academicJournals

Vol. 9(8), pp. 118-129, August 2017
DOI: 10.5897/JPBCS2017.0663
Article Number: ED81D6B65447
ISSN 2006-9758

http://www.academicjournals.org/JPBCS

Journal of Plant Breeding and Crop Science

Full Length Research Paper

Morphological distinctiveness between *Solanum aethiopicum* Shum group and its progenitor

Godfrey Sseremba[1,2], Pangirayi Tongoona[1], John Savior Yaw Eleblu[1]*, Eric Yirenkyi Danquah[1], Nahamya Pamela Kabod[2,3] and Elizabeth Balyejusa Kizito[2]

[1]West Africa Centre for Crop Improvement, University of Ghana, P. M. B. 30, Accra, Ghana.
[2]Department of Agricultural and Biological Sciences, Faculty of Science and Technology, Uganda Christian University, P. O. Box 4, Mukono, Uganda.
[3]Department of Agricultural Production, College of Agricultural and Environmental Sciences, Makerere University, P. O. Box 7062, Kampala, Uganda.

Received 7 May, 2017; Accepted 28 June, 2017

Use of morphological markers offers an alternative in germplasm discrimination of research-neglected crop species. A collection of 25 accessions including five wild progenitors was evaluated in screen house to identify the morphological difference between *Solanum aethiopicum* Shum and *Solanum anguivi*. An Unweighted Pair Group Method with Arithmetic mean hierarchical clustering revealed presence of moderate structure with a cophenetic correlation coefficient of 0.73. Five distinct clusters were produced; the progenitor accessions for the *S. aethiopicum* Shum were grouped in their own cluster. The Richness, Shannon-Weaver and Simpson indices were also different among qualitative variable categories. A 'prcomp' function based Principal component analysis (PCA) in R on quantitative variables indicated that days to germination and emergence, cotyledonous leaf length, cotyledonous leaf width, shoot biomass, plant height, petiole length, days to first flowering opening, plant width, plant branching, and number of leaves per plant are the major drivers of variability in the study accessions. Further, results from canonical discriminant analysis to discern between the *S. aethiopicum* and its progenitor accession groups showed that the days to germination and emergence provide the best separation; with the former emerging earlier than the latter. The mean values for flowering time, leaves per plant, number of branches per plant and plant height were more favorable for the Shum than its wild progenitor accessions. The study revealed that morphological markers are useful in distinguishing between the *S. aethiopicum* Shum and its progenitor accessions.

Key words: African indigenous vegetable species, genetic diversity, reordered hierarchical clustering, Principal component analysis (PCA), linear discriminant analysis.

INTRODUCTION

African indigenous vegetable species (AIVS) require genetic improvement in order to address constraints that curtail the crops' productivity and contribution to household income and food security in sub-Saharan

*Corresponding author. E-mail: jeleblu@wacci.edu.gh.

Africa (Abukutsa-Onyango, 2014; Cernansky, 2015). *Solanum aethiopicum* and *S*. *anguivi* are some of the major AIVS that are research-neglected. The *S*. *aethiopicum* is morphologically diverse with four recognized groups, of which Shum is a leafy type (https://avrdc.org/african-eggplant-solanum-aethiopicum/). The *S*. *aethiopicum* evolved from *S*. *anquivi* (Ebert, 2014; Sękara et al., 2007). Although the two species are domesticated, *S*. *anguivi* is grown for its fruits only. The *S*. *anguivi* exists both in the wild and at farmers' fields; indicating its environment robustness. The availability of germplasm of known diversity is important in variety improvement efforts (Gramazio et al., 2016). Morphological markers are cheap can be used for diversity analyses (Kouassi et al., 2014; Kubie, 2013). However, the usefulness of morphological traits in accession discrimination for the Shum and *S*. *anguivi* had not yet been investigated.

The most commonly used morpho-agronomic traits for genetic diversity studies in *Solanum* spp. have been published (Adeniji et al., 2013; Gramazio et al., 2016). Multivariate statistical methods such as multidimensional scaling (MDS), linear (canonical) discriminant analysis (LDA), cluster and principal component analysis are suited for use in understanding genetic diversity (Harding and Payne, 2012). MDS is a form of non-linear dimensionality reduction for visualizing the level of similarity of individual cases of a data set (information) that is contained in a distance matrix. Cluster analysis and PCA are the two most commonly used methods of genetic diversity analysis; the former handles both quantitative and qualitative variables while the latter is powerful and sensible with quantitative data sets (Zimisuhara et al., 2015). Clustering employs either the Ward's or "average" (Unweighted Pair Group Method with Arithmetic mean; UPGMA) method algorithms (Odong et al., 2011). Although both methods rely on coefficients such as cophenetic correlation coefficient (CPCC) to judge the strength (reliability) of subgroup differentiation, UPGMA is the most commonly used (Odong et al., 2011; Zimisuhara et al., 2015).

Further, diversity indices such as Richness, Shannon-Weaver, and Simpson can reveal groupings of accessions per level of qualitative variable (Altaye, 2015). A diversity index is a quantitative measure that reflects how many different types (in this case the levels of qualitative variables) there are in the dataset (community), and simultaneously takes into account how evenly the basic entities (such as individuals) are distributed among those types (Zimisuhara et al., 2015). On the other hand, the PCA serves to identify how different variables work together to reduce dimensionality and redundancy; thus helping to reveal hidden structure (Coghlan, 2017; Zimisuhara et al., 2015). LDA is aimed at finding linear combinations of original variables that gives the best possible separation among study groups (Coghlan, 2017; Harding and Payne, 2012). This study aimed to identify: hierarchical groups existing in the study accessions, major drivers of variability in collected data set for the study accessions, and variables that account for the greatest separation between the Shum and *S*. *anguivi* progenitors.

MATERIALS AND METHODS

Study site and germplasm

The study was carried out in screen house at West Africa Centre for Crop Improvement (WACCI) research farm, University of Ghana, Greater Accra, Ghana. During the experiment (October 2016 to April 2017), temperature ranged between 21-26°C (morning), 31-43°C (afternoon) and 29-35°C (evening hours). Relative humidity ranged between 74-81% (morning), 57-75% (afternoon) and 62-69% (evening hours). The germplasm was obtained from Department of Agricultural and Biological Sciences, Uganda Christian University, Bishop Tucker Road, Mukono, Uganda. The list of study accessions is as shown (Table 1).

Design

A total of 25 accessions which comprised 20 entries of *S*. *aethiopicum* Shum group and five of *S*. *anguivi* were used. Twelve plants of each accession were established in individual plastic pots of 5-L size in a completely randomized block design. Three seeds of an accession were directly sown in a pot on 10th Oct., 2016 followed by thinning to one plant per pot at 4-leaf stage (seedling stage) on 8th Nov., 2016. The potting soil was clay-loam. Optimum watering with uniform quantities of water on a daily basis, appropriate fertilizer application with NPK 17:17:17 at 4 g per pot on a fort-nightly basis and preventive pesticide sprays using mancozeb and dimethoate once every 2 weeks was carried out.

Data collection

Data was collected on; number of days to emergence, cotyledonous leaf length, cotyledonous leaf width, seedling leaf length, seedling leaf width, seedling fresh weight and seedling dry weight were recorded. Additional morphological traits were collected at flowering stage using a modified standard IBPGR *Solanum* species characterization manual.

Cluster analysis

The raw data was summarized in Excel to obtain means for different quantitative variables and cleaning up of the qualitative data followed by subsequent analysis in R (Everitt and Hothorn, 2014). A text delimited data frame of the mean values was imported into R followed by converting of traits to appropriate ordered (qualitative), nominal (qualitative) and numeric/integer (quantitative) variables. The target variables were then selected followed by installing and loading an R package cluster for cluster analysis. Because the data included both quantitative and qualitative variables, a function daisy() was used to group accessions based on a general coefficient of dissimilarity that combines and processes different types of variables according to their own mathematical type (Grum and Atieno, 2007; Zimisuhara et al., 2015). The hierarchical clustering was carried out using average (UPGMA) algorithm. A "re-ordered" dendrogram was plotted so that an accession in one cluster that has the smallest distance to accessions in the next cluster is the accession that is placed adjacent to the next cluster. In order to examine how well the

Table 1. List of accessions used in this study.

Entry	Code	Name (Pedigree)	Species name
1	168G	SAS168/G/2015	*S. aethiopicum* Shum
2	183G	SAS183/G/2015	*S. aethiopicum* Shum
3	163	SAS163/2015	*S. aethiopicum* Shum
4	163P	SAS/163/P/2015	*S. aethiopicum* Shum
5	157P	SAS/157/P/2015	*S. aethiopicum* Shum
6	160	SAS160/2015	*S. aethiopicum* Shum
7	163G	SAS163/G/2015	*S. aethiopicum* Shum
8	183P	SAS183/P/2015	*S. aethiopicum* Shum
9	108	SAS108/2015	*S. aethiopicum* Shum
10	157G	SAS157/G/2015	*S. aethiopicum* Shum
11	148	SAS/148/2015	*S. aethiopicum* Shum
12	145	SAS145/2015	*S. aethiopicum* Shum
13	168P	SAS/168/P/2015	*S. aethiopicum* Shum
14	184G	SAS184/G/2015	*S. aethiopicum* Shum
15	137	SAS137/2015	*S. aethiopicum* Shum
16	184P	SAS184/P/2015	*S. aethiopicum* Shum
17	141	SAS141/2015	*S. aethiopicum* Shum
18	108P	SAS108/P/2015	*S. aethiopicum* Shum
19	185G	SAS185/G/2017	*S. aethiopicum* Shum
20	185P	SAS185/P/2015	*S. aethiopicum* Shum
21	146	SAN146/2015	*S. anguivi*
22	177	SAN177/2015	*S. anguivi*
23	163W	SAN163/W/2015	*S. anguivi*
24	163C	SAN163/C/2015	*S. anguivi*
25	114	SAN114/2015	*S. anguivi*

distance (dissimilarity) matrix is represented graphically, a Mantel test that gives a cophenetic correlation coefficient (CPCC) was used. It measures the relationships between the original (true) pair-wise distance between accessions and pair-wise distances between accessions predicted using the dendogram. The CPCC is defined as a product-moment correlation coefficient between cophenetic distances and input distance matrix from the data; and the cophenetic distance between two accessions is the distance at which two accessions are first clustered together in a dendrogram going from bottom to top (Odong et al., 2011).

In order to ascertain that an optimum number of clusters was generated in the hierarchical tree, a Kelly-Gadner-Scutcliffe (1996) penalty function for pruning was calculated using a function kgs(). A function table() was then used to categorize the number of accessions in each cluster per level of qualitative variable. Diversity indices namely Richness, Shannon-Weaver and Simpson were also calculated. The Richness index defined here as the number of clusters represented in each qualitative variable was calculated using a function specnumber() that is dependent on (contained in) packages permute, lattice and vegan. The Shannon-Weaver index (swi) that combines a measure of richness with a measure of evenness was computed using a function diversity() whose default in R is set for the swi; otherwise an alternative index “simpson” that measures the evenness of group membership (Harding and Payne, 2012) was specified.

Principal component analysis

Principal component analysis (PCA) complements the cluster analysis in a way that the former helps to interrogate the data so as to understand the contribution of each variable to the existing diversity among accessions (Zimisuhara et al., 2015). The PCA procedure was performed in R on quantitative traits of the data frame using a base function prcomp() (Coghlan, 2017; Everitt and Hothorn, 2014). By default, the function prcomp() centres the variable to have mean equals to zero. Thus, the standard deviations were also set to 1 with the parameter scale=T in order to normalize the variables. The mean (center) and standard deviations (scale) of each variable, the principal component loadings (rotation) that constitute the rotation matrix which contains the principal component (PC) loading vector, and the matrix x that contains the PC score vectors were generated. A facility in the function prcomp() enables calculation of the standard deviation (sdev) of each PC (Coghlan, 2017). The variance (var) of each PC was then computed by squaring the sdev. The proportion of variance explained by each principal component (prop_varex) was calculated by dividing variance by total variance (that is, prop_varex= var/sum(var). In order to show the components that explain most of the variability in the data, a biplot and scree plots were used to plot the first two PCs, proportion of variance explained by each PC, and the cumulative proportion of variance explained.

Linear discriminant analysis

A multivariate analysis of variance (MANOVA) was conducted to identify variables that are significantly different between the two groups; *S. aethiopicum* Shum and *S. anguivi* at 99% confidence

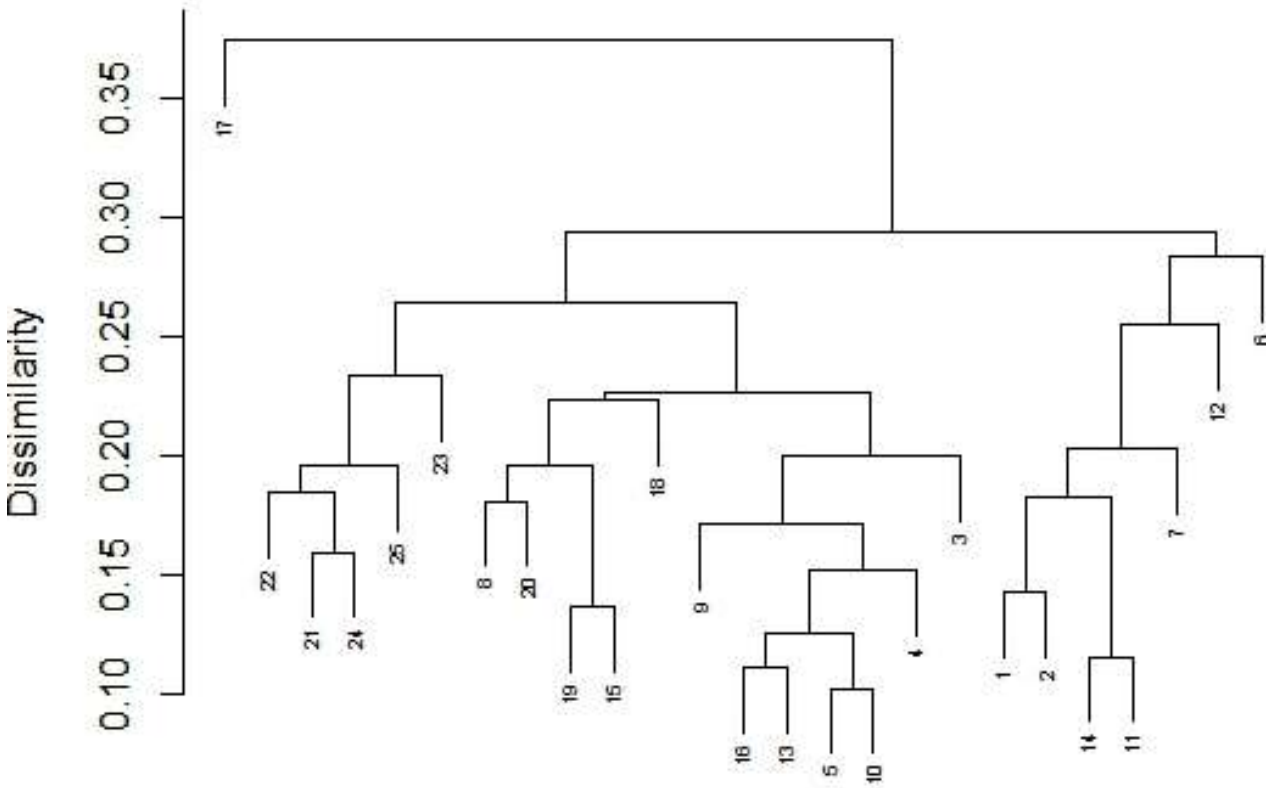

Figure 1. A re-ordered cluster dendrogram for the study accessions using the UPGMA method of agglomeration. The labels are the different study entries: 1-20 and 21-25 are the Shum and *S*. *anguivi* accessions, respectively.

level using GenStat Release 12.1 (VSN International Ltd). In MANOVA, the independent variables are the groups and the dependent variables are the predictors (Coghlan, 2017; Li and Wang, 2014). However, in LDA, the independent variables are the predictors and the dependent variables are the groups. The LDA was performed on data variables with significantly different means between the groups. Canonical vector loadings of discriminant function, correlations between data variates, and the correlations between the variates and discriminant function were generated. The maximum number of discriminant functions will be equal to the degrees of freedom, or the number of variables in the analysis, whichever is smaller (Coghlan, 2017); in this case the degrees of freedom is the smaller at 1 (that is, number of groups minus one). Thus, in this analysis, only one discriminant function was possible. The canonical loadings (standardized beta coefficients) were used to define the discriminant function (Harding and Payne, 2012). The larger the loading, the greater is the unique contribution of the respective variable to the discrimination between groups – without necessarily specifying the groups that the function discriminates (Coghlan, 2017). It is notable that in this case we are dealing with two groups; thus the loadings should give a reliable indication of the canonical data variable(s). Otherwise, a factor structure would be used to determine which variables define the discriminant function. The factor structure coefficients are the correlations between the data variates and the discriminant function; that denote the simple correlations between variables and the discriminant function (Harding and Payne, 2012). The pearsonian correlation coefficients between significant variables; and Mahalanobis (*D*-squared) intergroup distance (Harding and Payne, 2012; Zimisuhara et al., 2015) are also reported.

RESULTS

Clustering

Hierarchical clustering of the accessions based on the “average” (UPGMA) method produced five clusters (Figure 1). To test the goodness of the dendrogram by telling how well the distance (dissimilarity matrix) is represented graphically, the Mantel test revealed that the clusters were significantly distinct ($p<0.01$) with a cophenetic correlation coefficient of 0.73. On whether to prune the hierarchical cluster tree, a Kelly-Gardner Sutcliffe penalty function that compares the mean across all clusters with the mean within clusters of the dissimilarity measure further showed that the optimum number of clusters was five (Figure 2). The first and fifth clusters contained one accession each. The second, third and fourth cluster had five, twelve and six members, respectively. A summary of the cluster groups and their members are shown in Table 2. Entry 17 which comprises the first cluster was the only accession with spines. Table 3 shows how different qualitative traits are spread over the different clusters.

The Richness index (RI) was at maximum (at 5; the total number of clusters in the hierarchical tree) for

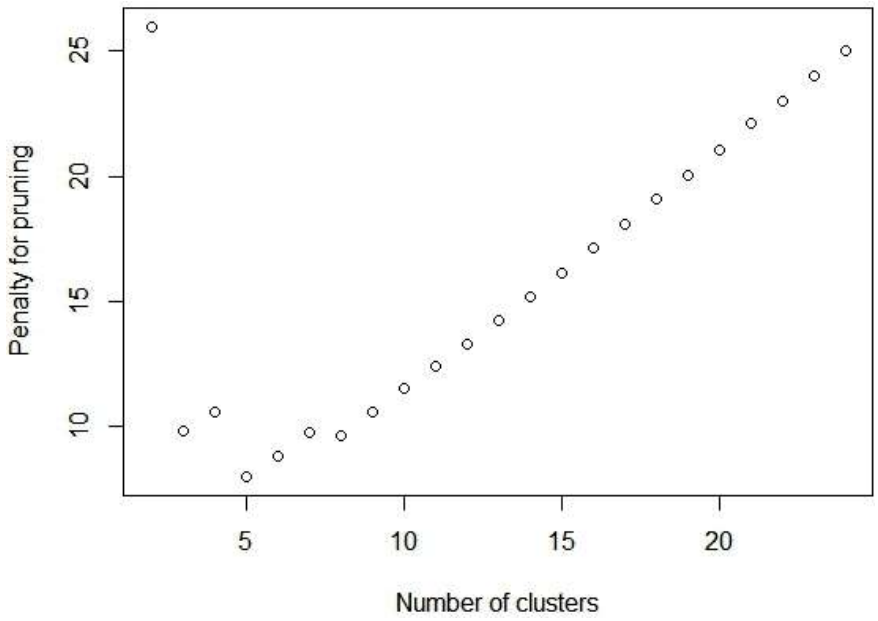

Figure 2. Kelley-Gardner-Sutcliffe penalty function for the cluster data showing that the optimum number of accessions is five.

Table 2. List of accessions per cluster.

Cluster name	Member (Entry)
1	17
2	22, 21, 24, 25 and 23 (all the *S*. *anguivi* accessions)
3	8, 20, 19, 15, 18, 19, 16, 13, 5, 10, 4 and 3
4	1, 2, 14, 11, 7 and 12
5	6

greenish white cotyledonous leaf color, poor seedling vigor, and acute leaf tip angle. The lowest RI of 1 (which implies that all accessions belonged to only one of the 5 clusters) was observed for many spines on stem, sparse stem pubescence, green stem color, green petiole color, very many prickles on lower leaf surface, very many prickles on upper leaf surface, many petiole prickles, weak leaf lobbing, very strong leaf lobbing, and purple leaf mid-rib color. Shannon-Weaver index (swi), a measure of richness and evenness, all clusters contained greenish white accessions based on cotyledonous leaf color (that is, swi is maximum at 1.27). The most diverse (in terms of number of clusters captured of out the total) and abundant (in terms of number of accessions represented) for spines on stem, seedling vigor, plant growth habit, stem pubescence, stem color, petiole color, leaf tip angle, leaf prickles was glabrous (swi=1.15), poor vigor (1.55), prostrate (1.33), medium (0.69), purple (0.95), acute (1.38), and glabrous (1.15). Generally, variable levels with high swi also had high values for the Simpson index (Table 4).

Principal component analysis

Sixteen principal components (PCs) were generated; the first two and ten PCs accounting for up to 51.53 and 96.83% of variation, respectively. The first PC that had higher loadings for days to germination (emergence; DG), leaf blade width (LBW), leaves per plant (LPP), leaf blade length (LBL) and leaf blade width (LBW) than the rest of the variables, accounted for 28.46% of variation. The second PC (23.07%) had high loadings for cotyledonous leaf length (CLBL), seedling fresh weight (SDFW), seedling leaf blade length (SLBL), seedling leaf blade width (SLBW) and days for first flower opening (FLW, Table 5). When represented on a scaled biplot such that the longer the arrows the higher the contribution to variation, it was shown that CLBL, LPP, DG, LBW, FLW, SDFW, SLBW, LBL, and plant width (PW) were shown to contribute to the highest variation among the study accessions (Figure 3). A scree plot showed that the first 10 PCs account for most of the variation at up to ~97% (Figure 4). Going by one variable per PC based on

Table 3. Number of members per cluster in the different levels of qualitative variables.

Variable / Levels	No. of accessions per cluster					Variable / Levels	No. of accessions per cluster				
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Cotyledonous leaf color											
Greenish white	1	5	12	6	1						
Spines on stem						**Petiole prickles**					
Glabrous	0	5	12	6	1	Glabrous	0	5	12	6	1
Many	1	0	0	0	0	Many	1	0	0	0	0
Visual seedling vigor						**Leaf blade lobbing**					
Intermediate	0	2	4	2	0	Weak	0	0	0	1	0
Poor vigor	1	2	1	2	1	Intermediate	0	0	6	3	0
Very poor vigor	0	1	5	1	0	Strong	1	5	6	2	0
Very vigorous	0	0	1	1	0	Very strong	0	0	0	0	1
Vigorous	0	0	1	0	0						
Plant growing habit						**Leaf blade color**					
Intermediate	0	3	5	1	1	Green	0	0	0	6	1
Prostrate	1	2	1	1	0	Pale purple	0	5	11	0	0
Upright	0	0	3	2	0	Purple	1	0	1	0	0
Very upright	0	0	3	2	0						
Stem pubescence						**Leaf midrib color**					
Glabrous	0	0	11	5	0	Green	0	0	0	6	1
Sparse	0	0	1	0	0	Pale purple	0	5	12	0	0
Medium	1	0	0	0	1	Purple	1	0	0	0	0
Dense	0	5	0	1	0						
Stem color						**Leaf pubescence on upper surface**					
Green	0	0	0	6	0	Glabrous	1	0	10	5	0
Pale purple	0	4	9	0	1	Sparse	0	1	2	0	1
Purple	1	1	3	0	0	Dense	0	4	0	1	0
Petiole color						**Leaf pubescence on lower surface**					
Green	0	0	0	6	0	Dense	0	4	0	1	0
Pale purple	0	5	10	0	1	Glabrous	1	0	10	5	0
Purple	1	0	2	0	0	Sparse	0	1	2	0	1
Leaf tip angle						**Leaf vein pigmentation**					
Acute	1	4	7	4	1	Green	0	0	0	6	1
Intermediate	0	1	5	2	0	Pale purple	1	5	12	0	0
Leaf prickles on lower surface						**Leaf prickles on upper surface**					
Glabrous	0	5	12	6	1	Glabrous	0	5	12	6	1
Very many	1	0	0	0	0	Very many	1	0	0	0	0

Diversity indices: Richness, Shannon-Weaver and Simpson.

Table 4. Richness, Shannon-Weaver (swi) and Simpson indices for the different variable levels.

Variable / levels	Richness	swi	Simpson	Variable / levels	Richness	swi	Simpson
Cotyledonous leaf color							
Greenish White	5	1.27	0.67				
Spines on stem				**Petiole prickles**			
Glabrous	4	1.15	0.64	Glabrous	4	1.15	0.64
Many	1	0.00	0.00	Many	1	0.00	0.00
Visual seedling vigor				**Leaf blade lobbing**			
Intermediate	3	1.04	0.63	Weak	1	0.00	0.00
Poor Vigor	5	1.55	0.78	Intermediate	2	0.64	0.44
Very Poor Vigor	3	0.80	0.45	Strong	4	1.20	0.66
Very Vigorous	2	0.69	0.50	Very strong	1	0.00	0.00
Vigorous	1	0.00	0.00				
Plant growing habit				**Leaf blade color**			
Intermediate	4	1.17	0.64	Green	2	0.41	0.24
Prostrate	4	1.33	0.72	Pale purple	2	0.62	0.43
Upright	2	0.67	0.48	Purple	2	0.69	0.50
Very Upright	2	0.67	0.48				
Stem pubescence				**Leaf midrib color**			
Glabrous	2	0.62	0.43	Green	2	0.41	0.24
Sparse	1	0.00	0.00	Pale purple	2	0.61	0.42
Medium	2	0.69	0.50	Purple	1	0.00	0.00
Dense	2	0.45	0.28				
Stem color				**Leaf pubescence (upper surface)**			
Green	1	0.00	0.00	Glabrous	3	0.83	0.51
Pale Purple	3	0.83	0.50	Sparse	3	1.04	0.63
Purple	3	0.95	0.56	Dense	2	0.50	0.32
Petiole color				**Leaf pubescence (lower surface)**			
Green	1	0.00	0.00	Dense	2	0.50	0.32
Pale Purple	3	0.83	0.51	Glabrous	3	0.83	0.51
Purple	2	0.64	0.44	Sparse	3	1.04	0.63
Leaf tip angle				**Leaf vein pigmentation**			
Acute	5	1.38	0.71	Green	2	0.41	0.24
Intermediate	3	0.90	0.53	Pale purple	3	0.79	0.48
Leaf prickles (lower surface)				**Leaf prickles (upper surface)**			
Glabrous	4	1.15	0.64	Glabrous	4	1.15	0.64
Very Many	1	0.00	0.00	Very many	1	0.00	0.00

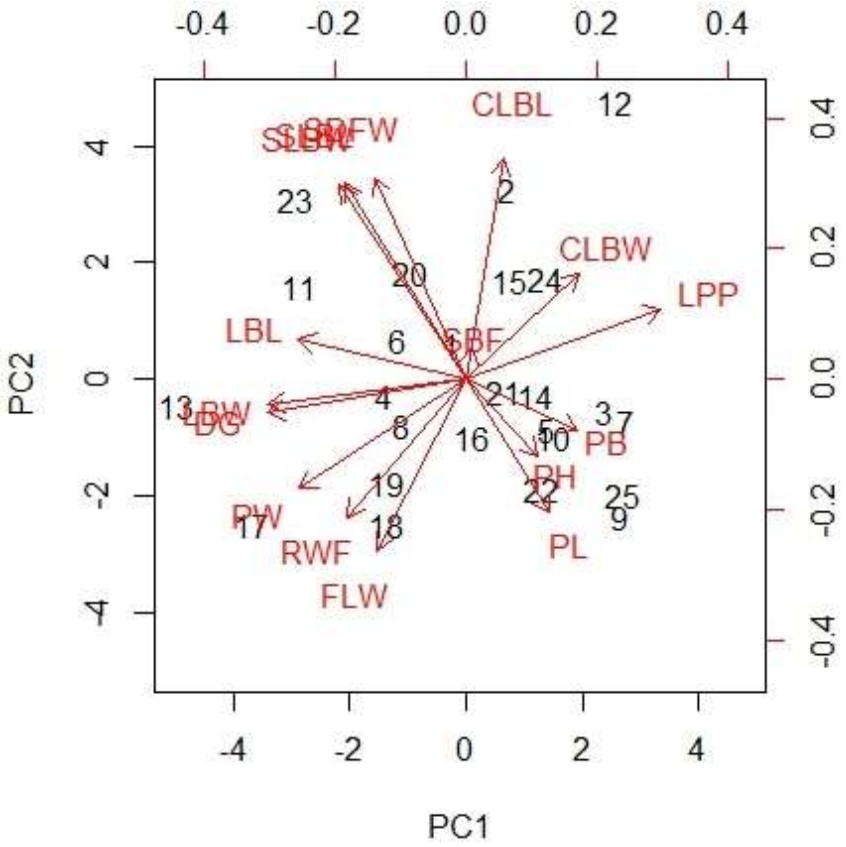

Figure 3. Biplot of the first two principal components of variation scaled by loadings (arrow length) to show contribution to variation. DG, days to germination and emergence; CLBL, cotyledonous leaf blade length (mm); CLBW, cotyledonous leaf blade width (mm); SLBL, seedling leaf blade length (mm); SLBW, seedling leaf blade width (mm); SDFW, seedling fresh weight (grams, g); LPP, number of leaves per plant; PH, plant height (cm); PB, number of branches per plant; PW, plant canopy width (cm); PL, petiole length (mm); LBL, leaf blade length (cm); LBW, leaf blade width (cm); SBF, shoot fresh biomass (g); RWF, root fresh weight (g); FLW, days to first flower appearance.

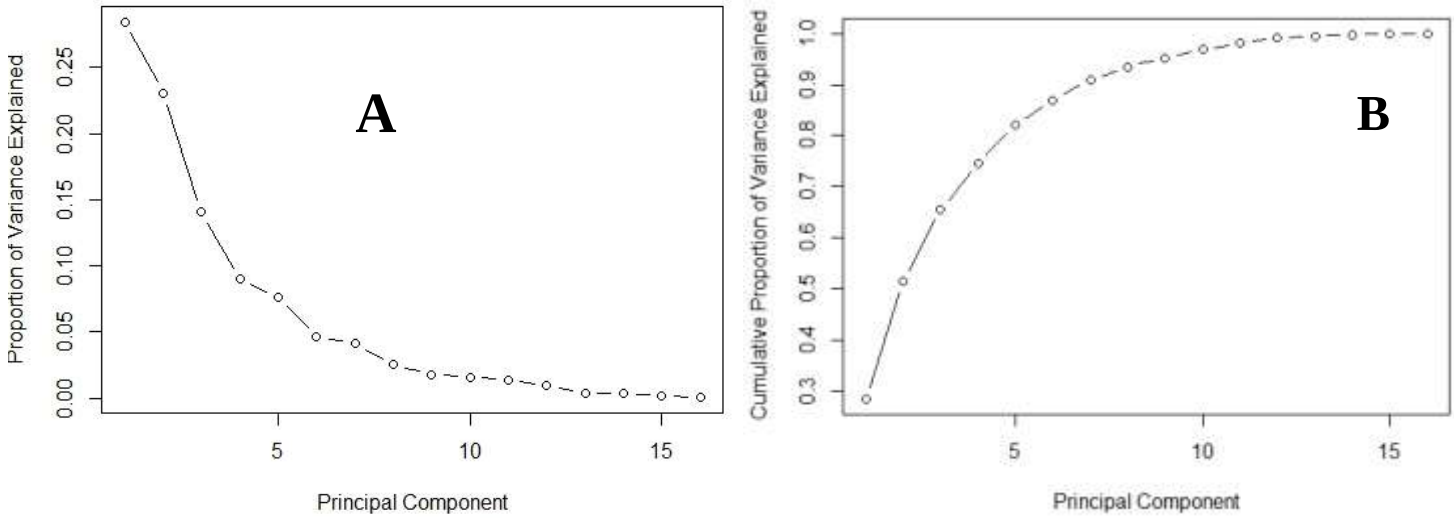

Figure 4. Scree plots for proportion (A) and cumulative proportion (B) of variance explained.

Table 5. The loadings and proportion of variation explained by each of 16 principal components among study accessions.

Variable	Mean	StDev	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
DG	7.60	3.24	-0.3802	-0.0649	0.1189	-0.0298	-0.2758	-0.3073	0.0391	-0.1807	0.5055	-0.0347	0.3806	0.2527	-0.1469	0.1179	-0.1958	-0.3037
CLBL	13.09	1.70	0.0718	0.4242	-0.1834	-0.1941	-0.1780	0.1842	0.0179	0.0618	0.0653	-0.4994	-0.3483	0.3363	-0.3531	0.2351	0.0527	-0.0554
CLBW	5.11	0.94	0.2164	0.2025	-0.4189	-0.2328	0.1742	0.0203	-0.2412	-0.0955	0.4711	0.0660	-0.0737	-0.1408	0.1404	-0.4985	-0.1373	-0.2141
SLBL	26.62	10.34	-0.2317	0.3756	0.2616	0.1378	0.1257	0.0673	-0.0281	-0.0915	0.0491	-0.0314	0.0928	-0.2125	-0.0303	-0.1388	0.7176	-0.3099
SLBW	19.00	7.20	-0.2422	0.3712	0.2708	0.1251	0.0677	0.0418	0.0356	-0.1101	0.2223	0.0164	-0.0167	0.0626	-0.0858	-0.2999	-0.1693	0.7176
SDFW	0.41	0.28	-0.1762	0.3860	0.2391	0.2278	0.1345	0.2222	0.0112	0.2223	-0.0263	0.0775	-0.1625	-0.1233	0.3157	0.2129	-0.5182	-0.3633
LPP	39.81	9.15	0.3737	0.1339	0.2214	0.0347	-0.0312	-0.0832	-0.1724	0.0100	0.2001	0.6794	-0.2243	0.3301	-0.1455	0.2261	0.1449	-0.0151
PH	29.55	8.38	0.1369	-0.1468	0.3035	-0.1712	0.6417	0.1245	0.1990	0.0689	-0.0611	-0.1270	0.1762	0.4422	-0.1697	-0.2286	-0.0859	-0.1784
PB	10.19	1.38	0.2147	-0.0966	0.3484	-0.4402	-0.1154	0.1760	0.2129	0.3861	0.3812	-0.0590	0.0940	-0.4420	-0.0546	0.1484	0.0391	0.1057
PW	48.58	4.15	-0.3191	-0.2092	-0.0100	0.0697	-0.2062	0.1804	-0.3888	0.6636	-0.0034	0.0733	-0.0522	0.2082	-0.1155	-0.3340	0.0731	-0.0340
PL	58.56	14.47	0.1597	-0.2546	0.2412	0.1610	-0.0203	0.4863	-0.5791	-0.3659	0.1457	-0.2477	0.0927	-0.0075	0.0521	0.1517	-0.0325	0.0380
LBL	20.95	2.80	-0.3212	0.0767	-0.2074	-0.4545	0.1737	0.1949	-0.0853	0.0161	0.0249	0.1414	0.1768	0.2743	0.5027	0.3236	0.2024	0.2007
LBW	14.59	2.42	-0.3789	-0.0488	-0.0983	-0.3190	0.1678	0.2150	-0.1135	-0.2280	-0.1862	0.3260	-0.0918	-0.2978	-0.5706	0.0902	-0.1660	-0.0632
SBF	137.75	32.75	0.0113	0.0620	0.4011	-0.4707	-0.4294	-0.0373	-0.0423	-0.2551	-0.3451	0.0187	-0.1650	0.1085	0.2274	-0.3536	-0.0807	-0.1384
RWF	51.26	12.23	-0.2292	-0.2637	0.1944	-0.0989	0.3108	-0.4619	-0.2100	-0.0047	0.1748	-0.2139	-0.6010	-0.0819	0.1136	0.1173	0.0657	0.0322
FLW	37.79	6.44	-0.1697	-0.3285	-0.0751	0.1611	-0.1229	0.4406	0.5202	-0.1973	0.2654	0.1325	-0.4043	0.1152	0.1187	-0.1275	0.1111	-0.0749
Standard deviation			2.1339	1.9211	1.5004	1.2013	1.1024	0.8597	0.8148	0.6329	0.5310	0.5034	0.4642	0.3834	0.2363	0.2281	0.1750	0.0809
Variance			4.5534	3.6908	2.2511	1.4430	1.2153	0.7390	0.6638	0.4006	0.2820	0.2534	0.2155	0.1470	0.0558	0.0520	0.0306	0.0065
Proportion			28.4589	23.0676	14.0696	9.0190	7.5954	4.6187	4.1490	2.5035	1.7625	1.5837	1.3470	0.9188	0.3489	0.3251	0.1915	0.0409
Cumulative			28.4589	51.5265	65.5962	74.6151	82.2105	86.8293	90.9783	93.4818	95.2442	96.8279	98.1749	99.0937	99.4426	99.7677	99.9591	100.00

DG, days to germination and emergence; CLBL, cotyledonous leaf blade length (mm); CLBW, cotyledonous leaf blade width (mm); SLBL, seedling leaf blade length (mm); SLBW, seedling leaf blade width (mm); SDFW, seedling fresh weight (grams, g); LPP, number of leaves per plant; PH, plant height (cm); PB, number of branches per plant; PW, plant canopy width (cm); PL, petiole length (mm); LBL, leaf blade length (cm); LBW, leaf blade width (cm); SBF, shoot fresh biomass (g); RWF, root fresh weight (g); FLW, days to first flower appearance.

scores (loadings), the ten principal components that account for the greatest differences among the study accessions include DG, CLBL, CLBW, shoot fresh biomass (SBF), plant height (PH), petiole length (PL), FLW, PW, PB, and LPP.

Discriminant analysis

There were significant differences (p<0.01) between the two groups of accessions; *S. aethiopicum* Shum, and *S. anguivi*) for CLBL, CLBW, DG, FLW, LPP, RWD, and SBD. The differences between the groups were however, non-significant (p>0.01) for LA, LBL, LBW, PB, PH, PL, PW, SDDW, SDFW, SLBL, SLBW and TBD (Table 6). Thus, subsequent linear discriminant analysis (LDA) was carried out on the seven significantly different variates between groups as predictors for the groups. The vector loadings (scores) that indicate unique contribution of each trait led to the following discriminant function: $Group = 0.0586CLBL - 0.2609CLBW + 0.6321DG + 0.0727FLW + 0.0627LPP + 0.0310RWD + 0.0200SBD$. In addition, simple correlations between variates and the discriminant function were generated. The variates with the highest loadings (0.6321 and -0.2609) and the highest correlations between the variates and discriminant function (0.7891 and -0.3491) were DG and CLBW, respectively (Table 7). The mean DG was 6.25 and 13.00 for *S. aethiopicum* Shum and *S. anguivi* accessions, respectively.

Table 6. Mean squares and probability values for rejecting a hypothesis of no difference (α=1%) between *S. aethiopicum* Shum and *S. anguivi* accessions

Variate	Mean for SAN	Mean for SAS	Between Groups MS	Within Groups MS	F.pr	Variate	Mean for SAN	Mean for SAS	Between Groups MS	Within Groups MS	F.pr
CLBL	12.10	13.34	51.028	3.527	<0.001	PH	27.41	30.09	163.89	81.64	0.163
CLBW	4.10	5.36	42.447	0.7524	<0.001	PL	60.67	58.04	524.1	339.2	0.22
DG	13.00	6.25	989.908	3.104	<0.001	PW	51.51	47.84	204.67	33.95	0.018
FLW	42.33	36.65	849.97	43.18	<0.001	RWD	13.60	9.84	313.76	29.88	0.002
LA	358.94	303.17	37735	23466	0.211	SBD	23.07	19.14	491.20	43.05	0.001
LBL	21.90	20.71	8.38	19.17	0.511	SDDW	0.04	0.04	0.00076	0.00168	0.505
LBW	16.23	14.18	85.45	14.19	0.018	SLBL	30.60	25.63	341.4	113.1	0.088
LPP	34.25	41.20	839.8	110.7	0.008	SLBW	23.20	17.95	385.54	53.55	0.01
PB	10.07	10.21	2.846	2.864	0.324						

SAN, *S. anguivi*; SAS, *S. aethiopicum*. Range for days to germination and emergence (DG) was 4-11 and 12-15 for SAS and SAN, respectively. CLBL, cotyledonous leaf blade length (mm); CLBW, cotyledonous leaf blade width (mm); LPP, number of leaves per plant; PB, number of branches per plant; LBL, leaf blade length (cm); LBW, leaf blade width (cmFLW, days to first flower appearance.

Table 7. Discriminant function (DF) and *r* for the correlations between variates and the DF.

Variate	Scores for DF	Coefficients for correlations between variates and the DF (*r*)
CLBL	0.0586	-0.1251
CLBW	-0.2609	-0.3491
DG	0.6321	0.7891
FLW	0.0727	0.1971
LPP	0.0627	-0.1378
RWD	0.031	0.15
SBD	0.02	0.1561

DG, days to germination and emergence; CLBL, cotyledonous leaf blade length (mm); CLBW, cotyledonous leaf blade width (mm); LPP, number of leaves per plant; SBD, shoot dry biomass (g); RWD, root dry weight (g); FLW, days to first flower appearance.

The Mahalanobis' intergroup distance (D^2) between the two groups was estimated to be 24.51. The discriminant scores for the group means for SAN and SAS were 3.864 and -1.087, respectively (Figure 5).

DISCUSSION

A moderately high CPCC obtained indicates presence of structure or strong group (subgroup) differentiation among the study accessions. It is notable however, that the higher the CPCC value up to over 0.9, the better the usefulness of dendrograms especially for taxonomic purposes (Coghlan, 2017; Odong et al., 2011). The clustering was unbalanced considering that one of the groups contained only one member compared to twelve in one of the four remaining groups. The dendrogram produced was the most appropriate with the data used considering that five clusters were shown; in concordance with optimum number read from the Kelley-Gardner-Sutcliffe penalty function (Grum and Atieno, 2007; Kelley et al., 1996). The UPGMA typically produces unbalanced dendrograms, leading to exposure of outliers (Grum and Atieno, 2007; Odong et al., 2011) like entry 17 that had leaf prickles. Leaf prickles are not a common attribute within the *S. aethiopicum* Shum and its progenitor *S. anquivi* (Adeniji et al., 2012, 2013). As a leafy vegetable, the *S. aethiopicum* Shum need not possess leaf spines unless prickliness is a marker associated with a yield, quality or other desired attribute like tolerance to a major productivity constraint.

The Richness index of 1 for leaf prickles further indicates that only one cluster of the five clusters was represented; implying that apart from entry 17, the rest of the 24 accessions spread in the 2-5th cluster did not have spines. For the Shannon-Weaver index (swi), the higher the value the higher the diversity and abundance of a certain category of qualitative variable. It is thus suggested that poor seedling vigor, acute leaf tip angle, prostrate plant growth habit, greenish white cotyledonous

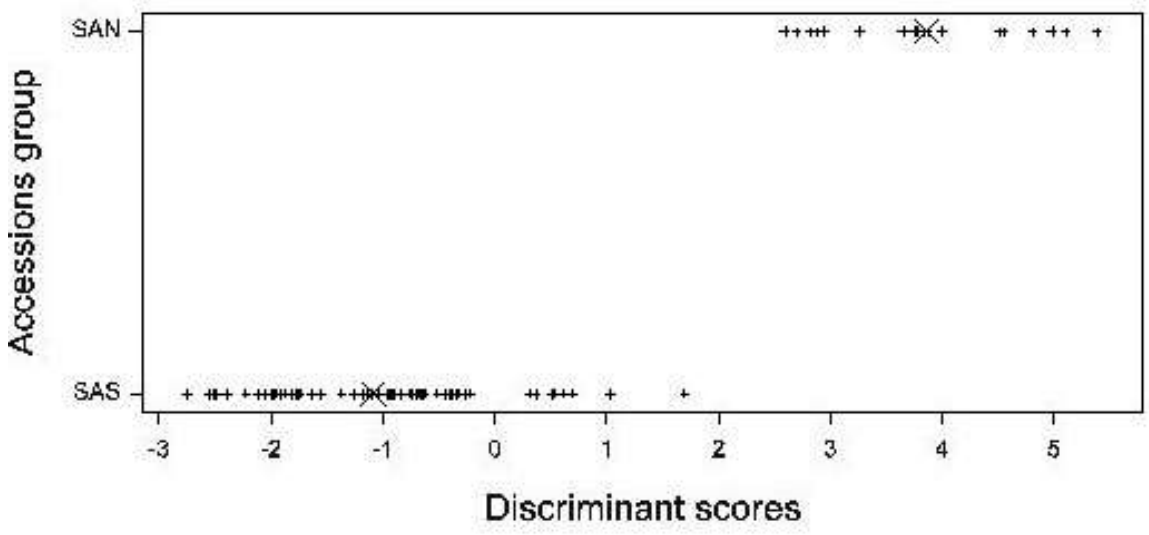

Figure 5. The discriminant scores for the two groups; SAS (*S. aethiopicum* Shum) and SAN (*S. anguivi*) accessions. The position of a group mean score is marked with an 'X'.

leaf color, strong leaf blade lobbing, glabrous stem prickliness, glabrous petiole prickliness, sparse leaf pubescence, and glabrous leaf prickliness are the most diverse and abundant attributes among the study accessions. The poor seedling vigor category had the highest Simpson index; suggesting a higher abundance of the accessions with poor vigor. The other variable categories that are highly abundant include prostrate plant growth habit, acute leaf tip angle, and greenish white cotyledonous leaf color. The Simpson index is a measure of evenness of group membership or the likelihood of two randomly selected accessions being different from each other (Coghlan, 2017; Harding and Payne, 2012).

The PCA indicated that by just selecting the variables leading for loadings in the first two PCs; at least half of the drivers of diversity among the accessions are captured. To this end, it is suggested that cotydonous leaf blade length (CLBL), number of leaves per plant (LPP), days to germination and emergence (DG), leaf blade width (LBW), days to first flower opening (FLW), seedling fresh weight (SDFW), seedling leaf blade width (SLBW), leaf blade length (LBL), and plant width (PW) greatly contribute to the variation captured by the first two PCs. Because the aim of conducting PCA is two pronged; reduce redundancy and retaining variables that explain as much variation as possible (Coghlan, 2017), a further scrutiny with help of the scree plots guided that the first 10 PCs explain up to ~97% of the diversity as revealed by hierarchical clustering. It is notable however, that the PCA was based on quantitative variables only. Therefore, DG, CLBL, CLBW, shoot fresh biomass (SBF), plant height (PH), petiole length (PL), FLW, PW, plant branching (PB), and LPP account for much of diversity (say ~97%) in the study accessions based on quantitative traits.

Based on group means for discriminant scores and Mahalanobis' intergroup distance (Harding and Payne, 2012), the morphological data clearly classified the Shum and *S. anguivi* accession groups as distinct. The two species can generally be distinguished based on variates; CLBL, CLBW, DG, FLW, LPP, RWD, and SBD in a discriminant function: $Group = 0.0586CLBL - 0.26CLBW + 0.63DG + 0.073FLW + 0.063LPP + 0.031RWD + 0.02SBD$. It is thus suggested that a clear distinction between the Shum and its progenitor can be made at seedling (CLBL, CLBW and DG), flowering (FLW) and harvest maturity (LPP, SBD and RWD). A correlation between the variates and discriminant function suggested that the DG, having a strong Pearsonian correlation coefficient (*r*) is a major canonical discriminant variate. The *r* obtained for DG was positive and strong; though a direction of the correlation would not affect interpretation in this case. It is thus suggested that a screening out of non-Shum genotypes can be done if germination and emergence exceeds eleven (min., 4; max., 11; mean, 6.25) days from the time of sowing; under conditions similar to those used in this study. Whereas morphological markers provide a good distinction among *S. aethiopicum* and between the Shum and *S. anguivi* accessions as reported earlier (Kouassi et al., 2014), a follow-up validation study with molecular markers could be complementary.

Conclusion

There is significant structure within the *S. aethiopicum*

Shum and *S*. *anguivi* accessions studied; with distinct clusters. Qualitative variables put aside, the principal component analysis revealed that diversity is mainly contributed by differences in days to germination and emergence, cotyledonous leaf length, cotyledonous leaf width, shoot fresh biomass, plant height, petiole length, days to flowering, plant width, plant branching, and number of leaves per plant. It was further revealed that the days to germination and emergence provide the greatest separation between the Shum and *S*. *anguivi* progenitors; with the former emerging earlier than the latter. Other traits which were more favorable among the Shum than the *S*. *anguivi* accessions include number of leaves per plant, number of branches per plant and plant height. This information is useful in suggesting germplasm conservation and breeding approaches for development of improved varieties of *S*. *aethiopicum* Shum.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

ACKNOWLEDGEMENTS

This study was funded at WACCI - University of Ghana Legon by the Intra-ACP – Mobility Project for Crop Scientists for Africa Agriculture (CSAA) coordinated at Makerere University; and co-funded by a top-up scholarship from the Germany Academic Exchange Service (WACCI/DAAD). The germplasm was obtained from the Department of Agricultural and Biological Sciences, Uganda Christian University.

REFERENCES

Abukutsa-Onyango M (2014). Strategic repositioning African indigenous vegetables and fruits with nutrition, economic and climate change resilience potential. In: Novel Plant Bioresources (Gurib-Fakim A. (Eds)). London: John Wiley and Sons. pp. 361-369.

Adeniji O T, Kusolwa P, Reuben S (2013). Morphological descriptors and micro satellite diversity among scarlet eggplant groups. Afr. Crop Sci. J. 21(1):37-49.

Altaye T (2015). Determination of genetic diversity and population structure in eggplant. Retrieved from http://openaccess.iyte.edu.tr/handle/11147/4291

Cernansky R (2015). Super vegetables. Nature 522(7555):146.

Coghlan A (2017). A little book for R for multivariate analysis. Retrieved from http://media.readthedocs.org/pdf/a-little-book-of-r-for-time-series/latest/a-little-book-of-r-for-time-series.pdf

Ebert A (2014). Potential of underutilized traditional vegetables and legume crops to contribute to food and nutritional security, income and more sustainable production systems. Sustainability 6(1):319-335.

Everitt B, Hothorn T (2014). A handbook of statistical analyses in R (3rd Edition). Chapman and Hall/CRC.

Gramazio P, Blanca J, Ziarsolo P, Herraiz FJ, Plazas M, Prohens J, Vilanova S (2016). Transcriptome analysis and molecular marker discovery in *Solanum incanum* and *S*. *aethiopicum*, two close relatives of the common eggplant (*Solanum melongena*) with interest for breeding. BMC Genomics 17(1).

Grum M, Atieno F (2007). Statistical analysis for plant genetic resources: clustering and indices in R made simple. Rome, Italy: Biodiversity International.

Harding S, Payne R (2012). A guide to multivariate analysis in GenStat (15th Edition). 5 The Waterhouse, Waterhouse Street, Hemel Hempstead, Hertfordshire HP1 1ES, UK: VSN International.

Kelley LA, Gardner SP, Sutcliffe MJ (1996). An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. Protein Eng. 9(11):1063.

Kouassi A, Béli-Sika E, Tian-Bi T, Alla-N'Nan O, Kouassi AB, N'Zi JC, … Tio-Touré B (2014). Identification of three distinct eggplant subgroups within the *Solanum aethiopicum* Gilo group from Côte d'Ivoire by morpho-agronomic characterization. Agric. 4(4):260-273.

Kubie LD (2013). Evaluation of genetic diversity in garden egg *(Solanum aethiopicum*) germplasm in Ghana. University of Ghana. Retrieved from http://ugspace.ug.edu.gh/handle/123456789/5825

Li C, Wang B (2014). Fisher Linear Discriminant Analysis. http://www.ccs.neu.edu. Retrieved from http://www.ccs.neu.edu/home/vip/teach/MLcourse/5_features_dimensions/lecture_notes/LDA/LDA.pdf

Odong TL, van Heerwaarden J, Jansen J, van Hintum TJL, van Eeuwijk FA (2011). Determination of genetic structure of germplasm collections: are traditional hierarchical clustering methods appropriate for molecular marker data? Theor. Appl. Genet. 123(2):195-205.

Sękara A, Cebula S, Kunicki E (2007). Cultivated eggplants–origin, breeding objectives and genetic resources, a review. Folia Hort. 19(1):97-114.

Zimisuhara B, Valdiani A, Shaharuddin N, Qamaruzzaman F, Maziah M (2015). Structure and Principal Components Analyses Reveal an Intervarietal Fusion in Malaysian Mistletoe Fig (*Ficus deltoidea* Jack) Populations. Int. J. Mol. Sci. 16(7):14369-14394.