



International Journal of Botany

ISSN: 1811-9700

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Genetic Diversity Assessment of Sorghum Landraces in Zimbabwe Using Microsatellites and Indigenous Local Names

¹E. Chakauya, ¹P. Tongoona, ²E.A. Matibiri and ³M. Grum

ACCI, Plant Pathology, University of Natal, P/ Bag X01, Scottville, KwaZulu Natal, South Africa

²Kutsaga Research Station, Airport Ring Road, Harare, Box 1909, Harare, Zimbabwe

³Sub-Saharan Africa IPGRI, c/o ICRAF, P.O. Box 30677, Nairobi, Kenya

Abstract: Landraces of sorghum are grown throughout Zimbabwe and frequently given names describing important traits they possess or their origin. Little is known about the genetic diversity of this germplasm, the consistency of the relationship between names and genetic identity and thereby the usefulness of names in developing strategies for germplasm conservation and use. This study analysed the genetic diversity of 52 uncharacterised genotypes from two districts of Zimbabwe, Nyanga North and Tsholotsho using two methods; microsatellites and indigenous local names. Seven microsatellite primers were used, giving 19 polymorphic bands. Genetic similarities (simple matching coefficients) were calculated and dendrograms were used to compare relationships based on indigenous local names and the origin of collections. Microsatellite markers showed that even when local names were similar, the genotypes those names represented were rarely similar genetically. Although collections from six of the seven clusters were identified in both districts, those in Nyanga North had a Simpson index of 0.728, compared with 0.533 for those in Tsholotsho. Based on the markers used, seven of the 15 collections from Tsholotsho were judged to be duplicates of material also found in Nyanga North. Diversity was found to be fairly highly concentrated in the hands of a few individual farmers. The results of this study indicate that farmer-given names are neither the best guide to the presence of diversity nor a good unit of sampling.

Key words: Sorghum, genetic diversity, indigenous local names, landraces, microsatellites, Zimbabwe

INTRODUCTION

Sorghum [*Sorghum bicolor* (L.) Moench], locally called mapfunde in the Shona and amabele in Ndebele, is ranked the third most important cereal crop in Zimbabwe after maize and wheat. Sorghum is a principal food crop for many Zimbabweans, particularly those situated in semi-arid areas. Sorghum already has a wide range of uses in Zimbabwean agriculture and industry. Human food uses include porridge made from types with cracking-grain and the brewing of beer from types with malting grain. Varieties with juicy stems (sorgos) are used for making syrups. The grain and other plant parts can be used as cattle feed and fodder. Sorghum can also become a source of industrial alcohol, adhesives, vegetable oil and waxes, construction materials and brooms. In short, its socio-economic importance is already vast and sorghum plays a pivotal role in Zimbabwe's hunger alleviation and economic development programmes.

One drawback to achieving this pivotal role is the fact that sorghum yields in Zimbabwe (500 kg ha^{-1}) are far below the world average, which is 1310 kg ha^{-1} .

About 82% of the nation's sorghum is produced on communal farms in semi-arid areas that are characterized by erratic rainfall, excessively high temperatures and infertile soils (Manyowa, 1993). Farmers in these areas depend on unimproved but well-adapted landraces because they know how to grow and store them and believe, despite their low yields, that these landraces are high in both nutritional value and tolerance to the stresses just mentioned. It is estimated that each sorghum-growing area has approximately ten landraces that are grown as mixtures, with two to four landraces per individual farmer. Unfortunately, relatively little attention has been accorded to study and conserve this diversity.

There is growing realization of the importance of information on genetic diversity. For example breeders, curators, and conservationists need this information for planning collections, screening sample duplicates in storage, for germplasm exchange, and for planning on-farm and off-farm storage and preservation. To improve diversity, scientists need to combine their own knowledge with that of the people who actually farm the land,

because the actions of those farmers play a major role in the distribution of truly diverse genotypes. We believe that in order to improve the genetic diversity of sorghum of these communal farms we first have to gain a better understanding of the socioeconomics of the various farmer groups.

Currently, there are about 1378 Zimbabwean sorghum accessions at the International Research Institute for Semi-arid tropics (ICRISAT), of which 830 have been characterised for agromorphological traits (IBPGR/ICRISAT, 1984). The Genebank of Zimbabwe holds additional partially uncharacterised sorghum germplasm collected since 1982 (Mafa, 1999). Rao and Mushonga (1987) presented a detailed catalogue of 285 accessions from the 1982 collection project but still a greater portion of the germplasm in storage locally and internationally needs 'formal' characterisation. In this study we used microsatellite analysis and indigenous local farmer-given names to assess genetic diversity in 52 landraces collected in 1998.

Literature on microsatellite use in genetic diversity studies is available (Dean *et al.*, 1999; Brown *et al.*, 1996). Microsatellites are particularly useful for providing a critical analysis of how genetic diversity is distributed within a population. They are highly polymorphic, informative and co-dominant markers. Microsatellites are often broadly applicable since loci are frequently conserved between related species and sometimes genera.

On the other hand, farmer classification of genetic diversity using indigenous local names has not been well studied. In Zimbabwean communal agriculture, indigenous local names play a major role in crop variety recognition and description. Generally, variety description is based on ethnic grouping, comparative morphology and geographic origin. Farmer variety names can indicate varietal traits such as stress resistance, panicle shape, response to bird attack or describe uses, the perceived origin of a variety, or be named after the person that introduced the variety to the community. No studies have been carried out in Zimbabwe to see whether there is any correspondence between farmer variety names and information from other genetic markers. In this study we investigated (a) the distribution of microsatellite marker diversity among landraces collected from two communal areas of Zimbabwe and (b) the consistency of farmer variety names from a molecular perspective.

MATERIALS AND METHODS

Germplasm: The study was carried out using 52 uncharacterised landraces of sorghum that had been collected by the National Genebank of Zimbabwe from two districts of that nation, Tsholotsho and Nyanga

Table 1: Collection identities, indigenous local names and origin of Sorghum landraces analysed for microsatellite polymorphism

Accession No.	Farmer-given name	Village	District
1343	Nhongoro	Samakande	Nyanga North
1345	pwa	Samakande	Nyanga North
1346	Ipwa	Samakande	Nyanga North
1355	Musoswe	Chibvende	Nyanga North
1378	Nzembe	Mangezi	Nyanga North
1384	Nzembe	Mangezi	Nyanga North
1385	Musoswe	Mangezi	Nyanga North
1392	Nhongoro	Mangezi	Nyanga North
1395	Mudhabhure	Mangezi	Nyanga North
1399	Nzembe	Mangezi	Nyanga North
1400	Nhongoro	Mangezi	Nyanga North
1401	Musoswe	Mangezi	Nyanga North
1402	Sorghum	Mangezi	Nyanga North
1409	Nhongoro	Mangezi	Nyanga North
1412	Sorghum	Mangezi	Nyanga North
1425	Musoswe	Kamunhukamwe	Nyanga North
1428	Musoswe	Kamunhukamwe	Nyanga North
1430	Shodhani	Kamunhukamwe	Nyanga North
1441	Musoswe	Kamunhukamwe	Nyanga North
1448	Musoswe	Kamunhukamwe	Nyanga North
1450	Shodhani	Kamunhukamwe	Nyanga North
1455	Ipwa	Kamunhukamwe	Nyanga North
1459	Sorghum	Renzva	Nyanga North
1460	Musoswe	Renzva	Nyanga North
1461	Sorghum	Renzva	Nyanga North
1472	Mutanda	Renzva	Nyanga North
1473	Malawi	Renzva	Nyanga North
1474	Malawi	Renzva	Nyanga North
1475	Shodhani	Renzva	Nyanga North
1477	Chipemu	Renzva	Nyanga North
1478	Sorghum	Renzva	Nyanga North
1480	Sorghum	Renzva	Nyanga North
1482	Sorghum	Renzva	Nyanga North
1483	Nyamwayawaya	Renzva	Nyanga North
1485	Nzembe	Renzva	Nyanga North
1487	Sorghum	Renzva	Nyanga North
1496	Ipwa	Renzva	Nyanga North
1499	Khaki	Siyabandela	Tsholotsho
1521	Imfe/yakayaka	Siyabandela	Tsholotsho
1523	Isigobane	Siyabandela	Tsholotsho
1527	Cimezile	Siyabandela	Tsholotsho
1535	Tsweta Red	Sizanani	Tsholotsho
1537	Tsweta khaki	Sizanani	Tsholotsho
1540	Red Swazi	Sizanani	Tsholotsho
1550	Cimezile	Sizanani	Tsholotsho
1555	Tsweta	Sizanani	Tsholotsho
1556	Hlusi	Sizanani	Tsholotsho
1557	Fethulitha	Sizanani	Tsholotsho
1572	Imfe	Siyazama	Tsholotsho
1592	Isifumbata	Siyazama	Tsholotsho
1593	Tsweta	Siyazama	Tsholotsho
1627	Tsweta red	Phakamani	Tsholotsho

North in 1998 (Mafa, 1999) (Table 1). The two districts are in almost opposite ends of Zimbabwe and isolated from each other by climates that are not very suitable for sorghum and their climates are significantly distinct.

Nyanga North was originally selected for this study because of the rich genetic variation observed during collection missions, while Tsholotsho was chosen because it represented the more marginal areas. In addition site selection was based on semi-aridity

Table 2: Comparison of monthly rainfall and mean temperature in Nyanga North and Tsholotsho showing the substantial differences that exist between the two sites.

Months	Rainfall (mm)		Temperature (°C)	
	Tsholotsho	Nyanga North	Tsholotsho	Nyanga North
Jan.	370	146	24	22
Feb.	320	131	23	21
Mar.	221	189	23	21
Apr.	100	33	22	21
May.	42	3	19	19
Jun.	32	0	16	18
Jul.	11	0	18	18
Aug.	15	0	19	18.5
Sept.	10	0	23.5	21.6
Oct.	55	4	26	23.9
Nov.	128	54	25	25
Dec.	346	149	25.8	24.5

Table 3: Characteristics of *Sorghum bicolor* microsatellite primers used in the study

Marker	Linkage group	Primer sequence	Annealing temperature
Xtxp8	B	F: ATA TGG AAG GAA GAA GCC GG R: AAC ACA ACA TGC ACG CAG G	60
Xtxp9	Unknown	F: AAT AGC ACC GCC GCG CG R: CAT TGT GGA GTC CCT GAT AC	55
Xtxp10	F	F: ATA CTA TCA AGA GGG GAG C R: AGT ACT AGC CAC ACG TCA C	50
Xtxp16	Unknown	F: TAG GGA AGA GCA AGT GCA GAC R: AAG AAA GGG CCC AGA GTT TC	60
Xtxp17	I	F: CGG ACC AAC GAC GAT TAT C R: ACT CGT CTC ACT GCA ATA CTG	55
Xtxp19	B	F: CTT TCA ATC GGT TCC AGA C R: CTT CCA CCT CCG TAC TC	55
Xtxp21	D	F: GAG CTG CCA TAG ATT TGG TCG R: ACC TCG TCC CAC CTT TGT TG	60

(Table 2), the abundance of phenotypic variation during a pilot study and presence of non-governmental organizations (NGO) already working with the communities in related projects (Mafa, 1999). Collections were made from two districts, three villages per district and several farmers per village. In this case a village was defined as a group of about 50 households under one social administrative authority that has a common obligation to conserve and manage genetic resources.

DNA extraction: DNA extracted by a modified cetyltrimethylammonium bromide (CTAB) described by Saghai-Marooof *et al.* (1984). Fresh leaf material was taken from 3 to 4 week old plants grown in a greenhouse. Five to eight plants of each accession were sampled and the material bulked.

Microsatellite analysis: Primer sets (Table 3) were provided by Texas A and M University, Texas. PCR reactions were performed in 25 µL volumes containing 30 ng template DNA, 1x PCR Buffer (50 mM KCl, 10 mM Tris-HCl, (pH 8.3), 1.5 mM MgCl₂), 0.25 mM deoxynucleotide triphosphates, 4.13 to 5.68 pmoles (30 ng) of each primer (Table 3) and 1.2 units Taq

polymerase. Temperature cycling was performed on a GeneAmp PCR system 9700 (Perkin-Elmer) programmed to run 4 min at 94°C, followed by 35 cycles of 94°C for 1 min, 1 min at annealing temperature (Table 2) and 72°C for 1 min. The final cycle was identical to the above but had a final elongation of 10 min at 72°C. PCR products were separated by electrophoresis using 1xTBE on 8% PAGE, stained with ethidium bromide and band sizes estimated with a 100 base pair ladder (Roche Diagnostics).

Data collection and statistical analysis: Gels were scored manually with each polymorphic band being treated as a unit character and scoring was for the presence (1) or absence (0) of a band. Two criteria were used for scoring bands: the band being scored had to stain strongly and there had to be an unambiguous difference between the allelic states of the band being scored (i.e., presence or absence of a band). The statistical analyses were carried out in the R statistics environment version 1.7.1. The degree of similarity between collections was calculated using the implementation of the simple matching coefficient in the R function daisy of the package “cluster” version 1.7.4 (Struyf *et al.*, 1996). Hierarchical clustering was done with the R function “hclust” in the package “mva”, version 1.7.1, using the unweighted pair-group method of arithmetic averages (UPGMA). The Kelley-Gardner-Sutcliffe penalty function for a hierarchical cluster tree was calculated using the function kgs in the package maptree (Kelley *et al.*, 1996) to suggest a number of clusters in the dataset. The cophenetic distance for the hierarchical clustering was calculated using the function cophenetic in the package mva, while the correlation between the two matrices used the function mantel in the package vegan. The Simpson diversity index was calculated using the function diversity also in the package vegan.

RESULTS

Out of the seven-microsatellite primers tested, 19 polymorphic bands were scored. The number of bands ranged from one to five; their corresponding sizes, from 140 to 280 bp.

The hierarchical clustering of the 52 accessions (Fig. 1) produced an agglomerative coefficient of 0.76 indicating a moderately good clustering structure. A correlation of $r = 0.71$ between the observed dissimilarities and the heights used in the dendrogram (the cophenetic distance) indicates that the plot is a reasonable representation of the dissimilarities. The Kelley-Gardner-Sutcliffe penalty function minimized at seven clusters and these clusters are used as the basis for much of the analysis below.

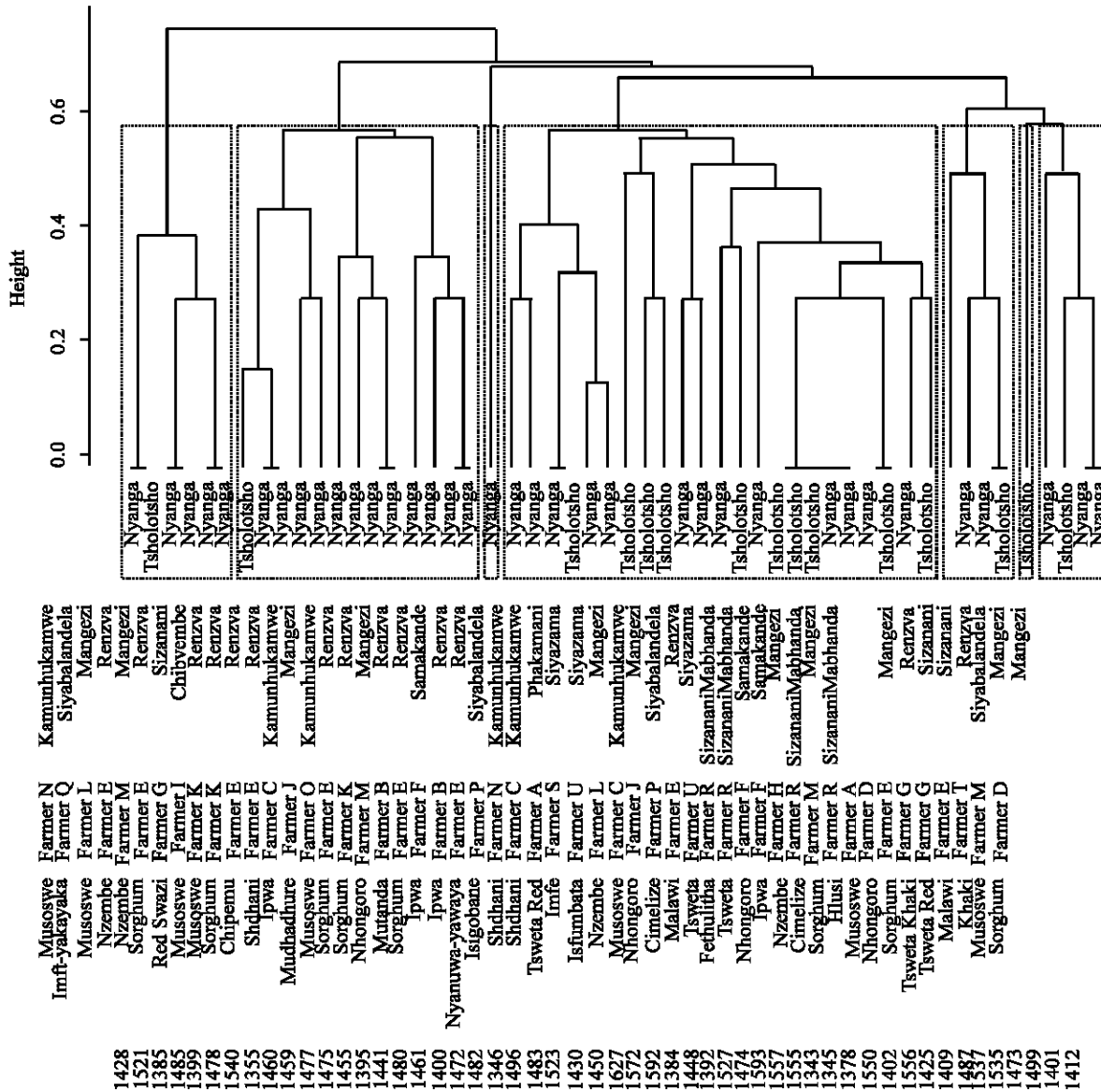


Fig. 1: Hierarchical clustering of 52 collections from Zimbabwe based on a simple matching coefficient and unweighted pair-group method of arithmetic averages (UPGMA) of seven microsatellite primers producing 19 polymorphic bands. The labels are, from left to right, accession number, farmer-given name, farmer, village and district of each accession.

Although both districts were represented by accessions from six of the seven clusters, Nyanga North showed considerably greater evenness of distribution among clusters, with a Simpson index of 0.728, compared with 0.533 for Tsholotsho. This index may be interpreted simply as the probability that a second accession sampled will belong to a different cluster from the first. Two thirds of all the accessions collected in Tsholotsho were from a single cluster, whereas only slightly over one third of the accessions from Nyanga North were from one cluster. Seven of the fifteen accessions from Tsholotsho were identical to accessions from Nyanga North for all the

markers used. Tsholotsho had one distinct accession (1535), which was unique in its cluster.

Within the district of Nyanga North, the village of Renzva had material from all six clusters represented in the district and was followed closely by the village of Mangezi with five of the clusters and Kamunhukamwe with three clusters represented. In the district of Tsholotsho, the villages of Siyabalandela and Sizanani both had material from three clusters.

One household in Renzva, Farmer E, had accessions from all six clusters collected from that village and more generally in Nyanga North, including accession 1482,

which was the only representative of cluster 3. In Tsholotsho, Farmer G had an accession from each of the three clusters represented in the village of Sizanani, including accession 1535, which was the sole representative of cluster 6. Farmer M from the village of Mangezi, with accessions from four different clusters, was the only other person with accessions from more than four clusters.

A total of 22 local names were recorded for 28 of the 36 genotypes were collected. Thirteen of these names were unique, while the rest were used for more than one accession. The local names used for more than one collection often fell in different clusters indicating marked genetic differences. Musoswe was the most common name, used for seven accessions from four different clusters, followed by Nhongoro, Nzembe and Ipwa, all used to name four collections, from three, two and two clusters, respectively.

There were two cases where the same farmer used the same name for two separate accessions. Farmer F called two accessions Ipwa (1345 and 1346), which means sweet sorghum and denotes varieties with stems for chewing. Although these accessions belonged to the same cluster, they were as far apart as they could be within the cluster. Surprisingly, Farmer F also had an accession (1343) that he called Nhongoro, which proved genetically similar to accession 1345 for the markers used in this study. Farmer R had two accessions (1555 and 1557) that were also genetically similar to accessions 1343, 1345 and 1593, yet she gave these genetically similar accessions two different names.

Farmer E had two accessions called Malawi that belonged to two different clusters. The description probably only indicated that these accessions were believed to originate from Malawi. In most cases where two accessions found to be genetically identical for the studied markers were collected from different people, they had also been given different names.

The data was also examined for relationships between the genetic structure of the population and the uses, but none was found. Even accessions found to be genetically identical for the markers used in this study, had different uses when used by different farmers.

DISCUSSION

The finding that Nyanga North is more diverse than Tsholotsho is in line with previous estimates based on casual observation. However, sorghum plays a much more crucial role in farmer livelihoods in Tsholotsho than in Nyanga North where maize is more common.

Considering that the two sampled districts are in opposite ends of Zimbabwe, it is somewhat surprising to

find that almost half the genotypes, as identified by the markers used, found in Tsholotsho are also found in Nyanga North. Since the germplasm is not yet taxonomically classified, we could not establish how different this material was. However, some of these genotypes might be releases from the national research system, though this is not reflected in their present names. The high similarity observed at molecular level could also be explained by the role of NGOs in seed dissemination. Some NGOs (such as the Community Technology Association and Intermediate Technology Development Group) were actively involved in organizing seed fairs, community seedbanks and in teaching local farmers storage techniques and utilization of traditional varieties (Mafa, 1999).

Huvio (1999) conducted farmer surveys indicating widespread germplasm exchange within and between villages. She observed that the spread of germplasm was both by design (to improve local crop) and by accident (in grains becoming mixed) and often new varieties were added to compliment old ones or completely replace them. About 80% of seed came from neighbours and 60% of the farmers walked long distances to buy or exchange seed. Sometimes abandoned germplasm could be revisited when the need arises.

Remarkably, although 20 households were sampled, only two, the most diverse households from each district, captured accessions from all seven clusters. These two households, Farmer E from Nyanga North and Farmer G from Tsholotsho, also had the two most distinct accessions (1482 and 1535, respectively).

Although there was a tendency for different collections with the same name to be genetically related, this was not consistent across the two districts, across villages and among farmers. Even within farmers there were several examples of the same name being used for genetically unrelated collections.

Discrepancies can be an indication of the complexity of the whole process of farmer classification of germplasm and its potential hazard in genetic diversity assessments. Variety names form an essential component of variety selection in communal agriculture particularly with traditional varieties and landraces. Farmers generally retain their own seeds and only in cases of loss would they resort to obtaining seed from the market or neighbours as discussed later. This is a common phenomenon with most resource-poor farmers in marginal areas who depend on landraces for food security (Rana *et al.*, 2002). Because of several cycles of selection, new lines develop with same name or different names. Because of insufficient information on farmer nomenclature and taxonomy for the current collection, it is difficult to conclude whether there is a standard criterion for naming

varieties. For example, some names seem to be based on uses (such as sweet sorghum or Imfe), others on origin (such as Malawi) or panicle shape (*Nyamuwa yawaya*). Moreover, variety names are at the farmers' discretion and the names are passed to next generation by oral tradition making it difficult to trace the origin of the names.

On one hand, this study showed that indigenous names cannot be considered a good guide to the presence of diversity, nor do they appear to be a good unit for sampling. Having established that, our next paper will relate agromorphological characters and microsatellite measurements for sampling genetic diversity in this germplasm and establish whether the molecular data translates into quantitative traits of the sorghum.

On the other hand, the results may help guide the development of optimal conservation strategies. The first matter would be to decide whether the best place to preserve the genotypes would be on farms where they are presently growing or in a new structure specifically designed for that purpose. Smale *et al.* (2004) suggest that on-farm conservation of plant genetic resources is most rational where both the public value of diversity and its private value are high, i.e. areas where there is a lot of genetic diversity and where it makes a substantial contribution to farmer livelihoods. They also argue that marker diversity is likely to be a good indicator of the public value of diversity. On the basis of the high levels of diversity that this paper found in Nyanga North relative to Tsholotsho, we would tentatively make the case that Nyanga North would be a better site for on-farm conservation than Tsholotsho. Further studies would be required to determine the private value, or the contribution of the diversity to farmer livelihoods, in order to determine the sustainability of on-farm conservation at the two sites. Given the relative importance of maize in Nyanga North, it is not certain that the conservation of sorghum landraces would be sustainable there. Even if conservation efforts were to focus on Nyanga North, we note that Tsholotsho has diversity not found in Nyanga North.

CONCLUSIONS

In conclusion, the study reported the existence of a striking cosmopolitan sorghum genotypes leading us to suggest that a conservation efforts concentrating primarily on high diversity areas, a high diversity of areas and recognise the importance of individuals with a special interest in diversity. Off-site conservation focuses on sampling from each situation, while on-farm activities might focus on increasing the number of individuals

holding diversity so as to minimize the risk related to individual households. Clearly both are needed to not only preserve existing genetic material but also to make that available for future breeding purposes.

ACKNOWLEDGMENTS

Grateful acknowledgments are given to Charles Mutengwa, Collins Musimwa and Dr. Sithole-Niang for comments during the study. The research was funded by IFAD in a collaborative project among the Genebank of Zimbabwe, IPGRI and FAO.

REFERENCES

- Brown, S.M., M.S. Hopkins, S.E. Mitchell, M.L. Senior, T.Y. Wang, R.R. Duncan, F. Gonzalez- Candelas and S. Kresovich, 1996. Multiple methods of the identification of polymorphic Simple Repeats Sequence (SSR) in sorghum (*Sorghum bicolor* L. Moench). *Theor. Applied Genet.*, 93: 190-198.
- Dean, R.E., J.A. Dahlberg, M.S. Hopkins, S.E. Mitchell and S. Kresovich, 1999. Genetic Redundancy and diversity among orange accessions in the US National sorghum collection as assessed with simple sequence repeats (SSRs) markers. *Crop Sci.*, 39: 1215-1221.
- Huvio, T., 1999. Women in Development Service. Development of strategies for *in situ* conservation of plant genetic resources for food and agriculture in the semi-arid areas of Zimbabwe. Ann. Rev. Workshop. Genebank of Zimbabwe, IPGRI and FAO Collaborative Project. DR and SS, Harare, Zimbabwe, pp: 23-45
- IBPGR/ICRISAT, 1984. Sorghum descriptors. IBPGR, Rome, Italy, pp: 49-70.
- Kelley, L.A., S.P. Gardner and M.J. Sutcliffe, 1996. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally-related subfamilies. *Protein Eng.*, 9: 1063-1065.
- Mafa, A., 1999. Germplasm Collection Report. Development of strategies for *in situ* conservation of plant genetic resources for food and agriculture in the semi-arid areas of Zimbabwe. Annual Review Workshop. Genebank of Zimbabwe, IPGRI and FAO collaborative Project. DR and SS, Harare, Zimbabwe, pp: 13-43.
- Manyowa, N.M., 1993. Maize breeding in Zimbabwe. Coping with drought stress in marginal agroecological zones. *Afri. Centre Fert. Dev. Occasional papers*, Harare, Zimbabwe, pp: 5-16

- Rana, R.B., D. Gauchan, D.K. Rijal, A. Subedi, M.P. Uphadyaya, B.R. Sthapit and D.I. Jarvis, 2002. Nepal Factors Influencing Farmers' Decision on Management of Local Diversity on Farm and Their Policy Implications. Nepal's contribution to agrobiodiversity conservation *in situ*: A scientific basis for policy recommendation. D. Gauchan and B.R. Sthapit., Eds.). IPGRI/NARC/LI-BIRD.
- Rao, S. and J.N. Mushonga, 1987. A catalogue of passport and characterisation data of sorghum, pearl millet and finger millet germplasm from Zimbabwe. IBPGR, Rome, pp: 15-46.
- Saghai-Marooif, M.A., K.M. Solima, R.A. Jorgenson and R.W. Allard, 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. Proc. Natl. Acad. USA., 81: 8014-8018.
- Smale, M., M.R. Bellon, D. Jarvis and B. Sthapit, 2004. Economic concepts for designing policies to conserve crop genetic resources on farms. Genet. Res. Crop Evol., 51: 121-135.
- Struyf, A., M. Hubert and P.J. Rousseeuw, 1996. Clustering in an object-oriented environment. J. Stat. Software, 1: 1-30.