

Molecular characterization of tropical sweet corn inbred lines using microsatellite markers

Pedram Kashiani¹, Ghizan Saleh^{1*}, Jothi Malar Panandam², Nur Ashikin Psyquay Abdullah¹, Ahmad Selamat¹

¹Department of Crop Science, Faculty of Agriculture, Universiti Putra Malaysia 43400 Serdang, Selangor, Malaysia

²Department of Animal Science, Faculty of Agriculture, Universiti Putra Malaysia 43400 Serdang, Selangor, Malaysia

*Corresponding author: E-mail: ghizan@agri.upm.edu.my

Abstract

Genetic variability among 13 tropical sweet corn inbred lines derived from source populations originating from five tropical countries was investigated using 95 polymorphic microsatellite markers. It was found that 92.9% of total molecular variance was due to variations among the inbred lines. This was further supported by the presence of high values of DST, FST, GST and FIT, indicating high genetic variation among the inbred lines. The average number of alleles and effective number of alleles were both close to one allele per locus per inbred line. Departure from Hardy-Weinberg equilibrium was observed with a drastic reduction of the observed heterozygosity, resulting in a lower index of genotypic diversity than expected for each of the inbred lines. The inbred lines were further assigned into six main heterotic groups based on their molecular characteristics using Fitch-Margoliash algorithm. High genetic variations among the inbred lines indicates the presence of different heterotic groups, while low level of genetic variability within the inbred lines indicates that they met the assumption of homozygosity of their loci to enable further diallel crosses to be made for analysis of their combining ability. It is therefore expected that high heterosis in yield and its components could be obtained from crosses among those inbred lines belonging to different heterotic groups.

Keywords: sweet corn (*Zea mays* L. *saccharata*), microsatellite marker, inbred line, genetic variation

Introduction

Corn is one of the most phenotypically diverse crop species. It can be grown in a wide range of environments, so long as its water requirements are in accordance (Kashiani et al, 2011). Corn breeders emphasize selection for a matrix of traits that are controlled by multigenes. These traits are known as quantitative traits which involve most of important agronomic characteristics which include yield and yield components (Kashiani and Saleh, 2010). Effectiveness of selection for different quantitative traits depends on the ability to develop effective screens for the traits individually and collectively (Hallauer and Carena, 2009). Variations among tropical sweet corn lines have previously been investigated using phenotypic markers. However, the performance of phenotypic markers is strongly influenced by environmental factors. In contrast, molecular markers which are not influenced by the environment can be utilized to detect total genetic variation among sweet corn lines at the DNA level. The use of molecular markers to assist the breeding process allows breeding goals to be achieved more efficiently with reduction in the need for field assays. During the last two decades, numerous molecular markers have been developed for most major crop species (Barcaccia, 2010). Genetic variation detected by molecular markers has been helpful for understanding the genome

dynamics of the particular crop as well as improving breeding efficiency (Varshney et al, 2006; Srdić et al, 2011a). Development of Polymerase Chain Reaction (PCR) methods has given a wide impact of using molecular techniques for detecting genetic variation in breeding lines. In the recent years, various molecular techniques, namely Amplified Fragment Length Polymorphisms (AFLPs), Random Amplified Polymorphic DNA (RAPDs), Restriction Fragment Length Polymorphisms (RFLPs), Simple Sequence Repeats (SSRs) and Single Nucleotide Polymorphism (SNP) have been used in genetic characterization of crop species.

Microsatellites, also known as SSRs, Short Tandem Repeats (STRs) or Sequence-tagged Microsatellite Sites (STMS), are tandemly repeated units of short nucleotide motifs that are 1 to 6 bp long. Di-, tri- and tetra-nucleotide repeats are widely distributed throughout the genomes of plants and animals (Tautz and Renz, 1984). One of the most important attributes of microsatellite loci is their high level of allelic variation, making them valuable and informative genetic markers (Xu, 2010). Microsatellite markers were developed and mapped on linkage maps in the 1990s (Kole, 2006). Currently, thousands of SSR primer pairs have been developed and mapped on the corn genome, the list of which is available in the Maize-GDB database (<http://www.maizegdb.org>). Most of

these primers were derived from cDNA sequences. In the last decade, SSR markers have been replacing other markers such as RFLP, AFLP and RAPD in genetic studies due to their advantages, namely being easy to detect, co-dominant, highly polymorphic and cost effective. Recently constructed corn linkage maps have been based on SSR markers (Kole, 2006). Microsatellite loci have proven their efficiency as genetic markers to assess genetic diversity in numerous plant species. To date, SSRs have been used on corn for mapping (Taramino and Tingey, 1996), genetic fingerprinting (Smith et al, 1997; Senior et al, 1998) and to assess genetic diversity among inbred lines (Lu and Bernardo, 2001; Enoki et al, 2002; Liu et al, 2003).

Corn inbred lines represent a fundamental resource for studies in genetics and breeding. While corn inbreds are used extensively in hybrid corn production (Troyer, 2001), they have also been critical for diverse genetic studies including the development of linkage maps (Burr et al, 1988), quantitative trait locus mapping (QTL) (Austin et al, 2001), molecular evolution (Henry and Damerval, 1997; Ching et al, 2002), developmental genetics (Fowler and Freeling, 1996), and physiological genetics (Crosbie et al, 1978). Johnson (2004) reported that marker-assisted selection (MAS) during the development of inbred lines was found to increase significantly the hybrid performance in advanced generations. A set of diverse corn inbreds was employed to perform the first phenotype-genotype association analyses in a plant species (Thornsberry et al, 2001) and to estimate linkage disequilibrium (LD) in corn (Remington et al, 2001; Tenaillon et al, 2001). Terra et al (2011) investigated genetic variability among field corn, sweet corn and teosinte using 25 SSR markers. They reported a mean value of 0.52 for polymorphic information content (PIC), which was similar to that reported by Amorim et al (2003) and Bered et al (2005) in the analysis of genetic diversity in sweet corn, and by Patto et al (2004) on commercial corn lines (0.56).

Recently, microsatellite loci were utilized to study genetic diversity among sweet corn populations and their relationships to field corn. Rupp et al (2009) investigated genetic structure and diversity among 13 sweet corn populations using 100 microsatellite loci. In another study, genetic variability in open-pollinated populations of sweet corn, field corn and teosinte was investigated using 10 microsatellite markers by Almeida et al (2011). Srdić et al (2011b) analyzed genetic similarity of six sweet corn inbreds based on 40 microsatellite markers, while Terra et al (2011) used 25 microsatellite loci to analyze genetic differences on two sweet corn populations, two field corn populations and one teosinte population.

Sweet corn can be grown in a wide range of environments as long as its water requirement is met (Kashiani et al, 2011). The tropical sweet corn inbred lines utilized in the current study were previously de-

veloped from the tropical source populations from Malaysia, Indonesia, Thailand, Taiwan and the USA. However, there is no information available on molecular characterization of the inbred lines as well as the source populations from which the inbred lines were derived. Previous phenotypic evaluation on some lines obtained from the same source populations after five generations of selfing and selection was done without accessing molecular variation among the inbred lines (Kashiani et al, 2010). For better understanding of inter- and intra-population genetic variation that existed among the inbred lines, 105 microsatellite markers distributed throughout the sweet corn genome were screened on the inbred lines. The main objective of the present study was to characterize genetic variation among and within a set of tropical sweet corn inbred lines using SSR markers.

Materials and Methods

A series of 13 selected tropical near-homozygous inbred lines previously developed from various source populations of tropical origin through a long-term breeding program conducted at Faculty of Agriculture, Universiti Putra Malaysia were used as plant materials in this experiment (Supplementary Table 1). Twenty seeds from each inbred line were germinated in jiffy cups and seedlings were grown to the two-leaf stage. Genomic DNA was extracted from young leaves of 10 seedlings per line as explained by Kashiani et al (2012).

Ninety nine microsatellite regions that are distributed throughout the corn genome were chosen from the MaizeGDB (<http://www.maizegdb.org.php>) based on their polymorphism information contents (PIC) and QTL information reported in previous investigations (Supplementary Figure 1). The amplifications were performed using volumes of 20 µl PCR reaction containing 5 µl (25–30 ng) of genomic DNA, 1.5 µl PCR 10x buffer, 1.5 µl of 25 mM MgCl₂, 0.3 µl dNTP Mix (10 mM each of dATP, dGTP, dCTP, dTTP), 0.2 unit of Taq polymerase, 1.8 µl (4 pmol/µl) of each primer (F and R primers), and 9.7 µl distilled water. PCR amplification was performed in an Eppendorf Mastercycler Gradient Thermal cycler (Eppendorf Scientific, Westbury, NY) using 96-well plate. The amplification conditions with the touchdown thermal cycling protocol were 95°C for 1 min, 30 cycles of 94°C for 1 min, 67°C annealing temperature decreasing by 0.4 °C per cycle for 2 min and 72°C for 2 min, and a terminal extension step at 72°C for 1 h. After amplification, 10 µl of the amplified DNA was mixed with 5 µl of formamide loading buffer and loaded into 4% (w/v) metaphore agarose 36-cm well-to-read gels with 1x TBE buffer. Electrophoresis was performed at 100 volts for 5 min, and followed by 55 volts for approximately 4 h until the bromophenol blue band of the loading dye was moved forward for 10 cm. All gels were visualized under UV light using an Alpha EaseR FC Imaging System (Alpha Innotech Corporation, CA)

Table 1 - Results of analysis of molecular variance (AMOVA) among 13 tropical inbred lines using 99 microsatellite DNA markers.

Source of Variation	d.f.	Sum Squares	Mean Squares	Estimated Variance	Percentage of Variance (%)
Before Pooling					
Among inbred lines	12	2,031.60	169.30**	28.07	93.6
Among plants within inbred lines	26	22.67	0.87 ^{ns}	-1.05	-3.5
Between alleles within individuals	39	116.00	2.97 ^{ns}	2.97	9.9
After Pooling					
Among inbred lines	12	2,031.60	169.30**	27.86	92.9
Within inbred lines	65	138.67	2.13 ^{ns}	2.13	7.1
Total	77	2,170.27			

** and ns: significant at $p \leq 0.01$ and non-significant, respectively.

after staining with ethidium bromide.

Analysis of Molecular Variance (AMOVA) was performed to estimate differentiation among the inbred lines and among individuals within inbred line using microsatellite data (Excoffier et al, 1992). The effect of spatial separation on genetic structure was tested by the Mantel test (Mantel and Valand, 1970) on genetics matrices (Nei, 1978) between populations. Mantel test was performed with 1000 random permutations using the software GenAlEx ver 6.2 (Peakall and Smouse, 2006).

Through locus-by-locus AMOVA, pairwise estimates of gene flow (Nm) (Nei, 1987) and proportion of differentiation among the inbred lines (F_{ST}) (Wright, 1978) were calculated independently to discard possible problematic loci according to the level of significance tested. A Principal Coordinate Analysis (PCoA) was further conducted on the pairwise F_{ST} matrix and the first two principal coordinates were plotted to infer the level of gene similarity among the inbred lines. Descriptive statistics for each inbred line including average number of alleles (n_a), number of effective alleles (n_e) (Kimura and Crow, 1964), observed and expected homozygosity (Hom_o and Hom_e, respectively), observed and expected heterozygosity (H_o and H_e, respectively) (Levene, 1949; Nei, 1987), average heterozygosity (\bar{H}), Shannon's information index (I) (Lewontin, 1972), Nei's expected heterozygosity (Nei's) (Nei, 1973), gene flow (Nm) (Nei, 1987), and coefficient of inbreeding (F) (Lukas and Donald, 2002) were estimated using the Population Genetic Analysis software (POPGENE) version 1.3.1 (Yeh et al, 1999).

Variance components, the sum of all squared differences and Wright's fixation indices (F_{IS}, F_{ST} and F_{IT}) (Wright, 1978), which are standardized variances in allele frequencies that detect departure from Hardy-Weinberg equilibrium test (HWT), were calculated to

estimate the population differentiation using the computer software Arlequin suite version 3.5 (Excoffier and Lischer, 2010). HWT was performed to determine if both gene and genotype frequencies were constant from generation to subsequent next generation using Chi-squared test (Guo and Thompson, 1992; Hartl and Clark, 2007). The inbred lines studied assumed not to reflect the most basal ancestor as they were developed from different source populations. Therefore, an unrooted tree based on the algorithm described by Fitch and Margoliash (Fitch and Margoliash, 1968) were constructed using Nei's standard genetic distance coefficients (Nei, 1972).

Results

Out of 99 microsatellite markers amplified, 95 were found to be polymorphic and produce visible and reproducible bands (Supplementary Figure 1). Results of analysis of molecular variance (AMOVA) showed the inbred lines evaluated were highly variable. It was found that 93.59% of total molecular variance was due to variations among the inbred lines (Table 1). As expected, there was no significant variation observed among the individual plants within inbred lines. Among the individual plants within inbred lines variance component was found to be negative and very small (-1.051). This indicates that there was no significant variation among individual plants within inbred lines, as the inbred lines have reached high level of homozygosity. This was further supported by non-significant variance for between alleles within individual plant (2.97), indicating high level of homozygosity at all the loci amplified. Hence, the total molecular variance was partitioned into among inbred lines and within inbred lines as sources of variation, as shown in Table 1. Result of the pooled AMOVA showed that 92.9% of the total variation occurred

Table 2 - Genetic differentiation statistics under assumption of Hardy-Weinberg disequilibrium for 13 tropical inbred lines, revealed by microsatellite DNA markers.

H _s	D _{ST}	H _T	F _{IS}	F _{IT}	F _{ST}	F _{ST} (pooled)	G _{ST}	Nm from F _{ST}	Nm from G _{ST}
0.040	0.546	0.586	-0.547	0.901	0.936	0.929	0.931	0.0182	0.0118

Table 3 - Pairwise estimates of Nm (above diagonal) and F_{ST} (below diagonal) values among 13 tropical inbred lines, using microsatellite DNA markers.

Inbred line	MAS-S8	BAK-S8	TSG-S8	TSN-S8	THI-S8	DFS-S8	IND-S8	MAN-S8	TSK-S8	TFZ-S8	MMIN-S8	HAW-S8	SBY-S8	Mean Nm
MAS-S8		0.0159	0.0119	0.0123	0.0126	0.0183	0.0068	0.0180	0.0075	0.0052	0.0066	0.0097	0.0082	0.0111
BAK-S8	0.9401		0.0275	0.0284	0.0255	0.0318	0.0189	0.0339	0.0222	0.0156	0.0175	0.0213	0.0198	0.0232
TSG-S8	0.9544	0.9010		0.0250	0.0202	0.0271	0.0125	0.0269	0.0145	0.0111	0.0123	0.0145	0.0143	0.0181
TSN-S8	0.9530	0.8979	0.9091		0.0256	0.0267	0.0130	0.0299	0.0142	0.0137	0.0121	0.0176	0.0144	0.0194
THI-S8	0.9519	0.9076	0.9253	0.9072		0.0303	0.0141	0.0317	0.0163	0.0144	0.0149	0.0167	0.0156	0.0198
DFS-S8	0.9317	0.8872	0.9022	0.9034	0.8919		0.0207	0.0372	0.0225	0.0220	0.0194	0.0244	0.0239	0.0254
IND-S8	0.9737	0.9298	0.9524	0.9507	0.9465	0.9236		0.0275	0.0114	0.0065	0.0078	0.0121	0.0099	0.0134
MAN-S8	0.9328	0.8805	0.9030	0.8933	0.8875	0.8705	0.9009		0.0311	0.0218	0.0222	0.0253	0.0223	0.0273
TSK-S8	0.9710	0.9186	0.9451	0.9462	0.9387	0.9173	0.9565	0.8893		0.0086	0.0091	0.0133	0.0116	0.0152
TFZ-S8	0.9796	0.9411	0.9574	0.9480	0.9456	0.9192	0.9747	0.9199	0.9668		0.0066	0.0100	0.0093	0.0121
MMIN-S8	0.9741	0.9347	0.9532	0.9538	0.9436	0.9279	0.9697	0.9184	0.9650	0.9741		0.0110	0.0094	0.0124
HAW-S8	0.9627	0.9214	0.9452	0.9344	0.9373	0.9112	0.9540	0.9080	0.9496	0.9617	0.9578		0.0156	0.0160
SBY-S8	0.9683	0.9267	0.9458	0.9454	0.9413	0.9127	0.9618	0.9180	0.9558	0.9640	0.9639	0.9411		0.0145
Mean FST	0.9578	0.9156	0.9328	0.9285	0.9270	0.9082	0.9495	0.9018	0.9433	0.9543	0.9530	0.9404	0.9454	

F_{ST} = proportion of the total inbreeding at a locus and Nm = gene flow.

among inbred lines and only 7.1% occurred within inbred lines. This confirms the presence of high level of genetic differentiation among the inbred lines evaluated.

High differences were shown among the inbred lines ($D_{ST} = 0.546$), while low genetic diversity was displayed within the inbred lines ($H_s = 0.040$) (Table 2). This indicates high inter-population diversity among the inbred lines and very low intra-population diversity. This was in agreement with the results revealed by AMOVA. Total genetic diversity was found to be high ($H_T = 0.586$) due to high differences among the inbred lines evaluated.

High and negative value of total inbreeding in an inbred line due to both inbreeding within inbred lines ($F_{IS} = -0.547$) indicates a significant deficiency of heterozygosity within the inbred lines studied.

This was due to low and negative variance component estimates obtained among samples within the inbred lines (-1.051). Total inbreeding coefficient in an inbred line due to both inbreeding within inbred lines and differentiation among inbred lines (F_{IT}) was found to be 0.901 (Table 2). It was, however, smaller than the proportion obtained only from differentiation among the inbred lines ($F_{ST} = 0.936$), indicating negative and non-significant variance among samples within inbred lines. Similar F_{ST} value was found from the pooled AMOVA estimate ($F_{ST} = 0.929$). Coefficient of gene differentiation which shows inter-population differentiation was found to be in close agreement with that of F_{ST} ($G_{ST} = 0.931$). The estimate of gene flow obtained from F_{ST} was also found to be in close agreement with that estimated from G_{ST} (0.0182 and 0.0118, respectively).

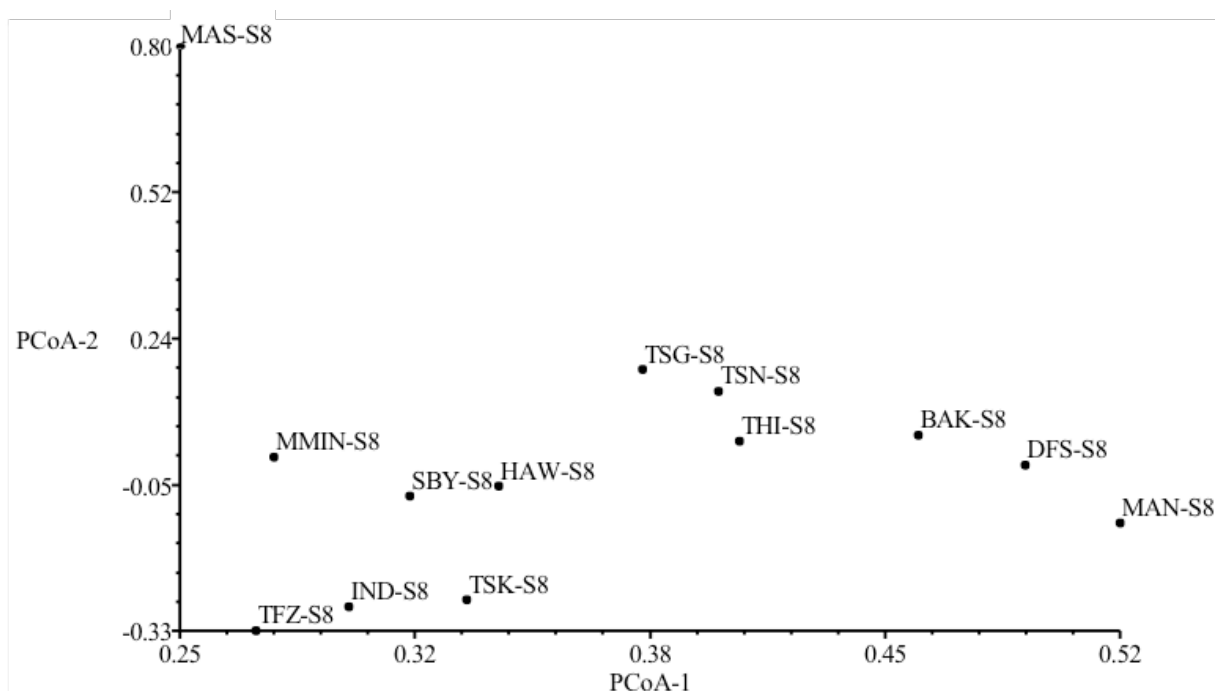


Figure 1 - Plot of the first two axes of a principal coordinate analysis of the pairwise F_{ST} matrix among 13 inbred lines studied.

Through locus-by-locus AMOVA, F_{ST} of each inbred line was calculated independently to discard possible problematic loci according to the level of significance tested. Permutation procedures (1,000 replicates) for the significance testing on variance components and F_{ST} values were performed. Inbred lines pairwise F_{ST} were used to infer what level of gene flow occurred among the inbred lines (Table 3). The F_{ST} among the 13 tropical inbred lines indicates a high level of genetic differentiation and suggests a low level of gene flow among the inbred lines. The limited gene flow which is an indication of genetic isolation, resulted from self-pollination and selection that the inbred lines had undergone. Mantel test was carried out to estimate the correlation and significance between pairwise F_{ST} and pairwise Nm values. The result showed that there was a strong negative correlation ($r = -0.999$, $t = -3.615$, $p = 0.0001$) between F_{ST} and Nm, indicating high level of genetic isolation among the inbred lines studied.

A principal coordinate analysis (PCoA) was conducted on the pairwise F_{ST} matrix and the first two principal coordinates accounting for 52.7% of variation were plotted (Figure 1). The result showed approximately equal distances among the inbred lines, indicating very low level of gene flow between any pair of inbred lines. However, MAS-S8 was separated from all other inbred lines, indicating its high level of genetic differentiation. This was also revealed by the pairwise F_{ST} and Nm values, where MAS-S8 showed the highest mean value for F_{ST} ($F_{ST} = 0.9578$) and the lowest mean value for Nm ($Nm = 0.0111$) compared to the other inbred lines evaluated. Results showed that DFS-S8 possessed the highest average number of alleles (n_a) and number of effective alleles (n_e) among the inbred lines evaluated, with values of 1.190 and 1.145, respectively, while TFZ-S8 was found to have the lowest n_a and n_e (1.032 and 1.025, respectively). However, the average number of alleles and number of effective alleles was found to be low among the inbred lines, with mean values of 1.091 and 1.068,

respectively (Table 4). This proves high homozygosity of the inbred lines studied at each locus amplified.

All the inbred lines studied had very high observed homozygosity (H_o) and very low observed heterozygosity (H_e), with mean values of 0.937 and 0.063, respectively. This indicates that the inbred lines reached homozygosity at all loci and can be maintained by selfing without further changes. The average observed homozygosities of the inbred lines DFS-S8, TFZ-S8, HAW-S8 and SBY-S8 which were developed from hybrid source populations were found to be higher than those derived from open-pollination source populations (0.939 and 0.937, respectively). TFZ-S8 was found to have the highest H_o and the lowest H_e among the inbred lines evaluated, with values of 0.975 and 0.025, respectively, while MAN-S8 revealed the lowest H_o and highest H_e (0.860 and 0.140, respectively). Expected homozygosity and expected heterozygosity (with mean values of 0.330 and 0.670, respectively) showed that none of the inbred lines studied was at Hardy-Weinberg equilibrium (Table 4). Departure from Hardy-Weinberg equilibrium was also revealed by the significant difference between average heterozygosity (0.666) and observed heterozygosity (0.063). Overall Nei's expected heterozygosity (Nei, 1973) and Shannon's information index (Lewontin, 1972), which are indices of gene diversity, were found to be very low among the inbred lines, with mean values of 0.0374 and 0.0543, respectively (Table 4). Similarly, average gene diversity over all loci amplified revealed very low diversity within the inbred lines evaluated (0.045).

Among the inbred lines, DFS-S8 was found to have the highest Nei's expected heterozygosity, Shannon's information index and average gene diversity (0.0784, 0.1136 and 0.094, respectively), while TFZ-S8 had the lowest values (0.0135, 0.0193 and 0.016, respectively). The inbred lines were found to have high inbreeding coefficient (F), with mean value of 0.904 (Table 4). This was due to high rate of self-pollination among the inbred lines. TFZ-S8 was found

Table 4 - Genetic characteristics of 13 tropical inbred lines revealed by microsatellite DNA markers.

Inbred line	n_a	n_e	Hom_o	Hom_e	H_o	H_e	\hat{H}	I	Nei's	h	\hat{h}	F	HW
MAS-S8	1.0316	1.0316	0.9684	0.7596	0.0316	0.2404	0.2392	0.0219	0.0158	0.0001	0.0189	0.8686	17.195**
BAK-S8	1.1684	1.1015	0.9053	0.3958	0.0947	0.6042	0.6010	0.0906	0.0602	0.0281	0.0723	0.8433	31.320**
TSG-S8	1.0947	1.0646	0.9404	0.3656	0.0596	0.6344	0.6311	0.0543	0.0368	0.0141	0.0442	0.9061	43.643**
TSN-S8	1.1053	1.0711	0.9404	0.3055	0.0596	0.6945	0.6909	0.0586	0.0392	0.0164	0.0470	0.9142	49.227**
THI-S8	1.1053	1.0881	0.9158	0.2805	0.0842	0.7195	0.7157	0.0667	0.0468	0.0094	0.0561	0.8830	49.764**
DFS-S8	1.1895	1.1448	0.8807	0.2745	0.1193	0.7255	0.7217	0.1136	0.0784	0.0328	0.0940	0.8356	37.602**
IND-S8	1.0526	1.0311	0.9719	0.2673	0.0281	0.7327	0.7288	0.0282	0.0187	0.0094	0.0225	0.9616	61.830**
MAN-S8	1.1684	1.1422	0.8596	0.2630	0.1404	0.7370	0.7331	0.1079	0.0760	0.0117	0.0912	0.8095	41.657**
TSK-S8	1.0421	1.0421	0.9579	0.3294	0.0421	0.6706	0.6671	0.0292	0.0211	0.0001	0.0253	0.9372	55.951**
TFZ-S8	1.0316	1.0251	0.9754	0.2744	0.0246	0.7256	0.7218	0.0193	0.0135	0.0024	0.0161	0.9661	63.031**
MMIN-S8	1.0316	1.0295	0.9719	0.3022	0.0281	0.6978	0.6941	0.0213	0.0152	0.0024	0.0182	0.9597	60.436**
HAW-S8	1.0842	1.0691	0.9404	0.2414	0.0596	0.7586	0.7546	0.0527	0.0368	0.0117	0.0442	0.9214	56.988**
SBY-S8	1.0737	1.0479	0.9579	0.2316	0.0421	0.7684	0.7644	0.0416	0.0281	0.0164	0.0337	0.9452	59.671**
Mean	1.0907	1.0684	0.9374	0.3301	0.0626	0.6699	0.6664	0.0543	0.0374	0.0119	0.0449	0.9040	48.332
SD	0.0555	0.0409	0.0370	0.1373	0.0370	0.1373	0.1365	0.0326	0.0224	0.0101	0.0269	0.0520	13.692

n_a = average number of alleles, n_e = number of effective alleles, Hom_o = observed homozygosity, Hom_e = expected homozygosity, H_o = observed heterozygosity, H_e = expected heterozygosity, \hat{H} = average heterozygosity, I = Shannon's information index, Nei's = Nei's expected heterozygosity, h = gene diversity, \hat{h} = average gene diversity, F = inbreeding coefficient, HW = departure from Hardy-Weinberg linkage equilibrium, and SD = standard deviation.

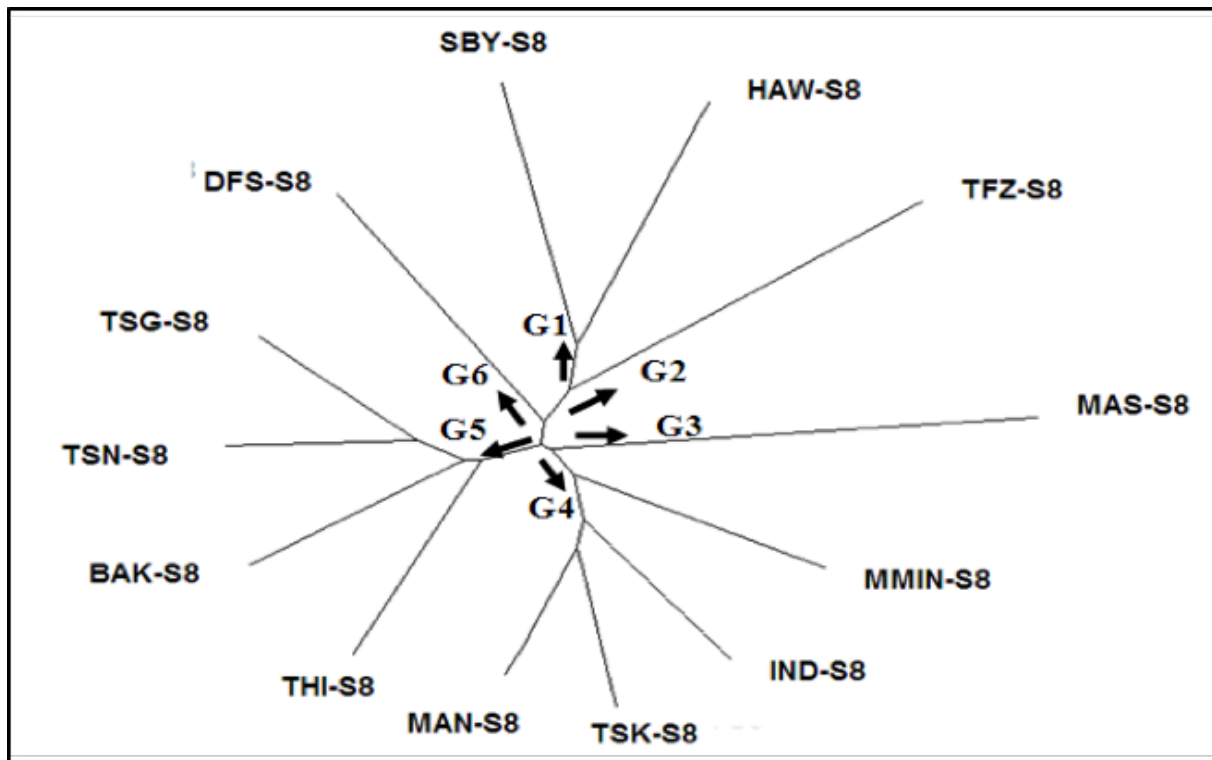


Figure 2- A Fitch-Margoliash unrooted tree indicating relationships among the tropical inbred lines based on molecular genetic distance coefficients obtained from amplification of 95 microsatellite DNA markers.

to have the highest inbreeding coefficient (0.966), while SBY-S8 had the lowest value (0.810) among the inbred lines.

The genetic relationship among the inbred lines was further investigated using an unrooted Fitch-Margoliash tree based on molecular genetic distances. Fitch-Margoliash method minimizes the weighted squared deviation of the tree path length distances from the distance estimates, and therefore is capable of estimating correct unrooted-branch lengths. The tree constructed showed that the inbred lines studied could be separated into six main heterotic groups (Figure 2). Based on the tree constructed, Group 1 consists of SBY-S8 and HAW-S8, Group 2 consists of TFZ-S8, Group 3 consists of MAS-S8, Group 4 consists of MMIN-S8, IND-S8, TSK-S8 and MAN-S8, Group 5 consists of THI-S8, BAK-S8, TSN-S8 and TSG-S8, and Group 6 consists of DFS-S6 (Figure 2). These relationships were in agreement with those obtained from principal coordinate analysis on the pairwise F_{ST} matrix, where MAS-S8 was separated from all other inbred lines, indicating its high level of genetic differentiation. In addition, the tree constructed was investigated for its goodness of fit using cophenetic correlation coefficients. The degree of fit between cophenetic correlations and molecular genetic distance coefficients was found to be very high (0.981) based on the scale reported by Rohlf and Fisher (Rohlf and Fisher, 1968) and Rohlf (Rohlf, 2002). Since the manifestation of heterosis depends

on the genetic divergence of the parental lines (Hallauer and Miranda, 1988; Zhong-hu, 1991), the inbred lines using for further hybrid production can be selected from each heterotic group.

Discussion

The genetic structure of the inbred lines reflects the interactions of different processes including mating system, gene flow and selection. The significant variation among the 13 tropical inbred lines revealed by microsatellites indicates high variability and high level of genetic differentiation among them. This could be due to differences in the initial gene pools from which the inbred lines were derived. AMOVA showed that the differences were mostly among the inbred lines rather than within. The inbred lines evaluated could preserve the genetic variation among themselves through continuous selection applied to them. In an investigation among important US inbred lines, Gethi et al (2002) reported that 87.8% of the total genetic variation was due to variations among the inbred lines. Solomon et al (2010) reported the significant role of selection methods such as single-hill and reciprocal recurrent selection in corn improvement, by maintaining adequate genetic diversity that ensures sustainable productivity improvement. In contrast, the negligible genetic variation observed within the inbred lines evaluated showed that they were homozygous at most of the loci amplified. This was due to the self-pollination mating system which caused

a reduction in intra-population heterozygosity (Kashiani et al, 2012). Similarly, Lu and Bernardo (2001) reported that genetic variation among corn inbred lines was reduced at the gene level (intra-population heterozygosity) but not at the population level (inter-population heterozygosity).

Genetic differentiation statistics, which were to assess the degree of relatedness among the inbred lines, also revealed high genetic variability among the inbred lines studied. High values of D_{ST} , F_{ST} , G_{ST} and F_{IT} showed that high proportion of total genetic diversity obtained was due to variation among the inbred lines. Many investigators reported high values of D_{ST} , F_{ST} , G_{ST} and F_{IT} when high genetic variability was mainly due to variations among populations (Gethi et al, 2002; Aga et al, 2005; Zhao et al, 2007). The estimate of overall population differentiation obtained from the G_{ST} value showed that the inbred lines were about 93% fixed for alleles involved. This might result in overwhelming performance of F_1 progenies obtained from crosses among the inbred lines, since they would receive different genes from each parent belonging to different heterotic groups. The high genetic differentiation of the inbred lines was also revealed through plotting of the first two principal coordinates obtained from pairwise F_{ST} and N_m in which the inbred lines were found to have approximately the same genetic differences from each other. Coefficient of inbreeding also showed high level of inbreeding within the inbred lines, resulting in high homozygosity within them.

Although corn is a cross-pollinating species, the genetic characteristics of the inbred lines revealed by microsatellites were similar to those of self-pollinating species. Aga et al (2005) reported that self-pollinating species are characterized by a high value of total genetic diversity ($H_T = 0.33$), a low value of gene diversity within populations ($H_S = 0.159$), and a high value of the coefficient of gene differentiation ($G_{ST} = 0.51$). This indicates that self-pollinating species maintain high genetic diversity at their polymorphic loci, and most of this variation is found among populations. The result obtained from the corn inbred lines in this study ($H_T = 0.586$, $H_S = 0.040$ and $G_{ST} = 0.931$) corresponds well to the genetic structure of a self-pollinating crop with most of its variation residing among the inbred lines rather than within the inbred lines. This observation showed the effect of severe selfing applied to the inbred lines over eight generations. Similarly, Hamrick and Godt (1989) reported that G_{ST} for outbreeding species was about 0.2, while that for inbreeding species was 0.5 and more. The coefficient of genetic differentiation (G_{ST}) reported for a self-pollinating species (*Elymus fibrosis*) was 0.65 with allozyme and 0.63 with RAPD (Díaz et al, 2000).

Average number of alleles and effective number of alleles estimated per locus were both close to one allele per locus per inbred line, which caused very high difference between observed and expected ho-

mozygosities. Departure from Hardy-Weinberg equilibrium was observed with a drastic reduction of the observed heterozygosity, resulting in a lower index of genotypic diversity than expected for each of the inbred lines evaluated. Nei's expected heterozygosity, Shannon's information index and average gene diversity showed very low variation within the inbred lines. This was caused by self-pollinations applied to the inbred lines, and not as a result of neutral allelic divergence on all loci within inbred lines. Falconer and Mackay (1996) reported that direct selection might result in reduction of number of effective alleles. A comprehensive study on a set of 260 corn inbred lines representing samples from the most important public lines from the United States, Europe, Canada, South Africa and Thailand, lines from the International Center for the Improvement of Maize and Wheat (CIMMYT), and the International Institute of Tropical Agriculture (IITA) showed that number of alleles per locus was 2.8 for corn inbred lines (Liu et al, 2003). This was rather higher than the number of alleles per locus estimated from the 13 tropical inbred lines used in this present study. This could be due to much bigger sample size (260 inbred lines) used and more diverse inbred lines utilized by Liu et al (2003) compared to that in the present study. In an investigation on genetic diversity among 15 sweet corn lines after five generations of selfing, Rupp et al (2009) reported that the average number of alleles per locus estimated on the lines was 2.61, indicating high heterozygosity within the lines. The lower number of alleles per locus estimated in the present study was due to higher number of generations of self-pollination (eight generations) applied on the inbred lines.

The inbred lines developed from hybrid source populations clustered separately from those developed from open-pollination and composite source populations based on their molecular characteristics. This indicates high dissimilarity in the initial gene pools between the introduced hybrid source populations and the local source populations. In addition, the three inbred lines developed from hybrid source populations were separated into three genetically diverse groups (Groups 1, 2, and 6), indicating higher genetic diversity among their source populations. This suggests that different alleles for any particular trait were present in those hybrid source populations. However, molecular characteristics of HAW-S8 were found to be similar with those of SBY-S8 which was developed from a composite source population (Surabaya). Kashiani et al (2010) reported similar phenotypic performances of the lines developed from Hawaiian and Surabaya source populations. Among the inbred lines developed from Thai Super Sweet source population in different environments, TSN-S8 and TSG-S8 were found to have similar molecular identities as BAK-S8, while TSK-S8 was more similar to MAN-S8 and MMIN-S8. This indicates that the source population from which TSN-S8 and TSG-S8

were developed was similar to the source population from which BAK-S8 was developed. This could be due to non-homogeneous rates of genetic change, resulting from founder effect and isolation by distance (Hedrick, 1999).

In conclusion, microsatellites were found to be informative markers to reveal genetic variation among the inbred lines studied. They indicate that total genetic variations that existed among the inbred lines were mostly due to differences between lines, showing the presence of different heterotic groups. Therefore, it is expected that high heterosis in yield and yield components could be obtained from crosses among the inbred lines from different heterotic groups.

Acknowledgements

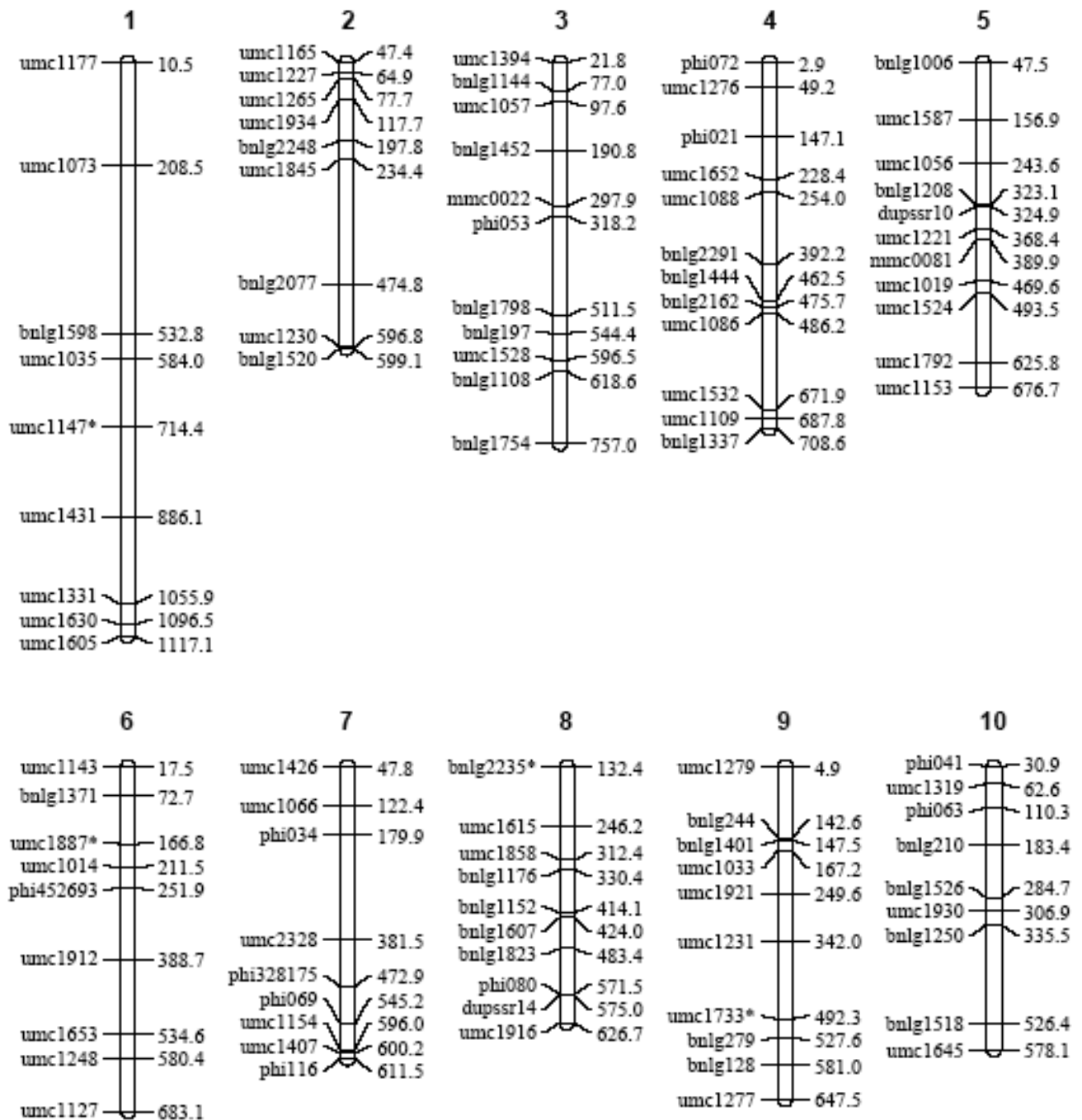
The authors would like to extend their gratitude to Ministry of Science, Technology and Innovation, Malaysia (MOSTI) for funding this project under e-Science Fund (Project No.: 05-01-04-SF0700). Special thanks go to Agro-Biotechnology Institute, Malaysia (ABI) for their hospitality during the research work.

References

- Aga E, Bekele E, Bryngelsson T, 2005. Inter-simple sequence repeat (ISSR) variation in forest coffee trees (*Coffea arabica* L.) populations from Ethiopia. *Genetica* 124(2): 213-221
- Almeida C, Amorim EP, Barbosa NJF, Filho JAC, Sereno MJC, 2011. Genetic variability in populations of sweet corn, common corn and teosinte. *Crop Breed. Appl. Biotechnol.* 11: 64-69
- Amorim EP, de Souza ACC, Sereno MJC, Bered F, Barbosa NJF, 2003. Genetic variability in sweet corn using molecular markers. *Maydica* 48: 177-181
- Austin DF, Lee M, Veldboom LR, 2001. Genetic mapping in maize with hybrid progeny across testers and generations: plant height and flowering. *Theor Appl Genet* 102(1): 163-176
- Barcaccia G, 2010. Molecular markers for characterizing and conserving crop plant germplasm, pp. 772. In: *Molecular techniques in crop improvement*. Mohan Jain S and Brar DS eds. Springer, New York, NY
- Bered F, Terra TF, Spellmeier M, Barbosa NJF, 2005. Genetic variation among and within sweet corn populations detected by RAPD and SSR markers. *Crop Breed Appl Biotechnol* 5: 418-425
- Burr B, Burr FA, Thompson KH, Albertson MC, Stuber CW, 1988. Gene mapping with recombinant inbreds in maize. *Genetics* 118(3): 519-526
- Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ, 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genet* 3: 19-33
- Crosbie TM, Mock JJ, Pearce RB, 1978. Inheritance of photosynthesis in a diallel among eight maize inbred lines from Iowa stiff stalk synthetic. *Euphytica* 27(3): 657-664
- Díaz O, Sun G-L, Salomon B, von Bothmer R, 2000. Levels and distribution of allozyme and RAPD variation in populations of *Elymus fibrosus* (Schrenk) Tzvel. (*Poaceae*). *Genet Resour Crop Evol* 47(1): 11-24
- Enoki, Sato, Koinuma, 2002. SSR analysis of genetic diversity among maize inbred lines adapted to cold regions of Japan. *Theor Appl Genet* 104(8): 1270-1277
- Excoffier L, Lischer HEL, 2010. Arlequin suite ver 3.5: A New Series of Programs to Perform Population Genetics Analyses under Linux and Windows. *Mol Ecol Resour* 10: 564-567
- Excoffier L, Smouse PE, Quattro JM, 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131(2): 479-491
- Falconer DS, Mackay TFC, 1996. *Introduction to quantitative genetics*. Longman Group Ltd, London
- Fitch WM, Margoliash E, 1968. The construction of phylogenetic trees. II. How well do they reflect past history? *Brookhaven Symp Biol* 21(1): 217-42
- Fowler JE, Freeling M, 1996. Genetic analysis of mutations that alter cell fates in maize leaves: dominant Liguleless mutations. *Dev Genet* 18(3): 198-222
- Gethi JG, Labate JA, Lamkey KR, Smith ME, Kresovich S, 2002. SSR variation in important US maize inbred lines. *Crop Sci* 42: 951-957
- Guo SW, Thompson EA, 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48(2): 361-72
- Hallauer AR, Carena MJ, 2009. Maize breeding, pp. 3-98. In: *Handbook of plant breeding: cereals*. Carena MJ eds. Springer, New York, NY
- Hallauer AR, Miranda FJB, 1988. *Quantitative genetics in maize breeding*. Iowa State University Press: Ames
- Hamrick JL, Godt MJW, 1989. Allozyme diversity in plant species, pp. 43-63. In: *Plant population genetics, breeding and genetic resources*. Brown AHD, Clegg MT, Kahler AL, Weir BS eds. Sinauer Associates Inc, Sunderland, MA
- Hartl DL, Clark AG, 2007. *Principles of population genetics*. Sinauer Associates, Inc, Sunderland, MA
- Hedrick P, 1999. Perspective: Highly variable loci and their interpretation in evolution and conservation. *Evolution* 53(2): 313-318
- Henry AM, Damerval C, 1997. High rates of polymorphism and recombination at the *Opaque-2* locus in cultivated maize *Mol Gen Genet* 256(2): 147-157

- Johnson GR, 2004. Marker-assisted Selection, pp. 293-309. In: Plant Breeding Reviews, Part 1. Janick J eds. John Wiley & Sons, Inc, Hoboken, NJ
- Kashiani P, Saleh G, 2010. Estimation of genetic correlations on sweet corn inbred lines using SAS mixed model. *Am J Agric Biol Sci* 5(3): 309-314
- Kashiani P, Saleh G, Abdullah NAP, Abdullah SN, 2010. Variation and genetic studies on selected sweet corn inbred lines. *Asian J Crop Sci* 2(2): 78-84
- Kashiani P, Saleh G, Jothi MP, Abdullah NAP, Ahmad S, 2012. Demarcation of informative chromosomes on tropical sweet corn inbred lines using microsatellite DNA markers. *Genet Mol Biol* 35(3): 614-621
- Kashiani P, Saleh G, Osman M, Habibi D, 2011. Sweet corn yield response to alternate furrow irrigation methods under different planting densities in a semi-arid climatic condition. *Afr J Agric Res* 6(4): 1032-1040
- Kimura M, Crow JF, 1964. The number of alleles that can be maintained in a finite population. *Genetics* 49: 725-38
- Kole C, 2006. *Cereals and millets* Springer, Verlag Berlin Heidelberg
- Levene H, 1949. On a matching problem arising in genetics. *Ann Math Stat* 20: 91-94
- Lewontin RC, 1972. Testing the theory of natural selection. *Nature* 236(5343): 181-182
- Liu K, Goodman M, Muse S, Smith JS, Buckler E, Doebley J, 2003. Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics* 165(4): 2117-2128
- Lu H, Bernardo R, 2001. Molecular marker diversity among current and historical maize inbreds. *Theor Appl Genet* 103(4): 613-617
- Lukas FK, Donald MW, 2002. Inbreeding effects in wild populations. *Trends Ecol Evol* 17(5): 230-241
- Mantel N, Valand RS, 1970. A technique of nonparametric multivariate analysis. *Biometrics* 26(3): 547-58
- Nei M, 1972. Genetic distance between populations. *Amer Natur* 106(949): 283-292
- Nei M, 1973. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* 70(12): 3321-3323
- Nei M, 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89(3): 583-590
- Nei M, 1987. *Molecular evolutionary genetics*. Columbia University Press, New York
- Patto MC, Satovic Z, Pêgo S, Fevereiro P, 2004. Assessing the genetic diversity of Portuguese maize germplasm using microsatellite markers. *Euphytica* 137(1): 63-72
- Peakall ROD, Smouse PE, 2006. *Genalex 6: Genetic analysis in Excel. Population genetic software for teaching and research*. Blackwell Science Ltd
- Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES, 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci USA* 98(20): 11479-11484
- Rohlf F, 2002. *NTSYS-pc: Numerical Taxonomy System, version 2.1*. Exeter Publishing Ltd., Setauket, New York, USA
- Rohlf F, Fisher D, 1968. Test for hierarchical structure in random data sets. *Syst Zool* 17: 407-412
- Rupp JV, Mangolin CA, Scapim CA, Machado MF, 2009. Genetic structure and diversity among sweet corn (su1-germplasm) progenies using SSR markers. *Maydica* 54: 125-132
- Senior ML, Murphy JP, Goodman MM, Stuber CW, 1998. Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. *Crop Sci* 38: 1088-1098
- Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Ziegler J, 1997. An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPS and pedigree. *Theor Appl Genet* 95(1): 163-173
- Solomon K, Martin I, Zeppa A, 2010. Temporal genetic structure patterns in tropical maize populations under reciprocal recurrent selection. *Euphytica* 176(2): 239-249
- Srdić J, Nikolić A, Pajić Z, Drinić SM, Filipović M, 2011a. Genetic similarity of sweet corn inbred lines in correlation with heterosis. *Maydica* 56(3): 251-256
- Srdić J, Pajić Z, Filipović M, Babić M, Sečanski M, 2011b. Inheritance of ear yield and its components in sweet corn (*Zea mays* L. *saccharata*). *Genetika* 43(2): 341-348
- Taramino G, Tingey S, 1996. Simple sequence repeats for germplasm analysis and mapping in maize. *Genome* 39(2): 277-87
- Tautz D, Renz M, 1984. Simple Sequences are Ubiquitous Repetitive Components of Eukaryotic Genomes. *Nucleic Acids Res* 12(10): 4127-4138
- Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS, 2001. Patterns of DNA Sequence Polymorphism along Chromosome 1 of Maize (*Zea mays ssp mays* L.). *Proc Natl Acad Sci USA* 98(16): 9161-9166
- Terra TF, Wiethölter P, de Souza ACC, Silva SD, Bered F, Sereno MJC, Barbosa NJF, 2011. Genetic variability in maize and teosinte populations estimated by microsatellites markers. *Cienc Rural* 41(2): 205-211
- Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES, 2001. Dwarf polymorphisms associate with variation in flowering time. *Nat Genet* 28(3): 286-289
- Troyer AF, 2001. Temperate corn, pp. 393-466. In: *Specialty Corns*. Hallauer AR eds. CRC Press,

- Boca Raton, FL
- Varshney RK, Hoisington DA, Tyagi AK, 2006. Advances in cereal genomics and applications in crop breeding. *Trends Biotechnol* 24(11): 490-499
- Wright S, 1978. *Evolution and the genetics of populations: Variability within and among populations*. University of Chicago Press, Chicago
- Xu Y, 2010. *Molecular plant breeding*. CAB International, Oxfordshire, UK
- Yeh FC, Yang RC, Boyle T, 1999. *Popgene: Microsoft Windows-based freeware for population genetic analysis release 1.31*. University of Alberta, Alberta, Canada
- Zhao W, Wang Y, Chen T, Jia G, Wang X, Qi J, Pang Y, Wang S, Li Z, Huang Y, Pan Y, Yang Y-H, 2007. Genetic structure of mulberry from different ecotypes revealed by ISSRs in China: An implications for conservation of local mulberry varieties. *Sci Hortic* 115(1): 47-55
- Zhong-hu H, 1991. An investigation of the relationship between the F_1 potential and the measures of genetic distance among wheat lines. *Euphytica* 58(2): 165-170



Supplementary Figure 1. Ninety nine microsatellite markers utilized for characterization of tropical sweet corn inbred lines, and their locations in the genome

** = Monomorphic marker*

Supplementary Table 1. List of inbred lines selected, their source populations, countries of origin and population structures

Inbred Line	Source Population	Country of Origin	Source Population Structure
BAK-S8	Bakti-1	Malaysia	Composite
TSG-S8	Thai Super Sweet	Thailand	Composite
TSK-S8	Thai Super Sweet	Thailand	Composite
TSN-S8	Thai Super Sweet	Thailand	Composite
TFZ-S8	Hybrid 240	USA	Single cross hybrid
DFS-S8	Hybrid D56	Taiwan	Single cross hybrid
IND-S8	Indo	Indonesia	Composite
MAS-S8	Mas Madu	Malaysia	Composite
MAN-S8	Manis Madu	Malaysia	Composite
MMIN-S8	Manis Madu × Indo	Malaysia	Cross between Manis Madu and Indo
THI-S8	Thailand	Thailand	Composite
SBY-S8	Surabaya	Indonesia	Composite
HAW-S8	Hawaiian	USA	Single cross hybrid