Journal of Agricultural Science and Technology (JAST), 28(1)
In Press, Pre-Proof Version

Assessing genetic diversity of soybean based on smartphone image-derived canopy parameter

Myong-Kwang Ri[1] , Kwang-O Jong[1]*, and Ye-Kwang Sin[1]

ABSTRACT

Convenient and accurate characterization of field-grown crops is essential for effective use of germplasm resources and breeding programs. In this study, we evaluated genetic relationships among 18 soybean accessions at the early growth stage using a smartphone image-derived canopy parameter, the canopy cover rate (CCR). Field experiments were conducted over two consecutive years (2021 and 2022). CCR was estimated from top-view images using image analysis software, providing a non-destructive and efficient indicator of plant morphology. CCR showed significant variation among accessions and was strongly correlated with traditional morphological / biomass traits (correlation coefficients >0.8). Multivariate analyses, including principal component analysis (PCA), hierarchical cluster analysis (HCA), and discriminant analysis (DA), revealed that CCR could effectively classify accessions, with DA achieving an average correct classification rate of 88.9%. The results suggest that CCR is a reliable index for assessing genetic diversity in field-grown soybean genotypes. This study introduces an innovative, simple, and accurate method for evaluating soybean genetic resources using image-derived parameter.

Keywords: Biomass**,** Canopy, Genetic diversity, Image-derived parameter, Phenotyping, Soybean

INTROIDUCTION

Soybean (*Glycine max* L.) is a globally important crop, valued for its protein and oil content as well as its role in sustainable agriculture through biological nitrogen fixation research (McDonald *et al*., 2023). From these reasons, effective conservation and utilization of soybean genetic resources are essential for breeding programs aimed at improving yield and resilience. Traditionally, genetic diversity has been assessed using morphological traits, which are often labor-intensive, subjective, and influenced by environmental factors (Khadivi, 2018; Shahid *et al*., 2021). Recent advances in digital phenotyping (Liang *et al*.,2018; Zhang *et al*., 2018; Zhou

[1] Department of Garden Plant Breeding, Faculty of Life Science, Kim Il Sung University, Pyongyang, Democratic People's Republic of Korea.
*Corresponding author; e-mail: life5@ryongnamsan.edu.kp

et al., 2018; Jong *et al*., 2021), particularly the use of smartphone-based imaging (Barman *et al*., 2020; Adhikari *et al*., 2020), offer promising alternatives for rapid, non-destructive, and objective crop characterization.

A good characterization of the plant materials is necessary for the effective use of germplasm resources and further for crop improvement (Zanklan *et al*., 2018). Because most of morphological and biomass traits may be affected by the genotype × environment interaction, it is essential to comprehensively and accurately evaluate the different phenotypes using image-derived phenotyping approach at growing stage. Wang *et al*. (2020) proposed a multiscale sliding chord matching method for characterising and recognising soybean cultivars from leaf images. Here, a chord was defined to slide along the leaf contour for measuring synchronized exterior shape features and interior appearance patterns of the soybean leaf image. However, the application of smartphone image-derived canopy parameters for genetic diversity assessment at the early growth stage in field-grown soybean remains unexplored. Because most of morphological / biomass traits may be affected by the genotype × environment interactions, it is essential to comprehensively and accurately evaluate the different phenotypes using image-derived phenotyping approach at the growth stage. Therefore, we hypothesized that the canopy cover rate (CCR), extracted from smartphone images, could serve as a reliable index for evaluating genetic diversity among soybean accessions. The objectives of this study were to: (i) evaluate the feasibility of extracting canopy parameters using image analysis software from smartphone images at the early growth stage, and (ii) assess genetic diversity among soybean accessions based on image-derived canopy parameters. This approach has the potential to enhance the efficiency and accessibility of genetic diversity evaluation in soybean breeding programs.

MATERIALS AND METHODS

1. Plant material

Seeds of eighteen soybean accessions were obtained from the Industrial Crops Institute, Academy of Agricultural Sciences, DPR Korea (Figure 1 and Table S1). All accessions were grown under field conditions.

2. Experiment site

Field experiments were conducted in experimental station (lat 39° 01′ 10 " N, long 125° 44′ 44" E, alt 30m asl) of life science faculty of Kim Il Sung university for two consecutive years

(2021 and 2022). The soil was classified as gray alluvial clay loam with a pH of 6.2. Maize was grown in the previous cropping system.

3. Weather conditions in experimental site

Weather data were recorded daily at a nearby meteorological station (2 km distance) and summarized in Table S2.

4. Experimental design

A randomized complete block design (RCBD) was used with three replicates per accession. Each plot measured 2 m² and consisted of two rows (2m length, 60cm between rows, 30cm between plants). The number of plants per plot was 12, and total number of replicates was three. The border plants were excluded from analysis.

5. Management Practices

Standard agronomic practices were followed, including pre-sowing fertilization (15-15-15 N-P_2O_5-K_2O at 200 kg ha^{-1}), manual weeding, and pest control with registered insecticides.

6. Plant measurements

Ten plants of twenty-day-old (2021) and 27-day-old (2022) plants were sampled for each accession, excluding border plants. Plant height (PH, ㎝) and root length (RL, ㎝) were measured with a ruler. For each root sampling, a block of soil (25 cm × 20 cm × 30 cm; length, width, and depth) around each individual hill was dug up using a sampling core. The roots of plant in each block of soil were carefully rinsed with a hydropneumatic elutriation device (Gillison 's Variety Fabrications, Benzonia, MI, USA). Root samples were used for the measurement of root length (RL, ㎝). Plants were oven-dried at 70°C for 48 hours to determine plant dry mass) (PDM, g), aboveground dry mass (ADM, g) and leaf dry mass (LDM, g).

7. Image Acquisition and Processing

Canopy cover (top-view) images were captured using a smartphone (Type 2428, Pyongyang, DPR Korea,48 MP camera) mounted on a selfie stick at 50 cm above each plant, between 11:30 and 12:30 h under natural light (Figure 2). Altogether, there were three digital images of ten plants for each accessions. Digital images stored in JPG file format. The cost of red–green–blue (RGB) image acquisition with smartphone camera is much lower than that with other optical instruments. Images were processed using IA software (Golden Field 2.0) developed using fuzzy c-means clustering algorithm (FCM) (Figure 3, 4 and 5). As one of the most widely used clustering methods, FCM introduces the fuzziness for the belongingness of each image pixel and can retain more information from the original image than the hard c-means clustering algorithm (Zhao *et al*., 2013). It is a pixels clustering process of dividing pixels into clusters so that pixels in the same cluster are as similar as possible and those in different clusters are as dissimilar as possible (Siang *et al*., 2010). FCM clustering algorithm tries to partition image pixels $\{x_k\}_{k=1}^{N}$ into c clusters. The standard FCM objective function was as follows.

$$J_m(U,v)=\sum_{i=1}^{n}\sum_{k=1}^{c}\mu_{ik}^{m}\left\|x_i-v_k\right\|^2$$

The fuzzy membership degree of a k th image pixel x_k to a specific cluster v_i was given by the membership value u_{ik} of the data point to that cluster. The membership value was calculated by minimization of a FCM function, which searches for the belongingness that gives the least error.

$$u_{ik}=\frac{1/d_{ik}^{2/(m-1)}}{\sum_{j=1}^{c}1/d_{ij}^{2/(m-1)}}$$

In equation above, m is a parameter that controls the fuzziness of the clustering process. The function needs approximate cluster centers v_i , as well as a metric for membership evaluation as input, e.g., the Euclidean distance:

$$d_{ik}=\left\|x_k-v_i\right\|$$

The minimization is an iterative process where new cluster centers are computed as weighted averages of all data points, where the membership values are the weights. After obtained R (red), G (green) and B (blue) values of each pixel from RGB images, these values were transformed into H(Hue), S(Saturation), V(Brightness) color system. HSV values of each pixel were used to distinguish green canopy cover pixels from background pixels using FCM clustering algorithm. Here, the number of the cluster c was 3, green pixels cluster count was 1, background pixels cluster count was 2. Canopy cover rate (CCR) was calculated as the ratio of green canopy cover pixels to input image pixel gross.

CCR%=(Canopy cover pixel gross / Input image pixel gross)×100

8. Statistical analyses

Data were analyzed using IBM SPSS Statistics v21. Means, variances, coefficients of variation, and Pearson correlations were calculated. One-way ANOVA was performed for each trait. Multivariate analyses (principal component analysis (PCA), hierarchical cluster analysis (HCA) and discriminant analysis (DA)) were conducted to assess genetic diversity and group accessions. Significance was determined at $p < 0.05$. To determine the comprehensive trait among the 6 traits investigated, arithmetic mean of sum of coefficients of determination ($\overline{R^2}$) was calculated using the following formula:

$\overline{R^2} = \frac{\sum r^2}{\mathrm{m}-1}$ (m=6; Number of the traits investigated)

RESULTS

1. Phenotypic variation

Statistical analysis revealed significant phenotypic variation among the 18 soybean accessions at the early growth stage (Table 1). PH and biomass traits including PDM, ADM and LDM exhibited highly significant differences ($p < 0.01$), with coefficients of variation (CV) ranging from 12.7% to 40.0% in 2021 and from 19.3% to 42.6% in 2022. In contrast, RL showed much lower variability (CV = 6.8% in 2021, 2.0% in 2022). No significant differences were observed in RL among accessions in both years.

2. Evaluation of CCR

CCR was estimated from top-view images using image analysis software, providing a non-destructive and efficient indicator of plant morphology. It varied significantly among accessions (Table 2). Duiguru17-1 had the highest CCR (14.01% in 2021, 29.89% in 2022), while Duiguru19-1 had the lowest (3.49% in 2021, 4.60% in 2022). CCR exhibited the highest CV among all traits (42.4% in 2021, 50.5% in 2022), indicating strong discriminatory power (Table 3).

3. Correlation between CCR and morphological/biomass traits

CCR showed strong, significant positive correlations ($p < 0.01$) with all measured morphological / biomass traits (Figure 5). In 2021, correlation coefficients ranged from 0.836 (PDM) to 0.943 (RL), while in 2022, they ranged from 0.878 (PH) to 0.943 (LDM).

Because $\overline{R^2}$ of CCR has the highest value (0.8023 in 2021, and 0.8403 in 2022) among $\overline{R^2}$ of 6 traits, CCR seems to be the comprehensive trait among all the investigated traits (Table 8).

4. Multivariate Analyses (PCA, HCA, DA)

Data on soybean plant descriptors including PH, RL and biomass traits were checked for KMO (Kaiser-Meyer-Olkin Measure) and homogeneity of variance (Bartlett's test). The KMO value (0.808 in 2021 and 0.870 in 2022) showed that it was good, while Bartlett's test of Sphericity with an associated p value of < 0.001 suggests that we can proceed with PCA.

PCA confirmed that two principal components explained over 95% of total variance in both years (Table 5). Table 6 showed the significant correlations were detected between PC1 and PC2 with CCR (0.608 and 0.749 in 2021, and 0.715 and 0.655 in 2022, respectively). Therefore, PC1 and PC2 could explain the characteristics of 18 soybean accessions, instead of morphological / biomass traits.

The individual component values in both years were calculated using the values from the component score coefficient matrix and the following equations, respectively:

$PC1=-0.741PH+(-0.148)RL+(0.680)PDM+(0.611)ADM+(0.361)LDM$ (in 2021)

$PC2=(1.194)PH+(0.509)RL+(-0.475)PDM+(-0.397)ADM+(-0.094)LDM$ (in 2021)

$PC1=-0.652PH+(-0.207)RL+(0.529)PDM+(0.533)ADM+(0.514)LDM$ (in 2022)

$PC2=(1.115)PH+(0.583)RL+(-0.306)PDM+(-0.312)ADM+(-0.289)LDM$ (in 2022)

The factor loadings that resulted from Varimax rotation were generated with PCs and morphological / biomass traits (Table 7). PC1 was strongly associated with biomass traits, while PC2 was linked to plant height and root length (Table 7). Moreover, PC1 had a strong positive correlation to biomass traits (PDM, ADM and LDM) which characterize the "*biomass*" of the plants, while PC2 showed the close relationships with quantitative traits such as PH and RL describing "*length*". Four main groups of accessions were identified based on PCA scatter plots (Figure 7). The two axes, namely, PC1 and PC2 accounted for 95.0 % (in 2021), and 96.5% (in 2022) of the variability in morphological / biomass traits.

According to scatter plots of PCA for morphological / biomass traits, four categories were identified among 18 accessions in both years. In detail, first category was the largest, consisting of 6 accessions, namely, Kong25-1, Kong27-1, Kong28-1, Kong29-1, KuNul5-1 and Gansokji1-1. Second category comprised Duiguru19-1, Duiguru20-1, Kangwon30-1, Haqjak40 and KuNul3-1. Third category was composed of Kong26-1, Kangwon11-1, Duiguru21-1 and Dongnong50, and fourth category contained Duiguru13-1, Duiguru14-1 and Duiguru17-1 (Figure 7).

In PCA, PC1 was positively correlated to biomass traits, respectively (0.883, 0.855 and 0.793 in 2021, and 0.882, 0.883 and 0.876 in 2022), while PC2 was positively correlated to morphological traits, respectively (0.894 and 0.726 in 2021, and 0.890 and 0.736 in 2022) (Table 7).

R-mode HCA was performed using between-groups linkage based on Pearson correlation coefficients to find out the relationships among the investigated traits. It showed that biomass traits had a high correlation coefficient one another, while CCR had a close similarity to biomass traits (Table 8).

Soybean accessions with the similar morphological / biomass traits were clustered together in Figure 8. When using the relative distance of 5.0 and 10.0 as a threshold, 18 accessions were clustered into three main categories and seven sub-categories. First category was the largest, consisting of 12 accessions, namely, Kangwon30-1, Duiguru20-1, Kong27-1, Gansokji1-1, Duiguru19-1, Haqjak40, KuNul3-1, Kong29-1, KuNul5-1, Dongnong50, Kong25-1 and Kong28-1 in 2021, and 8 accessions, namely, Duiguru19-1, Duiguru20-1, KuNul3-1, KuNul5-1, Kong28-1, Kangwon30-1, Haqjak40 and Gansokji1-1 in 2022. Second category composed of 3 accessions including Duiguru13-1, Duiguru14-1 and Duiguru17-1 in 2021, and 6 accessions including Kong25-1, Kong26-1, Kong27-1, Kong29-1, Kangwon11-1 and Dongnong50 in 2022. Third category consisted of 3 accessions, namely, Kangwon11-1, Duiguru 21-1 and Kong 26-1 in 2021, and 4 accessions, namely, Duiguru13-1, Duiguru14-1, Duiguru17-1 and Duiguru21-1 in 2022.

When 18 accessions were grouped based on CCR, the dendrograms obtained by HCA were shown in Figure 9. Four major categories could be detected using the relative distance of 5.0 and 10.0 as a threshold. First category was the largest, consisting of 6 accessions, namely, Kong27-1, Kong28-1, Kong29-1, KuNul5-1, Dongnong50 and Gansokji1-1 in 2021 and 2022. Second category consisted of 5 accessions, namely, Kangwon30-1, Duiguru19-1, Duiguru20-1, Haqjak40 and KuNul3-1 in 2021 and 2022. Third category comprised of Kong25-1, Kong26-1, Kangwon11-1 and Duiguru21-1 in 2021 and 2022. Fourth category included Duiguru13-1, Duiguru14-1 and Duiguru17-1 in 2021 and in 2022.

Because it was able to classify four major categories, results of HCA based on CCR were more similar to the ones suggested by the PCA than clustering based on morphological / biomass traits.

The group centroid of categories (the first, second, third, fourth category were -0.478, -3.276, 1.295, 4.690 in 2021 and -0.282, -2.502, 0.799, 3.668 in 2022, respectively) was calculated according to the following equations using unstandardized coefficients.

D = (0.907) CCR-6.796 (in 2021)

D = (0.306) CCR-4.529 (in 2022)

As results of DA for CCR, Kong25-1 of first category was classified into third category and Dongnong50 of third category was classified into first category in 2021. However, Kong25-1 was classified into third category and Dongnong50 was classified into second category in 2022. The percentage of correctly classified on the basis of CCR was 88.9% of grouped cases by PCA. For first category 83.3% of the cases were classified correctly, and 75.0% of the cases were classified correctly for third category. Especially, the classification rate of second category and fourth category was 100% (Table 9).

DISCUSSION

To our knowledge, although image-derived phenotyping has been explored in various crops, its application for genetic diversity assessment at the early growth stage in soybean remains largely unaddressed. In this study, we demonstrated that CCR derived from smartphone images is a robust and efficient index for evaluating genetic diversity among field-grown soybean accessions at the early growth stage. The high correlations observed between CCR and traditional morphological / biomass traits confirm the reliability of this image-derived parameter as an indirect measure of plant growth and architecture. These findings are consistent with recent reports on the utility of digital phenotyping in crop improvement (Zhou *et al*., 2018; Zhang *et al*., 2018), but to our knowledge, this is the first study to apply such an approach for genetic diversity assessment in soybean at early developmental stages.

Traditionally, several quantitative traits have been used to determine the genetic diversity and classify germplasm resources in many plants (Gadissa *et al*., 2020; Shahid *et al*., 2021). However, measuring quantitative traits such as plant height and biomass is the labor-intensive and time-consuming in large breeding populations and field environments (Jiang *et al*., 2016;

Amaral *et al*., 2015). Furthermore, conventional measuring on biomass traits has been obtained using the destructive method such as the drying in an oven (Wen *et al*. 2017).

In this study, CCR was estimated using IA software from RGB image without any significant alteration of plant morphology at the early growth stage. Moreover, estimation of CCR using IA software from canopy images taken by the smartphone camera seems to be suitable for young plants grown in field environment.

HCA produced similar groupings, especially when based on CCR. DA achieved an average correct classification rate of 88.9%, supporting the utility of CCR for distinguishing genetic diversity among accessions.

A multiscale sliding chord matching method was proposed to characterize and recognize soybean cultivars using joint leaf image patterns (Wang *et al*., 2020). Here, to obtain soybean cultivar leaf image database researchers used the destructive method to take the individual leaf image from the lower, middle and upper parts of the plants of one soybean cultivar, respectively. However, we employed the non-destructive method to take canopy image using a smartphone camera from individual soybean plant in field environment.

The results above provided the support for the hypothesis that that smartphone image-derived CCR could serve as a simple, accurate, and non-destructive index for evaluating genetic diversity among field-grown soybean accessions at the early growth stage.

Because the present approach using CCR as a genetic diversity index uses a smartphone camera for capturing digital images in the field environment, it is far simpler and lower-cost than the complex and expensive system using LiDAR-based Canopy Height Model, also known as a normalized Digital Surface Model (An *et al*., 2016) and Normalized Difference Vegetation Index using remote sensing (Rees *et al*., 2020) and Leaf Area Index (LAI) estimated by Terrestrial Laser Scanning (Chen *et al*., 2018).

The present approach is adequate to the early growth stage of crops, but is inadequate to the maturing period of high crops such as maize, sugarcane and sorghum, because capturing the top-view canopy image for high plants is difficult with smartphone camera.

Overall, our findings supported the use of smartphone image-derived CCR as a practical and effective tool for genetic diversity assessment in soybean, paving the way for more efficient phenotyping and breeding strategies.

Further validation in larger and more diverse populations, as well as at later growth stages, is recommended to confirm the generalizability of these findings.

CONCLUSIONS

In summary, this study demonstrated that CCR, extracted from smartphone images, is a robust and efficient index for evaluating genetic diversity among field-grown soybean accessions at the early growth stage. The strong correlation between CCR and traditional morphological / biomass traits, together with high classification accuracy in multivariate analyses, highlights the practical value of this approach. By leveraging accessible and low-cost smartphone technology, this method offers a rapid, non-destructive alternative to conventional phenotyping, making it particularly suitable for breeding programs and genetic resource management in resource-limited environments. While further validation across broader germplasm collections and developmental stages is warranted, our findings support the integration of image-derived canopy parameters into modern crop improvement pipelines.

REFERENCES

1. Adhikari, R., Li, C., Kalbaugh, K. and Nemali, K. 2020. A low-cost smartphone controlled sensor based on image analysis for estimating whole-plant tissue nitrogen (N) content in floriculture crops. *Comput. Electron. Agric.*, **169**:105173. https://doi.org/10.1016/j.compag.2019.105173
2. An, N., Palmer, C.M., Baker, R.L., Markelz, R.J.C., Ta, J., Covington, M.F., Maloof, J.N., Welch, S.M. and Weinig, C. 2016.Plant high-throughput phenotyping using photogrammetry and imaging techniques to measure leaf length and rosette area.*Comput. Electron. Agric.*,**127**:376-394,https://doi.org/10.1016/j.Compag.2016.04.002
3. Barman,U., Choudhury, R.D., Sahu,D. and Barman, G.G. 2020. Comparison of convolution neural networks for smartphone image based real time classification of citrus leaf disease.*Comput. Electron. Agric.,* **177**:105661. https://doi.org/10.1016/j.compag.2020.105661
4. Chen, Y.M., Zhang W.M., Hu R.H., Qi J.B., Shao J., Li D., Wan P., Qiao C., Shen A.J. and Yan, G.J. 2018. Estimation of forest leaf area index using terrestrial laser scanning data and path length distribution model in open-canopy forests. *Agri.Forest Meteo.*,**263**:323-333, https://doi.org/10.1016/j.agrformet.2018.09.006
5. Gadissa, Fekadu, Tesfaye, Kassahun, Dagne, Kifle, Geleta, Mulatu 2020. Morphological traits based genetic diversity assessment of Ethiopian potato

[*Plectranthus edulis* (Vatke) Agnew] populations from Ethiopia. *Genet. Resour. Crop Evol.*, **67**:809-829. https://doi.org/10.1007/s10722-019-00794-6

6. Jiang, Y., Li, C.Y. and Paterson, A.H. 2016. High throughput phenotyping of cotton plant height using depth images under field conditions. *Comput. Electron. Agric.*, **130**:57-68.
7. Jong, K.O., Han, K.M., Kawk, S.L., Jang, Y.J., Kim, K.P. and Ho, C. 2021. Simple estimation of green area rate using image analysis and quantitative traits related to plant architecture and biomass in rice seedling. *Theor. Exp. Plant. Physiol.*, **33**: 225-234. https://doi.org/10.1007/s40 626-021-00207-z
8. Khadivi,A.2018.Phenotypic characterization of *Elaeagnus angustifolia* using multivariate analysis.*Ind. Crops Prod.*, **120**, 155-161. https://doi.org/10.1016/j.indcrop.2018.04.050
9. Liang, W.Z., Kirk, K.R., Greene, J.K. 2018.Estimation of soybean leaf area, edge, and defoliation using color image analysis. *Comput. Electron. Agric.*, **150**:41-51. https://doi.org/10.1016/j.compag.2018.03.021
10. McDonald, S. C., Bilyeu, K., Koebernick, J., Buckley,B., Fallen, B. , Rouf Mian, M. A., Li, Z.L. 2023.Selecting recombinan ts to stack high protein with high oleic acid and low linoleic acid in soybean (*Glycine max*). *Plant Breed.*,**142**:477-488. https://doi.org//10.1111/pbr.13102
11. Rees, W. G., Golubeva , E. I. , Tutubalina, O. V., Zimin, M. V. and Derkacheva, A. A. 2020.Relation between leaf area index and NDVI for subarctic deciduous vegetation.*Inter. J. Remote Sensing*.**41**:8573-8589.https://doi.org/10.1080/01431161.2020.1782505
12. Shahid, A., Ayyub, C.M., Abbas, M. and Ahmad, R. 2021.Assessment of genetic diversity in round gourd (*Praecitrullus fistulosus*) germplasm of Pakistan considering morphological characters.*Genet.Resour.Crop Evol.*,**66**:215-224.https://doi.org/10.1007/s10722-018-0707-5
13. Siang, Tan, K., Mat Isa N.A. 2010.Color image segmentation using histogram thresholding– Fuzzy C-means hybrid approach. *Pattern Recognition*, **44**:1-15. https://doi.org/10.1016/j.patcog.2010.07.013
14. Wang, B., Gao, Y.S., Yuan, X.H., Xiong, S.W., Feng, X.Z.2020. From species to cultivar: Soybean cultivar

15. recognition using joint leaf image patterns by multiscale sliding chord matching. Biosystems engineering,**194**: 99-111. https://doi.org/10.1016/j.biosystemseng.2020.03.019
16. Wen, Z.F., Ma, M.H., Zhang, C.,Yi, X.M., Chen, J.L. and Wu, S.J. 2017. Estimating seasonal aboveground biomass of a riparian pioneer plant community: An exploratory analysis by canopy structural data. *Eco. Indic*., **83**:441-450. https://doi.org/10.1016/j.ecolind.2017.07.048
17. Zanklan, S., Becker, H.C., Sørensen, M., Pawelzik, E. and Grüneberg, W.J. 2018.Genetic diversity in cultivated yam bean (*Pachyrhizus* spp.) evaluated through multivariate analysis of morphological and agronomic traits.*Genet. Resour. Crop. Evol.*, **65**:811-843. https://doi.org/10.1007/s10722-017-0582-5
18. Zhao, F., Jiao, L.C., Liu, H.Q. 2013. Kernel generalized fuzzy c-means clustering with spatial information
19. for image segmentation. *Digital Signal Processing*. **23**:184-199. http://dx.doi.org/10.1016/j.dsp.2012.09.016
20. Zhang, C.Y., Si, Y.S., Lamkey, J. and Boydston, R.A., Garland-Campbell, K.A., Sankaran, S.2018. High-Throughput Phenotyping of Seed/Seedling Evaluation Using Digital Image Analysis. *Agronomy* **8**:63-77.
21. Zhou, J.F., Chen, H.T., Zhou, J., Fu, X.Q., Ye, H. and Nguyen, H.T. 2018.Development of an automated phenotyping platform for quantifying soybean dynamic responses to salinity stress in greenhouse environment. *Comput. Electron. Agric.,* **151**: 319-330. https://doi.org/10.1016/j.compag.2018.06.016

Table 1. Statistical parameters for morphological / biomass traits utilized at the early growth stage in 18 accessions.

Traits	Abbreviation	in 2021 Min	Max	Mean	SD	CV (%)	in 2022 Min	Max	Mean	SD	CV (%)
Plant height	PH	9.8	14.0	11.48	1.46	12.7	12.4	22.3	16.4	3.16	19.3
Root length	RL	10.9	13.4	12.01	0.82	6.8	21.5	23.1	22.3	0.44	2.0
Plant dry mass	PDM	0.34	1.22	0.72	0.24	33.3	0.64	2.82	1.55	0.64	41.3
Aboveground plant dry mass	ADM	0.27	0.90	0.56	0.19	33.9	0.51	2.30	1.22	0.52	42.6
Leaf dry mass per plant	LDM	0.16	0.71	0.40	0.16	40.0	0.32	1.45	0.78	0.32	41.0

SD: Standard Deviation, *CV:* Coefficient of Variation.

Table 2. CCRs at the early growth stage in 18 accessions.

Accessions	CCRs (%) in 2021	CCRs (%) in 2022
Kong 25-1	$8.80\pm0.13^{g,*}$	$19.32\pm0.65^{f,*}$
Kong 26-1	$10.11\pm0.09^{d,*}$	$18.20\pm0.47^{g,*}$
Kong 27-1	$7.02\pm0.09^{i,*}$	$15.54\pm0.38^{h,*}$
Kong 28-1	$7.61\pm0.10^{h,*}$	$12.60\pm0.31^{i,*}$
Kong 29-1	$6.52\pm0.02^{j,*}$	$12.79\pm0.37^{i,*}$
Kangwon11-1	$9.42\pm0.11^{f,*}$	$20.04\pm0.52^{e,*}$
Kangwon 30-1	$4.51\pm0.10^{m,*}$	$8.51\pm0.21^{l,*}$
KuNul 5-1	$6.11\pm0.02^{k,*}$	$11.49\pm0.26^{j,*}$
Duiguru13-1	$11.82\pm0.15^{c,*}$	$26.71\pm0.71^{b,*}$
Duiguru14-1	$12.2\pm0.11^{b,*}$	$23.67\pm0.65^{c,*}$
Duiguru17-1	$14.01\pm0.79^{a,*}$	$29.89\pm0.79^{a,*}$
Duiguru19-1	$3.49\pm0.10^{n,*}$	$4.60\pm0.12^{p,*}$
Duiguru20-1	$4.10\pm0.10^{n,*}$	$7.70\pm0.19^{m,*}$
Duiguru21-1	$9.61\pm0.10^{e,*}$	$21.14\pm0.34^{d,*}$
Haqjak40	$3.72\pm0.02^{p,*}$	$6.31\pm0.14^{n,*}$
Dongnong50	$6.60\pm0.16^{j,*}$	$10.19\pm0.17^{k,*}$
Gansokji1-1	$5.78\pm0.08^{l,*}$	$11.43\pm0.15^{j,*}$
KuNul3-1	$3.61\pm0.11^{o,*}$	$5.97\pm0.14^{o,*}$

Values are means±standard errors with results of statistical analysis, Total pixel counts= 84500 (n= 20), *Means in column followed by the same letters are not significantly different at $P< 0.05$ level by the Fisher's LSD test.

Table 3. Descriptive statistics for CCR among 18 accessions.

Trait	in 2021 Min	Max	Mean	SD	CV (%)	in 2022 Min	Max	Mean	SD	CV (%)
CCR	3.49	14.01	7.50	3.18	42.40	4.60	29.89	14.78	7.46	50.47

Table 4. The arithmetic mean of R^2 between the investigated traits.

Traits	$\overline{R^2}$ in 2021	$\overline{R^2}$ in 2022
PH	0.7029	0.6574
RL	0.7964	0.7330
PDM	0.7577	0.8322
ADM	0.7566	0.8326
LDM	0.7829	0.8347
CCR	0.8023	0.8403

Table 5. Percentage of variance and cumulative variance, and eigenvalues for two principal components

	in 2021			in 2022		
	Eigenvalues	**Percentage of variance (%)**	**Percentage of cumulative variance (%)**	**Eigenvalues**	**Percentage of variance (%)**	**Percentage of cumulative variance (%)**
PC1	4.457	89.134	89.134	4.484	89.683	89.683
PC2	0.295	5.898	95.032	0.341	6.822	96.505

Table 6. Correlation coefficients between PC1 and PC2 with CCR.

		in 2021		in 2022	
		Component 1	**Component 2**	**Component 1**	**Component 2**
CCR	Pearson Correlation	0.608**	0.749**	0.715**	0.655**
	Sig.(2-tailed)	0.007	0.000	0.001	0.003
	N	18	18	18	18

Results marked with **are significant at the 0.01 probability levels.

Table 7. Rotated component matrix with Varimax.

	in 2021		in 2022	
	Component 1	**Component 2**	**Component 1**	**Component 2**
PH	0.425	0.894	0.407	0.890
RL	0.640	0.726	0.593	0.736
PDM	0.883	0.435	0.882	0.464
ADM	0.855	0.451	0.883	0.462
LDM	0.793	0.550	0.876	0.469

Table 8. The agglomeration schedule between morphological / biomass traits and CCR.

	in 2021						in 2022					
Stage	**Cluster Combined**		**Coe**	**Stage Cluster First Appears**		**Next Stage**	**Cluster Combined**		**Coe**	**Stage Cluster First Appears**		**Next Stage**
	C1	**C2**		**C1**	**C2**		**C1**	**C2**		**C1**	**C2**	
1	RL	CCR	0.943	0	0	4	PDM	ADM	0.994	0	0	2
2	PDM	LDM	0.937	0	0	3	PDM	LDM	0.988	1	0	3
3	PDM	ADM	0.904	2	0	5	PDM	CCR	0.935	2	0	4
4	RH	RL	0.902	0	1	5	RL	PDM	0.862	0	3	5
5	PH	PDM	0.849	4	3	0	PH	RL	0.810	0	4	0

C1 Cluster 1, *C2* Cluster 2, *Coe* Coefficient

Table 9. Percentage of 18 accessions classified in each category by DA.

			in 2021					in 2022				
category			**Predicted category Membership**				**Total**	**Predicted category Membership**				**Total**
			1	**2**	**3**	**4**		**1**	**2**	**3**	**4**	
Original	Count	1	5	0	1	0	6	5	0	1	0	6
		2	0	5	0	0	5	0	5	0	0	5
		3	1	0	3	0	4	0	1	3	0	4
		4	0	0	0	3	3	0	0	0	3	3
	%	1	83.3	0.0	16.7	0.0	100.0	83.3	0.0	16.7	0.0	100.0
		2	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	100.0
		3	25.0	0.0	75.0	0.0	100.0	0.0	25.0	75.0	0.0	100.0
		4	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	100.0

88.9% of original grouped cases was correctly classified

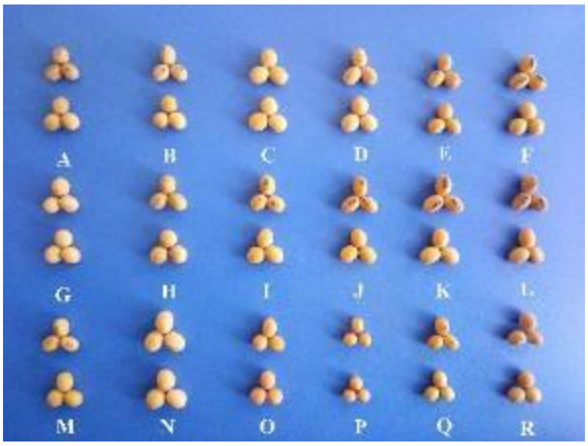


Figure 1. Soybean seeds in various accessions (online color). A) Kong25-1, B) Kong26-1, C) Kong27-1, D) Kong28-1, E) Kong29-1, F) Kangwon1-1, G) Kangwon30-1, H) KuNul5-1, I) Duiguru13-1, J) Duiguru14-1, K) Duiguru17-1, L) Duiguru19-1, M) Duiguru20-1, N) Duiguru21-1, O) Haqjak40, P) Dongnong50, Q) Gansokji1-1, R) KuNul3-1.

Figure 2. Acquiring canopy image using a smartphone camera fixed with the selfie stick under natural light in field (online color).

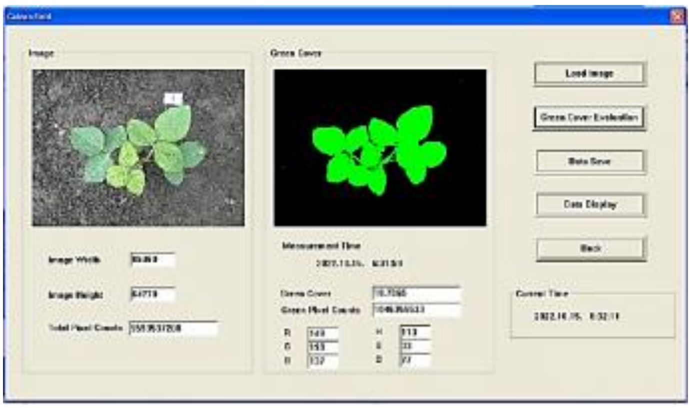
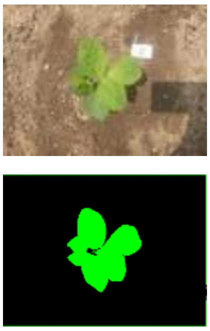
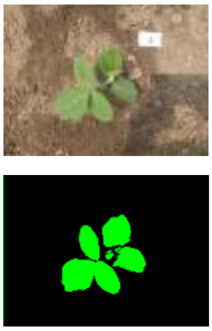
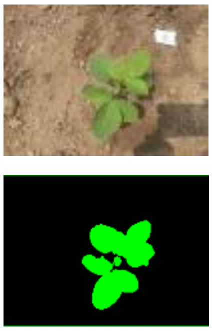
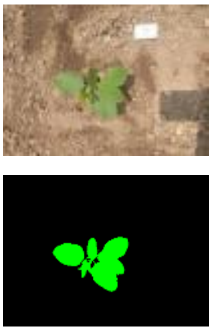


Figure 3. Canopy image processing using FCM algorithm (online color).


A

B

C

D

E

F

G

H

I

J

K

L

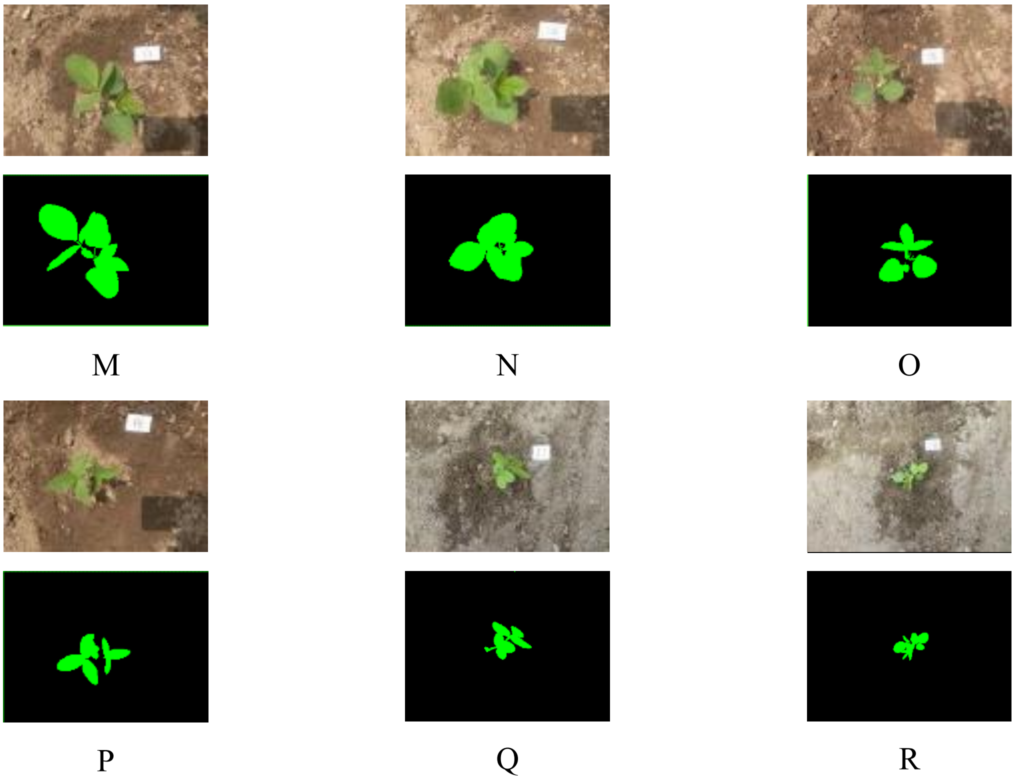
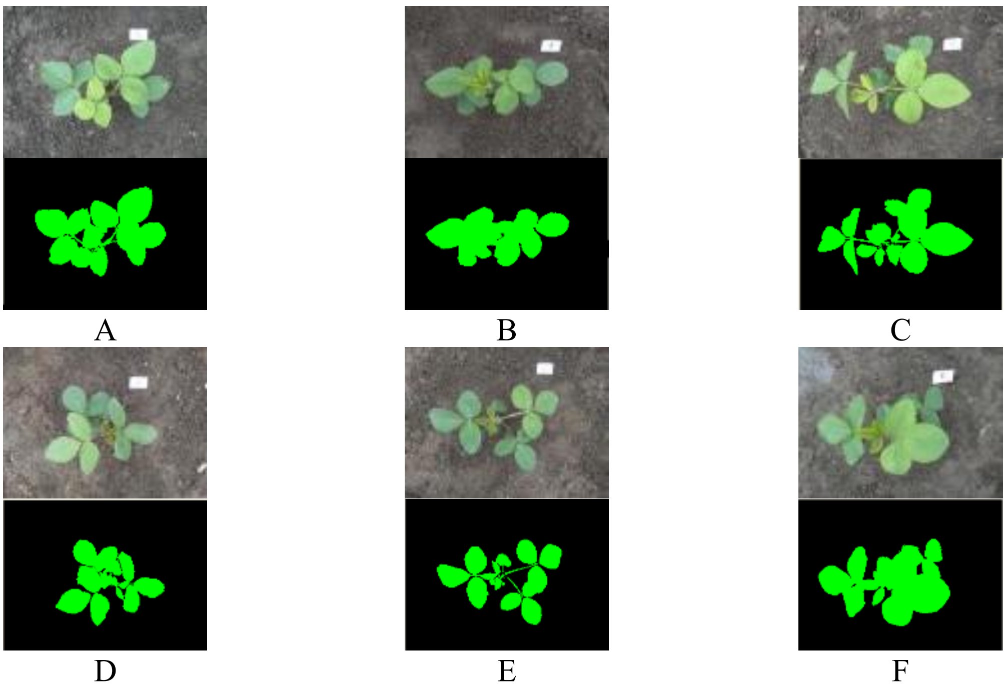


Figure 4. Original canopy image datasets taken from a plant grown during 20 days after sowing in 18 accessions (left) and corresponding canopy RGB images of processed with IA software (right) in 2021 (online color). A) Kong25-1, B) Kong26-1, C) Kong27-1, D) Kong28-1, E)Kong29-1, F) Kangwon11-1, G) Kangwon30-1, H) KuNul5-1, I) Duiguru13-1,J) Duiguru14-1, K) Duiguru17-1, L) Duiguru19-1, M) Duiguru20-1, N) Duiguru21-1, O) Haqjak40, P) Dongnong50, Q) Gansokji1-1, R) KuNul3-1.



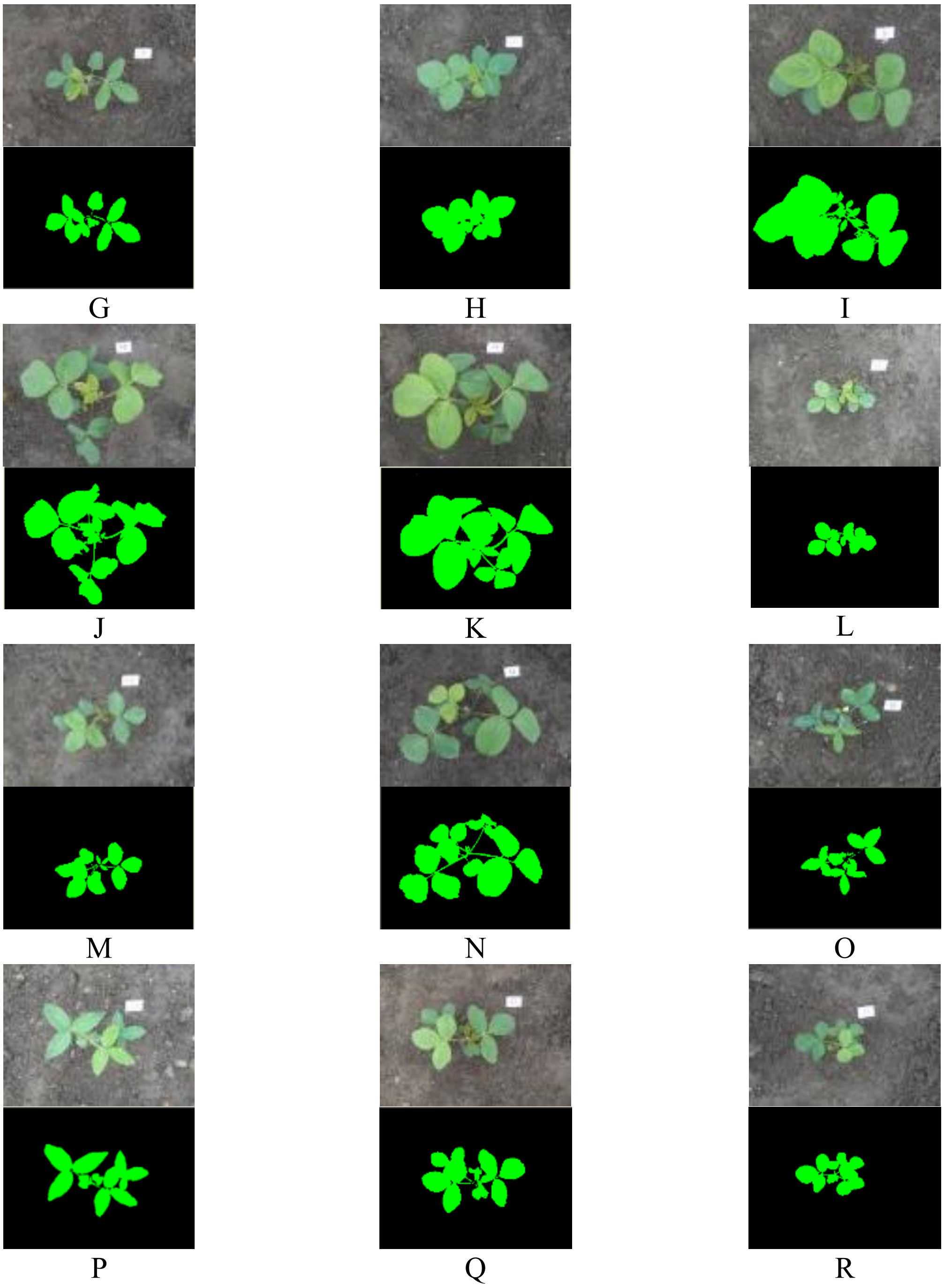


Figure 5. Original canopy image datasets taken from a plant grown during 27 days after sowing in 18 accessions (left) and corresponding canopy RGB images of processed with IA software (right) in 2022 (online color). A) Kong25-1, B) Kong26-1, C) Kong27-1, D) Kong28-1, E)Kong29-1, F) Kangwon11-1, G) Kangwon30-1, H) KuNul5-1, I) Duiguru13-1,J) Duiguru14-1, K) Duiguru17-1,L) Duiguru19-1, M) Duiguru20-1, N) Duiguru21-1, O) Haqjak40, P) Dongnong50, Q) Gansokji1-1, R) KuNul3-1.

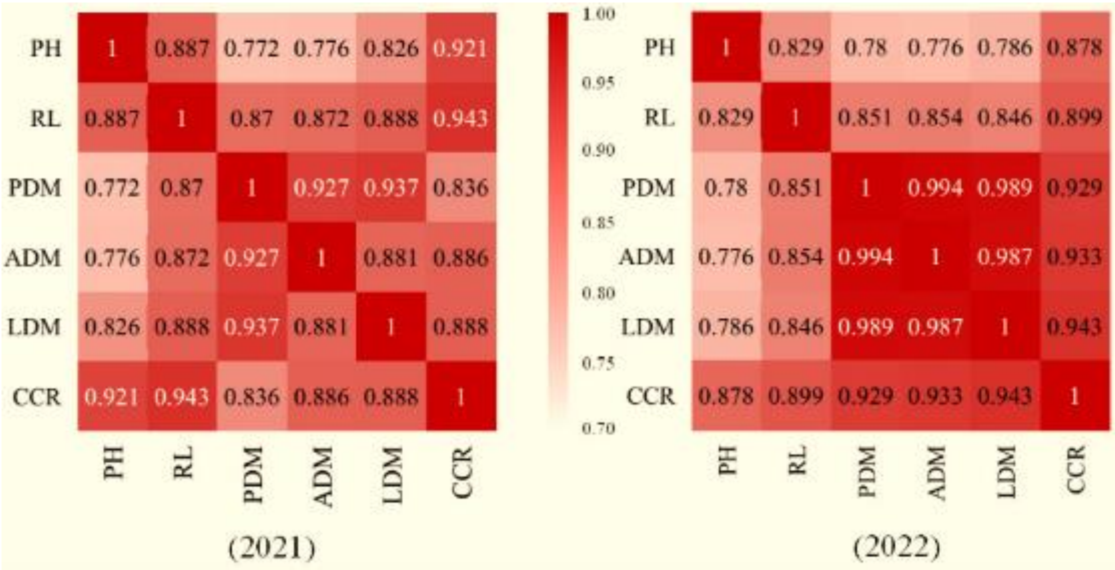
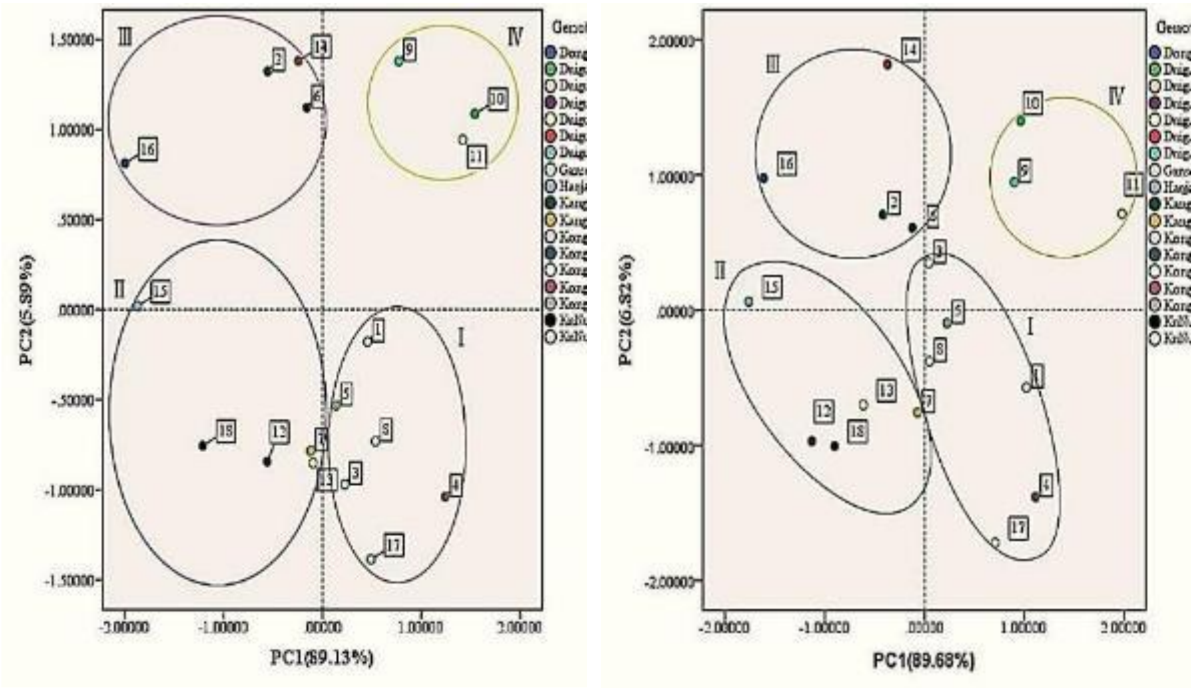
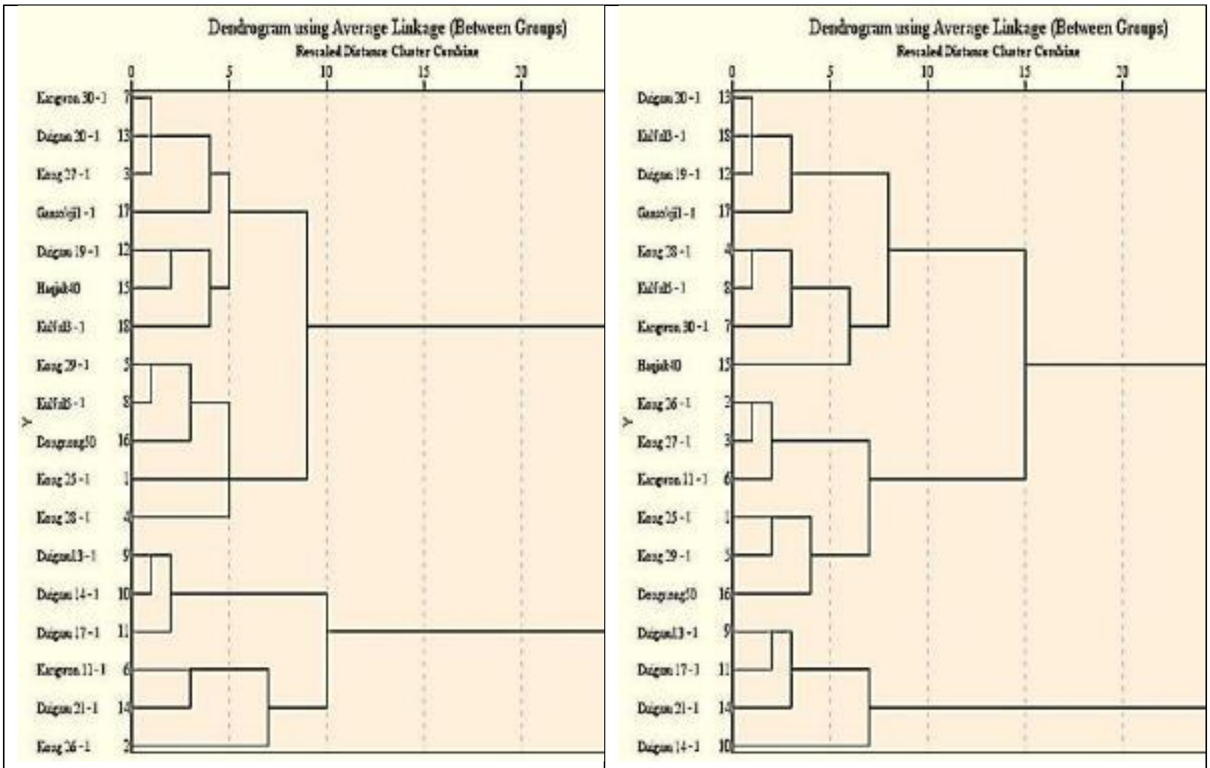


Figure 6. Heatmap of Pearson correlation coefficients between CCR and morphological/ biomass traits in 18 accessions.



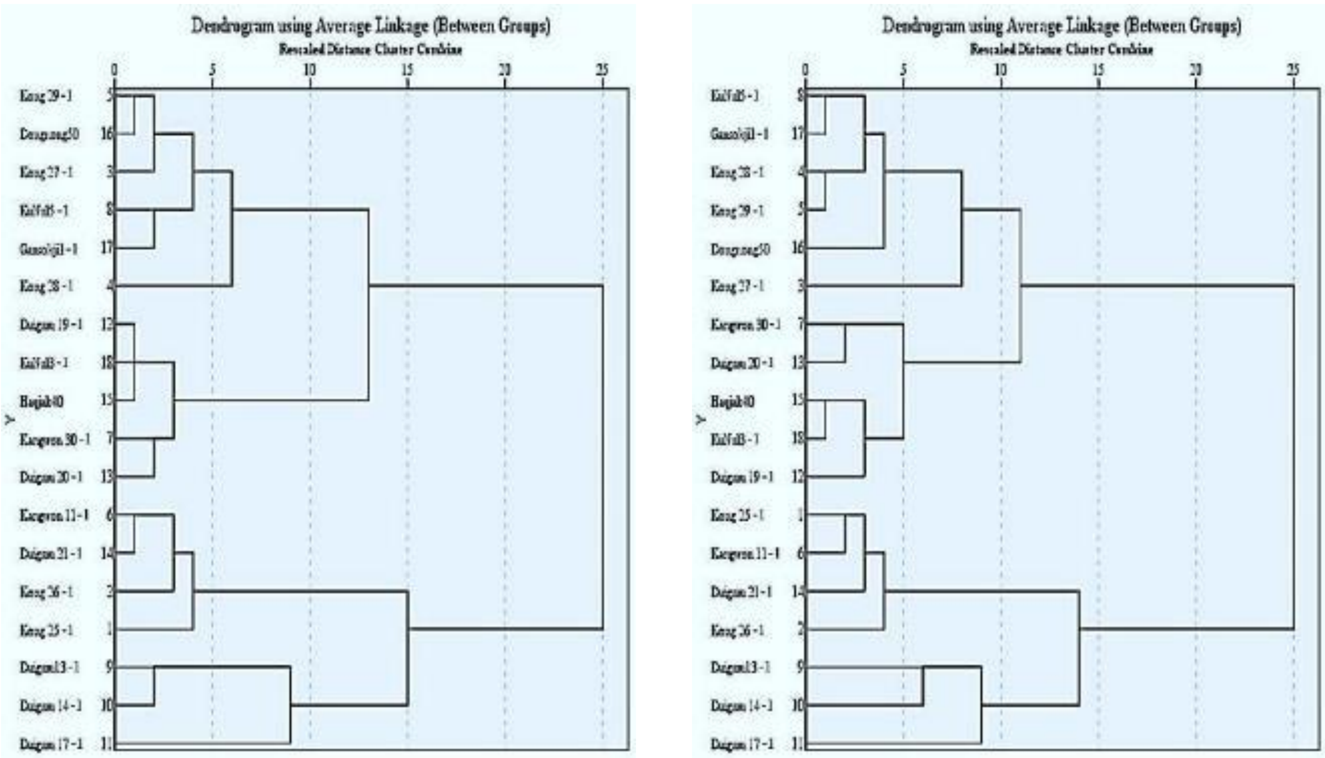
(in 2021) (in 2022)

Figure 7. Scatter plots by PCA based on morphological/biomass traits (online color).



(in 2021) in 2022)

Figure 8. Average linkage, rescaled distance cluster combine dendrograms obtained by HCA of the 18 accessions based on five morphological / biomass traits (online color).



(in 2021) (in 2022)

Figure 9. Average linkage, rescaled distance cluster combine dendrograms obtained by HCA of the 18 accessions based on CCR (online color).