



Genetic diversity among tropical provitamin A maize inbred lines and implications for a biofortification program

Journal:	<i>Plant Breeding</i>
Manuscript ID	Draft
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Sserumaga, Julius; National Agricultural Research Organisation, Cereals Makumbi, Dan; CIMMYT, Warburton, Marilyn; Mississippi State University, Opio, Stephen Asea, Godfrey; National Agricultural Research Organisation, Cereals Muwonge, Abu-baker; National Agricultural Research Organisation, Cereals Kasozi, Charles; National Crops Resources Research Institute, Cereals Programme
Keywords:	Maize, Inbred, SNP, Provitamin

Genetic diversity among tropical provitamin A maize inbred lines and implications for a biofortification program

Running Head: Diversity among maize inbred lines

Julius P. Sserumaga^{1*}, Dan Makumbi², Marilyn L. Warburton³, Stephen O. Opiyo⁴, Godfrey Asea¹, Abu-Baker Muwonge¹, and Charles L. Kasozi¹

¹Cereals Program, National Agricultural Research Organization; National Crops Resources Research Institute, Namulonge, P. O. Box 7084 Kampala, Uganda

²International Maize and Wheat Improvement Center (CIMMYT), P.O. Box 1041-00621, Nairobi, Kenya

³USDA ARS Corn Host Plant Resistance Research Unit, Mississippi State, MS 39762.

⁴Molecular and Cellular Imaging Center, Ohio State University, Columbus, OH, 43210

*Correspondence author: j.serumaga@gmail.com

Abstract

Insights into the diversity and relationships among elite breeding materials are an important component in maize improvement programs. We genotyped 63 inbred lines bred for high levels of provitamin A using 137 single nucleotide polymorphism markers. A total of 272 alleles were detected with gene diversity of 0.36. Average genetic distance was 0.36 with 56% of the pairs of lines having between 0.30 and 0.40. Eighty-six percent of the pairs of lines showed relative kinship values <0.50 , which indicated that the majority of these provitamin A inbred lines were unique. Relationship pattern and population structure analysis revealed presence of seven major groups with good agreement with NeighbourJoining clustering but not clearly correlated with pedigree. Utilization of this set provitamin A lines in a new biofortification program will be aided by information from both molecular-based grouping and pedigree analysis. The results should guide breeders in selecting parents for hybrid formation and testing as a short-term objective, and parents with diverse alleles for new breeding starts as a long-term objective in a provitamin A breeding program.

Key Words; Maize, inbred, SNP, Provitamin,

1.0 INTRODUCTION

In sub-Saharan Africa (SSA), maize is the principal staple crop, accounting for an average of 32% of consumed calories in Eastern and Southern Africa, rising to 51% in some countries (Cairns et al., 2012), and it is also a source of income. In the majority of countries in SSA the maize varieties grown are predominantly white grain with low levels of protein, fat, minerals, and micronutrients including provitamin A carotenoids. Although yellow maize is grown and consumed throughout the world including some countries in SSA, most of these varieties have less than $2 \mu\text{g g}^{-1}$ of provitamin A carotenoids (Pixley et al., 2013). In developing countries, vitamin A deficiency (VAD) affects up to 140 million children and pregnant women, and can cause blindness, immune system deficiency, and stunting of growth (West, 2002). Developing maize varieties with improved grain quality traits can be achieved through biofortification, which may either involve the use of natural genetic variation existing in the genepool, or by introduction of exotic germplasm. Biofortification of staple crops such as maize with high provitamin A carotenoids through conventional breeding leads to provitamin A enriched maize varieties that could be used to reduce VAD (Adeyemo et al., 2011). Provitamin A maize endosperm contains varying amounts of the provitamin A carotenoids, which include α -carotene, β -carotene, and β -cryptoxanthin, but the concentrations are very low (Adeyemo et al., 2011; Menkir et al., 2017). Some variation for specific carotenoid content has been reported in tropical adapted maize inbred lines (Adeyemo et al., 2011; Menkir et al., 2017). Since the tropical adapted yellow maize varieties grown in SSA contain small amounts of provitamin A carotenoids, there is a need to breed maize for high provitamin A to meet human nutritional requirements (Adeyemo et al., 2011; Brunson and Quackenbush, 1962). To achieve this goal, introduction of maize rich in provitamin A carotenoids into tropical maize breeding programs is necessary.

Effective utilization of completely new and uncharacterized maize inbred lines in an existing breeding program requires an understanding of their genetic relationships to identify possible heterotic group placement within the available germplasm. Use of pedigree information is beneficial for grouping of inbred lines when available and when lines are closely related (Liu et al., 2003; Olmos et al., 2014; Sserumaga et al., 2014). Genetic relationships among a group of unrelated lines can also be investigated using molecular genetic markers. Because of low cost per data point, high genomic abundance, locus-specificity, co-dominance, potential for high

throughput analysis, and lower genotyping error rates (Rafalski, 2002), single nucleotide polymorphism (SNP) markers are a powerful tool for many genetic applications, including genetic diversity studies, linkage and quantitative trait loci (QTL) mapping, and genome-wide association studies (Azmach et al., 2013; Farfan et al., 2015; Suwarno et al., 2015). There is a strong correlation between the molecular marker and pedigree-based distance measures in many plant species, including examples in maize (Liu et al., 2003; Olmos et al., 2014).

In Uganda, like many other countries in sub-Saharan Africa, biofortification has been viewed as a strategy for reducing micronutrient deficiencies. Although there has been great deal of discussion about biofortification as a tool for combating VAD, progress in establishing fortification programs has been slow (Fiedler and Afidra, 2010). In response to the emerging demand for biofortified maize, an array of provitamin A maize inbred lines from the International Maize and Wheat Improvement Center (CIMMYT) in Mexico were introduced into Uganda to start a provitamin A carotenoid maize breeding program. There is a need for molecular characterization of these inbred lines to assess their utility in hybrid breeding. The objectives of this study were to assess (i) the level of molecular diversity and population structure among 63 provitamin A maize inbred lines using 137 SNP markers, and (ii) the relationships among the set of maize lines for better exploitation in a breeding program.

2.0 MATERIALS AND METHODS

2.1 Genetic materials and SNP genotyping

Sixty-three maize inbred lines obtained from CIMMYT's provitamin A carotenoid maize breeding program in Mexico were used in this study (Table 1). These lines were selected based on high provitamin content and combining ability. The procedure for developing these inbred lines was described in detail by Pixley et al. (2013). A single seed per inbred line was grown in the greenhouse at the National Crops Resources Research Institute (NaCRRI), Namulonge, Uganda, up to the 3-4 leaf stage. Leaf samples were harvested following the leaf sampling protocol from LGC Genomics (<http://www.lgcgroup.com/our-science/genomics-solutions/#.WXpE7ITyu70>) using the plant sample collection kit from LGC Genomics (Middlesex, UK). The leaf samples were sent to LGC Genomics in the UK for DNA extraction using LGC's beadex™ extraction chemistry and genotyping using KASP platform. The 63 samples were genotyped with 142 SNPs and the results were visualized through the SNPviewer

software (<https://www.lgcgroup.com/products/genotyping-software/snpviewer>). The SNPs selected for this study were recommended for routine low-cost genotyping (Semagn et al., 2012) and were well distributed across all chromosomes.

2.2 Data analysis

Gene diversity, polymorphic information content (PIC), allele frequency, and relative kinship were computed using the TASSEL software package (Bradbury et al., 2007). Genetic distance (GD) between lines was calculated based on Rogers distance (Rogers, 1972) using PowerMarker version 3.25 (Liu and Muse, 2005). A dendrogram was constructed from the genetic distance matrix using the neighbor-joining technique in PowerMarker, and the resulting trees were visualised using MEGA version 5.0 (Tamura et al., 2011). The model-based clustering approach to analyze population structure was implemented in STRUCTURE software package (Pritchard et al., 2000). To estimate the correct number of clusters in the population of 63 lines using posterior probabilities (qK), a 100,000 burn-in period was used, followed by 100,000 iterations and a model allowing for admixture and correlated allele frequencies with no a priori location or population information. At least 10 independent runs of STRUCTURE were performed by setting the number of clusters (K) from 1 to 10, with 10 replicates for each K. The delta K was calculated for each value of K using the Structure Harvester software (Evanno et al., 2005). Structure analysis was run for different values of K. Each inbred line was assigned to a given cluster when the proportion of its genome in the cluster (qK) was higher than a threshold value of 50%. Finally, principal component analysis (PCA) was carried out to illustrate the relationships among the populations based on STRUCTURE results using a three-dimensional plot generated by R package Scatterplot3d (Ligges and Mächler, 2002).

3. RESULTS

3.1 SNP characterization, genetic distance and relationships

Of the 142 SNPs used for genotyping, 137 showed good amplification in allele calls and were used in further analyses. The 137 SNPs detected a total of 272 alleles, with an average of two alleles detected as expected. Minor allele frequency ranged from 0 to 0.49 (Supplementary Table 1). Gene diversity in this set of lines ranged from 0 to 0.50 with an average of 0.36 while heterogeneity ranged from 0 to 0.18 with an average of 0.05. The PIC value ranged from 0 to

0.37, with an average of 0.29. Genetic distance between pairs of inbred lines ranged from 0 to 0.54, with an average of 0.36. Most (34.2%) pairs of lines in the present study had genetic distances between 0.35 and 0.40 (Supplementary Figure 1). Relative kinship coefficients between pairs of lines ranged from 0 to 0.66, with an average of 0.42 (Supplementary Figure 2), but most values (86%) fell between 0.35 and 0.50.

3.2 Population Structure and Principle Component Analysis

The neighbor-joining (NJ) tree generated from Rogers GD matrix was constructed to gain more insight into the genetic diversity among the set of maize lines. The dendrogram grouped the 63 lines into three major groups with seven subgroups (Figure 1). The first major group was comprised of four lines. The second major group, which consisted of 35 lines, was split into four subgroups. The third major group was comprised of 24 lines and three subgroups. The model-based approach of STRUCTURE was also implemented to infer population structure for the 63 lines. The output from STRUCTURE based on $(\text{LnP}(D))$ and ΔK suggested the presence of either 2, 5, or 7 subgroups (clusters) among the provitamin A lines (Figure 2). When $K = 7$, there was excellent agreement with the seven subgroups of the NJ tree constructed with the genetic distance matrix (Figure 3c). The membership of the lines in the 7 subgroups were 6% (group 1), 25% (group 2), 16% (group 3), 19% (group 4), 22% (group 5), 3% (group 6) and 6% (group 7) (Figure 3c).

The results from $K = 7$ were compared with pedigree and breeding history to characterize the clusters, and it was seen that the STRUCTURE and NJ cluster classifications in many cases placed lines descended from a common ancestor in several groups. For example, lines derived from CML537 were in groups 2 and 4. Two lines in group 4 (entries 45 (CLHP0235) and 46 (CLHP0221)) involved a backcross to CML537 and the donor for carotenoid concentration, while the two lines in group 2 (entries 47 (CLHP0213) and 46 (CLHP0003)) differed from those in group 2 based on the donor for carotenoid concentration. Therefore, the distribution of these lines across the two groups could be attributed to the breeding history in carotenoid conversion program. A similar scenario was applicable to derivatives from line CLQRCWQ109 (entries 17, 32, and 38) that were spread across groups 2, 3 and 5, and for the majority of lines from CML506 (entries 8, 10, 14, 26, and 33) (Table 1). The exceptions were lines derived from CML496 and CML486 that were placed in two separate groups despite the fact that the same

donor line for carotenoid concentration was used in each case. Thus, clusters must be assumed to form based on the combined contribution of many parents, and are not formed based on any one ancestral line or phenotypic trait (such as carotenoid concentration). The clusters formed by both NJ dendrogram and STRUCTURE analysis were fairly clearly delineated, but a PCA of these data (first three PCs explained 25.5% of the total SNP variation among the lines for both K=5 and K=7) did not show clear separations at all (Figure 4).

4. DISCUSSION

Utilization of diverse sources of provitamin A carotenoid content is important for a new biofortified maize genetic enhancement program. Use of molecular markers to infer genetic diversity in yellow endosperm and provitamin maize been reported in many studies (Menkir et al., 2006; Suwarno et al., 2014; Adeyemo & Omidiji, 2014). The number of alleles recorded in this study was higher than that reported in studies with diverse tropical yellow inbred maize lines (Adeyemo and Omidiji, 2014; Badu-Apraku et al., 2013; Muthusamy et al., 2014). The average gene diversity recorded in this study was lower than that reported by Muthusamy et al. (2014) in yellow/carotenoid maize and Sserumaga et al. (2014) in tropical white maize but higher than that reported in other studies (Dao et al., 2014; Van Inghelandt et al., 2010). The GD between pairs of inbred lines in this study was smaller than that reported in most other studies (Adeyemo et al., 2011; Menkir et al., 2006; Muthusamy et al., 2014) but comparable to some (Suwarno et al., 2014; Semagn et al., 2012). Large GD estimates (> 0.5) between some of the pairs of provitamin A lines in this set suggests that there is reasonable diversity to choose from, and possibly high levels of heterosis between those pairs with high GD. Molecular marker distance may not predict heterosis accurately in cases where the lines are distantly related (Melchinger et al., 1990; Reif et al., 2003), but the pairs with very low distances will probably yield no heterosis, and these crosses can be avoided based on the molecular data. To determine how these exotic provitamin A lines should be classified for use in SSA, they must now be crossed to testers of known SSA maize heterotic groups. For example, the well-known white grain maize testers CML312 (heterotic group A) and CML444 (heterotic group B) that have since been converted to yellow endosperm (CIMMYT, unpublished data) would be some of the suitable testers to use. Suwarno et al. (2014) proposed the use of molecular marker-based GD as way to formulate heterotic groups among provitamin A maize lines developed at CIMMYT.

The relative kinship coefficient reflects the approximate degree of identity between two given individuals (Dao et al., 2014; Hardy and Vekemans, 2002), and the lines in this study appear to have a low proportion of common alleles. While the kinship between these lines are higher than in many studies (Wen et al. 2011; Semagn et al. 2012; Wu et al., 2016) the lines in this study were not chosen for diversity, but rather, for good expression of a single trait. This would normally cause a good deal of common alleles, but each line in this study is potentially contributing unique alleles at many of the loci surveyed which should reflect a more diverse genetic pool for breeding provitamin A maize for mid-altitude Africa. The clustering of maize lines based on pedigree or origin is rarely straightforward, unless a group of lines have been selected specifically to be diverse and represent different breeding programs, countries of origin, or growing environment (for example, Suwarno et al., 2015; Lu et al., 2009). Separations based on pedigree or selected traits are rarely found (Warburton et al., 2002; Xia et al., 2005), especially in smaller populations such as the present study. To structure a biofortification breeding program that would utilize this set of provitamin A lines, it will be important to make use of both molecular-based grouping and pedigree of the lines where molecular information does not distinguish lines. In addition, a few representative lines from each of the 7 clusters could be crossed with testers from existing African heterotic groups to help determine which will form the best hybrids; further testcrossing with the most productive heterotic patterns in an efficient and directed manner could identify new productive, adapted provitamin A hybrids. This should allow the use of the molecular data presented here to assist in decision making regarding the lines to be used for hybrid formation and testing as a short-term objective, and which lines to use for breeding starts as a long-term objective.

In conclusion, there was moderate genetic distance among some of the provitamin A lines used in this study, more than expected due to the common parents used as donor lines for carotenoid content. The results of this study indicated that a combination of molecular marker analyses, pedigree/breeding history, and directed testcrossing is needed to assign inbred lines to sub-groups for breeding decisions. There is substantial genetic variation in this set of provitamin A lines which will be useful for a new maize biofortification program in Uganda.

ACKNOWLEDGMENTS

This research was supported by the Government of Uganda through the Agricultural Technology and Agribusiness Advisory Services (ATAAS) project financed by the World Bank. Special thanks go to Dr. Thanda Dhliwayo at CIMMYT for providing the inbred lines, and to Dr. Natalia Palacios for providing the information on provitamin A content of the inbred lines. The authors would like to thank Moses Wasswa and Fred Ssemazzi for assistance during sample collection and preparation. We gratefully thank Drs. Sharon Mitchell and Alex Casa for thoughtful review of the manuscript.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

CONTRIBUTION OF AUTHORS

Julius P. Sserumaga and Dan Makumbi conceived and designed the study; Godfrey Asea was involved in the conception and design of the study. Abu-baker Muwonge and Kasozi selected the accessions to be included in the study and collected plant material for the study. Marilyn Warburton, Julius P. Sserumaga and Stephen O. Opiyo analyzed and interpreted the data. Dan Makumbi, Marilyn Warburton, Julius P. Sserumaga and Stephen O. Opiyo drafted and edited the manuscript and all authors helped to edit, reviewed and approved the final manuscript.

ORCID

- Dan Makumbi – 0000-0002-1801-5986
- Julius Pyton Sserumaga – 0000-0001-5152-2835
- Marilyn L. Warburton – 0000-0002-9542-9912
- Opiyo Stephen – 0000-0001-7933-9004
- Godfrey Asea – 0000-0003-1441-2199
- Charles Lwanga Kasozi – 0000-0002-8558-779X

FIGURE LEGENDS

FIGURE 1 Neighbor-joining tree for 63 inbred lines based on Rogers genetic distance calculated from 137 SNP markers. The subgroups are indicated with different colors. Details of the different subgroups are given in Supplementary Table 1.

FIGURE 2: Plot of changes in ΔK value with the number of subpopulations.

FIGURE 3a: Population structure among individuals with $\Delta K = 2$.

FIGURE 3b Population structure among individuals with $\Delta K = 5$.

FIGURE 3c Population structure among individuals with $\Delta K = 7$.

FIGURE 4 Principal component analysis (PCA) of 63 provitamin A maize inbred lines based on population structure when $\Delta K = 5$ (A) and $\Delta K = 7$ (B).

REFERENCES

- Adeyemo, O., Menkir, A., Melaku, G., & Omidiji, O. (2011). Genetic diversity assessment and relationship among tropical yellow endosperm maize inbred lines using SSR markers. *Maydica*, 56, 1703.
- Adeyemo, O., & Omidiji, O. (2014). SSR-based and carotenoid diversity assessment of tropical yellow endosperm maize inbred lines. *Plant Genetic Resources: Characterization and Utilization*, 12, 67-73.
- Azmach, G., Melaku, G., Menkir, A., & Spillane, C. (2013). Marker-trait association analysis of functional gene markers for provitamin A levels across diverse tropical yellow maize inbred lines. *BMC Plant Biology*, 13, 227.
- Badu-Apraku, B., Oyekunle, M., Fakorede, M.A.B., Vroh, I., O Akinwale, R., & Aderounmu, M. (2013). Combining ability, heterotic patterns and genetic diversity of extra-early yellow inbreds under contrasting environments. *Euphytica*, 192, 413-433.
- Bradbury, P., Zhang, Z., Kroon, D., Casstevens, T., Ramdoss, Y., & Buckler, E. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23, 2633-2635.
- Brunson, A., & Quackenbush, F. (1962). Breeding corn with high provitamin A in the grain. *Crop Science*, 2, 344-347.
- Cairns, J.E., Sonder, K., Zaidi, P.H., Verhulst, N., Mahuku, G., Babu, R., Nair, S.K., Das, B., Govaerts, B., Vinayan, M.T., et al. (2012). Maize production in a changing climate: Impacts, adaptation, and mitigation strategies. *Advances in Agronomy*, 114, 1-58.
- Dao, A., Sanou, J., Mitchell, S.E., Gracen, V., & Danquah, E.Y. (2014). Genetic diversity among INERA maize inbred lines with single nucleotide polymorphism (SNP) markers and their relationship with CIMMYT, IITA, and temperate lines. *BMC Genetics*, 15, 127.
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14, 2611-2620.
- Farfan, I.D.B., De La Fuente, G.N., Murray, S.C., Isakeit, T., Huang, P.-C., Warburton, M., Williams, P., Windham, G.L., & Kolomiets, M. (2015). Genome wide association study for drought, aflatoxin resistance, and important agronomic traits of maize hybrids in the subtropics. *PLoS One*, 10, e0117737.
- Fiedler, J.L., & Afidra, R. (2010). Vitamin A fortification in Uganda: Comparing the feasibility, coverage, costs, and cost-effectiveness of fortifying vegetable oil and sugar. *Food and Nutrition Bulletin*, 31, 193-205.

- Hardy, O., & Vekemans, X. (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, 2, 618–620.
- Ligges, U., & Mächler, M. (2002). Scatterplot3d - an R package for visualizing multivariate data. *Journal of Statistical Software* 8, 1–20.
- Liu, K., & Muse, S.V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, 21, 2128–2129.
- Liu, K., Goodman, M., Muse, S., Smith, J., Buckler, E., & Doebley, J. (2003). Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics*, 165, 2117–2128.
- Lu, Y., J. Yan, C. Guimarães, S. Taba, Z. Hao, S. Gao, S. Chen, J. Li, S. Zhang, B. Vivek, C. Magorokosho, S. Mugo, D. Makumbi, S. Parentoni, T. Shah, T. Rong, J. Crouch, and Y. Xu, 2009: Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. *Theoretical and Applied Genetics*, 120, 93–115.
- Melchinger, A., Lee, M., Lamkey, K., Hallauer, A., & Woodman, W. (1990). Genetic diversity for restriction fragment length polymorphisms and heterosis for two diallel sets of maize inbreds. *Theoretical and Applied Genetics*, 80, 488–496.
- Menkir, A., Maziya-Dixon, B., Mengesha, W., Rocheford, T., & Alamu, E.O. 2017. Accruing genetic gain in pro-vitamin A enrichment from harnessing diverse maize germplasm. *Euphytica*, 213:105.
- Menkir, A., Liu, W., White, W.S., Maziya-Dixon, B., & Rocheford, T. (2008). Carotenoid diversity in tropical-adapted yellow maize inbred lines. *Food Chemistry*, 109, 521–529.
- Menkir, A., Olowolafe, M.O., Ingelbrecht, I., Fawole, I., Badu-Apraku, B., & Vroh, B.I. (2006). Assessment of testcross performance and genetic diversity of yellow endosperm maize lines derived from adapted × exotic backcrosses. *Theoretical and Applied Genetics*, 113, 90–99.
- Muthusamy, V., Hossain, F., Thirunavukkarasu, N., Choudhary, M., Saha, S., Bhat, J.S., Prasanna, B.M., & Gupta, H.S. (2014). Development of β-carotene rich maize hybrids through marker-assisted introgression of β-carotene hydroxylase allele. *PLoS One* 9, e113583.
- Olmos, S., Delucchi, C., Ravera, M., Negri, M., Mandolino, C., & Eyherabide, G. (2014). Genetic relatedness and population structure within the public Argentinean collection of maize inbred lines. *Maydica*, 59, 16–31.
- Pixley, K., Palacios-Rojas, N., Babu, R., Mutale, R., Surles, R., & Simpungwe, E. (2013). Biofortification of maize with provitamin A carotenoids. In S.A. Tanumihardjo (Ed), *Carotenoids and Human Health*, (pp. 271–292). Springer Science + Business Media, New York, NY.
- Pritchard, J., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- Pritchard, J., Stephens, M., & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155, 945 - 959.
- Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology*, 5, 94–100.
- Reif, J., Melchinger, A., Xia, X., Warburton, M., Hoisington, D., & Vasal, S. (2003). Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. *Crop Science*, 43, 1275–1282.

- Rogers, J.S. (1972). Measures of genetic similarity and genetic distance. In: Studies in Genetics VII, University of Texas Publication 7213, Austin, 145-153.
- Semagn, K., Magorokosho, C., Vivek, B., Makumbi, D., Beyene, Y., Mugo, S., Prasanna, B., & Warburton, M. (2012). Molecular characterization of diverse CIMMYT maize inbred lines from eastern and southern Africa using single nucleotide polymorphic markers. *BMC Genomics*, 13, 113.
- Sserumaga, J. P., Makumbi, D., Hyeonso, J., Kiarie, N., James, W. M., George, N. C. W., Lee, S.-M., Godfrey, A., & Hakbum, K. (2014). Molecular characterization of tropical maize inbred lines using microsatellite DNA markers. *Maydica*, 59, 267–274.
- Suwarno, W.B., Pixley, K.V., Palacios-Rojas, N., Kaeppler, S.M., & Babu, R. (2014). Formation of heterotic groups and understanding genetic effects in a provitamin A biofortified maize breeding program. *Crop Science*, 54, 14–24.
- Suwarno, W.B., Pixley, K.V., Palacios-Rojas, N., Kaeppler, S.M., & Babu, R. (2015). Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. *Theoretical and Applied Genetics*, 128, 851–864.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, 28, 2731–2739.
- Van Inghelandt, D., Melchinger, A., Lebreton, C., & Stich, B. (2010). Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theoretical and Applied Genetics*, 120, 1289–1299.
- Warburton, M.L., Xia, X.C., Crossa, J., Franco, J., Melchinger, A.E., Frisch, M., Bohn, M. & Hoisington, D.A. (2002). Genetic characterization of CIMMYT maize inbred lines and open pollinated populations using large scale fingerprinting methods. *Crop Science* 42:1832–1840.
- West, K.P. Jr. (2002). Extent of vitamin A deficiency among preschool children and women of reproductive age. *Journal of Nutrition*, 132, 2857S–2866S.
- Wu, Y., San Vicente, F., Huang, K., Dhliwayo, T., Costich, D. E., Semagn, K., Sudha, N., Olsen, M., Prasanna, B. M., Zhang, X., & Babu, R. (2016). Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. *Theoretical and Applied Genetics*, 129, 753–765.
- Xia, X.C., Reif, J.C., Melchinger, A.E., Frisch, M., Hoisington, D.A., Beck, D., Pixley, K. & Warburton, M.L. (2005). Genetic diversity among CIMMYT maize inbred lines investigated with SSR markers. II. Subtropical, tropical midaltitude, and highland maize inbred lines and their relationships with elite U.S. and European maize. *Crop Science* 45:2573–2582.

Table 1 - List of inbred lines used in the study.

No.	Entry	Pedigree Information	Pro A µg/g DW
1	CML304	CML304-B-B-B-B	8.13
2	CML486	CML486-B-B-B	0.72
3	CML451	CML451-B-B-B	0.80
4	CLHP00306	(KUIcarotenoidsyn-FS17-3-1-B-B-B-B//((KUIcarotenoidsyn-FS17-3-1-B-B-B/(CML297-B×KUICarotenoidsyn-FS17-3-2-B/KUI3×SC55)))-plant19(HH)-3-1-B-B-B-B-B	13.34
5	CLHP00478	(CLQRCWQ97-B//((KUIcarotenoidsyn-FS17-3-2-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B)//CarotenoidSyn3-FS11-4-3-B-B-B))-B-25-6-B-B-B-B-B	11.71
6	CLHP00476	(CLQRCWQ97-B//((KUIcarotenoidsyn-FS17-3-2-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B)//CarotenoidSyn3-FS11-4-3-B-B-B))-B-25-2-B-B-B-B-B	No data
7	CLHP00446	((([B104xCLRQ00502]-F2)-03-01-K1/CarotenoidSyn3-FS8-4-3-B-B-B-B)//((([B104xCLRQ00502]-F2)-03-01-K1/CarotenoidSyn3-FS8-4-3-B-B-B-B)//(KUIcarotenoidsyn-FS17-3-2-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B)-B-4)-7-1-B-B-B-B-B	8.20
8	CLHP0310	B-B-B-B-B)-6-3-B-B-B-B-B	9.71
9	CLHP0290	([DTPYC9-F65-2-3-1-1-B-BxDTYPYC9-F65-2-2-1-1-B-B]-3-4-2-B-B-B//([DTPYC9-F65-2-3-1-1-B-BxDTYPYC9-F65-2-2-1-1-B-B]-3-4-2-B-B/(CML297-B×KUICarotenoidsyn-FS17-3-2-B/KUI3×B77)))-plant3(H)-1-3-B-B-B-B-B	No data
10	CLHP00308	((CML506//CML506/CML305)//((CML506//CML506/CML305)/(KU1409/DE3/KU1409)S2-18-2-B-B-B)-B)-11-1-B-B-B-B-B	No data
11	CLHP0302	(KUIcarotenoidsyn-FS17-3-1-B-B-B-B//((KUIcarotenoidsyn-FS17-3-1-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B))-B-plant4(H)-2-2-B-B-B-B-B	12.20
12	CLHP0352	([DTPYC9-F65-2-3-1-1-B-BxDTYPYC9-F65-2-2-1-1-B-B]-3-4-2-B-B/(CML297-B×KUICarotenoidsyn-FS17-3-2-B/KUI3×B77))-B-21-1-B-B-B-B-B	No data
13	CLHP00294	(KUIcarotenoidsyn-FS17-3-1-B-B-B-B//((KUIcarotenoidsyn-FS17-3-1-B-B-B/(CML297-B×KUICarotenoidsyn-FS17-3-2-B/KUI3×SC55)))-plant19(HH)-4-2-B-B-B-B-B	9.81
14	CLHP0312	((CML506//CML506/CML305)//((CML506//CML506/CML305)/(KU1409/DE3/KU1409)S2-18-2-B-B-B)-B)-7-1-B-B-B-B-B	No data
15	CLHP00328	([[[NAW5867/P30SR]-43-2/[NAW5867/P30SR]-114-1]-9-3-3-B-1-B/CML395-1]-B-13-1-B-4-#[BETASYN]BC1-8-1-1-1-B-B/(CML297-B×KUICarotenoidsyn-FS17-3-2-B/KUI3×B77))-B-1-1-B-B-B-B-B	No data

16	CLHP0334	((DTPYC9-F65-2-3-1-1-B-BxDTPYC9-F65-2-2-1-1-B-B)-3-4-2-B-B/(CML297-BxKUICarotenoidsyn-FS17-3-2-B/KUI3xB77))-B-12-1-B-B-B-B-B	13.23
17	CLHP0452	(CLQRCWQ109-B/////((CarotenoidSyn3-FS11-4-3-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B))/MAS[MSR/312]-117-2-2-1-B-B-B/(BETASYN)BC1-11-5-2-1-3-B))-B-2-3-B-B-B-B	No data
18	CLHP00286	(CML488/[BETASYN]BC1-13-1-1-1//G9BC1TSR-12P-1P-2P-3P-1-4-1-1-B-B-B-B-Bx CML488/[BETASYN]BC1-13-1-1//A.T.Z.T.V.C.82-1-3-1-1-1-B-B-B-B//A.T.Z.T.V.C.82-1-3-1-1-1-B-B-B-BGPO.Bx CML488/[BETASYN]BC1-13-1-1)-B-2-3-B-B-B-B-B	9.17
19	CLHP0303	(CML486/(CML297-BxKUICarotenoidsyn-FS17-3-2-B/KUI3xB77))-B-13-1-B-B-B-B-B	10.81
20	CLHP0301	(CML486/(CML297-BxKUICarotenoidsyn-FS17-3-2-B/KUI3xB77))-B-11-1-B-B-B-B-B	10.36
21	CLHP00448	((([B104xCLRQ00502]-F2)-03-01-K1/CarotenoidSyn3-FS8-4-3-B-B-B-B)////([B104xCLRQ00502]-F2)-03-01-K1/CarotenoidSyn3-FS8-4-3-B-B-B-B)////(KUIcarotenoidsyn-FS17-3-2-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B)-B-4)-7-3-B-B-B-B-B	No data
22	CLHP00340	([[[K64R/G16SR]-39-1/[K64R/G16SR]-20-2]-5-1-2-B*4/CML390]-B-38-1-B-7-#/BETASYN]BC1-1-1-1-#-B-B-B/(CML297-BxKUICarotenoidsyn-FS17-3-2-B/KUI3xSC55))-B-24-2-B-B-B-B-B	No data
23	CLHP0293	(CML496/(CML496/(CML297-BxKUICarotenoidsyn-FS17-3-2-B/A619xSC55)))-plant11(H)-1-1-B-B-B-B-B	No data
24	CLHP0289	(CML496/(CML496/(CML297-BxKUICarotenoidsyn-FS17-3-2-B/A619xSC55)))-plant11(HF)-1-3-B-B-B-B-B	No data
25	CLHP0299	(CML486/(CML297-BxKUICarotenoidsyn-FS17-3-2-B/KUI3xB77))-B-8-1-B-B-B-B-B	No data
26	CLHP00322	((CML506//CML506/CML305)////((CML506//CML506/CML305)/KUIcarotenoidsyn-FS25-3-2-B-B-B-B)-B)-5-1-B-B-B-B-B	4.25
27	CLHP00378	((([B104xCLRQ00502]-F2)-03-01-K1/CarotenoidSyn3-FS8-4-3-B-B-B-B)////([B104xCLRQ00502]-F2)-03-01-K1/CarotenoidSyn3-FS8-4-3-B-B-B-B)////(KUIcarotenoidsyn-FS17-3-2-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B)-B-4)-11-1-B-B-B-B-B	No data
28	CLHP0333	(CML491-B/////((KUIcarotenoidsyn-FS11-1-1-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B)////[NAW5867/P30SR]-40-1/[NAW5867/P30SR]-114-2]-16-2-2-B-2-B/CML395-6)-B-20-1-B-3-#/[BETASYN]BC1-3-1-1-#-B-B-B))-B-3-3-B-B-B-B	No data
29	CLHP0353	(CLQRCWQ97-B/////((KUIcarotenoidsyn-FS17-3-2-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B)////CarotenoidSyn3-FS11-4-3-B-B-B))-B-25-4-B-B-B-B-B	No data
30	CLHP00432	((([CLRQ00502xB109]-F2)-04-05-K1/KUIcarotenoidsyn-FS17-3-2-B-B-B-B)////([CLRQ00502xB109]-F2)-04-04-K1/CarotenoidSyn3-FS11-4-3-B-B-B-B	No data

		B)/(KUIcarotenoidsyn-FS11-1-1-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B)-B-2)-12-3-B-B-B-B-B	
		(CML491-B)/(KUIcarotenoidsyn-FS11-1-1-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B)/[[[NAW5867/P30SR]-40-1/[NAW5867/P30SR]-114-2]-16-2-2-B-2-B/CML395-6]-B-20-1-B-3-#[BETASYN]BC1-3-1-1-#-B-B-B)-B-3-1-B-B-B-B-B	3.04
31	CLHP0331	(CLQRCWQ109-B)/(KUIcarotenoidsyn-FS17-3-2-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B)/(CarotenoidSyn3-FS11-4-3-B-B-B)-B-9-3-B-B-B-B-B	7.95
32	CLHP0343	((CML506//CML506/CML305)/((CML506//CML506/CML305)/KUIcarotenoidsyn-FS17-3-2-B-B-B-B)-7-2-B-B-B-B-B	No data
33	CLHP0316	(CarotenoidSyn3-FS8-4-3-B-B-B/(CML297-BxKUICarotenoidsyn-FS17-3-2-B/A619xSC55))-B-36-2-B-B-B-B-B-B-B	3.77
34	CLHP0350	([[[NAW5867/P30SR]-40-1/[NAW5867/P30SR]-114-2]-16-2-2-B-2-B/CML395-6]-B-20-1-B-3-#[BETASYN]BC1-3-1-1-#-B-B-B/(CML297-BxKUICarotenoidsyn-FS17-3-2-B/A619xSC55))-B-24-1-B-B-B-B-B-B	6.38
35	CLHP0326	([DTPYC9-F65-2-3-1-1-B-BxDTPYC9-F65-2-2-1-1-B-B]-3-4-2-B-B/(CML297-BxKUICarotenoidsyn-FS17-3-2-B/KUI3xB77))-B-16-1-B-B-B-B-B-B	No data
36	CLHP0342	((KUIcarotenoidsyn-FS17-3-2-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B)/[[[NAW5867/P30SR]-40-1/[NAW5867/P30SR]-114-2]-16-2-2-B-2-B/CML395-6]-B-20-1-B-3-#[BETASYN]BC1-3-1-1-#-B-B-B)-B-18-1-B-B-B-B-B-B	No data
37	CLHP0360	(CLQRCWQ109-B)/(KUIcarotenoidsyn-FS17-3-2-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B)/(CarotenoidSyn3-FS11-4-3-B-B-B)-B-1Plant3(MAS:HF)-7-B-B-B-B-B-B	No data
38	CLHP0462	((([B100xCLQ06901]-F2)-02-04-K1/CarotenoidSyn3-FS8-4-3-B-B-B-B)/((([B100xCLQ06901]-F2)-02-04-K1/CarotenoidSyn3-FS8-4-3-B-B-B-B)/(KUIcarotenoidsyn-FS17-3-2-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B)-B-3)-4-B-B-B-B-B	No data
39	CLHP0390	((([CLRQ00502xB109]-F2)-04-05-K1/KUIcarotenoidsyn-FS17-3-2-B-B-B-B)/((([CLRQ00502xB109]-F2)-04-04-K1/CarotenoidSyn3-FS11-4-3-B-B-B-B)/(KUIcarotenoidsyn-FS11-1-1-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B)-B-2)-12-1-B-B-B-B-B	No data
40	CLHP00434	B-B	No data
41	CLHP0364	(KUIcarotenoidsyn-FS17-3-2-B-B-B-B/(CML297/(KU1409/DE3/KU1409)S2-18-2-B))-B-25-1-B-B-B-B-B-B	14.65
42	CLHP0404	((([B104xCLRQ00502]-F2)-03-01-K1/CarotenoidSyn3-FS8-4-3-B-B-B-B)/((([B104xCLRQ00502]-F2)-03-01-K1/CarotenoidSyn3-FS8-4-3-B-B-B-B)/(KUIcarotenoidsyn-FS17-3-2-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B)-B-3)-8-3-B-B-B-B-B	6.52

43	CLHP0159	((DTPYC9-F65-2-3-1-1 xDTPYC9-F65-2-2-1-1)xDTPYC9-F86-1-1-1-1 -B) -7-1 -B-B-B	No data
44	CLHP0102	[[GQL5/[GQL5/[MSRXPOOL9]C1F2-205-1(OSU23i)-5-3-X-X-1]F2-4sx]-8-4-B/[BETASYN]BC1-7-4-3-#[GQL5/[GQL5/[MSRXPOOL9]C1F2-205-1(OSU23i)-5-3-X-X-1]F2-4sx]-8-4-B/[BETASYN]BC1-7-4-3/KUISYN#]-4-1-1-3 -B-B	No data
45	CLHP0235	(CML537/[BETASYN]BC1-6-1-1-1-B//[CML537/[BETASYN]BC1-6-1-1-1/CML304)-5-1-2 -B-B (CML537/[BETASYN]BC1-11-3-1-#-B//[CML537/[BETASYN]BC1-11-3-1-#/CML300)-5-3-2 -	No data
46	CLHP0221	B-B-B	8.13
47	CLHP0213	(CML537/[BETASYN]BC1-10-2-1-#/CML-305-B)-B-11-1 -B-B	No data
48	CLHP0249	[CML226x[CATETODC1276/7619]]F2-25-1-B*7-B/([CML226x[CATETODC1276/7619]]F2-25-1-B*7/CML324//KUIcarotenoidsyn-FS11-1-1 -B)-3-3-1 -B-B	No data
49	CLHP0058	[[GQL5/[GQL5/[MSRXPOOL9]C1F2-205-1(OSU23i)-5-3-X-X-1-B-B]F2-4sx]-8-4-B/[BETASYN]BC1-7-4-3-#[GQL5/[GQL5/[MSRXPOOL9]C1F2-205-1(OSU23i)-5-3-X-X-1-B-B]F2-4sx]-8-4-B/[BETASYN]BC1-7-4-3/KUI+SC55SYN#]-6-1-1-1-B-B-B-B-B-B-B	4.44
50	CLHP0022	KUIcarotenoidsyn-FS25-3-2-B-B-B-B-B-B-B-B-B-B-B-B (KUIcarotenoidsyn-FS17-3-2-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B)-B-2(MAS:L4H1)-1-2-	8.75
51	CLHP0113	1-B-B-B-B-B (KUIcarotenoidsyn-FS11-1-1-B-B-B/(KU1409/DE3/KU1409)S2-18-2-B)-B-1(MAS:L4H1)-2-1-	8.35
52	CLHP0156	B-B-B-B (MAS[206/312]-23-2-1-1-B-B-B/[BETASYN]BC1-6-1-1-1-B//[MAS[206/312]-23-2-1-1-B-B-B/[BETASYN]BC1-6-1-1-1/CML304)-9-1-4-B-B-B-B-B-B	2.97
53	CLHP0237	MAS[206/312]-23-2-1-1-B-B-B/[BETASYN]BC1-11-3-1-#-B-B-B-B-B-B-B-B-B-B-B	9.71
54	CLHP0005	CML537/[BETASYN]BC1-10-3-#-B-B-B-B-B-B-B-B-B-B-B	5.21
55	CLHP0003	CarotenoidSyn3-FS8-4-3-B-B-B-B-B-B-B-B-B-B-B	6.35
56	CLHP0014	[CML197/N3//CML206]-X-32-1-4-B*5/[BETASYN]BC1-4-4-4-1-B-B-B-B-B-B-B-B-B-B	No data
57	CLHP0049	[[[NAW5867/P30SR]-43-2/[NAW5867/P30SR]-114-1]-9-3-3-B-1-B/CML395-1]-B-13-1-B-4-#[BETASYN]BC1-8-1-1-1-B-B-B-B-B-B-B	3.37
58	CLHP0068	KUIcarotenoidsyn-FS17-3-2-B-B-B-B-B-B-B-B-B-B-B	2.32
59	CLHP0020	[ZM305/BETASYN]BC2-133-1-3-B-B-B-B-B-B-B-B-B-B	10.46
60	CLHP0046	Ac8730SR-##-124-1-5-B-1-#[BETASYN]BC1-5-#-B-B-B-B-B-B-B-B-B-B	No data
61	CLHP0004	CML489/[BETASYN]BC1-2-#-B-B-B-B-B-B-B-B-B-B	No data
62	CLHP0002	CML488/[BETASYN]BC1-15-7-1-1-1-B-B-B-B-B-B-B	7.51
63	CLHP0006		No data

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

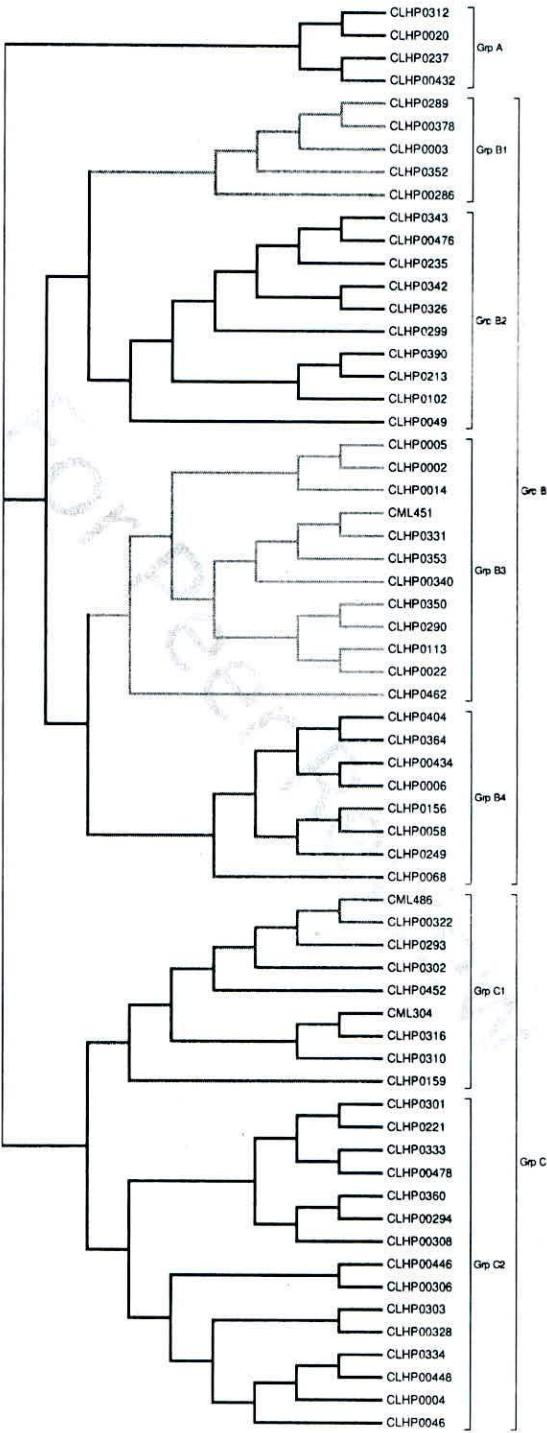


FIGURE 1 Neighbor-joining tree for 63 inbred lines based on Rogers genetic distance calculated from 137 SNP markers. The subgroups are indicated with different colors. Details of the different subgroups are given in Supplementary Table 1.

For Peer Review

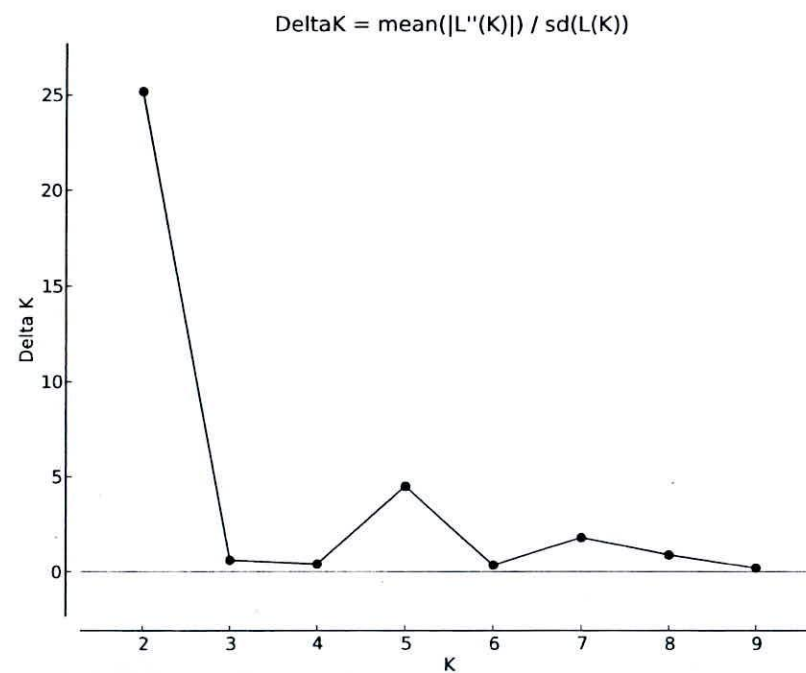


FIGURE 2: Plot of changes in ΔK value with the number of subpopulations.

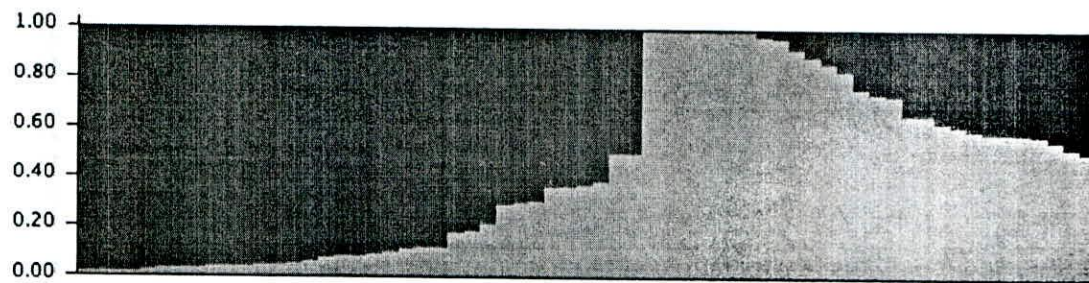


FIGURE 3a: Population structure among individuals with $K = 2$.

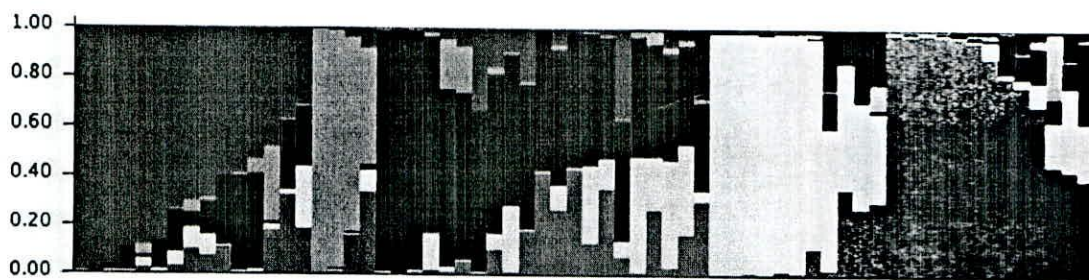


FIGURE 3b Population structure among individuals with $K = 5$.

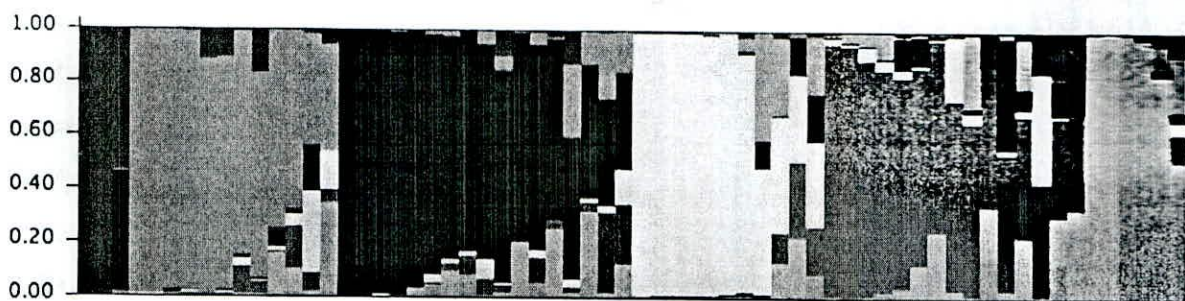


FIGURE 3c Population structure among individuals with $K = 7$.

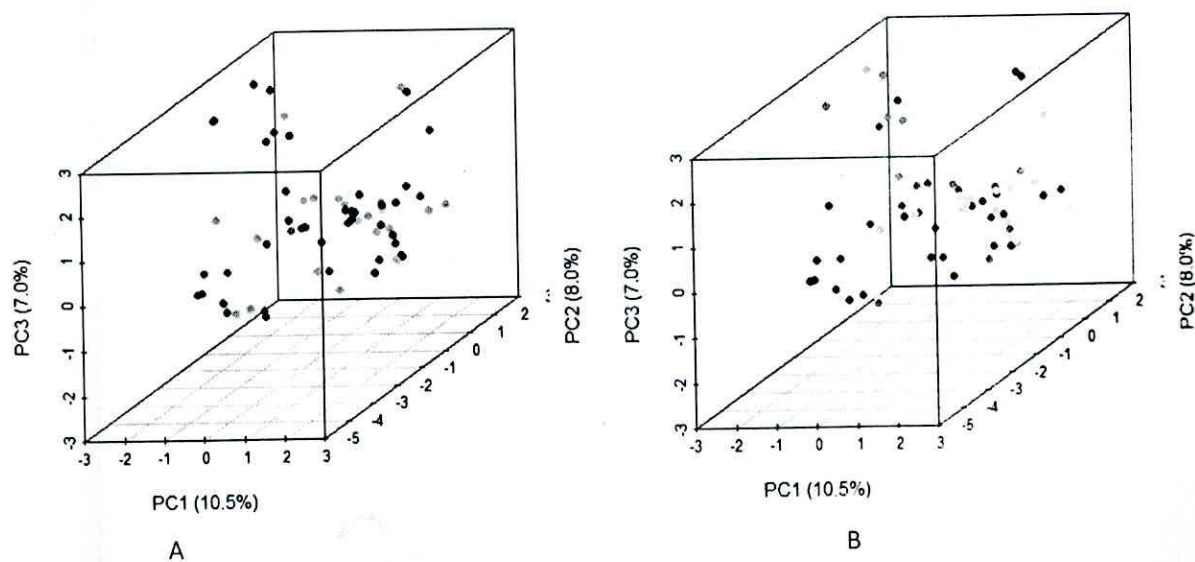


FIGURE 4 Principal component analysis (PCA) of 63 provitamin A maize inbred lines based on population structure when $K = 5$ (A) and $K=7$ (B).