

Institute of Plant Breeding, Seed Science, and Population Genetics

University of Hohenheim

Department of Population Genetics

Prof. Dr. H.H. Geiger

**Marker-assisted heterotic grouping  
of Sudanese sorghum landraces**

Dissertation submitted in fulfillment of the requirements for the degree

"Doktor der Agrarwissenschaften"

(Dr.sc.agr. / Ph.D. in Agricultural Sciences)

to the Faculty of Agricultural Sciences

Presented by

**Tahani Y. Elagib**

Sudan

Stuttgart-Hohenheim

November 2007

## SUMMARY

A heterotic group denotes a collection of related or unrelated germplasm that, when crossed with germplasm from a genetically distant group, tends to exhibit a higher degree of heterosis (on average) than when crossed with a germplasm of its own group. The objectives of the present study were to characterize the pattern of genetic diversity in a representative sample of Sudanese sorghum landraces, to determine genetically distinct gene pools, and to evaluate their potential for future hybrid breeding programs. For this purpose, eighty sorghum accessions from Sudan, ICRISAT, and the Plant Genetic Resources Conservation Unit (USDA/ARS) have been characterised using 15 genomic and 16 expressed sequence tag (EST)-derived SSR markers. The 31 SSR markers revealed a total of 242 alleles in the analysed collection with a mean of 7.8 alleles per locus. The polymorphic information content (PIC) of the markers ranged among accessions from zero to 0.902 with an average of 0.51. The average observed heterozygosity ( $H_o$ ) was 0.033. Generally, EST-derived SSRs showed a much lower information content and had considerably less alleles per locus than the SSR markers isolated from genomic libraries. This can be attributed to the different origin of the markers, the former mainly derived from more conserved coding regions of the genome whereas the latter from gene flanking non-coding regions.

Genetic diversity indices of different sorghum accession groups obtained with genomic SSR markers were much higher than those of EST-derived. Sudanese landraces showed the highest allelic richness and number of unique alleles of all accession groups analysed irrespective to the marker type. This indicated the great amount of genetic diversity that exist in these materials since the Sudan is considered as a major center of diversity for sorghum. Sudanese landraces and the world collection gave similar values of Nei's average gene diversities. This reflected the competitiveness of Sudanese landraces diversity to that found in

the world-wide sorghums. Inbred lines had the lowest percentages of polymorphic loci with both types of markers. The wild sorghums (*S. bicolor* ssp. *arundinaceum*) were more heterozygous than all other accessions, regardless of the marker type and showed 15 unique alleles not present in the cultivated sorghums.

A cluster analysis employing the UPGMA method revealed seven major groups, clustering by geographical origin or racial classification, or both. Almost all Sudanese sorghums clustered in the first four groups while most accessions from the world collection formed a separate cluster. Durra and guinea sorghums agreed with racial classification clustering.

For better comparability, a principal coordinate analysis of a representative sample of 33 accessions was performed. The analysis resulted in five major clusters. However, the principal coordinate analysis seems to support the UPGMA results and both showed that the Sudanese landraces are very distinct from other sorghums in the world collection.

To determine the potential of the Sudanese landraces for hybrid breeding, testcrosses and  $F_2$  diallel crosses were produced and evaluated in the Sudan. Based on the molecular genetic cluster analysis, representatives of Sudanese landraces and inbred lines from the world collection were selected for the production of testcrosses with two CMS tester lines. The two tester lines had the same  $A_1$  cytoplasm but were originating from different gene pools. Thirty one testcrosses with line Tx623A and 24 with line IS296A were produced and evaluated in two locations representing the major sorghum production extremes, irrigated and rain-fed. Assessment of male fertility restoration ability in the testcrosses was carried out for additional interest. Representative samples of landraces based on genetic distance obtained from the molecular marker study were manually crossed in a diallel manner. Parents were grouped according to maturity and were evaluated together with their  $F_2$  crosses under irrigation only.

Testcrosses with line Tx623A performed far better for grain yield under irrigation whereas those with line IS296A performed well under rain-fed conditions. This may be attributed to a better adaptation of line IS296A to arid conditions. On average, the testcrosses performed much better than the landraces in both experiments at both test sites. Therefore, single cross hybrids between landraces as restorers and improved adapted CMS lines could be a good strategy to develop superior sorghum varieties for Sudanese conditions. The Sudanese landraces proved to be effective restorers of fertility in hybrids with  $A_1$  cytoplasm. Significant phenotypic correlations between testcross and *per se* performance were observed for phenological traits but correlations for grain yield were only moderate.

Phenotypic correlation among traits showed that grain yield was positively correlated with head number in general. Plant height and straw yield were correlated throughout. The correlation between the two testcross series for grain yield was significant under rain-fed conditions only. However, site-specific differences in fertility restoration did not influence the correlation.

Line x tester analysis was performed using lattice-adjusted entry means. Variation due to lines was significant for most traits at both test locations whereas variation due to testers only occasionally reached statistical significance. Furthermore, important line x tester interactions occurred in almost all cases.

Mean grain yield of the early maturity diallel crosses was almost four times higher than that of the medium ones. In both maturity groups, relative  $F_2$  heterosis was positive for grain yield in three crosses and negative for all others. These many negative estimates obtained in the two  $F_2$  diallels may be explained by a breakage of coadapted gene arrangements ( $F_2$  breakdown) which could be expected in locally adapted landraces. Therefore, the use of  $F_1$  crosses for estimating heterosis in sorghum is recommended since estimation of heterosis in the  $F_2$  generation, neglecting epistasis, is expected to be half of that in the  $F_1$ .

Generally, positive heterosis for grain yield was obtained from crosses which had parents of similar ecological adaptation or racial classification. In the determination of heterotic groups, much weight should be given to these two criteria. In the medium maturity group, by comparing one pair of crosses, higher relative  $F_2$  heterosis and performance were observed in the inter- than in the intra-pool cross. Grain yield and relative  $F_2$  heterosis of dialleles were not significantly correlated with the marker-based genetic distance between parents. The lack of correlation may be due to sampling since only a small number of crosses were analysed.

The analysis of variance of the early maturity group did not indicate a significant difference among GCA effects for all traits analysed. SCA was the main source of variation. It was observed that with the predominance of GCA, superior crosses could better be predicted from their parents' GCA effects than from marker-based genetic distance measures. Therefore, the genetic distance approach was not a reliable method for predicting best cross combinations in our experiments.

Positive multiplicative interaction effects on grain yield were observed in one cross. Thus, selection of parents with contrasting grain yield components can be suggested to exploit this phenomenon as an important component of heterosis.

A principal component analysis of the 33 accessions based on five quantitative traits was performed. The results suggested three major clusters. The Sudanese sorghums grouped in two separate clusters. The third cluster mainly included inbred lines from the world collection. However, one-third of the accessions did not conform to any of these clusters. A better clustering of these accessions based on SSR molecular data than based on agronomic traits was observed. The correlation coefficient between the clusters of the 33 accessions generated from the molecular data and those from the agronomical traits turned out to be very low ( $r = 0.171$ ) but still was significant at  $P = 0.05$ . Therefore, genetic diversity analysis at DNA level is recommended as a guide for grouping accessions.

Based on the evaluation of heterotic response of the diallel crosses, the landraces Culum Brown and Geshesh in the early maturity group are considered as corner stones for development heterotic groups due to their good combining abilities and high grain yield of their cross. Similarly the landraces Wad Ahmed and Um Eltieman were identified in the medium maturity group. However, no distinct heterotic groups could be determined in the Sudanese landraces. Nevertheless, great genetic distances between clusters of Sudanese landraces represent a promising basis for heterotic grouping in view of hybrid breeding. Therefore, intensive field testing for a large number of diallel crosses among and within groups over a wide range of environmental conditions is proposed to identify heterotic groups and to select outstanding hybrid parents.

## ZUSAMMENFASSUNG

Eine heterotische Gruppe bezeichnet eine Zusammenstellung von verwandtem oder nicht verwandtem genetischen Material, welches durch Kreuzung mit dem genetischen Material einer genetisch divergenten Gruppe zu einem höheren Grad an Heterosis führt als durch Kreuzung mit Material der gleichen Gruppe. Das Ziel der vorliegenden Studie war, die genetische Diversität einer repräsentativen Stichprobe sudanesischer Sorghum (*Sorghum bicolor* (L.) Moench) Landrassen zu charakterisieren, genetisch verschiedene Genpools zu bestimmen und deren Potential für zukünftige Hybrid-Züchtungsprogramme zu evaluieren. Zu diesem Zweck wurden achzig Sorghum Akzessionen aus dem Sudan, von ICRISAT und der Plant Genetic Resources Conservation Unit (USDA/ARS) mit 15 genomischen und 16 auf Expressed Sequence Tag (EST) basierenden SSR-Markern charakterisiert. Die 31 SSR-Marker wiesen in der analysierten Gruppe 242 Allele auf mit einem Mittelwert von 7,8 Allelen pro Locus auf. Der Polymorphism Information Content (PIC) der Marker rangierte zwischen den Akzessionen zwischen 0 und 0,902 mit einem Mittelwert von 0,51. Die mittlere beobachtete Heterozygotie ( $H_0$ ) lag bei 0,033. Generell zeigten EST-basierende SSR Marker einen wesentlich niedrigeren Informationsgehalt und hatten deutlich weniger Allele pro Locus als die SSR Marker, die von genomischen Bibliotheken isoliert worden waren. Dies kann auf den unterschiedlichen Ursprung der Marker zurückgeführt werden, da erstere hauptsächlich von mehreren konservierten kodierenden Regionen des Genoms stammen, letztere jedoch von flankierenden nicht-kodierenden Regionen. Die Indizes der genetischen Diversität der unterschiedlichen Sorghumgruppen, die mit genomischen SSR Markern erhalten wurden, waren wesentlich höher als die der EST-abgeleiteten Marker. Die sudanesischen Landrassen zeigten von allen analysierten Sorghumgruppen die höchste allelische Diversität und Anzahl unterschiedlicher Allele unabhängig vom Typ der Marker. Dies illustriert die

hohe genetische Diversität, die in diesem Material existiert, da der Sudan als Vielfältigkeitszentrum für Sorghum gilt. Für die sudanesischen Landrassen und die Kollektion global verbreiteter Linien ergaben sich ähnliche Werte für die mittlere Gendiversität nach Nei. Dies verdeutlicht die Höhe der Diversität der sudanesischen Landrassen im Vergleich zu der Konkurrenzfähigkeit der Diversität im weltweiten Sorghummaterial. Der geringste prozentuale Anteil an polymorphen Loci wiesen bei beiden Markertypen die Inzuchtlinien auf. Die Wildsorghumakzessionen (*S. bicolor* ssp. *arundinaceum*) hatten unabhängig vom Markertyp einen höheren Heterozygotiegrad als andere Akzessionen und wiesen 15 private Allele auf, die in den kultivierten Sorten nicht nachweisbar waren.

Nach einer Clusteranalyse nach der UPGMA Methode konnten die Sorten in sieben Hauptgruppen eingeteilt werden, wobei die Cluster nach geographischem Ursprung, nach taxonomischer Klassifikation oder nach beiden Kriterien eingruppiert werden konnten. Nahezu sämtliche sudanesischen Sorghum-Sorten wurden in den ersten vier Clustern gefunden, während die Kollektion der weltweit verbreiteten Linien ein separates Cluster bildeten. Durra und Guinea Sorghum-Sorten konnten aufgrund ihrer Taxonomie eingruppiert werden.

Um das Potential der sudanesischen Landrassen für die Hybridzüchtung zu beurteilen wurden Testkreuzungen und  $F_2$ -Diallel-Kreuzungen produziert und im Sudan evaluiert. Aufgrund einer molekulargenetischen Clusteranalyse wurden repräsentative Genotypen von sudanesischen Landrassen und Inzuchtlinien der Kollektion global verbreiteter Linien für die Produktion von Testkreuzungen mit zwei CMS Testlinien selektiert. Die zwei Testlinien hatten das gleiche  $A_1$ -Cytoplasma, stammten aber aus verschiedenen Genpools. 31 Testkreuzungen mit der Linie Tx623A und 24 Testkreuzungen mit der Linie IS296A wurden produziert und an zwei verschiedenen Standorten evaluiert, wobei die Standorte Extremstandorte des Sorghumanbaus (bewässert und unbewässert) repräsentieren. Zusätzlich wurde eine Beurteilung der Fähigkeit



zur Restoration der männlichen Fertilität in den Testkreuzungen durchgeführt. Repräsentative Landrassen, deren genetische Distanz durch die Markerstudie bestimmt worden war, wurden ausgewählt und manuell in einem Diallel gekreuzt. Die Eltern wurden nach Reife gruppiert und zusammen mit ihren F<sub>2</sub>-Kreuzungen unter Bewässerung evaluiert. Testkreuzungen der Linie Tx623A zeigten einen weitaus höheren Ertrag bei Bewässerung; im Gegensatz dazu zeigten die Testkreuzungen mit der Linie IS296A eine höhere Leistung am unbewässerten Standort. Dies kann mit einer besseren Adaption der Linie IS296A an arides Klima erklärt werden. In beiden Experimenten und an beiden Standorten erzielten die Testkreuzungen im Schnitt eine höhere Leistung als die Landrassen. Demzufolge könnten Einfachhybriden zwischen Landrassen als Restorer und verbesserte adaptierte CMS-Linien eine gute Strategie sein, um höher entwickelte Sorghumsorten für die Anbaubedingungen im Sudan zu entwickeln. Die sudanesischen Landrassen bewährten sich als effiziente Restorer der Fertilität in Hybriden mit A<sub>1</sub> Cytoplasma. Signifikante phänotypische Korrelationen zwischen Testkreuzungen und deren Leistung an sich wurden für phänotypische Merkmale beobachtet, Korrelationen zum Korntrag waren jedoch lediglich moderat.

Phänotypische Korrelationen zwischen Ertrag und Rispenansätzen waren generell positiv. Die Wuchshöhe und Biomasse waren stets miteinander korreliert. Die Korrelation zwischen den beiden Testkreuzungsserien für Ertrag war nur unter unberechneten Bedingungen signifikant. Die Korrelation wurde jedoch nicht von standortspezifischen Unterschieden der Restoration der Fertilität beeinflusst.

Die Linie x Tester Analyse wurde anhand von gitteradjustierten Mittelwerten durchgeführt. Variationen aufgrund von Linien war für die meisten Merkmale an beiden Standorten signifikant, Variationen aufgrund der Tester erreichten nur gelegentlich statistische Signifikanz. Desweiteren traten in fast allen Fällen große Linie x Tester -Interaktionen auf.

Der mittlere Kornertrag der frühreifen Diallel Kreuzungen waren fast viermal höher als der der mittleren Reifegruppe. In beiden Reifegruppen war die relative  $F_2$  –Heterosis in drei Kreuzungen positiv für den Kornertrag, für alle anderen negativ. Die zahlreichen negativen Schätzungen der Diallele können durch Spaltung von ko-adaptierten Genanordnungen erklärt werden ( $F_2$ -breakdown), die in lokal adaptierten Landrassen erwartet werden können. Aus diesem Grunde wird der Gebrauch von  $F_1$ -Kreuzungen vorgeschlagen, um die Heterosis in Sorghum zu schätzen, da die Schätzung der Heterosis in der  $F_2$ -Generation (bei Vernachlässigung der Epistasie) als etwa der Hälfte von der der  $F_1$ -Generation erwartet wird. Generell wurde positive Heterosis für Kornertrag von Kreuzungen mit Eltern ähnlicher ökologischer Adaption, taxonomischer Klassifizierung oder beidem erzielt. Bei der Bestimmung der heterotischen Gruppen sollten diese beiden Kriterien besonders berücksichtigt werden. In der mittleren Reifegruppe wurden bei Betrachtung einzelner Kreuzungspaare in der Interpool-Kreuzung eine höhere relative  $F_2$  –Heterosis und Leistung beobachtet als in der Intrapool-Kreuzung. Kornertrag und relative  $F_2$ -Heterosis der Diallele korrelierten nicht signifikant mit der markerbasierenden genetischen Distanz zwischen den Eltern. Die fehlende Korrelation ist vermutlich auf die Probennahme zurück zu führen, da lediglich eine kleine Anzahl von Kreuzungen analysiert werden konnte.

Die Varianzanalyse der frühen Reifegruppe wies keinen signifikanten Unterschied zwischen den GCA-Effekten für alle untersuchten Merkmale auf. Die hauptsächliche Variationsursache war die SCA. Es konnte beobachtet werden, dass überlegene Kreuzungen anhand der GCA-Effekte der Eltern besser vorhergesagt werden konnten als anhand der marker-gestützten Messungen der genetischen Distanz. Der methodische Ansatz über die genetische Distanz war somit in dieser Studie keine verlässliche Methode, um die besten Kreuzungskombinationen vorherzusagen.

Effekte von positiven multiplikativen Interaktionen auf den Kornertrag wurden in einer Kreuzung beobachtet. Somit kann die Selektion von Eltern mit unterschiedlichen Ertragskomponenten für die Auswertung dieses Phänomens als wichtige Komponente der Heterosis vorgeschlagen werden.

Zusätzlich wurde eine Hauptkomponentenanalyse der Komponenten der 33 Akzessionen basierend auf 33 Merkmalen durchgeführt. Die Analyse ergab drei Hauptcluster. Die sudanesischen Sorghumsorten konnten in zwei separate Cluster eingruppiert werden. Das dritte Cluster beinhaltete hauptsächlich Inzuchtlinien der global verbreiteten Linien. Jedoch konnte ein Drittel der Akzessionen keinem der Cluster zugeordnet werden. Es konnte beobachtet werden, dass durch SSR Marker Daten eine bessere Gruppierung der Akzessionen erstellt werden konnte als durch agronomische Merkmale. Der Korrelationskoeffizient zwischen und agronomischen Daten war gering ( $r = 0,171$ ), dabei jedoch signifikant für  $P = 0,05$ . Deswegen kann die Analyse der genetischen Diversität auf DNA-Ebene als Basis zur Gruppierung von Akzessionen empfohlen werden.

aufgrund der guten Kombinationsfähigkeit und des hohen Kornertrages ihrer Kreuzungen können die Landrassen Culum, Brown und Geshesh der frühen Reifegruppe als Ausgangspunkt zur Entwicklung heterotischer Gruppen angesehen werden. In ähnlicher Weise sind die Landrassen Wad Ahmed und Um Eltieman in der mittleren Reifegruppe einzuschätzen. Es konnten jedoch keine eindeutigen heterotischen Gruppen innerhalb der sudanesischen Landrassen bestimmt werden. Dennoch repräsentieren hohe genetische Distanzen zwischen den Clustern der sudanesischen Landrassen eine vielversprechende Basis für heterotische Gruppen in Bezug auf die Hybridzüchtung. Hieraus resultiert die Empfehlung in ausführlichen Feldversuche für eine Vielzahl von Diallel-Kreuzungen zwischen und innerhalb dieser Gruppen unter mehreren Umweltbedingungen zur Identifizierung

heterotischer Gruppen und zur Selektion hervorragender Elterlinien für die Hybridzüchtung durchzuführen.