYS version 2.2v (Rohlf, 2000). The cophenetic correlation value was computed using the Mantel test (Mantel, 1967) to appraise the goodness of fit of a cluster analysis by comparing the original genetic similarity matrix with the cophenetic value matrix (computed from tree matrix) using COPH and MXCOMP modules in the NTSYS 2.2v. The genetic similarity matrix was also subjected to Principal Coordinate

Analysis (PCoA; otherwise known as metric-multidimensional scaling) to represent inter-individual and intergroup relationships in graphical form by using the *dcenter* and *eigen* modules in NTSYS pc 2.2v.

3.2.4.2 Model-based Bayesian Clustering (MBBC)

An allele frequency and model- based Bayesian clustering (MBBC) method was executed in the software program STRUCTURE v2.2 (Pritchard et al., 2000) to elucidate the structure of genetic variation and identify the number of genetically homogenous clusters within the local *S. spontaneum* collection using multilocus (AFLP) genotypic data. No prior information on the genotypes was used to define the clusters. The program was run with 100,000 iterations after a burn-in period of 100,000 steps for each number of genetic clusters (*K*) assuming an admixture model and correlated allele frequencies at each *K*. The *K* from 1 to 15 was chosen apriori where the genotypes are grouped into *K* clusters that have distinct allele frequencies. The final *K* (number of clusters) was decided upon comparing log probabilities (ln Pr (*X*|*K*)) of data estimated after each batch run in each of the four steps as suggested by Evanno et al. (2005). Based on the highest posterior probability value of inferred ancestry (q) in each individual (the greatest portion of one's genome), the genotypes were clustered into a gene pool (a cluster). Based on the assumption of the admixture model that any genotype could have inherited some fraction of its genome from any *K*, genetic admixture analysis was also performed on the wild *S. spontaneum* and the cultivated sugarcane genotypes to discern the level of introgression from *S. spontaneum* genotypes into the commercial hybrid parental genotypes. Also probed was the contribution of *S. spontaneum* genotypes that belong to different geographical origins.

Although Structure analysis groups genotypes based on the probability value, it does not assess the relatedness among the genotypes (Johnson et al., 2009; Lu et al, 2005). To calculate the relationships among the *S. spontaneum* genotypes in the collection, a distance matrix was

calculated using the proportion of membership values obtained for each genotype in the Structure analysis (Johnson *et al*., 2009). The distance matrix was computed using the formula, d_{ij}=1-(r_{ij}+1)/2, where r_{ij} is Pearson's correlation coefficient of the i^{th} and j^{th} genotypes based on the proportion of membership values. The distance matrix was finally inputted into the MEGA 4.0 software (Tamura et al., 2007) to obtain UPGMA clustering of the genotypes, where circular style of branching was chosen.

3.2.4.3 AMOVA Analysis

AMOVA (Analysis of Molecular Variance) was used to estimate the variance components attributable to differences among clusters and among individuals within a cluster using Nei's (1978) genetic distance. To test the significance of the variance components associated with possible levels of genetic structure, a nonparametric permutation procedure with 9999 pairwise-permutations was used. The clusters were previously defined based on the results obtained by UPGMA-CA, PCoA and MBBC clustering techniques. The AMOVA was performed in the excel-based software GenAlEx6 (Peakall and Smouse, 2006).

3.3 Results

3.3.1 AFLP Marker Diversity

A total of 51 *S. spontaneum* genotypes were fingerprinted using six AFLP primer pairs which produced distinct marker profiles with amplified fragments ranging from 56 to 569 bp (Table 3.3). Three hundred and ninety-four fingerprints were amplified ranging from 28 (E-AGG/M-CAG) to 90 (E-ACC/M-CAC) per primer pair with an average of 66. Out of 394 fragments, 305 (77.4%) were polymorphic with a range of 18 (64.2% in E-AGG/M-CAG) to 75 (83.3% in E-ACC/M-CAC) depending on the primer pair. The average number of polymorphic fragments per primer pair was 51.

The bootstrap analysis showed that the precision of the genetic similarity estimates

between genotypes increased with number of AFLP markers evaluated. The biplot (Figure 3.1) shows a relative diminishing mean CV (Coefficient of variation) value with an increasing sample size of markers. In this study, using all 394 AFLP markers reduced the CV value to 2.26. From the biplot, it is evident that a total of 83 polymorphic AFLP markers would be necessary to give a CV value of 5% (recommended value from earlier studies).

Table 3.3: AFLP-Diversity analysis of *S. spontaneum* genotypes

Primer pair	**Forward Primer**	**Reverse primer**	**Fragments amplified**	**Polymorphic bands**	**Per cent polymorphism**	**PIC**
1	E-ACC	M-CAA	89	70	78.65	0.187
2	E-AAC	M-CAA	68	50	73.5	0.182
3	E-ACC	M-CAC	90	75	83.3	0.180
4	E-AAC	M-CAC	55	39	70.9	0.163
5	E-AGA	M-CAG	64	53	82.81	0.201
6	E-AGG	M-CAG	28	18	64.28	0.188

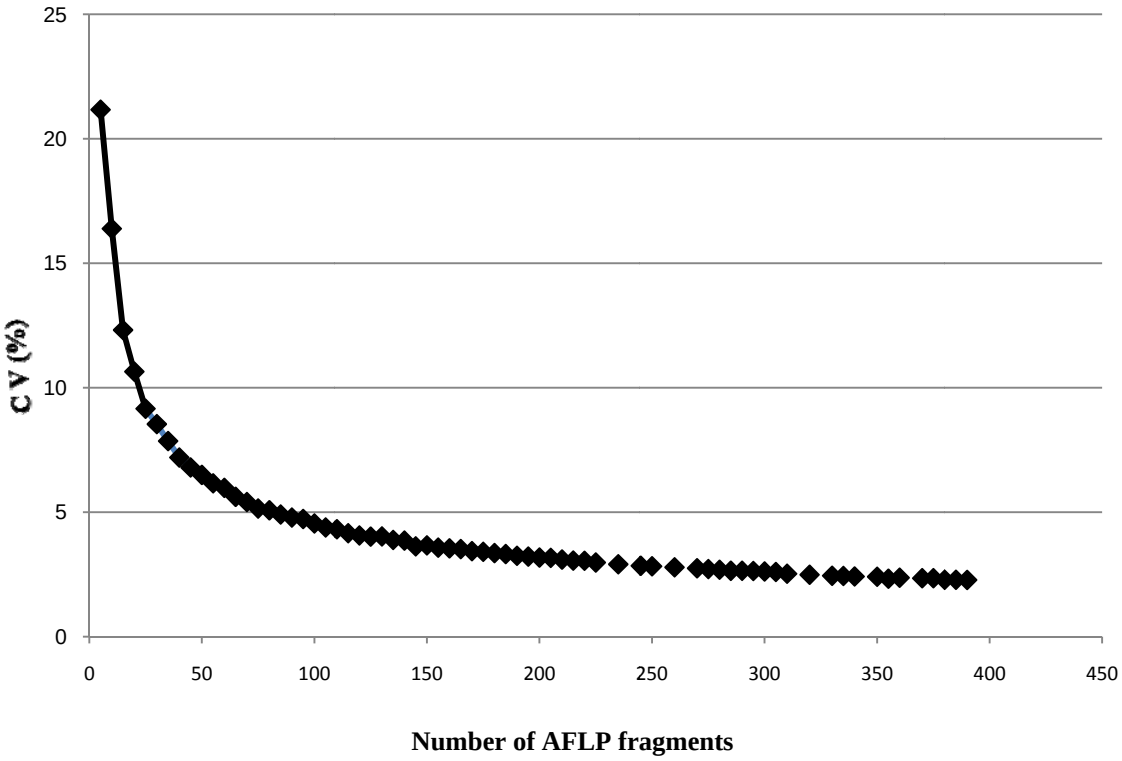


Figure 3.1: Biplot for bootstrap analysis showing the relationships between the % CV and number of AFLP polymorphisms in the study

3.3.2 Genetic Diversity and Structure of the *S. spontaneum* Genotypes

3.3.2.1 MBBC

Both model-based Bayesian clustering (MBBC) methods and distance-based multivariate exploratory (CA and PCoA) were performed to study underlying genetic relationships among the 51 *S. spontaneum* genotypes used in this study. All four steps employed (Evanno et al., 2005) on log probability values revealed four distinct genetic (*K*=4) clusters among the *S. spontaneum* genotypes. Based on the highest proportion of membership in the inferred ancestry (q), the genotypes were assigned into the respective clusters. Cluster I included mainly genotypes of Indian origin (IND- and SES-) plus a genotype from Indonesia (Djatiroto). The genotypes namely GUANGXI 8605, -8721, and -8722 of Chinese origin were assigned to Cluster II. Cluster III included genotypes mainly of Thailand (MPTH-), Philippines (PCAV-, PIN 841B, and PQ 843) and Taiwanese origin (S66-), Inonesian origin (IMP 9068, IMP 9089) and some geographically admixed genotypes (Coimbatore, IMP 9068 and -9089, SES 231 and -234B). Cluster IV contained genotypes (SES 234A, SH 249, SPONT 17, -24, and -37, Tainin, US 56-13-7, and US 56-15-8) from several geographical origins including, Malaysia, India, Iran, Taiwan, and Thailand (Table 3.1). For ease in visualizing the results, the individual relationships of the genotypes from the Structure analysis were explained by the UPGMA-dendrogram as suggested by Johnson et al., (2009 (Figure 3.2). From the dendrogram it could be seen that the genotypes in cluster II share the highest similarity and were found to be distinct, whereas the genotypes in the cluster IV were found to be the least related. The information in this dendrogram corroborated the information derived from CA as would be discussed later.

The Structure analysis identified the uppermost hierarchical level of genetic partitioning between genotypes at K=4. However, the multimodal distribution (peaks at different *K*) of the log probability values in the four steps indicated the presence of hierarchical structure in the

population (Pritchard et al., 2000). Consequently, the Structure results were further explored to test for the presence of sub-clusters in the main clusters. All the four main clusters were divided into a total of eight sub-clusters (Figure 3.3). Cluster I composed largely of genotypes originating from India, was divided into two sub-clusters (I-a, I-b) whereas genotypes in cluster III which were composed of genotypes largely from Thailand, Philippines and Taiwanese origin, were split into four sub-clusters (III-a, III-b, III-c, and III-d). The genotype 'SES 234B' originating from Malaysia, which grouped earlier in cluster III, was regrouped into I-b. The genotypes from China remained intact in cluster II and showed no sub-clustering. Most of the genotypes in cluster IV remained in that cluster except for SH 249 (India), SES 234A (Malaysia) and Tainin (Taiwan) which grouped into sub-clusters I-a, I-b, and III-c respectively. The individual relationships among the genotypes were then re-calculated based on the distance matrix using UPGMA-CA. The dendrogram from CA supported the hypothesis of eight sub-clusters in the collection at a distance value of 0.30 (Figure 3.3).

3.3.2.2 CA and PCoA

The CA identified two core clusters at a Dice similarity value of 0.82 (Figure 3.4). Cluster I included genotypes essentially from India, and a few clones from Iran (SPONT 17 and SPONT 37), Indonesia (Djatiroto), and Taiwan (SPONT 24). The majority of genotypes originating from China, Thailand, Philippines, Taiwan, Malaysia , and Indonesia grouped into cluster II. The genotype IND 81-161 was found to be distinct from the rest of the others. The cophenetic correlation (r) value of 0.82, a measure of goodness of fit for cluster analysis, fell above the threshold value of 0.80, above which clusters are considered a good fit of the data. For the purpose of comparison, the data were further analyzed using another metric namely, PCoA. The 3D-plot (Figure 3.5) explained 38.7% of the total variation among the genotypes. The grouping of genotypes in the CA and PCoA seemed to be consistent with each other.

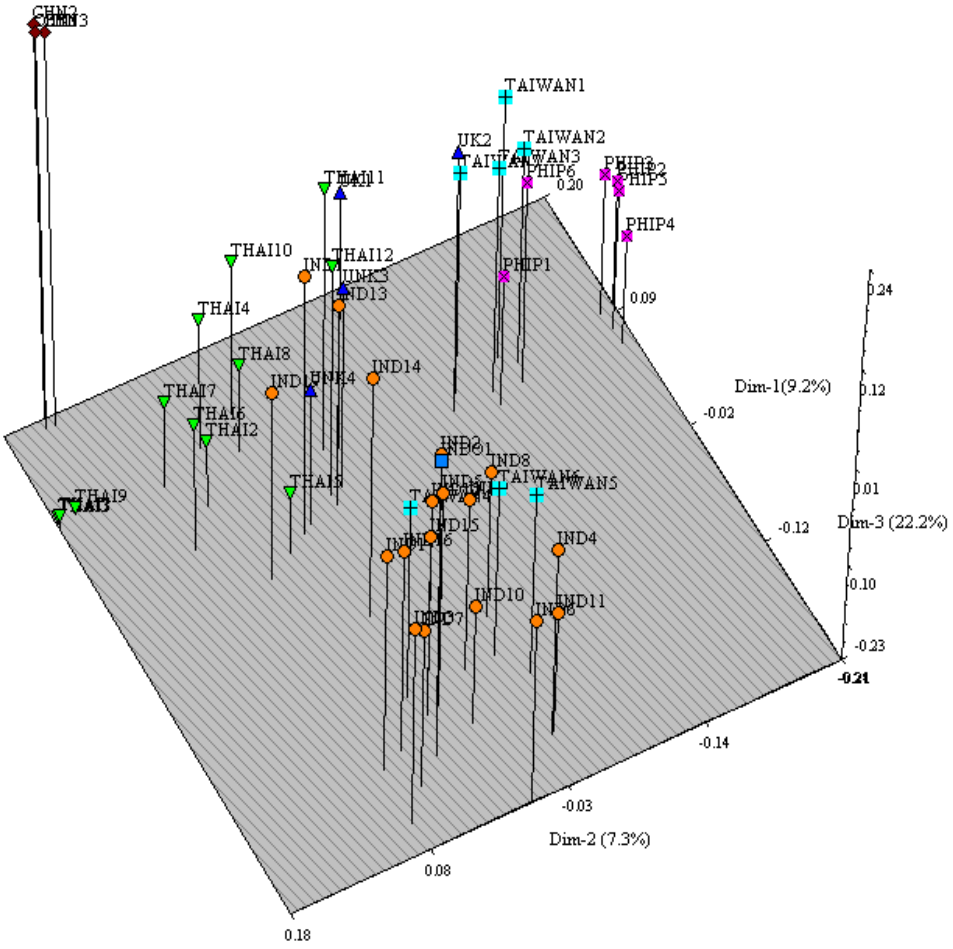


Figure 3.5: Results of Principal Coordinate Analysis on the Dice similarity coefficients between 51 *S. spontanaeum* genotypes (Table 3.1).

3.3.3 Genetic Differentiation of Clusters Using AMOVA

Based on an AMOVA analysis, it was evident that the differences among and within the main-clusters were significant (P<0.0005). The total genetic diversity found in *S. spontaneum* genotypes was mostly due to the variations within the main-clusters (83%) as compared to variations among main clusters (17%). The higher within cluster variation (83%) supports the hypothesis of sub-clusters within each main cluster as suggested by the multimodal distribution from Structure analysis (Table 3.4).

Table 3.4: Analyses of molecular variance (AMOVA) for four clusters confirmed by the clustering techniques. SS- sum of squares, MS- Mean sum of squares, Est.var- estimated variance

Source	df	SS	MS	Est. Var.	% of varaince	Prob
Among Pops	3	310.326	103.442	6.542	17%	< 0.0005
Within Pops	47	1534.027	32.639	32.639	83%	< 0.0005
Total	50	1844.353	136.081	39.180		

3.3.4 Wild Germplasm Contribution to Cultivated Sugarcane

The MBBC technique found that there were at least four major *S. spontaneum* clusters from the major geographical regions namely India, Thailand, Taiwan, Philippines, Indonesia, Malaysia, and China. The *S. spontaneum* genotypes were assigned into the clusters with the highest probability value in a cluster. The genotypes SES 234A and -234B, Coimbatore, and SH 249 were excluded from the final analysis owing to their genetic admixture with low probability values (< 0.75). The genotypes in cluster II were also excluded from the analysis because the sample size was small with only three genotypes. Forty-four *S. spontaneum* genotypes together with sixty-six sugarcane cultivars were considered for the analysis to assess the contribution of wild germplasm into cultivated sugarcane. The Structure analysis found two clusters (K=2) with the log probability values at each run using the four suggested methods (Evanno et al., 2005). One cluster contained the cultivated sugarcane genotypes, while the other cluster included the wild *S. spontaneum* genotypes originating from all geographical regions. However, the multimodal distribution of log probability values further alienated the two clusters into four sub-clusters.

The first sub-cluster included the early hybrids (old/ legendary) and the second sub-cluster included the modern sugarcane cultivars. The *S. spontaneum* genotypes grouped into two sub-clusters because some of the *S. spontaneum* genotypes that were previously assigned to the admixed Cluster IV (Figure 3.3) were reassigned into the Indian (Spont 17, SPont 24 and Spont

37) and Thailand (US 56-15-8 and US 56-13-7) clusters. The genotypes grouped into the respective clusters with more than 80% probability.

As indicated by the inferred ancestry, the old/legendary cultivars in the first sub-cluster (L 97-128, CP 77-407, HoCP 91-955, Q 160, NCO 310, POJ 2878, R 570, POJ 2725, and Black Cheribon) shared on average of 1% and 0.43% alleles from the *S. spontaneum* genotypes grouped in clusters I (India) and III (Thailand, Philippines, and Taiwan), respectively. Three cultivars in this group namely, L 97-128, CP 77-407, HoCP 91-955 can be classified as modern cultivars. The early hybrids, Q 160, NCO 310, POJ 2878, R 570, POJ 2725, and Black Cheribon did not share alleles from any *S. spontaneum* genotypes in the study. The cultivar 'CP 77-407' received the maximum number of alleles (3.1% and 9.6%) from the each of the three (cluster I and III) *S. spontaneum* clusters, respectively. The legendary cultivars shared an average of 24.6% genome with the modern sugarcane cultivars (genotypes in the second sub-cluster), with R 570 and Black Cheribon recording the minimum (0.02%) and maximum (47.3%) values, respectively.

Similarly, the modern sugarcane cultivars in the second sub-cluster shared 0.34% and 0.42% alleles from the two main clusters, namely, I (India) and III (Thailand), respectively. The cultivars LCP 81-010 (6.1%) and CP 86-916 (2.7%) were found to have inherited more alleles from the genotypes originating from cluster I. Similarly, the cluster III had maximum number of common alleles in the cultivars HoCP 01-553 (3.5%) and L 00-266 (1.9%). Relative to the *S. spontaneum* genotypes, the modern sugarcane cultivars shared a greater proportion of alleles (7.99%) with the old/legendary cultivars, where the cultivars C0 1148 (47.3%), Ho 95- 988 (37.9%), L 00-266 (33.4%), CP 85-830(29.3%), LCP 85-384 (27.55%), TucCP 77-42 (26.8%), US 93-16 (26.4%), US 93-15 (23.2%), HoCP 00-950 (22.8%), CP 70-321 (21%), CP 77-310 (20.2%), HoCP 85-845 (18%), and US 01-39 (16.3%) shared the highest number of alleles.

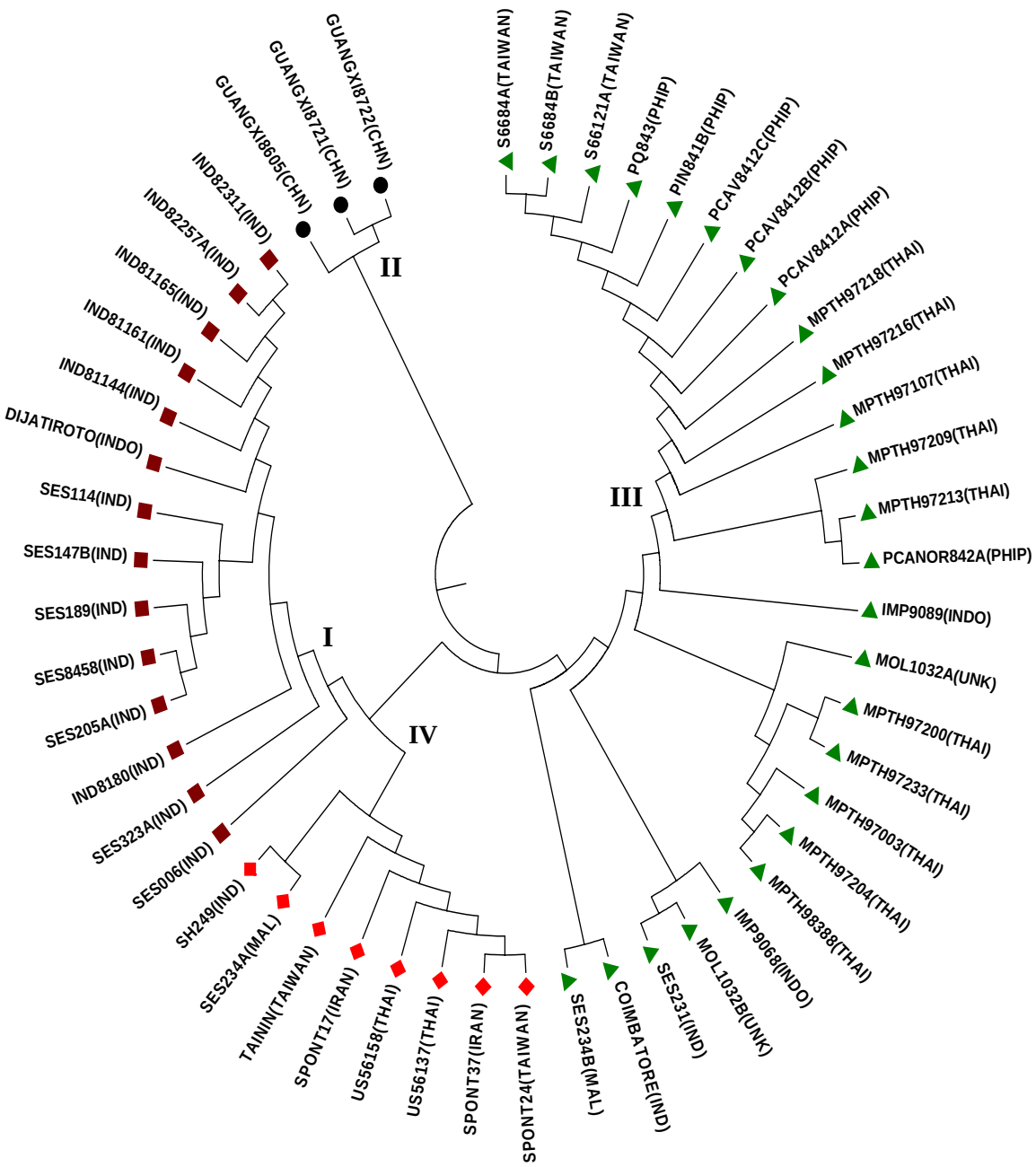


Figure 3.2: Dendrogram constructed based on the AFLP-derived distance matrix calculated using the membership values of 51 *S. spontaneum* geneotypes from structure analysis. Each cluster is indicated in different color symbol**.**

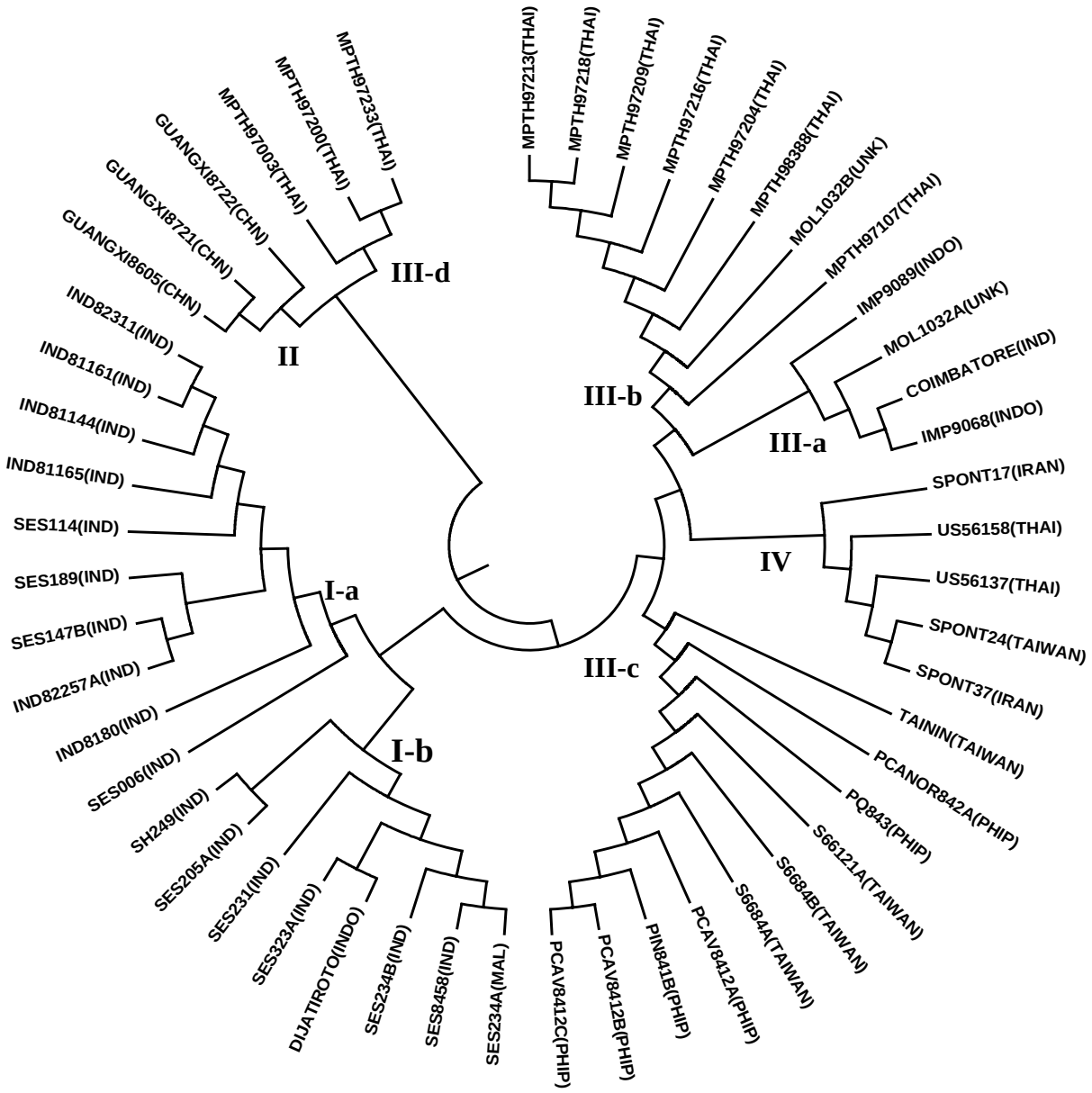


Figure 3.3: Dendrogram constructed based on the distance matrix calculated using the membership values of 51 *S. spontaneum* genotypes, which later grouped to eight sub-clusters using the AFLP markers.

3.4 Discussion

3.4.1 AFLP Polymorphism

The six AFLP primer combinations amplified 394 clearly discernable fragments, which together produced a unique profile for each of the 51 *S. spontaneum* genotypes in the collection. The level of polymorphism generated in this study was high and generally in agreement with

what has been reported when other molecular markers such as TRAP (Alwala et al., 2006), SRAP (Suman et al., 2008), SSR (Pan et al., 2003), and RAPD (Pan et al., 2004; Mary et al., 2006) which have been used to characterize the *Saccharum* germplasm.

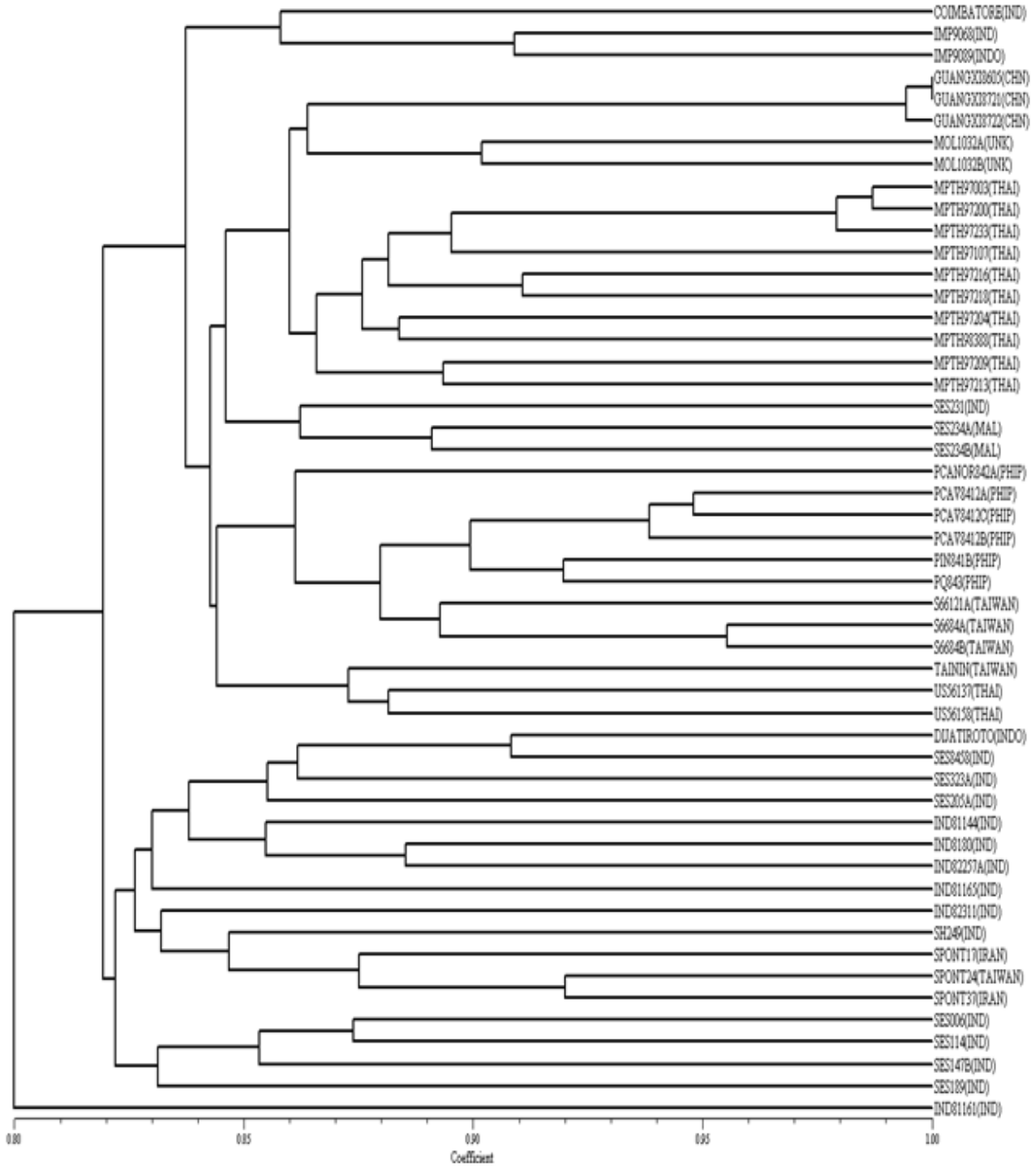


Figure 3.4: Dendrogram constructed from the UPGMA cluster analysis of 51 *S. spontaneum* genotypes (Table 3.1) based on the Dice similarity coefficients

The PIC value, a parameter indicative of the degree of informativeness of a marker, in this study ranged from 0.16 (E-AAC/M-CAC) to 0.20 (E-AGA/M-CAG) with an average value of 0.18. A dominant marker system such as AFLP can have a maximum PIC value of 0.5 because all AFLP fragments were scored as either present or absent. The PIC values in this study are comparable to previously reported values for dominant marker systems such as TRAP (Alwala et al., 2006) and SRAP (Suman et al., 2008) in sugarcane.

The AFLP markers in this study provided a high degree of reliability in differentiating among the *S. spontaneum* genotypes as indicated by the low CV value (5%) and low minimum number of polymorphic fragments (83) that would be required to detect differences among the 51 genotypes. Other marker systems required a much higher number of polymorphic markers to differentiate among over 60 sugarcane cultivars at the same level of precision (Lima et al., 2002; Arro et al., 2006) apparently because cultivars are less diverse than *S. spontaneum* and most of the diversity found among cultivars reportedly came from the *S. spontaneum* genome (Jannoo et al., 1999) because *S. officinarum* was used as the recurrent parent. .

3.4.2 Genetic Diversity among *S. spontaneum* Genotypes

S*accharum* genomes as a whole are characterized by high polyploidy, high levels of heterozygity and large genome sizes with *S. spontaneum* generally accepted as the most diverse of the *Saccharum* species in terms of geographical distribution, chromosome number (2n = 40–128) and morphology (Daniels and Roach, 1987). In the present study, however, the Dice genetic similarity coefficients ranged from 0.80 to 1.0 among the *S. spontaneum* genotypes. These values are high compared to those reported for *S. spontaneum* in other diversity studies using molecular markers (Mary et al., 2006; Arro et al., 2006; Pan et al., 2004). Since a second objective in this study was to find common fragments shared in cultivated genotypes, more emphasis was paid to polymorphic markers shared between *S. spontaneum* and cultivars such

that the full range of diversity among the *S. spontaneum* genotypes was probably not explored. However, a closer look at the dendrogram produced by CA (Figure 3.4) revealed that only a few nodes shared these high similarity values and the genotypes within these clusters were of the same geographical origin. Furthermore, these clusters were highly diverse relative to the others as depicted in the dendrogram and supported by the significant ($P<0.0005$; Table 3.4) among cluster variance from the AMOVA. The results, therefore, corroborate previous findings and suggest the presence of high levels of genetic variability among the *S. spontaneum* genotypes in the collection.

The MBBC technique divided the genotypes in this local collection *S. spontaneum* into four major clusters (K=4) largely according to the geographical origin specifically from the east and central zones (Figure 3.2; Table 3.1). Classifying *S. spontaneum* distribution on the basis of geographical proximity, Panje and Babu (1960) recognized three zones (East, Central and West) which were further divided into 7 regions (I – VII) among these zones. Genotypes from the West zone (Region VII) are not represented in the USDA, Houma working collection of *S. spontaneum* and probably need to be collected. The clusters I, II and III include genotypes that largely originated from India, China, and the geographically proximate S.E. Asian countries (Thailand, Philippines, and Taiwan), respectively. Similar clustering of these genotypes was reported by Pan et al. (2004) who genotyped a subset of the clones used in this study using RAPD markers. A fourth cluster was made up of genotypes from several geographic origins including India (SH 249), Iran (Spont 17 and Spont 37), Malaysia (SES 234A and SES 234B), Taiwan (Spont 24) and Thailand (US 56-15-8 and US 56-13-7). Based on the classification of Panje and Babu (1960) and the results of the first three clusters from this study, the genotypes from India and Iran should be assigned to cluster I while those from Malaysia, Taiwan and Thailand should be assigned to cluster III. Also in Cluster I, Djatiroto originating from

Indonesia should belong to cluster III while Coimbatore and SES 231 originating from India should belong to cluster I. Structure analysis revealed multimodal distribution (with another peak at *K* =*8*) of the log probability which is an indication that the data could be grouped into 8 clusters (Figure 3.3). We therefore reanalyzed the data with *K* = 8 clusters to determine if these discrepancies of misplaced clusters would be resolved. As expected, SH249 from cluster IV and SES 231 from cluster III were reassigned to cluster I. Also Tainin in cluster IV was reassigned to cluster III. These genotypes were, therefore, reassigned to their appropriate clusters due to the maximization of log probability values at *K*=8 (Pritchard et al., 2000). However, the genotypes Coimbatore from India and Djatiroto from Indonesia are probably mislabeled as these two genotypes remained in misplaced clusters relative to their country of origin. Furthermore, the genotypes SES 234A and -234B are known to have originated from Malaysia (source: http://www.ars-grin.gov) which according to Panje and Babu (1960) belongs to region III in the east zone. These genotypes were reclassified into the cluster I (central zone: region-IV) along with other clones from India. We suspect that these genotypes originated from India through germplasm exchange as it shares the SES nomenclature used to name other *S. spontaneum* genotypes from India. Five genotypes, namely Spont 17 and Spont 37 from Iran, Spont 24 from Taiwan and US 56-15-8 and US 56-13-7 from Thailand remained in cluster IV. Most of these individuals were grouped in cluster IV at a low probability (<0.75) level suggesting that the individuals in this cluster were geographically admixed (that is, share genomes from several geographical regions). It should be noted that the genotypes in this study could have been originally collected from locations other than those listed as country of origin and were subsequently transferred through germplasm exchange programs (Tai and Miller, 1992). Germplasm exchange is a common practice that has played a major role in sugarcane breeding programs all over the world (Tai and Miller, 1992). The disproportionate representation of

genotypes from the different geographical zones and regions could also be responsible for some of the discrepancies in clustering observed in cluster IV. It is likely some of these discrepancies could be resolved with a more proportional representation of a large number of genotypes from the different zones and regions.

India is considered as the center of origin as well as center of diversity for *S. spontaneum* (Mukherjee 1957; Roach and Daniels 1987). Many explorations have been conducted by the Sugarcane Breeding Institute, India since 1933 and germplasm collected from these trips has been distributed and been maintained separately at the Coimbatore campus, India (Sreenivasan et al. 2001; Amalraj et al., 2006) and at the USDA-ARS National Germplasm Repository, Miami, USA (Tai et al., 1995). Thus, it was not surprising that *S. spontaneum* genotypes from India were well represented in this local collection and cluster I, which comprised mostly of genotypes from India, was the most diverse as evidenced from the high percent of polymorphic loci (57%) and mean heterozysity (0.20) among genotypes. In the AMOVA, the percent variation (83%) contributed by individuals within clusters also revealed that a high level of diversity was retained within clusters except for cluster II which was made up of only a very few clones from China with high genetic similarity values (Figure 3.4).

The results obtained from the distance-based UPGMA-CA and PCoA and the MBBC methods of analysis complemented each other to provide a more meaningful analysis and interpretation of the data. The UPGMA-CA and PCoA are distance-based methods which use proportion of shared alleles between genotypes and subsequently plot the similarity coefficients in the form of a tree and a plot respectively. However, these methods are often criticized for their dependence on the distance measure chosen for the study, and arbitrary identification of clusters based on the user's judgment and without using a statistical parameter. Conversely, the Structure program utilizes a Bayesian clustering approach to probabilistically assign individuals to clusters

based on their genotypes and attempts to find the population structure in which each population is in linkage and Hardy-Weinberg equilibrium. The populations or clusters are characterized by a set of allele frequencies at each locus along with log likelihood for each run (*K*). This procedure finds the number of clusters and sub-clusters present in the population using the highest log likelihood value. Both the distance- and model-based- clustering techniques seemed to support the grouping of *S. spontaneum* genotypes in this collection according to their geographical origin. However, the Structure analysis was more useful in assigning some of the genotypes into their inferred correct clusters. For example, the genotype 'IND 81-161' originating from India was portrayed as a genetically distinctive individual in the UPGMA-CA (Figure 3.4) but Structure analysis grouped the genotype into cluster I at a high probability value. The MBBC technique is useful not only to deduce the structure of diversity among genotypes but also to assign individuals of unknown origin to discrete populations based on allelic frequencies. The MBBC successfully classified the genotypes MOL 1032A and -1032B (whose origin was unknown) into the cluster III (east zone: region-III). These genotypes grouped with Coimbatore in the distance based UPGMA CA and one would have thought they were from India.

3.4.3 Wild Germplasm Contribution to Cultivated Sugarcane

Prior to the 19th century only clones of S. *officinarum* and related species (*S. barberi* and *S. sinense*) were used for sugarcane cultivation but they all suffered from devastating effects of 'sereh' disease. To combat this disease, selections were made from among S. *officinarum* genotypes (Bandjarmasin hitam, Loethers, Black Cheribon, POJ 100, POJ 247B, EK 28, and DI 52) and these selections were found to be superior in sugar yields but not in disease resistance (Bremer, 1961). Later, the hybrid canes, such as Chunnee-cane (*S.baberi*) from India and 'Kassoer' (natural hybrid between *S. officinarum* and Java- *S. spontaneum*) from Java, were crossed to the cultivated noble canes and the disease resistant and highly productive hybrids were

developed. One such cross (POJ 2364 X EK 28) produced the Java –wonder cane (POJ 2878) and other several varieties (POJ 2725, -2727, and -2753) (Bremer, 1961). These early hybrids became the founder parents from which most sugarcane breeding programs were established. These legendarycultivars contributed a great proportion of their alleles to modern cultivars. The early derivatives of these parents are considered as legendary cultivars in this study.

The genetic admixture analysis using the Structure software clearly differentiated between the wild *S. spontaneum* and cultivated sugarcane genotypes. The analysis separated the cultivars into old/legendary and modern sugarcane cultivars and retained the original grouping of *S. spontaneum* genotypes into two major clusters (Central and East zones). Similar genotypic differentiation between wild and cultivated genotypes were reported by Pan et al. (2004), Suman et al. (2008), and Arro et al. (2006) in sugarcane.

The old/legendary cultivars did not share any alleles with the *S. spontaneum* genotypes in the working collection available in USDA-ARS, Houma, LA. This is probably because none of the actual *S. spontaneum* clones used in the original nobilization event is present in the collection. A few modern cultivars namely L 97-128, CP 77-407, and HoCP 91-555 grouped with the old/legendary cultivars. Among them, 'CP 77-407' inherited more alleles from the *S. spontaneum* genotypes than L 97-128 and HoCP 91-555. The inferred probability of assignment of 'CP 77-407' to the *S. spontaneum* clusters was 7.9% (cluster I, *S. spontaneum* from India) and 4.9% (cluster III, *S. spontaneum* from Thailand). 'CP 77-407' is a BC_3 derivative from the *S. spontaneum* genotype US 56-18-5 from Thailand and as recently as five generations ago, cultivars originating from the breeding station at Coimbatore, India were part of its parentage. Most of the original legendary cultivars were bred at Coimbatore, India. Also 'CP 77-407' is one of the grandparents of 'L 97-128' thus it was not surprising that they shared a close relationship. The inferred probability of assignment of 'L 97-128' to the *S. spontaneum* clusters

was 2 % (cluster I, *S spontaneum* from India) and 0.01 % (cluster III, *S spontaneum* from Thailand). These lowered values could be due to continuous crossing among modern cultivars that has occurred during the development of 'L 97-128'. For the third modern cultivar 'HoCP 91-555', the inferred probability of assignment to the *S. spontaneum* clusters was 0.01 % (cluster I, *S. spontaneum* from India) and 0.02 % (cluster III, *S. spontaneum* from Thailand). Compared to 'CP 77-407' and 'L 97-128', the pedigree of 'HoCP 91-555' shows no recent evidence of shared ancestry with any of the *S. spontaneum* clones in our collection but cultivars from Coimbatore were used in its parentage 6 generations ago.

The modern cultivars (CP-, LCP-, HoCP-, L-, and US-) on average shared almost an equal proportion of *S. spontaneum* alleles from the two major *S. spontaneum* clusters namely, India (0.34%) and Thailand (0.42%). There is a documented use of two *S. spontaneum* genotypes from this collection (US 56-15-8 from Thailand and SES 147B from India) in developing modern cultivars in Louisiana. The varieties 'HoCP 01-553' is a BC_5 progeny of US 56-15-8 with an inferred ancestry of 3.5% with genotypes from the Thailand cluster. However, it was surprising that LCP 81-010 was the cultivar that inherited the greatest proportion of alleles (6.1%) from the Indian cluster since there is no evidence suggesting that SES 147B was used in its parentage. The available pedigree information for LCP 81-010 suggest that it is two generations removed from a polycross with no information about the male parents used in the cross. On the other hand, SES 147B is listed as the grand parent of cultivar TucCP 77-42 yet the highest proportion of inferred ancestry (10%) was with the Thailand cluster. The reason for this discrepancy is unknown but could probably be a mix up or mislabeling of the two genotypes.

Generally, the proportion of *S. spontaneum* alleles from this collection that was traceable in the cultivars was very low. This was also true even for those cultivars that were derived from recent, documented crosses to *S. spontaneum* genotypes in the collection. Continuous crossing

among cultivars is probably responsible for this result. With aneuploidy in cultivars it may also be that some of the chromosomes amplified in the *S. spontaneum* genotypes were not present in the cultivars since the *S. spontaneum* genotypes were more diverse than the cultivars. Although AFLP markers were robust in revealing genetic diversity, they may not be the best tool for this type of analysis since they are not locus-specific and the polymorphism found on a gel may not truly represent a locus. Locus-specific microsatellite (SSRs) or RFLP markers may be more useful to estimate wild germplasm introgression by tracking allele inheritance over several generations.

3.5 Conclusions

The AFLP markers used in this study proved to be robust in revealing diversity among the *S. spontaneum* genotypes in the collection. Genotypes in the collection grouped largely according to their geographical origins. The exceptions in the clustering observed could be attributed to the admixed nature of the individual genomes, which could be due to gene flow between the genotypes in geographic proximity. In addition, the disproportionate representation of genotypes from different geographical regions within zones could be responsible for discrepancies in genotype grouping. For example, when cultivars were included in the analysis genotypes from cluster IV were reassigned into the two major clusters in agreement with what is expected based on their region of origin. Thus, a genetic diversity study should include reasonable and proportional representation of individuals from diverse origins. For the first time in sugarcane, we used the MBBC approach rather than regular distance-based clustering techniques to study the diversity and structure. The results are largely congruent with that of the distance- based clustering techniques. However, Structure analysis was more efficient in grouping the genotypes into their correct clusters (IND 81-161) and in assigning genotypes of unknown origin (MOL1032A and 1032B) with a probability value.

We assessed the percent contribution of *S. spontaneum* genotypes in both the old/legendary and modern cultivars based on the inferred ancestry of the each cultivar using Structure analysis. The percent contribution of *S. spontaneum* genotypes in the local collection was very low. The main reason attributed to this result was the fact that the original *S. spontaneum* genotypes that participated in the nobilization event were not included in this study. However, our analysis is appropriate relative to the working collection of *S. spontaneum* genotypes that are available in the working collection at the USDA, Sugarcane Research Unit, Houma, LA. The contribution of alleles from the *S. spontaneum* collection in the modern cultivars was also very low and about equal for the Central (India; cluster I) and East (Thailand; cluster III) zone with the two clusters contributing about 0.34% and 0.42%, respectively. It is evident that the commercial breeding program in Louisiana has not tapped the diversity from *S. spontaneum* genotypes in the collection except from the genotype 'US 56-15-8' and 'SES 147B'. However, we were unable to detect the relationship between the *S. spontaneum* in the collection and the ancestral *S. spontaneum* genotypes used in nobilization even when we included some old/legendary cultivars. This could be resolved in the future by using the ancestral *S. spontaneum* and *S. officinarum* genotypes. Continuous crossing and selection for sucrose content among successful cultivars likely eroded segments of the *S. spontaneum* genome that were not under selection and this could also be responsible for the low proportion of shared alleles recorded in this study. Also some of the chromosomal segments amplified in *S. spontaneum* genotypes might also be lost in modern cultivars due to aneuploidy. This study also allowed us to realize that *S. spontaneum* germplasm representing the west zone (African countries) was not present in the collection.

3.6 References

Aitken, KS, Jackson P, and McIntyre C (2006) Quantitative trait loci identified for sugar related

traits in a sugarcane (*Saccharum* spp.) cultivar × *Saccharum officinarum* population. *Theoretical and Applied Genetics* 112: 1306-1317.

Alwala S, Kimbeng CA, Veremis JC, and Gravois KA (2008) Linkage mapping and genome analysis in *Saccharum* interspecific cross using AFLP, SRAP and TRAP markers. *Euphytica* 164: 37–51.

Alwala, S., Suman A, Arro JA, Veremis JC, and Kimbeng CA (2006) Target Region Amplification Polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections. *Crop Science* 46:448–455.

Amalraj VA, Balakrishnan R, Jebadhas AW, and N. Balasundaram. (2006) Constituting a core collection of *Saccharum spontaneum* L. and comparison of three stratified random sampling procedures. *Genetic Resources and Crop Evolution* 53: 1563-1572.

Anzalone Jr. L, Paliatseas ED, Giamalva MJ, and Chilton SJP (1974) Registration of L 65-69 Sugarcane (Reg. No. 38). *Crop Science* 14: 605.

Arceneaux G (1967) Cultivated sugarcanes of the world and their botanical derivation. *Proceedings International Society of Sugarcane Technologists* 12:844-854.

Arro J A (2004) Genetic diversity among sugarcane clones using Target Region Amplification Polymorphism (TRAP) markers and pedigree relationships. Master's Thesis submitted to Louisiana State University, Baton Rouge, LA.

Arro JA, Veremis JC, Kimbeng CA, and Botanga C (2006) Genetic diversity and relationships revealed by AFLP markers among *Saccharum spontaneum* and related species and genera. *Journal of the American Society of Sugarcane Technologies*. 26: 101-115.

Artschwager E and Brandes EW (1958) 'Sugarcane (*Saccharum officinarum*), origin, characteristics and descriptions of representative clones'. *US Dept Agriculture Handbook* no. 122 (USDA, Washington, DC).

Baldoni L, Tosti N, Ricciolini C, Belaj A, Arcioni S, Pannelli G, Germana MA, Mulas M, and Porceddu A (2006) Genetic Structure of Wild and Cultivated Olives in the Central Mediterranean Basin. *Annals of Botany* 98: 935-942.

Berding N, and B.T. Roach (1987) Germplasm collection, maintenance and use. p. 143–210. *In* D.J. Heinz (ed.). *Sugarcane improvement through breeding*. Elsevier, New York.

Besse P, Taylor G, Carroll B, Berding N, Burner D, and Mc Intyre CL (1998) Assessing genetic diversity in a sugarcane germpalsm collection using an automated AFLP analysis. *Genetica* 104:143-153.

Brandes EW (1958) Origin, classification and characteristics. Pages1-3, 260-262. In: Sugarcane (S. officinarum), L.E. Artschwager and E. W. Brandes eds. USDA Agric. Handbook. 122p.

Breaux RD, Fanguy HP, and Garrison D (1982) Registration of CP 73-351 Sugarcane(Reg. No. 58). *Crop Science* 22: 688 .

Breaux RD, Fanguy HP, and James N I (1981) Registration of CP 72-356 Sugarcane (Reg. No. 56). *Crop Science* 21: 797.

Bremer G (1961) Problems in breeding and cytology of sugarcane. *Euphytica* 10:59–78.

Cai Q, Aitaken KS, Fan YH, Piperidis G, Kackson P, and McIntye CL (2005) A preliminary assessment of the genetic relationship between *Erianthus rockii* and "*Saccharum complex*" using microsatellite (SSR) and AFLP markers. *Plant Science* 169:976-984.

D'Hont A, Grivet L, Feldman P, Rao PS, Berding N, and Glaszmann JC (1996) Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Molecular and General Genetics* 250:405–413.

Daniels J and Roach BT (1987) Taxonomy and evolution. In: Heinz D.J (ed.) *Sugarcane Improvement Through Breeding*. Amsterdam: Elsevier, pp. 143–210.

Daniels J, Smith P, Paton N, and Williams CA (1975) The origin of the genus *Saccharum*. *Sugarcane Breeding Newsletter* 36: 24–39.

Deren CW (1995) Genetic base of U.S. mainland sugarcane. *Crop Science* 35:1195-1199.

Dunckelman PH and Breaux RD (1969) Agronomic characteristics of *Saccharum spontaneum* in culture in Houma, Louisiana. *International Sugar Journal* 71:333-334.

Dunckelman PH, and Breaux RD (1972) Breeding sugarcane varieties for Louisiana with new germplasm. *Proceedings of International Society of Sugar Cane Technologists* 14:233–239.

Evanno G, Regnaut S, and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* 14:2611–2620.

Fan YH, Chen H, Shi XW, Xai Q, Zhang M and Zhang YP (2001) RAPD analysis of *Saccharum spontaneum* from different ecospecific colonies in Yunnan. *Acta Botanica Yunnanica* 23:298-308.

Fanguy H P, Garrison D, and Breaux RD (1983) Registration of CP 74-383 Sugarcane. *Crop Science* 23: 1220.

Fanguy HP and Breaux RD (1981)Registration of CP 72-370 Sugarcane (Reg. No. 57) *Crop Science* 21: 798.

Fanguy HP, Dunckelman PH, and Breaux RD (1979a) Registration of 'CP 70–321' sugarcane. *Crop Science* 19:413.

Fanguy HP, Dunckelman PH, and Breaux RD (1979b) Registration of CP 70-321 Sugarcane (Reg. No. 48). *Crop Science* 19: 413.

Fanguy HP, Garrison DD, and Legendre BL (1989) Registration of 'CP 79-318' sugarcane. *Crop Science* 29(6):1574-1575.

Garrison DD, Breaux RD, and Fanguy H P (1985) Registration of CP 76-331 Sugarcane. *Crop Science* 25: 572-573.

Germplasm Resources Information Network (GRIN). http://www.ars-grin.gov.

Gravois KA and Bischoff KP (2008) New sugarcane varieties to the rescue. Louisiana Agriculture 51(2):14-16.

Gravois KA, Bischoff K., Milligan SB, Martin FA, Hoy JW, Reagan TE, Kimbeng CA, LaBorde CM, and Hawkins GL (2008) Registration of 'L 97-128' sugarcane. *Journal of Plant Registrations* 2:24–28.

Irvine JE (1999). *Saccharum* species as horticultural classes. *Theoretical and Applied Genetics* 98:186-194.

Jannoo N, Grivet L, Seguin M, Paul**e**t F, Domaingue R, Rao PS, Dookun A, D'Hont A, and Glaszmann JC (1999) Molecular investigation of the gen**et**ic base of sugarcane cultivars. *Theoretical and Applied Genetics* 99:171-184.

Johnson ES, Bekele FL, Brown BS, Song Q, Zhang D, Meinhardt LW, and Schnell RJ (2009) Population structure and genetic diversity of the Trinitario cacao (*Theobroma cacao* L.) from Trinidad and Tobago. *Crop Science* 49: 564-572.

Legendre BL, Grisham MP, White WH, Garrison DD, Dufrene EO, and Miller JD (1994) Registration of 'HoCP 85-845' sugarcane. *Crop Science* 34:820.

Legendre BL, Grisham MP, White WH, Garrison DD, Dufrene EO, and Miller JD (1994) Registration of 'HoCP 85-845' sugarcane. *Crop Science* 34:820.

Legendre BL, White WH, Grisham MP, Dufrene EO, Garrison DD, and Miller JD (2000) Registration of [HoCP 91-555] sugarcane. *Crop Science* 40:1506.

Lima MLA, Garcia AAF, Oliveira KM, Matsuoka S, Arizono H, de Souza CL Jr., and de Souza AP (2002) Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugarcane (*Saccharum* spp). *Theoretical and Applied Genetics* 104:30–38.

Lu H, Redus MA, Coburn RJ, Rutger JR, McCouch SR, and Tai TH (2005) Population structure and breeding patterns of 145 U.S. rice cultivars based on SSR marker analysis. *Crop Science* 45:66-76.

Lu YH, D'Hont A, Walker DIT, Rao PS, Feldman P and Glaszmann JC (1994) Relationships

among ancestral species of sugarcane revealed by RFLP using single-copy maize nuclear probes. *Euphytica* 78:7–18.

Mantel N (1967) Detection of disease clustering and generalized regression approach. *Cancer Research* 27:209-220.

Martin FA, Bischoff KP, Dufrene EO, Milligan SB, Quebedeaux JP, Hoy JW, Reagan TE, Giamalva MJ, Miller JD, Breaux RD and Legendre BL (1992) Registration of 'LCP 82-89' sugarcane. *Crop Science* 32: 499.

Martin FA, K.P. Bischoff, S.B. Milligan, E.O. Dufrene, K.L. Quebedeaux, J.W. Hoy, T.E. Reagan, J.D. Miller, and B.L. Legendre. 1996. Registration of 'LCP 86–454' sugarcane. *Crop Science* 36:206.

Mary S, Nair NV, Chaturvedi, Pankaj Kumar, and Selvi A (2006) Analysis of genetic diversity among *Saccharum spontaneum* L. from four geographical regions of India, using molecular markers. *Genetic Resources and Crop Evolution* 53:1221-1231.

Mariotti JA, Levi CA, Dunckelman PH, and Legendre BL (1991) Registration of 'TUCCP 77-42' sugarcane. *Crop Science* 31:492.

Miller JD and Tai PYP (1992)Use of plant introductions in sugarcane cultivar development. In: HL. Shands and LE. Weisner, Editors, *Use of Plant Introductions in Cultivar Development, Part 2*, Crop Science Society of America, Madison, WI. pp. 137–149.

Milligan SB, Martin FA, Bischoff KP, Quebedeaux JP, Dufrene EO, Quebedeaux KL, Hoy JW, Reagan TE, Legendre BL, and Miller JD (1994) Registration of 'LCP 85–384' sugarcane. *Crop Science* 34:819–820.

Ming R, Liu SC, Lin YR, Da Silva JAG, Wilson W, Braga D, van Devnze A, Wenslaff F, Wu KK, Moore PH, Burnquist W, Sorrells ME, Irvine JE, Paterson AH (1998) Detailed alignment of *Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and polyploid genomes. *Genetics* 150:1663–1682.

Mukherjee SK (1957) Origin and distribution of *Saccharum*. *Botanical Gazette* 199:55–61.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583-590.

Nei M and Li WH (1979) Mathematical modelling for studying genetic variation in terms of restriction endonuclease. *Proceedings National Academy of Sciences* 76:5268-5273.

Pan Y, Cordeiro GM, Richard Jr EP, Henry RJ (2003) Molecular genotyping of sugarcane clones with microsatellite markers. *Maydica* 48:319-329.

Pan YB, Burner DM, Legendre BL, Grisham MP, and White WH (2004) An Assessment of the genetic diversity within a collection *Saccharum spontaneum* L. with RAPD-PCR. *Genetic Resources and Crop Evolution* 51:895-903.

Panje R R, and Babu C N (1960) Studies in *Saccharum spontaneum* distribution angeographical association of chromosome numbers. *Cytologia* 25: 152-172.

Peakall R, and Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6:288-295.

Piperidis G and D'Hont A (2001) Chromosome composition analysis of various *Saccharum* interspecific hybrids by genomic *in situ* hybridization (GISH). *Proceedings of International Society of Sugar Cane Technologists* 11:565-566.

Price S (1963) Cytogenetics of modern sugarcane. *Economic Botany* 17:97-106.

Price S (1965) Cytology of *Saccharum robustum* and related sympatric species and natural hybrids. *USDA-ARS Technical Bulletin 1337*. U.S. Government Print Office, Washington, D.C.

Pritchard JK, Stephens M, and Donnell PJ (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.

Ramdoyal K and Badaloo GH (2002) Prebreeding in Sugarcane with an emphasis on the programme of Mauritius Sugar Industry Research Institute. In Managing Plant Genetic Diversity, Engels J.M.M. et al, eds . IPGRI. Rome Italy.

Reffay N, Jackson P, Aitken K, Hoarau J, D'Hont A, Besse P, and McIntyre CL (2005) Characterization of genome regions incorporated from an important wild relative into Australian sugarcane. *Molecular Breeding* 15:367–381.

Rice ER, Dunckelman P H, and. Hebert L P (1969) Registration of C.P. 57-614 Sugarcane (Reg. No. 10) *Crop Science* 9: 851.

Rice ER, Miller JD, James NI, and Dean JL (1978) Registration of 'CP 70–1133' sugarcane. *Crop Science* 18:256.

Roach B (1972) Nobilisation of sugarcane. *Proceedings of International Society of Sugar Cane Technologists* 14: 206-216.

Rohlf F (2000) NTSYS-PC numerical taxonomy and multivariate analysis system ver 2.11L. *Applied Biostatistics*, NY.

Rouf Mian MA, Hopkins AA, and Zwonitzer JC (2005) Determination of genetic diversity in tall fescue with AFLP markers. *Crop Science* 42:944-950.

SAS Institute. 2007. The SAS System, version 9.1.3 for windows. SAS Inst., Cary, NC.

Selvi A, Nair NV, Noyer J L, Singh N K, Balasundaram N, Bansal KC, Koundal KR, and Mohapatra T (2005) Genomic Constitution and Genetic Relationship among the Tropical

and Subtropical Indian Sugarcane Cultivars Revealed by AFLP**.** *Crop Science* 45: 1750-1757.

Selvi A, Nair NV, Noyer JL, Singh NK, Balasundaram N, Bansal KC, Koundal KR, and Mohapatra T (2006) AFLP analysis od the phonetic organization and gnetic diversity in the sugarcane complex, *Saccharum* and *Erianthum*. *Genetic Resources and Crop Evolution* 53:831-842.

Sreenivasan TV, Ahloowalia BS, and Heinz DJ (1987) *Cytogenetics*. p. 211–253. In D.J. Heinz (ed.) Sugarcane improvement through breeding. Elsevier, Amsterdam.

Sreenivasan TV, Amalraj VA, Jebadhas AW. 2001. Catalogue on Sugarcane Genetic Resources IV *Erianthus* species. SBI, Coimbatore, India. 98 p.

Stevenson GC (1965) *Genetics and Breeding of Sugarcane*. London: Longmans.

Suman A, Kimbeng CA, Edme` SJ, and Veremis J (2008) Sequence related amplified polymorphism (SRAP) markers for assessing genetic relationships and diversity in sugarcane germasplm collections. *Plant Genetic Resources: Characterization and Utilization* (2008), 6:222-231.

Tai PYP, Miller JD, and Legendre BL (1995) Evaluation of the world collection of *Saccharum spontaneum* L. *Proceedings of International Society of Sugar Cane Technologists* 21:250–260.

Tamura K, Dudley J, Nei M & Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24:1596-1599.

Tew TL, Burner DM, Legendre BL, White WH, Grisham MP, Dufrene EO, Garrison DD, Veremis JC, Pan YB, and Richard Jr EP (2005a) Registration of 'Ho95-988' sugarcane. *Crop Science* 45:1660–1661.

Tew TL, Dufrene EO, Garrison DD, White WH, Grisham MP, Pan YB, Richard, Jr. EP, Legendre BL, and Miller JD (2009) Registration of 'HoCP 00-950' sugarcane. *Journal of Plant Registrations* 3:42–50.

Tew TL, White WH, Legendre BL, Grisham MP, Dufrene EO, Garrison DD, Veremis JC, Pan YB, Richard Jr. EP, and Miller JD (2005b). Registration of 'HoCP96-540' sugarcane. *Crop Science* 45:785–786.

Ubi BE, Kolliker R, Fujimori M, Komatsu T (2003). Genetic diversity in diploid cultivars of rhodesgrass determined on the basis of Amplified fragment length polymorphism markers. Crop Science 43: 1516–1522.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Fornes M, Frijters A, Pot J, Peleman J, Kuiper M, and Zabeau M (1995) AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Research* 23:4407–4414.

Weir B (1990) Genetic data analysis: Methods for discrete population genetic data. Sinauer Assoc. Sunderland, MA. Nei, M., and W.H. Li. 1979. Mathematical Model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academic of Science USA* 96:5269–5273.

White W H, Burner DM, Legendre BL, and Miller JD (1998) Registration of 12 Sugarcane Germplasm Clones Resistant to Sugarcane Borer: HoCP 92-678, HoCP 93-775, and HoCP 93-776, US 93-15 to US 93-17, and US 96-1 to US 96-6. *Crop Science* 38: 1726-1727.

Zhang D, Gardini EA, Mischke S, Cernades LZ, Chavez AB, and Aguila JA (2006) Genetic diversity and structure of managed and semi-natural populations of cocoa (Theoborma cacao L.)in the Huallaga and Ucayali valleys of Peru. Ann Bot 98:647-655.

CHAPTER 4: GENETIC LINKAGE ANALYSIS OF THE SUGARCANE CULTIVAR 'LCP 85-384' USING THE SELFED PROGENY

4.1 Introduction

The sugarcane plant is a tall perennial grass, which is typically grown in tropical and sub-tropical climates for its stalks that accumulate sucrose. Sugarcane is a member of the Poaceae family like rice and Andropogoneae tribe like maize and sorghum. Up to the end of the 19th century, sugarcane varieties were mostly clones of *S. officinarum*, the species with high sugar content, domesticated from the species *S. robustum* in New Guinea (Brandes, 1958; Berding and Roach, 1987). Early in the 20th century, hybridization attempts between *S. officinarum* (2n=80) and its wild relative *S. spontaneum* (2n=40-128) in Java and India, and then backcrossing of hybrids to *S. officinarum* resulted in high sugar yields and disease resistance (Roach, 1972). However, an unequal transmission of chromosome number ('2n' from *S. officinarum* and 'n' from *S. spontaneum*) had taken place during the initial hybridization, and subsequent backcrossing events (Bremer 1961; Bhat and Gill 1985). As a result, modern sugarcane cultivars have chromosome numbers ranging between 2n = 100-130 with a strong prevalence of aneuploidy with the ancestral species, *S. officinarum* and *S. spontaneum*, contributing about 85% and 15%, respectively, of the genome (D'Hont et al., 1996) of modern sugarcane. Subsequent analysis of these ancestral species using the FISH (fluorescence *in situ* hybridization) technique has demonstrated that the basic chromosome number in *S. officinarum* is x=10 and x= in *S. spontaneum* (D'Hont et al., 1998).

Sugarcane is the source of 70% of the world's sugar, which is indeed a major export earning or commodity in some countries. In recent times, sugarcane has also received attention for its potential to produce bio-ethanol. However, despite its economic importance worldwide, the complexity of the genome limited classical genetic studies when other genetically simple

crops made remarkable gains (Barnes and Bester, 2000). The genetic complexity was due to coexistence of simplex and multiplex alleles, and irregular chromosome numbers in various homo(eo)logy groups caused by aneuploidy (Hoarau et al., 2001). The elevated ploidy levels, cytogenetic complexity of interspecific hybrids and the difficulty of controlled hybridization have further complicated genetic dissection studies (Hogarth, 1987). However, with the advent of a large number of molecular marker systems in recent times, the efficiency in developing genetic linkage maps in sugarcane has been increased and markers were eventually used in gene tagging, QTL mapping and map-based cloning (Cunff et al., 2008). The initial difficulty in mapping polyploids using molecular markers was due to the inability to identify the genotypes of marker phenotypes where a large number of genotypes for each marker phenotype are possible in a segregating population (Wu et al., 1992). Nonetheless, efforts in unraveling the sugarcane genome remain promising with the development of theoretical aspects of genetic mapping in polyploids by Wu et al. (1992) using single dose fragments (SDF).

Earlier efforts in developing linkage maps were successful in the ancestral species as well as in the commercial cultivars using the full-sib (F_1) individuals (pseudo-test cross strategy) based on RAPD, RFLP, AFLP, SSR, SRAP, TRAP, and EST-SSR markers (Da Silva et al., 1993; Al Janabi et al., 1993; Guimares et al., 1999; Atienza et al., 2002; Ming et al., 2002; Aitken et al., 2005; Raboin et al., 2006; Edme` et al., 2006; Aitken et al., 2007; Olievera et al., 2007; Alwala et al., 2008). However, few studies have been conducted using selfed populations. In France, Hoarau et al. (2001) developed a sugarcane map using the selfed progeny of the commercial hybrid R570 ($2n$ = 107-115). The map was based on 939 single dose markers distributed onto 120 co-segregating groups. The cumulative length of 5,849 cM was postulated to cover one-third of the genome. The same population had been used previously to construct a map using restriction fragment length polymorphism (RFLP) by the same group (Grivet et al.,

1996). This map consists of 408 RFLP loci on 96 co-segregation groups and 10 putative homologous groups. The mapping effort in this group led to the identification of the rust resistance gene '*Bru1*' (Daugrois *et al*., 1996; Asnaghi *et al*., 2000 and 2004), which was recently isolated by map-based cloning (Cunff et al., 2008).

LCP 85-384 is considered as one of the most successful sugarcane varieties in recent history in the Louisiana sugar industry and achieved significant monetary gains after its release in 1993. The sugar yields of LCP 85-384 were superior over the sugar yields of previously grown hybrids by about 25 per cent (Gravois and Bischoff, 2008). It was commercially successful and occupied 91% of the Louisiana sugarcane acreage (in 2004) because of its superior agronomic characters (good cane yield (tonnes of cane per hectare), ratooning ability and planting ratio), and resistance to various biotic and abiotic stresses (leaf scald (*Xanthomonas albilineans*), mosaic viral disease (Carla virus group) and post-freeze recovery). For this reason, LCP 85-384 has been frequently used as a parent in Louisiana breeding programs, although, the cultivar has recently become susceptible to the common rust (*Puccinia melanocephala*) (Gravois and Bischoff, 2008). A molecular genetic linkage map on the pseudo F_2 population of LCP 85-384 is, therefore, considered useful to understand the coexistence of genomic components derived from its parents and the genetic basis of the heterosis observed in the F_1 generation.

The objective of this study was to construct the molecular genetic linkage map of LCP 85-384 using AFLP (Amplified Length Polymoprhism), SSR (Simple Sequence Repeats), and TRAP (Target Region Amplification Polymorphism) markers based on the selfed progeny. The mapping population segregates in relation to various diseases and a number of agronomic traits. In this study, we present results concerning the development of the framework linkage map of LCP 85-384 as a first step towards the subsequent identification of QTLs for important agronomic traits.

4.2 Materials and Methods

4.2.1 Plant Materials

The S_1 progeny from selfing a sugarcane clone is considered as pseudo F_2 population. The progeny derived from self-fertilization of the modern cultivar LCP 85-384 was used to develop a molecular linkage map. LCP 85-384 was developed with the joint efforts of the Louisiana State University (LSU) AgCenter, Sugarcane Research Station, St. Gabriel, the USDA-ARS Sugarcane Research Unit, Houma, LA and the American Sugarcane League. The cultivar, LCP 85-384, was selected from the progeny of a cross between CP 77-310 x CP 77-407 (Milligan et al., 1994). More than 1000 progeny of true seed of the mapping population was germinated in flats in the glasshouse. These seedlings were transplanted to speedling trays after about three weeks and eventually to the field. The population was maintained as clones in field plots and as one-eye setts in the green house. A random sample of about 300 individual plants was selected from the population and used in the linkage mapping studies. The parents of LCP 85-384 (CP 77-310 x CP 77-407) were also included to develop grand parental maps of the pseudo F_2 population. The sampled seedlings were transplanted to the field and the agronomic data was collected in two consecutive years.

4.2.2 DNA Extraction

Young and actively growing leaves were collected from each individual plant in the green house, placed them on ice and stored in a refrigerator until DNA extraction. Genomic DNA was extracted using the Plant DNeasy Mini Kit (Qiagen, Valencia, CA) following the manufacturer's protocol with minor modifications. Concentrations of extracted DNA were estimated by the Nanodrop 1000 spectrophotometer (Nanodrop, Bethesda, MD) at 260 nm of UV wavelength and the DNA was stored at -20^0C. Quality of the DNA was checked by taking the ratio of UV wavelength at 260nm/280nm.

4.2.3 AFLP Protocol

AFLP marker analysis was performed on the DNA of all genotypes according to the protocol of Vos et al. (1995) with some modifications. The DNA (~ 200 ng /µL) of each genotype was double-digested with *Eco*RI and *Mse*I restriction enzymes. The restricted DNA fragments were ligated to adapters specific for the *Eco*RI and *Mse*I restriction sites. A pre-selective amplification was carried out with *Eco*RI+A and *Mse*I+C primers. The resultant PCR product was then 10-fold diluted and used as template for the selective amplifications. The selective amplifications were performed with three selective nucleotides in a final volume of 10 µL containing 1.5 µL of the diluted pre-amplification product, 1 µL of 1 µ*M* IR-labeled *EcoRI* primer, 1 µL of 1 µ*M MseI* primer, 0.25 µL of 5U/µLTaq DNA polymerase (Promega, Madison), 1.5 µL of 2.5m*M* dNTPs, 2.0 µL of 5x PCR buffer (supplied with Taq), 1.2 µL of 25m*M* $MgCl_2$, and 1.55 µL of nano pure water. The selective amplification products were mixed with 5 µL of 5x Bromo-phenol Blue loading dye. The mixture was denatured at 95^0C for 5 minutes and 0.75 µL was loaded on a 6.5% polyacrylamide denaturing gel in a LiCor 4300 DNA Analyzer (LiCor, Inc.). A total of 64 AFLP primer pairs were used to fingerprint all the individuals.

4.2.4 TRAP Protocol

The TRAP (Target Region Amplification Polymorphism) is a two primer based PCR technique. The design of the fixed/ forward primers used in this study was previously described in Alwala et al. (2006). The forward primers were designed using the gene/EST sequences of sucrose synthase (SuSy), soluble acid invertase (SAI), calcium dependent protein kinase (CDPK), sucrose phosphate synthase (SuPS), pyruvate orthophosphate dikinase (PODK), and starch synthase (StSy). The genes SuSy, SAI, SuPS, PODK, and StSy are associated with sucrose metabolism whereas CDPK is believed to be associated with cold tolerance. The forward

primers are listed in Table 3.1. The two reverse primers employed were IR labeled with IR dye -700 and -800. The PCR protocol used was as described by Alwala et al. (2006). A total of 12 TRAP primer pairs were used.

Table 4.1: Fixed / forward primer sequences (5'- 3') used in the TRAP PCR protocol

Primer	Gene/EST	Fixed primer sequence (5'-3')	NCBI GenBank accession number
Fixed/Forward primer	Sucrose Synthase (SuSy)	GGAGGAGCTGAGTGTTTC	AF263384
	Sucrose Phosphate Synthase (SuPS)	CGACAACTGGATCAACAG	AB001338
	Pyruvate Orthophosphate Dikinase (PODK)	CGTAAAGATTGCTGTGGA	AF194026
	Soluble Acid Invertase (SAI)	AGGACGAGACCACACTCT	AF062735
	Calcium Dependent Protein Kinase (CDPK)	ACAGAACCACCAAAGGAG	CF572977
	Starch Synthase (StSy)	GGCAAGAAGAAGTTCGAG	AF446084
Revere primer	R1	GACTGCGTACGAATTAAT	IR-700 dye
	R3	GACTGCGTACGAATTTGA	IR-800 Dye

4.2.5 SSR Protocol

The 19 SSR primers used in the study were obtained from the Sugarcane Microsatellite Consortium (Cordeiro et al., 2000) (Table 4.2). A robot was used to prepare 384-well PCR amplification reaction plates with each well containing a 5-µl PCR reaction mixture. The mixture consisted of 0.25 µl of DNA sample, 0.5 µl of 10X Buffer, 0.3 µl of 25 mM $MgCl_2$, 0.1 µl of 10 mM dNTPs, 0.41 µl each of 3 pM/µl forward and reverse primers, 0.5 µl of 10 mg/ml BSA-V, 0.5 µl of 100 mg/ml PVP-40, 0.025 µl of 5U/µl *Taq*, and 2.0 µl of PCR water. PCR amplification reactions were conducted on a DNA Engine Tetra equipped with four 384-well Alpha blocks with heated lids (Bio-Rad Laboratories, Hercules, CA) under the program of 95ºC for 15 min, 40 cycles of (94ºC for 15 sec, annealing for 15 sec, and 72ºC for 1 min), with a final extension at 72ºC for 10 min, and holding at 4ºC. The annealing temperature varies with SSR markers and is shown in Table 4.2. The robot was used again to prepare 384-well CE sample

plates by first diluting the amplified SSR DNA fragments and then dispensing in each well 1 µl of the diluted products and 9 µl Hi-Dye formamide solution premixed with the Rox^{TM} 500 size standards following the manufacturer's instructions (Applied Biosystems, Inc., Foster City, CA) (Pan et al., 2003).

4.2.6 Marker Scoring

The PCR fragments amplified by the AFLP and TRAP techniques were run on a LiCor 4300 DNA analyzer (LI-COR Inc., Lincoln, NE) while the SSR amplified fragments were run on an ABI3730 Genetic Analyzer (Applied Biosystems, Inc., Foster City, CA). The digital images of marker profiles were saved onto a computer hard drive after electrophoresis. The images from the AFLP and TRAP amplified fragments were manually scored as '1' for presence and '0' for absence of clear and unambiguous fragments. Individual GeneScan files from ABI3730 Genetic Analyzer were analyzed manually with the GeneMapper™ software (Applied Biosystems, Inc., Foster City, CA) (Pan et al., 2006). Presence of an SSR allele was given a score of "A" while its absence a score of "C" and then these scores were converted into binary scores like in other marker systems. AFLP markers were denoted by 'EM' for *ECo*RI – *Mse*I primer pairs with band size as suffix using the universal nomenclature according to Vuylsteke et al. (1999) where the numbers followed by each letter are codes for a primer pair. TRAP markers were denoted by the codes for forward and reverse primers as mentioned in Table 4.1 along with marker size as suffix. SSR markers were indicated with its name and identity number from the Sugarcane Microsatellite Consortium along with the allele size as suffix (Table 4.2).

4.2.7 Segregation Analyses

Both monomorphic and polymorphic fragments were produced by all three marker systems. Several segregation ratios are possible in the pseudo F_2 population. For example, in the absence of segregation distortion and in the presence of disomic inheritance, single dose

Table 4.2: International Sugarcane Microsatellite Consortium microsatellite primers with high polymorphism information content (PIC) values (Pan, 2006).

Name	SSR Repeat	# of fragments[1]	Size Range (bp)	PIC	°C	Forward Primer Sequence (5' to 3')	Reverse Primer Sequence (5' to 3')
mSSCIR3	$(GT)_{28}$	7 (6)	160-190	0.80	60	ATA GCT CCC ACA CCA AAT GC	GGA CTA CTC CAC AAT GAT GC
mSSCIR19	$(GA)_{23}$	8 (7)	130-160	0.80	48	GGT TCC AAA ATA CAC AAA	CAA TCT TAT CTA CGC ACT T
mSSCIR29	$(GA)_{26}$	6 (4)	110-130	0.72	48	AAT GGA AGG AGT TTT TGA	CTG CTT TCT GTG AGT GTG
mSSCIR66	$(GT)_{43}GC(GT)_6$	5 (4)	120-150	0.72	48	AGG TGA TTT AGC AGC ATA	CAC AAA TAA ACC CAA TGA
mSSCIR74	$(CGC)_9$	4 (2)	210-230	0.64	54	GCG CAA GCC ACA CTG AGA	ACG CAA CGC AAA ACA ACG
SMC119CG	$(TTG)_{12}$	6 (4)	100-160	0.72	58	TTC ATC TCT AGC CTA CCC CAA	AGC AGC CAT TTA CCC AGG A
SMC1604SA	$(TGC)_7$	5 (4)	100-130	0.80	58	AGG GAA AAG GTA GCC TTG G	TTC CAA CAG ACT TGG GTG G
SMC1751CL	$(TGC)_7$	4 (3)	140-160	0.72	60	GCC ATG CCC ATG CTA AAG AT	ACG TTG GTC CCG GAA CCG
SMC18SA	$(CGA)_{10}$	4 (2)	140-150	0.64	62	ATT CGG CTC GAC CTC GGG AT	AGT CGA AAG GTA GCG TGG TGT TAC
SMC278CS	$(TG)_{19}(AG)_{25}$	7 (6)	140-190	0.80	64	TTC TAG TGC CAA TCC ATC TCA GA	CAT GCC AAC TTC CAA ACA GAC T
SMC31CUQ	$(TC)_{10}(AC)_{22}$	7 (6)	130-180	0.80	62	CAT GCC AAC TTC CAA TAC AGA CT	AGT GCC AAT CCA TCT CAG AGA
SMC334BS	$(TG)_{36}$	5 (4)	140-170	0.72	60	CAA TTC TGA CCG TGC AAA GAT	CGA TGA GCT TGA TTG CGA ATG
SMC336BS	$(TG)_{23}(AG)_{19}$	7 (6)	140-190	0.80	62	ATT CTA GTG CCA ATC CAT CTC A	CAT GCC AAC TTC CAA ACA GAC
SMC36BUQ	$(TTG)_7$	3 (2)	110-120	0.56	64	GGG TTT CAT CTC TAG CCT ACC	TCA GTA GCA GAG TCA GAC GCT T
SMC486CG	$(CA)_{34}$	4 (4)	220-250	0.80	58	GAA ATT GCC TCC CAG GAT TA	CCA ACT TGA GAA TTG AGA TTC G
SMC569CS	$(TG)_{37}$	3 (2)	160-230	0.64	62	GCG ATG GTT CCT ATG CAA CTT	TTC GTG GCT GAG ATT CAC ACT A
SMC597CS	$(AG)_{31}$	8 (7)	140-170	0.80	64	GCA CAC CAC TCG AAT AAC GGA T	AGT ATA TCG TCC CTG GCA TTC A
SMC703BS	$(CA)_{12}$	7 (4)	200-220	0.72	62	GCC TTT CTC CAA ACC AAT TAG T	GTT GTT TAT GGA ATG GTG AGG A
SMC7CUQ	$(CA)_{10}(C)_4$	3 (2)	160-170	0.64	60	GCC AAA GCA AGG GTC ACT AGA	AGC TCT ATC AGT TGA AAC CGA

Name, SSR repeat, and the primer sequences were provided by Giovanni Cordeiro.
[1]Total of number (polymorphic) bands.

(simplex), double dose (duplex), and triple dose (triplex) markers segregate in 3:1, 15:1 and 63:1 ratios, respectively. Except for single dose markers, these ratios would become complicated in the presence of polysomic inheritance. Therefore, only simplex markers were considered for the linkage mapping because they are the most informative type (Wu et al., 1992; Grivet et al., 1996). The mapping population size in this study was confirmed to be large enough to differentiate the markers segregating in 3:1 from all the non-simplex markers (Bailey, 1961). However, we retained all the markers that have +/- ratio lower than 6.7:1 (√3 x 15:1) to ensure the selection of only simplex markers (Mather, 1957). This was accomplished by choosing the smallest non-simplex marker ratio (15:1) to cull out all non-simplex markers that were segregating in higher ratios. The markers selected from this analysis were again tested for the expected 3:1 theoretical Mendelian ratio using the χ^2 test (1*df*) at 5% error level (Type I). A Bonferroni correction was applied to limit the experiment-wide error rate associated with multiple testing (Sokal and Rohlf, 1995). The critical χ^2 values were calculated by dividing the alpha (0.05) by the number of markers. Markers which deviated from the theoretical expected ratio (3:1) even after the Bonferroni correction (α=6.47 x10^{-5}) were considered as distorted and marked with an asterisk.

4.2.8 Linkage Map Construction

Mapping of simplex markers onto co-segregation groups (Grivet et al., 1996) was implemented using the software JoinMap ver 3.0 (van Ooijen and Voorrips, 2001). Linkages in coupling phase were detected using an upper recombination fraction threshold of 0.44 (max*r*= {0.5 – *z* (α) √ 0.5 (1-0.5)/n}, where *z* (0.01) = 2.3264, and n=300) (Wu et al., 1992). Only coupling phase linkages were detected and included on the linkage map because all the coded dominant markers designate the same phase (D'Hont et al., 1994). The non-simplex markers both in coupling and repulsion phases were ignored irrespective of their inheritance pattern

(Grivet et al., 1996). In order to avoid false linkages, multiple two-point linkage analyses were performed at LOD score ≥ 4.0. Cosegregation groups (CGs), which correspond to a single chromosome among all the homo(eo)logous chromosomes, were identified by grouping the from linked markers. The genetic distances (in cM) between the markers on the map were computed recombination fractions using the Kosambi mapping function.

The CGs were assembled into homo(eo)logous groups (HGs) based on the alleles generated by the same SSR primer pair found on different CGs and also based on CG information from the parental maps. In addition to the S_1 map of LCP 85-384, maps were also constructed for each of the grandparents (female and male parent of LCP 85-384) from which the S_1 population was derived. The markers used to construct each parental map was based on an '*ao* x *oo*' (CP 77-310) and '*oo* x *ao*' (CP 77-407) configuration (that is, present in one parent and absent in the other) which segregated in a 3:1 fashion in the S_1 population. We also used markers which were present in both the parents, F_1, and segregating in a 3:1 ratio in the S_1 population. In using the parental maps to form homologous groups, markers belonging to a single parental CG and distributed in more than one CGs of the S_1 map allowed us to consider the S_1 CGs as part of a homo(eo)logous group.

Repulsion phase linkages between markers were tested to investigate pairing behavior (chromosome assortment) between CGs as described by Grivet et al. (1996) and Hoarau et al. (2001). Each of the three maps was tested separately. To test for repulsion phase linkages marker scores were converted (0→1 and 1 → 0). The original data was doubled by appending the converted scores to the original set of markers. Linkages between the original/original and original/converted markers were assessed by reconstructing the maps using the recombination threshold of 0.44 and LOD ≥ 4.0. This step would yield the markers in all possible pair-wise combinations. Linkages are said to be in coupling if the pairing occurs between the original/

original (or inverted /inverted) and in repulsion if they occur between the original/ inverted markers. In each CG, an observed ratio of repulsion to coupling linkages is tested against the theoretical ratio of 1:1 (for disomic pairing) and 0:1 (for random paring) using χ^2 value at (1 *df*) after Bonferroni correction (0.05/ n, where n=number of linkage groups) (Wu et al., 1992; Kreigner et al., 2003).

The JOINMAP 3.0 software is developed for diploid linkage mapping but can be extended for polyploid linkage mapping like sugarcane. Only coupling-coupling (C-C) and repulsion-repulsion (R-R) phase linkages are possible in a selfed (S_1) population. But in an F_1 population, all possible linkages are possible (C-C, R-R, C-R and R-C) (Kreigner et al., 2003) because any two loci can occur in two different (coupling and/or repulsion) phases in the two parents. For JOINMAP 3.0, the initial coding of d+/b- has to be re-coded as a/c (d → a and b → c : or vice-versa if the original coding was a/c). Two (b/d or a/c) coded markers are assumed to be in coupling phase, but an a/c (or b/d) coded marker is assumed to be in repulsion with any b/d (or a/c) coded marker. After appending the recoded matrix to the original data set, linkage mapping has to be performed again at the chosen values of recombination frequency and LOD. For each linkage group, this step computes the recombination frequency estimates between all possible pair-wise combinations of original/original (O/O), inverted/ inverted (I/I), original/inverted (O/I) and inverted/ original (I/O) markers. But recombination estimates between O/I of the same markers have to be ignored since the population type is selfed (S_1). The recombination frequency estimates between markers along with corresponding LOD values can be observed under the *strong linkages* tab. For markers to be in significant C-C phase linkage, pairs of O/O and I/I markers must have significant recombination estimates ($r < 0.5$). In contrast, pairs of I/O and O/I markers are in repulsion phase linkages if they have significant recombination frequency estimates ($r < 0.5$). In general recombination frequencies are non-

significant ($r > 0.5$) for coupling-coupling I/O (and O/I) and repulsion-repulsion O/O (and I/I) phase linkages. After this tedious procedure, C-C and R-R phase linkages are counted for each linkage group to test for the chromosome pairing during meiosis. The ratio 1:1 confirms (χ^2 at 0.05) allopolyploidy, whereas deviation (from 1:1) confirms either autopolyploidy or partial auto-allo polyploidy (Johan Van Ooijen, personal communication).

4.3 Results

4.3.1 Segregation Analyses

A total of 1113 polymorphic markers were produced from genotyping 300 S_1 progeny of the cultivar LCP 85-384 with 64 AFLP, 12 TRAP, and 19 SSR primer pairs. The number of polymorphic markers scored per primer pair ranged from 1 to 32 with a mean of 12. The Mather's criterion (Mather 1957) excluded 338 likely multiple dose markers and retained a total of 773 (69.45%) provisional single dose markers. The 773 single dose markers amplified ranged from 0 to 22 per primer pair with a mean of ~ 8. The 773 markers included 650 AFLP (84%), 94 TRAP (12.1%), and 29 SSR markers (3.75%). Of the 773markers, 224 (29%) markers did not fit the theoretical single dose markers ratio (3:1) using the chi-square test at 5% level (Type I error) (Table 4.3). However, only 32 of 224 markers deviated from theoretical expectations after the Bonferroni procedure ($\alpha=6.47 \times 10^{-5}$). The distribution of simplex markers was skewed toward the lowest ratios (less than 6.7), which implies that the majority (~70%) of the markers could be considered to be single dose markers (Fig 1).

The information on number of markers inherited by LCP 85-384 from its parents was limited and available from 60 AFLP primer pairs. The 773 markers were divided according to their parental origin. Markers that did not have information on parental origin were discarded from the analysis. The female parent (CP 77-310) contributed 210 AFLP markers with a mean of 3.5 markers per primer pair, whereas the male parent (CP 77-407) contributed 167 AFLP

markers with a mean of 2.78. However, the AFLP markers found in both parents ('*ao x ao*') which segregated in a 3:1 ratio in the S_1 were 230 with a mean of 3.84 markers per primer pair. The 230 markers were appended to both of the parent-specific markers, which yielded a total of 440 (210+230) for the female and 397 (167+230) markers for male parent. The number of distorted markers recorded at 5% level of significance in CP 77-310 and CP 77-407 were 127 (29%) and 130 (32.7%), respectively. After the Bonferroni procedure, the number of distorted markers in CP 77-310 was reduced from 127 to 19, whereas the reduction in CP 77-407 was from 130 to 18 (Table 4.4).

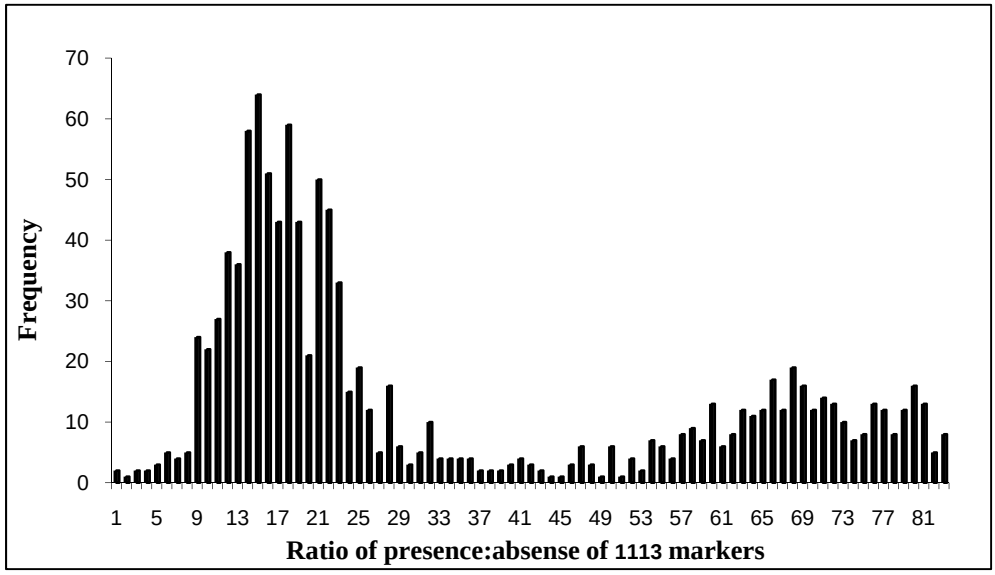


Figure 4.1: Frequency distribution showing the segregation ratio of presence: absence of all 1113 markers.

4.3.2 Linkage Map Construction and Segregation Distortion

4.3.2.1 Map of LCP 85-384

Out of the 773 simplex markers, 717 markers were assigned to 108 CGs (Figure 2) with a cumulative genome length of 5384 cM. Fifty six markers remained unlinked. The length of the CGs varied from 4cM (CG-102, and CG-68) to 147cM (CG-39) with an average of 7.5 cM between any two adjacent markers. The total number of markers per CG varied from 2 to 21. We

also counted the number of CGs containing markers from each primer pair. The range was from 9 to 19 CGs per primer pair. The marker from the primer pairs E36M61 (AFLP), CDPK_R3 (TRAP), and 18SA and CIR29 (SSR) covered the most number of CGs (Table 4.3). Although some of the CGs (CG-1, CG-4, CG-20, CG-39, and CG-75) were dense, the marker loci were not well distributed across CGs and gaps still remained with clustering of markers. For the recombination threshold (0.44) and the mapping function (Kosambi) used in this map, the theoretical maximum distance between any two adjacent markers can be 73.6 CM. However, no interval between two adjacent markers was observed to be greater than 43 cM (CG-97). Of the 32 distorted markers, 19 were scattered on 11 CGs and 13 remained unlinked. Extensive clustering of distorted markers was not observed on the CGs. The cumulative number of significant calls (skewed chromosomal regions) expected at the level of 0.05 after the Bonferroni procedure is 24, which coincided with the value observed (11+13=24) in the present linkage mapping study.

Table 4.3: Summary of AFLP, TRAP, and SSR polymorphic markers used for constructing the linkage map on a selfed progeny of LCP 85-384.

Primer pair	**Number of markers scored**	**Mather's[†] criterion (6.7:1)**	**Co-segregation groups covered**
64 AFLP markers			
E32M47	21	10	8,19,37,43,44,48,49,61,75,88 (10)
E32M48	16	14	2,3,15,20,34,36,37,46,54,55,63,83,93 (13)
E32M49	28	16	5,16,18,27,28,34,42,48,49,53,77,80,82,84 (14)
E32M50	20	13	2,17,18,51,67,72,73,75,78,90,96 (11)
E32M59	15	10	6,13,27,31,36,43,44,92,100 (9)
E32M60	21	11	1,6,10,17,22,49,62,65,96 (9)
E32M61	27	18	4,5,12,13,23,27,33,36,45,48,54,62,70,71, 76, 88,91(17)
E32M62	10	3	16,73,88 (3)
E33M47	9	6	21,40,41,49,50 (5)
E33M48	8	8	1,5,19,44,49,66,81,90 (8)
E33M49	12	11	12,16,17,31,39,49,75,89,103 (9)
E33M50	15	10	14,17,21,25,56,57,94,96,101,104 (10)
E33M59	20	15	1,7,16,20,22,24,35,44,45,47,76,78,83,89 (14)
E33M60	21	15	1,17,45,54,56,65,74,75,82,99,101 (11)
E33M61	18	15	10,13,14,50,52,53,61,74,78,84,87,88,89 (13)

TABLE 4.3 continued

E33M62	21	14	5,10,16,31,39,43,44,46,51,69,70,83,88 (13)
E36M47	26	11	4,7,9,12,18,55,55,75,85,100,104 (10)
E36M48	14	11	27,29,44,48,54,74,85 (7)
E36M49	31	20	1,2,9,17,20,21,23,39,51,60,67,75,78,93,97,99,108 (17)
E36M50	9	7	43,44,46,75,90,91,105 (7)
E36M59	32	18	1,4,8,9,10,12,17,27,31,45,47,50,53,79,107 (15)
E36M60	26	13	1,12,15,31,34,35,39,46,75,90,91 (11)
E36M61	28	22	1,4,7,20,22,23,34,43,50,56,71,73,76,78,79,83,85,88,96 (19)
E36M62	17	11	7,20,43,63,75,83,85,87,105 (9)
E37M47	15	11	11,20,34,35,43,49,55,57 (8)
E37M48	14	6	33,47,57,86,106,107 (6)
E37M49	15	14	5,24,25,31,42,44,47,70,76,78 (10)
E37M50	12	10	5,33,43,47,50,51,74,75,89,98,101 (11)
E37M59	14	10	2,20,31,37,39,52,78,98 (8)
E37M60	11	7	27,39,43,48,70,78 (6)
E37M61	15	12	1,16,18,25,26,27,28,35,46,54,74,85,89 (13)
E37M62	8	6	2,22,40,71,77,90 (6)
E38M47	7	5	4,12,24,35,53 (5)
E38M48	17	14	17,23,28,30,31,37,39,43,55,80,82,86 (12)
E38M49	14	10	8,23,37,49,68,70,76,83 (8)
E38M50	12	10	2,10,11,62,69,75,78,80,86, 97 (10)
E38M59	15	14	3,4,12,20,36,46,54,56,83,84,97,102 (12)
E38M60	10	10	20,21,25,39,52,78,83,94 (8)
E38M61	11	11	1,24,34,36,40,44,55,73,89,93 (10)
E38M62	8	8	11,17,28,47,73,75 (6)
E39M47	10	6	1,39,61,81,88,102 (6)
E39M48	20	14	10,11,20,21,25,26,30,39,49,58,80,81 (12)
E39M49	23	13	7,16,18,23,34,50,51,52,53,80,82,85 (12)
E39M50	11	8	4,35,37,70,92,94 (6)
E39M59	6	6	12,30,45,59,91,93 (6)
E39M60	23	13	1,4,6,15,22,23,32,47,70,76,84,98 (12)
E39M61	21	13	4,10,20,21,46,57,58,63,81,83,83,95 (11)
E39M62	9	4	15,37,52,77 (4)
E40M47	6	5	20,24,51,79,80 (5)
E40M48	10	7	2,43,68,77,92 (6)
E40M49	13	10	2,4,6,16,27,31,44,60,91,108 (10)
E40M50	8	5	2,40,81,85,87 (5)
E40M59	18	11	1,3,39,60,64,77,84 (7)
E40M60	13	8	9,18,34,39,41,65,80,90 (8)
E40M61	9	7	44,45,58,81,91 (5)
E40M62	14	9	1,12,16,17,18,23,50,86,103 (9)
E41M47	6	3	22,78,80 (3)
E41M48	10	8	1,20,37,51,57,74,84 (7)
E41M49	6	4	6,52 (2)
E41M50	8	6	42,43,59,76,85,106 (6)
E41M59	1	1	14 (1)

TABLE 4.3 continued

E41M60	17	9	4,7,14,31,32,42,64,78 (8)
E41M61	21	16	2,11,12,15,16,17,23,42,47,80,81,83,84 (13)
E41M62	6	4	1,8,15,38 (4)
Sub-total	**952**	**650**	
Mean	**14.87**	**10.15**	
Range	**1- 32**	**1-22**	
12 TRAP markers			
SuSy_R1	9	9	2,42,43,45,58,71,72,85,98 (9)
SuSy_R3	6	5	21,24,53,71 (4)
SuPS_R1	15	15	4,37,38,45,48,81,92 (7)
SuPS_R3	6	6	2,27,31,44,90 (5)
PODK_R1	8	7	4,10,15,37,42,45,81 (7)
PODK_R3	8	7	16,20,45,52,82,87 (6)
SAI_R1	7	6	31,45,59,75,86,100 (6)
SAI_R3	4	4	6,8,47 (3)
CDPK_R1	10	8	1,16,20,46,56,66,84 (7)
CDPK_R3	12	12	1,8,21,24,37,39,48,81,84,99 (10)
StSy_R1	5	5	12,20,33,73,83 (5)
StSy_R3	10	10	24,37,39,55,74,75,83,85,95 (9)
Sub-total	**100**	**94**	
Mean	**8.34**	**7.84**	
Range	**4-15**	**4-15**	
19 SSR markers			
119CG	2	1	70 (1)
1604SA	4	2	4,5 (2)
1751CL	4	2	44,49 (2)
18SA	4	3	12,23,40 (3)
278CS	3	2	10,11 (2)
31CUQ	3	2	10,11 (2)
334BS	2	1	13 (1)
336BS	3	2	10,11 (2)
36BUQ	2	1	20 (1)
486CG	3	1	0
569CS	3	3	49 (1)
597CS	4	1	64 (1)
7CUQ	3	1	73 (1)
CIR19	4	2	70,97 (2)
CIR29	5	3	27,29,75 (3)
CIR66	3	1	27 (1)
CIR74	3	1	93 (1)
703BS	5	0	0
CIR3	1	0	0
Sub-total	**61**	**29**	
Mean	**3.2**	**1.52**	
Range	**1-5**	**0-3**	

† Number of markers retained for linkage mapping analysis after applying Mather's† criterion (6.7:1) to select single dose markers.

Table 4.4: Summary of AFLP polymorphic markers used for constructing the female (CP 77-310) and male (CP 77-407) parental linkage maps based on molecular profiling of the selfed progeny of LCP 85-384.

Primer pair	**CP 77-310 (Female)**			**CP 77-407 (Male)**			***'ao x ao'***
	Mather's criterion	**Parent-specific**	**Co-segregation groups covered†**	**Mather's criterion**	**Parent-specific**	**Co-segregation groups covered †**	
E32M47	9	7	13,14,19,22,28,41,46,54 (8)	3	1	21,27,51 (3)	2
E32M48	8	3	3,6,14,35,36,48,60,73 (8)	11	6	4,12,13,28,29,35,45,46 (8)	5
E32M49	9	3	4,8,11,34,35,40,45,66 (8)	13	7	1,12,16,17,21,23,35,37,43,52, 55,60	6
E32M59	6	3	5,13,36,39,40,41 (6)	7	4	6,18,20,31,75,78 (6)	3
E32M60	7	4	1,5,9,28,44,56 (6)	7	4	9,10,25,40,47,59,72 (7)	3
E32M61	15	7	4,7,12,31,36,39,40,45,50,52,61,67 (12)	11	3	4,8,10,17,21,33,37,54,67 (9)	8
E32M62	2	2	8,54 (2)	1	1	49 (1)	0
E33M47	4	2	15,16,27,28 (4)	4	2	2,8,26,35 (4)	2
E33M48	8	5	1,4,13,19,28,57,58 (7)	3	0	27,77 (2)	3
E33M49	9	4	8,9,12,38,39,46,51 (7)	7	2	9,11,14,21 (4)	5
E33M50	10	2	7,9,18,27,29,30,68,70,81 (9)	8	0	9,22,24,26,30,56,59,71 (8)	8
E33M59	11	8	1,8,13,32,37,45,49,50,51,53,55 (11)	7	4	3,5,6,15,25,37 (6)	3
E33M60	10	4	1,9,18,46,50,56,68,78 (8)	11	5	4,9,32,36,44,47,55 (7)	6
E33M61	11	3	7,21,22,32,51,54,59,62,69 (9)	12	4	23,30,31,35,36,41,43,44,67,76 (10)	8
E36M47	10	2	10,12,23,46,47,63,70,77 (8)	9	1	5,8,11,35,70,71,75 (7)	8
E36M48	7	3	13,40,47 (3)	8	4	4,21,36 (3)	4
E36M49	15	7	1,2,9,17,20,27,46,60,65,76,78,80 (12)	13	5	8,9,14,26,33,39,48,62,66,68,72 (11)	8
E36M59	11	4	1,9,12,27,33,38,50,77,79 (9)	14	7	3,8,9,18,23,32,35,41,51,62,63,74 (12)	7
E36M60	10	5	1,6,12,37,39,46,48,58 (8)	8	3	12,14,29,32,38,47,54,79 (8)	5
E36M61	16	6	1,18,23,27,32,33,35,41,43,47,49,55,62,65,67 (15)	16	6	5,8,12,20,32,33,35,37,45,49,59,67 (12)	10
E36M62	7	5	35,41,46,47,49,59 (6)	6	4	5,12,28,45 (4)	2
E37M47	9	6	27,28,30,35,37,39,41,43 (8)	5	2	12,27,35 (3)	3
E37M48	4	1	30,39,79 (3)	5	2	3,24,57,63,73 (5)	3
E37M49	11	6	4,13,29,32,39,41,45,52,53 (9)	8	3	1,3,22,36,37,38 (6)	5
E37M50	7	5	4,17,39,46,51,68,75 (7)	5	3	3,35,36,58 (4)	2
E37M59	6	3	2,14,26,32,39,55 (6)	7	4	38,46,44,45,46,50,58,69 (8)	3
E37M60	6	3	32,37,40,41,52 (5)	4	1	15,21,36 (3)	3
E37M61	8	7	1,8,10,47,48,51,69 (7)	5	4	4,16,17,36,61 (5)	1
E37M62	4	3	15,45,58,67 (4)	3	2	2,25,46 (3)	1

TABLE 4.4 continued

E38M47	4	2	12,37,53 (3)	3	1	8,23,34 (3)	2
E38M48	8	4	9,34,38,39,41,42 (6)	10	6	9,13,16,19,20,23,33,35,55,57 (10)	4
E38M49	7	4	14,25,28,43,49,52,65 (7)	6	3	27,33,36,37,51 (5)	3
E38M50	8	1	32,34,46,76 (4)	9	2	10,41,46,53,57,68,70 (7)	7
E38M59	10	7	3,12,18,27,36,48,49,71 (8)	7	4	4,8,13,43,44,80 (6)	3
E38M60	8	5	21,27,29,32,49 (5)	5	2	22,25,44,50,56 (5)	3
E38M61	9	5	1,13,15,35,36,51,53,60,63 (9)	6	2	2,12,13,25,35,49 (6)	4
E38M62	6	1	43,46,64 (3)	7	2	3,16,42,49 (4)	5
E39M47	5	3	1,22,54,57,71 (5)	3	1	50,77,80 (3)	2
E39M48	8	4	27,28,29,34,43,44,74 (7)	10	6	19,22,26,40,45,50,61,74,77 (9)	4
E39M49	9	3	8,11,17,23,24,34,35,47 (8)	10	4	5,12,23,33,35,44,52,55,60,65 (10)	6
E39M50	5	2	14,37,52,81 (4)	6	3	8,36,51,56,78 (5)	3
E39M59	6	2	12,42,50,60,61,72 (6)	4	0	19,54 (2)	4
E39M60	8	3	1,5,6,12,39,45,52,75 (8)	10	5	3,8,25,29,33,36,37,43,58,60 (10)	5
E39M61	9	3	27,30,44,48,55,57,73 (7)	10	4	8,24,28,40,64,77 (6)	6
E39M62	4	1	6,14,21,45 (4)	3	0	29,44,69 (3)	3
E40M47	5	2	17,27,33,34,53 (5)	3	0	34,45 (2)	3
E40M48	5	3	25,26,41,45 (4)	4	2	46,78 (2)	2
E40M49	8	5	5,8,20,39,61,80 (6)	5	2	8,17,54,66 (4)	3
E40M50	4	2	3,15,47,59 (4)	3	1	2,65,77 (3)	2
E40M59	7	4	1,2,20,45 (4)	7	4	7,43,48,50 (4)	3
E40M60	7	4	11,16,34,38,53,56,58 (7)	4	1	8,14,79 (3)	3
E40M61	7	5	13,34,50,57,61,74 (6)	2	0	52 (1)	2
E41M47	2	1	32,34 (2)	2	1	25,76 (2)	1
E41M48	6	4	1,14,17,30 (4)	4	2	24,36,43,45 (4)	2
E41M49	4	0	5,24 (2)	4	0	44 (1)	4
E41M50	5	1	41,45,47,66,72 (5)	5	1	1,20,37,73 (4)	4
E41M59	0	0	0	1	1	30 (1)	0
E41M60	4	2	12,32,39 (3)	7	5	1,5,7,8,30,39 (6)	2
E41M61	10	3	6,8,12,43,49,57,66 (7)	13	6	1,3,9,29,33,42,43,46,77 (9)	7
E41M62	2	1	1,6 (2)	3	2	29,57 (2)	1
Total	**440**	**210**		**397**	**167**		**230**
Mean	**7.34**	**3.5**		**6.61**	**2.78**		**3.84**
Range	**0-16**	**0-8**		**0-16**	**0-7**		**0-10**

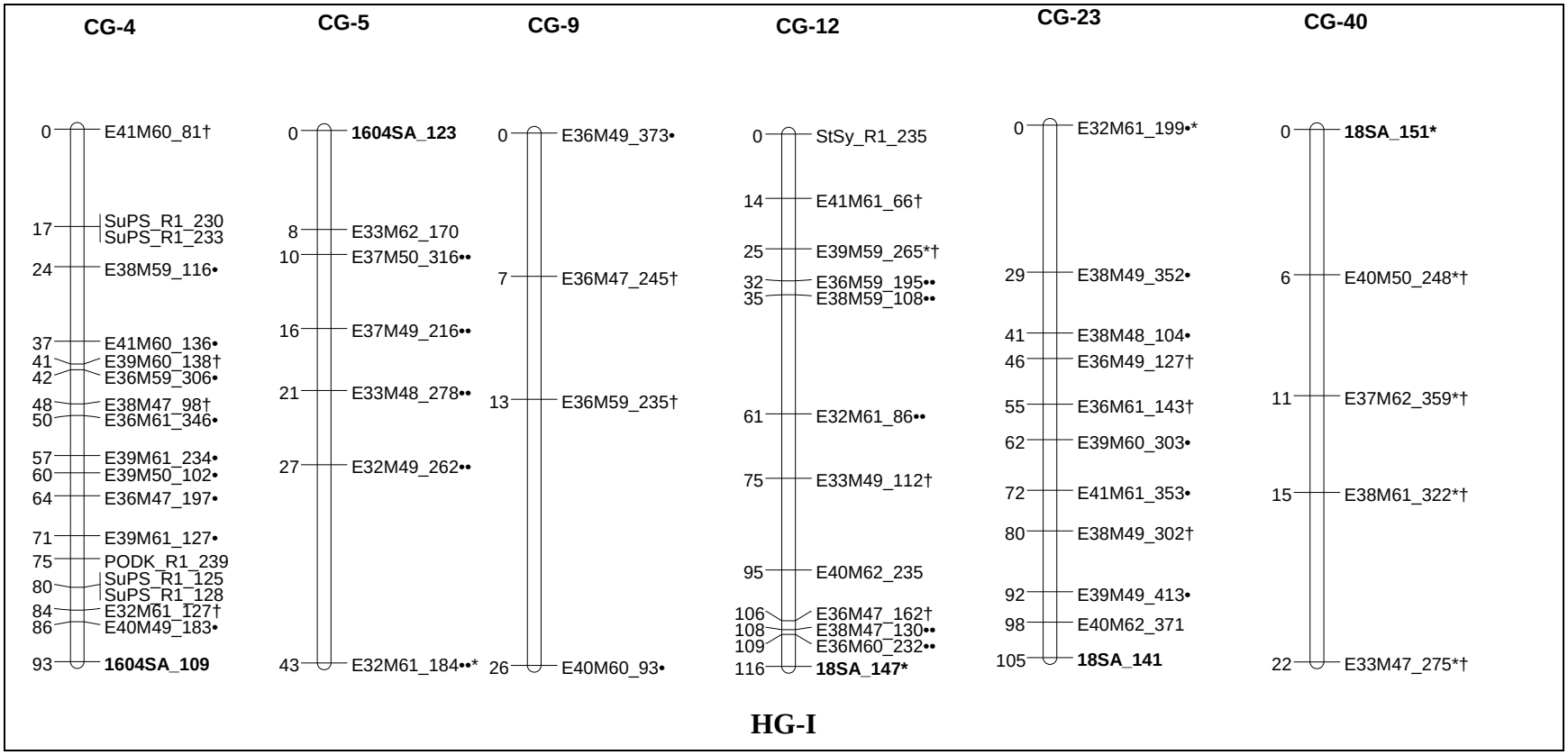


Figure 4.2: S_1 linkage map of sugarcane hybrid 'LCP 85-384' from a selfed progeny of 300 individuals. The map was constructed with a LOD score > 4.0 and a recombination fraction of 0.44 using AFLP, TRAP and SSR markers. A total of 773 single dose markers (3:1) were assigned onto 108 CGs. The vertical bars indicate CGs with markers in coupling phase linkages. The Kosambi map distances (cM) marker names are indicated on the left and right sides, respectively, of each CG. AFLP markers denoted by 'EM', TRAP markers are denoted as per the table 4.2. The remaining are SSR markers. CGs were grouped into HG based on SSR loci and information on grandparental maps. CGs belong to the same HG are enclosed in boxes and HGs are Roman numbered. The SSR alleles responsible in each HG are represented in bold. The grandparental specific markers are represented by one dot (·) or two dots (··) for CP 77-407 and CP 77-310, respectively. The markers present in both parents and segregated in S_1 population are denoted by the † symbol**.** The marker names with an asterisk (*) represent distorted markers. Independent CGs (do not belong to any HG) are not represented in boxes.

FIGURE 4.2 continued

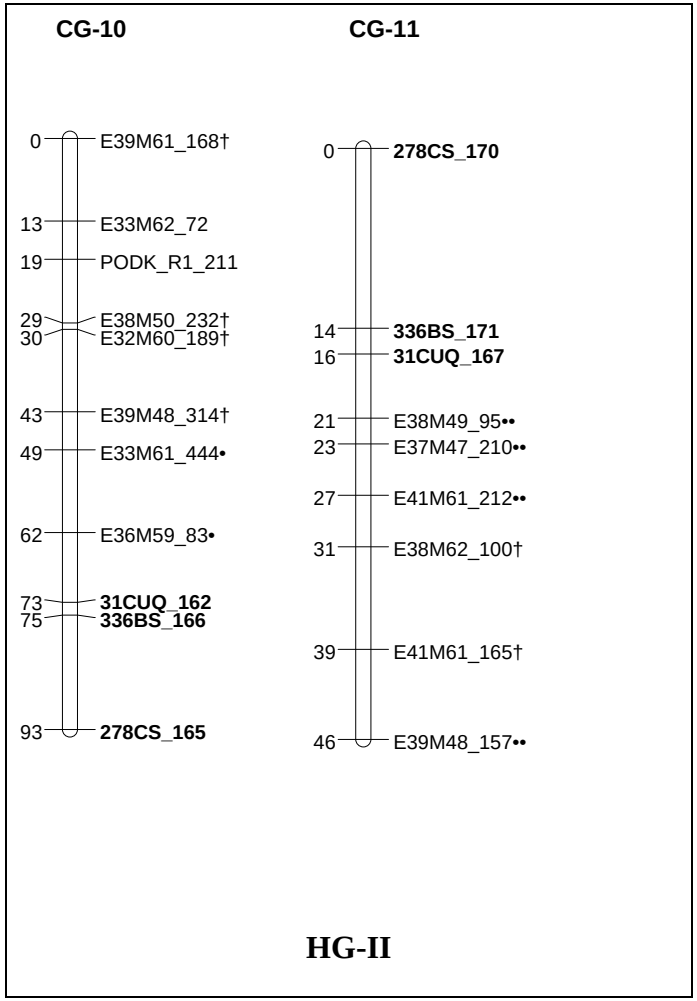


FIGURE 4.2 continued

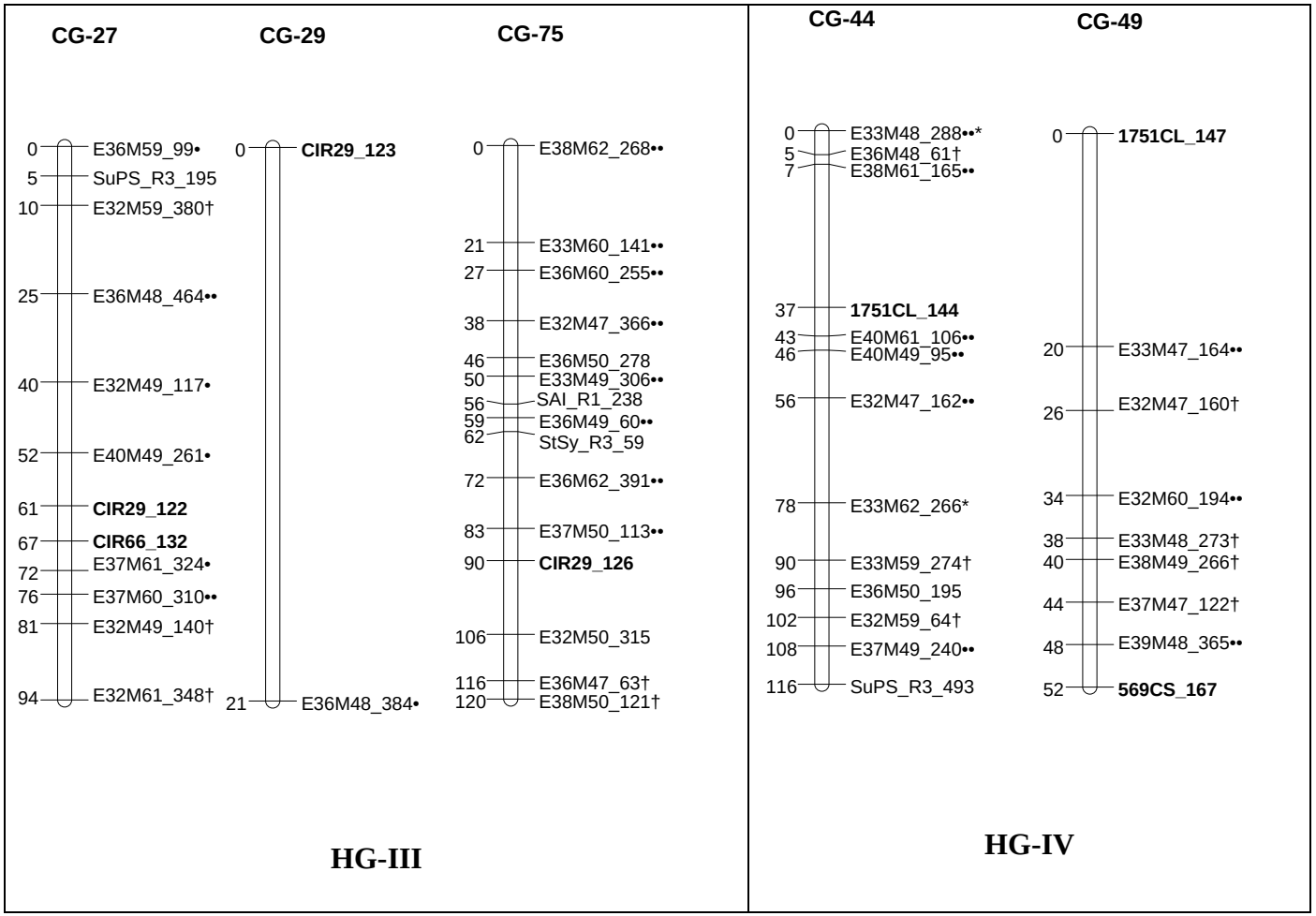


FIGURE 4.2 continued

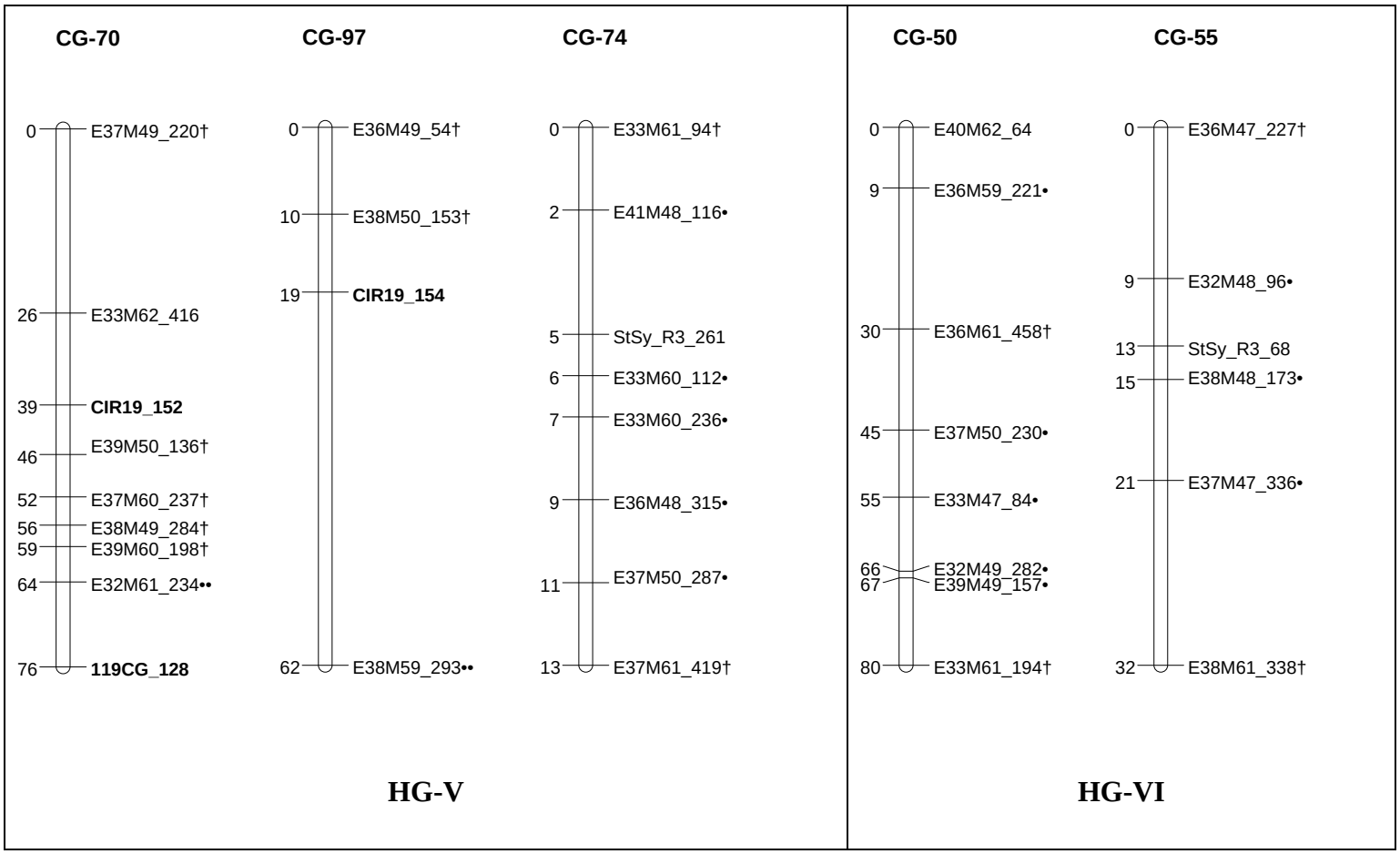


FIGURE 4.2 continued

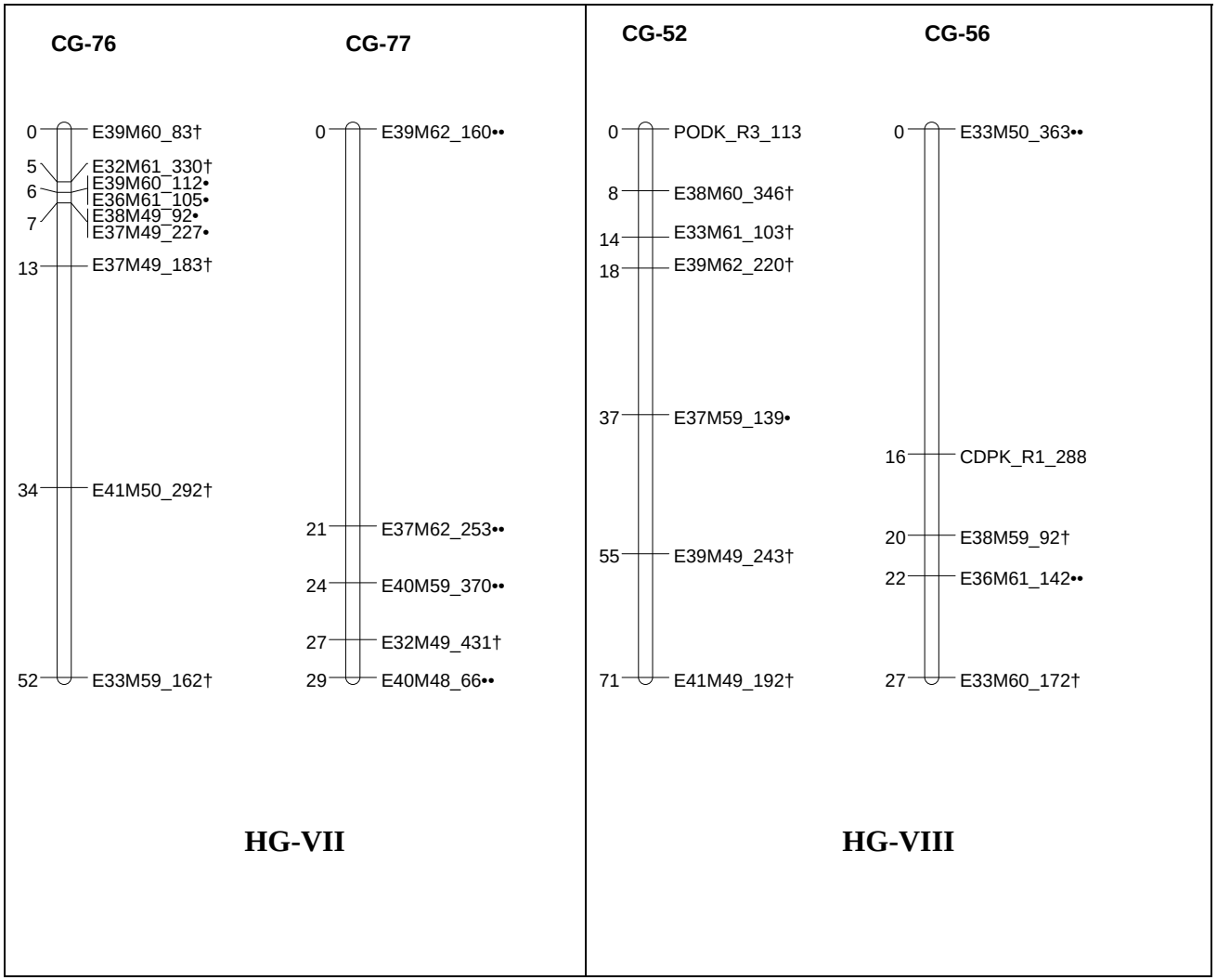


FIGURE 4.2 continued

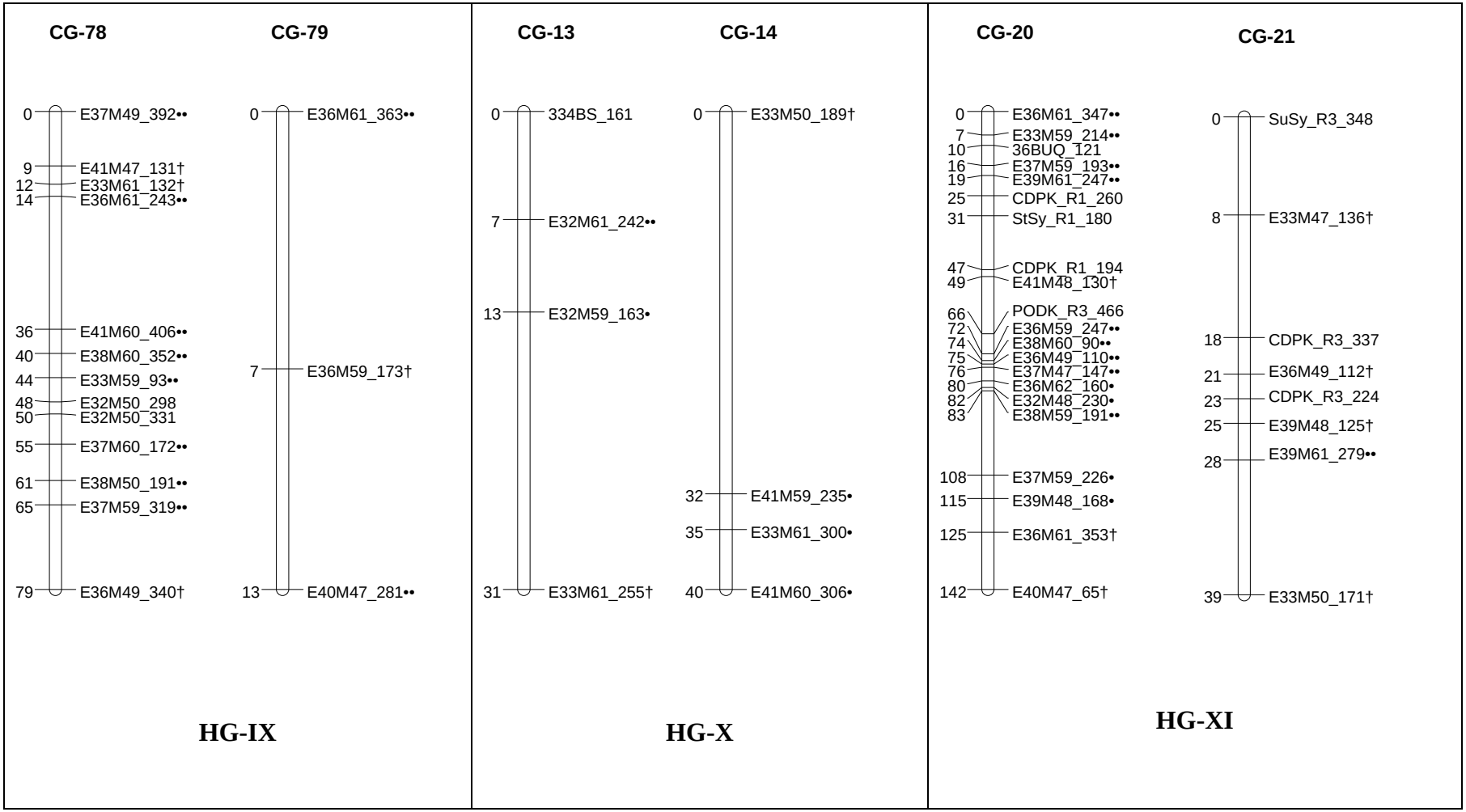
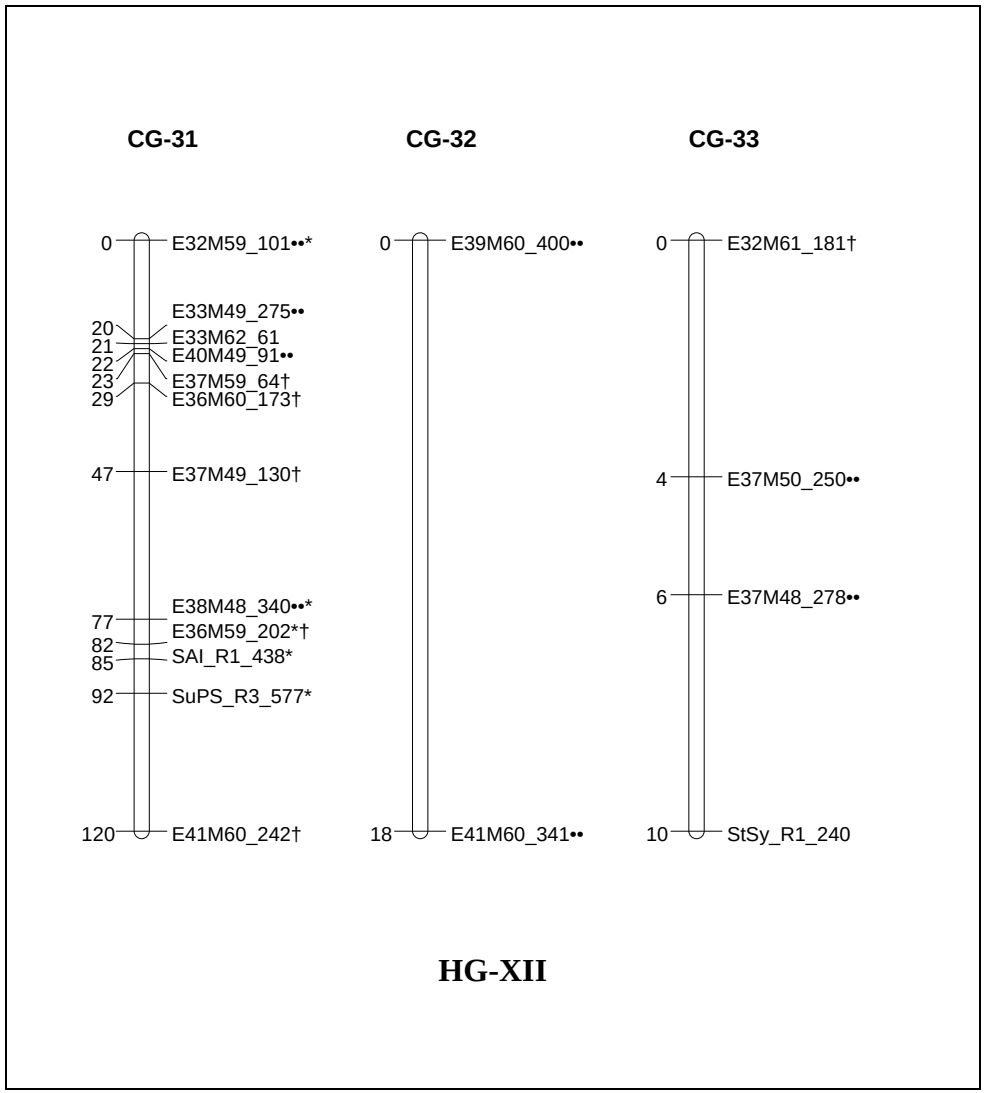


FIGURE 4.2 continued



HG-XII

FIGURE 4.2 continued

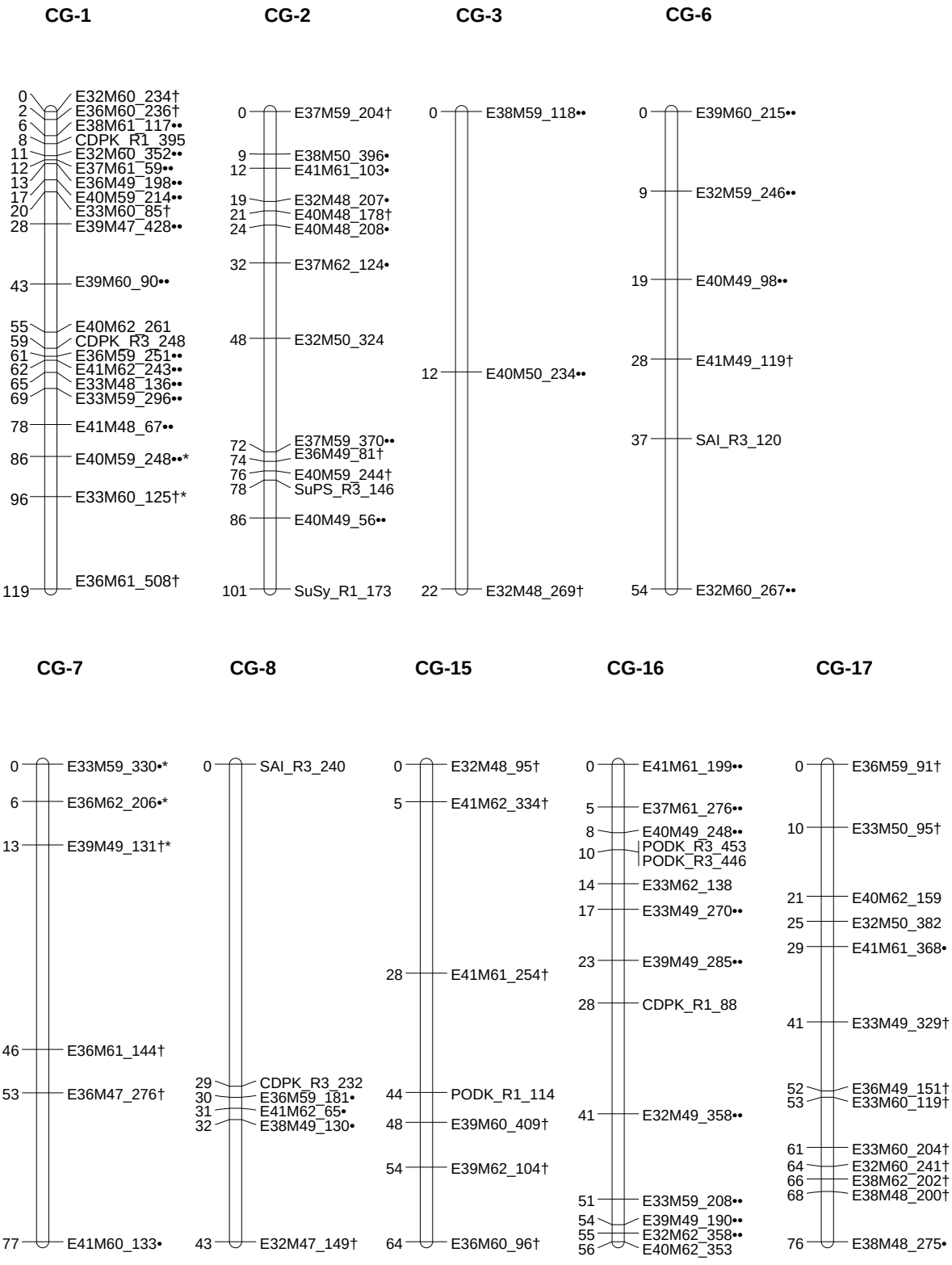


FIGURE 4.2 continued

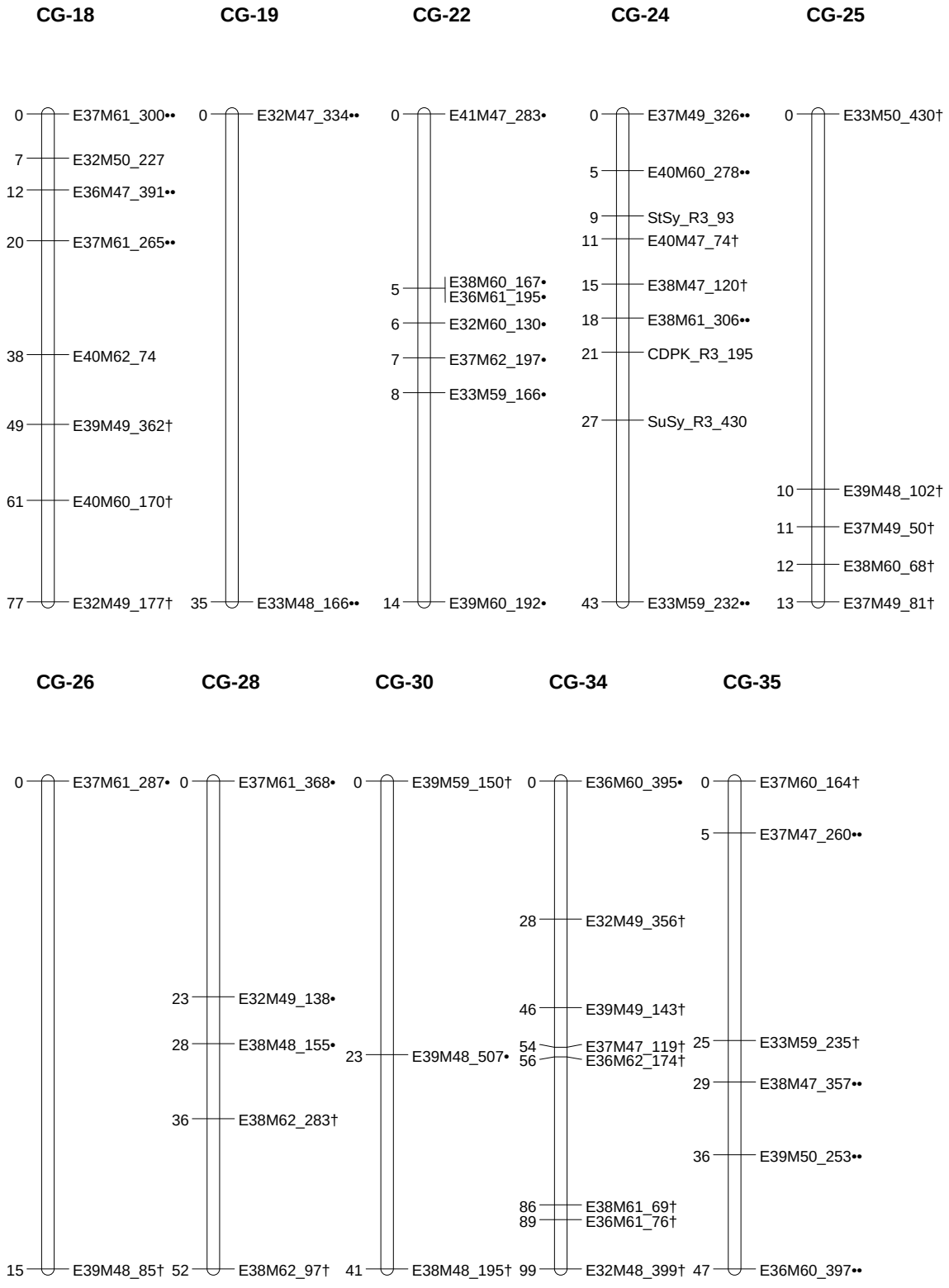


FIGURE 4.2 continued

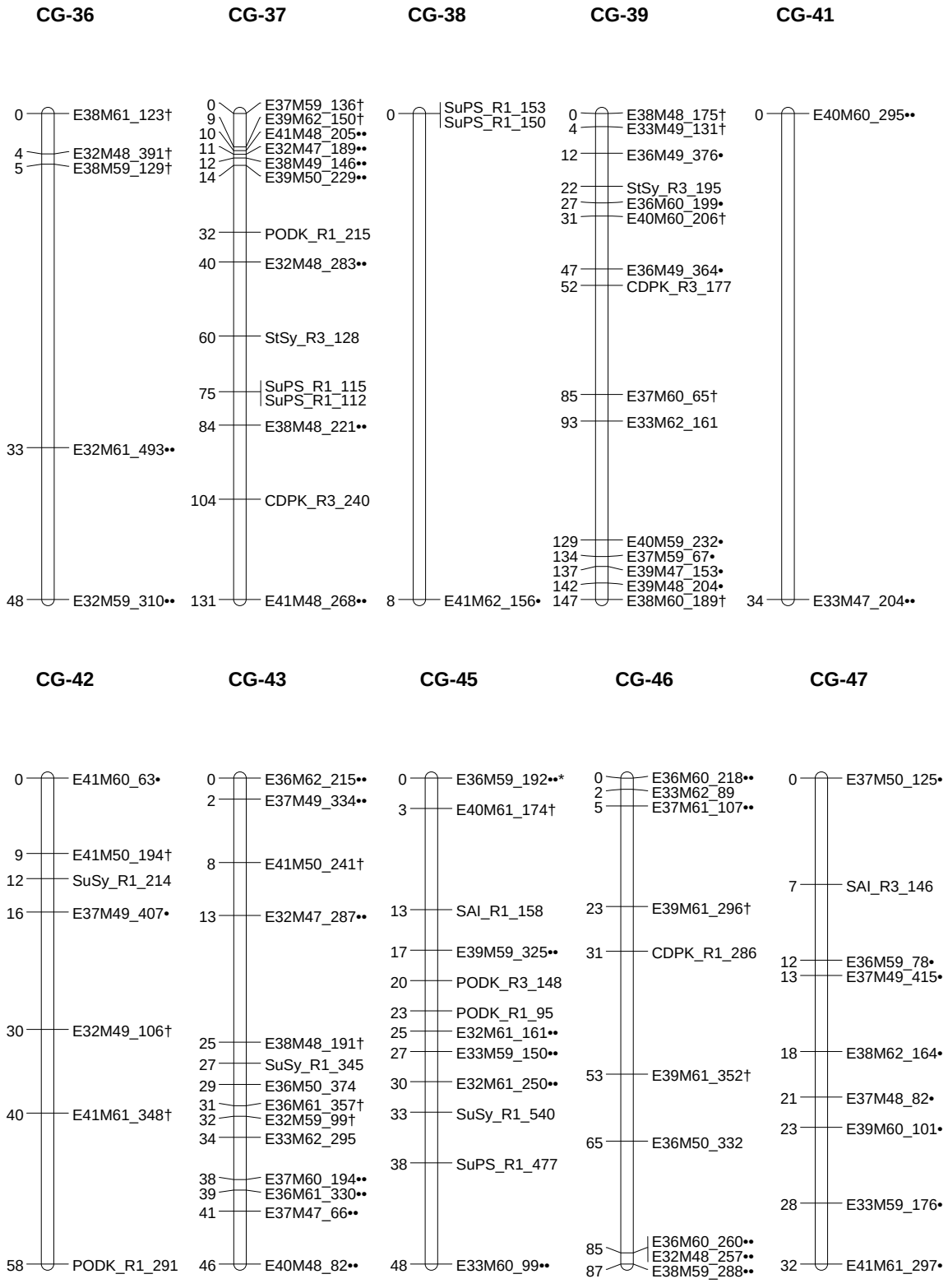


FIGURE 4.2 continued

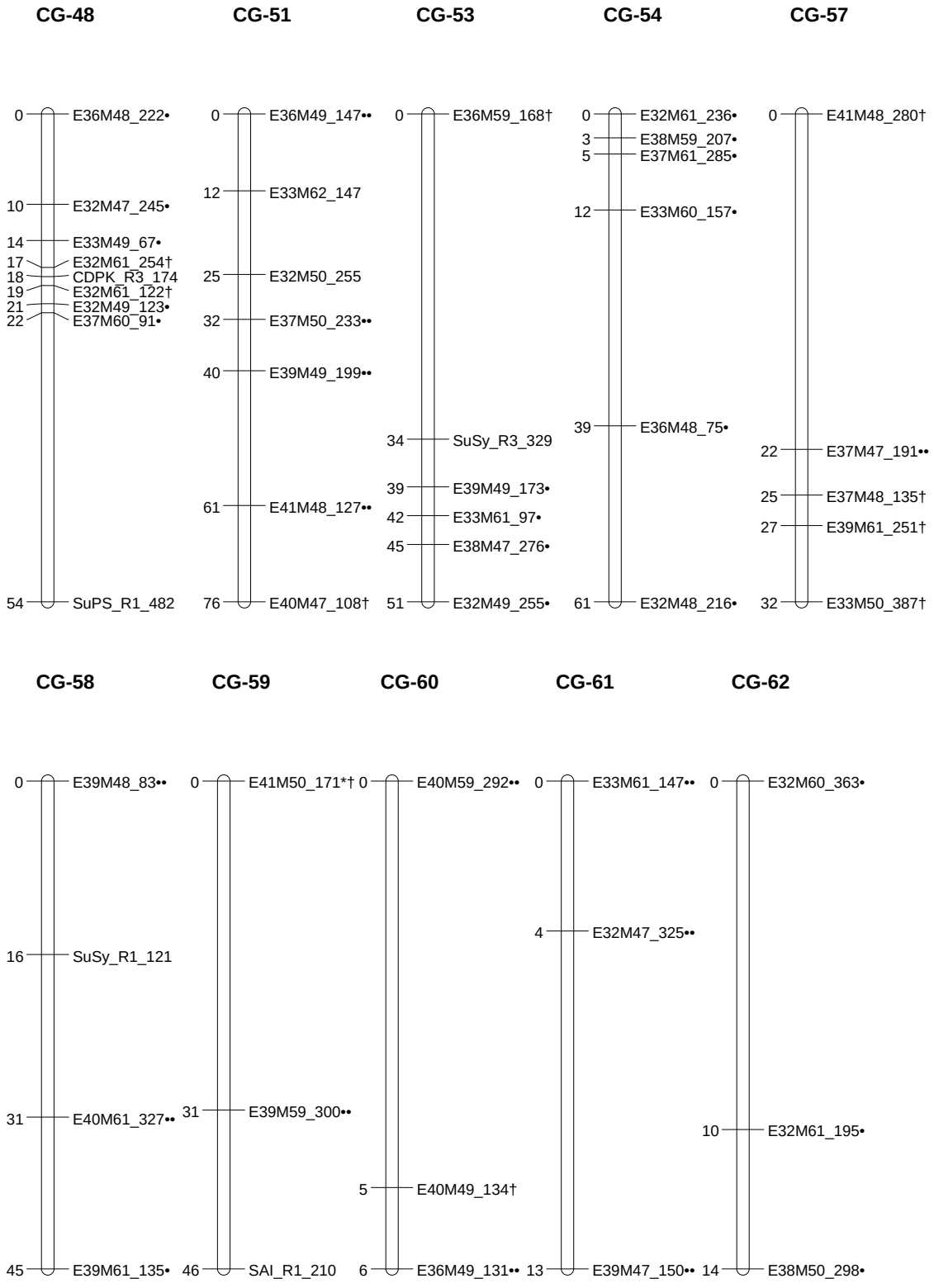


FIGURE 4.2 continued

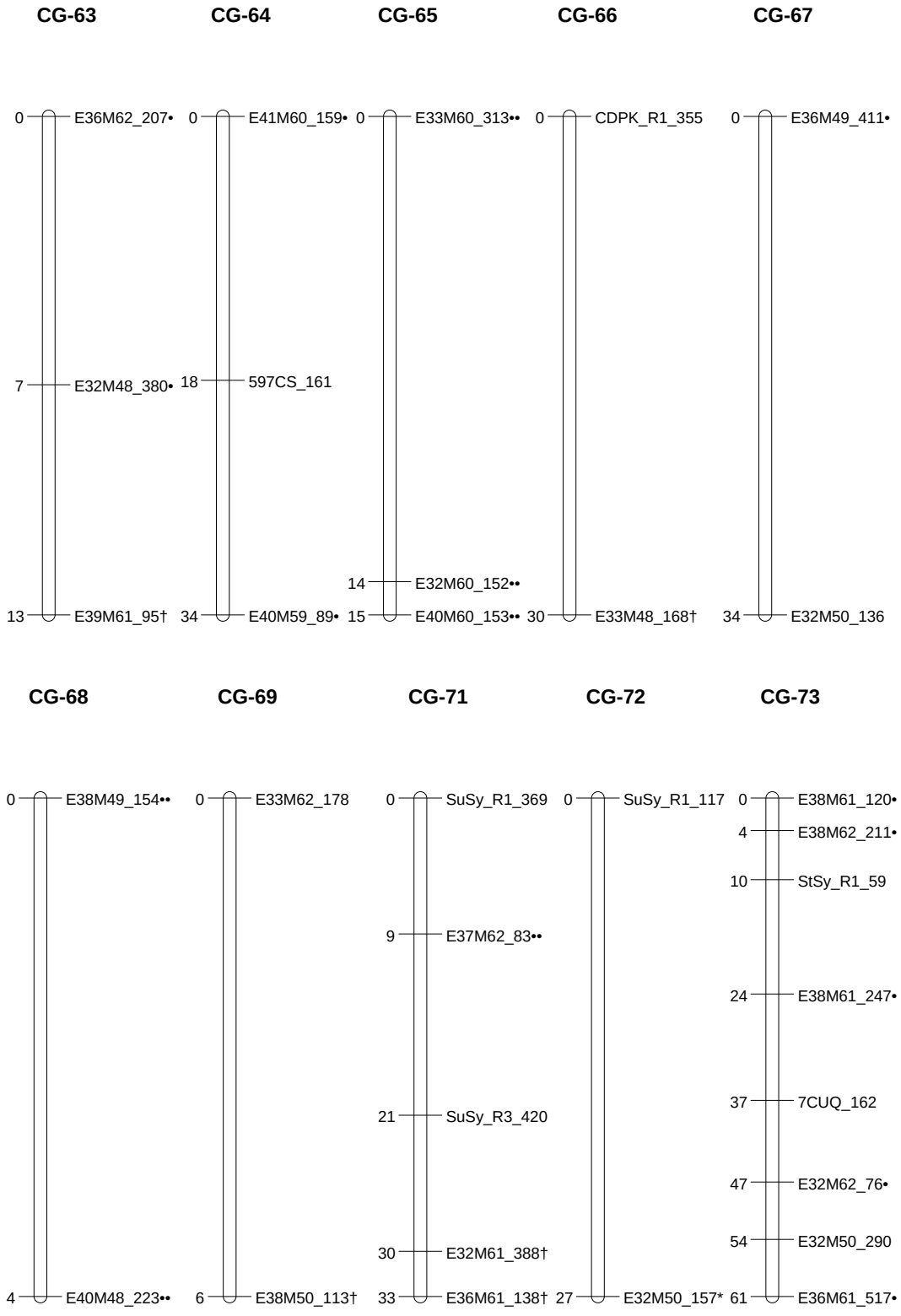


FIGURE 4.2 continued

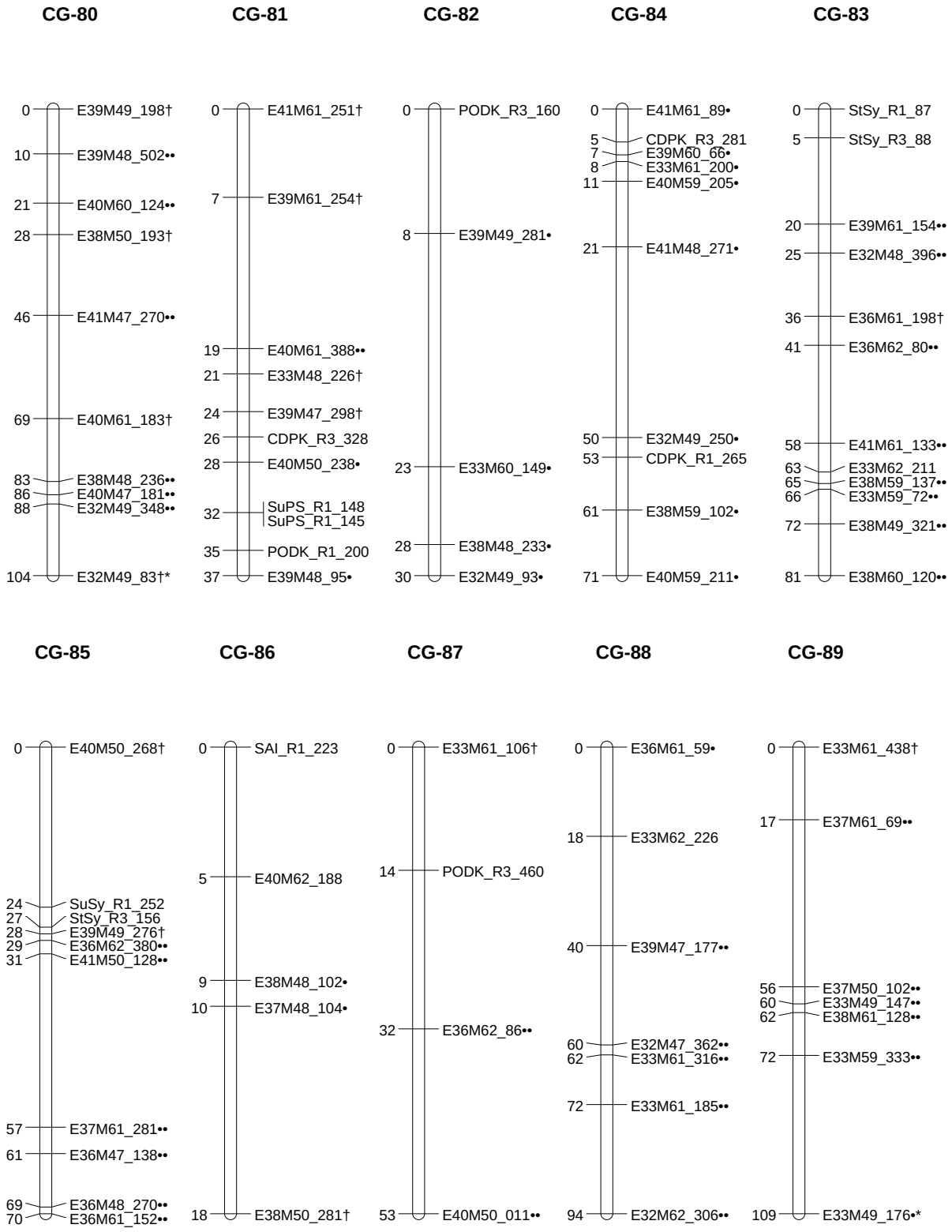


FIGURE 4.2 continued

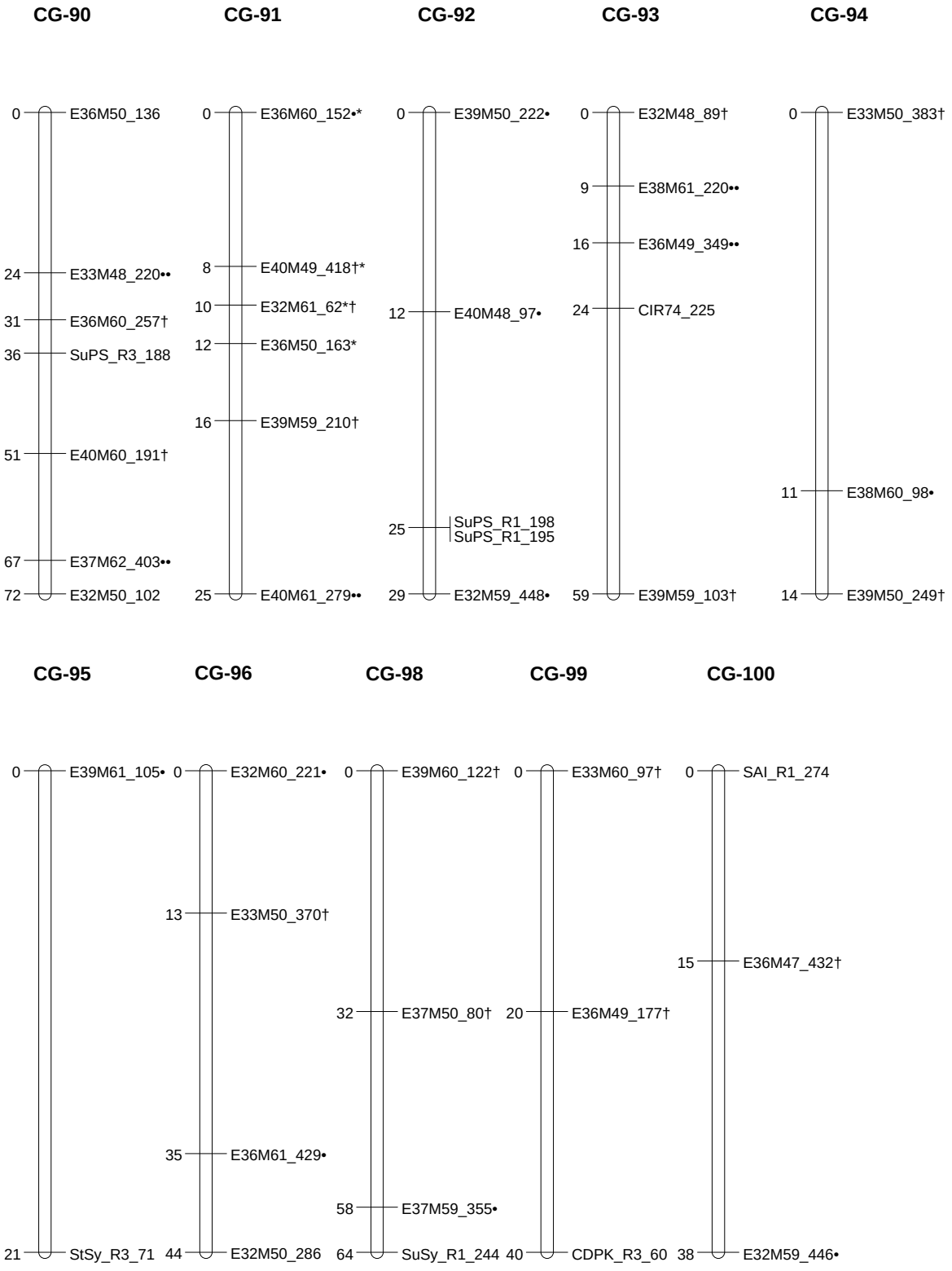
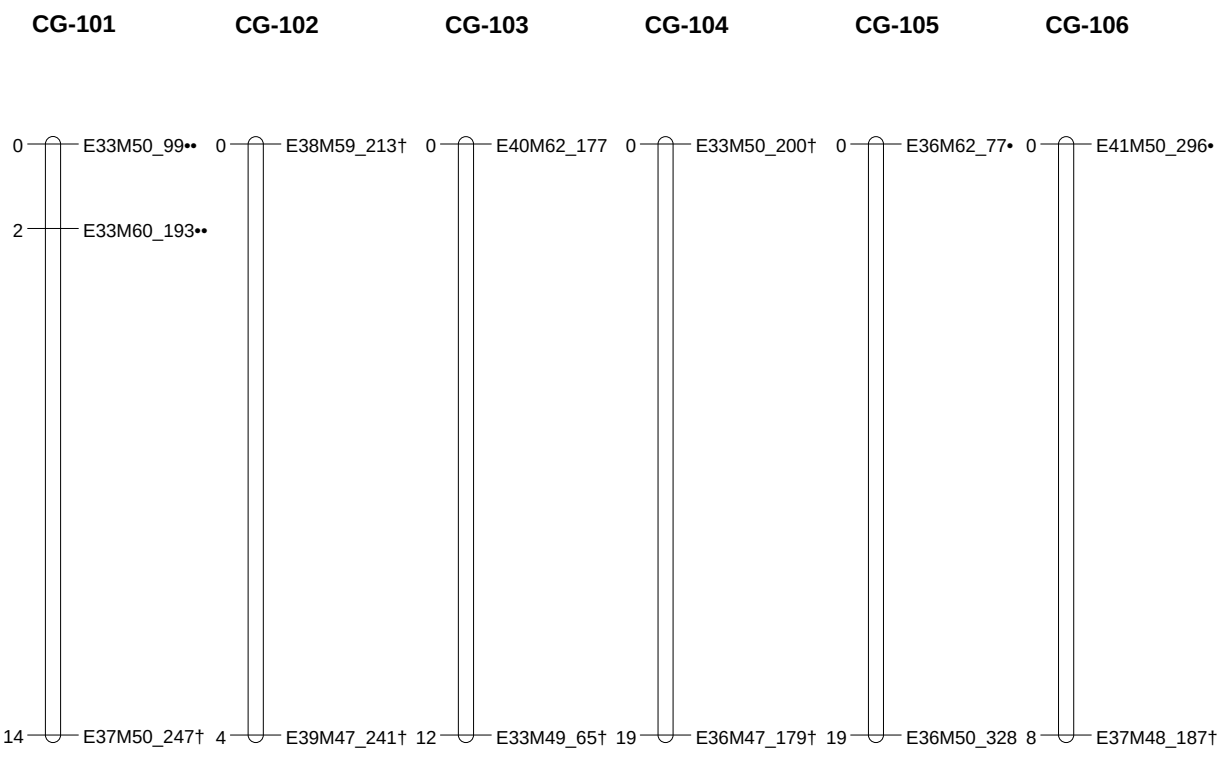
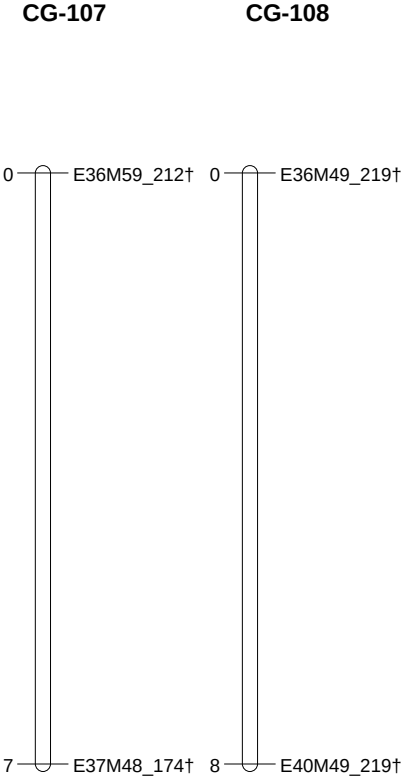


FIGURE 4.2 continued


CG-101
CG-102
CG-103
CG-104
CG-105
CG-106
0 E33M50_99••
2 E33M60_193••
14 E37M50_247†
0 E38M59_213†
4 E39M47_241†
0 E40M62_177
12 E33M49_65†
0 E33M50_200†
19 E36M47_179†
0 E36M62_77•
19 E36M50_328
0 E41M50_296•
8 E37M48_187†


CG-107
CG-108
0 E36M59_212†
7 E37M48_174†
0 E36M49_219†
8 E40M49_219†

4.3.2.2 LCP 85-384 Parental Maps

A total of 440 markers in the female (CP 77-310) and 397 markers in the male (CP 77-407) parents were used to develop parental maps of LCP 85-384. Information from only 60 AFLP markers was available for this analysis. The CP 77-310 map comprised of 391 linked markers, which spread over 81 CGs with a cumulative genomic length of 3476 cM, where 49 markers remained unlinked. The length of the 81CGs varied from 4 cM (CG310-26, -31, and -81) to 196 cM (CG310-39) with an average of 8.9 cM between any two adjacent markers. The highest number of loci (18) forming a CG were found in CG310-1. Markers generated by a primer pair covering CGs ranged from 0 (E41M59) to 15 (E36M61) (Figure 4.3 and Table 4.3).

In contrast, the CP 77-407 map comprised of 339 markers spanning 80 CGs with a cumulative genome length of 2777 CM. Fifty eight markers remained unlinked. The CGs varied in length from 4 cM (CG407-75 and -80) to 115 cM (CG407-44) with an average of 8.19 cM. The highest number of loci (18) forming a LG were found in CG407-1. Markers generated by a primer pair covering CGs ranged from 1 (E32M62, E40M61, E41M49) to 12 (E36M61, E36M59, and E32M49) (Figure 4.4).

The distorted markers, 16 out of 19 in CP 77-310 and 14 out of 18 in CP 77-407, were mapped to the grandparental maps. The distorted markers in CP 77-310 and CP 77-407 scattered over 10 CGs and 7 CGs, respectively. The cumulative skewed chromosomal regions in CP 77-310 (10+3=13) and CP 77-407 (4+7=11) were within the limit of expected number (24) by the random effect according to the Bonferroni procedure.

4.3.3 Homo(eo)logous Groups (HGs)

The co-dominance nature of SSR markers, and information on grandparental maps enabled the CGs to be assembled into putative (HGs) (Figure 4.2). The AFLP markers belonging to a single parental CG were distributed in more than one of the CGs of the LCP 85-384 map.

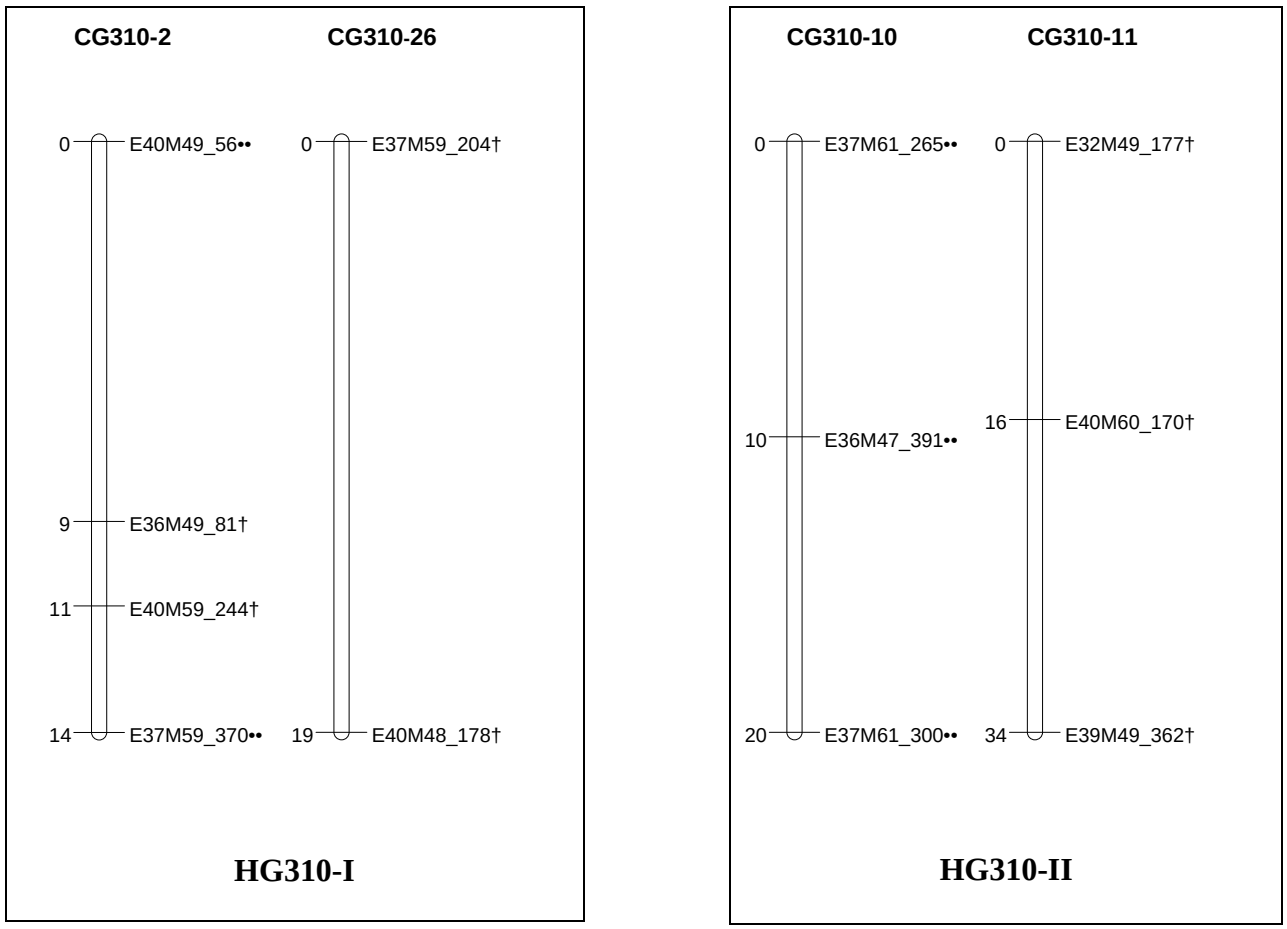


Figure 4.3: Female (CP 77-310) parent linkage map of the sugarcane hybrid 'LCP 85-384' based on S_1 mapping population. The map was constructed with a LOD score > 4.0 and a recombination fraction of 0.44 using AFLP markers. A total of 440 single dose markers (3:1) were assigned onto 81 CGs. The vertical bars indicate CGs with markers in coupling phase linkages. The Kosambi map distances (cM) and marker names are indicated on the left and right sides, respectively, of each CG. AFLP markers are denoted by 'EM'. CGs were grouped into HG based on S_1 map information. CGs of the same HG are enclosed in boxes and HGs are Roman numbered. The CP 77-310 specific markers are represented by one dot (·). The markers present in both parents and segregated in the S_1 population are denoted by the † symbol. The marker names with an asterisk (*) represent distorted markers. Independent CGs (do not belong to any HG) are not represented in boxes.

FIGURE 4.3 continued

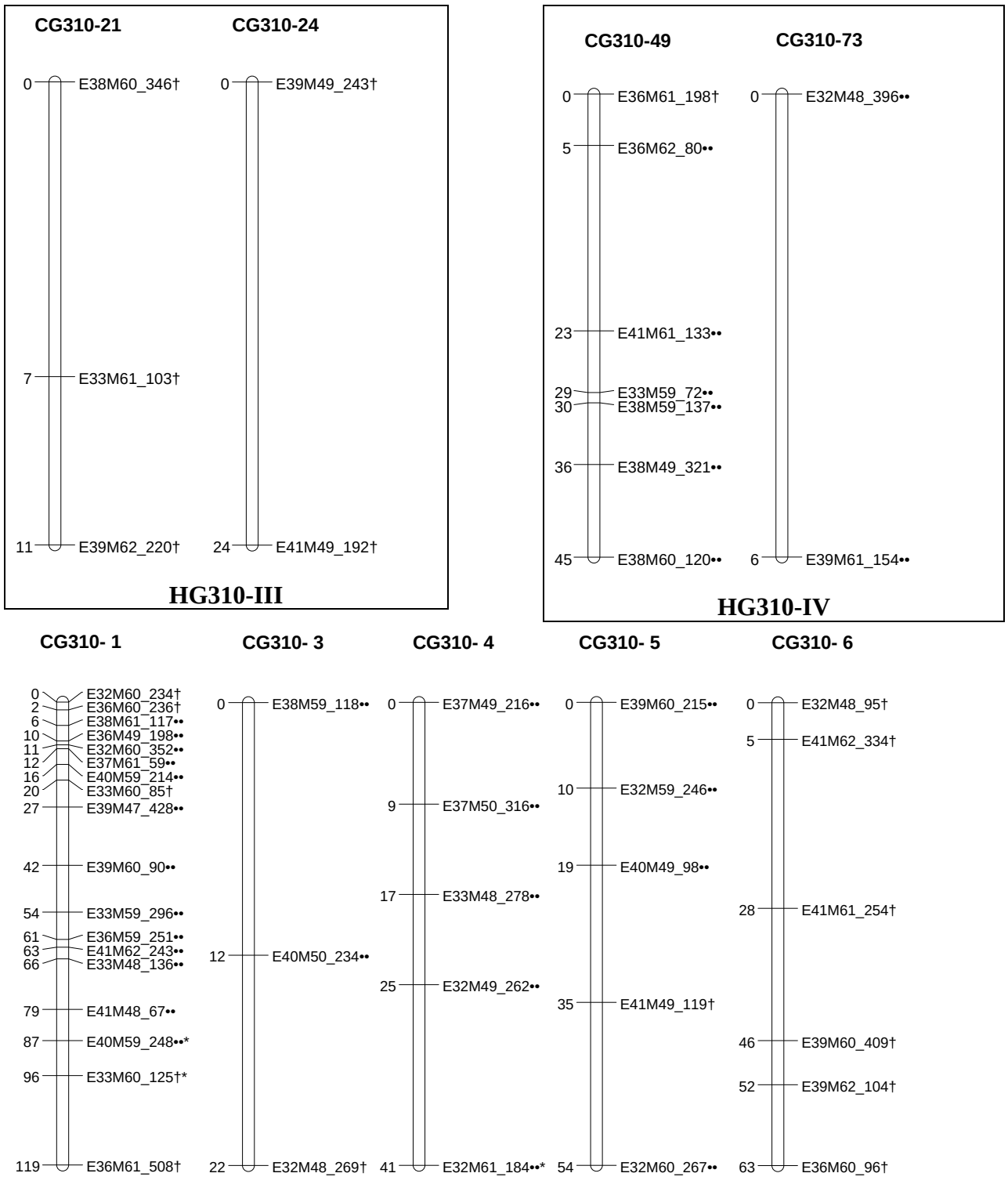


FIGURE 4.3 continued

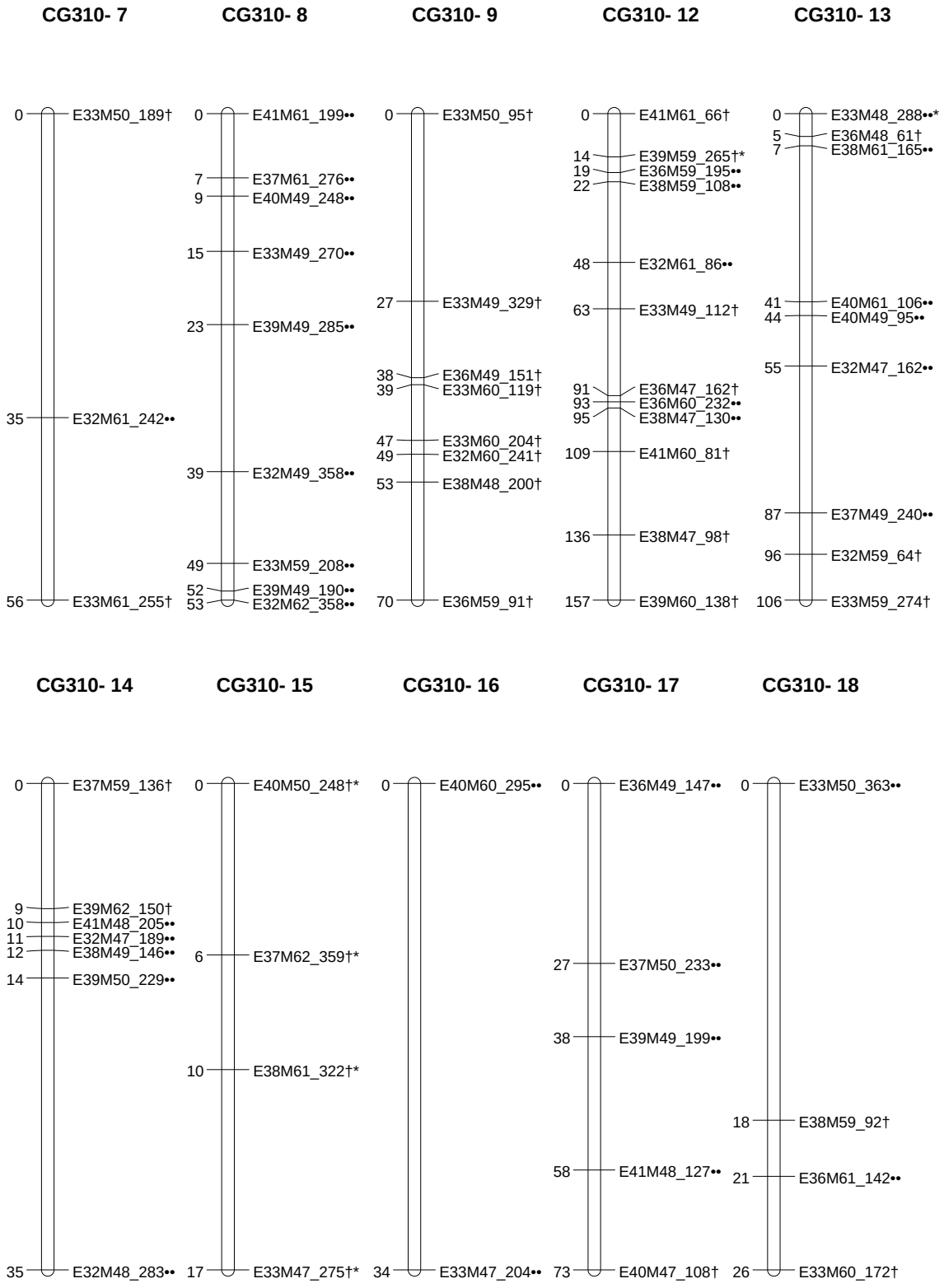


FIGURE 4.3 continued

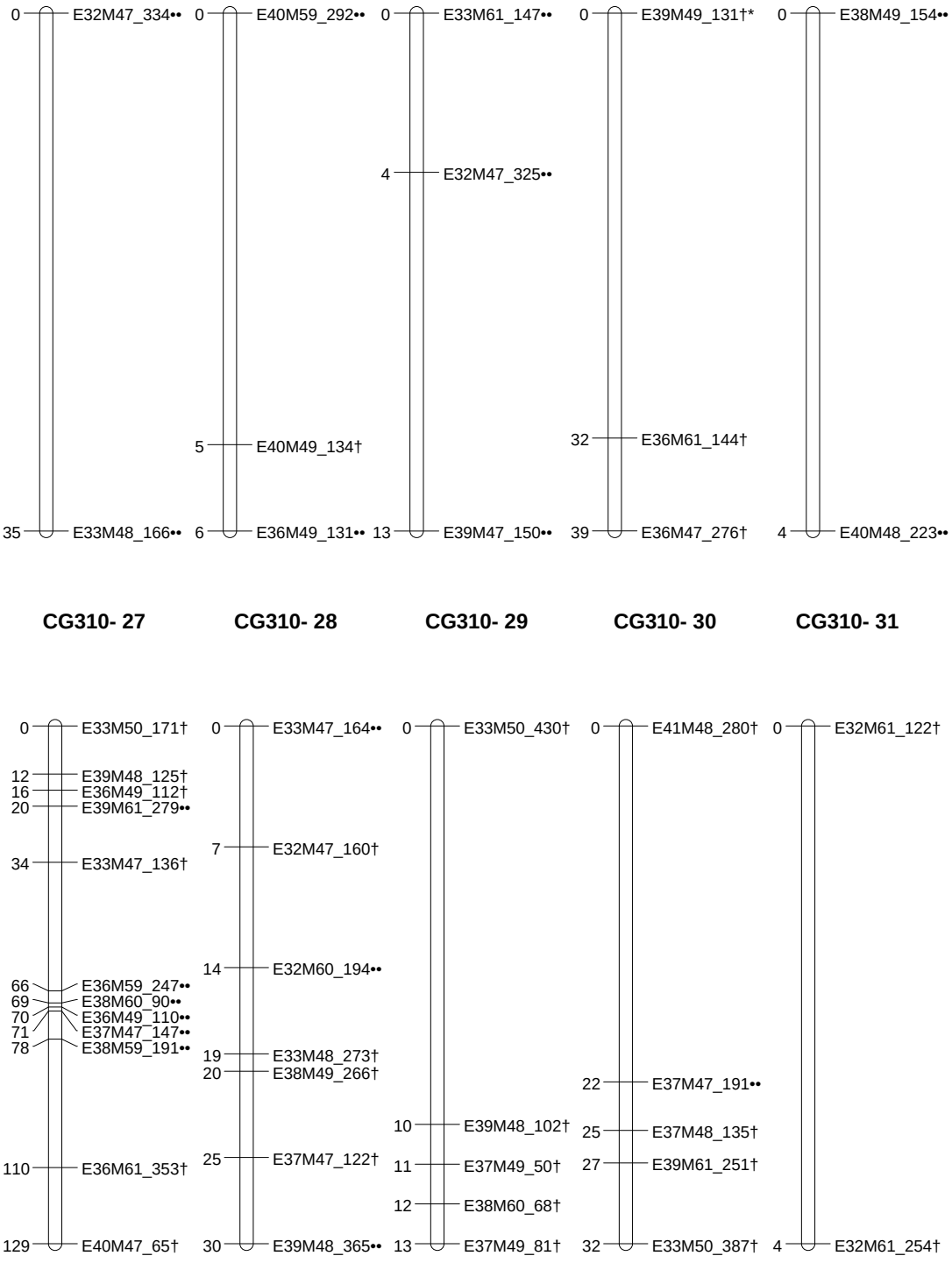


FIGURE 4.3 continued

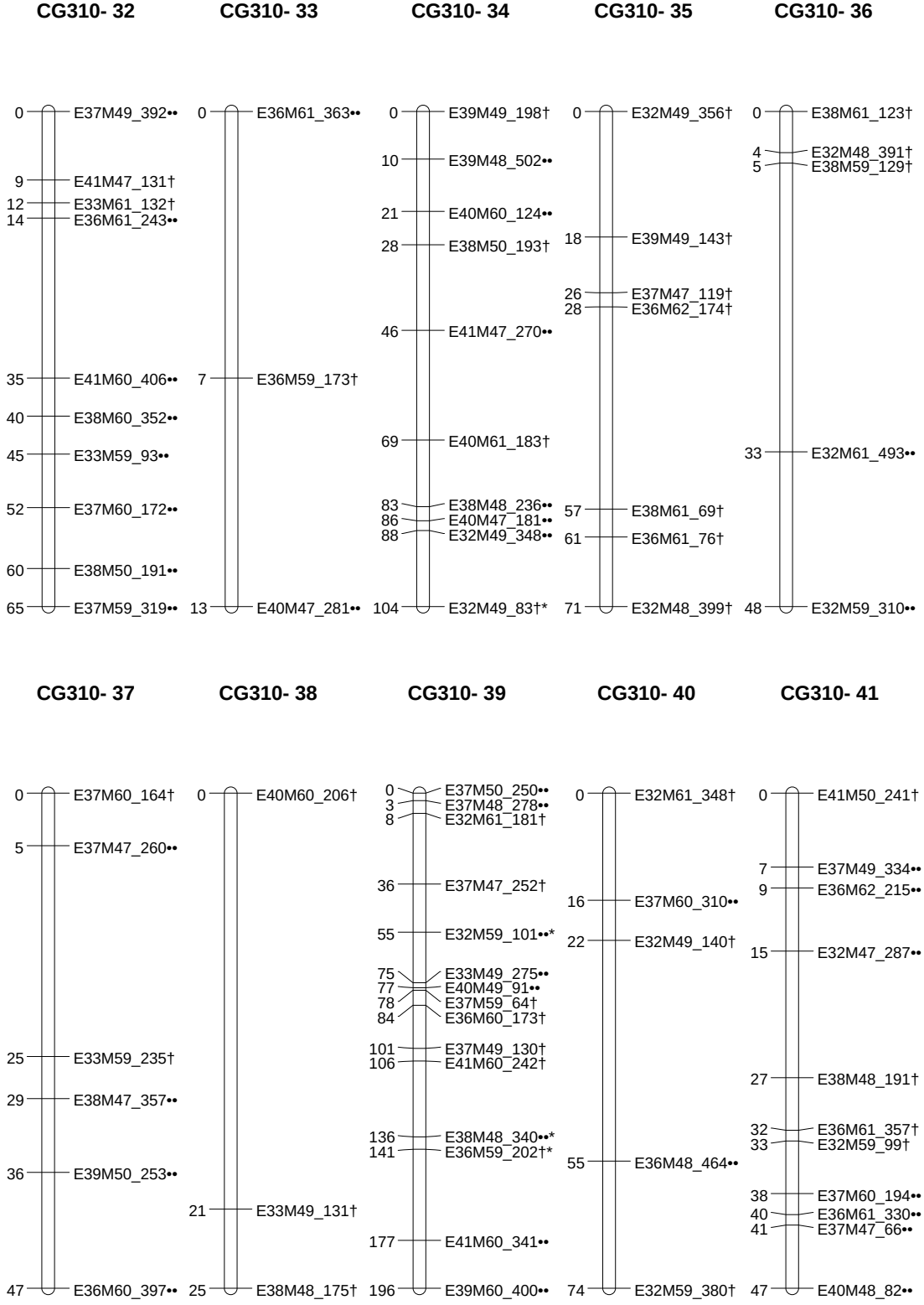


FIGURE 4.3 continued

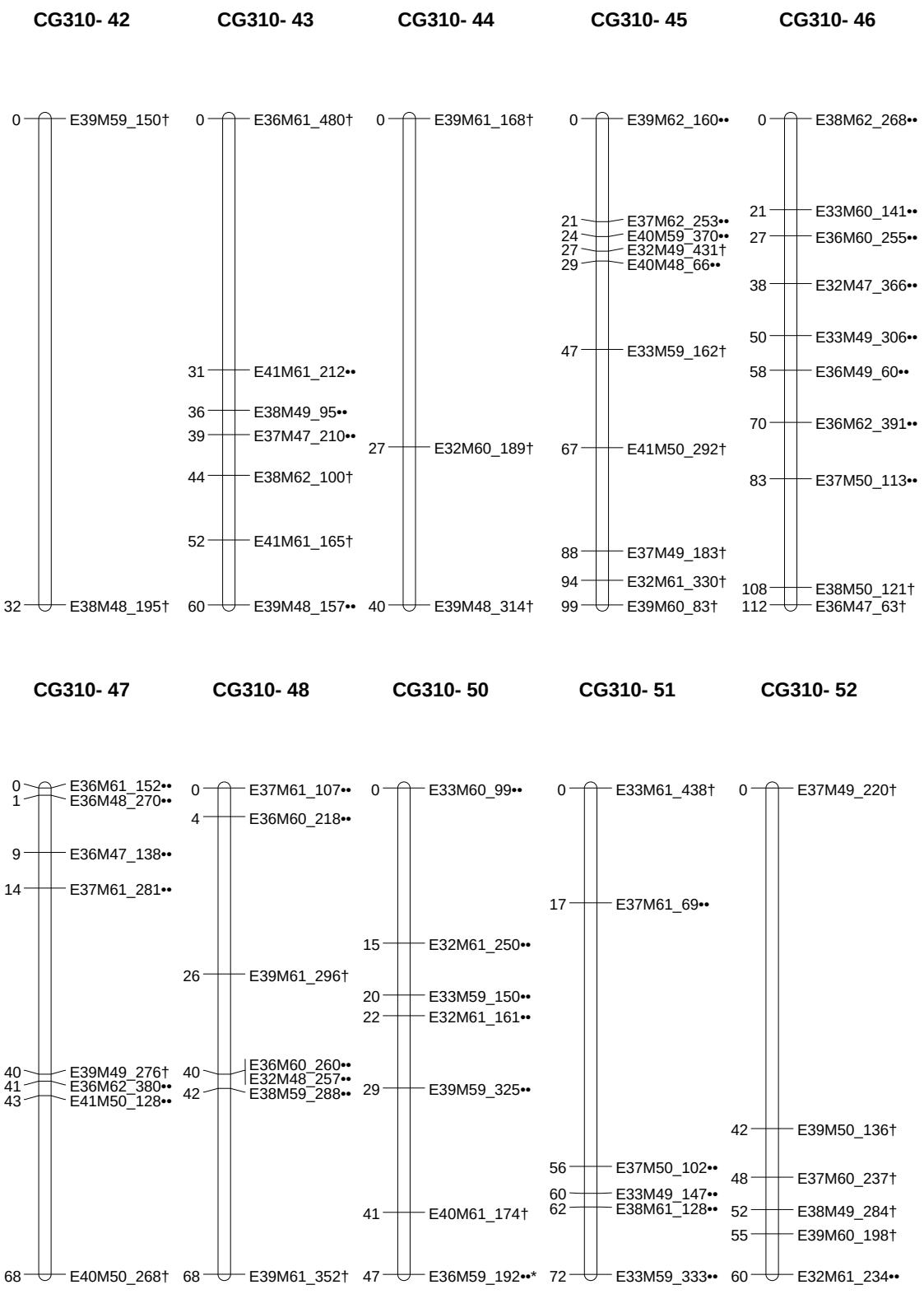


FIGURE 4.3 continued

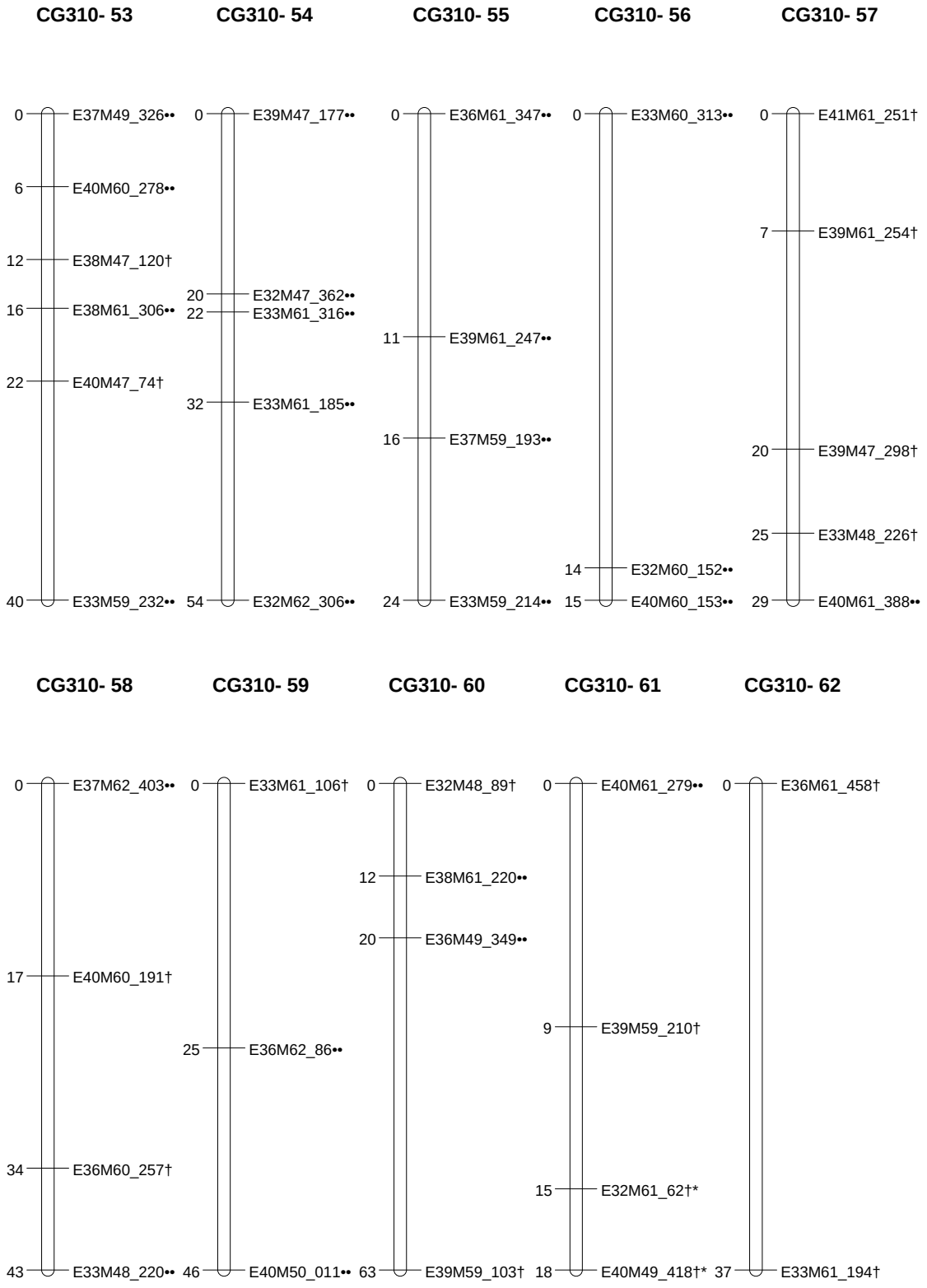


FIGURE 4.3 continued

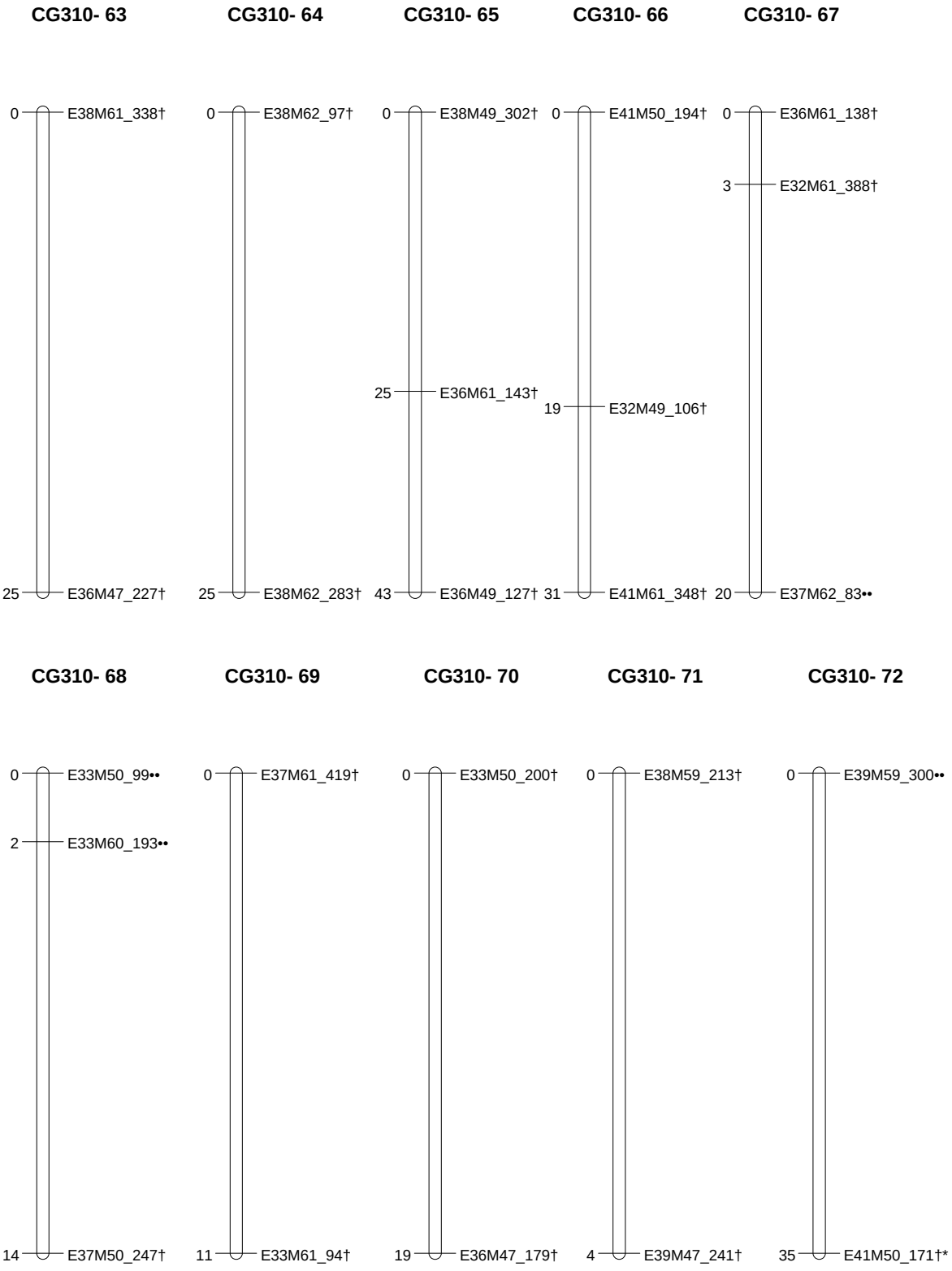
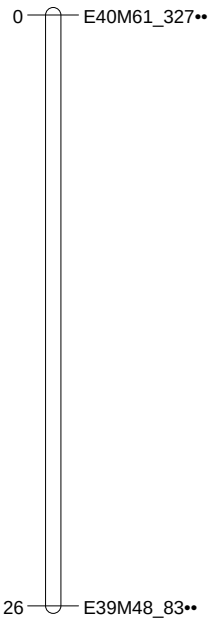
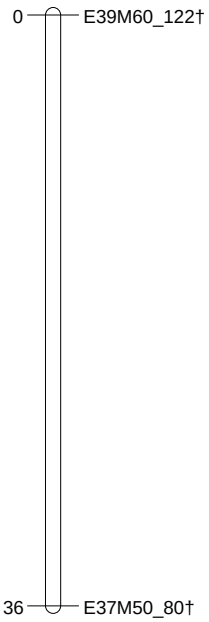


FIGURE 4.3 continued

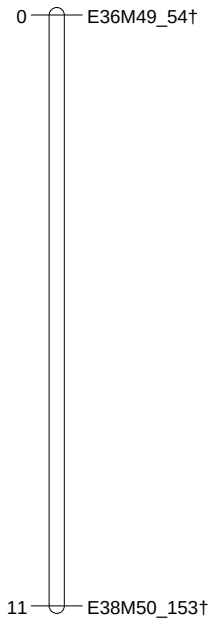
CG310- 74



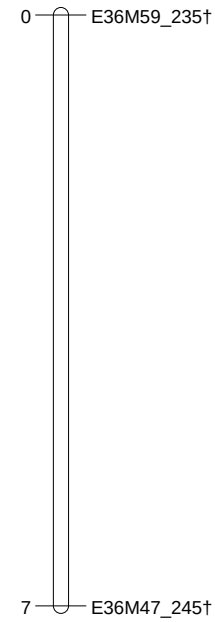
CG310- 75



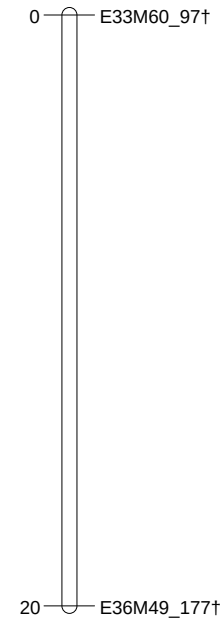
CG310- 76



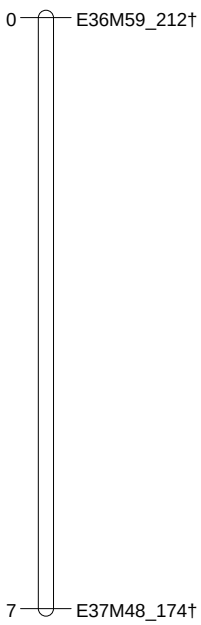
CG310- 77



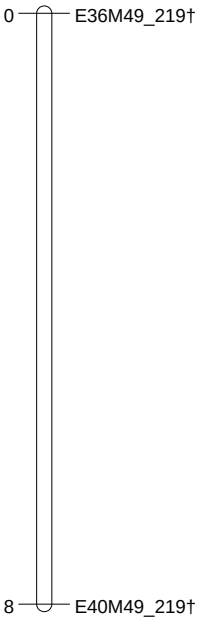
CG310- 78



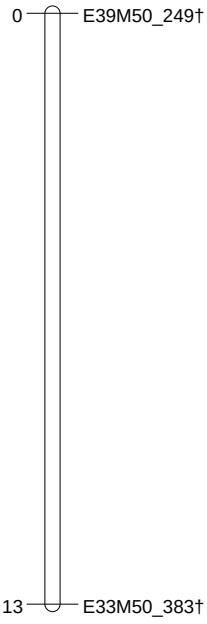
CG310- 79



CG310- 80



CG310- 81



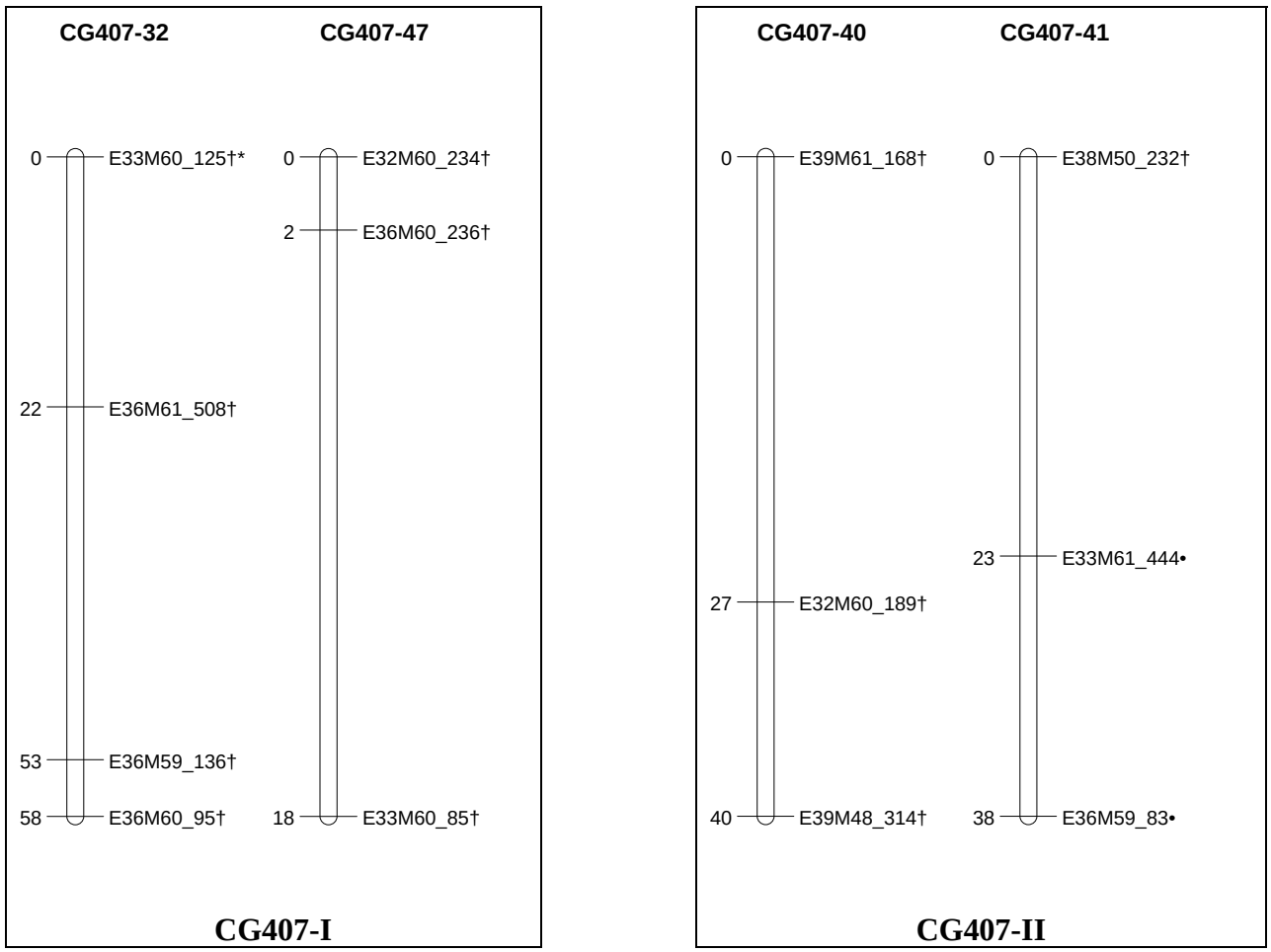


Figure 4.4: Male (CP 77-407) parent linkage map of the sugarcane hybrid 'LCP 85-384' based on S_1 mapping population. The map was constructed with a LOD score > 4.0 and a recombination fraction of 0.44 using AFLP markers. A total of 397 single dose markers (3:1) were assigned onto 80 CGs. The vertical bars indicate CGs with markers in coupling phase linkages. The Kosambi map distances (cM) and marker names are indicated on the left and right sides, respectively, of each CG. AFLP markers are denoted by 'EM'. CGs were grouped into HG based on S_1 map information. CGs of the same HG are enclosed in boxes and HGs are Roman numbered. The CP 77-407 specific markers are represented by one dot (·). The markers present in both parents and segregated in the S_1 population are denoted by the † symbol. The marker names with an asterisk (*) represent distorted markers. Independent CGs (do not belong to any HG) are not represented in boxes.

FIGURE 4.4 continued

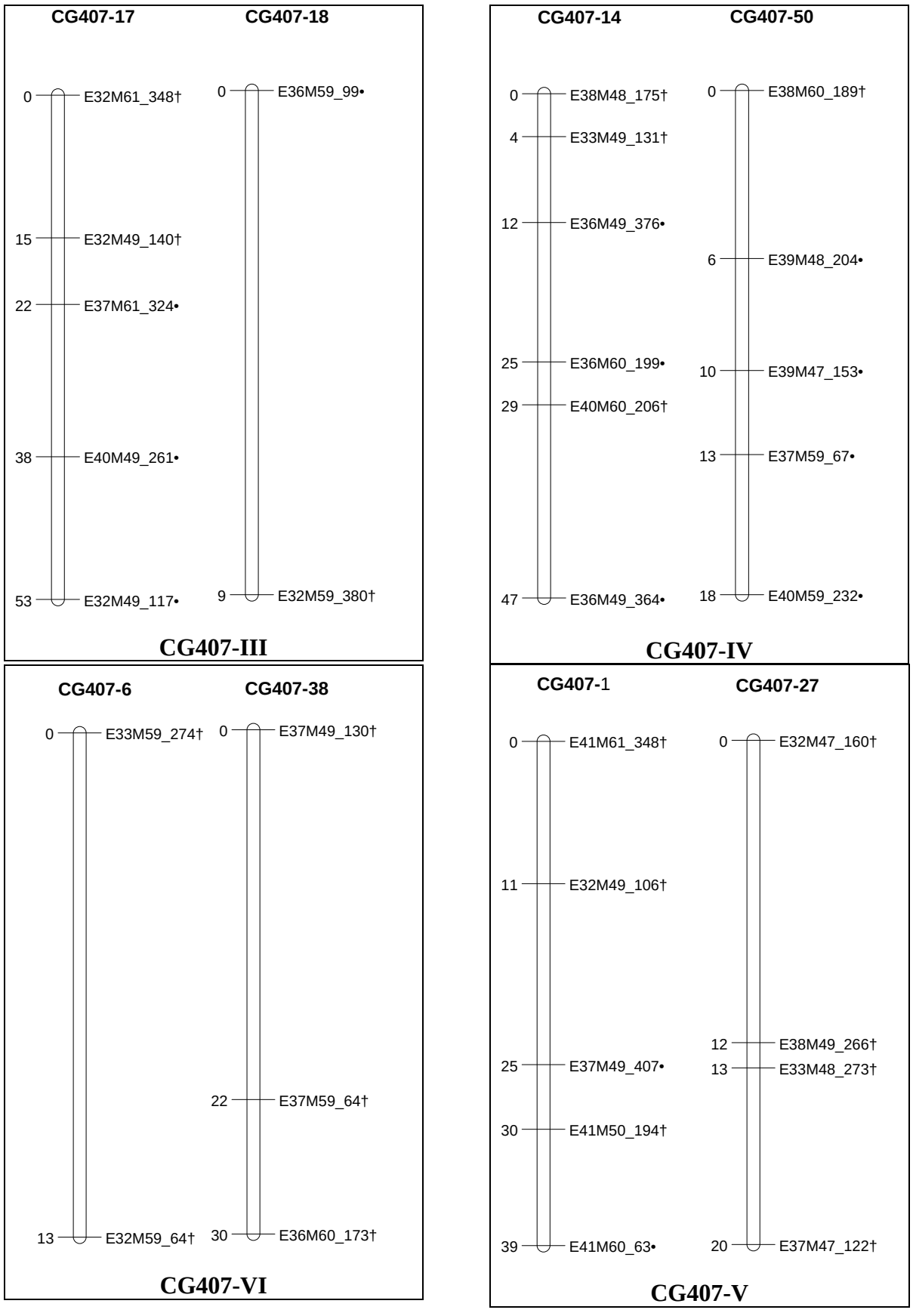


FIGURE 4.4 continued

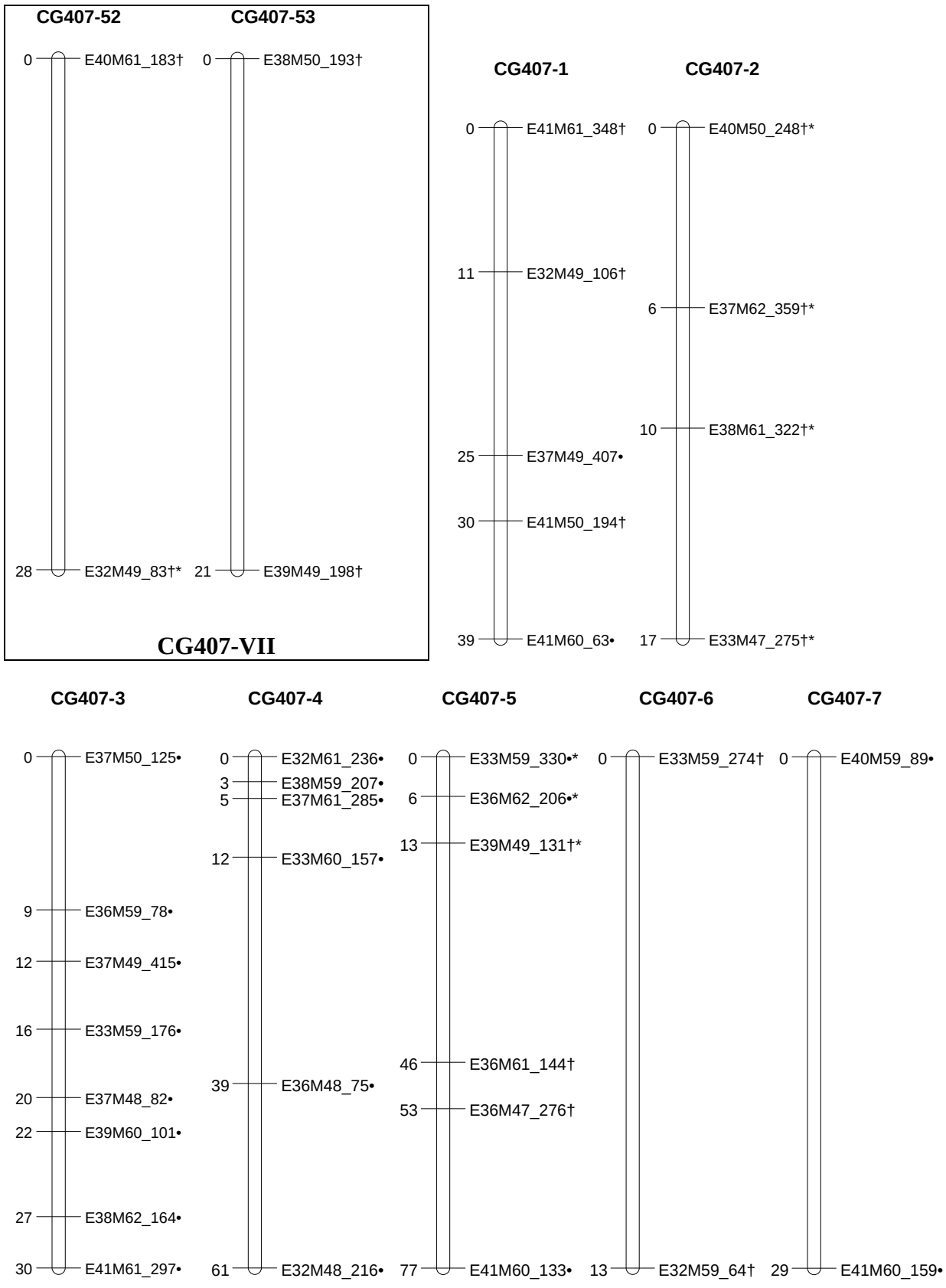


FIGURE 4.4 continued

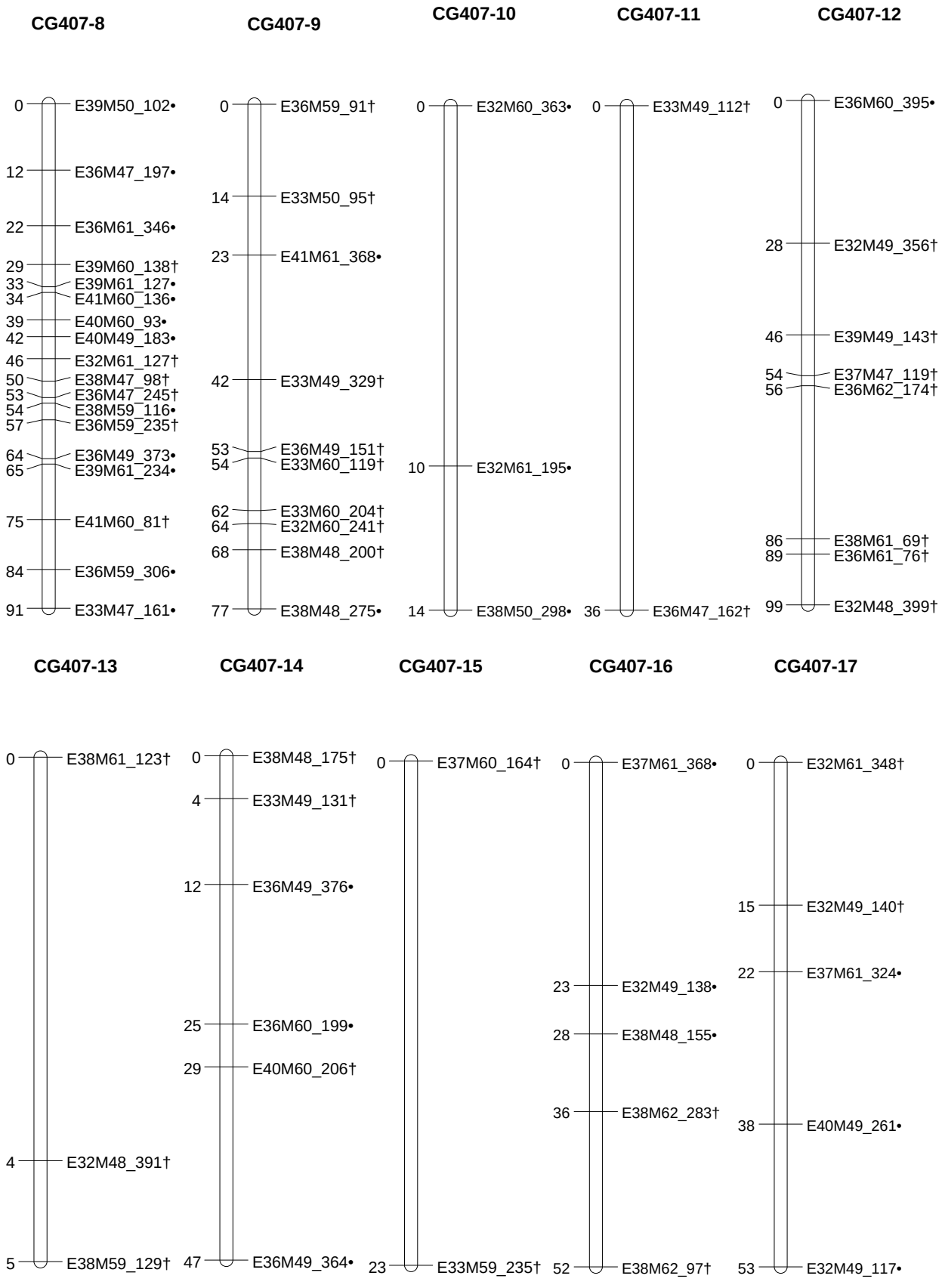


FIGURE 4.4 continued

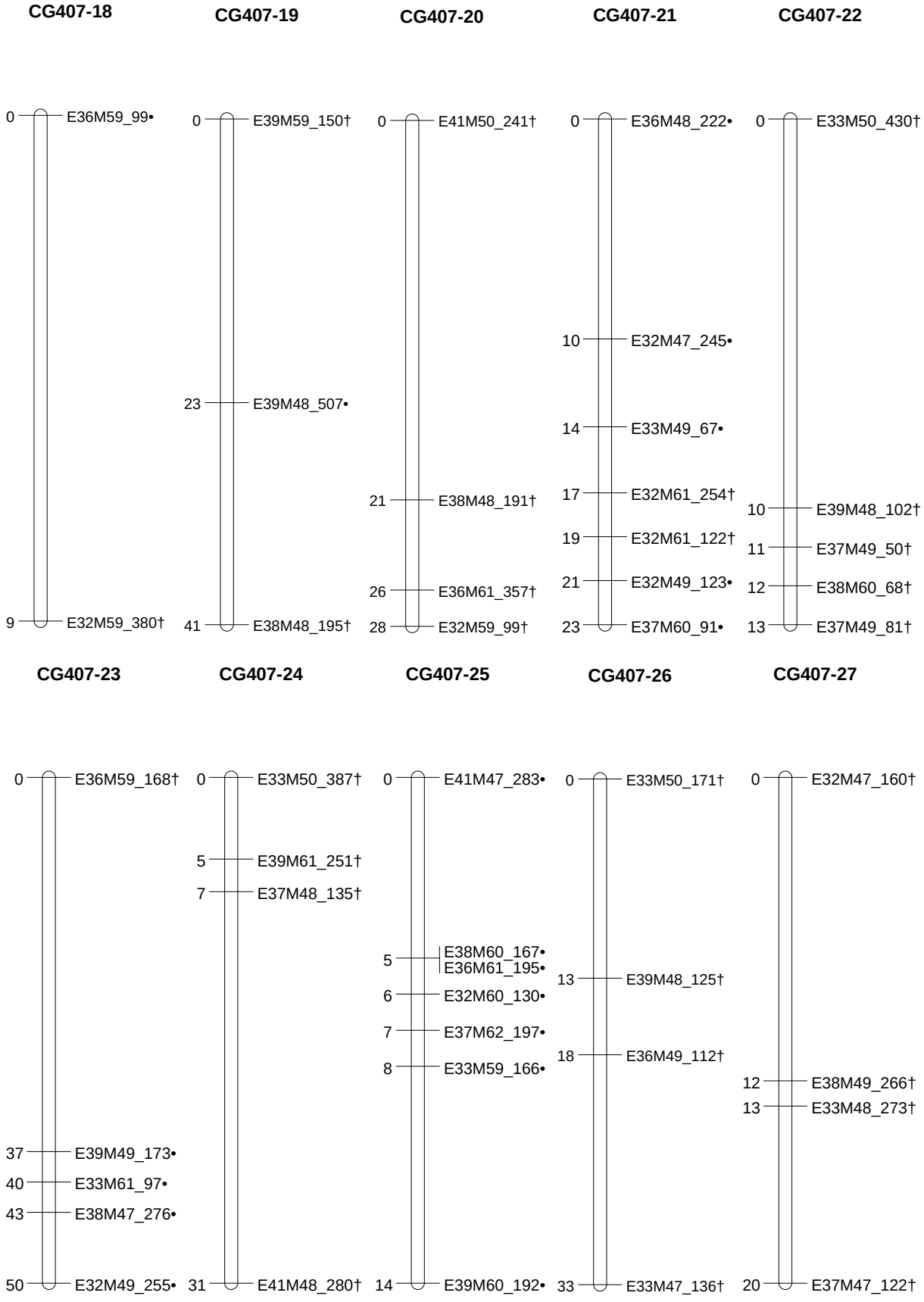


FIGURE 4.4 continued

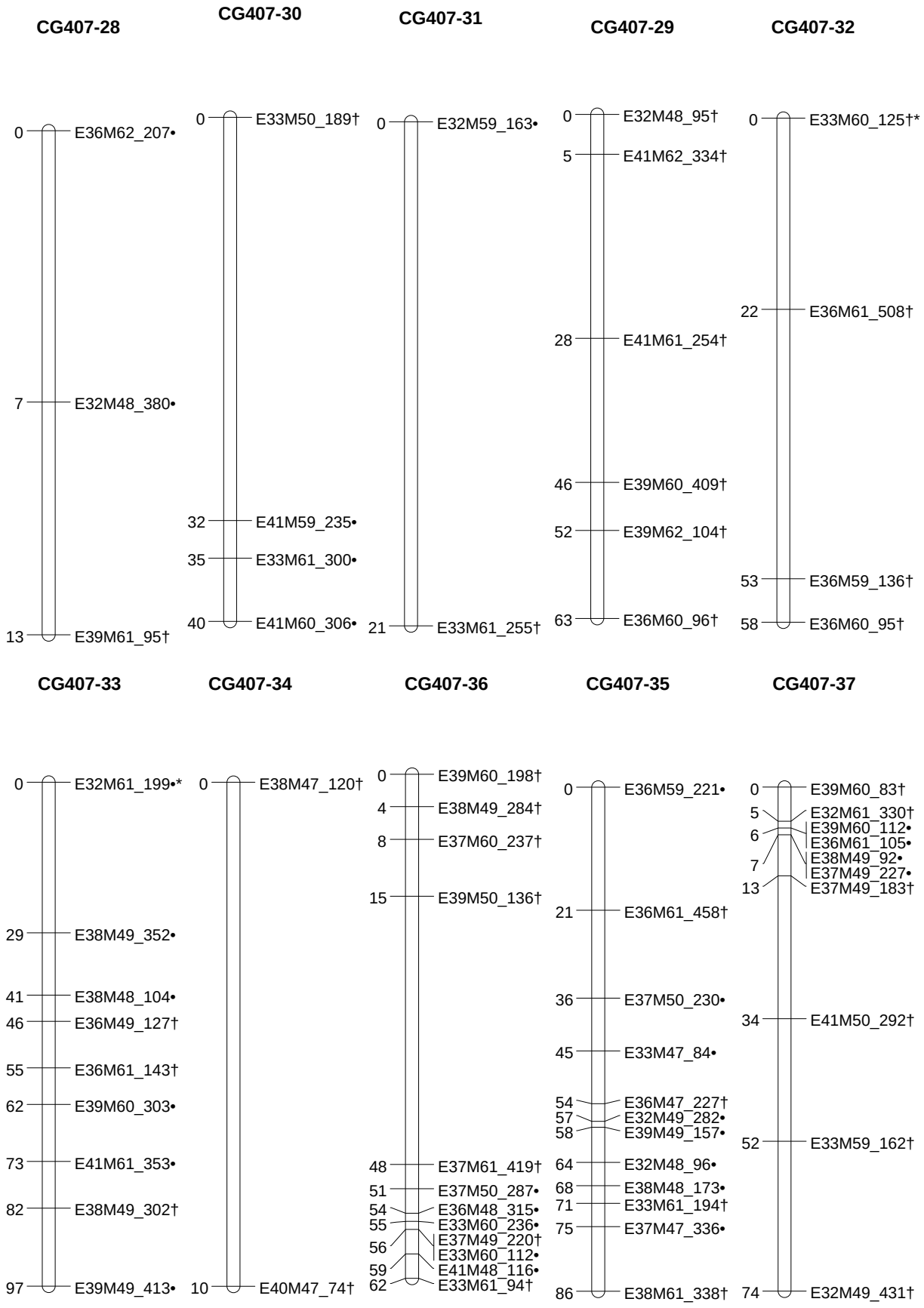


FIGURE 4.4 continued

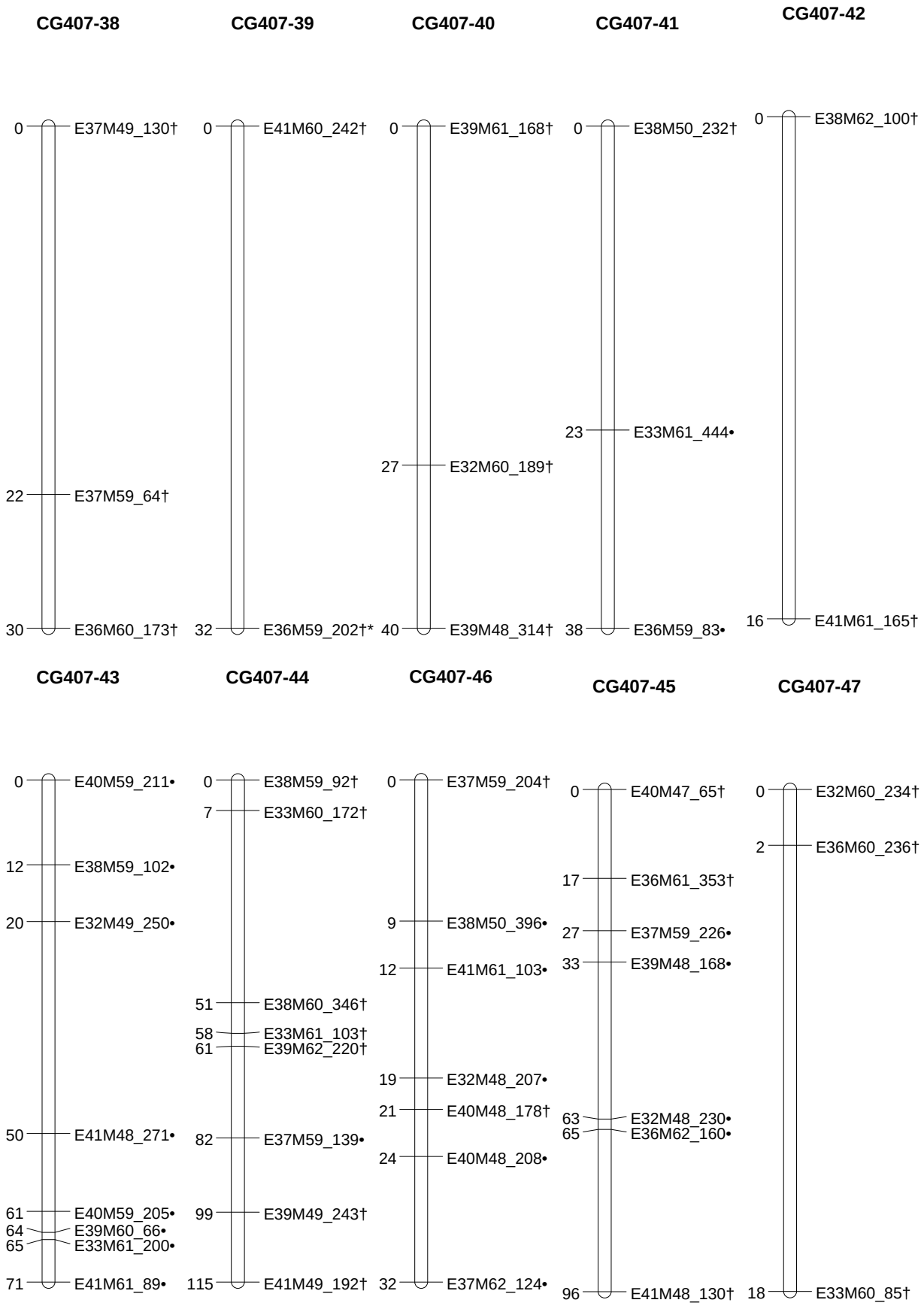


FIGURE 4.4 continued

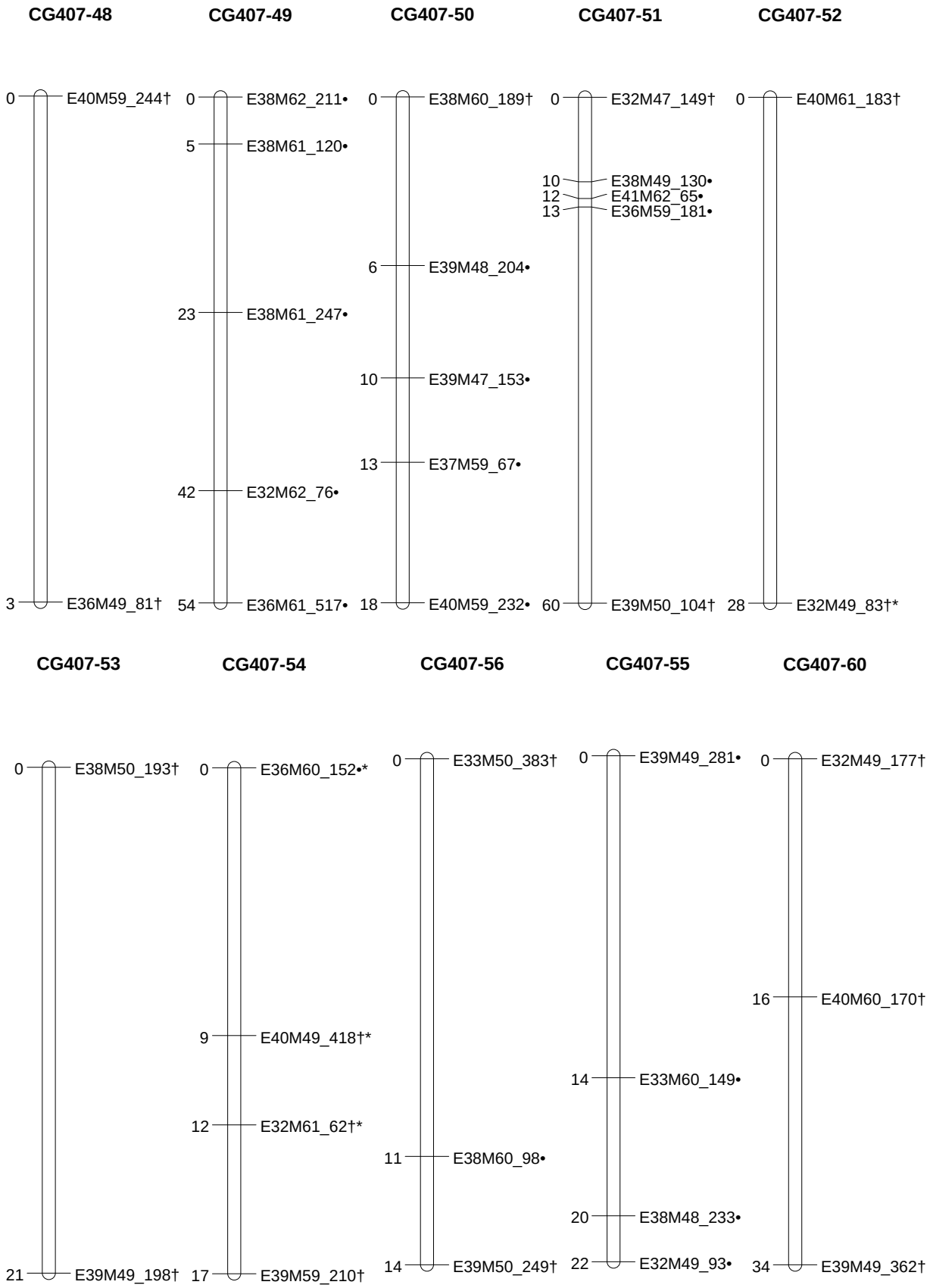


FIGURE 4.4 continued

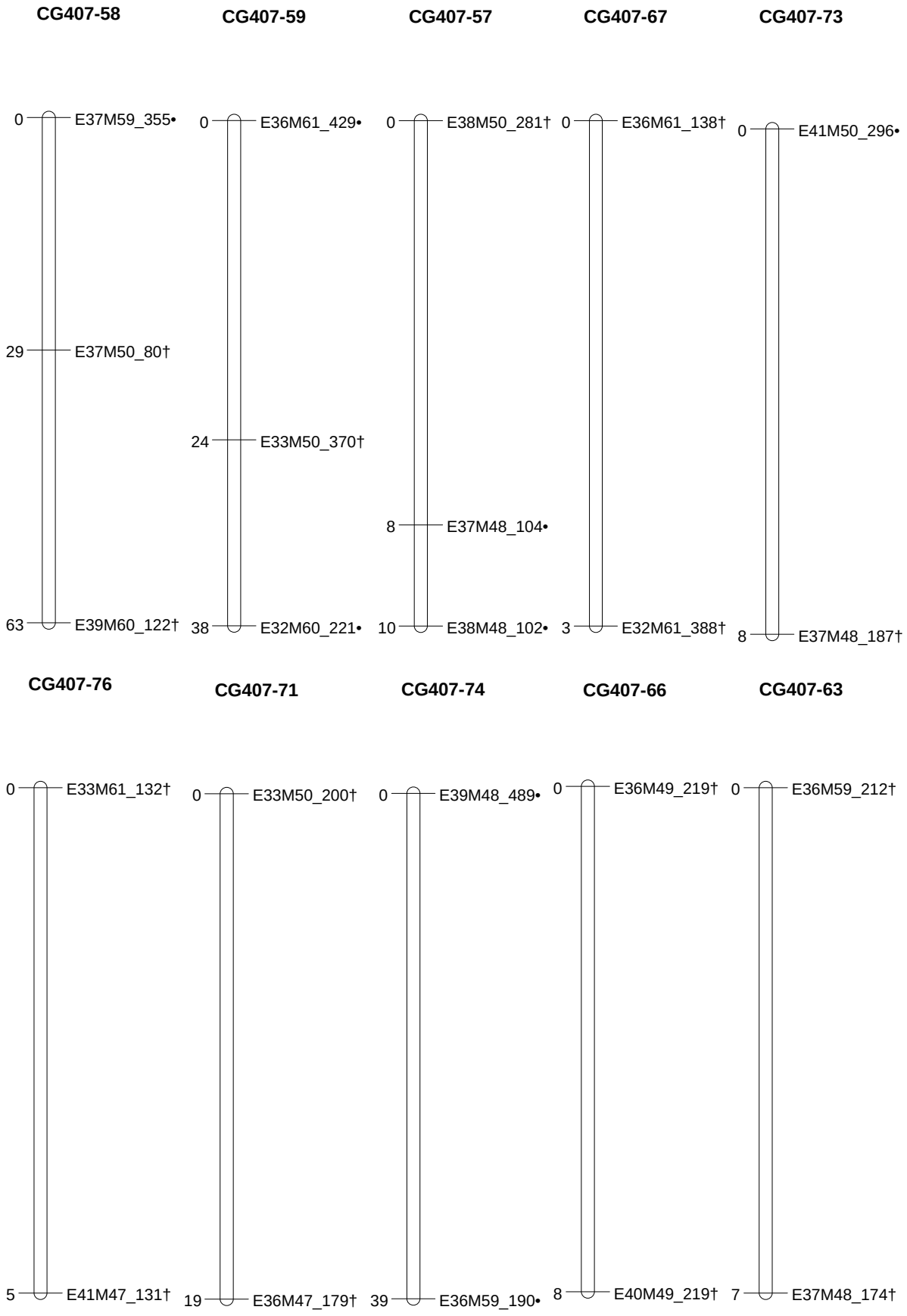


FIGURE 4.4 continued

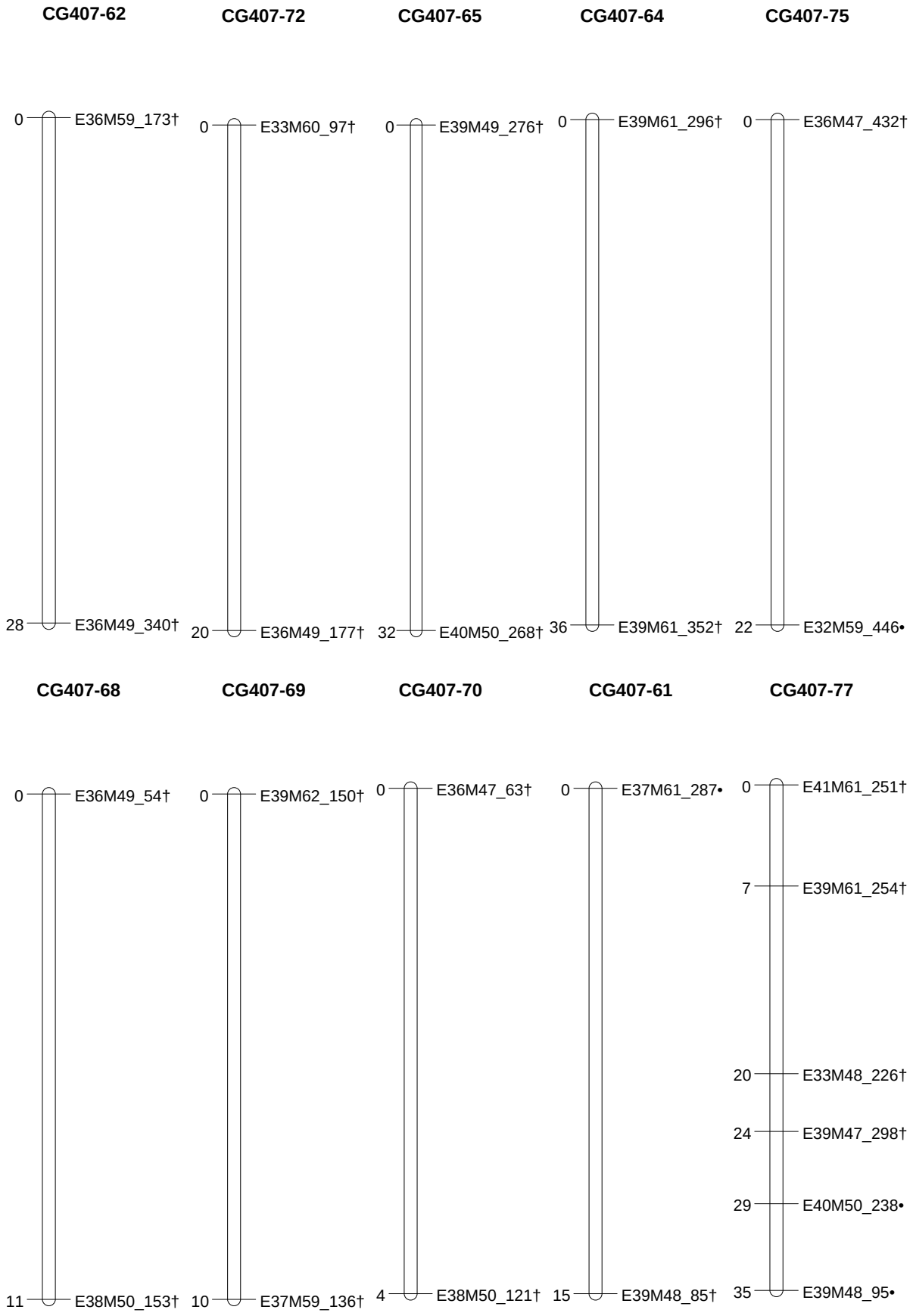
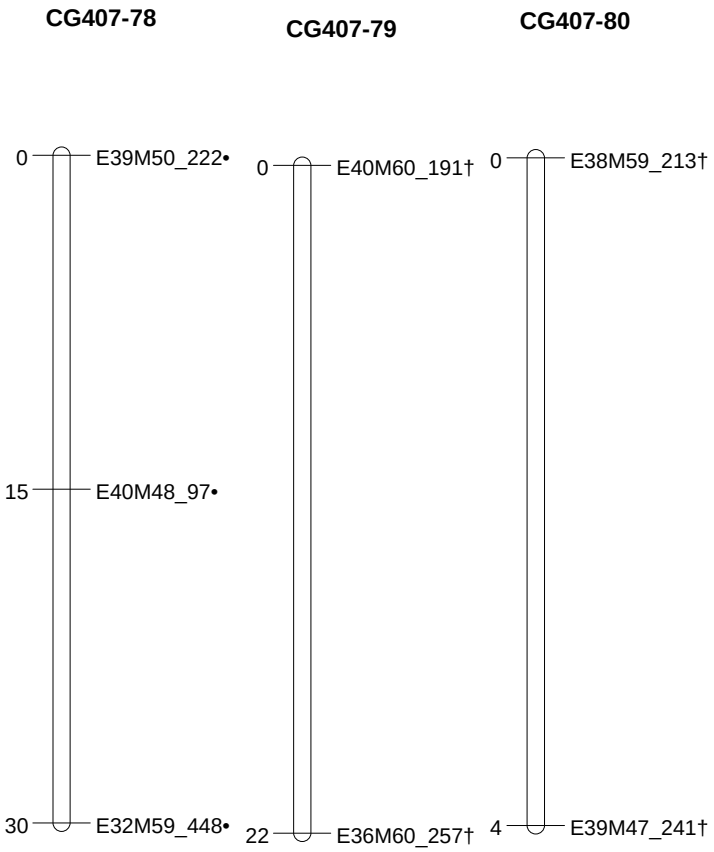


FIGURE 4.4 continued



Indeed, they were either part of a large CG or belong to a HG. From the final 108 CGs, a total of 31 CGs were assembled into 12 putative HGs using a total of 10 SSR loci (20 alleles), 5 CGs of 'CP 77-310' and 6 CGs of 'CP 77-407'. HG-1 was the biggest group with six CGs, where as the remaining of the HGs contained either two or three CGs. The CGs (CG-4, -5, 12, -23, and 40) in HG-I shared three SSR alleles from the locus 18SA, and two SSR alleles derived from the locus, 1604SA. However, a maximum of three loci (278CS, 336BS, and 31CUQ) which had consistent genomic position in CG-11 and CG-12 were found to have grouped in HG-II. A maximum of three CGs (CG-31, -32, and -33) were formed into HG-XII based on the CG310- 39 (Figure 4.2). The remaining CGs in the LCP 85-384 map either had no SSR markers in common or CGs from the parental maps, were considered to be independent groups (Table 4.5).

CGs in the parental linkage maps were also arranged into putative HGs using information from the LCP 85-384 map. In the 'CP 77-407' linkage map, a total of seven HGs were found

based on 7 CGs of the LCP 85-384 linkage map. In contrast, four HGs were found in the 'CP 77-310' using the information on three CGs of LCP 85-384linkage map. In both parental HGs, a maximum of two chromosomes were found in each HG. The remaining CGs in both parental maps were considered to be independent groups (Table 4.6).

4.3.4 Chromosomal Paring, Genome Size, and Genome Coverage

Repulsion phase linkages were found in CGs of LCP 85-384 and its parents. All together, recombination frequency estimates were computed for 597,529 (773*773), 193,600 (440*440), and 15,609 (397*397) pair-wise combinations for the LCP85-834, CP 77-310, and CP 77-407 linkage maps, respectively. A total of 4265:4306, 1639:1539, and 1362:1302 coupling to repulsion linkages were detected and confirmed the 1:1 ratio in LCP 85-384, CP 77-310 and CP 77-407 linkage maps, respectively. Moreover, within each CG of the LCP 85-384 and parental maps, the markers in C-C and R-R phase linkages confirmed the ratio 1:1(χ^2 at 0.05), which supports the prevalence of disomic inheritance (preferential pairing). Out of 108 CGs in the LCP 85-384 map, a total of 42 were of CP 77-310, and 44 of CP 77-310 chromosomal origin and 9 were recombinant. Thirteen CGs could not be assigned with a parental specific chromosomal origin probably due to lack of sufficient markers.

Approximate genome sizes of LCP 85-384 and its parents (CP 77-310 x CP 77-407) were estimated using the method followed by Aitken et al. (2005) and Hoarau et al. (2001). The estimated genome size of LCP 85-384, 106 x 120 =12720 cM, was obtained by multiplying the chromosome number (106) with the average size of the longest CGs (120 cM). Likewise, the estimated genome sizes of CP 77-310 and CP 77-407 were 14950 cM (115 x 130 cM), and 11500 cM (115 x 100 cM), respectively. The ratio between the cumulative genome length and estimated genome length indicated that approximately 42% (5384/12720), 23.2 % (3476/14950), and 24.1 % (2777/11500) of the LCP 85-384, CP 77-310, and CP77-410 genomes, respectively, have been covered in this study.

4.4 Discussion

Sugarcane is a highly heterozygous and genetically complex polyploid species, which shows severe inbreeding depression. For this reason, almost all published sugarcane linkage maps have been developed by application of the pseudo-testcross strategy on the full-sib populations using single dose markers (Garcia *et al.,* 2006). However, sugarcane genetic maps founded on the S_1 population of 'R570' were also developed using RFLP and AFLP markers to unravel genomic contributions by respective ancestral species (Grivet et al., 1996 and Hoarau et al., 2001). In this study, we report a reference framework genetic linkage map of Louisiana's popular cultivar 'LCP 85-384' using 300 S_1 progeny based on AFLP, TRAP and SSR markers.

In comparison with diploids, sugarcane has a large genome and generally requires a large number of progeny and markers to construct genetic linkage maps. The population size of 300 S_1 progeny used in this study is comparable to that used in other mapping efforts in sugarcane that used 295 S_1 progenies (Hoarau et al.,2001). High reliability in estimating useful genetic distances was assured in this study by using only SD markers because they are the most informative (Wu et al., 1992; Grivet et al., 1996; Hoarau et al., 2001). Single dose markers are abundant in polyploids and usually make up about 70% of polymorphic loci detected in sugarcane mapping studies (Da Silva et al., 1996; Hoarau et al., 2001; Aitken et al., 2005; Garica et al., 2006; Alwala et al., 2008). These results are in agreement with the expected number of SD markers (69.45%) found in this study. However, the total number of SD markers amplified in the present study is low compared to the SD markers found (939) by Hoarau et al. (2001) despite the similar population sizes were used. Several reasons including differences in the level of heterozygosity and chromosome numbers in the mapped hybrids could account for this difference. Also Hoarau et al. (2001) used the silver staining technique to visualize gels while in this study the gel images from the LiCor analyzer were saved onto a computer and scored manually. In our hands, we have been able to identify unambiguous bands ranging in size from 50 to 700 base pairs on silver

stained gels but not from the images saved from the LiCor Analyzer. A great number of bands beyond 550 base pairs obtained from the LiCor Analyser were unresolvable and therefore were not scored.

It is a common practice to develop parental maps using the pseudo-test cross strategy in polyploids to identify the origin of parental markers in F_1 (Grattapaglia et al., 1994; Edme et al., 2006; Alwala et al., 2008). Male and female parents undergo meiosis and produce gametes independently. Random union of such gametes constitutes an individual. Therefore, each parental data can be considered as independent (Kreigner et al., 2003). Using the same rationale,

Table 4.5: Summary of HGs detected in S_1 map based on SSRs, female (CP 77-310) and male (CP 77-407) grandparental maps.

Homologous groups in S_1 linkage map	**Based on**			**CGs of S1**
	SSRs	**CP 77-310**	**CP 77-407**	
HG-I	1604SA, 18SA	CG310-12	CG407-8	CG-4, -5, -9, -12, -23, -40
HG-II	336BS, 31CUQ	-	-	CG-10, -11
HG-III	CIR29, CIR66	-	-	CG-27, -29, -75,
HG-IV	1751CL	-	-	CG-44, -49
HG-V	CIR19, 119CG	-	CG407-36	CG-70, -74, -97
HG-VI	-	-	CG407-35	CG-50, -55
HG-VII	-	CG310-45	CG407-37	CG-76, -77
HG-VIII	-	-	CG407-44	CG-52, -56
HG-IX	-	-	CG407-62	CG-78, -79
HG-X	-	CG310-7	-	CG-13, -14
HG-XI	-	CG310-27	-	CG-20, -21
HG-XII	-	CG310-39	-	CG-31, -32, -33

Table 4.6: Summary of HGs detected in female (CP 77-310) and male (CP 77-310) parental maps based on LCP 85-384 map.

CP 77-407		
HGs	**Based on S_1**	**CGs of CP 77-407**
HG407-I	CG-1	CG407-32, -47
HG407-II	CG-10	CG407-40, -41
HG407-III	CG-27	CG407-17, -18
HG407-IV	CG-39	CG407-14, -50
HG407-V	CG-44	CG407-6,- 38
HG407-VI	CG-49	CG407-1, -27
HG407-VII	CG-80	CG407-52, -53
CP 77-310		
HGs	**Based on S_1**	**CGs of CP 77-310**
HG310-I	CG-2	CG310-2, -26
HG310-II	CG-18	CG310-10,11
HG310-III	CG-52	CG310-21,24
HG310-IV	CG-83	CG310-49,73

we have developed two independent parental maps for CP 77-310 and CP 77-407, the female and male parents of LCP 85-384, respectively. Compared to the parental specific markers present in '*ao x oo*' and '*oo* x *ao*' configuration, the high number of markers (230) present in both the grand-parents ('*ao* x *ao*') indicates the common ancestry shared by the parents (Garcia et al.2006).

4.4.1 Linkage Map Construction, Genome Coverage, and Segregation Distortion

4.4.1.1 LCP 85-384 and Parental Maps

LOD scores and upper recombination threshold generally determine the number of CGs present in a linkage map. Genetic linkage mapping in sugarcane has used LOD scores of ≥ 3.0 and recombination fraction values ranging between 0.25- 0.45 (Grivet et al., 1996; Al-Janabi et

al., 1993; Da Silva et al., 1993; Alwala et al., 2008). A maximum detectable recombination threshold of 0.44 and LOD score values of ≥ 4.0 (one allowed error in 10,000 linkages) were used in this study to avoid spurious linkages and give high confidence in the map by avoiding spurious linkages. However, the maximum detectable recombination generally depends on the size of the mapping population. The S_1-based linkage map of the commercial hybrid 'R570' reported by Grivet et al. (1996) and Hoarau et al. (2001) contained 96 CGs spanning 2,008 cM and 120 CGs spanning 5,849 cM, respectively. Compared to Hoarau et al. (2001) study, 108 CGs covered a cumulative map length of 5,384 cM in the current study. The number of CGs observed in the LCP 85-384 map was close to the expected number of chromosome number in LCP 85-384 (2n=106). Garcia et al. (2006) and Hoarau et al. (2001), obtained similar results. In contrast, the parental maps, 'CP 77-310' map had 81 CGs spanning 3476 cM and CP 77-407 had 80 CGs covering 2777cM. The number of CGs observed in the grandparents was not close to the expected chromosome number of CP hybrids (115-116; Edme, personal communication). Similar results have been observed in many biparental maps developed using interspecific F_1 mapping populations (Mudge et al., 1996; Ming et al., 1998; Edme et al., 2006).

The linkage maps of LCP 85-384, female and male parental maps were not saturated and covered only 42%, 23%, and 24% of the genome, respectively. The S_1 -based linkage map of 'R570' covered only 33% of the genome (Hoarau et al., 2001). The S_1 –based linkage map of LCP 85-384 could have achieved more genome coverage due to the smaller estimated genome of 'LCP 85-384' compared to 'R570' (Hoarau et al., 2001).

The linkage maps in the current study have uneven marker distribution along the CGs. Other studies have shown that the *S. spontaneum* portion of the genome is better mapped when compared to the *S. officinarum* portion of the genome (Lu et al., 1994; Grivet et al., 1996). For this reason, some of the CGs were probably dense as both *S. officinarum* and *S. spontaneum* chromosomes are expected in LCP 85-384. The uneven marker distribution on the linkage maps

could also be due to the use of only SD markers (by discarding the multiple dose markers) in coupling phase linkages (Ming et al., 1998; Garcia et al., 2006). Unsaturated linkage maps in the current study were also evident by the high number of unlinked markers coupled with short CGs (those with less than three markers per CG). A comparable number of unlinked markers to the present study were reported by Aitken et al. (2005) and Hoarau et al. (2001). In contrast, a significantly high number of unlinked markers was reported by Garcia et al. (2006) and Alwala et al. (2008) while mapping an F_1 population containing 100 individuals. The reason for this contrast could be the high number of progeny (300) and type of the population (S_1) used in the current study. Several short CGs, which may actually be part of larger CGs, could be a consequence of using the higher LOD values (≥ 4.0) while developing the linkage map (Alwala et al., 2008). However, spurious linkages were avoided by adopting the higher LOD values. Besides, gaps in sugarcane maps are expected because LCP 85-384 and its parents are complex poly-aneuploids and have a huge genome, which needs a large number of markers to saturate their linkage maps (Al-Janabi et al., 1993; Garcia et al., 2006). More markers are needed on the framework map for it to be saturated and to make it amenable to QTL discovery.

4.4.1.2 Segregation Distortion

Our segregation analysis of 773 SD markers for theoretical Mendelian ratio (3:1) revealed segregation distortions (~ 4.5% after the Bonferroni correction procedure), which could be reflective of genome disparities present in LCP 85-384 and its parents. Using a similar type of population, Grivet et al. (1996) and Hoarau et al. (2001) reported 2% and 8% distorted markers, respectively. Segregation distortion is an indication of divergence among the parents (Tanksley and Nelson, 1996). Interspecific hybrids derived from divergent parents are fit in F_1 and have a general tendency of exhibiting a high proportion of distorted markers upon selfing. Segregation distortion may also be more elevated in hybrid genomes within regions experiencing selection (Woram et al., 2004). LCP 85-384 originates from a cross between the CP 77-310 and CP 77-

407. The female parent of LCP 85-384 (CP 77-310) originates from the cross between CP52-068 and L65-69. The male parent 'CP 77-407' (CP71-021 x CP66-035) is a BC3 of *S. spontaneum* genotype 'US56-15-8'. Thus, the original grandparents of the S_1 segregating population came from very distant parents, which could have an impact on the segregation distortion observed. On the other hand, occurrence of chromosomal rearrangements between parental karyotypes would be the most classical way of explaining the segregation distortion. Deviations in expected ratios could also be related to the significantly different sizes and chromosome number of the parental genomes (Edme et al., 2006; Alwala et al., 2008).

Segregation distortion is also an indication of the linkage between molecular markers and distorting factors (deleterious recessive alleles) (Zamir and Tadmor, 1986). If the linkage is tight, they usually have similar segregation patterns and thus, skewed markers would appear to be clustered (Jenczewski et al., 1997). However, the distorted markers in this study did not show extensive clustering, which signifies a lower level of inbreeding depression. Similarly, in grapevine S_1 (Hvarleva et al., 2009) map, distorted markers did not cluster together. The clustering of distorted markers may not necessarily imply linkage, but linkage disequilibrium could be suspected (Jenczewski et al., 1997). Furthermore, the amplification of two fragments of the same length from non-allelic regions (homoplasy) and co-migration of two different fragments amplified at paralogus loci could also be responsible for some of the markers showing segregation distortion. However, more markers and larger population sizes would be required to identify distorted loci and determine their likely biological significance in sugarcane (Alwala et al., 2008).

4.4.2 Homo(eo)logous Groups (HGs)

Given the basic chromosome number of *Saccharum* species (~ x=10), it is predicted that 10-11 homo(eo)logus sets of chromosomes should exist within the genome if homo(eo)logus chromosomes are preserved during the diplodization process in meiosis. Previous studies have

identified 10 HGs using the S_1 population in the 'R570' map based on the joint segregation of RFLP (Grivet et al., 1996) and AFLP (Hoarau et al., 2001) markers. In the current study, we have found 12 HGs, which is slightly higher than the predicted number of HGs. However, only 7 and 4 HGs were found in the female and male parents of LCP 85-384, respectively, which is lower than the expected basic chromosome number for the genus *Saccharum*. It is apparent that (Figure 4.2, 4.3, and 4.4) HGs were under represented by CGs and many CGs were not identified as belonging to a homo(eo)logous group probably due to the lack of sufficient SSR markers. Most of the HGs were also formed with small and less dense CGs. Low levels of polymorphism in the regions of the *S. officinarum* part of the genome could be contributing to the lack of denser CGs (Ming et al., 1998; Hoaruau et al., 2001; Aitken et al., 2005). The number of CGs per HG as well as marker density will be enhanced upon including more locus specific markers. Accordingly, the HG number may be achieved closer to the basic chromosome number of *Saccharum* species.

Because of the sugarcane's double genome ancestry (D'Hont et al., 1998), CGs in the linkage map share homeologous affinities as well as homologous affinities to each other. Crossing over in homologues during meiosis plays an important role in diversifying of homologous regions (Wright et al., 1983). Thus, the highest levels of homologue affinity are expected to be found in regions that have the greatest degree of crossing over (at telomeric segments of chromosomes). In the current study, all the SSR loci which are responsible for the HG grouping appeared to group at telomeric positions. We did not find evidence of genome duplication due to limited information on SSR markers.

4.4.3 Chromosome Segregation

In the present study, probable disomic inheritance (allopolyploidy) was detected using the ratio between C-C and R-R phase linkages in both LCP 85-384 and its parents. We found 1:1 repulsion to coupling phase linkages at high LOD scores (> 3.0) which gave little chance for

artifacts. However, Grivet et al. (1996) and Hoarau et al. (2001) found evidence of partial preferential pairing of chromosomes in the hybrid 'R570' at high LOD scores. In the current study, the pairing of CGs in each HG could not be completely identified because of very few CGs within the each HG. In addition, most of the CGs in LPC85-384 and its parental maps were not covered with markers. Nonetheless, adding more markers to the maps will provide greater insight into the issue. Previous studies in sugarcane (D'Hont et al., 1994) have revealed very low rate of recombination (<10%) between homo(eo)logous chromosomes and suggested *in toto* transmission of chromosomes. Similarly, the current study found only 9 (~ 8%) recombinant CGs out of 108.

4.5 Future Studies

The S_1-based map of LCP 85-384 in the present study is not saturated and is based on 773 SD markers generated by 64 AFLP, 12 TRAP and 19 SSR primer pairs. The map covered only 43% of the genome and indeed, none of the published genetic maps of sugarcane are saturated. The main reason attributed to this is the genetic complexity of sugarcane. The density of the LCP 85-384 map in the current study is 7.5 cM per marker. However, gaps still exist. Filling such large gaps in the map, which could be from the less polymorphic *S. officinarum* fraction of the genome, will be an enormous task despite the rapid improvement of available marker technologies. However, more markers need to be added to the current map to make the map more resourceful in finding QTLs for several agronomic traits.

Most of the diversity found among modern cultivars was reportedly from the *S. spontaneum* genome, probably because *S. officinarum* was used as the recurrent parent during nobilization and transmitted 2n gametes to its progeny. In Louisiana, a few *S. spontaneum* genotypes have been used extensively to develop new cultivars. One such *S. spontaneum* genotype is US56-15-8 (Thailand origin), which is found in the pedigree of most of the popular cultivars (LCP 85-384, HoCP85-845, L97-128, and HoCP96-540) in Louisiana (Arro *et al*.,

2006). For that reason, it would be vital to estimate the genomic contribution of the *S. spontaneum* clone 'US56-15-8' and other ancestral clones (*S. officinarum* and legendary cultivars) to LCP 85-384. Tracking the source of each allele (from the ancestry) and eventually to efficiently identify and tag the markers in successful cultivars could be one way of mapping useful alleles in sugarcane.

4.6 Summary

In the current study, LCP 85-384 and its female (CP 77-310) and male (CP 77-407) parental framework linkage maps were developed using AFLP, TRAP and SSR markers. The current maps allowed us to study the segregation pattern in the mapping population and chromosome pairing during meiosis. LCP 85-384 possesses desirable agronomic traits and resistance to biotic and abitoic stresses. The established framework S_1 map in this study will provide an important background for mapping QTLs associated with sugar related traits and thus, information will be useful for crossing and in selecting the clones in the breeding program.

4.7 References

Aitken KS, Jackson PA, and McIntyre CL (2007) Construction of genetic linkage map for *Saccharum officinarum* incorporating both simplex and duplex markers to increase genome coverage. *Genome* 50: 742-756.

Aitken KS, Jackson PA, and McIntyre CL (2005) A combination of AFLP and SSR markers provide extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. *Theoretical and Applied Genetics* 110: 789-801.

Al-Janabi SM, Honeycutt RJ, McClelland M, and Sobral BWS (1993) A genetic linkage map of *Saccharum spontaneum* L. 'SES 208'. *Genetics* 134: 1249-1260.

Alwala S, Suman A, Arro JA, Veremis JC, and Kimbeng CA (2006) Target Region Amplification Polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections. *Crop Science* 46: 448-455.

Alwala S, Kimbeng CA, Veremis JC, and Gravois KA (2008) Linkage mapping and genome analysis in *Saccharum* interspecific cross using AFLP, SRAP and TRAP markers. *Euphytica* 164: 37–51.

Arro JA, Veremis JC, Kimbeng CA and Botanga C (2006) Genetic diversity and relationships

revealed by AFLP among a collection of *Saccharum spontaneum* and related species and genera**.** *Journal of the American Society of Sugar Cane Technologists* 26: 101-115.

Asnaghi C, Paulet F, Kaye C, Grivet L, Deu M, Glaszmann JC, and D'Hont A (2000) Application of synteny across Poaceae to determine the map location of a sugarcane rust resistance gene. *Theoretical and Applied Genetics* 101: 962-969.

Asnaghi C, Roques D, Ruffel S, Kaye C, Hoarau JY, Te`lismart H, Girard JC, Raboin LM, Risterucci AM, Grivet L, and D'Hont A (2004) Targeted mapping of a sugarcane rust resistance gene (B*ru*1) using bulked segregant analysis and AFLP markers. *Theoretical and Applied Genetics* 108: 759-764.

Atienza SG, Satovic Z, Peterson KK, Dolstra O, and Martin A (2002) Preliminary genetic linkage map of *Miscanthus sinensis* with RAPD markers. *Theoretical and Applied Genetics* 105: 946-952.

Barnes JM and Bester AE (2000) Genetic mapping in sugarcane: prospects and progress in the South African sugar industry. *Proceedings of South African Sugar Technologists Association* : 74.

Bailey NTJ (1961) Introduction to Mathematical Theory of Genetic Linkage**.** Oxford University Press, London.

Bhat SR, and Gill BS (1985) The implication of 2n egg gametes in nobilisation and breeding of sugarcane. *Euphytica* 34:377-384.

Brandes EW (1958) Origin, classification and characteristics. In: Artschwager E, Brandes EW (eds) Sugarcane (*S. officinarum* L.). *USDA Agriculture Handbook*, 122: 1–35, 260-262.

Bremer G (1961) Problems in breeding and cytology of sugarcane. *Euphytica* 10:59-78.

Berding N, and Roach BT (1987) Germplasm collection, maintenance, and use. In: Heinz DJ (ed.) *Sugarcane Improvement Through Breeding*. Amsterdam: Elsevier, pp. 143–210.

Breaux RD, Fanguy HP, Matherne RJ, and Dunckelman PH (1974) Registration of 'CP 65-357' sugarcane. *Crop Science* 14:605.

Cordeiro GM, Taylor GO, Henry RJ (2000) Characterisation of microsatellite markers from sugarcane (*Saccharum* sp.), a highly polymorphic species. *Plant Science* 155:161–168

Cunff LL, Garsmeur O, Raboin LM, Pauquet J, Telismart H, Selvi A, Grivet L, Philippe R, Begum D, Deu M, Costet L, Wing R, Glaszmann JC, and D'Hot A (2008) Diploid/polyploidy syntenic shuttle mapping and haplotype-specific chromosome walking toward a rust reistance gene (*Bru1*) in highly poyploid sugarcane (2n ~ 12x~115). *Genetics* 180: 649-660.

D'Hont A, Lu YH, de León DG, Grivet L, Feldmann P, Lanaud C, and Glaszmann JC (1994) A molecular approach to unraveling the genetics of sugarcane, a complex polyploid of the Andropogoneae tribe. *Genome* 37(2): 222- 230.

D' Hont A, Grivet L, Feldmann P, Rao PS, Berding N, and Glazmann JC (1996) Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Molecular General Genetics* 250: 405-413.

D' Hont A, Ison D, Alix K, Roux C, and Glazmann JC (1998) Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. *Genome* 41: 221-225.

Daugrois JH, Grivet L, Roques D, Hoarau JY, Lombard H, Glaszmann JC, and D'Hont A (1996) Putative major gene for rust resistance linked with a RFLP marker in sugarcane cultivar 'R570'. *Theoretical and Applied Genetics* 92 (8): 1059-1064.

Da Silva JAG, Sorrells ME, Burnquist W, and Tanksley SD (1993) RFLP linkage map and genome analysis of *Saccharum spontaneum. Genome* 36: 782–791.

Edmé SJ, Glynn NG, and Comstock JC (2006) Genetic segregation of microsatellite markers in *Saccharum officinarum* and *S. spontaneum. Heredity* 97: 366-375.

Garcia AAF, Kido EA, Meza AN, Souza HMB, Pinto LR, Pastina MM, Leite CS, Da Silva JAG, Ulian EC, Figueira A, and Souza AP (2006) Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. *Theoretical and Applied Genetics* 112: 298-314.

Grattapaglia D, and Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* 137: 1121-1137.

Grivet L, D'Hont A, Roques D, Feldmann P, Lanaud C, and Glaszmann JC (1996) RFLP mapping in a highly polyploid and aneuploid interspecific hybrid. *Genetics* 142: 987-100.

Gravois KA and Bischoff KP (2008) New sugarcane varieties to the rescue. *Louisiana Agriculture* 51(2):14-16.

Guimaráes CT, Honeycutt RJ, Sills GR, and Sobral BWS (1999) Genetic linkage maps of *Saccharum officinarum* L. and *Saccharum robustum* Brandes & Jew. Ex Grassl. *Genetics and Molecular Biology* 22: 125-132.

Hvarleva TD, Russanov KE, Bakalova AT, Zhiponova MK, Djakova GJ, Atanassov AI, and Atanassov II (2009) Microsatellite linkage map based on F2 population from Bulgarian grapevine cultivar Storgozia. Biotechnology & Biotechnology. Eq. 23/2009/1.

Hoarau JY, Offmann B, D'Hont A. Risterucci AM, Roques D, Glaszmann JC, and Grivet L (2001) Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). I. Genome mapping with AFLP markers. *Theoretical and Applied Genetics* 103:84-97.

Hogarth DM (1987) Genetics of Sugarcane, in *Sugarcane Improvement Through Breeding*, edited by D.J. Heinz, pp. 255-271. New York: Elsevier.

Jenczewski E, Gherandi M, Bonnin I, Prosperi JM, Oliveri I, and Hugget T (1997) Insight on segregation distortions in two intraspecific crosses between annual species of Medicago (Leguminosae). *Theoretical and Applied Genetics* 94:682-691.

Lu YH, D'Hont A. Paulet F, Grivet L, Arnaud M, and Glaszmann JC (1994) Molecular diversity and genome structure in modern sugarcane varieties. *Euphytica* 78:217-226.

Mather K (1957) The measurement of linkage in heredity. John Wiley & Sons, Inc., New York.

Kriegner A, Cervantes JC, burg K, Mwanga ROM, and Zhang D (2003) A genetic linkage map of sweetpotato [*Ipomoea batatas* (L.) Lam.] based on AFLP markers. *Molecular Breeding* 11: 169-185.

Milligan, SB, Martin FA, Bischoff KP, Quebedeaux JP, Dufrene EO, Quebedeaux KL, Hoy JW, Reagan TE, Legendre BL, and Miller JD (1994) Registration of 'LCP 85-384' sugarcane. *Crop Science* 34:819–820.

Ming R, Liu S-C, Bowers JE, Moore PH, Irvine JE, and Paterson AH (2002) Construction of *Saccharum* consensus genetic map from two interspecific crosses. *Crop Science* 42: 570-583.

Ming R, Liu SC, Lin YR, Da Silva JAG, Wilson W, Braga D, van Devnze A, Wenslaff F, Wu KK, Moore PH, Burnquist W, Sorrells ME, Irvine JE, and Paterson AH (1998) Detailed alignment of *Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and polyploid genomes. *Genetics* 150:1663–1682.

Mudge J, Andersen WR, Kehrer RL, and Fairbanks DJ (1996) A RAPD genetic map of *Saccharum officinarum*. *Crop Science* 36:1362-1366.

Oliveira KM, Pinto LR, Marconi TG, Margarido GRA, Pastina MM, Teixeira LHM, Figueira AV, Ulian EC, Garcia AAF, and Souza AP (2007) Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross. *Molecular Breeding* 20: 189-208.

Pan Y, Cordeiro GM, Richard Jr EP, and Henry RJ (2003) Molecular genotyping of sugarcane clones with microsatellite markers. *Maydica* 48:319-329.

Pan YB (2006) Highly polymorphic microsatellite DNA markers for sugarcane germplasm evaluation and variety identity testing. *Sugar Technology* 8(4):246-256.

Raboin LM, Oliveira KM, Lecunff L, Telismart H, Roques D, Butterfield M, Hoarau JY, and D'Hont A (2006) genetic mapping in sugarcane, a high polyploidy, using bi-parental progeny: identification of a gene controlling stalk colour and a new rust resistance gene. *Theoretical and Applied Genetics* 112: 1382-1391.

Roach BT (1972) Nobilization of sugarcane. *Proceedings International Society of Sugarcane Technologists* 14:206–216.

Sokal RR and Rohlf FJ (1995) Biometry. W.H. Freeman and Co., New York.

Tanksley SD, and Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theoretical and Applied Genetics* 92: 191-203.

Van Ooijen JW, and Voorrips RE (2001) JOINMAP 3.0. Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands.

Wright JE, Johnson k, Hollister A, and May B (1983) Meiotic models to explain classical linkage, pseudolinkage, and chromosomal pairing in tetraploid derivative salmonid genomes. Isozymes Curr. *Topics in Biological and Medical Research* 10: 239-260.

Woram RA, McGowan C, Stout JA, Gharbi K, Ferguson MM, Hoyheim B, Davidson EA, Davidson WS, Rexraod C, and Danzmann (2004) A genetic linkage map for Arctic char (*Salvelinus alpinus*): evidence for higher recombination rates and segregation distortion in hybrid versus pure strain mapping parents. *Genome* 47: 304-315.

Wu KK, Burnquist W, Sorrells ME, Tew TL, Moore PH, and Tanksley SD (1992) The detection and estimation of linkage in polyploids using single-dose restriction fragments. *Theoretical and Applied Genetics* 83: 294-300.

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot L, Peleman J, Kuiper M, and Zabeau M (1995) AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Research* 23:4407–4414.

Vuylsteke M, Mank R, Brugmans B, Stam P and Kuiper M **(**2000) Further characterization of AFLP data as a tool in gen**et**ic diversity assessments among maize (*Zea mays* L.) inbred lines. *Molecular Breeding* 6: 265–276.

Zamir D and Tadmor Y (1986) Unequal segregation of nuclear genes in plants. Botany Gazette 147:355-358.

CHAPTER 5: SUMMARY

5.1 Evaluation of SRAP Markers

Characterization of wild germplasm provides essential information on genetic diversity that breeders utilize for crop improvement. The potential of the sequence-related amplified polymorphism (SRAP) technique, which preferentially amplifies gene-rich regions, was evaluated to assess the genetic relationships among members of the *Saccharum* species. A panel of 31 SRAP primer combinations were used to score 30 genotypes of *S. officinarum*, *S. spontaneum*, *S. robustum*, *S. sinense*, *S. barberi*, and sugarcane hybrids, with *Miscanthus* and *Erianthus* included as outgroups, The amplifications produced 1364 DNA fragments for an average of 44 per primer combination, with 83 percent (1135) being polymorphic, and 8.7 percent (119) being species-specific. Based on the Dice index, all 30 genotypes were differentiated from each other with genetic similarity (GS) estimates ranging from 0.60 to 0.96 (mean=0.79). Both the dendrogram (obtained by the unweighted pairgroup method with arithmetic mean or UPGMA) and the non-metric multi-dimensional scaling (NMDS) method grouped the genotypes according to their phylogenetic relationships. *Erianthus* and *Miscanthus* were separated as two outgroups (at GS levels of 0.56 and 0.72, respectively) to two major clusters: Cluster I separated the *S. robustum*, *S. sinense, S. barberi*, and hybrids as different subgroups with each one including some *S. officinarum* clones, while Cluster II included the *S. spontaneum* clones, exclusively. A *S. officinarum- S. spontaneum* sequence comparison of some of the monomorphic and polymorphic bands revealed 65 to 90 percent homology with rice, corn, or sugarcane sequences deposited in databases. The possibility that most of the amplicons may be amplifying gene-rich regions of the genome coupled with a high discriminatory power makes SRAP a potentially robust tool for genetic mapping aimed at marker-assisted selection in sugarcane.

5.2 Molecular Diversity of *S. spontaneum* Working Collection and Its Genomic Contribution to Modern Cultivars

S. spontaneuam has been the most important source of wild germplasm for sugarcane development in Louisiana due to its diversity in both habitat and chromosome number. *S. spontaneum* is the source of genes for both biotic and abiotic stresses. However, very few *S. spontaneum* genotypes from the local collection have been successfully used in the development of sugarcane cultivars in Louisiana after the basic breeding was initiated in 1964. Fifty-one *S. spontaneum* genotypes in the collection grouped largely according to their geographical origins namely Central (India) and East (Thailand) zones. The exceptions in the grouping were attributed to admixed genome, which could be due to gene flow between the genotypes in geographic proximity. Also the disproportionate representation of genotypes from different geographical regions within zones could be responsible for the discrepancies in the grouping of genotypes. For the first time in sugarcane, we used the model-based Bayesian clustering (MBBC) approach rather than regular distance-based clustering techniques (UPGMA-CA and PCoA) to study the diversity and structure. The results are largely congruent with those of the distance-based clustering techniques.

We assessed the percent contribution of *S. spontaneum* genotypes in both the old/legendary and modern cultivars based on the inferred ancestry of the each cultivar using the Structure analysis. The percent contribution of *S. spontaneum* genotypes in the local collection was very low. The main reason attributed to this result was the fact that the original *S. spontaneum* genotypes that participated in the nobilization event were not included in this study. But our analysis is appropriate relative to the working collection of *S. spontaneum* genotypes that are available in the working collection at the USDA, Sugarcane Research Unit, Houma, LA. The contribution of alleles from the *S. spontaneum* collection in the modern cultivars was also very low and about equal for the Central (India; cluster I) and East (Thailand; cluster III) zone with

the two clusters contributing about 0.34% and 0.42%, respectively. It is evident that the commercial breeding program in Louisiana has not tapped the diversity from *S. spontaneum* genotypes in the collection except from the genotype 'US56-15-8' and 'SES 147B'. The low proportion of shared alleles recorded in this study could also be a result of continuous crossing and selection for sucrose content among successful cultivars. Some of the chromosomal segments amplified in *S. spontaneum* genotypes might also be lost in modern cultivars due to aneuploidy. We were unable to detect the relationship between the *S. spontaneum* in the collection and ancestral *S. spontaneum* genotypes used in nobilization even when we included some old/legendary cultivars. This study also allowed us to realize that *S. spontaneum* germplasm representing the west zone (African countries) was not present in the collection.

5.3 Linkage Mapping of LCP 85-384

Sugarcane hybrids are poly-aneuploids (2n=100 to 130) and derived from interspeficic hybridization between *S. officinarum* and *S. spontaneum*. Efforts in unraveling the sugarcane genome have recently been successful by using molecular marker technologies. A framework genetic linkage map of Louisiana's popular cultivar 'LCP 85-384' was constructed using the selfed progeny based on 64 AFLP, 12 TRAP and 19 SSR markers. The mapping population comprised of 300 individuals. A total of 773 out of 1113 polymorphic markers generated were single dose (SD) markers that segregated in the theoretical 3:1 ratio and these were used to construct the map. Thirty-two markers deviated from Mendilian segregation ratio after the Bonferroni correction procedure. Linkage map was constructed using a LOD value of > 5.0 and recombination threshold of 0.44. The genetic distances between pairs of markers linked in the coupling phase was computed using the Kosambi mapping function. Out of 773, 717 markers were assigned onto 108 co-segregation groups (CGs) with a cumulative map length of 5,384 CM. Fifty-six markers remained unlinked. In conjunction with the LCP 85-384 linkage map, maps of its parents (CP 77-310 x CP 77-407) were also developed. The 773 SD markers were divided

based on their parental origin and used for the construction of parental maps with the same linkage map parameters applied in the LCP 85-384 map. There were a total of 440 markers in the female parent (CP 77-310) and 397 markers in the male parent (CP 77-407). The CP 77-310 map comprised of 391 linked markers, which spread over 81 CGs with a cumulative genomic length of 3476 cM, where as the CP 77-407 map contained of 339 markers with a cumulative length of 2777 cM.

With the estimated genome size of 12,720 cM, the map covered an estimated 42% of the genome of LCP 85-384. The genome covered found to be 23.2% for CP 77-310 and 24.1% for CP 77-407. Of the 108 CGs formed in LCP 85-384, 31 CGs were assigned into 12 homo(eo)logous groups (HGs) based on the SSRs and parental maps information. Likewise, 4 HGs in CP 77-310 (using 8 CGs out of 81CGs) and 7 HGs in CP 77-310 (14 CGs out of 80 CGs) were formed based on the LCP 85-384 linkage map. Repulsion phase linkages studied in LCP 85-384 and its parents suggested the preponderance of preferential pairing (disomic segregation) between CGs within the homo(eo)logus chromosomes. The marker distribution is uneven on the map and gaps exist. However, more markers need to be added to the current map to make the map more resourceful for further studies. The framework map of LCP 85-384 established in this study will provide an important background for mapping QTLs associated with sugar related traits and thus, information will be useful for crossing and in selecting the clones in the breeding program.

APPENDIX

PERMISSION LETTER

CAMBRIDGE UNIVERSITY PRESS

The Edinburgh Building
Shaftesbury Road
Cambridge CB2 8RU, UK

www.cambridge.org

Telephone +44 (0)1223 312393
Fax +44 (0)1223 315052
Email information@cambridge.org

Suman Andru
PhD Candidate
School of Plant, Environmental and Soil Sciences
Louisiana State University
Baton Rouge
LA 70803
USA

October 15, 2009

Dear Suman Andru

Andru Suman, Collins A. Kimbeng, Serge J. Edmé and John Veremis, "Sequence-related amplified polymorphism (SRAP) markers for assessing genetic relationships and diversity in sugarcane germplasm collections", Plant Genetic Resouces, Volume 6(3), pp 222-231, (2008).

Thank you for your recent permission request to include the above extract(s) in:
your forthcoming PhD dissertation for non-commercial publication.

Non-exclusive permission is granted free of charge for this specific use on the understanding that you have checked that we do not acknowledge another source for this material.

Please ensure full acknowledgement (author, title, publication date, name of journal and Cambridge University Press).

Yours sincerely

C Taylor

Claire Taylor
Publishing Assistant
email ctaylor@cambridge.org

1584 · 2009
425 YEARS OF CAMBRIDGE PRINTING AND PUBLISHING

From: Suman Andru sandru1@tigers.lsu.edu>
To: plantgeneticresources@googlemail.com >
Date: Wed, Oct 14, 2009 at 12:18 PM
Subject: Request for permission to use published article as part of my thesis

Dear Robert Koebner,
Editor- in- chief
CropGen International, UK.

I kindly request you to grant me the permission to use below cited research article as part of my PhD dissertation.

Andru Suman, Collins A. Kimbeng, Serge J. Edmé and John Veremis. 2008. Sequence-related amplified polymorphism SRAP markers for assessing genetic relationships and diversity in sugarcane germplasm collections. Plant Genetic Resources, Volume 6, Issue 03, Dec 2008, pp 222-231.

Thank you very much,

Sincerely,
Suman Andru
PhD candidate,
School of Plant, Environmental and Soil Sciences,
LSU, Baton Rouge, LA-70803.

VITA

Suman Andru was born in 1977 in Nachina Pally village, Andhra Pradesh state, India. He attended Sri Saraswathi Vidya Nikethan, Nachina Pally, for primary school and then Andhra Pradesh Residential School, Bandaru Pally for high school education. He received a bachelor's degree in agriculture in ANGR Agricultural University, AswaraoPet campus, India. He then moved to ANGR Agricultural University main campus, Hyderabad, India, to pursue master's degree in agriculture majoring in genetics and plant breeding. He worked as a senior research fellow in Directorate of Rice Research, ICRISAT campus, Hyderabad, for one year. In 2004, Suman admitted into the doctoral program in sugarcane genetics under the guidance of Dr. Collins A Kimbeng, Louisiana State University. He finished master's degree in applied statistics in Louisiana State University, 2009 whilst pursuing his doctorate.

Suman was a recipient of National Merit Scholarship for scoring the highest percentage in high school. He also received Telugu Vignana Parithoshikam (by Govt. of Andhra Pradesh) scholarship consecutively for five years in high school. Suman is currently a member of Crop Science Society of America, and Sigma Xi honor society.

Suman is married to Swathi and blessed with a baby girl, Nayana, while pursuing his advanced degree in LSU, Baton Rouge.