Principal component analysis for assessment of variability in phenological and morphological traits in French bean (*Phaseolus vulgaris* L)

Rani Shama, Nayeema Jabeen and Parvaze A. Sofi

ISSN: 0975-928X

Volume: 10

Number:4

EJPB (2019) 10(4):1569-1575

DOI:10.5958/0975-928X.2019.00201.1

https://ejplantbreeding.org

Electronic Journal of Plant Breeding, 10 (4): 1569-1575 (Dec 2019) DOI: 10.5958/0975-928X.2019.00201.1
ISSN 0975-928X

Research Article

Principal component analysis for assessment of variability in phenological and morphological traits in French bean (*Phaseolus vulgaris* L)

Rani Shama[1], Nayeema Jabeen[1] and Parvaze A. Sofi [2*]
[1]Division of Vegetable Sciences, SKUAST-K, Shalimar, 190025
[2] Dry land Agricultural Research Station, SKUAST-K, Budgam, 190007
***E-Mail:** parvazesofi@gmail.com

(Received: 17 Feb 2019; Revised: 10 Dec 2019; Accepted: 10 Dec 2019)

Abstract

Forty four french bean (*Phaseolus vulgaris* L.) genotypes were evaluated for phenological, morphological, pod and seed traits in an augmented block design and variability was assessed through principal component analysis. PCA concentrated variability in first six principal components. The total variance explained with the first six PC's was 80.366 %. Latent roots (Eigen values) for significant PCs ranged from 3.390 (PC1) to 1.056 (PC5). The first two PC's that were used for constructing biplot graphs explained 41.259 %. The first PC contributed 22.601% of total variation mainly contributed by SYPP, NPPP, followed by PYPP, PGR and SPP. The second component explained 18.658 % of variation contributed largely by DF, followed by DPF, DM, PGR and PL, while as the third, fourth, fifth and sixth component explained 13.615, 10.390, 8.064 and 7.038 % of variation respectively. In the present study, seven variables including SYPP, PYPP, PGR, NPPP, DF, DPF, and DM contributed above the expected average to the variability in PC1 and PC2. Based on the factor loading graph, seed yield per plant is strongly correlated with number of pods and seeds per pod. Similarly pod yield per plant is strongly correlated with pod growth rate, pod length and number of pods per plant. Seed yield per plant is negatively correlated with seed length and plant height while as no correlation with days to maturity can be visualized. The genotypic profiles in biplot revealed that genotypes WB-129, WB-371, WB-1187, WB-642 and WB-1518 have high pod yield per plant as all of them have high pod growth rate, high pod number as well as higher pod length.

Key words

French bean, Augmented block design, Principal component analysis, Biplot

Introduction

French bean (*Phaseolus vulgaris* L.), 2n=22 is one of the important leguminous vegetable crop. It belongs to the family Leguminosae (Fabaceae), Sub-family Papilionaceae, Order Leguminosales (Fabales), Subtribe Phaseolinae, Tribe Phaseolae. It is also known by many names as kidney bean, common bean, field bean, garden bean, bush bean, navy bean, haricot bean, pinto bean, string bean, marrow bean and snap bean *etc.* It is one of the most significant leguminous vegetable, which is grown for fresh pod consumption and for processing as a frozen vegetable in many countries. This vegetable plays a vital role in the nourishment of human population (Ram, 2014). It is an important source of carbohydrate (61.4 per cent), proteins (17.5 - 28.5 per cent) and mineral matter (3.2 – 5.0 per cent). Among Vitamins, Vitamin C and pro Vitamin A are found in it. It has significant amount of fibre and supplies minerals like iron, potassium, phosphorus, magnesium, copper etc. Seeds are protein rich and are compared closely with meat, but nutritionally, this lack in amino acids like tryptophan and methionine. Green beans are eaten around the world. The green immature pods are cooked and eaten as a vegetable. Immature pods are marketed fresh, canned, frozen, whole, cut or uncut. It may be cooked as whole or eaten raw as salads or lightly steamed or sautéed. Green Bean Casserole is popular throughout United States as thanksgiving. Green bean chips and bean paste are quite famous among youngsters. French bean not only plays a vital role in nourishment of human population, but also improves soil fertility to a greater extent by virtue of being nitrogen fixing crop. It is widely cultivated in Tropics, Sub Tropics and Temperate regions. Globally French bean is grown over an area of 15.57 million hectares, with a production of 23.59 million tonnes (FAO, 2016). India accounts for about 0.23 million hectares of French bean area and 0.66 million tonnes of french bean production (FAO, 2016). In India, it is mainly grown in Himachal Pradesh, Jammu and Kashmir, Punjab, Haryana, Uttar Pradesh, Bihar, Gujarat, Madhya Pradesh, Maharashtra, Karnataka, Andhra Pradesh and Tamil Nadu. In Kashmir, French Bean is grown over an area of 24360 hectares, with a production of 14380 tones with the productivity of 600kg/hectare (Anonymous, 2016).

The major limitation with using seed yield *per se* as the selection criteria for crop improvement has proved to be less reliable on account of the complexity of its inheritance as well as low heritability. Therefore, the realised gains for yield have not been very encouraging by direct selection for yield. As a result, there has been greater emphasis across breeding programmes to elucidate the nature of relationship and trait associations between yield and other traits which are relatively less complex and have comparatively better heritability for use as indirect selection criteria for improving yield. The phenotypic correlation among traits reflects the observed relationship between traits arising out of both genetic and environmental factors; while as the genotypic correlations arise from linkage and pleiotropy. The knowledge of trait associations in breeding materials is essential for a variety of reasons: (i) it enables us to perceive the diversity of breeding material and identify the trait through which a crop plant is able to grow successfully in a given ecological condition with optimum productivity and to avoid characters that have little or no breeding value in combination with PCA; (ii) it also enables us to narrow down to a very few traits that not only account for large amount of variation but have a breeding value correlated with seed yield. The principal component analysis (PCA), one of multivariate analysis methods elucidates among a set of the traits which ones are decisive in genotypic differentiation (Kovacic, 1994). PCA enables easier understanding of impacts and connections among different traits by identifying them and explaining their roles. This method is a powerful multiple method to apply for evaluating yield component (Guertin and Bailey, 1982), identify biological relationships among traits (Acquaah *et al.*, 1992), decrease associated-traits to a few factors (Johnson and Wichern, 1996) and description of correlations among variables.

Materials and Methods

The present study was undertaken in *Kharif* 2018 in the research farm of Dryland Agricultural Research Station at Budgam (34^0 -01'North and 74^0 -47'E at an altitude of 5201 metres above sea level). The minimum and maximum temperature during experimental period was 14.87 and 28.97 ^{0}C, the minimum and maximum relative humidity was 55.68 and 80.14 per cent and the total rainfall recorded was 336.2 mm.

The material used in present study comprised of 44 genotypes including 40 test germplasm accessions (comprising local landraces a well as accessions procured from national and international gene banks) and four checks namely Shalimar French Bean-1, Arka Sharath, Arka Komal and Arka Anoop. Shalimar French Bean-1 is the variety released by SKUAST-Kashmir, whereas, Arka Sharath, Arka Komal and Arka Anoop are released by IIHR, Bangalore.

The material was evaluated in augmented block design (Federer, 1956). The design consisted of five blocks containing 12 genotypes in each with eight test entries and four check entries. Each genotype was represented by a plot size of 3 x 1.5 meter dimensions with four lines. The plants were space planted for optimal expression of traits. Data was collected from ten randomly selected competitive plants on various phonological (Days to flowering, days to maturity, days to pod set, days to pod fill, pod growth rate) and morphological (Plant height, number of pods per plant, pod length, seeds per pod, 100-seed weight, pod yield per plant**,** seed yield per plant, pod wall biomass, seed length, seed breadth) traits. In each block the checks were allotted randomly.

The analysis of variance for augmented block design was was done using SPAD (Statistical Package for Augmented Design) software developed by IASRI, New Delhi. Principal component analysis was based on Pearson correlation matrix and Euclidean distances. Principal component analysis was done using XLSTAT version 2018 (Addinsoft). The fact that Eigen values are above one indicates that the evaluated principle component weight values are reliable (Mohammadi and Prasanna, 2003). In the principal components where the values of Latent roots (Eigen values) was reduced to less than unity, which in present study occurred after the sixth principal component together accounting for more than 80 per cent of total variance and as such, the rest of the components were not considered

Results and Discussion

Principal component analysis (PCA) was done for fifteen traits (phenological, morphological, pod and seed traits) scored in the field experiment laid in an augmented block design. The number of principal components calculated from correlation matrix is 15 which is similar to number of observed traits. PCA concentrated variability in first six principal components. The total variance explained with the first six PC’s was 80.366 per cent (Table 1). The variance explained with last ten PC’s is irrelevant. In the principal components where the values of latent roots (Eigen values) was reduced to less than unity, which in present study occurred after the sixth principal component together accounting for more than 80% of total variance in the present experimental material, the rest of the components were not considered. Latent roots (Eigen values) for significant PCs ranged from 3.390 (PC1) to 1.056 (PC5). The first two PC’s that were used for constructing biplot graphs explained 41.259 per cent The first PC contributed 22.601 per cent of total variation mainly contributed by seed yield per plant (21.244 per cent), number of pods per plant

(20.325 per cent) followed by pod yield per plant (18.898 per cent), pod growth rate (12.585 per cent) and seeds per pod (7.916 per cent). The second component explained 18.658 per cent of variation contributed largely by days to flowering (24.172 per cent), followed by days to pod fill (23.222 per cent), days to maturity (15.991 per cent) , pod growth rate (10.146 per cent.) and pod length (9.072 per cent), while as the third, fourth, fifth and sixth component explained 13.615, 10.390, 8.064 and 7.038 of variation respectively (Table 2). Panchbhaiya *et al* (2017) also reported similar results in French bean with six PC's accounting for about 79 per cent of variation days to flowering, days to maturity and seed yield as important contributors. In the present study, if the contribution of the variables were uniform, the expected value would be 1/length of variables 1/15 = 6.67 per cent. The total contribution of a given trait (contribution), on explaining the variations retained by two PCs (PC1 and PC2) is given by contribution= [(C1 x Eig1) + (C2 x Eig2)]/(Eig1 + Eig2), where: C1 and C2 are the contributions of the variable on PC 1 and PC 2, respectively and Eig 1 and Eig 2 are the eigen values of PC1 and PC2, respectively (Kassambara, 2017). In the present study, seven variables including seed yield per plant, pod yield per plant, pod growth rate, number of pods per plant, days to flowering, days to pod fill, and days to maturity contributed above the expected average to the variability in PC1 and PC2.

The Pearson correlations among the traits are presented in Table 3. The GT data are approximately displayed in a GT biplot (Fig. 1a and 1b), which can be used to visualize the trait associations and the trait profiles of the genotypes. In terms of the trait-standardized GT data, when two vectors are close, forming a small angle (acute, $< 90^0$), the two variables they represent are strongly positively correlated. If vector rays meet each other at 90°, they are not likely to be correlated. Similarly, if the rays diverge and form a large angle (close to 180°), they are negatively correlated. In the present study, based on the factor loading graph (Figure 1a and 1b), seed yield per plant is strongly correlated with number of pods and seeds per pod. Similarly pod yield per plant is strongly correlated with pod growth rate, pod length and number of pods per plant. Seed yield per plant is negatively correlated with seed length and plant height while as no correlation with days to maturity can be visualized. Similar results have also been reported in GT biplot analysis in common bean by Oliveira *et al* (2018). The earlier studies also reported that pods per plant contributed to PC1 (Panchbhaiya *et al*., 2017, Araujo and Vivas, 2018). Similarly, among other trait correlations, number of pods per plant is negatively correlated with pod length, 100-seed weight and seeds per pod. The Genotype by Trait (GT) biplot also enables us to compare genotypes on the basis of the measured multiple variables and also identifies genotypes that are particularly superior in certain traits. The GT biplot can thus be effectively used as independent selection criteria based on several traits and in yield trials for grain yield evaluation (Yan and Rajcan, 2002). The vector length (*i.e*., the distance to the biplot origin) of a trait indicates how well the trait is represented in the biplot; a relatively short vector indicates that the variation of the trait across genotypes is either small or not well presented in the biplot, which is due to its weak or lack of correlation with other traits (Yan and Fregeau-Reid, 2018). This invariably occurs due to poor goodness of fit of the biplot as the two PCs (PC1 and PC2) account for only a part of total variation (the goodness of fit of the GT biplot in Fig. 1a and 1b is 41.259 per cent). The vector length of a genotype indicates whether it is intermediate for all traits or has clear strengths and/or weaknesses in its trait profile. The GT biplot in Fig. 1a and 1b also shows the trait profiles of the genotypes, the accuracy of which also depends on the goodness of fit of the biplot. Figure 1b representing genotypic profiles reveals that genotypes WB-129, WB-371, WB-1187, WB-642 and WB-1518 have high pod yield per plant as all of them have high pod growth rate, high pod number as well as higher pod length. Genotypes with low pod yield as they have lower values for these traits and fall opposite to these traits.

Genotype by Trait (GT) biplots is highly effective in elucidating the interrelationships among the traits and provides an effective tool for visual comparison among genotypes on the basis of multiple traits. Also, it can be used in independent culling based on multiple traits and in comparing selection strategies (Yan and Rajcan, 2002). The method of calculation to find the relation between traits in GT biplot is different from simple correlation coefficient because the GT biplot approach is constructed by the first two principal components (PC1 and PC2), so in this approach, the total yield variation cannot be explained. In addition, GT biplots graphically describes the interrelationships among all measured traits on the basis of overall pattern of the data for different genotypes and traits, whereas simple correlation coefficients only describe the relationships between two traits (Yan and Rajcan, 2002). Therefore, simple correlation between traits does not agree with those of relation between traits in GT biplot completely. In light of above, interrelationships among measured traits on the basis of overall pattern of the data (GT biplot) are better than the simple correlation coefficients that only describe the relationships between two traits. Genotypes by Trait (GT) biplots also provide a tool for visual comparison among genotypes on the basis of multiple traits. Yan and Tinker (2006) stated that the length of the genotype vector, which is the distance between a genotype and the biplot origin, measures the difference of the genotype from the "average" genotype. Therefore, genotypes or any treatment or variables with the longest vectors are either the best or the poorest genotypes. Despite being located on the vertex of the polygon, they are not always the best answer. If they are located on the left side of the biplot, these genotypes show the worst values, and care should be taken to not have an erroneous interpretation.

References

Acquaah, G., Adams, M.W. and Kelly, J.D. 1992. A factor analysis of plant variables associated with architecture and seed size in dry bean, *Euphytica.,* **60:** 171-177.

Anonymous. 2016. Ministry of Agriculture and Farmers Welfare Annual Report.

Araujo, K. C., and Vivas, M. 2018. Multivariate analysis used as a tool to select snap bean (*Phaseolus vulgaris* L.) genotypes. *Australian Journal of Crop Science.,* **12(1)**: 67.

FAO. 2016. FAOSTAT. www.fao.org

Federer, W. 1956. Augmented designs. Hawaiian Planter Recorder., **55:** 191-208

Guertin, W.H. and Bailey J.P. 1982. Introduction to modern factor analysis. Edwards Brothers (ed) Inc Michigan USA.

Johnson, R.A. and Wichern, D.W. 1996. Applied multivariate statistical analysis. Sterling Book House New Delhi.

Kassambara, A. 2017. Multivariate analysis: *Practical guide to cluster analysis in R: unsupervised machine learning. (*www.sthda.com)

Kovacic, Z. 1994. Multivarijaciona analiza. Univerzitet u Beogradu, Ekonomski fakultet, 283 pp.

Mohammadi, S.A. and Prasanna, B.M. 2003. Analysis of Genetic Diversity in Crop Plants—Salient Statistical Tools and Considerations. *Crop Sci*., **43:** 1235-1248

Oliveira, T. R. A. D., Gravina, G. D. A., Oliveira, G. H. F. D., Araújo, K. C., Araújo, L. C. D., Daher, R. F., and Cruz, D. P. D. 2018. The GT biplot analysis of green bean traits. *Ciência Rural.,* **48(6)**: e20170757

Panchbhaiya, A., Singh, D.K., Priyanka Verma, Jatav, V. and Maurya, A.K. 2017. Genetic analysis of French bean (Phaseolus vulgaris l.) germplasm through principal component analysis and D2 cluster analysis. *Journal of Pharmacognosy and Phytochemistry.,* **6(3):** 537-542

Ram, H.H. 2014. Vegetable breeding principles and practices. Kalyani Publishers, New Delhi. pp 439

Yan, W. and I. Rajcan, 2002. Biplot analysis of test sites and trait relations of soybean in Ontario. *Crop Sci.,* **42:** 11-20.

Yan, W., and Tinker, N. A. 2006. Biplot analysis of multi-environment trial data: Principles and applications. *Canadian journal of Plant Science.,* **86(3):** 623-645.

Yan, W., and Frégeau-Reid, J. 2018. Genotype by Yield* Trait (GYT) Biplot: a novel approach for genotype selection based on multiple traits. *Scientific reports.,* **8:** 8242

Table 1. Eigen values (Latent roots) and rotated component loadings (values of principal component traits of french bean)

Component	Eigen value	Variability (%)	Cumulative %
PC1	3.390	22.601	22.601
PC2	2.799	18.658	41.259
PC3	2.042	13.615	54.874
PC4	1.559	10.390	65.264
PC5	1.210	8.064	73.328
PC6	1.056	7.038	80.366
PC7	0.726	4.841	85.208
PC8	0.648	4.317	89.525
PC9	0.543	3.620	93.145
PC10	0.383	2.552	95.696
PC11	0.294	1.960	97.656
PC12	0.167	1.111	98.767
PC13	0.134	0.893	99.660
PC14	0.032	0.211	99.871
PC15	0.019	0.129	100.000

Table 2. Trait contributions to principal components and total contribution to PC1 and PC2

Trait	PC1	PC2	PC3	PC4	PC5	PC6	Total contribution to PC1 and PC2
DF	0.617	**24.172**	0.277	0.646	6.373	3.279	**11.269**
DM	3.478	**15.991**	2.100	0.000	24.565	1.940	**9.137**
DPS	0.189	0.047	22.612	3.061	5.218	21.591	0.124
DPF	0.440	**23.222**	2.248	3.802	0.007	7.745	**10.743**
PGR	**12.585**	**10.146**	3.128	4.062	8.105	0.726	**11.481**
PH	1.868	0.156	11.297	6.676	1.465	22.805	1.093
PL	3.423	9.072	10.677	0.585	6.573	7.933	5.977
NPPP	**20.325**	0.609	0.095	0.002	11.362	4.592	**11.408**
SPP	**7.916**	1.618	12.101	0.002	1.968	11.150	5.067
PYPP	**18.898**	4.022	1.079	1.736	6.121	0.034	**12.170**
PWBM	0.051	4.191	19.728	7.430	5.395	2.274	1.923
SL	3.587	0.072	3.032	36.268	0.601	1.845	1.997
SB	5.221	2.321	2.660	16.781	0.060	4.795	3.909
100SW	0.160	2.366	8.821	18.816	11.259	6.743	1.157
SYPP	**21.241**	1.994	0.146	0.133	10.929	2.547	**12.536**

ABBREVIATIONS: DF: Days to flowering, DM: Days to maturity, DPS: Days to pod set, DPF: Days to pod fill, PGR: Pod growth rate, PH: Plant height, NOP: Number of pods per plant, PL: Pod length, SPP: Seeds per pod, 100SW: 100-seed weight, PYPP: Pod yield per plant**,** SYPP: seed yield per plant, PWBM: Pod wall biomass, SL: Seed length, SB: Seed breadth

Table 3. Correlation matrix for 15 traits in French bean (values in bold are significant p≤0.05)

Trait	DF	DM	DPS	DPF	PGR	PH	PL	NPPP	SPP	PYPP	PWBM	SL	SB	100SW	SYPP
DF	1	**0.705**	0.095	**-0.642**	0.216	0.089	0.168	-0.101	-0.144	0.092	0.240	0.040	-0.160	-0.196	-0.181
DM		1	**-0.314**	**-0.540**	0.019	-0.003	0.003	-0.212	-0.097	-0.144	0.248	0.065	0.004	-0.115	-0.303
DPS			1	0.072	-0.042	-0.084	0.212	0.060	**-0.346**	-0.005	**0.304**	-0.003	-0.213	0.151	0.079
DPF				1	**-0.494**	-0.097	-0.218	0.163	0.233	-0.162	0.002	-0.110	-0.034	0.082	0.230
PGR					1	-0.112	**0.401**	**0.342**	0.253	**0.898**	-0.126	-0.121	-0.177	-0.126	0.296
PH						1	-0.274	-0.161	0.038	-0.180	-0.216	-0.195	0.135	-0.215	-0.178
PL							1	0.106	-0.059	**0.380**	**0.411**	0.193	**-0.372**	-0.015	0.101
NPPP								1	**0.328**	**0.507**	0.019	-0.304	-0.197	0.197	**0.935**
SPP									1	**0.309**	-0.308	-0.168	-0.142	-0.104	**0.476**
PYPP										1	-0.077	-0.184	-0.253	-0.081	**0.455**
PWBM											1	-0.088	-0.203	0.016	-0.033
SL												1	**0.352**	**0.333**	-0.260
SB													1	0.107	-0.217
100SW														1	0.280
SYPP															1

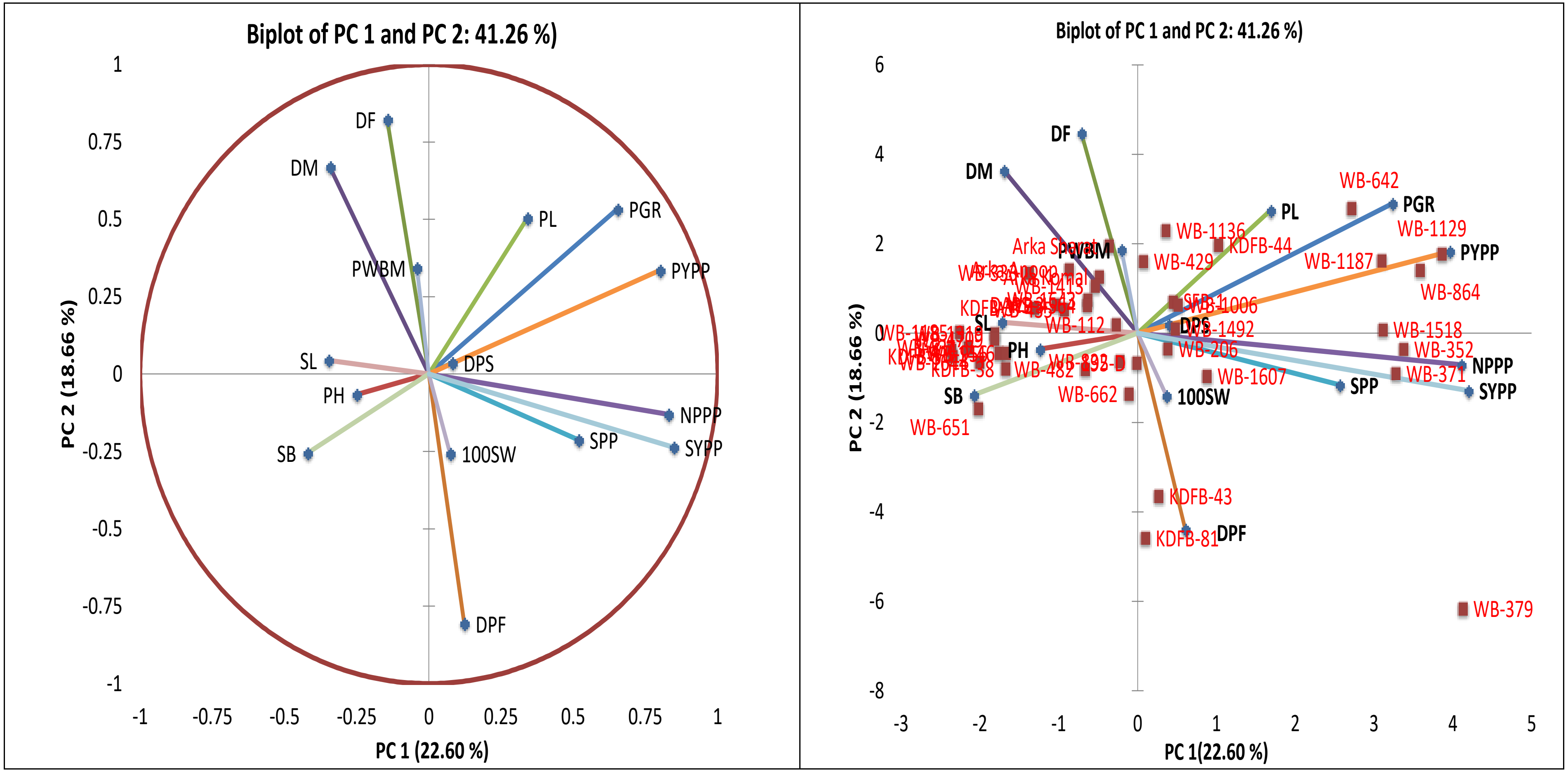

Fig. 1a and 1b: Trait biplot and GT biplot based on PC1 and PC2

https://ejplantbreeding.org