

INTERNATIONAL JOURNAL OF THE FACULTY OF AGRICULTURE AND BIOLOGY,
WARSAW UNIVERSITY OF LIFE SCIENCES – SGGW, POLAND

REGULAR ARTICLE**The integrated phenotype and plasticity of Cuphea PSR23:
a semi-domesticated oilseed crop****Abdullah A. Jaradat**

United States Department of Agriculture (USDA)-Agricultural Research Service (ARS) 803 Iowa
Avenue, Morris, MN 56267, USA.

E-mail: abdullah.jaradat@ars.usda.gov

CITATION: Jaradat A.A. (2016). The integrated phenotype and plasticity of Cuphea PSR23: a semi-domesticated oilseed crop. *Communications in Biometry and Crop Science* 11, 10–30

Received: 22 August 2015, Accepted: 9 November 2015, Published online: 21 December 2015

© CBCS 2016

ABSTRACT

Cuphea PSR23, a semi-domesticated potential oilseed crop, is a selection from an interspecific cross between the wild species *Cuphea lanceolata* and *C. viscosissima*. Understanding the extent to which its phenotype is integrated, by studying complex trait interactions and interdependencies, is critical for its full domestication, as well as for advancing our knowledge of its developmental plasticity, and adaptation to new environments and alternative management practices. Phenotypic plasticity is necessary for Cuphea's adaptation to changing environments and may become inversely related to phenotypic integration, especially under abiotic stress. Phenotypic integration was quantified using several multivariate statistical procedures on log- or z-transformed raw data, or on latent variables derived from phenotypic or nutrient variables within structural, metabolic and reproductive plant modules. Several estimates of the level of phenotypic integration were used to generate phenotypic integration indices (PIIs) as the deviation of functionally linked phenotypic, eco-physiological, and nutrient traits from the means of respective traits in its wild parents. These traits were assessed in Cuphea PSR23 in response to directional selection for high seed weight, seed yield, and oil content under cultivation. Under managed agro-ecosystems and in comparison with its wild parents, Cuphea PSR23 displayed complex phenotypic traits that are naturally highly dimensional; it invested more in phenotypic and eco-physiological traits that are related to larger biomass and to stronger inter-plant competition under high population density; and displayed wide variation in the scale and significance of bivariate and multivariate trait (dis)associations. However, Cuphea PSR23 may have become less phenotypically integrated due to directional population selection for a few agronomic traits under cultivation. Reduction in phenotypic integration if continued under cultivation, may render PSR23, as a semi-domesticated oilseed crop, more vulnerable to abiotic stresses, and may delay its full domestication unless systemic germplasm enhancement and breeding programs are launched for its genetic and agronomic improvement.

Key Words: *oilseed crop; integrated phenotype; plasticity; multivariate statistics.*

INTRODUCTION

Phenotypic integration (PI) in wild and crop plants is concerned with the study of the ecology and evolution of complex phenotypes (Pigliucchi 2003); it signifies the multidimensional relationships within and between multiple quantitative and qualitative traits across interrelated plant modules (e.g., structural, metabolic, and reproductive interrelated groups of functional traits) of a complex plant phenotype. The PI is also concerned with correlations (or co-variation) among traits within functional units as well as the phenotypic correlation structure of multiple traits (Chitwood and Topp 2015). Furthermore, the PI is a dynamic process, whereby interactions among traits within and among plant modules may change during plant ontogeny and across environments (Pigliucchi 2003, Adhikaru and Wallace 2014). On the other hand, morphological integration in a crop plant, which is a specific aspect of phenotypic integration, indicates interdependences among morphological traits to produce an organized functional plant (Ferreira et al. 2013). Understanding these complex interactions and interdependencies is critical to advancing our knowledge of developmental plasticity, adaptation of crop plants to new environments, and to new or alternative management practices (Tetsana et al. 2014), as well as for crop domestication (Ladizinsky 1998). In the latter case, developmental plasticity (i.e., when different phenotypes of a single trait are expressed in response to the environment) is crucial for the survival and productivity of newly-domesticated crops such as Noug (*Guizotia abyssinica*) (Dempewolf et al. 2015), Evening Primrose (*Oenothera* spp.) (Vilela et al. 2008), Lesquerella (*L. fendleri*) (Ploschuk et al. 2005), and *Cuphea* PSR23.

Plant phenotyping, defined as the process of capturing or recording a set of quantitative and qualitative traits, has been the backbone of most agronomic, ecological, and eco-physiological studies to explore plant functional diversity, compare the performance of species and crop cultivars, or study plant responses to abiotic stresses (Granier and Vile 2014). Most pioneering phenotypic studies have been performed and statistically analyzed separately to answer specific questions; however, there is sufficient, yet untapped information in databases that needs to be integrated in multivariate and meta-analyses (Chen et al. 2014, Milla et al. 2014). Therefore, the dynamic variation in plant phenotypes and their plasticities (caused by the environment or due to ontogeny) should be included in such analyses (Tetsana et al. 2014).

Plants, especially those undergoing domestication, reveal complex phenotypic traits that are naturally extremely highly dimensional (Chen et al. 2014, Chitwood and Topp 2015); therefore, increasing the number of phenotypic measurements (e.g., by image feature extraction) is an important goal in phenomics and phenotypic integration studies. Large numbers of measurements on individual plants, but not on populations, are needed to accurately estimate phenotypic variability due to typically large sampling errors on trait variances, and when phenotypic data is heteroscedastic (Geiler-Samerotte et al. 2013). Consequently, a powerful and appropriate experimental design is required that enables the separation of multiple factors (i.e., fixed and random factors, and their interactions) that may affect phenotypic variance.

Ecological literature suggests that breeding and selection for specific traits, especially under agricultural management (i.e., high-resource managed agro-ecosystems), might lead to reduced phenotypic integration (Milla et al. 2014). Under managed systems, crop plants invest more in traits involved in competition for light and other renewable and non-renewable inputs; and under higher population density, crop plants become larger as compared to their wild progenitors or wild relatives. However, crop plants may become less phenotypically integrated, especially at the later stages of crop improvement (Ciaccio et al. 2014, Milla et al. 2014). Reduction in phenotypic integration may affect the possibility of improving newly domesticated or semi-domesticated crops (e.g., *Cuphea* PSR23), to

withstand biotic and abiotic stresses (Pinet et al. 2009, Bloomfield et al. 2014). Unlike traits in plants with poorly-integrated phenotypes, plant traits with high phenotypic integration respond in a synchronized manner to changes in the environment, whether these changes are due to external inputs or to abiotic stresses (Milla et al. 2014).

Current methods of collecting phenotypic data are being enhanced by new statistical methods that may lead to higher capacity for collecting, collating and processing phenotypic data and the development of phenotypic integration models of crop plants (Chitwood and Topp 2015). More sophisticated statistical algorithms and procedures to discern the quantitative structure of a crop plant will be required as we approach measuring the totality of a plant "phenotype." Structural equation modelling (SEM; Lamb et al. 2011, Warton et al. 2012) together with model selection approaches, combine aspects of researcher-defined and data-driven perspectives required for investigating complex plant traits and how they are allocated to different, but interrelated modules. Most phenotypic studies involve fixed and random effects (i.e., mixed); in which case, statistical modelling can separate multiple sources of variation that contribute to phenotypic distributions (Geiler-Samerotte et al. 2013). The values of fixed effects are repeatable, determined by the experimenter, and of inherent interest (e.g., the genotype); whereas random effects take values that are sampled from a potentially infinite set of measurements (i.e., with measurement error). Other multivariate statistical analyses procedures aim at reducing the dimensionality of the phenotypic space while preserving the major axes of functional variation. These methods provide an efficient way to extract specific traits relevant for the discrimination of groups that can be further investigated using conventional statistical procedures. It is highly desirable to use a combination of multivariate methods, embedded in a pipeline of statistical analyses, in order to increase the discrimination efficiency of biologically meaningful variables (Granier and Vile 2014). However, interpreting the phenotype in a biologically expressive and meaningful manner remains as a challenge, even if the statistical and biometrical tools are available to capture its complex and multidimensional structure (Chitwood et al. 2015). The overall objective of this study was to apply several multivariate statistical analyses procedure in studying phenotypic integration of a semi-domesticated potential oilseed crop (Cuphea PSR23) and its wild parents (*C. lanceolata* and *C. viscosissima*); whereas, the specific objectives were to (1) quantify the level of phenotypic integration in all three Cuphea spp. and in Cuphea PSR23 in comparison with its wild parents, (2) quantitatively verify if Cuphea PSR23 did evolve high-resource use strategies during several planting-harvesting-seed selection cycles, and (3) quantify the temporal level of coordination and co-variation between resource-use phenotypic and eco-physiological traits in Cuphea PSR23 during cultivation.

MATERIALS AND METHODS

GERMPLASM

The germplasm material used in this study was characterized and evaluated during several (2003-2008) years of growth chamber, greenhouse and field studies. Information on germplasm sources and evaluation are available elsewhere (Jaradat 2012, Jaradat and Rinke 2014 and references therein). A brief description of inputs and measured (primary) or derived (secondary) variables is presented in Table 1 for the purpose of this study.

MEASUREMENTS

Phenotypic measurements on single mature plants from each of four replicates per entry, experiment, and year (3-5 single plants, depending on experiment, species and year) were classified into five major groups of latent variables (Chitwood et al. 2015) that maximized the separation of *a priori* categories (Table 1). These were plant structural, metabolic, and reproductive traits; and, capsule and seed physical, and seed chemical composition traits, including oil content. Basic statistics (including mean and standard deviation, SD; among others) were estimated for data collected for each trait, species,

experiment and year. The mean vs. SD plots were developed and examined as a diagnostic procedure of heteroscedasticity and to determine whether and to what extent variation depends on the mean, in which case appropriate data transformation was applied where needed (Zar 1996). Morphometric data, describing different morphotypes, was evaluated by Kruskal-Wallis test because it makes no assumptions in reference to variable distribution (StatSoft Inc. 2012).

Table 1. List of traits measured or estimated on single plants or replicated samples of *Cuphea* PSR23, *C. lanceolata*, and *C. viscosissima* germplasm.

Traits based on (tissue)	Measured Trait (Primary measured variables)	Secondary (derived) variables
Structural	Plant height (PH), Primary (PB) and secondary (SB) branches, Total Branch length (TBL); stem dry weight, root dry weight (Growth chamber experiment)	Plant size (volume), and Fractal dimension (Do), Relative Growth Rate (RGR).
Metabolic	Leaf length and width, Leaves/plant, leaf dry weight,	Leaf area/plant (LA), Specific leaf area (SLA), Leaf colour space (L^* , a^* , b^*). Relative Growth rate (RGR), Net Assimilation Rate (NAR).
Reproductive	Flowers/plant (FP), Capsules/plant (CP), Seed/capsule (SC), 1000-Seed weight (SWT),	Harvest Index (HI), Flower density/main branch, capsule density/main branch,
Capsule and seed physical traits	Weight, and dimensions (Length, width, perimeter, area, major axis, minor axis).	Seeds/capsule, packaging cost (capsule weight/seed). Capsule and Seed colour space (L^* , a^* , b^*).
Seed chemical composition	Oil content, protein content (estimated as N x 6.25), Macro- and micro-nutrients (Al, B, Ba, Ca, Cr, Cu, Fe, K, Mg, Mn, Na, P, S, Si, Se, Zn).	C:N, N:P, C:P ratios (in stems, leaves and seeds, [roots only in growth chamber experiment]).

Structural traits were captured using digital imagery of single plants, then converted into quantitative measurements using ImageJ software (Rasband 2015); then plant volume (cm^3) and plant fractal dimension (Do) were estimated on digital images using special modules or macros in ImageJ. Specific leaf area (SLA) was estimated as the ratio between leaf area and dry leaf weight. Net assimilation rate (NAR), which is related to net whole-plant carbon fixation ($\text{g cm}^{-2} \text{d}^{-1}$) given the equation RGR at time (t)

$$\text{NAR} = [\text{NAR}(t)] \times [\text{SLA}(t)] * [\text{LMR}(t)];$$

where RGR, NAR and SLA are as defined before, and LMR is leaf mass ratio (Shipley 2004). Data on SLA, RGR and NAR were z-transformed prior to statistical analyses (Urbaniak 2010). The remaining secondary statistics (Table 1) were calculated from raw data by simple mathematical formulae. Chemical analyses of structural, metabolic and reproductive tissues were carried out using LECO instrument for carbon and nitrogen estimation (Rayan 2013) and the Varian Vista-Pro CCD (charged coupled device) (Varian Inc.) Simultaneous ICP-OES (inductively coupled plasma-optical emission spectroscopy) instrument was used for macro- and micro-nutrients (Masson et al. 2010). The elemental standards “MNUSDA-STD 1A” and “MNUSDA-STD 2 for this analysis were prepared according to Inorganic Ventures, Lakewood, NJ.

In order to evaluate reproducibility among replicates, and to avoid introducing low quality or weak phenotypic traits into subsequent statistical analyses, each quantitative trait was transformed into three categories as follows. The mean and standard deviation (SD) calculated for each of 42 quantitative traits measured on plants or modules in growth chamber, greenhouse and field experiments during 2004-2006 (Table 1) were used to classify each trait into one of three discrete categories: low, medium, or high. If a trait value was \leq the mean - 1 SD, it was classified as “low”; if it was \geq the mean + 1 SD, it was classified as

“high”; otherwise it was classified as “medium” (Zar 1996). Then the number of “effective” traits (i.e., N_e in the analysis of genetic diversity) was calculated using POPGEN (Yeh et al. 2000). A total of 1,890 (or about 83.0%) of the original year-experiment-replicate-species-trait combinations showed high reproducibility after removing outliers (Chen et al. 2014).

Quantitative data on the final number of traits, after removing outliers, were z-transformed (mean of 1 and zero variance). The transformation satisfied assumption of univariate and multivariate analyses of variance; improved normality, and removed the effect of scale of measurements (Zar 1996). Phenotypic relationships among the final set of traits were explored by principal components analyses (PCA; Payne et al. 2007) in order to capture global phenotypic variation in the whole database and to extract specific phenotypic traits relevant for further multivariate statistical analyses.

PAIRWISE SIMILARITIES

Morphological or phenotypic pair-wise similarities among Cuphea PSR23 (averaged for each year; from 2004 to 2006 in field experiments and 2007-2008 in growth chamber experiments) and among all three Cuphea species (i.e., Cuphea PSR23, *C. lanceolata* and *C. viscosissima*; averaged for each year and species; 2003-2004 in field experiments) were quantified for each trait by the shared phenotypes index (Z_{ij}) as follows (Ferreira et al. 2013):

$$Z_{ij} = [(Y_i - U)(Y_j - U)]/V;$$

where Y_i and Y_j are the values of the quantitative trait Y , respectively, for two individuals ‘ i ’ and ‘ j ’, and U and V are, respectively, the sample mean and variance of Y in the whole sample as described above. The structure of the matrices of phenotypic similarities was compared between traits by means of Mantel tests with 10,000 permutations (Mantel 1967) and was performed by NTSYSpc v. 2.21o. (Rohlf 2011).

TRAIT RELATIVE SHIFTS OVER TIME

Relative change in plant trait relationships was estimated for Cuphea PSR23 on the basis of 42 traits integrated over field experiment (2004-2006) in comparison with its wild parents. The log-trait of oil%, seed weight, seeds/capsule, packaging cost and C:N ratio) or the first (PC1) and second (PC2) principal components of seed physical traits; capsule physical traits; and seed nutrients (other than C and N), were used in a comparative cluster analyses (Ward clustering method and Euclidean linkage distance). Data from field experiments during 2004, 2005 and 2006 on Cuphea PSR23, and from field experiment on wild Cuphea species (*C. lanceolata* and *C. viscosissima*), were used to estimate the level of integration measured by the Euclidean linkage distances between these variables. Another phenotypic integration in Cuphea PSR23 was estimated on the basis of the z-transformed fractal dimension, leaf specific area, oil%, structural, reproductive and metabolic C:N, N:P and C:P ratios, net assimilation rate (NAR), and relative growth rate (RGR) in each of three years (2004-2006). Mean values, standard error of the means and variances, as well as a Euclidean distance using the Unweighted Pair Group Method and Arithmetic Mean (UPGMA) clustering method. Common PC analysis (CPCA) was conducted on the variance-covariance matrices of each entry in order to detect analogous patterns of trait correlation. The statistical test is based on the Flury hierarchy of hypothesis tests (Flury 1988), a set of algorithms for determining which eigenvectors differ significantly between covariance matrices (Phillips and Arnold 1999). The objective of the test was to examine the functional links among multiple phenotypic and nutrient traits (listed in Table 1), and their assumed coordinated evolution during cultivation. The CPCA Analysis examines whether matrices share more complex relationships than expected based on their eigenvectors while being different in the eigenvalues associated with their principal components (Flury 1988, Ackerman and Cheverud 2000). The maximum number of partial CPCs used in this analyses was six; this is a conservative number of PCs which was based on a t-test indicating that no significant amount of variation was captured beyond the 5th PC (Payne et al. 2007).

LATENT VARIABLES

Structural equation modelling (SEM) was used to examine ‘hidden’ functional linkages between multiple traits and their presumed coordinated development during domestication under cultivation (Crisci 2012, Zhang et al. 2014). Furthermore, SEM also provides additional empirical support for the general validity of *a priori* inter-trait relationships (Lamb et al. 2011); these relationships (Fig. 1) were conceptualized on the basis of prior knowledge of PCR23 and its wild parents (Jaradat 2012, Jaradat and Rinke 2014, and references therein), and further based on eco-physiological and biological theory (Martre et al. 2015). The degree of fit between the observed and expected covariance structures in SEM was tested by a χ^2 goodness-of-fit test. A non-significant goodness-of-fit test indicates that the model does fit the data globally (Lamb et al. 2011, Crisci 2012).

PHENOTYPIC INTEGRATION INDICES

The log-transformed trait values (Log-trait) and PCA eigenvalues (i.e., latent variables) for groups of trait (listed in Table 1) were used in estimating a matrix of phenotypic integration as the difference between average values of the wild species from that of *Cuphea* PSR23 (i.e., the difference for a particular trait between *Cuphea*_{PSR23} and *Wild*_{Trait}). These differences represented trait shifts as a result of evolution under cultivation of *Cuphea* PSR23 (Milla et al. 2014).

Two matrices were developed on the basis of Spearman correlation coefficients, the first was based on “among-traits” and PCA eigenvalues as an indicator of phenotypic integration; while the second was based on the differences in log-traits or PCA eigenvalues between PSR23 and its wild parents; this matrix was used as an indicator of changes in phenotypic integration during *Cuphea* PSR23 evolution under cultivation. Finally, a phenotypic integration index (PII), adjusted for plant size, and a relative phenotypic integration index, adjusted for plant size (RelPII_{sc}) were estimated using PHENIX, an R package to calculate confidence intervals, and a method to simulate a null distribution and test the statistical significance of phenotypic integration (Torices et al. 2015). The procedure and the resulting PII are modifications of the classical index developed by Wagner (1984). The PII quantifies the overall level of correlation of a multivariate distribution as an estimation of functional relationships between phenotypic traits.

STATISTICAL SOFTWARE

Relevant modules in several statistical analyses packages were used in performing basic and advanced univariate, bivariate and multivariate statistical analyses; these included The Unscrambler (CAMO 2011) to build and validate principal components; CPCA (Flury 1988, Phillips and Arnold 1999) test the significance of shared or common principal components between species; NTSYSpc v.2.21o (Rohlf 2011) to carry out and test the significance of matrix correlations (Mantel 1967); and GenStat v. 10 (Payne et al. 2007), and STATISTICA v. 12 integrated with the R package v. 3.1.2. (StatSoft, Inc. 2013, R Core Team, 2015) to carry out several univariate, bivariate and multivariate statistical tests; PHNIX (Torices and Muñoz-Pajares 2015) to estimate phenotypic integration indices; and ImageJ (Rasband 2015) to perform digital imagery statistics.

RESULTS

The traits measured or estimated on replicates of single plants or samples of *Cuphea* PSR23, *C. lanceolata*, and *C. viscosissima* germplasm (Table 1) covered all life history traits, as well as secondary (or derived) traits which captured the complex multivariate nature of both *Cuphea* PSR23 and its wild parents. In addition, nutrient ratios, estimated in structural, metabolic, and reproductive plant organs (or tissues), served as indicators of nutrient stoichiometry. Singly, and in combinations, these traits were used in describing the temporal

evolution of, and to what extent the Cuphea PSR23 phenotype was integrated under cultivation, and in comparison with its wild parents.

A conceptual *a priori* model for Cuphea PSR23 plant phenotype (Fig. 1) was constructed on the basis of prior experience and phenotypic analysis. Direct and indirect effects on the plant phenotype were hypothesized on theoretical basis, and were meant to serve as a prototype for further multivariate statistical analyses and modelling. The plant volume and fractal dimension both summarize the structural components of the plant as the outcome of genotype x environment interaction; the specific leaf area (SLA), net assimilation rate (NAR), and relative growth rate (RGR) are eco-physiological traits that summarize input acquisition and its conversion rate into biomass through the metabolic component of the phenotype; while the reproductive components (i.e., flowers, seeds and capsules) is the final expression of resource use efficiency as reproductive effort (i.e., seed and oil yield). The three-way interactions between these components, and their effects on biomass and on seed size, nutrients, and oil% define, to a large extent, the final phenotype.

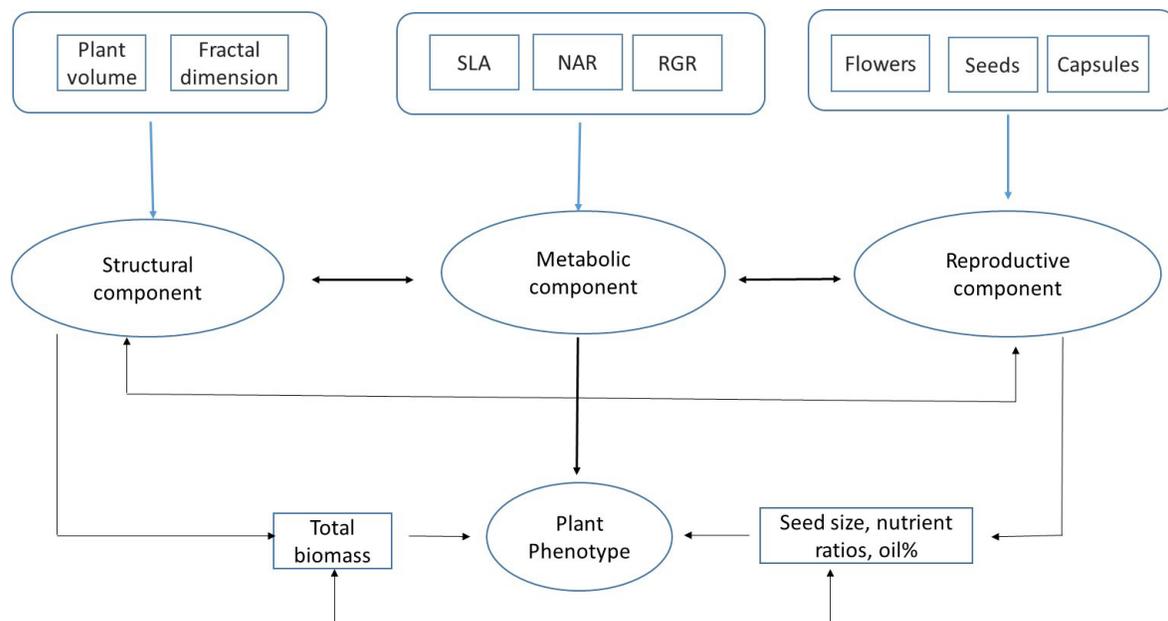


Figure 1. Conceptual *a priori* model for a Cuphea PSR23 plant phenotype (SLA, specific leaf area; NAR, net assimilation ratio; RGR, relative growth rate).

When all five groups of traits (i.e., structural; metabolic; reproductive; capsule, and seed physical traits; and seed chemical composition traits) were subjected to principal components analyses (PCA), the first validation PC explained 45.0, 52.0, and 57.0% of total variation in Cuphea PSR23, *C. lanceolata*, and *C. viscosissima*, respectively (Table 2). The best common principal components solution for the number of shared PCs (Table 2, above diagonal), based on Akaike Information Criterion (AIC) suggested that Cuphea PSR23 shared three and four, out of five maximum CPCs with *C. lanceolat* and *C. viscosissima*, respectively; whereas, both wild species shared three PCs in common. The largest correlation coefficient (Table 2, below diagonal) was between the standardized PC1 of Cuphea PSR23 and that of *C. viscosissima* (0.83; $p < 0.0001$), followed by the one with *C. lanceolata* (0.75; $p < 0.01$); while the correlation between both wild species was the smallest (0.62; $p < 0.001$). Four eco-physiological traits describing plant architecture (i.e., Do, the fractal dimension), and yield determining traits (i.e., SLA, RGR and NAR; Table 3) exhibited medium levels of variation (C.V., 15.3-26.8%), except NAR (49.0-66.4%).

Table 2. Percent validation variance explained by the first principal component in *Cuphea* PSR23, *C. lanceolata* and *C. viscosissima*; and best common principal components analyses (CPC) solutions based on AIC test for comparison of covariance matrices between PSR23 and each of its wild parents (*C. lanceolata* and *C. viscosissima*) using Flury's hierarchy tests (Flury, 1988) on five principal components derived from plant structural, metabolic, and reproductive traits; seed and capsule physical traits, and seed chemical composition traits (above diagonal); and correlation coefficients between the first standardized principal components in each *Cuphea* spp. (PC1, below diagonal).

Species	<i>Cuphea</i> PSR23	<i>C. lanceolata</i>	<i>C. viscosissima</i>
Percent validation variance Q ² , PC1	45.0**†	52.0**	57.0**
		<u>CPC analysis</u>	
<i>Cuphea</i> PSR23	Equality	CPC(3)	CPC(4)
<i>C. lanceolata</i>	0.75**	Equality	CPC(3)
<i>C. viscosissima</i>	0.83***	0.62**	Equality

†; **, ***: significant at the 1 and 0.1% level of probability, respectively.

Table 3. Basic statistics (mean and standard deviation) and bivariate correlation coefficients (above diagonal) between fractal dimension (Do, 1.0 to 2.0) and three eco-physiological traits (specific leaf area [SLA, cm²g⁻¹], relative growth rate [mg g⁻¹d⁻¹], and net assimilation rate [NAR, mg cm⁻²day⁻¹]) in *Cuphea* PSR23 and its wild parents (*C. lanceolata* and *C. viscosissima*).

Species	Traits	Mean	C.V.%	Specific leaf area (SLA)	Relative growth rate (RGR)	Net assimilation rate (NAR)
<i>C. lanceolata</i>	Do	1.538b†	16.6	0.95**§	0.31ns	-0.56*
	SLA	269.5a	16.4		0.42*	-0.66**
	RGR	119.6c	18.2			0.81**
	NAR	1.25a	49.6			
<i>C. viscosissima</i>	Do	1.618ab	15.7	0.89**	-0.21ns	-0.44*
	SLA	218.9b	17.2		0.42*	-0.63**
	RGR	148.83b	15.3			0.89**
	NAR	0.89b	49.0			
<i>Cuphea</i> PSR23	Do	1.79a	29.8	0.79**	0.16ns	-0.32ns
	SLA	259.9a	24.0		0.51*	-0.82**
	RGR	188.65a	26.8			0.93**
	NAR	1.25a	66.4			

†; A trait followed by the same letter do not differ significantly (p<0.05) between species.

§; ns, *, **: not significant, significant at the 5 and 1% level of probability, respectively.

Table 4. Correlations between correlation matrices for *Cuphea* PSR23 in each of 5 years of field experiments. Raw matrix correlations (above diagonal), adjusted correlations (below diagonal), and matrix repeatability (**in bold**) on the diagonal. Each calculation was based on annual population correlation matrix of phenotypic and nutrient traits listed in Table 1 (p< 0.001).

	Year1	Year 2	Year 3	Year 4	Year 5
Year 1	0.87	0.74	0.65	0.67	0.58
Year 2	0.82	0.83	0.85	0.79	0.71
Year 3	0.72	0.76	0.89	0.84	0.72
Year 4	0.69	0.64	0.75	0.91	0.82
Year 5	0.61	0.64	0.78	0.74	0.86

Cuphea PSR23 generally had significantly larger values of these traits than its wild parents. However, relationships between these traits followed the same pattern. The exception was a non-significant relationship between Do and NAR in Cuphea PSR23 ($r = -0.32$, $p > 0.05$). The fractal dimension was strongly ($p < 0.01$) correlated with SLA, but not with RGR, and negatively and significantly with NAR in *C. lanceolata* and *C. viscosissima*. As expected, RGR was positively correlated with NAR ($p < 0.01$) and with SLA ($p < 0.05$); while SLA was negatively correlated with NAR ($p < 0.01$) in all three species. Relationships between correlation matrices (Table 4) illustrate multivariate correlations between the same set of traits for five years of field experimentation, after being permuted 10,000 times. The adjusted correlation coefficients (below diagonal) were slightly improved when compared with the raw coefficients (above diagonal). The raw and adjusted matrix correlation coefficients were highly significant for all pair-wise combinations; however, a gradual decrease was observed in their numerical values over time; the drop in adjusted matrix correlation was largest from 0.82 in Year 1 to 0.61 in Year 5, and smallest between Year 4 and Year 5 (0.78 to 0.74). Matrix repeatability was generally large and ranged from 0.83 to 0.91 (bold, diagonal) with no obvious temporal trend.

The same set of data was used in performing canonical discriminant analysis among germplasm produced in each of five years of field experimentation (Table 5). A decreasing trend in percent correct classification was observed as the time lag between years increased (83.5 to 66.0% correct classification in years 1 and 5, respectively; diagonal, in bold). Percent misclassification of germplasm decreased as the time lag between years increased from 1 year (e.g., 15.6% of germplasm in Year 1 was misclassified as belonging to Year 2) to four years (e.g., 4.8% of germplasm in Year 1 was classified as belonging to Year 5). The statistical tests (i.e., Mahalanobis Squared Distances between population centroids; D^2 , below diagonal) were all significant, suggesting that germplasm population from each year differed at the multivariate level from the remaining germplasm populations, in spite of the less-than-perfect annual correct classification. An increasing temporal trend was found in D^2 , with the smallest distance between Years 1 and 2 ($D^2 = 45.9$), and the largest between Years 2 and 5 ($D^2 = 64.9$).

Table 5. Percent correct classification of germplasm in each of five years (diagonal, in bold), percent misclassification of germplasm from each year as belonging to the preceding year (e.g., germplasm entries in Year 3 being classified as germplasm from Year 1 and Year 2; above diagonal), and multivariate distances between PSR23 populations (below diagonal) estimated as Mahalanobis squared distances (D^2) and based on standardized phenotypic and nutrient data in each of five years of field experiments.

	Year1	Year 2	Year 3	Year 4	Year 5
Year 1	83.5	15.6	5.6	4.7	4.8
Year 2	45.9**†	75.0	16.8	12.6	10.9
Year 3	63.4***	55.3**	77.6	16.8	8.3
Year 4	65.1***	59.5***	56.7**	65.9	10.0
Year 5	64.7***	64.9***	63.4***	60.2***	66.0

†; **, and ***, significant at the 1, and 0.1% level of probability, respectively.

Changes (i.e., decrease or increase) in phenotypic integration of Cuphea PSR23 (Table 6), in comparison with its wild parents (above diagonal), and after five years of field experimentation (below diagonal) was expressed as the deviation in r-values between traits (i.e., Cuphea PSR23 – Wild species; and Cuphea PSR in 2008 – Cuphea PSR23 in 2003). Sixty-seven percent of Cuphea PSR23 deviations from its wild parents were positive and significant; 13% of which were negative (between SLA and NAR, and between C:N and N:P), while the remaining 33% of the deviations in r-values were not significant. On the other

hand, 80% of deviations in r-values during five years of cultivation were significant, 20% of which were negative; while the remaining 20% were non-significant. A thorough analysis of the plant-size controlled integration index (PIIsc) for the same traits (Table 7) suggested that only RGR and N:P out of the 10 traits evaluated for this index were not significant. Similar results were obtained when the relative integration index corrected for plant size was estimated. In both cases, C:N ratio in metabolic tissues had the highest, and N:P in structural tissues had the smallest PIIsc and RelPIIsc values.

Table 6. Changes in phenotypic integration in *Cuphea* PSR23 in comparison with its wild parents (*C. lanceolata* and *C. viscosissima*) (above diagonal), and for five years of *Cuphea* PSR23 experimentation in 2003 as compared with 2008 (below diagonal).

Trait	Do	SLA	RGR	NAR	C:N	N:P
Do†		0.28*§	0.23*	0.33*	0.11ns	0.09ns
SLA	0.31*		0.25*	-0.23*	0.21*	0.08ns
RGR	0.30*	0.51*		0.09ns	0.33*	0.07ns
NAR	0.45*	-0.32*	0.26		0.43*	0.21*
C:N	0.09ns	0.32*	-0.23*	0.24*		-0.34*
N:P	0.03ns	0.12ns	0.21*	0.25*	-0.44*	

†; Do, Fractal dimension; SLA, Specific leaf area; RGR, relative growth rate; NAR, Net assimilation ratio; C:N, Carbon-to-nitrogen ratio; N:P, Nitrogen-to-phosphorus ratio.

§; * significant at the 5% level of probability; ns: not significant.

Table 7. Plant-size controlled index (PIIsc), and relative integration index corrected for plant size (RelPIIsc) for eco-physiological traits and nutrient ratios in structural, metabolic and reproductive tissues of *Cuphea* PSR23.

PII	Phenotypic integration index for:									
	Eco-physiological traits				Nutrient ratios					
	Do	SLA	RGR	NAR	Structural		Metabolic		Reproductive	
					C:N	N:P	C:N	N:P	C:N	N:P
PIIsc	0.42*†	0.60*	0.28	0.65*	0.49*	0.21	0.75*	0.33	0.83*	0.47*
RelPIIsc	15.5*	29.7**	8.3	21.8**	17.7*	7.2	25.8**	10.2	28.2**	16.2*

†; *, **, and *** significant at the 5, 1 and 0.1% level of probability, respectively.

Table 8. Validation coefficients of determinations for each of PC1 and PC2 derived from capsule, seed, plant phenotypic, and nutrient traits in each of three years (2004–2006) of field experiments on *Cuphea* PSR23 and in averaged over two years of filed experiments (2003–2004) of wild *Cuphea lanceolata* and *C. viscosissima*.

Species and Year	Capsule		Seed		Phenotypic traits		Nutrients		
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2	
	Validation coefficient of determination, Q ²								
A: <i>Cuphea</i> PSR23 (2004)	0.43*†	0.24	0.57	0.18	0.32	0.16	0.53	0.22	
B: <i>Cuphea</i> PSR23 (2005)	0.47	0.22	0.59	0.21	0.32	0.15	0.46	0.20	
C: <i>Cuphea</i> PSR23 (2006)	0.45	0.18	0.58	0.25	0.27	0.11	0.49	0.25	
D: Wild <i>Cuphea</i> spp.	0.44	0.24	0.43	0.18	0.45	0.14	0.56	0.22	

†; *, **, significant at the 5 and 1% level of probability, respectively.

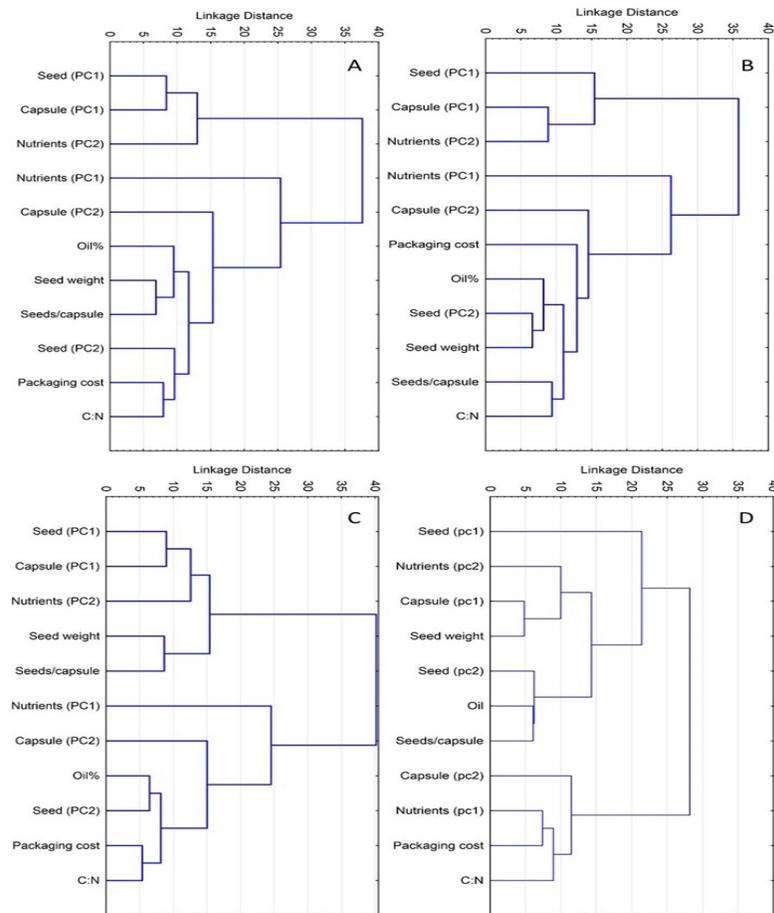


Figure 2. Integrated phenotype of *Cuphea* PSR23 [(A): field experiments in 2004, (B): 2005, and (C): 2006] and its wild parents, *C. lanceolata* and *C. viscosissima* [(D): average field experiments in 2004-2006] based on Euclidean linkage and UPGMA clustering method of four z-transformed phenotypic traits (seed weight, seeds/capsule, oil% and C:N ratio in seeds).

The level of phenotypic integration decreased slightly between 2004 and 2006 as measured by Euclidean distances (i.e., larger overall linkage distance) between five phenotypic traits (z-transformed packaging cost (i.e., weight of capsule tissue in relation to seed weight), oil%, seed weight, seeds/capsule, and C:N ratio) and the first 2 PCs derived from capsule, seed and nutrient traits, in *Cuphea* PSR23 during three successive years (Fig. 2A-C); and was lower than the average estimate of its wild parents (Fig. 2D). Although the general patterns of trait linkages were more or less the same during three years of phenotyping, there were differences in trait associations at the sub-group level and in how clusters were amalgamated. There were two main clusters in Fig. 2A-C, with PC1 derived from nutrients being linked at an intermediate distances between both. Trait structural and amalgamation distances in wild parents differed from those in *Cuphea* PSR23, with a maximum linkage distance of 27 in the wild species (Fig. 2D) as compared to 40 in PSR23. Nutrients (PC1) in *Cuphea* PSR23 formed a separate sub-cluster 25 Euclidean linkage distance; while seed (PC1) was separate from the rest of traits in the wild species at 20 Euclidean linkage distance. The combined validation coefficients of determinations (Q^2) for PC1 and PC2 derived from capsule, seed, plant phenotypic, and nutrient traits in each of three years (2004-2006) of field experiments on *Cuphea* PSR23 and in averaged over two

years of field experiments (2003-2004) of wild *Cuphea lanceolata* and *C. viscosissima* (Table 8) ranged from a low value of 0.47 (phenotypic traits in 2005) to a high value of 0.75 for nutrients in each of 2004 and 2006. The respective values for the wild species were 0.59 for phenotypic traits to 0.78 for nutrients. However, there was no clear temporal trend of Q^2 , although numerically Q^2 values were smaller for phenotypic traits and larger for capsule and seed traits.

Multivariate assessment of phenotypic integration in *Cuphea lanceolata* (Fig. 3), the wild parent of *Cuphea* PSR23, was assessed by trait loadings and the amount of variance explained by two PCs (Fig. 3A); bi-variate correlations among traits (Fig. 3B); and Euclidean linkage distances among traits using UPGMA clustering method (Fig. 3C). Both PCs explained 75.67% of total variation (Fig. 3A); most traits, especially the eco-physiological traits, had large loadings (i.e., correlations with a PC) on both PCs, except reproductive C:P, metabolic C:N, and reproductive C:N, in decreasing order. Positive loadings of NAR and RGR, and negative loadings of the remaining traits on PC1, accounted for 47.82% of total variance, and largely characterized *C. lanceolata*'s phenotype. The joint trait clustering produced a highly variable heatmap (Fig. 3B), with correlation coefficients among traits (range from -0.8 to +1.0) suggesting that Do, SLA, RGR and NAR, displayed the largest number of variable correlation coefficients, followed by those within and among other (i.e., nutrient ratio) traits, while the smallest variability in correlation coefficients was displayed by nutrient ratios. There were 32% non-significant pairwise correlations and only 23% of all pairwise correlation coefficients were $>|0.75|$. Two sub-clusters were produced by the UPGMA clustering method and were joined at a Euclidean linkage distance of 65.0 (Fig. 3C). Components of each sub-cluster were joined at a much lower (<12.0) Euclidean distance. Two of the eco-physiological traits (i.e., SLA and RGR) clustered with most nutrient ratios; whereas, Do and NAR clustered with N:P ratios in structural, metabolic and reproductive tissues.

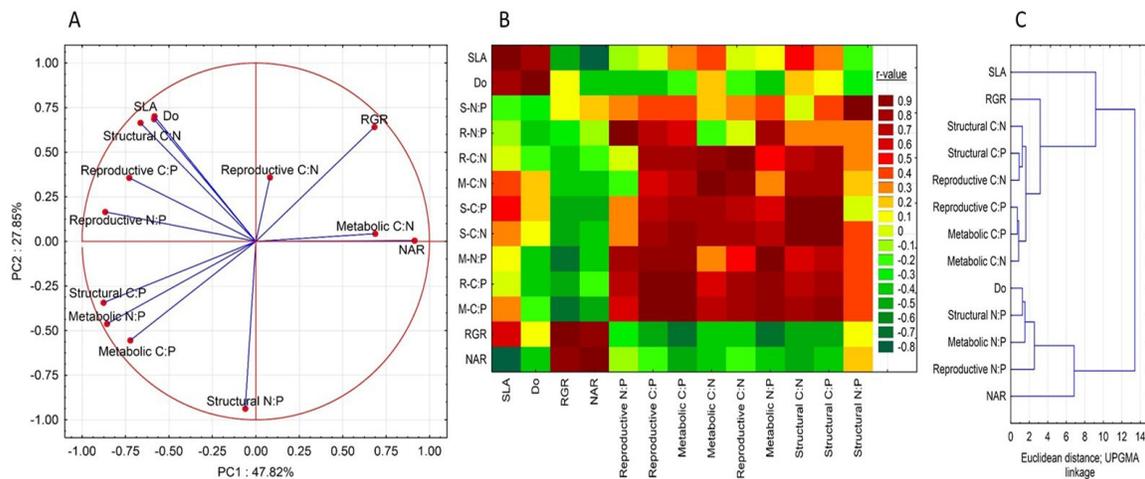


Figure 3. Phenotypic integration of *Cuphea lanceolata* averaged over two years of field experiments as measured by: (A) trait loadings and variance explained by the first principal component (PC1: accounted for 47.82% of total variance) and the second principal component (PC2: accounted for 27.85% of total variance); (B) covariation among traits (range from 1.0 to 25.0) between eco-physiological (Do, fractal dimension; SLA, specific leaf area; RGR, Relative Growth Rate; and NAR, Net Assimilation Rate); and nutrient ratios (C, carbon, N, nitrogen, and P, phosphorus) in structural (S-), metabolic (M-), and reproductive (R-) plant tissues; and (C) Euclidean linkage distances between traits using UPGMA clustering method.

Multivariate assessment of phenotypic integration in *Cuphea viscosissima* (Fig. 4), the second wild parent of *Cuphea* PSR23, was also assessed by the amount of variance explained by two PCs and trait loadings on them (Fig. 4A); bi-variate correlation coefficients among traits (Fig. 4B); and by Euclidean linkage distances among traits using UPGMA clustering method (Fig. 4C). Both PCs explained 77.04% of total variance (Fig. 4B), with PC1 explaining 55.85% and was dominated by positive loadings of RGR, NAR, and metabolic N:C ratio; and mainly by the negative loadings of structural, metabolic, and reproductive C:P ratios; and by metabolic and reproductive N:P ratios. Two of the eco-physiological traits (i.e., Do and SLA), followed by RGR, dominated the positive loadings of PC2; whereas reproductive C:N followed by metabolic C:P ratio dominated the negative loadings. The heatmap produced by joint trait clustering of *C. viscosissima* (Fig. 4B), resembled to a large extent, that of *C. lanceolata*, with a few differences and a slightly stronger maximum negative correlation between traits (e.g., Do and NAR ≥ -0.90). Three eco-physiological traits (SLA, NAR, and RGR), in decreasing order, had the largest number of dis-association with, and largest number of non-significant correlation coefficients nutrient ratios; whereas, N:P ratios in all three plant tissues, generally, had moderate levels of correlation estimates. There were 31% non-significant pairwise correlations and 31% of all pairwise correlation coefficients were $> |0.75|$. A slightly different configuration of two sub-cluster were formed by the same traits (Fig. 4C), although they merged at the same (~ 14.0) Euclidean linkage distance. The SLA and RGR were separated from the C-dominated nutrient ratios; while NAR was separated from Do and from all N:P ratios.

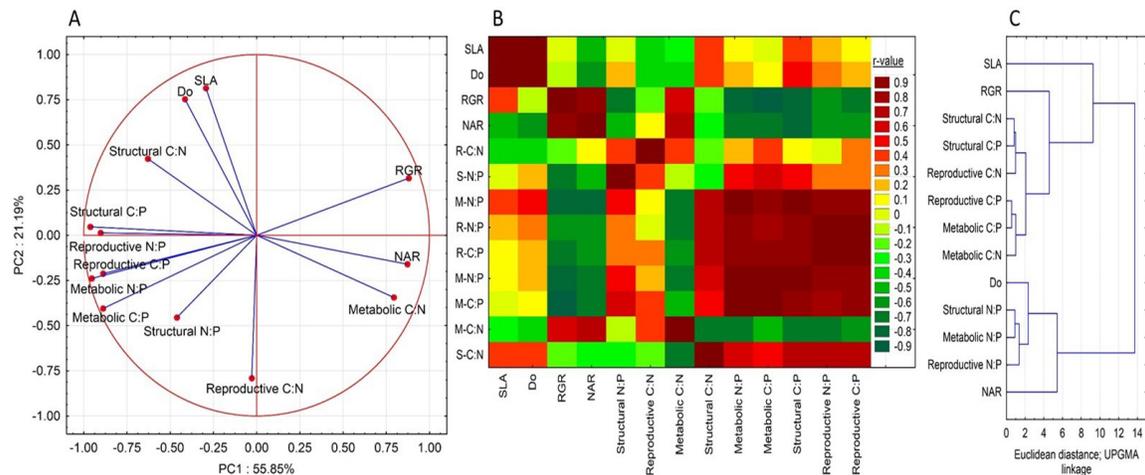


Figure 4. Phenotypic integration of *Cuphea viscosissima* averaged over two years of field experiments as measured by: (A) trait loadings and variance explained by the first principal component (PC1: accounted for 55.85% of total variance) and the second principal component (PC2: accounted for 21.19% of total variance); (B) covariation (range from 1.0 to 21.0) between eco-physiological (Do, fractal dimension; SLA, specific leaf area; RGR, Relative Growth Rate; and NAR, Net Assimilation Rate); and nutrient ratios (C, carbon, N, nitrogen, and P, phosphorus) in structural (S-), metabolic (M-), and reproductive (R-) plant tissues; and (C) Euclidean linkage distances between traits using UPGMA clustering method.

Cuphea PSR23, in comparison with its wild parents, displayed different trait associations and loadings on PC1 which explained 39.65%, and PC2 which explained 23.77% of total variation (Fig. 5A). The smaller total variation explained by these PCs (59.42%) is reflected on the wider dispersion of traits and their shorter eigenvectors on both PCs. Positive loadings of NAR and RGR on PC1, and negative loadings of the remaining traits, largely defined the overall structure of phenotypic integration in *Cuphea* PSR23. However, positive loadings and eigenvectors of NAR and RGR; and negative loadings and

eigenvectors of Do and SLA on PC2 contributed a smaller portion to the explained variance and in determining the level of phenotypic integration. The contributions of most nutrient ratios were much smaller than those of the eco-physiological traits. Negative loadings of structural and reproductive N:P on PC1 contrasted with large loadings of most other nutrient ratios. The joint clustering patterns (i.e., heatmap) produced by *Cuphea* PSR23 (Fig. 5B) differed substantially from those produced by its wild parents; several trait sub-groupings displayed a wide range of (dis)associations, the largest of which were those involving SLA, RGR, and NAR. Intermediate values were observed where N:P- and C:P-based nutrient ratios are involved. There were 41% non-significant pairwise correlations and only 19% of all pairwise correlation coefficients were $>|0.75|$. Trait composition and Euclidean linkage distances of two sub-clusters (Fig. 5C) resembled those of its wild parents, with the exception of a slightly larger linkage distance (14.0) and closer linkage between *Do* and NAR.

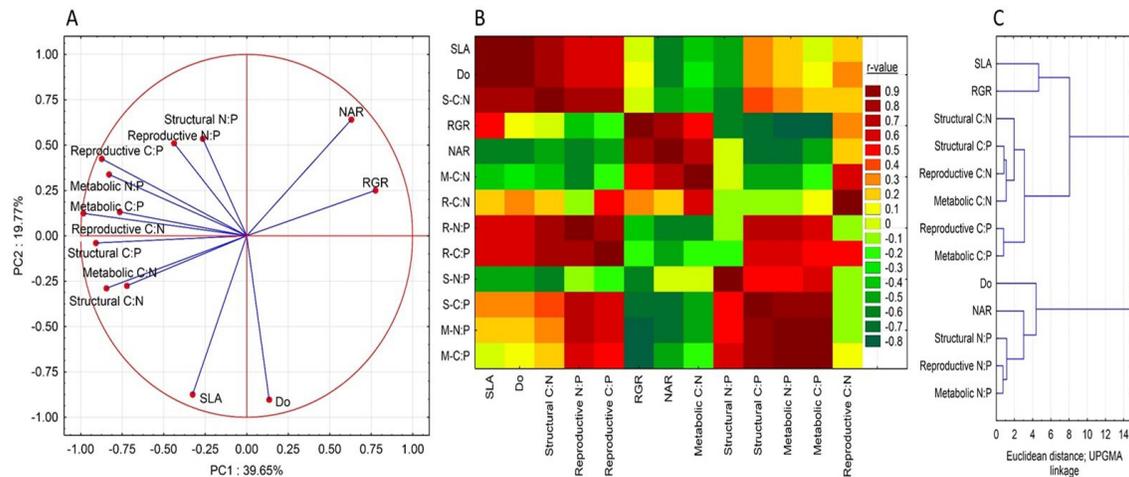


Figure 5. Overall phenotypic integration of *Cuphea* PSR23 after three years of field experimentation as measured by: (A) trait loadings and variance explained by the first principal component (PC1: accounted for 39.65% of total variance) and the second principal component (PC2: accounted for 19.77% of total variance); (B) covariation (range from 1.0 to 41.0) between eco-physiological (*Do*, fractal dimension; SLA, specific leaf area; RGR, Relative Growth Rate; and NAR, Net Assimilation Rate); and nutrient ratios (C, carbon, N, nitrogen, and P, phosphorus) in structural (S-), metabolic (M-), and reproductive (R-) plant tissues; and (C) Euclidean linkage distances between traits using UPGMA clustering method.

DISCUSSION

Large numbers of non-redundant phenotypic traits, measured on single plants or on plant and seed samples, are required to accurately estimate phenotypic variability and phenotypic integration, especially when phenotypically highly variable semi-domesticated crops are considered (Geiler-Samerotte et al. 2013). Such requirement was largely satisfied in the current study and provided the necessary primary and secondary variables to achieve the stated objectives. The concept of joint morphological integration and modularity in crop plants have provided valuable understanding of the evolution of complex biological structures (Chen et al. 2014, Chitwood and Topp 2015) such as those encountered in *Cuphea* PSR23. Phenotypic integration in crop plants (i.e., the pattern and magnitude of trait correlations; or the tendency for covariation) plays a major role as an internal constraint to phenotypic plasticity (Gianoli et al. 2009, Armbruster et al. 2014); the latter is indispensable for crop adaptation to changing environments and may become inversely related to the former, especially under abiotic stress (Granier and Vile 2014). However, if inputs are

inadequate and environmental factors are adverse under field conditions, the expression of phenotypic plasticity may be limited, causing crop plants to fall short of producing the ideal phenotype for a given environment (Pigliucchi 2003, Adhikaru and Wallace 2014). In its simplest form, phenotypic integration can be quantified as the number of significant phenotypic correlations between a set of traits (Murren 2012, Milla et al. 2014). Other statistical approaches that have been used to study phenotypic integration, include, but are not limited to, PCA, CPCA, canonical discriminant analysis, and matrix correlations; all of which have been employed in analyzing multivariate data in the current study to quantify the phenotypic integration in *Cuphea* PSR23 and its wild parents.

Phenotypic integration has a major role in plant and crop evolutionary diversification at the long- (Ladizinsky 1998) and short-term (Murren 2012, Milla et al. 2014) scales; plant traits are expected to function together within a module (e.g., reproductive module) as a prerequisite for the proper functioning of complex plant phenotypes, especially in semi-domesticated or newly-domesticated crop species where some, if not most traits of the wild progenitor(s) are still retained in such crop plants (Bloomfield et al. 2014). A clear understanding of how multiple quantitative (and qualitative) plant traits interact to function in an integrated manner is critical to advance our understanding of the evolutionary developmental plasticity of plants and crops under natural ecosystems (Adhikaru and Wallace 2014) and managed agro-ecosystems (Milla et al. 2014). Morphological integration in crop plants, which is a specific feature of phenotypic integration, can be empirically examined, or described *a priori* (Pigliucchi 2003) such that a theoretical model (Fig. 1) can be tested and refined. Several modules in *Cuphea* PSR23 and its wild parents (Table 1) have been defined by data collected over several years in an effort to study its phenotypic integration by examining patterns of trait variation and covariation using a wide range of multivariate statistical analyses methodologies. On the other hand, a few functional modules have been defined *a priori* and used to study phenotypic integration in *Cuphea* PSR23 using structural equation modelling. The two approaches were instrumental in integrating information at successive levels of organization (e.g., structural, metabolic, and reproductive levels).

Plant dry weight proved to be a reliable indicator of plant size or plant volume (Chen et al. 2014) when predicted by digital imagery; however, predictability of plant size was highly improved when the fractal dimension was included ($R^2 = 0.78$ vs. 0.89 ; data not presented). The mean-variance relationships in heterogeneous populations, such as PSR23 and its wild parents, although potentially informative, can be challenging in quantifying phenotypic integration due to the confounding effect of variance on the mean of a trait. Data transformation and the use of principal components eigenvalues substantially reduced this effect (Zar 1996, Payne et al. 2007) by extracting variances independent of their means.

High-input agro-ecosystems encourage natural selection of trait combinations (i.e., integrated phenotype) in crop plants that support fast resource acquisition and rapid canopy closure (Milla et al. 2014). However, when selecting and breeding for high yield and adaptation under high-input agro-ecosystems (especially of semi-domesticated crops), specific, rather than a wide spectrum of agronomic trait syndromes are selected (Dempewolf et al. 2015). This approach, as predicted by ecological literature (Galloway and Burgess 2009), may lead to reduced phenotypic integration in the new crop at the expense of its adaptation to the target environment. Larger competitiveness for natural resources (i.e., larger plant size), was more obvious during the early years of the current study; whereas lower phenotypic integration was documented towards the latter years. The larger plant size and complexity (i.e., branching) of PSR23, predicted by digital imagery and fractal dimension (when compared to plant size of its wild parents) resulted in a larger relative growth rate but was not associated with a larger net assimilation rate or a higher seed yield (Table 2). Such a reduction in phenotypic integration is predicted to negatively impact crop adaptation to future climate change (Milla et al. 2014).

As a result of sequential and indeterminate plant growth and development, yield distribution (number of capsules and seeds) among primary and secondary branches was uneven (estimated at 10, 25 and 65% at the lower, middle and upper thirds of the plant height, respectively; data not presented). Due to a largely ineffective apical dominance in an indeterminate plant, it is expected that growth in the lower main branches was mainly involved in compensation (e.g., due to environmental damage), and that any increase in capsule yield could be related to increased dry biomass with no concomitant increase or modification of the efficiency in biomass partitioning and allocation to reproductive effort (i.e., seed yield) (Pinet et al. 2009, Jaradat 2012).

The best CPC solution based on AIC test (Table 3) suggested that *Cuphea* PSR23 is phenotypically closer to *C. viscosissima*, (with four out of a maximum of five common PCs and $r = 0.83$ between their standardized PC1s), than to *C. lanceolata*. Both wild parents differed widely in a number of morphological and phenotypic traits, including plant height, and flower and seed traits in this and earlier studies (Hirsinger and Knowles 1984); however, early selections may have favored those phenotypes resembling *C. viscosissima* (Knapp and Crane 2000). In this type of multivariate analysis, some traits may have had disproportionate influence on the overall phenotype than others; and either directional or natural selection over time may have caused large idiosyncratic changes in the PC structures (Chitwood et al. 2015).

The RGR, combined with SLA, which scales positively and linearly with RGR (Poorter et al. 2009), can be optimized for faster plant growth and opportunistic resource acquisition (Grotkopp and Rejmánek 2007) thus leading to high yield potential. The magnitude of both eco-physiological traits is expected to be large due to the “weedy” nature of both wild parents and the weedy ancestry of *Cuphea* PSR23; the later had larger or comparable values for both traits (Table 2). However, *Cuphea* PSR23 had larger RGR and SLA ($188.65 \pm 50.6 \text{ mg g}^{-1} \text{ d}^{-1}$ and $259.9 \pm 62.4 \text{ cm}^2 \text{ g}^{-1}$), respectively) as compared with other Lathyraceae species (113.4 ± 9.1 and 279.7 ± 18.8 , respectively), and Moraceae species (154.8 ± 7.1 and 220.2 ± 14.8 , respectively), but comparable values to Rosaceae species (186.6 ± 13.5 and 261.2 ± 11.1 , respectively). Some other life-history traits of “weedy” *Cuphea* PSR23, besides fast RGR, include short life cycle (~90 days from planting to full maturity), small seed mass, and long duration of flower production due to indeterminacy (Grotkopp and Rejmánek 2007, Dempewolf et al. 2015). These differences are consistent with earlier findings in that “weeds” have faster RGR and larger SLA than “domesticated” crop plants (Poorter et al. 2009), especially under higher soil fertility conditions, typical of soils in field experiments and agricultural fields. However, both RGR and SLA exhibited larger variation in *Cuphea* PSR23 in comparison with the other species. Plant species may vary widely in their potential growth rate; especially when grown under standardized conditions (e.g., greenhouse); for example, RGR may vary threefold or more (Shipley 2002, Poorter et al. 2009). The moderate positive relationship between RGR and SLA in all three species (Table 3) contrasts with earlier studies (Shipley 2002), where weak and negative correlations were reported. Nevertheless, direct and indirect effects of SLA on RGR and NAR if decomposed, may lead to trade-offs between SLA and NAR as functions of daily irradiance; therefore, these relationships and effects may have major impact on the level of phenotypic integration.

Several multivariate statistical tests have been performed using correlation matrices as input data in this study (Table 4, Fig. 3-5); the only exception was the CPCA test in order to detect analogous patterns of trait correlation (Flury 1988). The CPCA examines whether matrices share more complex relationships than expected based on their eigenvectors while being different in the eigenvalues associated with their PCs (Wagner 1984, Phillips and Arnold 1999). Correlation matrices are preferred over variance-covariance matrices in morphological integration studies because correlations are independent of scaling factors (Wagner 1984). In addition, large and significant eigenvalues >1 in PCA can only be expected in randomly organized correlation matrices if there is a bias towards positive correlations

(Flury 1988). The magnitude and significance of all adjusted matrix correlation coefficients (Table 4) indicated that the phenotypic integration of trait components was a dynamic temporal process, in line with the predictions of ecological theory (Milla et al. 2014) and as confirmed by the repeatability tests (Rohlf 2011). Results of the canonical discrimination test (Table 5), which is basically a multivariate analysis of variance (Payne et al. 2007), identified which traits contributed to the large discrimination of germplasm over time, and were in line with the matrix correlation test as indicated by the D^2 estimates. These results have been corroborated by changes in phenotypic integration of Cuphea PSR23 based on each of the eco-physiological and nutrient ratio traits (Table 6), and rigorously verified by estimates of the plant-size controlled index (PIIsc) and the relative integration index (RelPIIsc) when corrected for plant size (Table 7) in all traits estimated in structural, metabolic and reproductive tissues. Correction for plant size was emphasized due to its large variability under conditions of high plant density (Ciaccio et al. 2014, Milla et al. 2014), which was documented in Cuphea PSR23 under field conditions (Jaradat and Rinke 2014).

Results of a series of multivariate statistical tests, comparing Cuphea PSR23 with its wild parents (Figs. 3-5), can be summarized by the differences between total variance explained by two PCs in *C. lanceolata* (75.67%) (Fig. 3A), *C. viscosissima* (77.04%) (Fig. 4A), and Cuphea PSR23 (59.42%) (Fig. 5A). The PC analyses suggested that five and six PCs were needed to explain ~90% of total variation in *C. lanceolata* and *C. viscosissima*, respectively; whereas, eight PCs were needed to explain the same (~90%) portion of variance in Cuphea PSR23 (data not presented). Similarly, 41% of the pairwise correlation coefficients were not significant among the 13 traits in Cuphea PSR23 (Fig. 5B), compared to 32 and 31% in *C. lanceolata* (Fig. 3B) and *C. viscosissima* (Fig. 4B), respectively. The same trend was observed in percent pairwise correlations $>|0.75|$, with the smallest (19%) in Cuphea PSR23, compared to 23 and 31% in *C. lanceolata* and *C. viscosissima*, respectively. These percentages are indicative of the level of phenotypic integration in each of the three Cuphea species and confirm that managed agroecosystems promote lower levels of phenotypic integration (Milla et al. 2014). However, linkage patterns, but not maximum linkage distances (Figs. 3C, 4C, and 5C) illustrate, in a semi-quantitative manner, trait relationships in all three species. Typically, when traits co-vary tightly (highly significant r -values), they produce highly variable eigenvalues; however, eigenvalues will be mostly similar when co-variation between traits is weak (Wagner 1984).

When expressed as the number of significant correlations ($p < 0.05$) of each trait with all others, phenotypic integration may identify which traits displayed decreased (or increased) plasticity, either over time (Fig. 2), or in comparison with the respective traits of its wild parents (Fig. 3B, 4B and 5B); such comparisons can help quantify the departure from the optimal plant phenotype for the experimental location or the production environment (Murren et al. 2002, Tetsana et al. 2014). Cuphea PSR23 and similar semi-domesticated plants reveal complex phenotypic traits that are naturally extremely highly dimensional (Chen et al. 2014, Chitwood and Topp 2015); therefore, the large number of phenotypic measurements, such as those used in this and similar studies on Noug (*Guizotia abyssinica*) (Dempewolf et al., 2015), Evening Primrose (*Oenothera* spp.) (Vilela et al. 2008), and Lesquerella (*L. fendleri*) (Ploschuk et al. 2005) is an important goal in phenomics and phenotypic integration studies. Large numbers of measurements on individual plants, made it possible to accurately estimate phenotypic thus adjusting for variability due to the typical large sampling errors on trait variances, and when phenotypic data is heteroscedastic (Geiler-Samerotte et al. 2013). The use of a combination of multivariate methods, embedded in a pipeline of statistical analyses, is highly desirable (Granier and Vile 2014); it usually leads to increased discrimination efficiency of biologically meaningful variables. However, the challenge of interpreting the phenotype in a biologically expressive manner remains, even if the statistical and biometrical tools are available to capture its complex and multidimensional structure (Chitwood et al. 2015).

CONCLUSIONS

Interspecific hybridization, such as the one between *C. lanceolata* and *C. viscosissima* to produce *Cuphea* PSR23, historically played an important role in plant and crop evolution as a means of promoting crop diversity. Understanding the interactions and inter-dependencies of plant traits is important to advancing our knowledge of a crop developmental flexibility, adaptation of crop plants to new environments, and to new or alternative management practices. *Cuphea* PSR23 is a semi-domesticated potential oilseed crop that produces high-quality oil appropriate for several industrial and food applications. Several cycles of planting and harvesting may result in indirect selection for certain morphological and yield-related traits under agricultural management. Phenotypic integration was quantified as the deviation of functionally linked traits of the whole phenotype of *Cuphea* PSR23 from the mean of its wild parents; and how these traits evolved during cultivation in a non-coordinated manner during cultivation. High resource-use strategies, as well as more efficient resource-use traits may have evolved in a coordinated manner during years of cultivation after hybridization and selection of *Cuphea* PSR23. It is postulated that directional selection during *Cuphea* PSR23 cultivation, may have only targeted a few beneficial traits (e.g., fractal dimension, seed size, and oil content) rather than the whole “phenotype” and resulted in reduced phenotypic integration during cultivation. This process may lead to a reduced level of trait combinations necessary for plant adaptation under varying environmental conditions. Under managed agricultural systems, *Cuphea* PSR23 plants became larger because they invested more in traits involved in competition for light and other renewable and non-renewable resources. Reduction in trait association and integration may affect the possibility of improving *Cuphea* PSR23 as a semi-domesticated crop, to withstand biotic and abiotic stresses. The compiled information on this semi-domesticated oilseed crop may be of value to agronomists and oilseed breeders in defining which plant traits are of essential for the full domestication of PSR23 and to improve its seed and oil yield. Directional selection during several cycles of planting, harvesting and seed selection for larger phenotypic expression (e.g., larger biomass combined with more complex plant architecture) favoured reduced integration through modified relationships between quantitative traits. A compromise between less integration and more expression of a particular set of traits will be inevitable in order to achieve optimum yield of *Cuphea* PSR23 under specific environmental and management conditions.

REFERENCES

- Adhikaru B., Wallace L.E. (2014). Does the *Platanthera dilatata* (Orchidaceae) complex contain cryptic species or continuously variable populations? *Plant Systematics and Evolution* 300, 1465–1476.
- Bloomfield J.A., Rose T.J., King G.J. (2014). Sustainable harvest: managing plasticity for resilient crops. *Plant Biotechnology* 12, 517–533.
- CAMO (2011) *The Unscrambler software, release 10.1*. CAMO Software Inc. One Woodbridge Center, Suite 319. Woodbridge, NJ, USA.
- Chen D., Neumann K., Friede S., Kilian B., Chen M., Altmann T., Klukas C. (2014). Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. *The Plant Cell* 26, 4636–4655.
- Chitwood D.H., Topp C.N. (2015). Revealing plant cryptotypes: defining meaningful phenotypes among infinite traits. *Current Opinion in Plant Biology* 24, 54–60.
- Ciaccio M.F., Finkle J.D., Xue A.Y., Bagheri N. (2014). A systems approach to integrative biology: An overview of statistical methods to elucidate association and architecture. *Integrative and Comparative Biology* 54, 296–306.

- Corwin D.L., Lesch S.M. (2003). Application of soil electrical conductivity to precision agriculture: theory, principles, and guidelines. *Agronomy Journal* 95,455–471.
- Crisci A. (2012). Estimation methods for the structural equation models: Maximum likelihood, partial least squares and generalized maximum entropy. *Journal of Applied Quantitative Methods* 7, 1–15.
- Dempewolf H., Tesfaye M., Teshome A., Bjorkman A.D., Andrew R.L., Scascitelli Black S., Bekele E, Engels J.M.M., Quentin C. B., Cronk Q.C.B., Rieseberg L.H. (2015). Patterns of domestication in the Ethiopian oil-seed crop noug (*Guizotia abyssinica*). *Evolutionary Applications* 8, 464–475.
- Ferreyra L.I., Vilaridi J.C., Verga A., Lopez V., Saidman B.O. (2013). Genetic and morphological markers are able to differentiate three morphotypes belonging to section Algarobia of genus *Prosopis* (Leguminosae, Mimosoideae). *Plant Systematics and Evolution* 299, 1157–1173.
- Flury B. (1988). *Common principal components and related multivariate models*. New York: Wiley.
- Galloway L.F., Burgess K.S. (2009). Manipulation of flowering time: phenological integration and maternal effects. *Ecology* 90, 2139–2148.
- Geiler-Samerotte K.A., Bauer C.R., Li S., Ziv, N., Gresham D., Siegal M.L. (2013). The details in the distributions: why and how to study phenotypic variability. *Current Opinion in Biotechnology* 24, 752–759.
- Granier C., Vile D. (2014). Phenotyping and beyond: modeling the relationships between traits. *Current Opinion in Plant Biology* 18, 96–102.
- Grotkopp E., Rejmánek M. (2007). High seedling relative growth rate and specific leaf area are traits of invasive species: phylogenetically independent contrasts of woody angiosperms. *American Journal of Botany* 94, 526–532.
- Hersinger F., Knowles P.F. (1984) Morphological and agronomic description of selected *Cuphea* germplasm. *Economic Botany* 38, 439–451.
- Jaradat A.A. (2012). Reproductive allocation and nutrient relationships in *Cuphea*: a semi-domesticated oilseed crop. *Journal of Plant Nutrition* 35, 1579–1599.
- Jaradat A.A., Rinke J.L. (2014). Empirical modeling of the impact of Mollisol soils variation on performance of *Cuphea*: a potential oilseed crop. *Australian Journal of Crop Science* 8, 1093–1113.
- Kerkhoff A.J., Fagan W.F., Elser J.J., Enquist B.J. (2006). Phylogenetic and growth form variation in the scaling of nitrogen and phosphorus in the seed plants. *The American Naturalist* 168, E104–E122.
- Kim K-I., Gesch R.W., Cermak S.C., Phippen W.B., Berti M.T., Johnson B.L., Marek L. (2011). *Cuphea* growth, yield, and oil characteristics as influenced by climate and soil environments across the upper Midwest USA. *Industrial Crops and Products* 33, 99–107.
- Ladizinsky G. (1998). *Plant evolution under domestication*. Kluwer Academic publishers, New York.
- Lamb E.G., Shirliffe S.J., May W.E. (2011). Structural equation modeling in the plant sciences: an example using yield components in oat. *Canadian Journal of Plant Science* 91, 603–619.
- Mantel N.A. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research* 27, 209–220.
- Masson P., Dalix, T., Bussière S. (2010). Determination of major and trace elements in plant samples by inductively coupled plasma-mass spectrometry. *Communications in Soil Science and Plant Analysis* 43, 231–243.
- Martre P., Quilot-Turion, B., Luquet, D., Memmah, O-S., Chenu, K., Debaeke P. (2015). Model-assisted phenotyping and ideotype design. In: Sadras V.O., Calderini D.F. (Eds.) *Crop Physiology, Second Edition*. Elsevier 349–373.

- Milla R., Morente-Lopez J., Alonso-Rodrigo J.M., Martín-Robles N., Chapin F.S. III. (2014). Shifts and disruptions in resource-use trait syndromes during the evolution of herbaceous crops. *Proceedings of the Royal Society B* 281, 20141429.
- Murren C.J. (2012). The integrated phenotype. *Integrative and Comparative Biology* 52, 64–76.
- Payne R.W., Harding S.A., Murray D.A., Soutar D.M., Baired D.B., Welham S.J., Kane A.F., Gilmore A.R., Thompson R., Webster R., Wilson G.T. (2007). *The Guide to GenStat Release 10. Part 2: Statistics*. VSN Int., Hemel Hempstead, UK.
- Phillips P.C., Arnold S. J. (1999). Hierarchical comparison of genetic variance-covariance matrices. I. Using the Flury hierarchy. *Evolution* 53, 1506–1515.
- Pigliucchi M. (2003). Phenotypic integration: studying the ecology and evolution of complex phenotypes. *Ecology Letters* 6, 265–272.
- Pinet A., Jullien A., Allirand J.-M., Mathieu A., Ney B. (2009). Are yield and biomass distribution affected by sink organ clipping during reproductive phase of winter oilseed rape (*Brassica napus* L.)? *Plant Growth Modeling, Simulation, Visualization and Applications (PMA)*, Third International Symposium, 9-13 November 2009, 222–225.
- Ploschuk E.L., Slafer G.A., Ravetta D.A. (2005). Reproductive allocation of biomass and nitrogen in annual and perennial Lesquerella crops. *Annals of Botany* 96, 127–135.
- Poorter H., Niinemets Ü., Poorter L., Wright I.J., Villar R. (2009). Causes and consequences of variation in leaf mass per area (LMA): a meta-analysis. *New Phytologist* 182, 565–588.
- R Core Team (2015). *R: A language and environment for statistical computing*. R version 3.1.2. <http://R-project.org>.
- Rasband W. S. (2015) *ImageJ v 1.49u: Image processing and analysis in Java*. U.S. National Institute of Health, Bethesda, MD. <http://Imagej.nih.gov/ij>.
- Rayan A. (2013). *VARIAN-Rapid measurements of major, minor and trace elements in plant and food materials using Varian 730-ES*, www.varianinc.com.
- Rohlf F.J. (2011). *NTSYSpc: Numerical Taxonomy System. v. 2.21o*. Exeter Software, Setauket, New York.
- Shipley B. (2002). Trade-offs between net assimilation rate and specific leaf in determining relative growth rate: relationship with daily irradiance. *Functional Ecology* 16, 682–689.
- StatSoft, Inc. (2013). *STATISTICA (data analysis software system). Version 12*. www.statsoft.com.
- Tetsana N., Pedersen, H.A.E., Sridith, K. (2014). Character inter-correlation and the potential role of phenotypic plasticity in orchids: a case study of the epiphyte *Liparis resupinata*. *Plant Systematics and Evolution* 300, 517–526.
- Torices R., Muñoz-Pajares A.J. (2015). *PHENIX: An R package to estimate a size-controlled phenotypic integration index* (<https://cran-r-project.org/web/packages/PHENIX.pdf>).
- Urbaniak J. (2010). Analysis of morphological characters of *Chara baltica*, *C. hispida*, *C. horrida*, and *C. rudis* from Europe. *Plant Systematics and Evolution* 286, 209–221.
- Vilela A., Gonzalez-Paleoa L., Rondanini D., Ravetta D. (2008). Biomass allocation patterns and reproductive output of four *Oenothera* L. accessions native to Argentina. *Industrial Crops and Products* 27, 249–256.
- Wagner G.P. (1984). On the eigenvalue distribution of genetic and phenotypic dispersion matrices: evidence for a non-random organization of quantitative character variation. *Journal of Mathematical Biology* 21, 77–95.
- Warton D.I., Dursma R.A., Falster D.S., Taskinen S. (2012). SMATR 3 – an R package for estimation and inference about allometric lines. *Methods in Ecology and Evolution* 3, 257–259.
- Yeh F.C., Yang R.C., Boyle T.B.J., Ye Z.H., and Mao J.X. (2000). *POPGENE, the user-friendly shareware for population genetic analysis*. Molecular Biology and biotechnology Center, University of Alberta, Edmonton, AB, Canada.

Zar J.H. (1996). *Biostatistical analysis*, 3rd ed. Prentice Hall, Englewood Cliffs, NJ.

Zhang T., Lamb E. G., Soto-Cerda B., Duguid S., Cloutier S., Rowland G., Diederichsen A., Booker H.M. (2014). Structural equation modeling of the Canadian flax (*Linum usitatissimum* L.) core collection for multiple phenotypic traits. *Canadian Journal of Plant Science* 94, 1325-1332.