


Genetic diversity in cultivated yam bean (*Pachyrhizus* spp.) evaluated through multivariate analysis of morphological and agronomic traits

A. Séraphin Zanklan  · Heiko C. Becker · Marten Sørensen · Elke Pawelzik · Wolfgang J. Grüneberg

Received: 22 June 2016 / Accepted: 7 October 2017 / Published online: 28 December 2017
© The Author(s) 2017. This article is an open access publication

Abstract Yam bean [*Pachyrhizus* DC.] is a legume genus of the subtribe Glycininae with three root crop species [*P. erosus* (L.) Urban, *P. tuberosus* (Lam.) Spreng., and *P. ahipa* (Wedd.) Parodi]. Two of the four cultivar groups found in *P. tuberosus* were studied: the roots of ‘Ashipa’ cultivars with low root dry matter (DM) content similar to *P. erosus* and *P. ahipa* are traditionally consumed raw as fruits,

whereas ‘Chuin’ cultivars with high root DM content are cooked and consumed like manioc roots. Interspecific hybrids between yam bean species are generally completely fertile. This study examines the genetic diversity of the three crop species, their potentials for breeding and the identification of useful traits to differentiate among yam bean genotypes and accessions. In total, 34 entries (genotypes and accessions) were grown during 2000–2001 at two locations in Benin, West Africa, and 75 morphological and agronomical traits, encompassing 50 quantitative and 25 qualitative characters were measured. Diversity between entries was analyzed using principal component analysis, cluster analysis, multivariate analysis of variance and discriminant function analysis. Furthermore, phenotypic variation within and among species was investigated. Intra- and interspecific phenotypic diversity was quantified using the Shannon–Weaver diversity index. A character discard was tested by variance component estimations and multiple regression analysis. Quantitative trait variation ranged from 0.81 (for total harvest index) to 49.35% (for no. of storage roots per plant). Interspecific phenotypic variation was higher than intraspecific for quantitative traits in contrast to qualitative characters. Phenotypic variation was higher in overall for quantitative than qualitative traits. In general, intraspecific phenotypic variation ranged from 0.00 to 82.61%, and from 0.00 to 80.03% for quantitative and qualitative traits, respectively. Interspecific phenotypic variation ranged from

Electronic supplementary material The online version of this article (doi:<https://doi.org/10.1007/s10722-017-0582-5>) contains supplementary material, which is available to authorized users.

A. S. Zanklan (✉)
Département de Biologie Végétale, Faculté des Sciences et Techniques, Université d’Abomey-Calavi, 01 BP 526, Cotonou, Benin
e-mail: azanklan@daad-alumni.de

H. C. Becker
Department of Crop Sciences, Göttingen University,
Von-Siebold-Str. 8, 37075 Göttingen, Germany

M. Sørensen
Section for Organismal Biology, Faculty of Life Sciences,
Copenhagen University, Thorvaldsensvej 40, 3,
1870 Frederiksberg C, Denmark

E. Pawelzik
Department of Crop Sciences, Göttingen University,
Carl-Sprengel-Weg 1, 37075 Göttingen, Germany

W. J. Grüneberg
Department of Genetic Resources and Crop Improvement,
International Potato Center, P.B. 1558, Lima 12, Peru

0.00 to 95.02%, and 0.00–81.58% for the two trait types, respectively. The Shannon–Weaver diversity index (H') was in general high and over 0.80 for most of the trait. Diversity within *P. tuberosus* was higher than within *P. erosus* and *P. ahipa*. Across the 50 quantitative and 25 qualitative traits, the Shannon–Weaver diversity index of intra- and interspecific variation was around 0.83 and 0.51, respectively and was lower for qualitative than for quantitative traits. Monomorphism was observed in eight qualitative traits and one quantitative character. The first, second and third principal components explained, respectively, 39.1, 21.3 and 8.3% of the total variation in all traits. *Pachyrhizus erosus*, *P. ahipa*, and *P. tuberosus* ('Chuin' and 'Ashipa') were clearly separated from each other by these analyses. Multivariate analysis of variance indicates significant differences between *Pachyrhizus* species for all individual or grouped traits. Discriminant function analysis revealed that the first two discriminant functions were almost significant. Biases due to unbalanced sample size used per species were small. Within each species a similar amount of diversity was observed and was determinable to 70% by only ten traits. We conclude that the cultivated yam bean species represent distinct genepools and each exhibits similarly large amounts of genetic diversity.

Keywords Agronomic traits · Genetic diversity · Yam bean · West Africa

Abbreviations

BIOM	Total biomass
DFA	Discriminant function analysis
MVA	Multivariate analysis
MANOVA	Multivariate analysis of variance
PV	Phenotypic variation
SEY	Seed yield
SRDY	Storage root dry matter yield

Introduction

Many thousands of plant species can be used by humanity, and around a hundred have been developed into crops. However, as only a few crops are widely grown today research interest into the so-called underutilized crops is rapidly growing—among them the yam beans (*Pachyrhizus* spp.). The nearest relative of economic importance is the soybean

(*Glycine max* (L.) Merr.) and the levels of oil and protein of yam bean seeds resemble those typical of soybean (Grüneberg et al. 1999). Formerly, the genus *Pachyrhizus* was placed in the subtribe Diocleinae in close relationship to the subtribe Glycininae and Phaseolinae (Lackey 1977; Ingham 1990), but based on chloroplast DNA restriction site mapping, it was transferred to the subtribe Glycininae (Bruneau et al. 1994; Polhill 1994). Within the Glycininae, the yam bean shows a close relationship to tropical kudzu (*Pueraria phaseoloides* (Roxb.) Benth.) and other genera with a chromosome base number of $x = 11$ (Lee and Hymowitz 2001; Kumar and Hymowitz 1987). The yam bean species are diploid ($2n = 22$), self-pollinating (up to 8% cross pollination) and native to South and Central America (Sørensen 1990). The genus is defined as a homogeneous entity due to the stigma structure having a median to subterminal globular process on the adaxial side, the short hairs on the adaxial side of the ovary extending almost to the stigma, and the formation of storage roots (Sørensen 1988). Unlike its close relative, the soybean, the yam bean is exclusively used for its storage roots (Ramos-de-la-Peña et al. 2013). The use of yam bean seeds as source of biodegradable insecticide is also of potential economic interest due to their high rotenone contents (Lautié et al. 2012). The crop is the most important storage-root-forming legume, as its productivity is high and it has also high protein content in the storage roots (NRC 1979). In the cultivated species, due to the roots' high moisture content, and their traditional raw consumption, they have been considered exclusively as fruity vegetables.

The genus *Pachyrhizus* encompasses two wild (*P. ferrugineus*, *P. panamensis*) and three cultivated species: Amazonian yam bean (*P. tuberosus*), Mexican yam bean (*P. erosus*), and Andean yam bean (*P. ahipa*). The cultivated species are separated taxonomically on morphological and physiological traits using univariate statistics (Sørensen 1988; Sørensen et al. 1997a, b): (1) *P. ahipa*—in contrast to *P. tuberosus* and *P. erosus*—is bushy or semi-erect with generally entire leaflets and with short racemes, which are only basally dibotryoid; it is day length insensitive and only found cultivated in cool tropical and subtropical Andean valleys within 1800–2900 m a.s.l. (2) *P. tuberosus*—in contrast to *P. erosus*—has wing and keel petals that are ciliolate and rarely

glabrous; the legume at maturity is 13–14 cm long and the seeds are plump and reniform with the exception of the square seeds of the ‘Chuin’ cultivar group (Sørensen et al. 1997a); usually plants are larger than *P. erosus* and *P. ahipa*, i.e. the stem can reach up to 10 m in length, but semi-erect types can be found that exhibit growth type similar to those of semi-erect *P. ahipa*; the habitat is wet tropical lowlands of Central and South America and the slopes of the Andean mountain range within an altitude range from sea level to 2000 m a.s.l. (Sørensen et al. 1997b) (3) *P. erosus* has wing and keel petals that are glabrous; the pod is glabrous to strigose at maturity and 6–13 cm long; the seeds are flat and square to round; it is widely distributed throughout many tropical and subtropical regions in South and Central America, South and East Asia and the Pacific from sea level to 2200 m a.s.l. (Sørensen 1988).

Interspecific crosses among all three cultivated yam bean species result in fertile and vigorous hybrids with one exception, i.e. *P. ahipa* × *P. tuberosus* ‘Ashipa’ yielded non-functional seeds (Grum 1990; Sørensen 1991; Grüneberg et al. 2003; Agaba et al. 2017). From the breeder’s perspective, the species form one primary gene pool. In a validation of the taxonomic separation of Sørensen (1988), no clear overall separation between the three cultivated species was found by Døygård and Sørensen (1998), based on 18 morphological characteristics as characteristics of the tuberous roots is not included in the herbarium material analysed by principal component analysis. The authors concluded that flower and inflorescence characters appeared to be the major differences, followed by leaf, legume and seed characteristics among accessions of the yam bean gene pool. Using molecular characterization [random amplified polymorphic DNA (RAPD) markers], Estrella et al. (1998) observed a clear gene pool separation between *P. tuberosus* and *P. erosus*. However, that study did not consider *P. ahipa*.

Pachyrhizus erosus and *P. ahipa* are not subdivided into cultivar groups, but for *P. tuberosus* four cultivar groups are distinguished: ‘Chuin’, ‘Ashipa’, ‘Yushpe’ and ‘Jíquima’ (Sørensen et al. 1997a, b; Tapia and Sørensen 2003; Oré-Balbin et al. 2007). Both ‘Jíquima’ and ‘Ashipa’ have low storage-root dry matter content similar to that of *P. erosus* and *P. ahipa*, whereas ‘Chuin’ and ‘Yushpe’ cultivars

exhibit a high storage root dry matter content (Grüneberg et al. 1998; Oré-Balbin et al. 2007). The Peruvian ‘Chuin’ type of *P. tuberosus* was first reported by Tessmann (Tessmann 1930; Sørensen et al. 1997a, b) and is cooked and consumed like cassava from the root of the manioc plant (Sørensen et al. 1997a, b; Grüneberg et al. 2003). Its existence has caused researchers to conclude that the yam bean could be used and developed as a protein-rich starchy staple also outside its current area of cultivation along the Río Ucayali, Peru. Due to the later discovery, the ‘Yushpe’ cultivar group was not included in the study presented here. Studies of the genetic diversity within the cultivated species have been performed in *P. erosus* (Heredia-Zepeda and Heredia-Garcia 1994; Estrella et al. 1998) and in *P. tuberosus* (Sørensen et al. 1997a, b; Estrella et al. 1998; Tapia and Sørensen 2003). There is no study on the genetic diversity in *P. ahipa*, except those which had involved univariate descriptive statistics (Ørting et al. 1996) and no investigation of the genetic diversity comprising all the three species (*P. erosus*, *P. tuberosus* and *P. ahipa*), except the study by Santayana et al. (2014). This is of interest because breeding is aiming to combine the wide adaptation of *P. erosus*, the storage root quality of the ‘Chuin’ type in *P. tuberosus* and the bushy-erect growth type and day-length insensitivity of *P. ahipa*.

A good description of the plant materials is necessary for the effective use of germplasm resources and for crop improvement. Therefore, curators of genebanks characterize their materials, recording selected traits of an accession. Traditionally, these data are limited to highly heritable morphological and agronomic traits (Acquaah 2007). With increases in germplasm sizes and data on molecular, biochemical, morphological, and agronomic traits, multivariate statistical analysis (MVA) methods are receiving increasing interest. Some MVAs (e.g. multivariate analysis of variance, MANOVA, discriminant function analysis, DFA, and partial least square, PLS) are extensions of uni- and bivariate statistical methods appropriate for significance tests of statistical hypotheses. MVA methods classify and order large numbers of breeding material and, trait combinations. Unfortunately, handling data with multi-collinearity can be unwieldy and variables are best considered together since they may be interdependent (Becker 2011; Acquaah 2007; Hill

et al. 1998; Falconer 1989; Wricke and Weber 1986; Wricke 1962). Multivariate statistical analyses (MVA) will be advantageous in analyzing genetic diversity and classifying germplasm collections (Acquaah 2007; Mohammadi and Prasanna 2003). Recently, several studies have been conducted to assess the genetic diversity based on morphological traits and using multivariate procedures including principal component analysis (PCA) (Badu-Apraku et al. 2011; Zeba and Isbat 2011; Hailu et al. 2006), discriminant function analysis (DFA) (Francisco et al. 2011; Safari et al. 2008), multivariate analysis of variance (MANOVA) (Arms et al. 2016; Ukalska et al. 2006), and partial least squares (PLS) (Jaradat and Weyers 2011). MANOVA analyses differences among populations for a given trait, and the distinctiveness is studied with a number of vector variables combined (Acquaah 2007; Zhu 1990). Population or genotype discrimination can be achieved by discriminant function analysis (DFA). DFA as a post-cluster analysis method was able to recognize the accuracy of clustering when used by several researchers (Safari et al. 2008). The magnitude of each trait in the genetic diversity of tall fescue (*Festuca arundinacea* Schreb.) (Vaylay and Van Santen 2002), hairy vetch (*Vicia villosa* Roth) (Yeater et al. 2004) and groundnut (*Arachis hypogaea* L.) (Safari et al. 2008) was identified using discriminant functions.

Understanding the level of genetic diversity is helpful for the selection of parental genotypes and important in widening the genetic base of crops (Molosiwa et al. 2016). Assessment of crop diversity also allows efficient sampling and proper management of the germplasm (Van Hintum et al. 2000). Reports on application of MVA in yam bean research are still limited. Da Silva et al. (2016) characterized 64 yam bean accessions maintained in genebank in Brazilian Amazon in a location using PCA and cluster analysis. Tapia and Sørensen (2003) studied the morphological variation in a germplasm collection of *P. tuberosus* grown at CATIE (Costa Rica) by canonical statistics. Similar reports were highlighted on recent performance investigations in yam beans using descriptive statistics and PCA in Rwanda and Uganda, respectively (Ndirigwe et al. 2017; Agaba et al. 2016).

This study had three objectives: (1) to investigate the genetic variation comprising all species and cultivar groups currently of interest for breeding

using morphological as well as agronomic traits measured under field conditions by different assessment procedure; (2) to determine and compare the amount of diversity present between and within the three species by diverse multivariate analysis (MVA) methods; and (3) to determine the morpho-agronomic quantitative and qualitative traits that contribute most to the differentiation between and within species.

Materials and methods

Plant material

In total, 34 entries were studied: 14 *P. ahipa* entries [13 seed sample accessions; one genotype from each seed sample accession, except for sample AC214 from which two genotypes (AC214-109 and AC214-110) that differed in flower colour were selected], 14 *P. erosus* entries, and 6 *P. tuberosus* entries, including five ‘Chuin’ and one ‘Ashipa’ seed sample accession (Table 1).

Study sites and experimental design

The field experiments were carried out between June 2001 and January 2002 at two locations in Benin: the Centre Songhai in Porto-Novo (02°37'E, 06°29'N) and at the experimental station of INRAB (Institut National des Recherches Agricoles du Bénin) in Niaouli (02°18'E, 06°66'N). The soils at both locations were well-drained sandy red loams. Each species was grown in one of three separate but adjacent experiments in the same field, and each experiment was laid out as a randomized block design with two plot replications for each entry. The experimental plots consisted of four rows, each containing 24 plants; the distance between plots was 1 m. In the experimental plots, the planting distance was 0.75 m between rows and 0.25 m within rows. Two seeds were sown per planting station at a depth of approx. 2 cm. Five weeks after sowing surplus plants were removed leaving only one per planting station. In the case of *P. tuberosus* and *P. erosus*, two stakes were used as trellises. Weeds were manually removed every 2 weeks, and no fertilisers or pesticides were applied. The rainfall during the crop growth period was 460 mm and 393 mm at Songhai and Niaouli, respectively. The trials at Porto-

Table 1 List of tested entries and their passport data; AC = Andean yam bean (*Pachyrhizus ahipa*)

Entry code	CIP ^a code	Collector(s)	Country	Region	Longitude	Latitude
AC102-153	209006	Sørensen	Bolivia	Tarija	64°43'W	21°31'S
AC201-19	n.a.	Ørting and Grüneberg	Bolivia	Luribay	67°38'W	17°00'S
AC202-27	n.a.	Ørting and Grüneberg	Bolivia	Luribay	67°38'W	17°00'S
AC203-43	209001	Ørting and Grüneberg	Bolivia	Luribay	67°36'W	17°00'S
AC205-68	n.a.	Ørting and Grüneberg	Bolivia	Machaca	66°53'W	17°8'S
AC208-72	209002	Ørting and Grüneberg	Bolivia	Machaca	66°53'W	17°6'S
AC209-73	n.a.	Grum and Grüneberg	Bolivia	Tirata	67°46'W	16°46'S
AC213-92	n.a.	Ørting	Bolivia	Irupana	67°27'W	16°34'S
AC214-109	n.a.	Grüneberg	Bolivia	Arce	67°35'W	16°49'S
AC214-110	n.a.	Grüneberg	Bolivia	Arce	67°35'W	16°49'S
AC215-129	n.a.	Grüneberg	Bolivia	Arce	67°34'W	16°44'S
AC216-139	n.a.	Grüneberg	Bolivia	Arce	67°34'W	16°44'S
AC524	209004	Meisen ^d	Unknown	Unknown	Unknown	Unknown
AC525	209007	Valio	Bolivia	Ayopaya	66°10'W	17°23'S
EC006	n.a.	Sørensen	Mexico	Oaxaca	96°42'W	17°03'N
EC032	n.a.	Unknown	Mexico	Yucatan	89°01'W	20°48'N
EC033	n.a.	Unknown	Mexico	Yucatan	88°49'W	20°42'N
EC040	n.a.	Sørensen	Guatemala	Jutiapa	90°01'W	14°12'N
EC041	209016	Sørensen	Guatemala	Jutiapa	90°01'W	14°12'N
EC042	n.a.	Sørensen	Guatemala	Jutiapa	90°02'W	14°03'N
EC104	n.a.	CATIE no 17137	Mexico	Yucatan	88°58'W	20°13'N
EC114	209017	Unknown	Brazil	Para	51°51'W	3°23'S
EC204	n.a.	Sørensen	Mexico	Vera Cruz	96°57'W	19°25'N
EC253	n.a.	Hue Anh.	Vietnam	Tan An	106°39'E	10°59'N
EC533	209018	CheangKeong	Macao	Macao	113°54'E	22°2'N
EC550	n.a.	CAEB	Mexico	Guanajuato	100°53'W	20°31'S
EC557	n.a.	INIFAP/CIFAP	Mexico	Guanajuato	100°53'W	20°31'N
ECKEW	209019	Kew Botan. Garden	Mexico	Unknown	Unknown	Unknown
TC118 ^c	n.a.	Hyvert	Haiti	Nord Este	72°19'W	19°50'N
TC353 ^b	209008	Jensen and Thirup	Peru	Requena	73°53'W	5°05'S
TC354 ^b	209013	Jensen and Thirup	Peru	Requena	73°59'W	4°59'S
TC355 ^b	209014	Jensen and Thirup	Peru	Requena	73°50'W	5°07'S
TC361 ^b	209015	Huanta	Peru	Requena	73°53'W	5°05'S
TC362 ^b	209020	Huanta	Peru	Rio-Ucayali	73°12'W	3°42'S

EC = Mexican yam bean (*P. erosus*). TC = Amazonian yam bean (*P. tuberosus*)

^a CIP International Potato Center; ^b 'Chuin' type; ^c 'Ashipa' type; *n.a.* not available; ^d Jardin Botanique. Meisen. Belgium. no. 0494 of unknown origin (growth type similar to AC102)

Novo were irrigated during August and September (approx. 800 mm irrigation), whereas at Niaouli no irrigation was applied. The average annual temperature at the experimental sites ranges from 23 °C in

August to 28 °C in May. The average temperatures during this study were 28.1 °C (Songhai) and 27.2 °C (Niaouli).

Traits recorded

In total, 75 morpho-agronomic characters were measured (Tables 2, 3). The descriptors developed for *Phaseolus* spp., *Vigna* spp. and *Ipomoea batatas* (L.) Lam. were used with small modifications (IBPGR 1985, 1987, 1991). Data were recorded on a plot basis, six individual plants selected randomly from the two central rows of each plot and six plant parts of six individual plants selected randomly from the two central rows of each plot. Measurement unit and measurement/sampling procedure of each trait are given in Tables 2 and 3. A calliper was used to measure most of the morphological characters. Storage root crude protein content was determined by measuring the N content according to the Dumas method (Sweeney and Rexroad 1987) and multiplying the N content by 6.25; starch content was determined using the polarimetric standard analysis No. 123/1 (ICC 1999), while sucrose, glucose and fructose contents were analysed enzymatically as described by Boehringer (1983).

Quantitative traits

Fifty (50) quantitative traits were surveyed in each accession or line. Traits were recorded on variable sets represented by the following plant organs: storage root, seed, pod, flower, stem, leaf, and composite characters evolving more than a single plant organ in their magnitude. The quantitative traits encompassed 26 morphological and 24 agronomic and quality attributes. All the traits evaluated are listed in Table 2.

Qualitative traits

Twenty-five (25) qualitative traits were studied (Table 3). They were treated as quantitative since they showed continuous variation within and among species. All qualitative characters were morphological (Table 3) and scaled diversely according to the visual observation on frequency among genotypes (Table 3).

To sum up, investigated traits encompassed 50 quantitative (Table 2) and 25 qualitative characters (Table 3). Qualitative characters were described as exhibiting continuous or quantitative variation.

Statistical analyses

Phenotypic variation estimates

Statistical analyses were conducted using SAS6.12 (SAS 1997). Initially data were classified relative to the experimental factors: species (S), entries (A), and location (L). For each trait x_i the variance components were estimated relative to: species (σ_S^2), entries within species ($\sigma_{A(S)}^2$), locations (σ_L^2), and the error term (σ_e^2) comprising the genotype-location interactions and the plot error. Variance component estimations were performed using the SAS procedure REML (SAS 1997) and the statistical model: $Y_{ijkl} = \mu_i + s_{ij} + a(s)_{ik(j)} + l_{il} + \varepsilon_{ijkl}$, where Y_{ijkl} is the observed value of the i th trait of the j th species, for the k th entry and l th location, μ_i is the trial mean of the i th trait; s_{ij} , $a(s)_{ik(j)}$, and l_{il} are, respectively, the effects of species, entries within species, and locations; ε_{ijkl} is the error, comprising effects of genotype-location interactions and the plot error.

We estimated the phenotypic variation within ($PV_{A(S)}$) and between species (PV_S) as:

$$PV_{A(S)} = \sigma_{A(S)}^2 / \sqrt{V_p} \quad \text{and} \quad PV_S = \sigma_S^2 / \sqrt{V_p}$$

where $\sigma_{A(S)}^2$ is the sum of variances represented by within and between genotypes in a given species, σ_S^2 the sum of variances within and between species in the genus, σ_e^2 is the experimental error term including $G \times E$ interactions to take into account the inter-annual variation in the expression of a trait due to environmental different effects on genotypes.

The phenotypic and average genetic diversity (H') at species and genus levels for all traits were further assessed using the Shannon–Weaver (1949) index of diversity as applied by Al Khanjari et al. (2008) and were calculated based on phenotypic frequency of alleles controlling each trait category of descriptors. The differences among the groups for the levels of diversity were tested using the Wilcoxon non-parametric test implemented in JMP 7.0 software (SAS Institute, Inc. 2007). Each character was categorized into specific class states. The 25 qualitative and 50 quantitative characters were assigned to classes ranging from 1 to 22, and analyzed using the Shannon–Weaver diversity index (H' ; Shannon and Weaver 1949) as defined by Jain

Table 2 Variable sets and 50 observed yam bean quantitative characters, codes, measurement units and measurement procedures

Variable set (Plant organ)	Character	Code	Measurement unit and measurement/ sampling procedure
Storage root	Storage root fresh yield	SRFY	Fresh weight in $t\ ha^{-1}$ —measured at physiological maturity, for 24 plants from 4 rows
	Storage root dry matter content	SRDM ^c	Weight of sun-dried samples expressed as a % of SRFY
	Storage root dry matter yield	SRDY ^c	$SRDY = SRFY \times SRDM$
	Harvest index for storage root yield	HIR	$HIR = (SRDY/BIOM) \times 100$
	Length of storage roots	LS ^{a,c}	cm—measured at harvest; 6 plants within plot centre
	Width of storage roots	WS ^{a,c}	cm—measured at harvest; 6 plants within plot centre
	Length to maximum width of storage roots	MWS ^{a,c}	cm—measured at harvest; 6 plants within plot centre
	Number of storage roots per plant	NSP	Counted at harvest, for 6 plants within plot centre
	Protein content of storage roots	PRO ^c	Protein content calculated as N content (Dumas method) $\times 6.25$, then expressed as a % of SRDM
	Starch content of storage roots	STA ^c	% of SRDM—polarimetric analysis (ICC Standard No. 123/1)
	Sucrose content of storage roots	SUC ^{a,c}	% of SRDM—enzymatic analysis
	Glucose content of storage roots	GLUC ^c	% of SRDM—enzymatic analysis
	Fructose content of storage roots	FRUC	% of SRDM—enzymatic analysis
	Seed	Seed yield	SEFY ^b
Thousand-seed weight		TSW ^a	g—measured at physiological maturity, on two samples of 100 seeds
Harvest index for seed yield		HIS ^a	$HIS = (SEFY/BIOM) \times 100$
Seed number per pod		SNP	Counted at harvest, for 6 pods per plant from 6 plants within plot centre
Seed length, width and height		SL ^a , SW ^a , SH	mm—measured after 6 days sun drying, post-harvest, 6 plants within plot centre (5 pods per plant and 3 seed per pod)
Pod	Pod yield	PODY	$t\ ha^{-1}$ —measured at physiological maturity, for 24 plants from 4 rows
	Shell weight	SHEL ^a	$t\ ha^{-1}$ —measured at physiological maturity, for 24 plants from 4 rows
	Time of maturity	TM	No. of days from sowing to physiological maturity (80% of pods dry within the 2 centre rows)
	Pod length (including beak), width, and height	PL ^a , PW, POH	mm—measured after 6 days sun drying, post-harvest, 6 plants within plot centre (6 pods per plant)

Table 2 continued

Variable set (Plant organ)	Character	Code	Measurement unit and measurement/ sampling procedure
	Number of pods per plant	PN	Counted at harvest, for 6 plants within plot centre
	Pod degree and shape of curvature	PDS ^a	Angle—measured after 6 days' sun drying, post-harvest; 6 plants within plot centre (6 pods per plant)
	Pod beak length and curvature	PBL, PBC	mm and angle—measured after 6 days' sun drying, post-harvest; 6 plants within plot centre
Flower	Start of flowering	BF ^{a,b}	No. of days from sowing to start of flowering
	Time of flowering	TF ^a	No. of days from sowing to the time that 50% of plants within the centre rows were flowering
	Period of flowering	PF ^c	No. of days from start of flowering to end of flowering; 6 plants within plot centre
	Inflorescence length	IL	cm—measured at full flowering; 6 plants within plot centre (6 inflorescences per plant)
Stem	Time of emergence	TE	No. of days from sowing to the time that plant emergence was 50% within the centre rows
	Early vigour (width of first leaf)	EV	mm—at time of development of third leaf; two measurements per plant for 6 plants within plot centre
	Start of climbing	SC ^a	No. of days from sowing to begin of climbing; 6 plants within plot centre
	Plant height	PH ^a	cm—measured at time of full flowering, for 6 plants within plot centre
Leaf	Terminal leaflet length, width, and ratio of length to maximum width	TLL ^{b,c} , TLW ^c , TLMW	cm—assessed at full flowering, on 6 plants within plot centre (6 leaves per plant)
	Lateral leaflet length, width, and ratio of length to maximum width	LLL ^{b,c} , LLW ^c , LLMW ^a	cm—assessed at full flowering, on 6 plants within plot centre (6 leaves per plant)
	Number of leaves	LN	Counted at full flowering; 6 plants within plot centre
	Terminal leaflet lobe number	TLLN	Counted at full flowering, on 6 plants within plot centre (6 leaflets per plant)
	Lateral leaflet lobe number	LLL ^a	Counted at full flowering, on 6 plants within plot centre and 6 leaflets per plant
Composite agronomic traits (Root + Stem + Leaf + Pod)	Total biomass	BIOM ^{a,b,c}	BIOM = SRDY + VLW + PODY
	Weight of vines and leaves	VLW	Sun-dried weight in t ha ⁻¹ measured at physiological maturity for 24 plants from 4 rows
	Total harvest index	HITOT ^a	HITOT = ((SRDY + SEEY)/BIOM) × 100

^a Differences between *Pachyrhizus* accessions among species relied mainly upon those traits in MANOVA

^b These traits were sufficient in MANOVA to analyze the variability existing within *P. ahipa*

^c Traits that were necessary in analyzing the variability within *P. erosus* and *P. tuberosus* in MANOVA

Table 3 Variable sets and 25 observed yam bean qualitative characters; codes; measurement units and measurement procedures

Variable set (Plant organ)	Character	Code	Measurement unit and measurement/ sampling procedure
Storage root	Damage to storage roots by nematodes	DSN ^a	Scores from 0 to 6; 0 = no damage, 6 = high level of damage
	Damage to storage roots by insects	DSI	Scores from 0 to 6; 0 = no damage, 6 = high level of damage
	Shape of storage roots	SS	9 scores from 1 to 9: 1 = round, 2 = round elliptic, 3 = elliptic, 4 = ovate, 5 = obovate, 6 = oblong, 7 = long oblong, 8 = long elliptic, 9 = long irregular or curved—assessed at harvest; 6 plants within plot centre
	Colour of storage roots	CSR ^a	5 scores from 1 to 5: 1 = white, 2 = yellow, 3 = brown, 4 = purple-red, 5 = dark purple—assessed at harvest; 6 plants within plot centre
	Surface defects of storage roots	SDS ^a	9 scores from 0 to 8: 0 = absent, 1 = alligator-like skin, 2 = veins, 3 = shallow horizontal constrictions, 4 = deep horizontal constrictions, 5 = shallow longitudinal grooves, 6 = deep longitudinal grooves, 7 = deep constrictions and deep grooves—assessed at harvest; 6 plants within plot centre
	Secondary flesh colour of storage roots	SFC ^a	10 scores from 0 to 9: 0 = absent, 1 = white, 2 = cream, 3 = yellow, 4 = orange, 5 = pink, 6 = red, 7 = purple-red, 8 = purple, 9 = dark purple—assessed at harvest; 6 plants within plot centre
	Distribution of secondary flesh colour of storage roots	DSFC ^a	10 scores from 0 to 9: 0 = absent, 1 = narrow ring in cortex, 2 = broad ring in cortex, 3 = scattered spots in flesh, 4 = narrow ring in flesh, 5 = broad ring in flesh, 6 = ring and other areas of flesh, 7 = in longitudinal sections, 8 = covering most of the flesh, 9 = covering all flesh—assessed at harvest; 6 plants within plot centre
	Stalk of storage roots	SSR	6 scores from 0 to 9: 0 = sessile or absent, 1 = very short (< 2 cm), 3 = short (2–5 cm), 5 = intermediate (6–8 cm), 7 = long (9–12 cm), 9 = very long (> 12 cm)—measured at harvest; 6 plants within plot centre
	Cracking of storage roots	SCR	4 scores from 0 to 7: 0 = absent, 3 = few cracks, 5 = medium number of cracks, 7 = many cracks—assessed at harvest; 6 plants within plot centre
Seed	Colour of seeds	CSE	9 scores from 1 to 9: 1 = olive, 2 = brown, 3 = orange-red, 4 = dark red, 5 = pink, 6 = purple, 7 = purple or black with white mottled, 8 = black, 9 = other—assessed at harvest
Pod	Pod green colour	PC	Very light to very dark (5 scores)—assessed after 7 weeks of full flowering, on 6 plants within plot centre
	Dehiscence of pods	DP	3 scores from 3 to 7: 3 = absent, 5 = a little dehiscent, 7 = dehiscent—assessed at harvest; 6 plants within plot centre
	Colour of mature pods	CMP	3 scores from 1 to 3: 1 = yellow, 2 = brown, 3 = dark brown—assessed at harvest; 6 plants within plot centre
Flower	Flower colour of sepals	CS	Green, a little purple, purple; 6 plants within plot centre (12 flowers per plant)
	Flower colour of standard and wing	FCS, FCW ^{b,c}	White, pink, violet—6 plants within plot centre (12 flowers per plant)

Table 3 continued

Variable set (Plant organ)	Character	Code	Measurement unit and measurement/ sampling procedure
Stem	Plant type	PT ^{a,b,c}	Scores from 3 to 9; 3 = erect, 5 = semi-erect, 7 = spreading, 9 = pronounced spreading
	Stem colour	SCO	9 scores from 1 to 9: 1 = green, 3 = green with few purple spots, 4 = green with many purple spots, 5 = green with many dark purple spots, 6 = mostly purple, 7 = mostly dark purple, 8 = totally purple, 9 = totally dark purple—assessed at full flowering on 6 plants within plot centre
Leaf	Leaf green colour	LC	Very light to very dark (5 scores)—assessed at full flowering on 6 plants within plot centre
	Shape of central terminal leaflet lobe	SCTLL ^a	Absent to linear (narrow) (10 scores from 0 to 9)—assessed at full flowering on 6 plants within plot centre (6 leaflets per plant)
	Shape of central lateral leaflet lobe	SCLLL ^a	Absent to linear (narrow) (10 scores from 0 to 9)—assessed at full flowering, on 6 plants within plot centre (6 leaflets per plant)
	Terminal leaflet lobe type	TLLT ^a	Entire to very deep (6 scores from 0 to 9: 0 = entire, 1 = very slight lobes, 3 = slight, 5 = moderate, 7 = deep, 9 = very deep)—assessed at full flowering, on 6 plants within plot centre (6 leaflets per plant)
	Lateral leaflet lobe type	LLLT ^a	Entire to very deep (6 scores from 0 to 9: 0 = entire, 1 = very slight lobes, 3 = slight, 5 = moderate, 7 = deep, 9 = very deep)—assessed at full flowering, on 6 plants within plot centre (6 leaflets per plant)
Composite agronomic traits (Root + Stem + Leaf + Pod)	Damage to stem and leaves by insects	DSL ^a	Scores from 0 to 6; 0 = no damage, 6 = high damage
	Damage to stem and leaves by fungi	DSL ^f	Scores from 0 to 6; 0 = no damage, 6 = high level of damage

^a Differences between *Pachyrhizus* accessions among species relied mainly upon those traits in MANOVA

^b Those traits were sufficient in MANOVA to analyze the variability existing within *P. ahipa*

^c Traits that were necessary in analyzing the variability within *P. erosus* and *P. tuberosus* in MANOVA

et al. (1975) to calculate phenotypic variation of each accession:

$$H = - \sum_{i=1}^n P_i \ln P_i$$

where n is the number of phenotypic classes for a character and P_i is the genotype frequency or the proportion of the total number of entries in the i th class.

H was standardized by converting it to a relative phenotypic diversity index (H') after dividing it by $H_{max} = \log_e^{(n)}$

$$H' = \frac{\sum_{i=1}^n P_i \ln P_i}{H_{max}}$$

Principal component and cluster analyses

Principal component analysis was performed using SAS procedure PRINCOM (SAS 1997). The spatial relationships among entries (accessions and genotypes, respectively) were presented by plotting the scores of the first, second and third principal components in a three-dimensional space. Correlations of all traits with the first five principal components were calculated using the SAS procedure CORR, (SAS 1997) using the Pearson correlation coefficient. A cluster analysis was carried out using the SAS procedure CLUSTER (SAS 1997). All traits were standardized by their mean value and standard deviation [$z = (x - \bar{x})/s$] using the STD option of

the CLUSTER procedure. Euclidian distances were calculated and a cluster analysis, involving the unweighted group average linkage method (UPGMA), was conducted using the AVE option of the CLUSTER procedure. Cluster summaries were plotted using the SAS Macro DENDRO (Nicholson 1995). All traits with estimated ratios of $\sigma_{A(S)}^2/\sigma_\epsilon^2 > 2$ and significant correlation with at least one of the first five principal components were analysed by a multiple regression analysis to select useful traits to differentiate among entries. The multiple regression analysis was performed by SAS procedure REG with the selection option STEPWISE (SAS 1997). The dependent variables in the multiple regression model were the first five principal components, whereas the regressor variables were those traits with estimated variance component ratios of $\sigma_{A(S)}^2/\sigma_\epsilon^2 > 2$ (traits with considerable genetic variation between entries and low genotype-location interactions and plot errors). In a final analysis, an environment search was carried out to identify—on the basis of temperature and rainfall range (including irrigation) at the experimental sites—some relevant target set of yam bean production environments in the regions of the world. The environment search was carried out with ArcGIS software and the options: (1) temperature range of 23–28 °C, (2) a rainfall range of 400–1200 mm and (3) at least six consecutive months of the temperature and rainfall parameters.

Multivariate analysis of variance and discriminant function analysis

Further multivariate analyses including multivariate analysis of variance (MANOVA) and discriminant function analysis (DFA) methods were applied to distinguish within- and between-species variations on the 34 accessions and lines of *Pachyrhizus* studied. Analyses were based on four different data sets consisting of all the *P. ahipa*, *P. erosus*, *P. tuberosus* accessions, as well as the 34 genotypes taken together. MANOVA and DFA were fulfilled separately in *P. ahipa*, *P. erosus*, *P. tuberosus*, and on the other hand in all genotypes taken together with the raw data from all observed variables. A comparison of the outputs has led to conclude on the consequences of unbalanced entry number size upon the accuracy of the results. To test the power of discrimination across locations, the within- and

among-diversity were estimated considering $G \times L$ and $S \times L$ interactions, where G and S represent Genotype and Species, respectively. MANOVA and DFA were performed by partitioning the raw data into variable sets according to the plant organ they are inferred to Tables 2 and 3.

To determine the number of groups representing the optimal partition in the hierarchical tree, a multivariate analysis of variance (MANOVA) was performed. MANOVAs using the Wilks' lambda statistic and Hotelling test as well as Pillai's trace and Roys' Max root were performed using the raw data for all 75 variables with the MANOVA statement in JMP 7.0 (SAS Institute, Inc. 2007). In the MANOVA, sources of variation are as described by Lázaro-Nogal et al. (2015) and Ukalska (2006), and the followed model was used:

$$Y = 1_N m + XG + ZR + E$$

where Y is the $(N \times k)$ -dimensional observation matrix with k , the number of response traits; 1_N is the $(N \times 1)$ -dimensional unit vector; N , the total number of not empty subclasses in the two-way data set; m is the k -dimensional vector of the general means; X is the $(N \times a)$ -dimensional design matrix for genotypes; G is the $(a \times k)$ -dimensional matrix of the random genotypic effects; Z is the $(N \times b)$ -dimensional design matrix for locations; R is the $(b \times k)$ -dimensional matrix of the random location effects; and E is the $(N \times k)$ -dimensional matrix of the residuals. These estimates were calculated using the MANOVA option of JMP 7.0 (SAS Institute, Inc. 2007). A significant effect of genotype indicates genetically based phenotypic differences. This model was repeated with mixed models using restricted maximum likelihood (REML), testing for the fixed effects of location and the random effects of genotype and interaction.

A discriminant function analysis (DFA) was carried out in JMP 7.0 (SAS Institute Inc. 2007) to identify which variables best differentiate the groups identified in the hierarchical classification. The correlation of each variable with each discriminant function based on the structure matrix was used to create the discriminant function. These Pearson coefficients are structure coefficients or discriminant loadings and function like factor loadings in factor analysis. By determining the largest loadings for each discriminant function, insights were gained into how to name each function.

To avoid biases, which might occur due to heterogeneity of variance from difference in genotype size between the three species surveyed (e.g. among *P. tuberosus* represented by 6 accessions, and *P. ahipa* and *P. erosus* with 14 genotypes within each species), response variables were log transformed before the multivariate analyses, which as suggested by Hughes et al. (2009) are necessary to meet assumptions of normality and homogeneity of variance. Linear relationships among the variables investigated were not improved by logarithmic transformation and therefore untransformed data was used afterwards in the following analysis.

To analyze possible sampling error bias—that may due to unbalanced data or unequal genotype size among species analyzed here—three effect size estimators (η^2 , \mathcal{E}^2 , ω^2) were generated from the ANOVA, to take into account a possible lack in accuracy of our results, which relied upon an heterogeneity of variances. This analysis was performed following the procedure and the software elaborated by Skidmore and Thompson (2013). The total sample or entry size of 34 in the present study falls between the smaller (24) and the larger (48) sample sizes studied by Skidmore and Thompson (2013) with a number k of groups equal to 3, corresponding here to the three species involved in the study.

Results

Phenotypic variation

Variance component estimations (Tables 4, 5) show that for a large number of traits σ_S^2 or $\sigma_{A(S)}^2$ are larger than σ_e^2 . For many traits, σ_S^2 is larger than σ_e^2 , and $\sigma_{A(S)}^2$ is smaller than σ_e^2 , i.e. comprising most agronomical traits, but also several morphological traits; this includes situations with zero estimates for $\sigma_{A(S)}^2$ and genetic variation within species, respectively. For 40 traits, estimates of $\sigma_{A(S)}^2$ are larger than σ_e^2 . Among these are 33 traits for which the estimated ratios of $\sigma_{A(S)}^2/\sigma_e^2$ were > 2 and nearly all these are morphological traits.

Yam bean accessions and lines used appeared to be well differentiated from one another for all the 75 characters investigated, except for some qualitative traits. Table 6 reports the trait variation for each quantitative attribute. Significant differences were

observed among species as well as inside them for each variable, particularly for quantitative character; trait variation ranged from 0.81 (for total harvest index) to 49.35% (for number of storage roots per plant). Lower trait variation inferior to 10% was noted in 16 characters. High differences were, however, observed in many other traits such as beginning of flowering with a phenological trait variation of 20.93% (Table 6). Agronomic and quality characters were also highly varied (Table 6). The trait pod number per plant showed further a high variation (46.39%) within and among species of the germplasm collection under evaluation.

Phenotypic variation estimates among genotypes within species ($PV_{A(S)}$) in quantitative characters ranged from 0.00 to 82.61% (Table 7). $PV_{A(S)}$ was high and above 10% for most traits (Table 7). Among species, phenotypic variation (PV_S) ranged from 0.00 (for storage root sucrose content) to 95.02% (for ratio of lateral leaflet length to maximum width) indicating that interspecific variability is higher. Furthermore, that variability was clearly above 10% in all characters with the exception of storage root dry matter yield, ratio of storage root length to maximum width, vine start of climbing, and leaf number per plant (Table 7).

For qualitative characters, $PV_{A(S)}$ was in general higher than PV_S for most of the traits. It ranged from 0.00 (for damage of storage root by insects and nematodes) to 80.03% (for shape of storage root). For most of the qualitative traits, $PV_{A(S)}$ presented values above 50%, except for damage of stem and leaves by fungi, leaf green colour, and damage of stem and leaves by insects (Table 8). Among species variation (PV_S) for qualitative characters was rather smaller than at the intraspecific level (Table 8). The highest value for PV_S was scored in leaf green colour (81.58%), and the lowest in shape of storage root (16.12%). Most of the qualitative traits showed PV_S values between 25 and 50%; thus lower than $PV_{A(S)}$, where values are often clearly above 50% (Table 8).

Over all fourteen *P. ahipa* accessions and lines—with regard to quantitative characters—(Table 9), the mean Shannon–Weaver diversity index (H') value was highest for pod yield (1.00) followed by period of flowering (0.99), seed width (0.99), seed yield (0.99), time of maturity (0.96), and seed height (0.95). In all traits, H' is higher than 0.50, except for pod degree and shape of curvature (0.37) and pod beak degree curvature, which presented

Table 4 Variance component estimations of species (σ_s^2), entries within species ($\sigma_{A(S)}^2$) and the error term [(σ_e^2) comprising genotype-location interactions and plot errors] for 50 morphological and agronomic quantitative traits measured in 34 *Pachyrrhizus* entries

Variable set	Trait	Variance component estimates			Variable set	Trait	Variance component estimates		
		σ_s^2	$\sigma_{A(S)}^2$	σ_e^2			σ_s^2	$\sigma_{A(S)}^2$	σ_e^2
Storage root	SRFY	30.150	57.880	50.510	Storage root	SRDM	23.690	0.000	15.580
	SRDY	0.540	2.020	3.090		HIR	15.920	76.600	38.990
	LS	20.460	1.230	18.140		WS	4.790	5.380	2.190
	MWS	< 0.001	2.045	4.004		NSP	0.026	0.042	0.055
	PRO	1.940	0.640	1.030		STA	35.060	9.950	23.710
Seed	SUC	0.000	2.710	5.360	Seed	GLUC	1.690	1.170	1.710
	FRUC	2.890	1.330	1.500		-	-	-	-
	SEFY	3.010	0.480	0.810		TSW	5623.60	839.40	8.000
	HIS	13.240	0.940	25.970		SNP	1.033	0.125	0.410
	SL	1.281	0.181	0.186		SW	0.271	0.053	0.164
Pod	SH	1.436	0.071	0.083	Pod	-	-	-	-
	PODY	21.030	1.320	6.310		SHEL	8.130	0.000	3.040
	TM	471.760	43.100	28.500		PL	314.300	39.600	45.990
	PW	2.446	0.633	0.513		POH	3.193	0.350	0.447
	PN	216.570	56.790	60.560		PDS	0.016	0.003	< 0.001
Flower	PBL	26.520	2.600	2.390	Flower	PBC	0.002	0.011	< 0.001
	BF	94.710	16.630	7.900		TF	170.780	28.810	7.990
	PF	52.310	214.310	0.630		IL	477.200	46.600	16.980
	TE	1.030	0.000	3.400		EV	7.220	24.930	10.870
	SC	5.530	98.920	15.280		PH	25604.0	1484.4	2585.7
Leaf	TLL	40.720	0.530	2.620	Leaf	TLW	16.960	1.420	2.840
	TLMW	9.850	0.000	1.000		LLL	29.090	0.070	2.029
	LLW	4.490	0.650	0.850		LLMW	5.730	0.000	0.300
	LN	15.810	46.430	146.840		TLLN	3.027	4.496	< 0.001
	LLLN	1.437	2.132	< 0.001		-	-	-	-
Composite traits (Root + Stem + Pod + Leaf)	BIOM	53.590	9.510	26.290	Composite traits (Root + Stem + Pod + Leaf)	VLW	12.620	4.560	18.660
	HITOT	22.260	54.440	32.490		-	-	-	-

Table 5 Variance component estimations of species (σ_S^2), entries within species ($\sigma_{A(S)}^2$) and the error term [(σ_e^2) comprising genotype-location interactions and plot errors] for 25 morphological qualitative traits measured in 34 *Pachyrhizus* entries

Variable set	Trait	Variance component estimates			Variable set	Trait	Variance component estimates		
		σ_S^2	$\sigma_{A(S)}^2$	σ_e^2			σ_S^2	$\sigma_{A(S)}^2$	σ_e^2
Storage root	DSN	1.810	0.000	1.520	Storage root	DSI	1.750	0.000	1.500
	SS	0.508	2.521	0.121		SDS	0.598	0.960	< 0.001
	SFC	1.424	2.091	< 0.001		DSFC	2.667	4.100	< 0.001
	SSR	0.176	0.292	< 0.001		SCR	0.945	1.550	< 0.001
	CSR	0.072	0.110	< 0.001		–	–	–	–
Seed	CSE	0.872	1.274	< 0.001	–	–	–	–	
Pod	PC	0.056	0.094	< 0.001	Pod	DP	0.379	0.570	< 0.001
	CMP	0.065	0.082	< 0.001	–	–	–	–	
Flower	CS	0.119	0.181	< 0.001	Flower	FCS	0.124	0.190	< 0.001
	FCW	0.124	0.190	< 0.001	–	–	–	–	
Stem	PT	0.360	0.500	< 0.001	Stem	SCO	0.123	0.169	< 0.001
Leaf	LC	0.802	0.174	0.007	Leaf	SCTLL	0.605	0.918	< 0.001
	SCLLL	0.388	0.568	< 0.001		TLLT	1.280	1.870	< 0.001
	LLLT	0.304	0.504	< 0.001	–	–	–	–	
Composite traits (Root + Stem + Pod + Leaf)	DSLJ	0.130	0.150	0.060	Composite traits (Root + Stem + Pod + Leaf)	DSLJ	0.580	0.190	0.400

monomorphism ($H' = 0.00$) (Table 9). The average diversity index value within *P. ahipa* was 0.83 comparable to the observed mean in *P. erosus* (0.87) and *P. tuberosus* (0.86), where the highest values for Shannon–Weaver index ($H' = 1.00$) were noted in characters inflorescence length, pod length and storage root glucose content; and the lowest values were observed for pod beak degree and shape of curvature (0.47) in *P. erosus*, while in *P. tuberosus* the highest H' appeared for harvest index of storage root yield ($H' = 1$) and the lowest for storage glucose content ($H' = 0.61$) (Table 9). The Shannon–Weaver diversity index was in general high and over 0.80 for most of the traits evaluated. At the generic level, monomorphism in *P. ahipa* scored similarly for the trait pod beak curvature. Except character pod degree and shape of curvature ($H' = 0.09$), the diversity index was equal to or higher than 0.65. Across the 50 quantitative traits, Shannon–Weaver diversity index in the yam bean germplasm is quite similar and around 0.83 both within and among the three species studied (Table 9).

Diversity index estimates with the 25 qualitative characters indicated generally lower values compared

to quantitative attributes ($H' = 0.51$; 0.61; 0.46 and 0.51 within *P. ahipa*, *P. erosus*, *P. tuberosus* and the genus *Pachyrhizus*, respectively) (Table 10). Monomorphism ($H' = 0.00$) was observed for the traits: colour of storage root, surface defect of storage root, cracking of storage root, dehiscence of pod, mature pod colour, shape of central terminal leaflet lobe and shape of central lateral leaflet lobe inside *P. ahipa*. In *P. erosus*, only the trait mature pod colour exhibited the monomorphism. Within *P. tuberosus*, the characters: surface defect of storage root, cracking of storage root, mature pod colour, storage root colour, flower colour of petal, flower colour of wing and plant type were monomorphic ($H' = 0.00$) (Table 10), whereas among the three species, monomorphism was observed in seven traits. Diversity index ranged from 0.00 to 1.00 in *P. ahipa*; from 0.00–0.83 in *P. erosus*, and 0.00–0.93 in *P. tuberosus*. Amount of diversity within and between yam bean species using qualitative traits was lower than with quantitative characters (Tables 9, 10). However, *P. tuberosus* yielded higher H' values for most of the qualitative traits than *P. erosus* and *P. ahipa*, respectively.

Table 6 Percentage (%) of trait variation and significance levels for each quantitative morphological and agronomic trait considered in *Pachyrhizus*

Variable set	Trait	Percentage of variation	Variable set	Trait	Percentage of variation	Variable set	Trait	Percentage of variation
Storage root	SRFY	12.46***	Storage root	SRDM	10.88***	Storage root	SRDY	26.07**
	HIR	14.12**		LS	11.84***		WS	24.12***
	MWS	23.14***		NSP	49.35***		PRO	23.82***
	STA	9.45***		SUC	33.75*		GLUC	30.45***
Seed	FRUC	23.68**						
	SEFY	26.17***	Seed	TSW	2.77**	Seed	HIS	12.46*
	SNP	12.79***		SL	12.88**		SW	10.03*
	SH	8.91***						
Pod	PODY	19.45**	Pod	SHEL	17.51**	Pod	TM	24.19***
	PN	2.81***		PL	2.75**		PW	8.7***
	POH	15.49***		PDS	11.19*		PBL	46.39*
	PBC	18.91**						
Flower	BF	20.93***	Flower	TF	4.8**	Flower	PF	4.96**
	IL	6.57***						
	TE	18.29**		EV	14.76*	Stem	SC	8.57
	PH	6.95***						
Leaf	TLL	5.57***	Leaf	TLW	12.46***	Leaf	TLMW	26.75***
	LLL	14.1***		LLW	12.73***		LLMW	29.19***
	LN	5.11**		TLLN	31.03**		LLLN	39.4**
	BIOM	0.83*	Composite traits (Root + Stem + Seed + Pod + Leaf)	VLW	0.99**	Composite traits (Root + Stem + Seed + Pod + Leaf)	HITOT	0.81
Average over all traits								
		0.87***						

Table 7 Phenotypic variation within ($PV_{A(S)}$) and among (PV_S) the three cultivated yam bean species evaluated at two locations in West Africa for 50 morphological and agronomic quantitative traits measured in 34 *Pachyrhizus* entries

Variable set	Trait	PV_{AS} (%) ^a	PV_S (%) ^a	Variable set	Trait	PV_{AS} (%) ^a	PV_S (%) ^a
Storage root	SRFY	41.78	21.76	Storage root	SRDM	0.00	60.32
	SRDY	35.75	9.55		HIR	58.24	12.10
	LS	3.09	51.36		WS	43.52	38.75
	MWS	33.80	0.016		NSP	34.14	21.13
	PRO	17.73	53.74		STA	14.48	51.02
	SUC	33.58	0.00		GLUC	25.60	36.98
	FRUC	23.25	50.52		–	–	–
Seed	SEY	11.16	70.00	Seed	TSW	12.97	86.90
	HIS	2.341	32.97		SNP	7.97	65.88
	SL	10.98	77.73		SW	10.86	55.53
	SH	4.465	90.31		–	–	–
Pod	PODY	4.60	73.37	Pod	SHEL	0.00	72.78
	TM	7.93	86.82		PL	9.90	78.59
	PW	17.62	68.09		POH	8.77	80.02
	PN	17.00	64.85		PDS	15	80.00
	PBL	8.25	84.16		PBC	78.57	14.28
Flower	BF	13.94	79.42	Flower	PF	80.19	19.57
	TF	13.88	82.27		IL	8.61	88.24
Stem	TE	0	23.25	Stem	EV	57.94	16.78
	SC	82.61	4.62		PH	5.002	86.28
Leaf	TLL	1.208	92.82	Leaf	TLW	6.69	79.92
	TLMW	0	90.78		LLL	0.22	93.27
	LLW	10.85	74.95		LLMW	0.00	95.02
	LN	22.20	7.56		TLLN	59.75	40.23
	LLLN	59.72	40.25		–	–	–
Composite traits (Root + Stem + Pod + Leaf)	BIOM	10.64	59.95	Composite traits (Root + Stem + Pod + Leaf)	VLW	12.72	35.21
	HITOT	49.85	20.38		–	–	–

^a Phenotypic variation among genotypes across locations was calculated within ($PV_{A(S)}$) and between (PV_S) species for all traits in the interannual environmental changes

Principal component analysis

The first ten principal components of the analysis explained 90% of the total variation. The first, second, third, fourth and fifth principal component accounted for 39.3, 21.3, 8.3, 5.7 and 3.8% of the total variance, respectively. The first component was highly correlated with eight agronomic traits and more than eleven morphological characters (Tables 11, 12). The second component was mainly related to variation in three agronomic and seven morphological traits (Tables 11, 12). The third component was highly correlated with five agronomic

traits and only one morphological attribute, while the fifth principal component was significantly correlated only with two morphological traits (Tables 11, 12).

The plot of the scores of the first three principal components (Fig. 1) derived from the principal component analysis showed that the three species are clearly separated, and that within *P. tuberosus*, the ‘Chuin’ and ‘Ashipa’ cultivar groups can be distinguished on the basis of the third principal component. All *P. ahipa* entries had negative scores for the first and second principal components. Entries of *P. tuberosus* from all cultivar groups sensu lato had positive scores for the first and second principal

Table 8 Phenotypic variation within ($PV_{A(S)}$) and among (PV_S) the three cultivated yam bean species evaluated at two locations in West Africa for 25 morphological qualitative traits measured in 34 *Pachyrhizus* entries

Variable set	Trait	PV_{AS} (%) ^a	PV_S (%) ^a	Variable set	Trait	PV_{AS} (%) ^a	PV_S (%) ^a
Storage root	DSN	0.00	54.35	Storage root	DSI	0.00	53.84
	SS	80.03	16.12		SDS	61.57	38.35
	SFC	59.47	40.50		DSFC	60.57	39.40
	SSR	62.26	37.52		SCR	62.09	37.86
	CSR	60.11	39.34		–	–	–
Seed	CSE	59.33	40.61	–	–	–	
Pod	PC	62.25	37.08	Pod	DP	60.00	39.89
	CMP	55.40	43.91	–	–	–	
Flower	CS	60.13	39.53	Flower	FCS	60.31	39.36
	FCW	60.31	39.36	–	–	–	
Stem	PT	58.07	41.81	Stem	SCO	57.67	41.98
Leaf	LC	17.70	81.58	Leaf	SCTLL	60.23	39.70
	SCLLL	59.35	40.54	TLLT	59.34	40.62	
	LLLT	62.30	37.57	–	–	–	
Composite traits (Root + Stem + Pod + Leaf)	DSLJ	44.11	38.23	Composite traits (Root + Stem + Pod + Leaf)	DSLJ	16.24	49.57

^a Phenotypic variation among genotypes across locations was calculated within ($PV_{A(S)}$) and between (PV_S) species for all traits in the interannual environmental changes

components, whereas *P. erosus* entries showed positive scores for the first principal component, but negative scores for the second principal component. The ‘Ashipa’ entry of *P. tuberosus* was the only entry with a large negative score for the third principal component.

Cluster analysis

In general, the results of the cluster analysis (Fig. 2) were similar to those from the principal component analysis. Each species entry was usually clustered at the first fusion steps and the average Euclidian distance between entries within species was large (> 0.25). *Pachyrhizus erosus*, *P. ahipa* and the *P. tuberosus* of the ‘Chuin’ cultivar group formed three main groups. At the final fusion steps, the *P. tuberosus* ‘Chuin’ group was aggregated with the *P. erosus* group; this ‘*P. tuberosus* ‘Chuin’–*P. erosus*’ cluster was then merged with the *P. ahipa* group and, following this, with the *P. tuberosus* ‘Ashipa’ cultivar group. The *P. tuberosus* ‘Ashipa’ and ‘Chuin’ fell into two distinct clusters, and the average Euclidian distance between *P. tuberosus* ‘Chuin’ and *P. erosus* was smaller than the average Euclidian distance

between the two *P. tuberosus* types. Within each of the three species, a similar amount of diversity was observed and several subgroups could be identified. The cluster structure obtained for *P. erosus* only partly reflects the geographic origins of the entries (Table 1). Thus, some clusters were formed by entries with the same origin (e.g. EC040, EC041 and EC042 from Guatemala), while other clusters combined entries with different origins (e.g. ECKEW from Mexico and EC533 from Macau in Asia).

Only a few traits remained in the final multiple regression model among all agronomic and morphological traits with variance component estimations of $\sigma_{A(S)}^2/\sigma_e^2 > 2$ that entered the stepwise multiple regression analysis as regressor variables (principal components as dependent variables) (Table 13). Inflorescence length, pod degree and shape of curvature, and pod green colour showed a coefficient of determination (R^2) of 0.976 for the first principal component. Thousand seed weight, start of flowering, shape of central terminal and lateral leaflet lobe had a coefficient of determination (R^2) of 0.972 for the second principal component. The third, fourth and fifth principal components were determined to 50.1% by pod beak curvature, to 69.8% by pod beak

Table 9 Estimation of the standardized Shannon–Weaver diversity index (H') for 50 quantitative morphological and agronomic characters of yam bean among and within 3 species in the genus *Pachyrhizus*

Variable set	Trait	Species			Average for the genus
		<i>P. ahipa</i>	<i>P. erosus</i>	<i>P. tuberosus</i>	
Storage root	SRFY	0.76	0.92	0.89	0.83
	SRDM	0.96	0.85	0.87	0.87
	SRDY	0.60	0.83	0.94	0.66
	HIR	0.88	0.70	1.00	0.83
	LS	0.86	0.94	0.96	0.87
	WS	0.76	0.84	0.85	0.77
	MWS	0.67	0.95	0.94	0.93
	NSP	0.70	0.67	0.84	0.95
	PRO	0.63	0.69	0.67	0.65
	STA	0.72	0.86	0.83	0.89
	SUC	0.93	0.8	0.79	0.86
	GLUC	0.86	1.00	0.61	0.82
	FRUC	0.77	0.8	0.88	0.74
Seed	SEYY	0.99	0.99	0.84	0.97
	TSW	0.91	0.87	0.88	0.89
	HIS	0.87	0.85	0.96	0.87
	SNP	0.82	0.84	0.88	0.94
	SL	0.94	0.84	0.86	1.01
	SW	0.99	0.93	0.89	0.93
	SH	0.95	0.93	0.93	0.89
Pod	PODY	1.00	0.95	0.89	0.84
	SHEL	0.87	0.81	0.85	0.86
	TM	0.96	0.91	0.84	0.99
	PN	0.77	0.94	0.89	0.93
	PL	0.90	1.00	0.86	0.84
	PW	0.89	0.95	0.84	0.86
	POH	0.91	0.91	0.90	0.95
	PDS	0.37	0.48	0.82	0.09
	PBL	0.82	0.86	0.83	0.88
	PBC	0.00	0.47	0.64	0.00
Flower	BF	0.82	0.98	0.90	0.82
	TF	0.85	0.92	0.90	0.88
	PF	0.99	0.75	0.81	1.00
	IL	0.81	1.00	0.82	0.84
Stem	TE	0.86	0.95	0.8	0.84
	EV	0.91	0.91	0.82	0.94
	SC	0.80	0.91	0.73	0.74
	PH	0.66	0.87	0.93	0.77
Leaf	TLL	0.84	0.85	0.92	0.8
	TLW	0.78	0.97	0.83	0.78
	TLMW	0.76	0.99	0.89	0.75
	LLL	0.90	0.86	0.97	0.92

Table 9 continued

Variable set	Trait	Species			Average for the genus
		<i>P. ahipa</i>	<i>P. erosus</i>	<i>P. tuberosus</i>	
	LLW	0.90	0.87	0.97	0.91
	LLMW	0.89	0.98	0.82	0.88
	LN	0.92	0.92	0.96	0.96
	TLLN	0.94	0.88	0.65	0.98
	LLLN	0.90	0.87	0.63	0.88
Composite traits (Root + Stem + Seed + Pod + Leaf)	BIOM	0.82	0.83	0.89	0.86
	VLW	0.77	0.99	0.79	0.88
	HITOT	0.87	0.81	0.91	0.90
Average over all traits	–	0.83	0.87	0.86	0.83

curvature and width of storage roots, and to 37.1% by terminal leaflet lobe type, respectively. In total only 10 traits were significant at the 0.01 level for the first five principal components in the regression analysis. These principal components explained 78.7% of total variation observed in this study, and the 10 selected traits (i.e. IF, PDS, PC, TSW, BF, SCTLL, PBC, WS and TLLT) explained 69.6% of the total variation observed.

The global search for environmental conditions characterized by comparable temperature and rainfall parameters of the experimental sites, i.e. average annual temperature from 23 to 28 °C and precipitation from 400–1200 mm (including both rainfall and irrigation) for at least six consecutive months, i.e. the crop duration at experimental sites, revealed that similar conditions can be found in 85 countries around the world (Fig. 3). The relevant target set of yam bean production environments are mainly located in Central and South America, West, Central, and South East Africa, South East Asia and the Pacific. Albeit, the environmental survey did fail to identify the known yam bean production areas in Central Mexico, South India and South China.

Multivariate analysis of variance

The behaviour of the genotypes and species was the same regardless of the environment used, considering all the 50 quantitative and 25 qualitative variables simultaneously. The main effect of each factor (G or L) was then investigated separately. For the factor G, significant differences were noticed between all

genotypes or species. All entries react also in the same way since $P \leq 0.0001$. Furthermore, the other multivariate contrasts were significant (Table 14).

The MANOVA against all 50 quantitative and 25 qualitative traits measured variables revealed then significant Wilks' Lambda ($P \leq 0.0001$). They yielded also significant results with similar statistics (Pillai's trace, Hotelling-Lawley trace, and Roys' Max root) (Table 14). Wilks' Lambda was transformed as an F approximation. Strong significant differences were detected among all genotypes within as well as between species, which suggested the need for discriminant analysis for centroid comparison between groups. All parallel statistical tests resulted from MANOVA (Hotelling-Lawley, Pillai's trace and Roys Max root) were treated in the way as Wilks' Lambda for the F test. The tested morpho-agronomic and quality traits can be efficiently utilized in further breeding programs. MANOVAs were conducted for sources of variation Location, Genotype (G), Species (S), $G \times L$, and $S \times L$ in the full MANOVA, excepted for qualitative characters where little variability is exhibited within each species (Table 14). In agreement with the PCA and cluster analysis, MANOVA indicated that the main components of the total phenotypic variance were due to almost 28 characters of the 75 evaluated across two environments. Comparison of the 50 quantitative and 25 qualitative traits using MANOVA showed a significant difference between accessions of *Pachyrhizus* species at all taxonomic levels examined. The difference is based mainly on 19 quantitative traits and 9 qualitative characters which showed positive

Table 10 Estimation of the standardized Shannon–Weaver diversity index (H') for 25 qualitative morphological characters of yam bean among and within 3 species in the genus *Pachyrhizus*

Variable set	Trait	Species			Average for the genus
		<i>P. ahipa</i>	<i>P. erosus</i>	<i>P. tuberosus</i>	
Storage root	DSN	0.75	0.67	0.59	0.79
	DSI	0.89	0.80	0.81	0.80
	SS	0.83	0.61	0.86	0.75
	CSR	0.00	0.47	0.89	0.00
	SDS	0.00	0.51	0.00	0.00
	SFC	0.73	0.47	0.78	0.69
	DSFC	0.72	0.47	0.63	0.76
	SSR	0.68	0.47	0.93	0.59
	SCR	0.00	0.82	0.00	0.00
Seed	CSE	0.69	0.63	0.93	0.69
Pod	PC	1.00	0.59	0.72	1.01
	DP	0.00	0.43	0.81	0.00
	CMP	0.00	0.00	0.00	0.00
Flower	CS	0.73	0.64	0.00	0.74
	FCS	0.36	0.73	0.00	0.47
	FCW	0.38	0.74	0.00	0.46
Stem	PT	0.75	0.83	0.00	0.71
	SCO	0.74	0.47	0.77	0.75
Leaf	LC	0.87	0.75	0.00	0.87
	TLLT	0.61	0.73	0.00	0.61
	LLLT	0.54	0.72	0.00	0.67
	SCTLL	0.00	0.58	0.73	0.00
	SCLLL	0.00	0.46	0.58	0.00
Composite traits (Root + Stem + Pod + Leaf)	DSLJ	0.60	0.75	0.69	0.50
	DSLJF	0.86	0.69	0.65	0.86
Average over all traits	–	0.51	0.61	0.46	0.51

Eigenvectors related to genotype and species distinctiveness (Tables 2 and 3). The MANOVA applied to the 34 accessions exhibited significant difference between the 3 species pairwise ($F = 3.90$; $P \leq 0.0001$). This result relied mainly on the 19 aforementioned quantitative characters of the 75 measured. However, some differences appeared among the three species. In *Pachyrhizus ahipa*, 8 traits were sufficient to analyze variation (Zanklan, unpublished data), while 18 traits were necessary in *P. erosus* and *P. tuberosus* (Zanklan, unpublished data) as indicated in Tables 2 and 3. Differences between accessions inside a given species are as large as differences among species as described by

phenotypic variation estimates and Shannon–Weaver diversity index analyses (Tables 9, 10).

Discriminant function analysis

Tables 15–23 are provided in Electronic Supplementary Material Table ESM1.

Tables 15–17 present the discriminant function analysis (DFA) structure for the 75 morpho-agronomic characters, the 50 quantitative and the 25 qualitative traits effects respectively, and related statistics including eigenvalues, proportion of total discriminant power accounted for by the first five canonical functions as well as cumulative amount of

Table 11 Pearson correlation coefficients for the relationship between each of the first five principal components (PC) and each of 50 quantitative morphological and agronomic traits measured in 34 *Pachyrhizus* entries

Variable set	Trait	PC1	PC2	PC3	PC4	PC5
Storage root	SRFY	− 0.464*	− 0.201	− 0.506*	− 0.630**	− 0.172
	SRDM	− 0.086	− 0.826**	− 0.232	− 0.223	− 0.157
	SRDY	− 0.368	− 0.135	− 0.615**	− 0.603**	− 0.128
	HIR	− 0.421	− 0.168	− 0.696**	− 0.503*	− 0.105
	NSP	− 0.488*	− 0.077	− 0.063	− 0.292	− 0.429
	PRO	− 0.746**	− 0.015	− 0.050	− 0.252	− 0.155
	STA	− 0.142	− 0.692**	− 0.122	− 0.087	− 0.072
	SUC	− 0.219	− 0.164	− 0.200	− 0.165	− 0.135
	GLUC	− 0.397	− 0.516*	− 0.287	− 0.023	− 0.065
	FRUC	− 0.014	− 0.704**	− 0.267	− 0.082	− 0.094
	LS	− 0.807**	− 0.037	− 0.249	− 0.319	− 0.098
	WS	− 0.451*	− 0.382	− 0.461*	− 0.579**	− 0.118
	MWS	− 0.244	− 0.043	− 0.245	− 0.488*	− 0.033
	Seed	SEFY	− 0.795**	− 0.413	− 0.078	− 0.119
TSW		− 0.568**	− 0.726**	− 0.157	− 0.230	− 0.071
HIS		− 0.084	− 0.672**	− 0.071	− 0.488*	− 0.302
SNP		− 0.882**	− 0.017	− 0.045	− 0.014	− 0.122
SL		− 0.552**	− 0.703**	− 0.155	− 0.219	− 0.138
SW		− 0.265	− 0.675**	− 0.044	− 0.427	− 0.355
SH		− 0.912**	− 0.086	− 0.136	− 0.139	− 0.284
Pod	PODY	− 0.793**	− 0.452*	− 0.080	− 0.081	− 0.038
	SHEL	− 0.784**	− 0.473*	− 0.081	− 0.056	− 0.036
	TM	− 0.842**	− 0.395	− 0.271	− 0.126	− 0.028
	PL	− 0.184	− 0.840**	− 0.221	− 0.186	− 0.120
	PW	− 0.812**	− 0.337	− 0.148	− 0.112	− 0.043
	POH	− 0.892**	− 0.034	− 0.185	− 0.186	− 0.275
	PN	− 0.752**	− 0.454*	− 0.061	− 0.234	− 0.064
	PDS	− 0.939**	− 0.256	− 0.045	− 0.042	− 0.128
	PBL	− 0.958**	− 0.164	− 0.072	− 0.039	− 0.039
	PBC	− 0.053	− 0.223	− 0.686**	− 0.464*	− 0.027
Flower	BF	− 0.739**	− 0.516**	− 0.239	− 0.179	− 0.107
	TF	− 0.829**	− 0.418	− 0.216	− 0.116	− 0.124
	PF	− 0.092	− 0.429*	− 0.657**	− 0.413	− 0.262
	IL	− 0.953**	− 0.109	− 0.056	− 0.021	− 0.063
Stem	TE	− 0.646**	− 0.054	− 0.124	− 0.082	− 0.256
	EV	− 0.318	− 0.255	− 0.669**	− 0.329	− 0.056
	SC	− 0.331	− 0.055	− 0.033	− 0.026	− 0.207
Leaf	PH	− 0.919**	− 0.300	− 0.169	− 0.013	− 0.008
	TLL	− 0.606**	− 0.760**	0.130	− 0.063	− 0.012
	TLW	− 0.843**	− 0.423	− 0.144	− 0.130	− 0.134
	TLMW	− 0.584**	− 0.794**	− 0.001	− 0.011	− 0.021
	LLL	− 0.488*	− 0.842**	− 0.111	− 0.059	− 0.026
	LLW	− 0.910**	− 0.117	− 0.189	− 0.178	− 0.103

Table 11 continued

Variable set	Trait	PC1	PC2	PC3	PC4	PC5
	LLMW	– 0.675**	– 0.708**	– 0.065	– 0.035	– 0.087
	LN	– 0.354	– 0.387	– 0.332	– 0.086	0.219
	TLLN	– 0.607**	– 0.641**	– 0.088	– 0.089	– 0.059
	LLLN	– 0.668**	– 0.581**	– 0.047	– 0.119	– 0.060
Composite traits (Root + Stem + Pod + Leaf)	VLW	– 0.623**	– 0.337	– 0.587**	– 0.109	– 0.218
	BIOM	– 0.883**	– 0.010	– 0.226	– 0.160	– 0.134
	HITOT	– 0.510*	– 0.137	– 0.740**	– 0.324	– 0.029

*, ** Significant with $P \leq 0.01$ and 0.001 respectively

Table 12 Pearson correlation coefficients for the relationship between each of the first five principal components (PC) and each of 25 qualitative morphological and agronomic traits measured in 34 *Pachyrhizus* entries

Variable set	Trait	PC1	PC2	PC3	PC4	PC5
Storage root	DSN	– 0.861**	– 0.273	– 0.010	– 0.088	– 0.065
	DSI	– 0.862**	– 0.254	– 0.109	– 0.111	– 0.071
	SS	– 0.049	– 0.400	– 0.283	– 0.332	– 0.151
	CSR	– 0.573**	– 0.555**	– 0.233	– 0.331	– 0.006
	SDS	– 0.601**	– 0.180	– 0.307	– 0.235	– 0.131
	SFC	– 0.789**	– 0.417	– 0.228	– 0.188	– 0.072
	DSFC	– 0.579**	– 0.568**	– 0.322	– 0.013	– 0.035
	SSR	– 0.033	– 0.533**	– 0.536**	– 0.125	– 0.024
	SCR	– 0.579**	– 0.307	– 0.202	– 0.143	– 0.313
Seed	CSE	– 0.841**	– 0.331	– 0.148	– 0.185	– 0.211
Pod	PC	– 0.635**	– 0.063	– 0.344	– 0.030	– 0.186
	DP	– 0.688**	– 0.212	– 0.298	– 0.287	– 0.393
	CMP	– 0.817**	– 0.357	– 0.080	– 0.125	– 0.369
Flower	CS	– 0.304	– 0.722**	– 0.088	– 0.271	– 0.286
	FCS	– 0.479*	– 0.633**	– 0.129	– 0.153	– 0.368
	FCW	– 0.479*	– 0.633**	– 0.129	– 0.153	– 0.368
Stem	PT	– 0.875**	– 0.401	– 0.032	– 0.088	– 0.149
	SCO	– 0.301	– 0.828**	– 0.057	– 0.063	– 0.011
Leaf	LC	– 0.763**	– 0.416	– 0.092	– 0.110	– 0.040
	TLLT	– 0.505*	– 0.428	– 0.132	– 0.101	– 0.616**
	LLLT	– 0.423	– 0.485*	– 0.060	– 0.200	– 0.593**
	SCTLL	– 0.559**	– 0.562**	– 0.421	– 0.249	– 0.227
	SCLLL	– 0.514**	– 0.683**	– 0.431*	– 0.200	– 0.070
Composite traits (Root + Stem + Pod + Leaf)	DSLII	– 0.365	– 0.457**	– 0.467**	– 0.165	– 0.183
	DSLFI	– 0.678**	– 0.224	– 0.095	– 0.228	– 0.063

*, ** Significant with $P \leq 0.01$ and 0.001 respectively

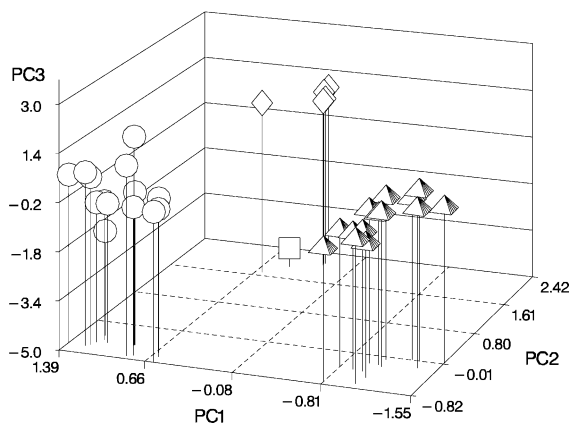


Fig. 1 Plot of the scores for the first (PC1), second (PC2) and third principal component (PC3) of the principal component analysis of 34 entries of three *Pachyrhizus* species determined from 75 morphological and agronomic traits; triangle = *P. ahipa*, circle = *P. erosus*, diamond = *P. tuberosus* ('Chuin'), square = *P. tuberosus* ('Ashipa')

discrimination power of functions. Tables 18–20 show results on DFA performed to dissect differences between entries at the interspecific scale.

Similarly, Tables 21–23 show standardized canonical coefficients against the first three functions for comparison of the performance of genotypes within and among the three yam bean species studied, again at both intra- and interspecific levels. Graphically, the mean scores of the first two canonical variates were scatter plotted to show visually differences between genotypes within and among species across two environments to indicate the original variables, which easily explained the variability. These are showed in Figs ESMs 1–5, respectively. Figs ESMs 6–8 show discrimination variability between the two locations used in the present study. Differences between genotypes were examined within and among the species.

Discriminant function analysis (DFA) carried out on the entire 75 morpho-agronomic characters scored with emphasis on traits recorded on different plant organs (Table 15) showed that the cumulative variance explained by the first five canonical variates accounted for 98.90% of the total variance with respect to both quantitative and qualitative storage root traits. The first two functions accounted for 96.77%. The variables which most contributed to those canonical variates were metrical as well as visual descriptor traits as shown in Table 21 listing

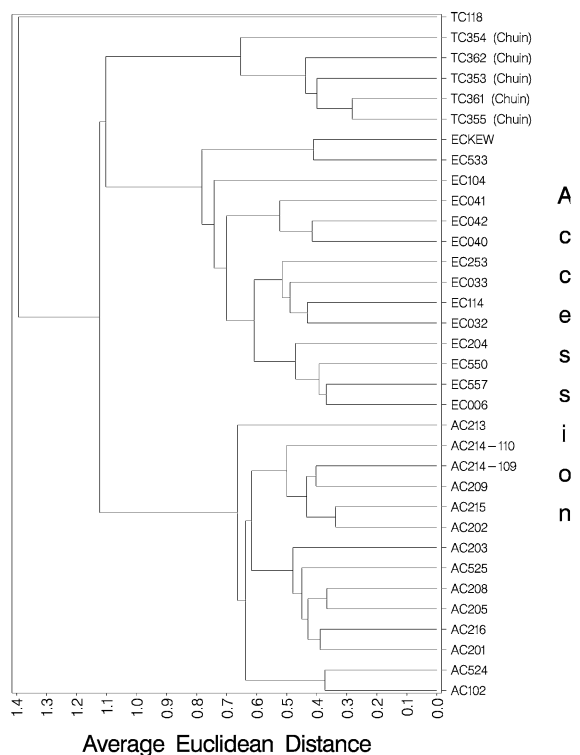


Fig. 2 Cluster analysis of 34 entries of three *Pachyrhizus* species based on 75 agronomic and morphological traits; TC = *P. tuberosus*, AC = *P. ahipa* and EC = *P. erosus*

the standardized canonical discriminant function coefficients between the first three canonical scores of discriminant ordinations and 75 morphological and agronomic traits in *Pachyrhizus* spp. The first discriminant function, which accounted for 94.96% of the total variance, was negatively correlated with 27 characters. Positive association with CAN1 was due to nine traits (Table 21). Examination of the second function suggested it was mainly associated with eight characters (with positive correlation) and negatively linked to twenty further traits (Table 21). The order in which the variables were included in the discriminant analysis indicates their relative importance in classifying entries.

Canonical analysis to find divergent trends for root related traits resulted in five main variates that accounted together for 98.90% of the total variation in *Pachyrhizus* (Table 15). The first and second variates contributed for 94.96 and 1.80% of the variation, respectively.

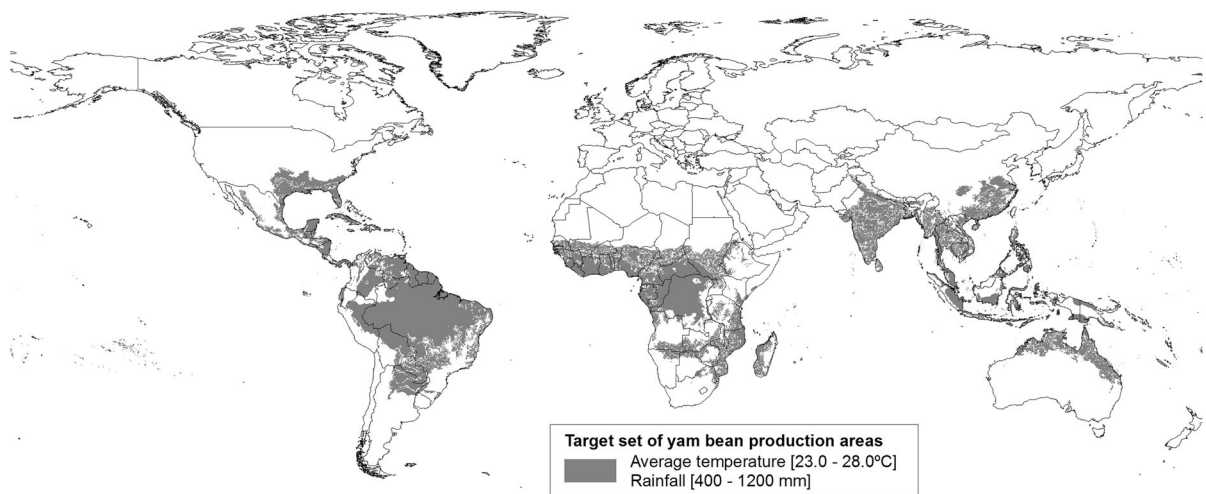
With regards to seed traits, the first five canonical variates extracted from DFA were responsible for

Table 13 Morphological and agronomic traits measured in 34 *Pachyrhizus* entries with variance component estimations of $\sigma_{A(S)}^2/\sigma_e^2 > 2$ entered and left in multiple regression analysis for variable reduction in future *Pachyrhizus* classificatory studies

Dependent variables	Regressor variables entered in the model	Regressor variables left in the model ^a	R ^{2b}
PC1	BF, TF, TSW, PT, LC, TLLT, SCTL, SCLL, TLLN, LLLN, IL, FCS, FCW, PC, PDS, DP, C MPCSE, CSR, SDS, SFC, DSFC, WS, SCR	IL, PDS, PC	0.9756
PC2	BF, TSW, DSLI, PF, SCTL, SCLL, TLLN, LLLN, CS, FCS, FCW, SCO, CSR, DSFC, SSR	TSW, SCLL, BF, SCTL	0.9718
PC3	DSLI, EV, PF, PBC, WS, SSR	PBC	0.5011
PC4	PBC, WS	PBC, WS	0.6976
PC5	TLLT, LLLT	TLLT	0.3711

^a All variables left in the model are significant at the 0.01 level

^b Coefficient of determination

**Fig. 3** Target set of yam bean production areas in the tropics, which correspond to the average annual temperature range (23–28 °C) and rainfall range [400–1200 mm (including irrigation)] at experimental sites for at least six consecutive months

99.98% of total variation (Table 15). Canonical loadings showed that CAN1 was determined and dominated by traits presented in Tables 15 and 21. The first variate represented 98.37% of the total variation explained by DFA and was highly correlated to most of the original variables aforementioned ($P < 0.0001$), but it was negatively correlated to SNP and SW. The second variate explained 1.16% of the seed parameter variation between genotypes and was negatively correlated to most variables.

DFA performed with focus on pod characters indicated that the first five variables demonstrated 99.79% of the total variation. CAN1 powered 90.93% and CAN2 explained 5.25% of the total variance. The original variables contributing to the variation observed are stressed in Tables 15 and 21.

Canonical analysis to identify genotypic differences for flower traits resulted in five variates that accounted together for 100% of the variability among genotypes inside a given species. The first canonical

Table 14 Wilks' λ , Hotelling-Lawley, Pillai's trace and Roy's Max root results from the MANOVA analysis of the 75 traits investigated for the sources of variation (Location, Species, Genotype) and their interactions

Sources of variation	Wilks' λ -test	F value	Hotelling-Lawley	F value	Pillai's trace	F value	Roy's Max root	F value
Location (L)	1.81E-33	5.94***	4968.79	38.27***	22.70	2.12***	2159.77	1874.52***
Species (S)	9.86E-06	232.80***	698.19	251.34***	1.99	215.54***	596.21	494.53***
Genotype (G)	6.17E-31	3.90***	3758.95	33.50 ^a ns	23.14	1.75***	2652.16	1980.28***
G \times L	7.07E-19	1.45***	1282.79	9.88***	16.49	0.92***	1081.86	938.97***
G (in L1) ^b	0.019	37.73***	50.53	37.73***	0.98	37.73***	50.53	37.73**
G (in L2) ^b	2.23E-20	4.32***	790.67	23.98***	17.62	1.92***	432.58	726.74***
S \times L	0.003	12.05***	1.880	8.592***	36.10	18.08***	24.77	18.08***
S (in L1)	0.003	31.41***	49.81	48.68***	1.81	19.99***	44.70	91.44***
S (in L2)	0.0001	134.10***	167.02	136.02***	1.97	131.28***	104.82	176.10***

For each analysis, Wilks' λ , Hotelling, Pillai's trace and Roy's Max root. F values and significance levels are included

*** Significant at the 0.001 probability level

^a Non significant

^b L1 = Songhai, L2 = Niaouli

variate described 94.96% of the total variation (Table 15). Distribution of genotypes through canonical axes 1 and 2 showed a conspicuous divergence between the groups formed by accessions within each species with high correlation to traits beginning of flowering and period of flowering (Tables 15, 21).

Stem traits showed big variation within as well among species. Canonical outcomes indicated that five variates explained 100% of the total variation observed. Original variables with CAN1 contributed jointly to 97.85% of the variation, while CAN2 was demonstrated only 1.73% of the total variation (Table 15). The first variate was negatively linked with stem colour, early vigour and plant height. CAN2 was caused mainly by time of emergence (Table 21).

All original leaf variables associated with the five first variates from DFA contributed to dissect 99.53% of the amount existing in the germplasm studied. The first two caused 66.16% of the total variability observed, whereas the second explained 30.70% showing among all other variable sets the higher percentage for variation attributed to the second canonical variate. Main traits causing the powers of CAN1 and 2 are labelled in Table 21.

Traits related to many plant organs together, considerable variation was also observed in the yam bean collection. Thus, the first variate explained most

of the total amount of variation (99.57%). The remaining variates were responsible for less than 1% (Table 15). The second variate and others explaining only additional variability discriminated accessions on the basis of weight of vines and leaves.

Using only the 50 quantitative or the 25 qualitative traits (Tables 16 and 17) indicated generally the same trends as previously described for the entire 75 characters investigated. As showed in Tables 22 and 23, all traits positively or negatively associated with CAN1 and CAN2 were highly variable and dispersion of genotypes within and between species differed significantly according to the plant organ considered.

Canonical ordinations of variables and genotypes along the first two significant axes presented also in general similar trends, as mentioned above, and their distribution was highly variable (Figs ESMs 1–3).

Among species, canonical analysis extracted five variates that were responsible for nearly 100% of total variation of the entire 75 quantitative and qualitative traits data with attention on storage root, seed, pod, flower, stem, leaf and composite characters (Tables 18–23). Distribution of whole species through canonical axes from DFA showed a conspicuous divergence between the groups formed by them. Their ordinations on the significant axes were attained by each assembling either all the 75

morpho-agronomic, or the 50 quantitative, or 25 qualitative characters; even by considering traits related to each plant organ separately (Figs ESMs 4 and 5). Species showed unequivocally distinct characteristics by appearing sometimes on both negative values for CAN1, or on high mean values for CAN1 and 2, or showing differences for diverse other combinations of variables along these first two canonical axes (Figs ESMs 4 and 5).

Even though species were well discriminated by DFA across both locations (Songhai and Niaouli), differences appeared in the importance of original variables and variable sets between environments as described above with the MANOVA results. As shown in Figs ESMs 6–8, DFA displayed diverse results for CAN1 and CAN2 in location Songhai or Niaouli. Moreover, these differences in the discriminatory of the CANs were linked to the characters related to varied plant organ of interest. Trait variation was displayed most inversely according to the location indicating high $S \times L$ interactions (Figs ESMs 6–8).

Effect size estimators and bias due to unequal sample size per species

In the study reported here, the total number of 34 entries was placed in three groups. The estimated bias due to unbalanced data error was computed and three effect size estimates were generated for all 50 quantitative and 25 qualitative traits evaluated, and summarized in Tables 24 and 25. In our experiment, it was noted that for forty-one traits, ϵ^2 and ω^2 estimates were negative as indicated in Tables 24 and 25 provided in Table ESM1.

Discussion

Phenotypic variation

The coefficient of variation provides a measure of diversity for quantitative traits (Ferguson and Robertson 1999). A part of the results on descriptive statistics, including 33 traits, has been reported earlier in Zanklan et al. (2007). Clear variability existed in the yam bean germplasm studied for all quantitative inherited traits with somewhat reduced variation in characters related to biomass production, pod number

per plant—a trait well associated to other agronomic traits—terminal leaflet length, inflorescence length, thousand seed weight, pod length, time of flowering, weight of vines and leaves, pod width, period of flowering, vine start of climbing and total harvest index. Trait variation ranged from 0.81 (for HITOT) to 49.35% (for number of storage root per plant) considering all traits and species evaluated. This indicated that genotypes belonging to all three cultivated yam bean species possess a high potential for biomass yield production and its components such as NSP, SEEY and SRDY. A remarkable diversity for most morphological and agronomic characters is shown at the intraspecific scale (Table 6), whereas overall diversity at the interspecific level was somewhat lower. Understanding the mechanisms making some sites confer more variability to the germplasm would be desirable to plan collecting missions, and to efficiently exploit the available genetic diversity in genebanks (Pecetti and Piano 2002). Our current results indicate that there is a wide differentiation among accessions and lines within and between the three cultivated yam bean species, both for quantitative and qualitative traits that can be used to breed for higher biomass, storage root dry matter content and yield.

The grouping of similar genotypes relies on the dissimilarity among them, which can be determined by a phenotypic diversity index (Upadhyaya et al. 2002). The average diversity index was similar in the three species, The Shannon–Weaver diversity index was calculated to compare phenotypic diversity index (H') among traits and within- and between-groups. A low H' indicates extremely unbalanced frequency classes for individual traits and a lack of diversity (Upadhyaya et al. 2002). Diversity estimates were performed for each trait and the three species as well as in the entire genus. H' was then pooled across traits, species and the genus *Pachyrhizus* (Tables 9 and 10). The average H' across traits was quite similar for species and the genus for quantitative characters. For qualitative traits, *P. erosus*, *P. tuberosus* and the genus presented the same trends, while in *P. ahipa*, the diversity was lower. The three species value of the H' index for each trait or averaged across all quantitative or qualitative characters was neither correlated with the number of constituent accessions and lines per species, nor with environmental differences. The significant variation at the interspecific as

well as the intraspecific scales suggests a differentiation of the species, likely related to the selective pressures in the environments of origin. Mean adjustment of adaptive traits takes place—in the long term—according to the prevailing environmental conditions of the location of origin (Pecetti and Piano 2002; Piano et al. 1996). In the present study, from the relative importance of among- and within-species variation (Tables 7, 8), such adjustment to a given environment may be realized. A relative high level of intraspecific variation, which is a primary factor of adaptation, can provide a buffering effect to the population to cope with the unpredictable, seasonal climatic fluctuations (Pecetti and Piano 2002).

Phenotypic variation estimated with Shannon–Weaver diversity index seemed to be equal in importance for quantitative as well as qualitative characters (Tables 9, 10) even though some differences were recorded between the two types of variables. This is indicating the equal usefulness of both variable types in studying genetic diversity in yam bean.

Døygard and Sørensen (1998) observed no single trait separation among the three species when analysing 18 quantitative morphological traits in herbarium material, i.e. excluding agro-ecological traits, using principal component analysis. However, our results, based on 75 morphological and agronomic characters (50 quantitative and 25 qualitative) and field trials, showed a clear separation between *P. erosus*, *P. ahipa*, *P. tuberosus* ‘Chuin’ and *P. tuberosus* ‘Ashipa’ in both the principal component analysis and the cluster analysis. A clear separation between *P. tuberosus* and *P. erosus* had previously been observed by Estrella et al. (1998), who used random amplified polymorphic DNA (RAPD) markers; however, that study did not consider *P. ahipa*. A clear separation between *P. tuberosus* ‘Chuin’ and *P. tuberosus* ‘Ashipa’ was also observed by both Tapia and Sørensen (2003) and Delêtre et al. (2017), the first using a canonical analysis based on 70 morphological and agronomic traits, and the second chloroplast DNA and microsatellite markers. Tapia and Sørensen (2003) found a large genetic distance in *P. tuberosus* between ‘Chuin’ and ‘Ashipa’. The same was observed in this study and additionally, that the genetic distance between the ‘Chuin’ and ‘Ashipa’ group within *P. tuberosus* is as large as the genetic

distance between ‘Chuin’ and *P. erosus* as well as between ‘Chuin’ and *P. ahipa*.

One limitation of the current study is the fact that we were only able to use one accession of the *P. tuberosus* of the cultivar group ‘Ashipa’ (TC118); however, this entry can be considered representative of the ‘Ashipa’ group (Tapia and Sørensen 2003). With five ‘Chuin’ entries we gave emphasis to the agronomically most important cultivar group within *P. tuberosus* due to the quality characteristics of the storage root. Together the *P. tuberosus* of the cultivar group ‘Ashipa’, entry TC118 from Haiti, and the five *P. tuberosus* of the cultivar group ‘Chuin’ from Peru, the clear differences concerning growth type and storage root nutrient content within *P. tuberosus* can be assessed in this study. The domestication of the *P. tuberosus* cultivar groups was studied by Delêtre et al. (2017) and the cultivar groups are still found in cultivation among many indigenous farmer communities in the Amazonas basin of Peru, Ecuador, Colombia, Brazil, Bolivia, and Venezuela, so that the existing ex situ germplasm collections of *P. tuberosus* maybe improved through renewed collection efforts. Such collection initiatives should be directly linked with the maintenance of the material in professionally managed genebanks. For this study it was only possible to obtain five ‘Chuin’ and three ‘Ashipa’ accessions and two of these ‘Ashipa’ samples could not be used because of lack of seed vigour. In contrast to the samples of *P. tuberosus*, the *P. ahipa* samples used in this study represent a very good coverage of the current diversity of this species.

Pachyrhizus ahipa had been considered extinct, because it had not been observed in the Andean fields in Peru and Ecuador (Grüneberg, pers comm.). However, in 1994/95 farmers in the remote valleys of Bolivia and Northern Argentina were found to be growing *P. ahipa*. This material was collected during two field trips of several months duration in 1994 and 1996 (Ørting et al. 1996). All the *P. ahipa* material used for this study—with the exception of AC102, AC524 and AC525—trace back to these two collection trips (Ørting et al. 1996) and were selected to cover the diversity between and within accessions. Most of this material is today available from CIP’s genebank (Table 1).

In contrast to *P. tuberosus* and *P. ahipa*, the species *P. erosus* can be easily found in cultivation—often on a commercial scale—in many tropical regions of the

world. Nevertheless, like *P. tuberosus* there is no professionally managed genebank with a mandate for *P. erosus* and germplasm acquisition depends upon research institutions and botanical gardens providing seed samples, excepted of “INIFAP genebank *Campo Experimental Bajio*” at Celaya, Guanajuato, Mexico. The *P. erosus* material used for this study represents a broad sample for this species from different regions of the world. The sample of yam bean genotypes were assessed in two field trials in Benin, West Africa, and both experimental sites had the same soil type (well-drained sandy red loam). Average annual temperature at the experimental sites ranges from 23 °C in August to 28 °C in May and the sites are characterized by usually dry conditions. However, using irrigation we generated a wide range of water supply in these field trials ranging from about 400 mm to 1200 mm during the crop growth period of about six months. Using ArcGIS software to search environments in the world with a temperature range from 23 °C to 28 °C and a precipitation range from 400 to 1200 mm within at least six consecutive months we found such conditions present in 85 countries worldwide, mainly in Central and South America, West, Central, and South East Africa, South East Asia and the Pacific (Fig. 3). This relevant target set of yam bean production environments corresponds well with the known yam bean distribution, with the exception of the cultivation areas in Central Mexico, South India and South China. However, it was observed that in this environmental survey all areas of origin of the three species were included in the output. This might explain why all three yam bean species were performing well at our experimental sites, although they originate from different eco-geographic regions. The two experimental sites used in this study are representative of a very wide range of yam bean production environments in West Africa as well as in other tropical regions around the world.

Principal component and cluster analyses

The average Euclidian distance observed between *P. tuberosus* group ‘Chuin’ and *P. tuberosus* group ‘Ashipa’ was larger than the average Euclidian distance between *P. tuberosus* group ‘Chuin’ and *P. erosus* as well as that between *P. tuberosus* group ‘Chuin’ and *P. ahipa*. This is a clear indication that

the three species are very closely related. That the relationships between the three cultivated yam bean species are very close is further indicated by the fact that the species are not, with one exception, separated by crossing barriers (Grum 1990; Sørensen 1991; Grüneberg et al. 2003). The diversity observed within *P. ahipa* and the *P. tuberosus* of the cultivar group ‘Chuin’ type (Figs. 1, 2) was nearly as large as the diversity observed within *P. erosus*. This was unexpected; because all *P. tuberosus* of the cultivar group ‘Chuin’ entries originate from Peru and all *P. ahipa* entries from Bolivia, whereas the *P. erosus* entries originated from several different countries in Central and South America and Asia. Moreover, within *P. erosus*, no clear subgroups related to geographic origin was detected which agrees with the results of Estrella et al. (1998). This suggests that in both Peru and Bolivia an extremely large amount of diversity can likely be found for cultivated yam beans. This also suggests that the yam bean was introduced to Asia from the Americas, which is in agreement with historical reports that *P. erosus* was introduced into Asia in the sixteenth century by the Spaniards via the Acapulco—Manila trade route (Sørensen 1996).

The first three principal components explained about 70% of the total variation in all traits (Tables 4, 5). This figure is high considering the large number of variables recorded in this study. Principal component analysis showed that, for the first principal component, all *P. erosus* entries had positive scores and all *P. ahipa* entries had negative scores (Fig. 1). This reflects the fact that *P. erosus* matures later, has a generally higher yield potential than *P. ahipa* and can morphologically be clearly separated from *P. ahipa*, because the first principal component was mainly associated with yield-related traits. With regard to yield potential and time to maturity, *P. tuberosus*, fell between *P. erosus* and *P. ahipa* (Zanklan et al. 2007). This was reflected by the positive scores obtained for the first principal component in *P. tuberosus*, which were always lower than those obtained for *P. erosus* and higher than those obtained for *P. ahipa*. However, *P. tuberosus* was mainly separated from *P. erosus* and *P. ahipa* by the high positive values obtained for the second principal component, which were associated with leaf and seed morphological characteristics, high storage-root dry matter and starch content and low storage-root fructose and glucose content as well as leaf length and shape.

In total, 75 characters were used in the present study. Recording such a large number of traits is both labour-intensive and time-consuming. This study shows that a character discard can be performed in describing cultivated yam bean genotypes, because many characters were highly correlated. The variation of the entire primary yam bean gene pool considered by 75 agronomic and morphological traits can be determined by a few traits, which express low genotype-environment interactions and plot errors (Tables 4, 5, 13). Only seven highly heritable characters—namely inflorescence length, legume angle and shape of curvature, pod green colour, thousand seed weight, start of flowering, outline of terminal central lobe and lateral leaflet lobe—explain a large part of the total pattern of variation observed by the first two principal components. With three further characters—namely the curvature of the persistent style at the tip of the legume, width of storage roots, and terminal leaflet lobe type—nearly 70% of the genetic diversity may be determined in yam beans. Our results will be useful to identify characters that could be used as descriptors for cultivated yam bean accessions, genotypes and varieties. Nevertheless, the development of a descriptor list would require a greater number of accessions including new genotypes. However, to obtain a large diversity of cultivated yam beans for studies is difficult, because no professionally managed genebank has the genus *Pachyrhizus* as a mandate. Only for *P. ahipa* is there a clear mandate at CIP within the frame of the mandate for Andean Root and Tuber Crops. However, this collection has started to include a few *P. erosus* and *P. tuberosus* accessions that might be of interest for agronomy and breeding (Table 1).

Multivariate analysis of variance

Our results showed morpho-agronomic and quality trait heterogeneity within the collection of *Pachyrhizus* evaluated. With MANOVA, differences among the three species were significant (Table 14). The power of MANOVA declines with an increase in the number of response variable (Scheiner 2001). Unequal sample sizes are not a large problem for MANOVA, but may bias the results for factorial or nested designs (Chahouki 2011).

To take into account these observations, we performed two sets of statistical analyses, either with

all 75 characters or the quantitative and qualitative separately. Our results showed no significant difference with the methodology, and no weakness of analysis was observed. Differences between single genotypes and species are based mainly on 21 traits (Tables 2, 3).

In all analyses involving the factor G (Genotype) and S (Species) including their interactions with other sources of variation, phenotypic variance was distributed across all eigenvectors. In the full MANOVA, the primary root accounted for about 43.41% of the variance (source G). Across locations, the primary root accounted for 43.50% (source G) (at location Songhai) and 43.47% (at location Niaouli) of variance associated with G. Applying MANOVA to tomato trial, Lounsbury et al. (2016) reported minor rank changes among genotypes across different locations. These findings are consistent with the results presented here in yam bean. Our results indicate that genotypes of yam bean with favourable phenotypic trait expression in terms of yield (storage root, seed and fodder), and related other morpho-agronomic characters exist and simple morphs characterized by BIOM, SRDM, SRDY, HIR, HIS, SEEY, TSW, BF, PF, TF, PH, PT, LS, WS, MWS, PRO, STA, SUC, GLUC, FRUC and high tolerance and resistance towards abiotic and biotic stresses could be selected for further breeding activities.

Discriminant function analysis

To classify accessions, a discriminant function analysis (DFA) was conducted using the entire set of 75 morpho-agronomic traits including 50 quantitative and 25 qualitative characters. Variables which have relatively high positive regression weights on a variate are positively inter-correlated as a group. Similarly, those having high negative weights are also positively inter-correlated, but negatively with those showing positive weights. The magnitude of the weights indicates the relative contribution of the original variables to each canonical variate. The total amount of variability were explained by two to five canonical variates considering the factor levels species, genotypes, location, and their interactions. DFA revealed a clear separation between genotypes within and among species. The discriminant function analysis based on the entire 75 traits explored and either on the 50 quantitative or 25 qualitative

characters identified correctly nearly 100% of *P. ahipa*, *P. erosus*, and *P. tuberosus* with an overall average minor error (Tables 18–20, as presented in Table ESM1).

The two dimensional plots (Figs ESMs 4 and 5) obtained from the first two variates indicated the formation of three distinct groups represented by each of the species. We found no earlier morphometric studies in yam beans conducted under field conditions into that scale for comparison with the data analyzed in this report. Floral traits exhibited clear differences among species, but narrow variation within them. Similar observations were made for leaf, stem and pod characters. Among and within species floral variability was confirmed by DFA completing MANOVA results. Species were clearly separated into distinct groups by first canonical variate with almost all the variable sets investigated as shown in all graphics presented. Tables 9, 10, 11, 12, 13 and 14 show the relative contribution of each morpho-agronomic trait involved in the discriminant functions for genetic dissimilarity. Results indicate that all traits were with greater contribution to the genetic diversity. DFA results demonstrates further that genotypes varied in their phenotype dependent upon the environment, and the magnitude of that variation was very diverse relying on the trait, indicating the need for further researches on stability analysis upon the most important agronomic traits since the results presented here are highlighting significant $G \times L$ interactions.

Results from DFA in combination with those from MANOVA were more useful and powerful statistical tools than simple ANOVA, because considering variables in combination. With the DFA following the MANOVA, the complex interrelationships among dependent traits could not only be revealed, but could also be taken into account in statistical inference, which is not done in a simple ANOVA.

General assessment and comparison of multivariate analysis techniques

Investigation on intra- and interspecific variation in yam bean for as many traits as in this study using MVAs has not been undertaken to date. However, Tapia and Sørensen (2003) reported a high level of diversity in a germplasm collection of *P. tuberosus* using hierarchical grouping (Mahalanobis distances)

and Duncan test. The present study indicates the existence of genetic diversity with superior characteristics that could be used in diverse breeding programs. The results presented here demonstrate clearly the congruity between the patterns of morpho-agronomic and quality characters along with genetic variation among the three *Pachyrhizus* species. All MVAs (PCA, cluster and regression analyses, MANOVA, and DFA) clearly separated the three species from one another. Using 50 quantitative and 25 qualitative traits with application of MVAs are in accordance with earlier findings indicating clear divergence between all yam bean genotypes investigated (Zanklan et al. 2007; Tapia and Sørensen 2003). Many genotypes possess high root and seed yields with quality suitable for diverse nutritional purposes. The information on diversity provides breeders with the ability to develop desirable types having high yield as well as better nutritional profiles. The reduction in the number of variables makes it easy to characterize and evaluate the performance of *Pachyrhizus* genotypes. Surprisingly, only one canonical variable accounted for about 90% of the total variation in the interspecific level apart from the results obtained on leaf traits (Tables 15, 17–20, as shown in Table ESM1), in contrast to the intraspecific scale, where more canonical variates were involved to elucidate the total variation. A good visualization of discrimination between species and accessions within and among the taxa is presented in scatter plots (Figs ESMs 1–5). The MANOVA and DFA were important in the study of morpho-agronomic and quality characteristics, allowing the simultaneous analysis of the most important attributes. Moreover, they facilitated the distinction of genotypes regardless of their taxonomic origin. Utilization of the multivariate techniques is therefore recommended in further studies in yam bean breeding.

Results from PCA, cluster and regression analyses, MANOVA, and DFA indicated that all those techniques have excellent predictive power for distinguishing among yam bean genotypes whichever taxa they are belonging to. However, we cannot definitely conclude that one method is better than the other since a judgment of these classification methods depends on the completeness of the data and the objectives of the study. With our results DFA was slightly more powerful than the others methods in classifying and discriminating *Pachyrhizus* genotypes

on the basis of their morpho-agronomic and quality traits. Results from DFA also showed a range of possibilities to use diverse types of traits to discriminate between genotypes.

Compared to PCA (76.4%), the discriminant function analysis accounted for nearly 100% of the within and among variance when considering five axes. The discriminant analysis identified more clearly a number of traits to be used in future studies. A combination of all techniques would be most appropriate for describing the variation in yam bean germplasm and to design a collection strategy.

Conclusion

The study allows a better knowledge of the cultivated yam bean germplasm collection. Morpho-agronomic characterization using MVAs demonstrated significant intra- and interspecific variation and indicates significant differences between *Pachyrhizus* species for all individual or grouped traits. The statistical analysis was useful in identifying the most divergent variables within yam bean species and can be helpful in the future to advance progress in breeding programs.

In conclusion, the study's results demonstrate that within each cultivated species a similar amount of diversity may be found and that the genetic distance between species is limited. Moreover, considerable diversity may exist within *P. ahipa* and *P. tuberosus* grown at both sides of the Andean mountain range. Since interspecific hybridisation is possible (Grum 1990; Sørensen 1991; Grüneberg et al. 2003), all three cultivated yam bean species may constitute an important source for breeding. The close relationship among species further supports the proposition that only a few highly heritable characters are required to describe the diversity within the yam bean gene pool. The list of these traits may serve breeders and curators in germplasm management, acquisition and distribution.

Acknowledgements This research was supported by a scholarship from the German Academic Exchange Service (DAAD), which was awarded to the first author. We would like to acknowledge the facilities provided by the Centre Songhai in Porto-Novo and the INRAB (Institut National des Recherches Agricoles du Bénin) in Niaouli during the fieldwork undertaken

in Benin. Thanks are also due to Bo Ørting and Friedrich Kopsisch-Obuch, for their valuable comments.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Acquaah G (2007) Principles of plant genetics and breeding, 1st edn. Backwell Publishing Ltd, New York, p 584p
- Agaba R, Tukamuhabwa P, Rubaihayo P, Tumwegamire S, Ssenyonjo S, Mwanga ROM, Ndirigwe J, Grüneberg WJ (2016) Genetic variability for yield and nutritional quality in yam bean (*Pachyrhizus* sp.). HortScience 51(9):1079–1086
- Agaba R, Tukamuhabwa P, Rubaihayo P, Mwanga ROM, Ssenyonjo A, Ndirigwe J, Tumwegamire S, Grüneberg W (2017) Heritability, combining ability and inheritance of storage root dry matter in yam beans. Afr Crop Sci J 25 (1):83–95
- Al Khanjari S, Filatenko AA, Hammer K, Buerkert A (2008) Morphological spike diversity of Omani wheat. Genet Resour Crop Evol 55:1185–1195
- Arms EM, Lounsbury JK, Bloom AJ, St. Clair DA (2016) Complex relationships among water use efficiency-related traits, yield, and maturity in tomato lines subjected to deficit irrigation in the field. Crop Sci 56:1698–1710
- Badu-Apraku B, Akinwale RO, Ajala SO, Menkir A, Fakorede MAB, Oyekunle M (2011) Relationships among traits of tropical early maize cultivars in contrasting environments. Agron J 103:717–729
- Becker HC (2011) Pflanzenzüchtung. UTB Verlag, Stuttgart
- Boehringer (1983) Methodenbuch der enzymatischen Lebensmittelanalytik. Comp. Boehringer, Mannheim
- Bruneau A, Doyle JJ, Doyle JL (1994) Phylogenetic relationships in Phaseoleae: evidence from chloroplast DNA rearrangements in reconstructing plant phylogeny. In: Crips M, Doyle JJ (eds) Advances in legume systematics. Royal Botanic Gardens, Kew, pp 309–330
- Chahouki MAZ (2011) Multivariate analysis techniques in environmental science. In: Dar IA 2011, environmental sciences—ISBN 978-953-307-468, p 630
- da Silva ES, da Silva Filho DF, Ticona-Benavente CA (2016) Diversity of yam bean (*Pachyrhizus* spp. Fabaceae) based on morphoagronomic traits in the Brazilian Amazon. Acta Amazon 46(3):233–240

- Delêtre M, Soengas B, Vidaurre PJ, Meneses RI, Delgado Vásquez O, Oré Balbín I, Santayana M, Heider B, Sørensen M (2017) Ecotypic differentiation under farmers' selection: molecular insights into the domestication of *Pachyrhizus* Rich. ex DC. (Fabaceae) in the Peruvian Andes. *Evol Appl* 10(5):498–513. doi:[10.1111/eva.12472](https://doi.org/10.1111/eva.12472)
- Døyggaard S, Sørensen M (1998) Principal component analysis of morphological characters in the genus *Pachyrhizus* DC. In: Proceedings of 2nd international symposium on Tuberous Legumes. Celaya, Guanajuato—Mexico, 5–8 August 1996. MacKeenzie Press, Copenhagen (Denmark), pp 33–42
- Estrella JS, Phillips R, Abbott Gillies A, Sørensen M (1998) Genetic variation and relationships in agronomically important species of yam bean (*Pachyrhizus* DC.) based on RAPD markers. In: Proceedings of 2nd international symposium on Tuberous Legumes. Celaya, Guanajuato—Mexico, 5–8 August 1996. MacKeenzie Press, Copenhagen (Denmark), pp 43–59
- Falconer DS (1989) Introduction to quantitative genetics, 3rd edn. Longman, Harlow
- Ferguson ME, Robertson LD (1999) Morphological and phonological variation in the wild relatives of lentil. *Genet Resour Crop Evol* 46:3–12
- Francisco M, Velasco P, Lema M, Cartea ME (2011) Genotypic and environmental effects on agronomic and nutritional value of *Brassica rapa*. *Agron J* 103:735–742. doi:[10.2134/agronj2010.0439](https://doi.org/10.2134/agronj2010.0439)
- Grum M (1990) Breeding *Pachyrhizus* Rich. ex DC. A review of goals and methods. PhD Thesis, Department of Crop Husbandry and Plant Breeding, The Royal Veterinary and Agricultural University, Copenhagen, Denmark
- Grüneberg WJ, Büttner G, Delgado-Licon E (1998) Protein and starch quality of yam bean tubers. In: de Ron AM (ed) EUCARPIA international symposium on breeding of protein and oil crops, 1–4 April 1998, Pontevedra, Spain. MBG-CSIC, Pontevedra, Spain, pp 95–97
- Grüneberg WJ, Goffman FD, Velasco L (1999) Characterization of yam bean (*Pachyrhizus* spp.) seeds as potential sources of high palmitic acid oil. *JAOCs* 76:1309–1312
- Grüneberg WJ, Freynhagen-Leopold P, Delgado-Vázquez O (2003) A new yam bean (*Pachyrhizus* spp.) interspecific hybrid. *Genet Resour Crop Evol* 50:757–766
- Hailu F, Merker A, Singh H, Belay G, Johansson E (2006) Multivariate analysis of diversity of tetraploid wheat germplasm from Ethiopia. *Genet Resour Crop Evol* 53:1089–1098
- Heredia-Zepeda A, Heredia-García E (1994) “San Miguelito”, “San Juan” y “Vega de San Juan” nuevas variedades de jicama (*Pachyrhizus erosus* (L.) Urban) para el Bajío, Guanajuato, México. In: Sørensen M (ed) Proceedings of the first international symposium on Tuberous Legumes; Guadeloupe, F.W.I., 21–24 April 1992. Jordbrugsforlaget, Copenhagen (Denmark), pp 257–272
- Hill J, Becker HC, Tigerstedt PMA (1998) Quantitative and ecological aspects of plant breeding. Chapman and Hall, Boca Raton, p 275
- Hughes AR, Stachowicz JJ, Williams SL (2009) Morphological and physiological variation among seagrass (*Zostera marina*) genotypes. *Oecologia* 159:725–733
- IBPGR (International Board for Plant Genetic Resources) (1985) *Phaseolus acutifolius* descriptors. IBPGR Secretariat, Rome
- IBPGR (International Board for Plant Genetic Resources) (1987) Descriptors for bambara groundnut (*Vigna subterranea*). IBPGR Secretariat, Rome
- IBPGR (International Board for Plant Genetic Resources) (1991) Descriptors for sweet potato (*Ipomoea batatas*). IBPGR Secretariat, Rome
- ICC (1999) Standard methods of the International Association for Cereal Science and Technology. ICC Vienna
- Ingham JL (1990) Systematic aspects of phytoalexin formation within tribe Phaseoleae of the Leguminosae (subfamily Papilionoideae). *Biochem Syst Ecol* 18:320–343
- Jain SK, Qualset CO, Bhatt GM, Wu KK (1975) Geographical patterns of phenotypic diversity in a world collection of durum wheat. *Crop Sci* 15:700–704
- Jaradat AA, Weyers SL (2011) Statistical modelling of yield and variance instability in conventional and organic cropping systems. *Agron J* 103:673–684
- Kumar PS, Hymowitz T (1987) Where are the diploid ($2n = 2x = 20$) genome donors of *Glycine* Willd. (Leguminosae, Papilionoideae)? *Euphytica* 40(3):221–226
- Lackey JA (1977) A revised classification of the tribe Phaseoleae (Leguminosae: Papilionadeae), and its relation to Canavanine distribution. *J Linn Soc Bot* 74:163–178
- Lautié E, Rozet E, Hubert P, Quetin-Leclercq J (2012) Quantification of rotenone in seeds of different species of yam bean (*Pachyrhizus* spp.) by a SPE HPLC-UV method. *Food Chem* 131:1531–1538
- Lázaro-Nogal A, Matesanz S, Hallik L, Krasnova A, Traveset A, Valladares F (2015) Population differentiation in a Mediterranean relict shrub: the potential role of local adaptation for coping with climate change. *Oecologia*. doi:[10.1007/s00442-015-3514-0](https://doi.org/10.1007/s00442-015-3514-0)
- Lee J, Hymowitz T (2001) A molecular phylogenetic study of the subtribe Glycinae (Leguminosae) derived from the chloroplast DNA rps16 intron sequences. *Am J Bot* 88(11):2064–2073
- Lounsbury J, Arms EM, Bloom AJ, St. Clair DA (2016) QTL for water stress tolerance traits localize on chromosome 9 of wild tomato *Solanum habrochaites*. *Crop Sci*. doi:[10.2135/cropsci2015.07.0432](https://doi.org/10.2135/cropsci2015.07.0432)
- Mohammadi SA, Prasanna BM (2003) Analysis of genetic diversity in crop plants—salient statistical tools and considerations. *Crop Sci* 43:1235–1248
- Molosiwa OO, Gwafila C, Makore J, Chite M (2016) Phenotypic variation in cowpea (*Vigna unguiculata* [L.] Walp.) germplasm collection from Botswana. *Int J Biodivers Conserv* 8(7):153–163
- Ndirigwe J, Rubaihayo P, Tukamuhabwa P, Agaba R, Rukundo P, Mwanga OMR, Tumwegamire S, Kamarirwa V, Kayinamura E, Grüneberg WJ (2017) Evaluation of performance of introduced yam bean (*Pachyrhizus* spp.) in three agro-ecological zones of Rwanda. *Trop Plant Biol*. doi:[10.1007/s12042-017-9188-5](https://doi.org/10.1007/s12042-017-9188-5)
- Nicholson P (1995) Using SAS/GRAPH software to plot dendrograms. *Observ Tech J SAS Softw Users*. 5(1):4–17 ed. SAS Inst. Inc., Cary, NC, USA

- NRC (1979) Tropical legumes. Resources for the future. National Academy of Science, Washington, DC, USA, NRC, National Research Council
- Oré-Balbin I, Sørensen M, Kvist LP, Delgado Vasquez O (2007) Review of the *Pachyrhizus tuberosus* (Lam.) Spreng. Cultivar groups in Peru. Plant Genet Resour Newsl 151:1–13
- Ørting B, Grüneberg WJ, Sørensen M (1996) Ahipa (*Pachyrhizus ahipa* (Wedd.) Parodi) in Bolivia. Genet Resour Crop Evol 43:435–446
- Pecetti L, Piano E (2002) Variation of morphological and adaptative traits in subterranean clover populations from Sardinia (Italy). Genet Resour Crop Evol 49:189–197
- Piano E, Pecetti L, Carroni AM (1996) Climatic adaptation in subterranean clover populations. Euphytica 92:39–44
- Polhill RM (1994) Classification of the Leguminosae and complete synopsis of legume genera. In Bisby FA, Buckingham J, Harborne JB (eds) Phytochemical dictionary of the Leguminosae. Plants and their constituents, vol 1. Chapman and Hall, London, pp 35–57
- Ramos-de-la-Peña AM, Renard CMGC, Wicker L, Contreras-Esquivel JC (2013) Advances and perspectives of *Pachyrhizus* spp. in food science and biotechnology. Trends Food Sci Tech 29:44–54
- Safari P, Honarnejad R, Esfehiani M (2008) Assessment of genetic variation in peanuts (*Arachis hypogaea* L.) cultivars using canonical discriminant analysis. Iran J Field Crops Res 6:327–334
- Santayana M, Rossel G, Núñez J, Sørensen M, Delêtre M, Robles R, Fernández V, Grüneberg WJ, Heider B (2014) Molecular characterization of cultivated species of the genus *Pachyrhizus* Rich. ex DC. by AFLP markers: calling for more data. Trop Plant Biol 7:121–132
- SAS (1997) SAS user's guide: statistics, version 6.2, SAS-Institute, Cary, NC, USA
- SAS Institute (2007). JMP User Guide, Release 7. SAS Institute Inc., Cary, NC, USA. ISBN 978-1-59994-408-1
- Scheiner SM (2001) Theories, hypotheses, and statistics. In: Scheiner SM, Gurevitch J (eds) Design and analysis of ecological experiments. Oxford University Press, Oxford, pp 3–13
- Shannon CE, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urbana
- Skidmore ST, Thompson B (2013) Bias and precision of some classical ANOVA effect sizes when assumptions are violated. Behav Res 45:536–546
- Sørensen M (1988) A taxonomic revision of the genus *Pachyrhizus* (Fabaceae—Phaseoleae). Nord J Bot 8:167–192
- Sørensen M (1990) Observations on distribution, ecology and cultivation of the tuber-bearing legume genus *Pachyrhizus* Rich. ex DC. (Fabaceae: Phaseoleae). Wagening Pap 90 (3):1–38
- Sørensen M (1991) Cross breeding experiments: compatibility, pollen fertility, and germination percentages of interspecific hybrids, F₁, F₂ & F₃ in the genus *Pachyrhizus* Rich. ex DC. (Fabaceae: Phaseolae). pp 597–624 in Caribbean Food Crops Society, Twenty-Fifth Annual Meeting, Guadeloupe, Dossier, 1–6 July 1989, Vol 25
- Sørensen M (1996) Yam bean (*Pachyrhizus* DC.) Promoting the conservation and use of underutilized and neglected crops. 2 (Heller, J. series ed.) Institute of Plant Genetics and Crop Plant Research, Gatersleben/International Plant Genetic Resources Institute, Rome, p 141
- Sørensen M, Døyggaard S, Estrella JE, Kvist LP, Nielsen PE (1997a) Status of the South American tuberous legume *Pachyrhizus tuberosus* (Lam.) Spreng. Biodivers Conserv 6:1581–1625
- Sørensen M, Grüneberg WJ, Ørting B (1997b) Ahipa (*Pachyrhizus ahipa* (Wedd.) Parodi). In: Hermann M, Heller J (eds) Andean roots and tubers. Promoting the conservation and use of underutilized and neglected crops. 21. Institute of Plant Genetics and Crop Plant Research, Gatersleben/International Plant Genetic Resources Institute, Rome, pp 13–73
- Sweeney RA, Rexroad PR (1987) Comparison of LECO FP-228 “Nitrogen Determinator” with AOAC copper catalyst Kjeldahl method for crude protein. JAOAC 70:1028–1030
- Tapia C, Sørensen M (2003) Morphological characterization of the genetic variation existing in a neotropical collection of yam bean, *Pachyrhizus tuberosus* (Lam.) Spreng. Genet Resour Crop Evol 50:681–692
- Tessmann G (1930) Die Indianer Nordost-Perus. Grundlegende Forschungen für eine systematische Kulturkunde. Friedrichsen, de Gruyter & Co, Hamburg
- Ukalska J, Mądry W, Ukalski K, Masny A, Żurawicz E (2006) Patterns of variation and correlation among traits in a strawberry germplasm collection (*Fragaria* × *ananassa* Duch.). J Fruit Ornament Plant Res 14:5–22
- Upadhyaya HD, Ortiz R, Bramel PJ, Singh S (2002) Phenotypic diversity for morphological and agronomic characteristics in chickpea core collection. Euphytica 123:333–342
- Van Hintum TJJ, Brown AHD, Spillane C, Hodgkin T (2000) Core collections of plant genetic resources. IPGRI Technical Bulletin No. 3. International Plant Genetic Resources Institute. Rome, Italy
- Vaylay R, Van Santen E (2002) Application of canonical discriminant analysis for the assessment of genetic variation in tall fescue. Crop Sci 42:534–539
- Wricke G (1962) Über eine Methode zur Erfassung der ökologischen Streubreite in Feldversuchen. Z Pflanzenzüchtung 47:92–96
- Wricke G, Weber WE (1986) Quantitative genetics and selection in plant breeding. Walter de Gruyter Verlag, Berlin
- Yeater KM, Bollero GA, Bullock DG, Rayburn AL, Rodriguez-Zas S (2004) Assessment of genetic variation in hairy vetch using canonical discriminant analysis. Crop Sci 44:185–189. doi:10.2135/cropsci2004.0185
- Zanklan AS, Ahouangonou S, Becker HC, Pawelzik E, Grüneberg WJ (2007) Evaluation of the storage root forming legume yam bean (*Pachyrhizus* spp.) under West African Conditions. Crop Sci 47(5):1934–1946
- Zeba N, Isbat M (2011) Multivariate analysis for yield and yield contributing traits in F₀ and F₁ generations in tobacco (*Nicotiana tabacum*). J Exp Biosci 2:101–106
- Zhu D (1990) Phenotypic and genotypic variation in Kiwifruit (*Actinidia deliciosa* (A. Chev.) C.F. Liang et A.R. Ferguson) seedling populations. PhD Dissertation, Massey University, USA, p 187