

Leveraging phenotypic diversity in a worldwide panel from the Korean genebank to identify new resources for multi-purposes sesame breeding

Yedomon Ange Bovys Zoclanclounon^{1,2}, Sung-Up Kim³, Hwa-Jin Jung¹, Senthil Kumar Thamil Arasan¹, Marème Niang⁴, Daniel Fonceka⁴, Ndjido Ardo Kane⁴, Ndiaga Cissé⁴, Nam-Jin Chung² & Keunpyo Lee¹

¹Genomics Division, National Institute of Agricultural Sciences, RDA, Jeonju, 54874, Republic of Korea

²Department of Crop Science and Biotechnology, Chonbuk National University, Jeonju 54896, Republic of Korea

³Department of Southern Area Crop, National Institute of Crop Science, RDA, Miryang, 50424, Republic of Korea

⁴Centre d'Etudes Régional pour l'Amélioration de l'Adaptation à la Sécheresse (CERAAS), BP 3320 Route de Khombole, Thiès, Sénégal

Corresponding author: Keunpyo Lee, kplee@korea.kr

Abstract

Introduction

Over the last few years, global hunger has begun a rising challenge in the world due to the increasing population growth. By 2050, the global population is expected to exceed 9 billion, increasing the food demand about 70%¹. When coupled with the current adverse effects of climate change, the Zero Hunger objective of the Food and Agricultural Organization (FAO) is seriously jeopardized with the addition of more than 80 million of undernourished people in the recent COVID-19 pandemic context². The projections of the climate change variability on agricultural sector threaten the attainment of the food security and poverty reduction ambitions in developing countries^{3,4}. Therefore, there is an urgent need to develop and deploy crops that combine high yield, nutritional values and strong ability to grow in harsh environments. Besides, crops that can improve the human health through the alleviation of diseases are getting more importance⁵. The example of the barley cultivar BARLEYTM revealed the presence of resistant starch metabolite that has positive effects on the alleviation of the type-2 diabetes and coronary heart diseases⁶.

Considered as an orphan crop⁷, sesame (*Sesamum indicum*) is a nutritional food⁸ and an excellent source of lignans⁹ that showed a wide range of benefits for human health including lowering blood cholesterol and glucose, cardiovascular disease prevention¹⁰, tumor growth suppression¹¹, anti-carcinogenic properties¹² and metabolic syndrome alleviation¹³. Lignans become marketable compounds with a high economical value estimated to \$351.6 million dollars in 2019. The lignans market might approximate more than \$500 million dollars by 2027 ([Link 1](#)). This interest for lignans is noticeable with a growing number of patents in healthy food additives as well as skin care sectors^{14–17}.

Genebank is a reservoir of genetic diversity allowing the identification of promising sources for crop improvement¹⁸. South Korea has the second largest sesame genebank in the world with **7,698** accessions. At the early stage of the Korean genomic-assisted sesame breeding programme, the selection of valuable high-quality nutritional and health-beneficial materials is a crucial prerequisite.

From the Korean genbank, sesame core collection development using 2,751 worldwide accessions, was previously performed by including only one African country¹⁹. As broad geographic coverage is, genetic resources origin can provide a heuristic view for detection of novel desirable traits. Therefore, we enlarge geographic basis of genotypes covering 35 countries throughout the world by reaching 22 African countries (**Supplementary Table S1**). Despite, the characterization of some African accessions from the Chinese genebank²⁰ for

lipid content, the two major lignans *viz* sesamin and sesamolin assessment have been neglected.

The present study aimed to i) assess the variability of agronomic traits in a worldwide sesame panel, ii) infer a core collection, iii) screen the core collection for sesamolin, sesamin, oil and protein contents. Ultimately, this research global goal is to provide relevant resources that can serve as starting materials for the Korean genomic-assisted breeding initiative for boosting sesame yield, nutritional quality and health benefits.

Materials and methods

Plant material and field experiment

A total number of 507 accessions (**Supplementary Table S1**) provided by the Korean genebank were tested during summer season 2018 (May-September). The selected plant materials contained a total of **10** landraces, **10** cultivars and **three** wild relatives (*Sesamum alatum*, *Sesamum schinzianum* and *Sesamum radiatum*). The experiment was laid out following Federer's augmented design²¹ with 8 blocks at Jeonju (35° 49' 50.37''N latitude, 127° 3' 52.79''E longitude) and Miryang (35° 29' 29.70''N latitude, 128° 44' 31.98''E longitude). The inter-row and inter-plant distances were 0.7m and 0.2m respectively. The length of a row was 1.4 m with eight plants. During the experiment, 14 quantitative and five qualitative traits (**Table 1**) were measured on 5 randomly-selected healthy plants. However, for flowering and days to maturity parameters, biomass and yield data were recorded on the unit plot basis. NPK fertilization was provided by the ratio 2.9:3.1:3.2 kg per acre. The recommended cultivation practices were followed during the experiment.

Lipid extraction and quantification

Sesamin and sesamolin quantification

Data analysis

The collected data were fully analyzed with the open-source statistical software R v.4.0.2²². For reproducibility purpose, the data as well as the accompanying R code are made accessible (See **Data and code availability** section).

Data diagnosis and descriptive statistics. Using the dlookr v.0.4.2²³ and pastecs²⁴ packages, the outliers identification, data frequency distribution, Shapiro-wilk normality test and descriptive statistics were performed using *find_outliers()*, *plot_normality()*, *normality()*, and *stats.desc()* functions, respectively.

Diversity index analysis. For the qualitative traits, (branching type, capsule hairiness, flower color, inflorescence type and seed color) Shannon-Weaver²⁵ and Simpson²⁶ diversity indexes

were calculated with the function *diversity()* of the package *vegan*²⁷. The equations of the two indexes are:

$$\text{Shannon index (H)} = - \sum_{i=1}^p \left(\frac{n_i}{N} \right) \log \left(\frac{n_i}{N} \right)$$

and

$$\text{Simpson index (D)} = \frac{1}{\sum_{i=1}^n \left(\frac{n_i}{N} \right)^2}$$

where n is the number (n) of observations regarding one particular qualitative traits modality i divided by the total number of observations (N)

Variability of agronomic traits: Considering that Africa or Asia continents are predicted to be the center of diversity candidates, we performed a one-way analysis of variance with the factor continent of origin. A generalization of Welch's method using trimmed means was employed since homoscedasticity assumption of our dataset was not satisfied. The one-way analysis of variance was run using the function *ggbetweenstats()* of the *ggstatsplot* v.0.7.0 package²⁸ with the option *type = "robust"*.

Correlation among traits. In order to assess correlation between agronomic parameters, Spearman correlation test were performed with the function *ggcorrmat()* of the package *ggstatsplot* v.0.70²⁸.

Path coefficient analysis for yield and yield relative components. As correlation alone doesn't automatically mean causative effect, we executed the path coefficient analysis following Dewey and Lu²⁹, to unravel direct or indirect effect between dried seed weight and relative yield components with the function *path.analysis ()* of the package *agricolae* v.1.3-3³⁰.

Classification of the accessions based on the agronomic traits. In order to group accessions based on agronomic traits, a principal component analysis followed by hierarchical agglomerative clustering were done using the function *PCA()* and *HCPC()* of the packages *FactoMineR* v.2.4³¹ and *factoextra* v.1.0.7³² respectively. The Euclidean distance-based similarity followed by Ward classification method was employed for the clustering stage. Prior the principal component analysis, data were standardized using the *scale()* function of the *rstats*²² package. The visualization of the multivariate analysis was rendered using the function *fviz_pca_biplot()* of the *factoextra* package v.1.0.7³².

Core collection inference and quality evaluation. The R version of Core Hunter 3³³ viz *corehunter* v3.2.1 was employed to determine a core collection by applying the average-entry-to-nearest-entry distance scheme based on Gower's distance metric³⁴. The Core Hunter

phenotypic data was generated from the comma-separated values excel file format of the data using the *phenotypes()* function. Then, the core collection was inferred with the function *samplecore()*. The quality of the inferred core collection with regard to the whole worldwide panel was assessed with the following metrics suggested by Hu et al.³⁵ and Kim et al.³⁶:

- the coincidence rate of range CRR (%),

$$\text{CRR (\%)} = \frac{1}{n} \sum_{i=1}^n \frac{R_{Ci}}{R_{Wi}} \times 100$$

where R_{Ci} is the range of the core collection for the agronomic trait i , and R_{Wi} is the range of the worldwide panel for the trait i ;

- the variable rate VR (%)

$$\text{VR (\%)} = \frac{1}{n} \sum_{i=1}^n \frac{CV_{Ci}}{CV_{Wi}} \times 100$$

where CV_{Ci} is the coefficient of variation of the core collection for the agronomic trait i , and CV_{Wi} is the coefficient of variation of the worldwide panel for the trait i ;

- the variance difference percentage VDP (%)

$$\text{VDP (\%)} = \frac{1}{n} \sum_{i=1}^n \frac{|\sigma_{Wi} - \sigma_{Ci}|}{\sigma_{Ci}} \times 100$$

where σ_{Ci} is the variance of the core collection for the agronomic trait i , and σ_{Wi} is the variance of the worldwide panel for the trait i ;

- the mean difference percentage MDP (%)

$$\text{MDP (\%)} = \frac{1}{n} \sum_{i=1}^n \frac{|\mu_{Wi} - \mu_{Ci}|}{\mu_{Ci}} \times 100$$

where μ_{Ci} is the variance of the core collection for the agronomic trait i , and μ_{Wi} is the variance of the worldwide panel for the trait i .

Besides, the means difference significance between the core and the whole accessions sets were computed following a Student t-test (for productive axis length), Wilcoxon test (for capsule number, plant height, branch number, stem diameter, dried biomass, dried seed weight, thousand seed weight, number of days to 50% flowering, number of days to maturity, number of days between flowering and maturity, capsule length, capsule width) or generalized linear model with a poisson error distribution (for harvest index, number of capsule per leaf axil, number of locules per capsule).

Geographical map: The map was rendered with *sf*³⁷, *ggplot2*³⁸ packages. The world shape file was retrieved from the University of California UC DAVIS geographic map data web repository ([Link 2](#)).

Results

Natural traits variation in the worldwide sesame panel

The range, mean, standard deviation, and coefficient of variation of measured traits are presented in the **Table 1**. The highest coefficient of variation was observed for dried seed weight (74.10%) followed by harvest index (72.40%), number of branches per plant (71.54%), number of capsules per plant (64.50%), and number of capsule per leaf axil (53.87%). Most of yield-related traits exhibited a wide range of variation indicating that phenotypic-based selection is appropriate for those traits. Interestingly, we identified the accession T218 that present exceptional 8 locules per capsule (**Fig. 1-A**).

Plant architecture including branching type is economically important trait that can affect crop productivity³⁹. In the present panel, 61.73% highly branched ($n > 10$ branches), 5.12% bi-branched and only 0.2 % (the accession TN42) unbranched accessions were recorded (**Table 1**). Most of the tested accessions exhibited white flower (97.4%), followed by pink (2.76%) and purple (0.4%) flowers. The purple color was observed only for the wild relative *Sesamum radiatum* whereas a typical pronounced pink color was showed by the wild *Sesamum alatum* (**Fig. 1-B**).

None of the accessions showed determinate growth habit except the induced determinate mutants dt-sel and dt-45 originated from Turkish sesame breeding programme⁴⁰.

A remarkable diverse ($H = 1.72$, $D = 0.78$) seed color was also noted (**Table 2**, **Fig. 1-C**), with 33.72% of white seed followed by 21.49% and 9.66 % of black seeds (**Table 1**).

An investigation of the variability of the studied traits among continent of origin revealed a significant variation between African- and Asian-originated accessions for both yield- and phenological-related traits (**Fig. 2**, **Fig. 3**, **Fig4**, **Fig 5**, **Fig 6**, **Fig. 7**).

Relationship among traits

The coefficients of correlation matrix among agronomic traits were summarized in the **Fig. 8**. A total of 41 positive ($p < 0.05$) and 45 negative ($p < 0.05$) coefficients of correlation were highlighted. For yield aspect, the strongest positive correlation was found between dried seed weight and harvest index ($r = 0.85$, $p < 0.001$). Similarly, number of days to flowering and number of days to maturation also exhibited a high relationship ($r = 0.85$, $p < 0.001$). In term of biomass, a positive relationship was revealed between stem diameter and dried biomass ($r = 0.69$, $p < 0.001$). The similar tendency was confirmed between plant height and stem

diameter ($r = 0.73$, $p < 0.001$) and plant height and dried biomass ($r = 0.61$, $p < 0.001$). Meanwhile, the highest negative relationship was detected between plant height and harvest index ($r = -0.71$, $p < 0.001$), followed by stem diameter and harvest index ($r = -0.69$, $p < 0.001$), and plant height and dried seed weight ($r = -0.54$, $p < 0.001$). Overall, biomass traits augmentation seems to have a reduction effect on yield and yield components traits.

In order to clarify the effect of the studied traits with a focus on dried seed weight, a path coefficient analysis was carried out. Details results were presented in the **Table 3**. The results showed that the number of capsules (0.18), the productive axis length (0.14), the thousand seed weight (0.02), and number of locules per capsule (0.02) exerted a positive direct effect on dried seed weight. However, number of days to flowering (-1.12), flowering to maturity days (-0.57) exhibited negative effect on dried seed weight. Branch number, plant height and stem diameter has also a negative influence on dried seed weight.

Altogether, both correlation and path coefficient analysis pinpointed the positive contribution of the number of capsules, number of locules per capsule, and the productive axis length for dried seed weight.

Phenotypic-based clustering

The principal component analysis performed on the quantitative traits revealed that 66.2% of the overall variability was retained by the first four principal components with 36.6%, 11.5%, 9.4% and 8.7% for the components 1, 2, 3 and 4 respectively (**Supplementary Fig. S2**).

The most contributing trait for the first principal component was harvest index (13.62 %) (**Supplementary Fig. S3-A**). This first dimension is characterized by some yield-related variables including harvest index ($r = 0.86$, $p < 0.0001$), capsule number per leaf axil ($r = 0.68$, $p < 0.0001$), and dried seed weight ($r = 0.64$, $p < 0.0001$) (**Supplementary Table S2**). The second principal component was highly correlated with the number of capsule per plant ($r = 0.78$, $p < 0.0001$) (**Supplementary Table S2**) with the highest contribution (35.7%) to the construction of the factorial axis 2 (**Supplementary Fig. S3-B**). Therefore, the factorial plan (1×2) depicted high-yielding variables (**Fig. 9-A**).

The third principal component was strongly correlated to two phenological traits including days to flowering ($r = 0.74$, $p < 0.0001$) and days to maturation ($r = 0.70$, $p < 0.0001$) (**Supplementary Table S2**) with their relative contribution of 38.22% and 35.11% (**Supplementary Fig. S3-C**) respectively. Thus, the factorial plan (1×3) highlighted high values of flowering and maturation days (late flowering and maturation characteristics) (**Fig. 9-B**).

Productive axis length, plant height, capsule width and capsule length contributed to construction of the fourth principal component with 23.19%, 19.61%, 12.08% and 11.05% respectively (**Supplementary Fig. S3-D**). The fourth principal component is positively correlated to the productive axis length ($r = 0.55$, $p < 0.0001$), plant height ($r = 0.51$, $p < 0.0001$), capsule width ($r = 0.40$, $p < 0.0001$), and capsule length ($r = 0.37$, $p < 0.0001$) (**Supplementary Table S2**). This dimension of the principal component analysis presented sesame plant growth rate-related parameters (**Fig. 9-C**).

The hierarchical classification of the accessions resulted in three clusters (**Fig. 10**). The cluster 1, 2, and 3 grouped 34.32%, 38.07%, and 27.61% of the total number of accessions, respectively. The quantitative traits that described each cluster are summarized in the **Table 4**. The cluster 1 encompassed the accessions that exhibited high biomass and low yield. The cluster 2 is characterized by late maturing and moderately yield-performing accessions while the cluster 3 representing the elite accessions harboring high-yield attributes (**Table 4, Fig. 10**). Most (61.43%) accessions of the cluster 3 originated from eastern Asia (**Fig. 11**). Interestingly, African representatives (17.14% for northern Africa, 3.57% for eastern Africa, 3.57% for western Africa, 2.14 % for southern Africa) are the second largest group that exhibited high-yield performance (**Fig. 11**).

Core collection inference

From 507 accessions, the Core Hunter 3 program generated a core collection encompassing 102 accessions. The number of retained accessions following geographical position is presented in the **Fig. 12**. Among the 102 accessions, a total number of 31, 37, and 34 belonged to the cluster 1, cluster 2 and cluster 3 respectively. The top 3 most contributing regions were eastern Asia ($n = 34$), followed by the northern Africa ($n = 30$) and the eastern Africa ($n = 10$) (**Fig. 12**).

The evaluation of the core collection quality revealed a variation of the coincidence rate of range (CRR%) per trait from 50% to 100% with an overall value of 78.04% (**Table 5**). The variable rate per (VR %) trait was ranging from 89.65% to 116.9% with an overall VR% of 100.49%. More interestingly, there was no significant difference ($p > 0.05$) between the core and the worldwide collection for all traits. This result was supported by the low overall mean difference (3.60%) and variance difference (14.39%) percentages (**Table 5**).

Identification of candidate genotypes for oil, protein, sesamin and sesamolin

Out of 102 accessions of the core collection, we were able to extracted and quantify oil and lignans content for 72 accessions due to the minimum seed weight (10 g) requirement criteria.

The oil and lignans content of the 72 accessions are presented in the **Supplementary Table S3**.

The hierarchical classification based on oil, lignans, and the others agronomic traits resulted in three clusters (**Fig. 13-A**). The cluster 1 grouped the accessions that show not only high-yield attributes but also are rich in proteins and alpha linoleic acid content (**Table 6**). The cluster 2 highlighted the accessions that exhibit higher content in stearic acid and sesamol while the cluster 3 is characterized by high-biomass yield accessions with higher sesamin, palmitic and linoleic acid content (**Table 6**).

Interestingly, the principal component analysis pinpointed the accessions T218, T077, T419, T170, T042, and T148 (**Fig. 13-B**) as valuable candidates for lignan-oriented breeding. The accession TN03, TN13, and T356 were highlighted for oleic acid, protein content and oleic acid content respectively (**Fig. 13-B**).

Discussion

The present study reports a comprehensive view of the phenotypic variability of a worldwide sesame panel from the Korean genbank and the development of a multi-purpose core collection regarding agronomic and nutritional traits.

The wide range variability observed among the accessions for the studied traits provide a scope for selection and set a path for the identification of novel genotypes with desirable traits. The study showed that some accessions (most from western and central African continent) exhibited important leafy biomass, are taller and less productive compared to the Asian representatives. Despite sesame leaves nutritional values have been neglected (mainly due to the oilseed trait) by the scientific community⁴¹, it is widely consumed in some African countries including Benin, Togo, Niger, Burkina-Faso, Nigeria, Sudan as leafy-vegetables and employed as remedy⁴¹.

The delay of the flowering and maturation days for these African accessions was also observed. More interestingly, certain did not enter in the maturity stage or even not flower. These observations indicate the photoperiodism sensitivity of some African accessions in our experimental environment (35° N latitude). Therefore, the photosensitivity appears as an adaptative trait for discriminating some African genotypes. Similarly, Bedigian et al.⁴² reported some tropical accessions that did not flower at the latitude of 40°N. Despite this phenomenon, valuable African representative performed well at 35° N Latitude with early flowering and high yield and nutritional characteristics. Most of those accessions are from eastern and northern Africa, representing acceptable resources as parental genotype for population development in the tested environment.

Wide-range seed color was also observed in the worldwide panel. As suggested by Bedigian et al.⁴³, sesame seed color may undergo intensive selection by human so far. As a result, extensive seed color variability occurred. This observation was in line with others agronomic traits including number of locules per capsule, branching type and number of capsule per leaf axil.

For yield-oriented breeding, the knowledge about the yield components traits that has a direct or indirect impact of yield is paramount for efficient yield-related breeding. Herein, we investigated the cause and effect relationship among yield and yield-component traits. The results highlighted the number of capsules, number of locules per capsule, and the productive axis length as key traits that has a direct effect on dried seed weight. Similar studies carried out in India⁴⁴, Turkey⁴⁰, and China⁴² support the present findings. Therefore, these traits may be considered as index for parental material selection for yield improvement. Specifically, the accession T218 that harbors 8 locules per capsule constitutes a valuable candidate as parental genotype.

Meanwhile, we were able to define a core collection that preserve the phenotypic variability from the whole set. The core collection size is about 20% from the initial worldwide set, suggesting that the inferred core set contains the minimum of repetitiveness. Comparable core set size was previously reported for sorghum⁴⁵ and safflower⁴⁶ with 24% and 31% respectively. Moreover, a non-significant difference between whole and core set for all traits was noted, supporting the fact that the core set maintained the genetic diversity. It is also valuable to mention a geographical broad representativeness of the inferred core collection in contrast with the previous core collection developed by Park et al.¹⁹. These observations support the good quality of the inferred core collection for effective usage in sesame breeding through genome-wide association studies for the dissection of the genetic basis of the desirable traits.

The study presents the first investigation of lignans content from a worldwide set of sesame accessions. The lignans is known to have multiple health benefits for human⁴³. The candidate genotypes provided by this study constitute a valuable resource regarding lignan-oriented breeding. Besides, for nutritional purposes, we also identified candidate rich-protein and lipid content accessions that can serve as parental material for population development. We also suggest as further investigation, to screen the high leafy biomass accessions for leaf nutritional content. It may be a valuable fiber dietary alternative that can help to feed people in the current context of the increasing population.

As part of the Korean genomic-assisted sesame breeding, these initial results pave the way for the identification of genomic regions responsible of the expression of the desirable agronomic traits. Ultimately, the proposed core collection would lay a foundation for association mapping studies for effective sesame breeding regarding oil, protein and lignans contents.

Data and code availability

Code and data can be found at: **xxxx [Ref zenodo]**.

Acknowledgments

The authors acknowledge the National Institute of Agricultural Sciences of Korea for supporting through the project **xxxxx**

Author contributions

YABZ, SUK and KL conceived and designed this study. YABZ, SUK, HJJ and KL conducted the experiments and collected data. YABZ performed data analysis and drafted the manuscript. SUK, SKTA, MN, NAK, NC, NJC and KL supervised the study, provided funding and technical support, and revised the draft of the manuscript. All authors have read and approved the final version of this manuscript.

Competing interests

The authors declare no competing interests.

References

1. Tripathi, A. D., Mishra, R., Maurya, K. K., Singh, R. B. & Wilson, D. W. Estimates for world population and global food availability for global health. in *The role of functional food security in global health* (eds. Watson, R., Singh, R. & Takahashi, T.) 3–24 (Academic Press, Cambridge, 2018).
2. FAO, IFAD, UNICEF, WFP, W. *The State of Food Security and Nutrition in the World 2020. Transforming food systems for affordable healthy diets* (FAO, IFAD, UNICEF, WFP and WHO, 2020). doi:10.4060/ca9692en.
3. Molotoks, A., Smith, P. & Dawson, T. P. Impacts of land use, population, and climate change on global food security. *Food Energy Secur.* **10**, 1–20 (2021).
4. Nhemachena, C. *et al.* Climate Change Impacts on Water and Agriculture Sectors in Southern Africa: Threats and Opportunities for Sustainable Development. *Water* **12**, 2673 (2020).
5. Yu, H. & Li, J. Short- and long-term challenges in crop breeding. *Natl. Sci. Rev.* **8**, 2021 (2021).
6. Morell, M. K. *et al.* Barley *sex6* mutants lack starch synthase IIa activity and contain a starch with novel properties. *Plant J.* **34**, 173–185 (2003).

7. Dossa, K. *et al.* The emerging oilseed crop *Sesamum indicum* enters the “Omics” era. *Front. Plant Sci.* **8**, 1–16 (2017).
8. Bedigian, D. Feeding the Forgotten: Wild and Cultivated *Ceratotheca* and *Sesamum* (Pedaliaceae) That Nourish and Provide Remedies in Africa. *Econ. Bot.* **72**, 496–542 (2018).
9. Moazzami, A. A., Haese, S. L. & Kamal-Eldin, A. Lignan contents in sesame seeds and products. *Eur. J. Lipid Sci. Technol.* **109**, 1022–1027 (2007).
10. Huang, H., Zhou, G., Pu, R., Cui, Y. & Liao, D. Clinical evidence of dietary supplementation with sesame on cardiovascular risk factors: An updated meta-analysis of randomized controlled trials. *Crit. Rev. Food Sci. Nutr.* **0**, 1–11 (2021).
11. Harikumar, K. B. *et al.* Sesamin Manifests Chemopreventive Effects through the Suppression of NF- κ B–Regulated Cell Survival, Proliferation, Invasion, and Angiogenic Gene Products. *Mol. Cancer Res.* **8**, 751–761 (2010).
12. Yokota, T. *et al.* Sesamin, a lignan of sesame, down-regulates cyclin D1 protein expression in human tumor cells. *Cancer Sci.* **98**, 1447–1453 (2007).
13. Farajbakhsh, A. *et al.* Sesame oil and vitamin E co-administration may improve cardiometabolic risk factors in patients with metabolic syndrome: a randomized clinical trial. *Eur. J. Clin. Nutr.* **73**, 1403–1411 (2019).
14. Forse, A. R. & Chavali, S. R. Sesamol Inhibitor of Delta-5-Desaturase Activity and Uses Therefor. U.S. Patent No. 2001/0031275. (2001).
15. Kojima, A., Yuasa, I., Kiyomoto, K. & Omura, A. Composition for promoting collagen production, for promoting keratinocyte migration and usage therefor. U.S. Patent No. 2016/0175280. (2020).
16. Yamada, D. *et al.* Oil-in-Water Emulsions Containing Lignan-Class Compounds and Compositions Containing the Same. U.S. Patent No. 8,685,455. (2020).
17. You, J. W., Rho, H. S., Kim, D. H., Chang, I. S. & Lee, O. S. Sesamol Derivatives and Their Salts, the Process for Preparing the Same, and the Skin External Composition Containing the Same. U.S. Patent No. 7,943,599. (2011).
18. Smale, M. & Jamora, N. Valuing genebanks. *Food Secur.* **12**, 905–918 (2020).
19. Park, J.-H. *et al.* Development and Evaluation of Core Collection Using Qualitative and Quantitative Trait Descriptor in Sesame (*Sesamum indicum* L.) Germplasm. *Korean J. Crop Sci.* **60**, 75–84 (2015).
20. Dossa, K. *et al.* Near-infrared reflectance spectroscopy reveals wide variation in major components of sesame seeds from Africa and Asia. *Crop J.* **6**, 202–206 (2018).

21. Federer, W. T. & Raghavarao, D. On Augmented Designs. *Biometrics* **31**, 29 (1975).
22. R Core Team. R: A Language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>. (2020).
23. Ryu, C. dlookr: Tools for data diagnosis, exploration, transformation. R package version 0.4.2. <https://CRAN.R-project.org/package=dlookr>. (2021).
24. Grosjean, P., Ibanez, F. & Etienne, M. pastecs: Package for Analysis of Space-Time Ecological Series. R package version 1.3.21. <https://CRAN.R-project.org/package=pastecs>. (2018).
25. Shannon, C. E. & Weaver, W. *The mathematical theory of communication*. (Urbana: University of Illinois Press, 1949).
26. Simpson, E. H. Measurement of Diversity. *Nature* **163**, 688–688 (1949).
27. Oksanen, J. *et al.* vegan: Community Ecology Package. (2020).
28. Patil, I. ggstatsplot: ‘ggplot2’ based plots with statistical details. R package version 0.7.0. <https://CRAN.R-project.org/package=ggstatsplot>. (2018).
29. Dewey, D. R. & Lu, K. H. A Correlation and Path-Coefficient Analysis of Components of Crested Wheatgrass Seed Production 1. *Agron. J.* **51**, 515–518 (1959).
30. de Mendiburu, F. agricolae: Statistical procedures for agricultural research. R package version 1.3-3. <http://CRAN.R-project.org/package=agricolae>. (2020).
31. Lê, S., Josse, J. & Husson, F. FactoMineR : An R Package for Multivariate Analysis. *J. Stat. Softw.* **25**, 1–18 (2008).
32. Kassambara, A. & Mundt, F. factoextra: Extract and visualize the results of multivariate data analyses. R package version 1.0.7. <http://CRAN.R-project.org/package=factoextra>. (2019).
33. De Beukelaer, H., Davenport, G. F. & Fack, V. Core Hunter 3: flexible core subset selection. *BMC Bioinformatics* **19**, 203 (2018).
34. Gower, J. C. A General Coefficient of Similarity and Some of Its Properties. *Biometrics* **27**, 857 (1971).
35. Hu, J., Zhu, J. & Xu, H. M. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theor. Appl. Genet.* **101**, 264–268 (2000).
36. Kim, K.-W. *et al.* PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics* **23**, 2155–2162 (2007).
37. Pebesma, E. *et al.* sf: Simple Features for R. R package version 0.9-7. <http://CRAN.R-project.org/package=sf>. (2021).

38. Wickham, H. *ggplot2: Elegant graphics for data analysis*. (Springer International Publishing, 2016).
39. Teichmann, T. & Muhr, M. Shaping plant architecture. *Front. Plant Sci.* **6**, 1–18 (2015).
40. Uzun, B. & Çağırhan, M. İ. Comparison of determinate and indeterminate lines of sesame for agronomic traits. *F. Crop. Res.* **96**, 13–18 (2006).
41. Bedigian, D. Feeding the Forgotten: Wild and Cultivated *Ceratotheca* and *Sesamum* (Pedaliaceae) That Nourish and Provide Remedies in Africa. *Econ. Bot.* **72**, 496–542 (2018).
42. Yingzhong, Z. & Yishou, W. Genotypic correlations and path coefficient analysis in sesame. *Sesame safflower Newsl.* **17**, 10–12 (2002).
43. Andargie, M., Vinas, M., Rathgeb, A., Möller, E. & Karlovsky, P. Lignans of Sesame (*Sesamum indicum* L.): A Comprehensive Review. *Molecules* **26**, 883 (2021).
44. Subramanian, S. & Subramanian, M. Correlation Studies and Path Coefficient Analysis in Sesame (*Sesamum indicum* L.). *J. Agron. Crop Sci.* **173**, 241–248 (1994).
45. Girma, G. *et al.* A comprehensive phenotypic and genomic characterization of Ethiopian sorghum germplasm defines core collection and reveals rich genetic potential in adaptive traits. *Plant Genome* **13**, 1–17 (2020).
46. Kumar, S. *et al.* Utilization of Molecular, Phenotypic, and Geographical Diversity to Develop Compact Composite Core Collection in the Oilseed Crop, Safflower (*Carthamus tinctorius* L.) through Maximization Strategy. *Front. Plant Sci.* **7**, 1–14 (2016).