

This is the peer reviewed version of the following article:

Terzić S, Zorić M., Seiler G. Qualitative traits in sunflower breeding: UGA-SAM1 phenotyping case study. *Crop Science* 2020, 60 (1), 303 – 319.

which has been published in final form at [<https://doi.org/10.1002/csc2.20059>]. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. This article may not be enhanced, enriched or otherwise transformed into a derivative work, without express permission from Wiley or by statutory rights under applicable legislation. Copyright notices must not be removed, obscured or modified. The article must be linked to Wiley's version of record on Wiley Online Library and any embedding, framing or otherwise making available the article or pages thereof by third parties from platforms, services and websites other than Wiley Online Library must be prohibited.



This work is **licensed** under

Creative Commons - Attribution-Noncommercial-NoDerivative Works 4.0 International

Plant Genetic Resources

Qualitative Traits in Sunflower Breeding: UGA-SAM1 Phenotyping Case Study

Sreten Terzić,* Miroslav Zorić, and Gerald J. Seiler

S. Terzić, Sunflower Dep., Institute of Field and Vegetable Crops, Maksima Gorkog 30 21000 Novi Sad, Serbia; M. Zorić, Maize Dep., Institute of Field and Vegetable Crops, Maksima Gorkog 30 21000 Novi Sad, Serbia; G.J. Seiler, USDA-ARS, Edward T. Schafer Agricultural Research Center, Northern Crop Science Lab., 1616 Albrecht Blvd. N., Fargo, ND 58102-2765, USA. Received 18 Feb. 2019. Accepted 9 Sept. 2019. *Corresponding author (sretenterzic@gmail.com). Assigned to Associate Editor Eric Olson.

Abbreviations: BRAN, branching; CMS, cytoplasmic male sterile; DUS, distinctness, uniformity, and stability; FLO, flowering time; HA, maintainer lines; HOMALS, homogeneity analysis in R; LATV, lateral veins angle; LAUR, leaf auricles; LBLIS, leaf blistering; LCROS, leaf shape of cross-section; LD, linkage disequilibrium; LGCOL, leaf green color; LSDIS, shape of the distal part of the leaf; LSIZ, leaf size; LTIP, height of the tip of the leaf blade compared with insertion of petiole; LWIN, leaf wings; MAT, maturity groups; NCRPIS, Northcentral Regional Plant Introduction Station; NO, non-oil lines; O, oil lines; OPV, open-pollinated variety; PC, principal component; PHE, plant height; QTL, quantitative trait locus/loci; RHA, restorer lines; RIL, recombinant inbred line; SCOL, seed color; SSCO, seed stripes color; SSHP, seed shape; SSIZE, seed size; SSTR, seed stripes; UPOV, International Union for the Protection of New Varieties of Plants.

ABSTRACT

The production of the first permanent, publicly available sunflower (*Helianthus annuus* L.) association mapping population (UGA-SAM1) provided material to test the usability of morphological descriptors for discriminating germplasm accessions. The objective of this study was to evaluate the phenotypic diversity in the UGA-SAM1 population and evaluate the usefulness of qualitative traits for the discrimination of genotypes. The SAM1 population consists of 285 accessions characterized for 20 morphological traits. The Shannon–Weaver diversity index (H') was used to determine phenotypic diversity, whereas HOMALS (homogeneity analysis by alternating least squares) and association statistics were used to determine discriminative power of the descriptors. Phenotypic diversity was moderately high for the traits (0.74). The highest diversity was found in the less developed genotypes followed by non-oil genotypes. Pronounced associations in individual genotype groups and category traits, most notably for the maturity trait in the restorer line (RHA) oil group, resulted in both high association and diversity index values. The association test proved to be a useful addition to HOMALS analysis for determining the trait discriminative power. Adequate selection of traits used in germplasm evaluation can improve the efficiency of breeding programs, whereas the loss of variability could be lowered if diversity focused traits were used including leaf, seed, and certain flower traits, instead of only focusing on yield and quality.

Sunflower (*Helianthus annuus* L.) is one of the major oil crops in the world, but also a model species for research on plant speciation (Whitney et al., 2015). Use of mapping populations has greatly increased the available genotypic data and made identification of quantitative trait loci (QTL) possible for important traits (Bowers et al., 2012). For traditional QTL mapping in biparental crosses, recombinant inbred lines (RIL) with high levels of homozygosity are developed, allowing comparison of performance of genotypes between sites (Baack et al., 2008). Some sunflower RIL populations are in the process of being made available for distribution from the USDA, National Plant Germplasm System (NPGS), Northcentral Regional Plant Introduction Station (NCRPIS), such as RHA280 × RHA801 (confectionary × oilseed), HA89 × ANN1238 (oilseed × wild), and Hopi RIL (oilseed × landrace) (Tang et al., 2002; Baack et al., 2008; Bowers et al., 2012).

TOC Head: ; Section Head: ; Article Type: ARTICLE

Germplasm collections or natural populations can also be used to associate genotype with phenotype using linkage disequilibrium (LD) between closely linked loci. The advantage of LD vs. biparental QTL mapping is that analysis of a larger and more representative gene pool is possible. Depending on the available genotypic data, either “candidate gene approach” can be used, or a whole-genome analyses in search of LD between known marker genes and the desired trait (genome-wide association study; Neumann et al., 2011; Álvarez et al., 2014).

The usual approach in plant breeding for introducing a new trait is to exploit the existing variability and select pure lines. The desired trait can be introduced from cultivars, landraces, or crop wild relatives, if not found within elite inbred material (Atlagić and Terzić, 2014). Using exotic germplasm to introduce a novel trait usually implies costly evaluation of many genotypes because the amount and distribution of genetic diversity is not known. To increase efficiency of finding the source of desired genes in large collections and promote their use, Frankel (1984) proposed a “core collection” approach, where the genetic diversity of a large collection is evaluated by including only accessions with a minimum of repetitiveness.

For a crop species, a core collection consists of a limited number of accessions chosen to represent the genetic diversity of the whole crop species and its crop wild relatives. It is assembled cooperatively by national and international genebanks and supplemented where needed to fill the gaps (Brown, 1995; Van Hintum et al., 2000). The first large synthetic core collection was the international Barley Core Collection (Knupffer and van Hintum, 1995). The first cultivated sunflower core collection was based on selected accessions from the USDA-ARS-NCRPIS gene bank (Brothers and Miller, 1999). Since then, the first international sunflower core collection (UGA-SAM1) was described by Mandel et al. (2011). UGA-SAM1 represents 90% of the allelic diversity in the cultivated sunflower gene pool (Mandel et al., 2011). It is, as such, an invaluable resource for further research and improvement of cultivated sunflower using mapping techniques appropriate for large populations.

The choice of mapping technique depends on the available genotypic data, but also the diversity of the desired trait (McKhann et al., 2004). The genotypic data for the UGA-SAM1 population are increasing (Hübner et al., 2019) and are mostly available through the sunflower genome database (<https://sunflowergenome.org/>).

Evaluation of a large numbers of phenotypic traits is labor intensive and time consuming. Using subsamples of the whole collection can reduce the time and number of people involved in phenotyping but increases the risk of missing allelic diversity for specific traits as the number of accessions in the subsample is reduced (Foulley and Ollivier, 2006).

Validating the diversity in core collections is important, since they are also attractive as a starting material for discovering desired traits and genotyping (Mandel et al., 2013; Nambeesan et al., 2015). Phenotyping is predominantly used in the evaluation of heterotic potential, whereas relationships between plant morphological traits and disease intensity are of increasing importance in crops like sunflower (Schwanck et al., 2016). Proper evaluation of morphological traits is thus important for all breeding steps from early line development to comparison of hybrid performance.

The objectives of this study were (i) to evaluate the present phenotypic diversity in the first publicly available association mapping population (UGA-SAM1) using the Shannon–Weaver

TOC Head: ; Section Head: ; Article Type: ARTICLE

diversity index (H') according to the frequency of genotypes in each descriptor category; (ii) to determine whether association test or HOMALS (homogeneity analysis by alternating least squares)-based analysis is more informative for determining trait discriminative power; and (iii) to assess the usefulness of individual qualitative traits for the discrimination of sunflower accessions.

MATERIALS AND METHODS

Plant Materials

The sunflower population used in this study was composed of 285 lines from a total of 288 lines previously described as a core collection of the cultivated sunflower gene pool (Mandel et al., 2011). This population is composed of accessions from the collections held by both the USDA-ARS-NCRPIS, and the French National Institute for Agricultural Research (INRA) (Supplemental Table S1).

All accessions were assigned to specific categories based on their breeding history (maintainer [B] lines = HA; restorer [Rf] lines = RHA) and agronomic use (oil vs. non-oil). The following categories were defined: HA non-oil = HA-NO (48 lines), HA oil = HA-O (80), RHA non-oil = RHA-NO (25), RHA oil = RHA-O (77), and less developed (29) containing introgressed lines, landraces, open-pollinated varieties (OPVs), cultivars, and populations. An “all other” category (26) was created for lines classified as “other oil” and “other non-oil” by Mandel et al. (2011), where B (HA maintainer lines) vs. Rf (RHA restorer lines) designation could not be made, including accessions with no breeding background information. Mandel et al. (2013) designated accessions as “non-oil” when they were either confectionery types or could not be clearly defined as being oil types. The “introgressed” category included accessions with a recent history of introgression from wild *Helianthus* species. The OPV category included named sunflower accessions that represent OPVs of the pre-hybrid era of sunflower breeding, along with two Native American landraces, Hopi and Mandan.

Field Design

The study was established during 2013 near Novi Sad in Rimski šančevi, Serbia (45°19'34.46" N 19°49'39.50" E) on dark chernozem soil (81 m asl). The genotypes were tested in an unreplicated grid plot design (Müller et al., 2010). All genotypes were considered as unreplicated, except the inbred lines HA89 (Miller et al., 1992) and HA26 (IFVCNS, inbred line). These lines were replicated in systematic arrangement throughout the experimental grid. The design grid size was 20 rows × 16 columns. Plants were grown in rows 3.6 m long with 12 plants per plot. The distance between rows was 70 cm and between plants was 30 cm, which provided a plant density of 47,000 plants ha⁻¹. Plants were irrigated and fertilized to minimize effects of selection in the field. Locally recommended sunflower agricultural practices were applied.

Phenotyping

Phenotypic diversity of the tested material was determined using distinctness, uniformity, and stability (DUS) test guidelines (UPOV, 2000). The selected qualitative traits are nominal and

TOC Head: ; Section Head: ; Article Type: ARTICLE

ordinal categorical variables referred to in the paper as descriptors and each of the trait categories as descriptor categories. The traits where International Union for the Protection of New Varieties of Plants (UPOV) descriptor categories did not cover the present variability or were less precise (did not have defined ranges for trait categories like plant height [PHE]) were described using the International Board for Plant Genetic Resources (IBPGR, 1985) descriptor list (Table 1).

Visual ratings were performed in the field for each genotype on a whole-plot basis. Observations of leaf traits were made at the flowering stage R5.5 according to Schneiter and Miller (1981). Branching type and plant height (cm) were rated at the R9 stage regarded as physiological maturity, whereas the seed traits were evaluated after harvest. Days to 50% flowering were recorded as days from sowing to 50% of plants having begun flowering, whereas maturity groups (MAT) were determined based on days to maturity (sowing to 90% of plants at the R9 stage (Table 1).

Statistical Analysis

Diversity Index

Since the variables collected throughout the experiment are categorical nominal traits, the Shannon and Weaver (1949) diversity index (H') was used to estimate the morphological diversity of the genotypes for a particular trait. The estimated H' value was normalized by dividing it by its maximum value ($\log 2_n$), which ensured that all H' values were in the range of 0–1.

Multivariate Analysis of Categorical Traits

The estimation of the diversity of sunflower genotypes was performed by multivariate homogeneity analysis by means of alternating least squares (HOMALS) procedure, also known as principal component analysis (PCA) for categorical nominal data. The HOMALS procedure is based on contingency tables, in which principal results of decomposition are the coordinates of lines and columns, further used for visualization of the results by a two-dimensional plot. The HOMALS plots enable close grouping of objects (genotypes) of the same category, whereas different objects are located at a certain mutual distance. The distance between points representing two objects depends on the similarities of their profile. Categories divide objects into homogenous groups, so that each object is in close proximity to the points, which represent the object category (Michailidis and De Leeuw, 1998).

Association among Categorical Traits and Genotype Groups

To further clarify the associations among the categorical traits and genotype groups defined by Mandel et al. (2013), the two-way contingency or $I \times J$ tables were created for each particular trait with observed genotype frequencies (n_{ij}). The contingency tables were created so that rows contain genotype classification with $i = 1, \dots, I$ categories, whereas columns of the tables contain $j = 1, \dots, J$ categories of the nominal traits. Under the H_0 that the rows and columns of the contingency table are independent, the expected cell frequencies are $m_{ij} = n\pi_i.\pi_{.j}$, $i = 1, \dots, I, j = 1, \dots, J$, where n is the total sample size ($n = \sum_{i,j} n_{ij}$) and π_i ($\pi_{.j}$) is the i th row (j th column) marginal

TOC Head: ; Section Head: ; Article Type: ARTICLE

of the underlying probabilities matrix $\pi = (\pi_{ij})$. The hypothesis of independence is tested through the chi-square test statistic:

$$c^2 = \sum_{i,j} \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}}$$

The distribution of the χ^2 statistics under the H_0 of independence is closely approximated by a χ^2 distribution with $(I - 1)(J - 1)$ degree of freedom, assuming that the expected cell frequencies are larger than five.

The χ^2 values were further examined by contingency coefficient to measure the strength of the association of two categorical variables. The observed contingency coefficient is defined by the following formulae:

$$C_{\text{obs}} = \sqrt{\frac{c^2}{c^2 + n}}$$

Since the contingency coefficient vary from 0–1, which depends on the number of rows and columns of the table, we used the following relationship:

$$C_{\text{max}} = \sqrt{\frac{r-1}{r} \times \frac{c-1}{c}}$$

For cases where the number of rows and columns are equal, the C_{max} value can be estimated as

$$C_{\text{max}} = \sqrt{\frac{r-1}{r}}$$

Finally, we used the ratio among the observed and maximum value of the contingency coefficient for the interpretation of the strength of the association of the categorical variables. The standardized coefficient (C_{stand}) of the contingency defined as $C_{\text{stand}} = C_{\text{obs}}/C_{\text{max}}$ ranges from 0 (complete independence) to 1 (complete dependence). To quantify the effect size based on value of the contingency coefficient, the w index (Cohen, 1988) was estimated using the following equation:

$$w = \sqrt{\frac{C^2}{1 - C^2}}$$

Since there is no general recommendation in the literature, the w index was estimated for both types of the contingency coefficient (i.e., $w_{C_{\text{obs}}}$ and $w_{C_{\text{max}}}$). According to Cohen (1988), the following scale was proposed for the interpretation of the effect size: small effect size from 0.1–0.3, medium effect size from 0.3–0.5, and large effect size >0.5 .

After conducting a formal test of the independence, detailed information about the associations in the contingency table can be gathered by direct visual display of Pearson residuals, such as

TOC Head: ; Section Head: ; Article Type: ARTICLE

$$r_{ij} = \frac{n_{ij} - \hat{m}_{ij}}{\sqrt{\hat{m}_{ij}}}$$

which are standardized raw residuals. In an association plot (Cohen, 1988), each cell is represented by a rectangle that has a height that is proportional to the corresponding Pearson residuals (r_{ij}) and a width that is proportional to the square root of the expected counts ($\sqrt{\hat{m}_{ij}}$). The area is proportional to the raw residuals $n_{ij} - \hat{m}_{ij}$. The sign is visualized by its position relative to the baseline of the association plot.

All statistical analyses and data visualizations were performed using the *vcd* and *homals* libraries from the R computing environment (R Core Team, 2018).

RESULTS

Genotype Diversity

Diversity index (H') is determined by the number of categories and evenness in the distribution of genotypes in each category of a descriptor (Shannon and Weaver, 1949). Uniform distributions of genotypes within the descriptor categories are revealed by higher values of the H' index, while an unbalanced distribution and a lack of diversity is indicated by a low H' index (Moreno et al., 2013). Morphological descriptor categories, genotype distribution and diversity indices of the analyzed population are presented in Table 1.

The mean diversity index of all trait groups was 0.74 and varied from 0.39–0.98. The index value was the lowest for flowering time (FLO), seed color (SCOL), leaf shape of cross-section (LCROS) and branching (BRAN), whereas it was highest for a group of leaf descriptors (leaf size [length \times width, LSIZ], height of the tip of the leaf blade compared with insertion of petiole [LTIP], leaf wings [LWIN]) and MAT. Moderately high to high diversity index (0.72–0.79) was found for most of the seed traits (Figure 1).

Discriminative Power

The diversity for a trait (descriptor) in a population can be high according to the Shannon index, but it does not necessarily imply high discriminative power. To further check discrimination power, distance of descriptors from the biplot origin was analyzed using HOMALS analyses to detect descriptors with unique profiles.

The HOMALS plot revealed that the first two dimensions accounted for 37.9% of total variability (Figure 2a). The first dimension was related to descriptors differentiating genotypes according to BRAN, seed stripes color (SSCO), seed stripes (SSTR), MAT, LTIP, LWIN, and seed size (SSIZE) (Figure 2a). The second dimension was also related to SSCO, SSTR, BRAN, and LTIP. Besides those, leaf blistering (LBLIS) and leaf auricles (LAUR) were most informative for genotype discrimination, together with LSIZ, shape of the distal part of the leaf (LSDIS), leaf serration (LSERR), and SCOL. Morphological descriptors with shorter vectors, positioned close to the plot origin (PHE, LCROS, leaf green color [LGCOL], lateral veins angle [LATV], petiole length [PET], FLO, and seed shape [SSHP]), were less variable among genotypes, less informative, and thus had lower discriminative power for identification of

TOC Head: ; Section Head: ; Article Type: ARTICLE

sunflower genotypes (Figure 2a). The best eight traits analyzed accounted for 58.7% of the variability, whereas BRAN, SSCO, SSTR, LBLIS, and MAT were positioned furthest from the origin (Figure 2b).

Association Statistics

Association statistics were used to check for the association between trait categories and genotype groups as an indication of traits most useful for genotype discrimination (Supplemental Table S2, Figure 1). Interpretation of the effect size in the association test was performed according to interval values defined by Cohen (1988) for social sciences but was used in the current study as guidance and can be regarded as a suggestion. The highest association between descriptor categories and genotype groups was for BRAN, SSCO, MAT, and SSIZE (Table 2).

Each rectangle in the association plots is shaded according to the value of the Pearson residual, so that higher residuals are marked with a darker color using the scale shown in the legend (Figures 3–6). The highest positive association in BRAN was found for the RHA-O group and overall branched plants, whereas the highest negative association was for the same group and nonbranched plants. Similar but less significant associations were found for RHA-NO group. Opposite associations were found for the HA-O and HA-NO groups, with positive associations for nonbranched plants and negative associations for branched plants (Figure 3). Seed stripes color is a trait important in non-oil sunflowers as confirmed by residuals >4 for white stripes in HA-NO and brown stripes in RHA-NO, whereas most of the RHA-O genotypes had no stripes on seeds (Figure 4). Maturity group is similar to branching, a trait where fertility restorer lines RHA-O and RHA-NO differ from all other groups. The RHA-O group had a high positive association with early maturity that gradually declined to a significant negative association with late maturity (Figure 5). Considering SSIZE, both fertility restorer groups are positively associated with smaller seeds, but only significantly in the RHA-O group. The highest positive association for very large seed was seen in the HA-NO group, which also had a negative association for small seeds (Figure 6).

If the association values are presented as an exponential trend line, a negative correlation is visible in comparison with the diversity index (Figure 1). Except partly for SSIZE, SSHP, and MAT descriptors, the association between descriptor categories and genotype groups was less pronounced with the increase of morphological diversity and more uniform distribution of genotypes across descriptor categories (Figure 1). A chi-square test of independence showed the strongest associations between descriptor categories and genotype groups for BRAN, SSCO, MAT, and SSIZE, implying that those descriptors have the highest discriminative power among genotypes. To visualize grouping effect of each of those descriptors, separate association plots are presented for each (Supplemental Figure S1a–S1d).

Genotype Grouping and Distribution

Relationships between genotypes were examined based on HOMALS analysis of 20 morphological descriptors. The present variation resulted in a wide dispersion of genotypes in the HOMALS distribution plot. The first dimension explained 24.97% of the total variability, whereas the second explained 12.96% in the original dataset (Figure 7).

TOC Head: ; Section Head: ; Article Type: ARTICLE

A study of the distribution of 286 genotypes revealed their division along the first dimension into two large groups of HA and RHA lines, whereas the division of oil and non-oil was less pronounced, but still present in the second dimension. Most of the genotypes were clearly separated, indicating their distinct grouping pattern. No complete overlapping was found, whereas partial overlaps were mostly found in the upper left quadrant including genotype pairs such as 62 and 213, 90 and 240, 52 and 269 (Figure 7).

Descriptor categories for BRAN were clearly separated from lower left to upper right quadrant, starting with nonbranched HA lines that included only several lines with basal branching, and continuing with the fully branched and upper stem branching, typical for RHA lines (Supplemental Figure S1a). Seed stripes color was distributed from lower right to upper left quadrant starting with seeds without stripes, followed by a mixed group of white and gray stripes, whereas brown and black were positioned furthest along the second dimension (Supplemental Figure S1b). The MAT biplot was specific as it revealed each of the four genotype groups—HA-NO (including less developed and other), HA-O, RHA-NO and RHA-NO—had all maturity categories represented, whereas they were most distinctly grouped in the RHA-O followed by RHA-NO (Supplemental Figure S1c). Seed size again contributed to the basic four groups as restores were found to have smaller seeds. The RHA-O and RHA-NO groups had no genotypes in Category 9 and only two in Category 7. On the other hand, HA-O had all categories represented, whereas the greatest variability was found in the expanded HA-NO group (Supplemental Figure S1d).

DISCUSSION

Diversity

Morphological traits important for sunflower vegetative growth and yield include plant height; size and position of the head; leaf number, size, and position; days from sowing to maturity; stem diameter; and seed size and weight, among others (Fick and Miller, 1997; Škorić, 2012b). Since plant breeding is focused on improving plant heredity in relation to its economic utility (Fehr, 1987), it is important to have adequate starting material and selection methods.

Proper selection of progenitor is the key to successful breeding, both in advanced cultivars and exotic germplasm from gene banks. Qualitative traits play an important role in the cultivar development process, depending on the desired phenotype as the ultimate breeding goal. They are codified by one or a few genes whose phenotypic expression is under little influence by the environment. Evaluation of qualitative traits is thus frequently used as a tool to confirm the effectiveness of crosses made between selected germplasm (de Jesus Verneti and de Jesus Verneti, 2017). Since the material selected for the current study is expected to contain high diversity, it was used to check the applicability of selected traits as tools in germplasm discrimination, as well as phenotypic diversity for each of the descriptor categories.

The mean diversity index value was moderately high for the material studied. Some descriptor categories were nonetheless underrepresented with frequencies of <0.5% of the total analyzed genotypes as defined by Cavalli-Sforza and Bodmer (1981). Descriptor categories of specific genotype groups with such low frequencies were not used for the calculation of

TOC Head: ; Section Head: ; Article Type: ARTICLE

association statistics to avoid misleading results. Lines with such traits are, on the other hand, practically important as sources of variability.

Diversity for most of leaf traits was high in the current study, even though they are simply inherited depending on one up to several genes (Škorić et al., 2012b). Diversity was not only high at the population level, but also within most of the genotype groups. Leaf size was small to medium in the group of restorer lines mostly due to their branching habit, whereas slightly increased presence of plants with medium-sized leaves was found in the HA-O group resulting from the change in breeding goals since the introduction of hybrids to avoid very tall or very late material (Panković et al., 1991; Vear, 2016). A higher proportion of genotypes with large leaf size was found in the less developed group, and to some extent in HA-NO, resulting from the presence of old cultivars with tall leafy plants and confectionary lines. Lower plant density of confectionary hybrids results in plants with larger leaf area, but such ideotypes are not desired in some programs that are trying to obtain maximum leaf area faster by producing a larger number of smaller leaves (Yang et al., 2010).

The full extent of variation is visible in the populations and its subgroups as each contains plants with very low to very high leaf tip position. Distinctiveness between the categories was highest in the RAH-NO group when the HOMALS plot used all 20 traits. However, category distinction was highest between HA lines when only eight most discriminative traits were used for plot construction. This result shows how selection of descriptors can affect efficiency (lower number of evaluated traits) and precision of selection, leading eventually to better breeding results.

There were no plants with sessile leaves, even though it would be a beneficial trait for increasing the number of plants per hectare (Breton et al., 2010), while extremely long petioles were rarely present since such plant types have been eliminated in modern, high-density cultivated sunflower.

For maximum photosynthetic activity, corrugated leaves oriented downward are preferable (Škorić, 2012a). Leaf blistering was present, but categories of absent and weak dominated the UGA-SAM1 population, except for a slightly higher frequency in HA lines and the less developed group. Considering shape, leaves of sunflower are generally optimized for maximizing photosynthetic rate, which is why most of the genotypes had flat and wide leaves with broad triangular to rounded distal parts.

Leaf serration was mostly fine to medium, as expected for cultivated sunflower (Miller and Fick, 1997). Leaf auricles were mostly small to medium, with only oil lines (both RHA and HA) having a higher frequency of large auricles that may be a consequence of increasing leaf area.

Plant height is important for production stability, since an improper balance with stem diameter increases the chance of lodging (Škorić, 2012b; Kaya, 2015). Most of the plants in the current study were in the medium height category. The HA lines and the less developed categories were taller than the RHA lines, similar to the findings of Hladni et al. (2017) where plant height did not have high discriminative power because of the high variability in each of the genotype groups.

Branching was one of traits with lowest diversity (0.58), including three out of five categories with <0.7% of total genotypes. On the other hand, discriminative power was high, as

TOC Head: ; Section Head: ; Article Type: ARTICLE

HA lines were clearly separated from branched RHA lines. Branching is determined by dominant genes in wild species, whereas in cultivated lines it is under control of recessive genes (Miller and Fick, 1997). Primarily apically branched RHA-O and RHA-NO genotypes were also visible as a group corresponding most probably to the *b3* recessive gene action (Hockett and Knowles, 1970).

Seed shape and color traits are most important in non-oil confection sunflower breeding, but the current study confirmed that other seed traits can also be used as good descriptors of diversity (Hladni et al., 2017). The diversity index was only low for SCOL. In the current study, seeds were mostly black, small to medium size, with stripes both on margins and intermarginal. Anthocyanin seed color was only found in one RHA-O line as the rarest category, indicating lack of pigments in the first layer (Putt, 1940). Since only confectionary sunflower lines were intentionally selected to have seed stripes together with larger seeds (Miller and Fick, 1997; Škorić, 2012b), there was a significantly higher frequency of lines without seed stripes in both HA-O and RHA-O lines, whereas those two groups also contained the highest number of genotypes with small seeds.

The highest frequency of rounded seed was found in HA and less developed lines, whereas on the other hand, highest frequency of elongated was found in the group of RHA-O lines. That could be a side effect of restorer line selection where seed size and shape were less important than the restorer genes and the desired oil content. Seed size depends on the seed number per head, determined by genotype and environment, whereas a clear difference exists in the 1000-seed weight between oil (40–100 g) and non-oil (>100 g) cultivars (Miller and Fick, 1997; Škorić, 2012b).

Understanding phenological development is important for breeding since cultivated sunflower is under environmental control. Days to flowering and maturity shows how various genotypes react and adapt to specific environments. Earliness is a trait that also implies a changed ideotype (reduced plant height, leaf number, etc.) so that the plant is adapted to achieve an acceptable yield in a shorter time (Connor and Hall, 1997). Similar to other traits, variability of the initial material is very important when selecting for earliness (Škorić, 2012a). Wild species as progenitors of cultivated sunflower have longer vegetative periods and are generally photoperiod sensitive, but occasionally insensitive plants can be found (Rogers et al., 1982; Terzić et al., 2012).

Diversity in the UGA-SAM1 population was very low for days to flowering because >80% of lines were in the medium category (50–59). Only the HA-O category contained one very early genotype with <40 d to flowering. Extension of the preanthesis phase does increase yield (Connor and Hall, 1997). The concentration of genotypes in the medium category is most likely a result of compromise between the desired shorter phase length and less significant yield decrease (Kirichenko, 2005; López Pereira et al., 2008).

Days to maturity trait was specific but at the same time showed high diversity (0.91) and distinctive grouping of genotypes. Medium late genotypes were the most frequent, followed by medium and late. The most uniform distribution and best grouping was found for RHA-O genotypes, which had a significantly higher proportion of early maturing genotypes. The RHA lines are branched, have longer flowering time than single-headed cytoplasmic male sterile

(CMS) lines, but also shorter grain-filling period that results in smaller seed size and weight, all a consequence of optimizing RHA lines for *Rf* genes, and less seed yield (Radić et al., 2013).

Trait Discriminative Power

The HOMALS analysis explained a high proportion of the variability for the first two principal components (PCs) of the biplot (Da Silva and Padovani, 2006). The first component was positively associated with branching type (BRAN), grouping nonbranched genotypes to the negative and branched (mostly restorer lines) to the positive side of PC1. The PC2 was closely related with SSTR and SSCO, grouping lines along the axis from the lower right (black seeds no stripes) to upper left corner of the biplot (striped seed), similar to the findings of Hladni et al. (2017). Discriminative power of BRAN, SSCO, and SSTR is even more obvious when the HOMALS biplot is made using only eight descriptors positioned furthest from the biplot origin. Maturity and seed size descriptors were closer to the biplot origin, but they influenced wider genotype dispersion and facilitated discrimination, as they were positioned on the opposite side of BRAN- and SSTR-related traits.

The test of association provided a clear separation of traits that had more distinctive grouping than HOMALS-based estimation (Bertin, 1981). Besides BRAN, SSCO, and SSTR, the association analysis revealed that MAT, SSIZE, and SSHP were the most discriminative. Discrimination power (association) was negatively correlated with the increase of morphological diversity and more uniform distribution of genotypes. The lowest discrimination power was in the uniform distribution among genotype groups as in LATV, or among both genotype groups and trait categories as in LWIN, LSDIS, and LGCOL. Maturity group had high diversity with close to uniform distribution across trait categories, but pronounced differences between categories in individual genotype groups, most noticeably visible in RHA-O, resulting in high association values. Branching on the other hand, also had high discriminative power, but significantly lower H' , as genotype distribution was not homogenous according to descriptor categories. Like MAT, it was significantly related to genotype groups (HA vs. RHA lines).

Genotype Grouping

The phenotype of sunflower inbred lines depends on the goal of the selection and the desired ideotype (Vear, 2016). Early selection for increased oil content and hybrid sunflower development resulted in recognizable divergent groups, confirmed in the current study. The source materials for the development of sunflower are local populations and landraces together with OPVs and materials introgressed from the crop wild relatives (Atlagić and Terzić, 2014). As the source material, genotypes from the less developed group had the largest diversity and were positioned in the upper left part of the HOMALS biplot. The selection for increased oil content and introduction of another dark seed layer for resistance to sunflower moth (*Homoeosoma electellum*, Hulst) resulted in a separation of HA-O lines with black seeds and no stripes that were high in oil (Carlson et al., 1972). The HA-NO group is positioned together with the less developed group, having mostly lighter color, striped seeds (Dedio, 1995). The position of HA-NO can also be a result of less intensive selection than for the oil type (Hladni, 2016), resulting in a larger proportion of retained diversity for the analyzed traits, as in the less developed material.

TOC Head: ; Section Head: ; Article Type: ARTICLE

Distinctive groups of HA and RHA lines were observed in the current study with the RHA genotypes branched with smaller elongated seeds. The positioning of HA and RHA lines on the opposite sides of the first PC confirmed the desired heterogeneity of parental lines. Similar grouping of RHA and CMS confectionary sunflower lines was found by Hladni et al. (2017). Traits like LBLIS and LSIZ were less variable and less expressed in Rf lines, pointing to increased selection pressure and loss of variability compared with HA lines. On the other hand, all five MAT categories were clearly visible only in RHA-O (and to some extent RHA-NO) group. Such result can be attributed to the requirement for higher yield in HA line seed production and negative correlation of yield and short vegetation cycle (Connor and Hall, 1997). Therefore, HA lines have low percentage of genotypes in the early MAT categories to maximize seed yield, whereas RHA lines were under lower selection pressure for late MAT and probably exhibit different breeder strategies in crosses between CMS and Rf lines (Siniša Jocić, personal communication, February 2018).

Appropriate Descriptor Traits

Sunflower inbred lines are developed from genetically divergent populations including interspecies hybrids and certain gene pools (Škorić, 2012a). During the early generations of inbreeding (S_0 – S_1 or F_2 – F_3), nondesirable material is eliminated based on a set of yield-, plant architecture-, and disease resistance-related traits, where higher diversity of the starting material usually leads to higher proportion of eliminated material. Elimination of lines is also due to lower combining ability, which is tested in F_4 – F_5 again by evaluating desirable traits (Miller, 1987).

The current study confirmed how increased selection pressure increasingly changed population structure (Supplemental Table S2) and resulted in lower variability (Mandel et al., 2011). The loss of variability in breeding programs could be lowered if, in addition to the yield and quality, diversity-focused traits were also included. Leaf traits (LTIP, LAUR, and LBLIS) are relatively easy to evaluate and are among the most informative together with seed coloration traits (SMC, SSTR, and SSCO), even though they are more related to confectionary sunflower breeding programs. Hladni et al. (2017) also found LBLIS informative for genotype discrimination together with floret array and anthocyanin coloration of stigma. Evaluation of genetic diversity based on allelic richness is useful, as it allows expectation of selection limits determined by the initial allelic composition more than by heterozygosity (Foulley and Ollivier, 2006). The lack, or very low frequencies, of genotypes in certain trait categories should be further investigated to determine if it represents a consequence of selection or a natural occurrence requiring trait category modification.

The use of appropriate qualitative traits is important, even though efficient molecular marker selection techniques are now in regular use. Most of the modern breeding programs use several parallel approaches for material evaluation (Cheres et al., 2000). Classical evaluation of combining abilities is the most used method for predicting hybrid sunflower performance (Yao et al., 2013; Reif et al., 2013), although molecular tools are in development for sunflower (Dimitrijevic and Horn, 2018). Phenotype traits are thus used for both early line selection and the choice of hybrid parents.

The current, most widely adopted test for DUS is based mostly on phenotype traits (UPOV, 2000). Improvement of the guidelines involving methods of observation, types of trait expression

TOC Head: ; Section Head: ; Article Type: ARTICLE

(trait categories), etc., is a constant effort led by the Community Plant Variety Office (CPVO). Preparations for the use of biomolecular techniques in DUS testing has started but is still under development. One of the main concerns is the increased distinctness discoverable by genetic analysis of individual plants and not visible by field phenotypic observation. One of the proposed solutions is to “translate” genetic distances to UPOV characteristics and add them to the current DUS guideline with existing phenotype traits (UPOV, 2016). Such a solution should be viable, as it resembles the current state of the art in the evaluation of heterotic potential of parental lines where phenotype analysis after line crosses (general and specific combining ability) is still the most used approach supplemented by molecular marker screening (Darvishzadeh, 2012).

In summary, phenotypic diversity was moderately high for the studied traits and confirmed the UGA-SAM1 as a valuable resource for sunflower research. The highest phenotypic diversity was found in the less developed group, followed by non-oil genotypes, resulting from less intensive selection compared with oil lines and even more compared with the RHA-O group, which had the lowest diversity. On the other hand, the RHA-O group had the highest diversity for specific traits like days to maturity, resulting from the optimization of HA lines for pollination and less intensive exclusion of early genotypes due to lower yields. Some descriptor categories were nonetheless underrepresented, pointing to the possible need to supplement the population when working on specific traits.

Discriminative power based on association between genotype groups and category traits was generally negatively correlated with the increased uniformity in distribution of genotypes among groups and/or trait categories (H'). On the other hand, close-to-uniform distribution across trait categories (H') but pronounced differences between categories in individual genotype groups, most visible for MAT, resulted in both high association values and index of diversity (H'). Branching also had high discriminative power, but significantly lower uniformity (H'), as genotype distribution was not homogenous according to descriptor categories, but instead to genotype groups. The association test proved to be a useful tool, in addition to HOMALS-based analysis, for determining trait discriminative power.

Selection for increased oil content followed by hybrid sunflower development, grouped genotypes to oil vs. non-oil and HA maintainer group vs. RHA restorer group. Overlapping of HA-NO and less developed categories is most likely a result of less intensive selection than for the oil type, resulting in a larger proportion of retained diversity for the analyzed traits. The lowest variability found in the RHA oil group suggests the strongest selection pressure. Nonetheless, RHA-O had the most uniform and distinctive grouping according to MAT categories due to specific breeder requirements for restorer lines. Distinctiveness between trait category groups was also affected by the number of traits used for the HOMALS analysis, signifying the importance of descriptor selection for germplasm evaluation.

Appropriate descriptor evaluation is important both during early inbred line development and during evaluation of combining abilities. The loss of variability in breeding programs could be lowered if, besides focusing on yield and quality, diversity-focused traits were used including leaf, seed, and certain flower traits. Efficient molecular techniques are now in regular use, but sunflower breeding programs are still based on phenotyping with markers mostly used for disease resistance testing. The combined use of pheno- and genotyping seems to also be the

TOC Head: ; Section Head: ; Article Type: ARTICLE

future of UPOV DUS test, where genetic distances translated to UPOV characteristics are proposed as an addition to existing phenotype traits.

Supplemental Material

Supplemental material is available online for this article.

Conflict of Interest

The authors declare that there is no conflict of interest.

ACKNOWLEDGMENTS

We thank USDA-ARS-NCRPIS for providing seeds of the UGA-SAM1 population. We also thank the reviewers for their valuable suggestions and help with improving our manuscript. The financial support from the Institute of Field and Vegetable Crops in Novi Sad, the Ministry of Science and Education of Republic of Serbia (TR31025), and the Provincial Secretariat for Higher Education and Science of Vojvodina (114-451-2126/2016-03) are greatly appreciated.

REFERENCES

- Álvarez, M.F., T. Mosquera, and M.W. Blair. 2014. The use of association genetics approaches in plant breeding. *Plant Breed. Rev.* 38:17–68. doi:10.1002/9781118916865.ch02
- Atlagić, J., and S. Terzić. 2014. Sunflower genetic resources: Interspecific hybridization and cytogenetics in prebreeding In: J.I. Arribas, editor, *Sunflowers: Growth and development, environmental influences and pests/diseases*. Nova Sci. Publ., New York. p. 95–130.
- Baack, E.J., Y. Sapir, M.A. Chapman, J.M. Burke, and L.H. Rieseberg. 2008. Selection on domestication traits and quantitative trait loci in crop–wild sunflower hybrids. *Mol. Ecol.* 17:666–677. doi:10.1111/j.1365-294X.2007.03596.x
- Bertin, J. 1981. *Graphics and Graphic Information Processing*. Translated by W.J. Berg and P. Scott. Walter de Gruyter, Berlin. doi:10.1515/9783110854688
- Bowers, J.E., E. Bachlava, R.L. Brunick, L.H. Rieseberg, S.J. Knapp, and J.M. Burke. 2012. Development of a 10,000-locus genetic map of the sunflower genome based on multiple crosses. *G3: Genes, Genomes, Genet.* 2:721–729. doi:10.1534/g3.112.002659
- Breton, C., H. Serieys, and A. Bervillé. 2010. Gene transfer from wild *Helianthus* to sunflower: Topicalities and limits. *OCL* 17:104–114. doi:10.1051/ocl.2010.0296
- Brothers, M.E., and J.F. Miller. 1999. Core subset for the cultivated sunflower collection. In: L. Kroh, editor, *Proceedings of the 21st Sunflower Research Workshop*, Fargo, ND. 14–15 June. 1999. Natl. Sunflower Assoc., Bismarck, ND. p. 124–127.
- Brown, A.H.D. 1995. The core collection at the crossroads. In: T. Hodgkin, et al., editors, *Core collections of plant genetic resources*. John Wiley & Sons, UK. p. 3–9.
- Carlson, E.C., P.F. Knowles, and J.E. Dillb. 1972. Sunflower varietal resistance to sunflower moth larvae. *Calif. Agric.* 26:11–13.
- Cavalli-Sforza, L.L., and W.F. Bodmer. 1981. *Genética de las Poblaciones Humanas*. Ed. Omega, Barcelona.
- Cheres, M.T., J.F. Miller, J.M. Crane, and S.J. Knapp. 2000. Genetic distance as a predictor of heterosis and hybrid performance within and between heterotic groups in sunflower. *Theor. Appl. Genet.* 100:889–894. doi:10.1007/s001220051366
- Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. L. Erlbaum Assoc., Hillsdale, NJ.

TOC Head: ; Section Head: ; Article Type: ARTICLE

- Connor, D.J., and A.J. Hall. 1997. Sunflower physiology. In: A.A. Schneiter, editor, Sunflower technology and production. Agron. Monogr. 35. ASA, CSSA, SSSA, Madison, WI. p. 113–182. doi:10.2134/agronmonogr35.c4
- Da Silva, N.R., and C.R. Padovani. 2006. Utilização de componentes principais em experimentação agrônômica. Energ. Agric. 21:98–113.
- Darvishzadeh, R. 2012. Phenotypic and molecular marker distance as a tool for prediction of heterosis and F₁ performance in sunflower (*Helianthus annuus* L.) under well-watered and water-stressed conditions. Aust. J. Crop Sci. 6:732–738.
- de Jesus Vernetti, F., and F. de Jesus Vernetti, Jr. 2017. Qualitative traits in breeding. In: F. Lopes da Silva, et al., editors, Soybean breeding. Springer, Cham, Switzerland. p. 55–79. doi:10.1007/978-3-319-57433-2_5
- Dedio, W. 1995. Association of sunflower achene color with other achene characters and bird preference. Can. J. Plant Sci. 75:377–380. doi:10.4141/cjps95-063
- Dimitrijevic, A., and R. Horn. 2018. Sunflower hybrid breeding: From markers to genomic selection. Front. Plant Sci. 8:2238. doi:10.3389/fpls.2017.02238
- Fehr, W.R. 1987. Principles of cultivar development. Vol. 1 Theory and technique. Macmillan, New York.
- Fick, G.N., and J.F. Miller. 1997. Sunflower breeding. In: A.A. Schneiter, editor, Sunflower technology and production. Agron. Monogr. 35. ASA, CSSA, and SSSA, Madison, WI. p. 395–439. doi:10.2134/agronmonogr35.c8
- Foulley, J.L., and L. Ollivier. 2006. Estimating allelic richness and its diversity. Livest. Sci. 101:150–158. doi:10.1016/j.livprodsci.2005.10.021
- Frankel, O.H. 1984. Genetic perspectives of germplasm conservation. In: W. Arber, et al., editors, Genetic manipulation: Impact on man and society. Cambridge Univ. Press, Cambridge, UK. p. 161–170.
- Hladni, N. 2016. Present status and future prospects of global confectionery sunflower production. In: Y. Kaya and S. Hasancebi, editors, Proceedings of the 19th International Sunflower Conference, Edirne, Turkey. 29 May–3 June 2016. Int. Sunflower Assoc., Paris. p. 45–59.
- Hladni, N., S. Terzić, B. Mutavdžić, and M. Zorić. 2017. Classification of confectionary sunflower genotypes based on morphological characters. J. Agric. Sci. 155:1594–1609. doi:10.1017/S0021859617000739
- Hockett, E.A., and P.F. Knowles. 1970. Inheritance of branching in sunflowers, *Helianthus annuus* L. Crop Sci. 10:432–436. doi:10.2135/cropsci1970.0011183X001000040036x
- Hübner, S., N. Bercovich, M. Todesco, J. R. Mandel, J. Odenheimer, E. Ziegler, et al. 2019. Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. Nat. Plants 5:54–62. doi:10.1038/s41477-018-0329-0
- IBPGR. 1985. Sunflower descriptors. Int. Board Plant Genet. Resour., Rome.
- Kaya, Y. 2015. Sunflower. In: S.K. Gupta, editor, Breeding oilseed crops for sustainable production. Elsevier, London. p. 55–82.
- Kirichenko, V.V. 2005. Sunflower breeding and seed production. (In Russian.) Inst. Field Crops, Kharkov, Ukraine. p. 1–385.
- Knupffer, H., and Th.J.L. van Hintum. 1995. The barley core collection: An international effort. In: T. Hodgkin, et al., editors, Core collections of plant genetic resources. John Wiley & Sons, New York. p. 71–78.
- López Pereira, M., A. Berney, A.J. Hall, and N. Trapani. 2008. Contribution of preanthesis photo assimilates to grain yield: Its relationship with yield in Argentine sunflower cultivars released between 1930 and 1995. Field Crops Res. 105:88–96. doi:10.1016/j.fcr.2007.08.002
- Mandel, J.R., J.M. Dechaine, L.F. Marek, and J.M. Burke. 2011. Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, *Helianthus annuus* L. Theor. Appl. Genet. 123:693–704. doi:10.1007/s00122-011-1619-3

TOC Head: ; Section Head: ; Article Type: ARTICLE

- Mandel, J.R., S. Nambeesan, J.E. Bowers, L.F. Marek, D. Ebert, L.H. Rieseberg, et al. 2013. Association mapping and the genomic consequences of selection in sunflower. *PLoS Genet.* 9:e1003378. doi:10.1371/journal.pgen.1003378
- McKhann, H.I., C. Camilleri, A. Bérard, T. Bataillon, J.L. David, X. Reboud, et al. 2004. Nested core collections maximizing genetic diversity in *Arabidopsis thaliana*. *Plant J.* 38:193–202. doi:10.1111/j.1365-313X.2004.02034.x
- Michailidis, G., and J. De Leeuw. 1998. The Gifi system of descriptive multivariate analysis. *Stat. Sci.* 13:307–336. doi:10.1214/ss/1028905828
- Miller, J.F. 1987. Sunflower. In: W.R. Fehr, editor, *Principles of cultivar development*. Vol. 2. Macmillan Publ. Co., New York. p. 626–668.
- Miller, J.F., and G.N. Fick. 1997. The genetics of sunflower. In: A.A. Schneiter, editor, *Sunflower technology and production*. Agron. Monogr. 35. ASA, CSSA, and SSSA, Madison, WI. p. 441–495. doi:10.2134/agronmonogr35.c9
- Miller, J.F., G.J. Seiler, and C.C. Jan. 1992. Introduced germplasm use in sunflower inbred and hybrid development. In: H.L. Shands and L.E. Weisner, editors, *Use of plant introductions in cultivar development*, Part 2. CSSA, Madison, WI. p. 151–166.
- Moreno, M.V., V. Nishinakamasu, M.A. Loray, D. Alvarez, J. Giéco, A. Vicario, et al. 2013. Genetic characterization of sunflower breeding resources from Argentina: Assessing diversity in key open-pollinated and composite populations. *Plant Genet. Resour.* 11:238–249. doi:10.1017/S1479262113000075
- Müller, B.U., A. Schutzenmeister, and H.-P. Piepho. 2010. Arrangement of check plots in augmented block designs when spatial analysis is used. *Plant Breed.* 129:581–589. doi:10.1111/j.1439-0523.2010.01803.x
- Nambeesan, S.U., J.R. Mandel, J.E. Bowers, L.F. Marek, D. Ebert, J. Corbi, et al. 2015. Association mapping in sunflower (*Helianthus annuus* L.) reveals independent control of apical vs. basal branching. *BMC Plant Biol.* 15:84. doi:10.1186/s12870-015-0458-9
- Neumann, K., B. Kobiljski, S. Denčić, R.K. Varshney, and A. Borner. 2011. Genome-wide association mapping: A case study in bread wheat (*Triticum aestivum* L.). *Mol. Breed.* 27:37–58. doi:10.1007/s11032-010-9411-7
- Panković, D., Z. Sakač, M. Plesničar, T. Čupina, and D. Škorić. 1991. Leaf expansion and photosynthesis during growth and development of NS sunflower hybrids and inbred lines. *Helia* 4:55–62.
- Putt, E.D. 1940. Observation on morphological characters and flowering process in sunflower (*Helianthus annuus* L.). *Sci. Agric.* 21:167–169.
- R Core Team. 2018. R: A language and environment for statistical computing. R Found. Stat. Comput., Vienna. <https://www.r-project.org/> (accessed 10 Sept. 2019).
- Radić, V., J. Mrđa, M. Jocković, P. Čanak, A. Dimitrijević, and S. Jocić. 2013. Sunflower 1000-seed weight as affected by year and genotype. *Ratar. povrt.* 50:1–7. doi:10.5937/ratpov50-3214
- Reif, J.C., Y. Zhao, T. Würschum, M. Gowda, and V. Hahn. 2013. Genomic prediction of sunflower hybrid performance. *Plant Breed.* 132:107–114. doi:10.1111/pbr.12007
- Rogers, C.E., T.E. Thompson, and G.J. Seiler. 1982. *Sunflower species of the United States*. Natl. Sunflower Assoc., Bismarck, ND.
- Shannon, C.E., and W. Weaver. 1949. *The mathematical theory of communication*. Univ. Illinois Press, Urbana.
- Schneiter, A.A., and J.F. Miller. 1981. Description of sunflower growth stages. *Crop Sci.* 21:901–903. doi:10.2135/cropsci1981.0011183X002100060024x
- Schwanck, A.A., S. Savary, P. Debaeke, P. Vincourt, and L. Willocquet. 2016. Effects of plant morphological traits on phoma black stem in sunflower. *Eur. J. Plant Pathol.* 145:345–361. doi:10.1007/s10658-015-0848-8

TOC Head: ; Section Head: ; Article Type: ARTICLE

- Škorić, D. 2012a. Sunflower breeding. In: D. Škorić and Z. Sakač, editors, Sunflower genetics and breeding. Int. Monogr. Serbian Acad. Sci., Novi Sad, Republic of Serbia. p. 165–354.
- Škorić, D. 2012b. The genetics of sunflower. In: D. Škorić and Z. Sakač, editors, Sunflower genetics and breeding. Int. Monogr. Serbian Acad. Sci., Novi Sad, Republic of Serbia. p. 1–164.
- Tang, S., J.K. Yu, B. Slabaugh, K. Shintani, and J. Knapp. 2002. Simple sequence repeat map of the sunflower genome. *Theor. Appl. Genet.* 105:1124–1136. doi:10.1007/s00122-002-0989-y
- Terzić, S., J. Atagić, I. Maksimović, T. Zeremski, S. Petrović, and B. Dedić. 2012. Influence of photoperiod on vegetation phases and tuber development in topinambour (*Helianthus tuberosus* L.). *Arch. Biol. Sci.* 64:175–182. doi:10.2298/ABS1201175A
- UPOV. 2000. Guidelines for the conduct of tests for distinctness, uniformity and stability: Sunflower (*Helianthus annuus* L.). TG/81/6. Int. Union Prot. New Varieties Plants, Geneva. <http://www.upov.int/edocs/tgdocs/en/tg081.pdf> (accessed 10 Sept. 2019).
- UPOV. 2016. Report by the Community Plant Variety Office (CPVO) to the UPOV BMT meeting in Moscow, 24–27 May 2016 on CPVO IMODDUS. Int. Union Prot. New Varieties Plants, Geneva. http://www.upov.int/edocs/mdocs/upov/en/bmt_15/bmt_15_27.pdf (accessed 10 Sept. 2019).
- Van Hintum, Th.J.L., A.H.D. Brown, C. Spillane, and T. Hodgkin. 2000. Core collections of plant genetic resources. *Tech. Bull. 3. Int. Plant Genet. Resour. Inst., Rome.*
- Vear, F. 2016. Changes in sunflower breeding over the last fifty years. *OCL* 23:D202. doi:10.1051/ocl/2016006
- Whitney, K.D., K.W. Broman, N.C. Kane, S.M. Hovick, R.A. Randell, and L.H. Rieseberg. 2015. Quantitative trait locus mapping identifies candidate alleles involved in adaptive introgression and range expansion in a wild sunflower. *Mol. Ecol.* 24:2194–2211. doi:10.1111/mec.13044
- Yang, D., K.J. Niklas, S. Xiang, and S. Sun. 2010. Size-dependent leaf area ratio in plant twigs: Implication for leaf size optimization. *Ann. Bot.* 105:71–77. doi:10.1093/aob/mcp262
- Yao, W.H., Y.D. Zhang, M.S. Kang, H.M. Chen, L. Liu, L.J. Yu, and X.M. Fan. 2013. Diallel analysis models: A comparison of certain genetic statistics. *Crop Sci.* 53:1481–1490. doi:10.2135/cropsci2013.01.0027

FIGURE 1. Morphological diversity index (H') compared with the chi-square association test for independence of descriptor categories and genotype groups. Higher contingency coefficient indicates a higher association. C.standarized, standardized coefficient. For definitions of morphological descriptors, see Table 1.

FIGURE 2. Discriminative power of (a) all analyzed traits and (b) eight traits furthest from the biplot origin. For definitions of morphological descriptors, see Table 1.

FIGURE 3. Association between descriptor categories and genotype groups for BRAN (branching). Less dev. (less developed), RHA-NO (restorer [Rf] line, non-oil), HA-O (maintainer [B] line, oil), HA-NO (maintainer [B] line, non-oil), P. basal (predominantly basal), P. apical (predominantly apical).

FIGURE 4. Association between descriptor categories and genotype groups for SSCO (seed stripe color). Less dev. (less developed), RHA-NO (restorer [Rf] line, non-oil), HA-O (maintainer [B] line, oil), HA-NO (maintainer [B] line, non-oil).

FIGURE 5. Association between descriptor categories and genotype groups for MAT (maturity group). Less dev. (less developed), RHA-NO (restorer [Rf] line, non-oil), HA-O (maintainer [B] line, oil), HA-NO (maintainer [B] line, non-oil).

TOC Head: ; Section Head: ; Article Type: ARTICLE

FIGURE 6. Association between descriptor categories and genotype groups for SSIZE (seed size). Less dev. (less developed), RHA-NO (restorer [Rf] line, non-oil), HA-O (maintainer [B] line, oil), HA-NO (maintainer [B] line, non-oil).

FIGURE 7. Homogeneity analysis in R (HOMALS) genotype distribution plot based on 20 analyzed morphological traits. LD (linkage disequilibrium), Less dev. (less developed), RHA-NO (restorer [Rf] line, non-oil), HA-O (maintainer [B] line, oil), HA-NO (maintainer [B] line, non-oil).

TABLE 1. Morphological descriptor categories and codes, genotype distribution and diversity index of a UGA-SAM1 population.

No.	Descriptor	Trait code	IBPGR† trait code	UPOV‡ trait code	Category names, ranges, and un
Vegetation					
1	Flowering time	FLO	IBPGR 4.2.1 continuous	Trait 14 (1,3,5,7,9)	1, very early, <40 d 3, medium early, 40–49 d 5, medium, 50–59 d 7, medium late, 60–69 d 9, very late, >70 d
2	Maturity groups	MAT	IBPGR 4.2.3 continuous	–	1, early, <90 d 3, medium early, 90–99 d 5, medium, 100–109 d 7, medium late, 110–119 d 9, late, >120 d
Leaf					
3	Leaf size (length × width)	LSIZ	IBPGR 4.1.1 (1,3,5,7,9)	Trait 3 (3,5,7)	3, <500 cm ² 5, 500–1000 cm ² 7, >1000 cm ²
4	Petiole length	PET	IBPGR 6.1.17: (0,1,5,9)	–	1, <10 cm 3, 10–14 cm 5, 15–19 cm 7, 20–25 cm 9, >25 cm
5	Lateral veins angle	LATV	IBPGR 6.1.15 (3,5,7)	Trait 11 (1,2,3)	1, acute 2, right 3, obtuse
6	Height of the tip of the leaf blade compared to insertion of petiole	LTIP	IBPGR 6.1.16: (1,3,5,7,9)	Trait 12 (3,5,7)	1, extremely low 2, low 3, medium

TOC Head: ; Section Head: ; Article Type: ARTICLE

					4, high 5, extremely high
7	Leaf green color	LGCOL	IBPGR 6.1.7 (3,5,7)	Trait 4 (3,5,7)	3, light 5, medium 7, dark
8	Leaf blistering	LBLIS	IBPGR 6.1.10 (1,3,5,7,9)	Trait 5 (1,3,5,7)	1, extremely weak 3, weak 5, medium 7, strong 9, extremely strong
9	Leaf serration	LSERR	(IBPGR 6.1.11-Leaf margin)	Trait 6 (1,3,5,7,9)	1, very fine 3, fine 5, medium 7, coarse 9, very coarse
10	Leaf, shape of cross-section	LCROS	IBPGR 6.1.13 (3,5,7) Concave, flat, convex	Trait 7 (1,2,3,4,5)	1, strongly concave 3, weakly concave 5, flat 7, weakly convex
11	Leaf, shape of distal part	LSDIS	(IBPGR 6.1.5-leaf shape)	Trait 8 (1–9)	1, lanceolate 3, narrow triangular 5, broad triangular 7, broad triangular rounded 9, rounded
12	Leaf, auricles	LAUR	(IBPGR 6.1.14-leaf base)	Trait 9 (1,3,5,7,9)	1, none or very small 3, small 5, medium 7, large 9, very large
13	Leaf, wings	LWIN	–	Trait 10 (1,2,3)	1, none 2, weak 3, strong expression
Stem					
14	Plant height	PHE	IBPGR 4.1.3 (1–9)	Trait 28 1,3,5,7	2, 40–80 cm 3, 80–120 cm 4, 120–160 cm 5, 160–200 cm 6, >200 cm

TOC Head: ; Section Head: ; Article Type: ARTICLE

15	Branching	BRAN	IBPGR 4.2.9–10 (0,1,2,3,4)	Traits 29, 30 0,1,2,3,4,5	0, none 1, only basal 2, pred. basal 3, overall 4, pred. apical 5, only apical
Seed					
16	Seed color	SCOL	IBPGR 4.3.1 (1–5)	Trait 38 (1–8)	1, white 2, gray 3, brown 4, black 5, anthocyanin
17	Seed stripes	SSTR	IBPGR 6.3.6 (1,2,3)	Traits 39 and 40	0, none 1, marginal 2, lateral 3, both
18	Seed stripes color	SSCO	IBPGR 4.3.3 (1,2,3)	Trait 41 1–4	0, none 1, white 2, gray 3, brown 4, black
19	Seed size	SSIZE	IBPGR 6.3.1 (3,5,7)	Trait 35 3,5,7,9	1, <9 mm 3, 9–10 mm 5, 11–12 mm 7, 13–14 mm 9, >14 mm
20	Seed shape	SSHP	IBPGR 6.3.2 (1,2,3,4)	Trait 36 1,2,3,4	1, elongated 2, narrow ovoid 3, broad ovoid 4, rounded

† IBPGR, International Board for Plant Genetic Resources.

‡ UPOV, International Union for the Protection of New Varieties of Plants.

§ *H* index, The Shannon–Weaver diversity index

TABLE 2. The chi-square test of independence presenting association between evaluated categorical variables (as described in the Materials and Methods section).

Trait‡	χ^2	<i>P</i>	Variable†				
			<i>C</i> _{obs}	<i>C</i> _{max}	<i>C</i> _{stand}	<i>w</i> _{<i>C</i>obs}	<i>w</i> _{<i>C</i>stand}
FLO	27.85	.113	.298	.904	.330	.312	.349
MAT	95.02	.000	.500	.904	.553	.577	.664
LSIZ	25.08	.005	.284	.863	.329	.296	.348
PET	25.22	.193	.285	.904	.315	.297	.332

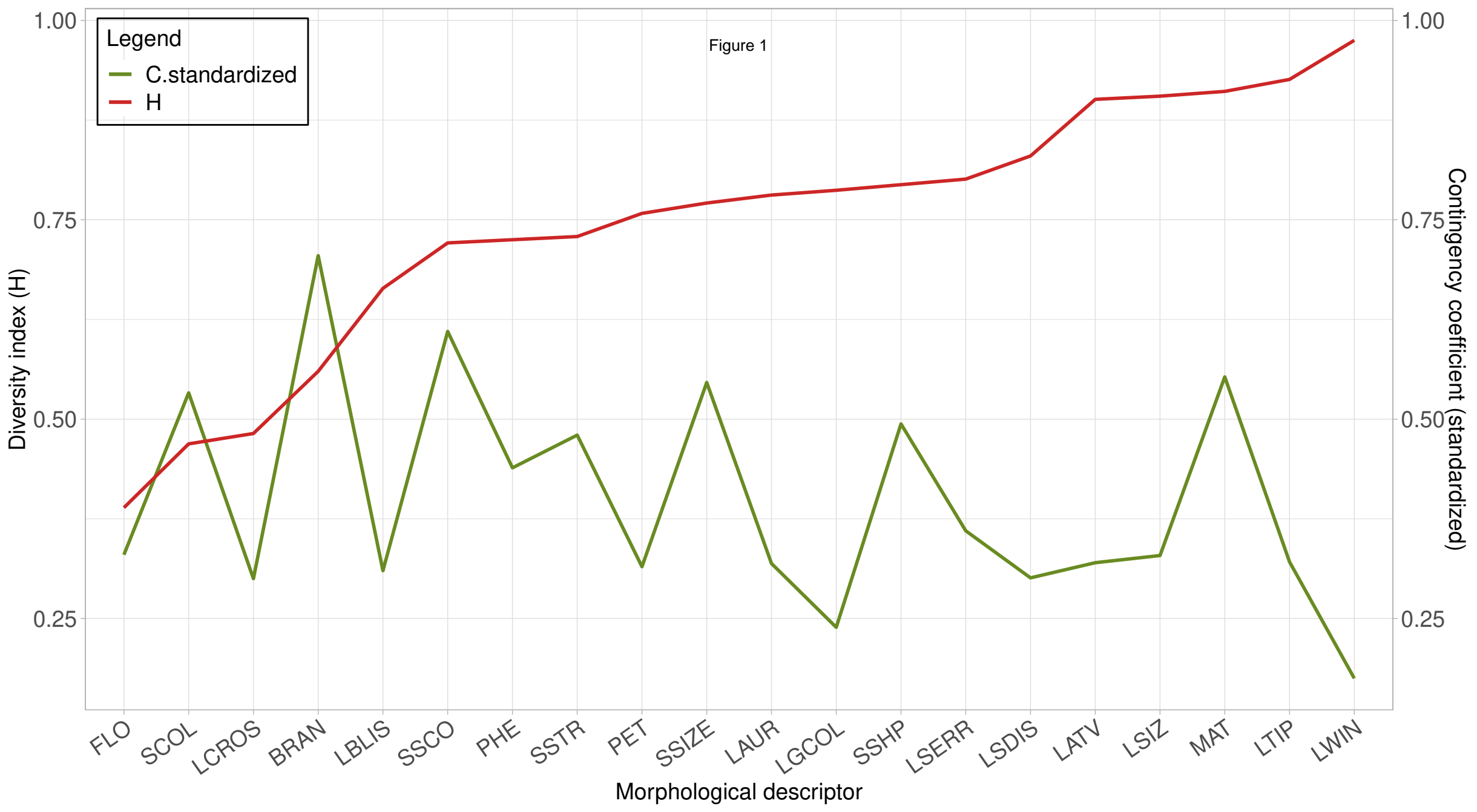
<txtPII>

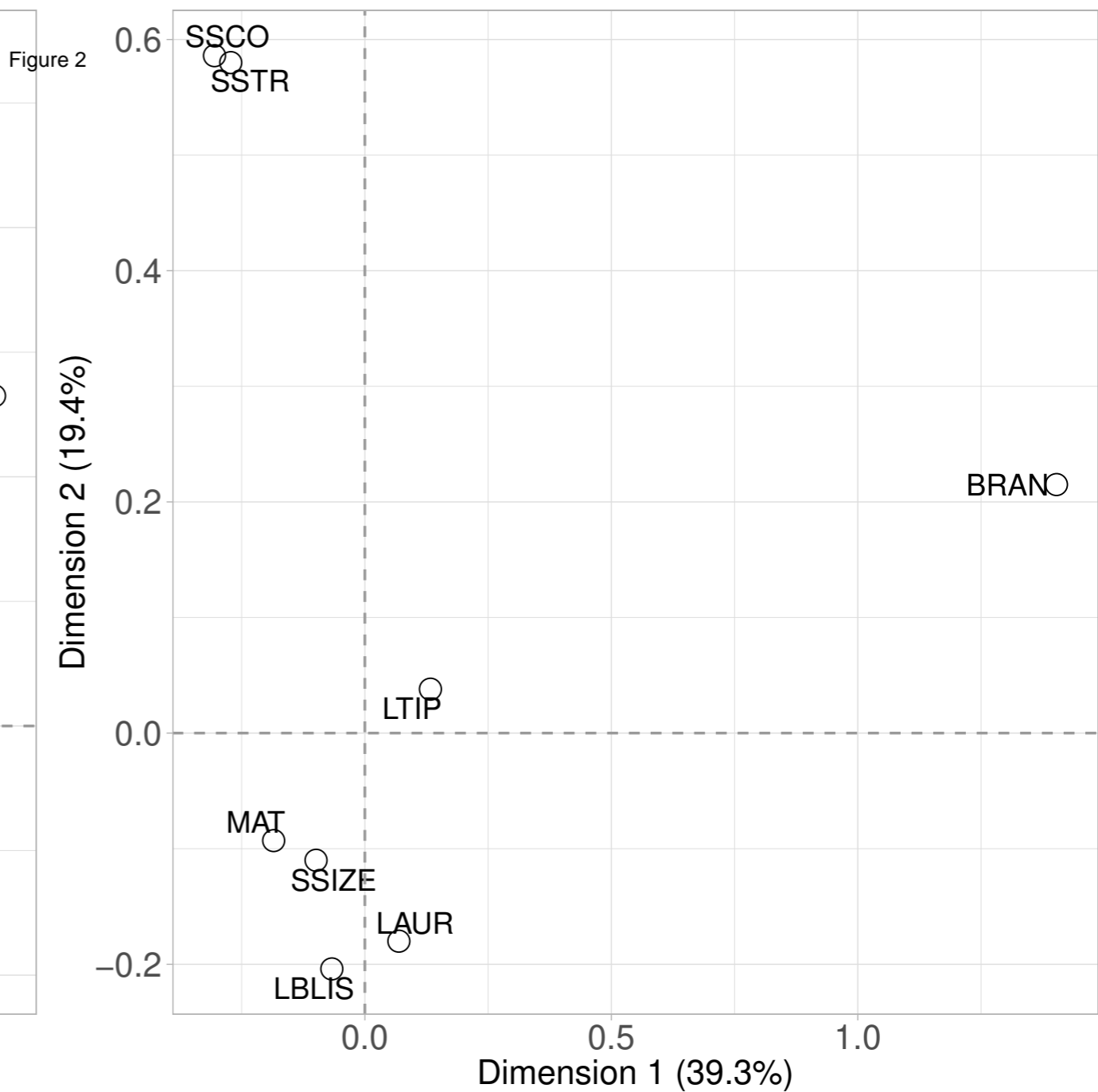
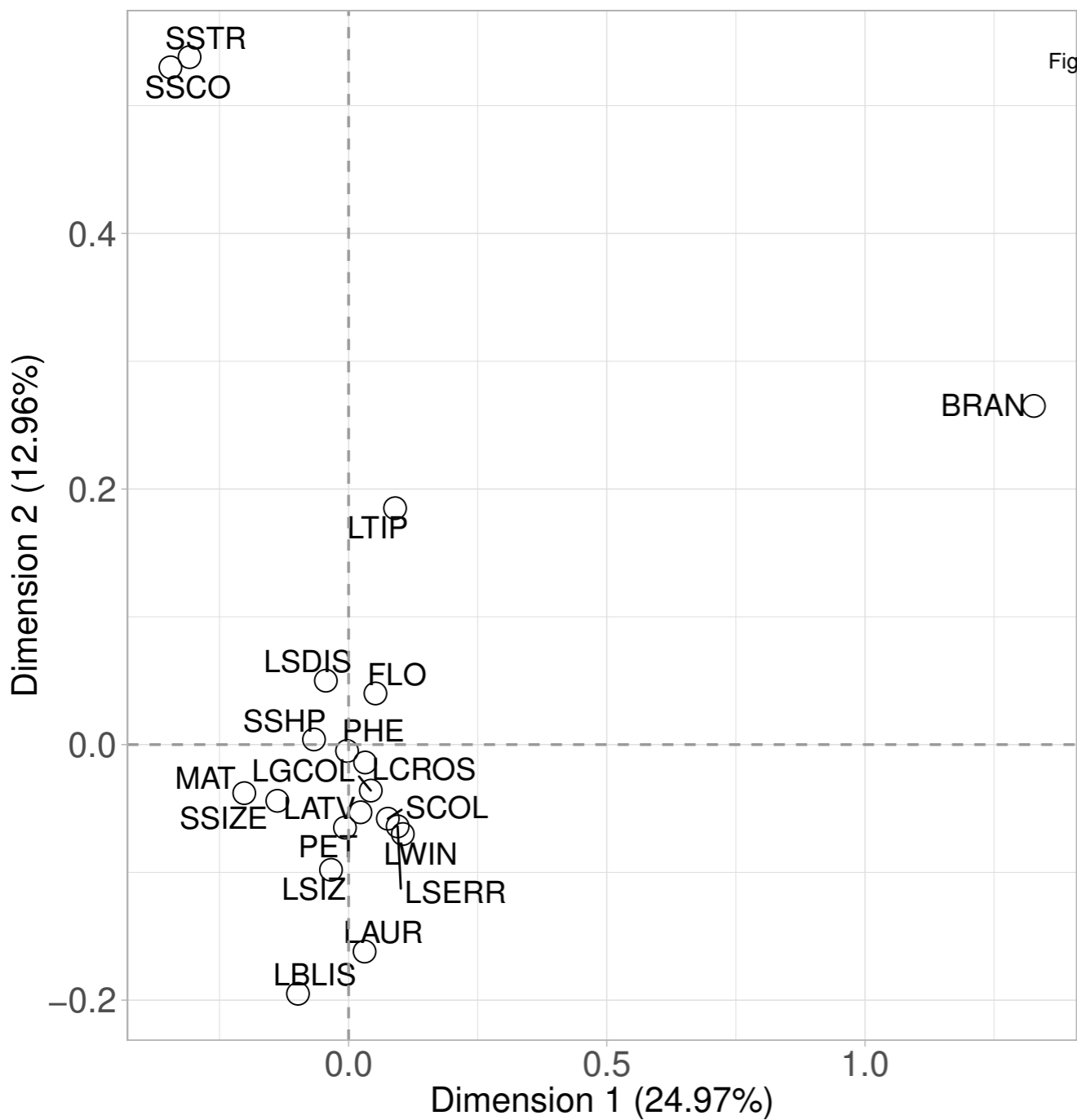
TOC Head: ; Section Head: ; Article Type: ARTICLE

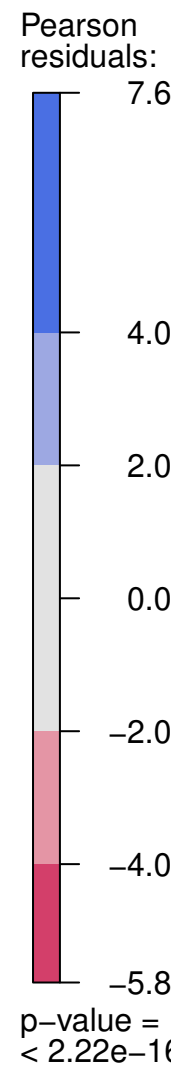
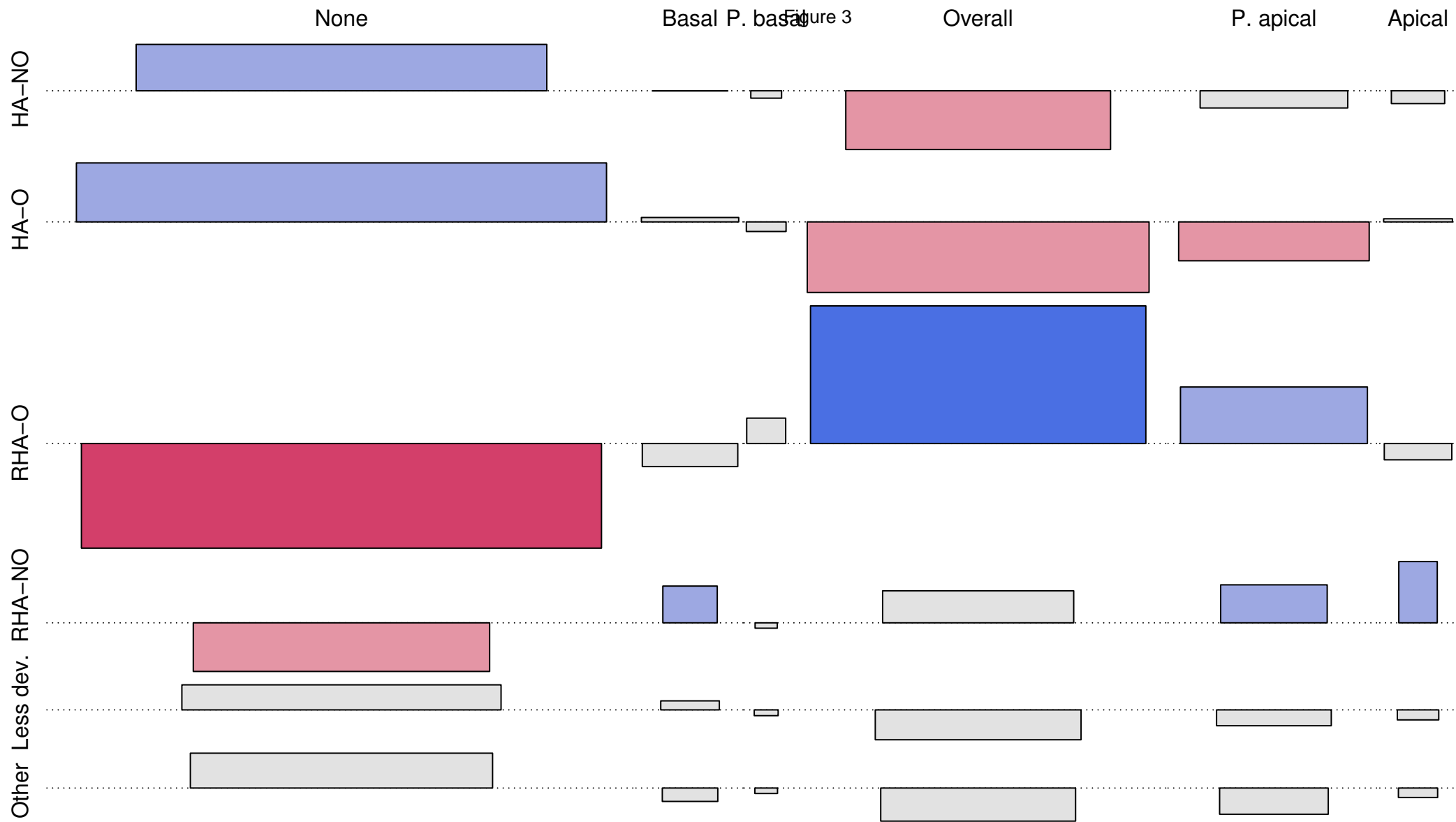
LATV	23.42	.009	.276	.863	.320	.287	.337
LTIP	26.12	.162	.290	.904	.321	.303	.339
LGCOL	12.61	.246	.206	.863	.239	.211	.246
LBLIS	24.31	.229	.280	.904	.310	.292	.326
LSERR	33.55	.029	.325	.904	.360	.344	.385
LCROS	22.64	.307	.271	.904	.300	.282	.314
LSDIS	22.79	.299	.272	.904	.301	.283	.316
LAUR	25.83	.171	.288	.904	.319	.301	.336
LWIN	6.64	.759	.151	.863	.175	.153	.178
PHE	53.45	.000	.397	.904	.439	.433	.489
BRAN	201.47	.000	.644	.913	.705	.842	.995
SCOL	86.10	.000	.482	.904	.533	.550	.631
SSTR	63.577	.000	.427	.889	.480	.472	.548
SSCO	124.38	.000	.551	.904	.610	.660	.769
SSIZE	91.575	.000	.493	.904	.546	.567	.651
SSHP	67.993	.000	.439	.889	.494	.489	.568

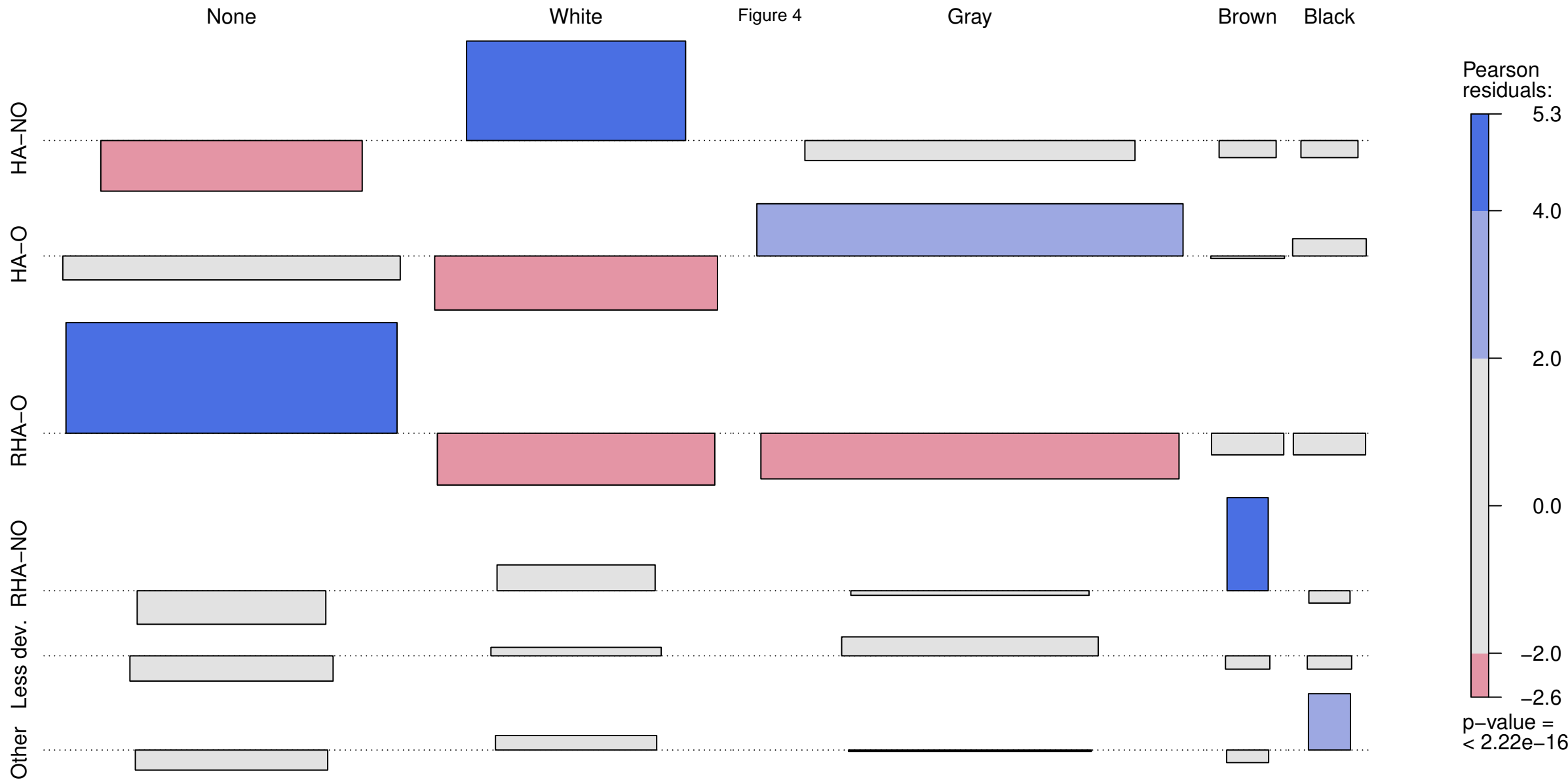
† C_{obs} , observed contingency coefficient; C_{max} , maximum contingency coefficient; C_{stand} , standardized coefficient;
 $w_{C_{obs}}$, W index of effect size based on the value of C_{obs} ; $w_{C_{stand}}$, W index of effect size based on the value of C_{stand} .

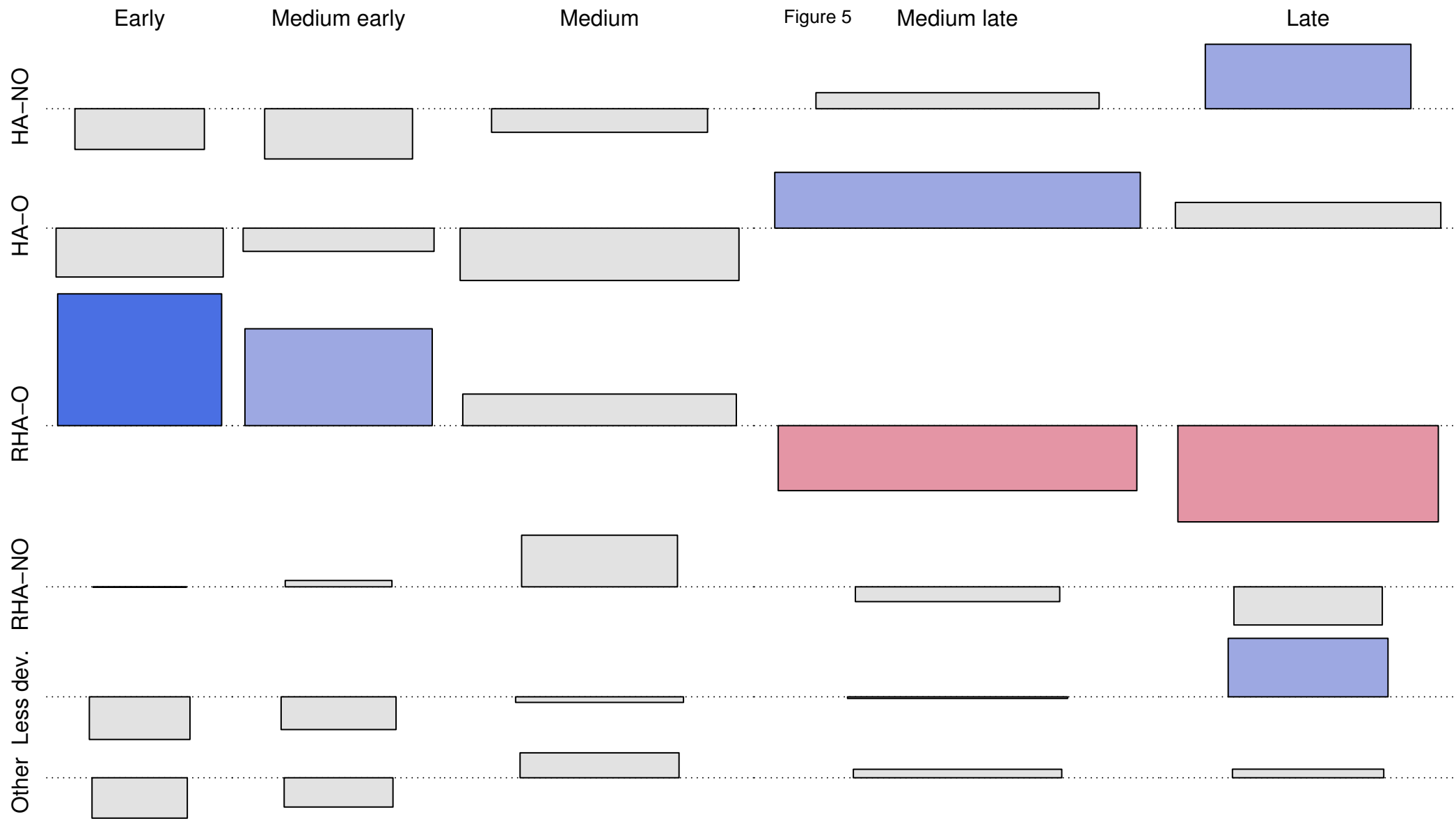
‡ For definitions of morphological descriptors, see Table 1.



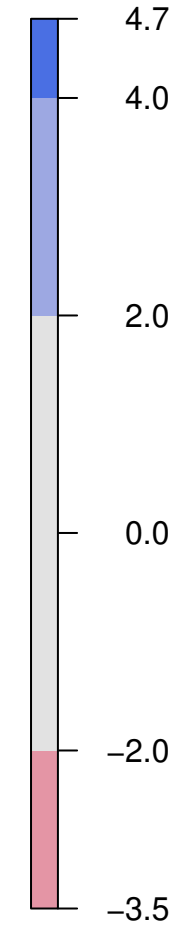








Pearson residuals:



p-value = 9.7082e-12

