

**Application of GGE biplot analysis to evaluate Genotype (G),  
Environment (E) and GxE interaction on *P. radiata*: a case  
study**

Meimei Ding<sup>1</sup>, Bruce Tier<sup>1</sup> and Weikai Yan<sup>2</sup>

e-mail: [mding@une.edu.au](mailto:mding@une.edu.au); Tel: +61 2 67733780

<sup>1</sup> Animal Genetics and Breeding Unit\*, University of New England, Armidale, NSW, 2351, Australia

<sup>2</sup> Eastern Cereal and Oilseed Research Centre, Agriculture and Agri-Food Canada, 960 Carling Ave, Ottawa, ON, Canada, K1A 0C6

---

\* AGBU is a joint venture of NSW Department of Primary Industries and the University of New England

## **Abstract**

GGE biplot analysis is an effective method which is based on principal component analysis (PCA) to fully explore multi-environment trials (METs). It allows visual examination of the relationships among the test environments, genotypes and the genotype-by-environment interactions (GxE interaction). The METs of *P. radiata* with 165 families in five environments was used to demonstrate GGE biplot analysis. The graphics reveal two non-overlapping clusters of environments. Group one includes three sites of PT5459, RAD211 and VRC060, RS27A and RS27B in the second group. The graphics also identifies the best family for each environment. Correlations range from 0.98 to -0.50 between the five sites.

Key words: GGE biplot, PCA, GxE interaction, *P. radiata*

---

Abbreviations: G, genotype; E, environment; GxE interaction, genotype-by-environment interaction; PCA, principal component analysis.

## Introduction

Phenotypes are a mixture of genotype (G) and environment (E) components and interactions (GxE) between them. GxE interactions complicate the process of selecting of genotypes with superior performance. Consequently, METs are widely used by plant breeders to evaluate the relative performance of genotypes for target environments (Delacy et al., 1996). Numerous methods have been developed to reveal patterns of GxE interaction, such as joint regression (Finlay and Wilkinson, 1963; Eberhart and Russel, 1966; Perkins and Jinks, 1968), additive main effects and multiplicative interaction (AMMI, Gauch, 1992) and type B genetic correlation (Burdon, 1977; Yamada, 1962). These methods are commonly used to analyse METs data and have also been applied on GxE interaction in *P. radiata* (Ades et al., 1997; Johnson et al., 1989; Wu et al., 2005).

GGE biplot analysis was recently developed to use some of the functions of these methods jointly. In total phenotypic variation, E explains most of the variation and G and GE are usually small (Yan, 2002). However, only the G and GxE interaction are relevant to cultivar evaluation particularly when GxE interaction is identified as repeatable (Hammer et al., 1996). Hence, Yan et al. (2000) deliberately put the two together and referred to the mixture as GGE. Following the proposal of Gabriel (1971), the biplot technique was used to display the GGE of a METs data, referred to as a GGE biplot (Yan, 2001; Yan et al., 2000).

GGE biplot is a data visualization tool, which graphically displays a GxE interaction in a two way table (Yan, 2000). GGE biplot is an effective tool for: 1) mega-environment analysis (e.g. “which-won-where” pattern), whereby specific genotypes can be recommended to specific mega-environments (Yan, 2003; Yan, 2006), 2) genotype evaluation (the mean performance and stability), and 3) environmental evaluation (the power to discriminate among genotypes in target environments).

GGE biplot analysis is increasingly being used in GxE interaction data analysis in agriculture (Butron, 2004; Crossa, 2002; Dehghani et al., 2006; Hunt, 2002; Kaya et al., 2006; Samonte et al., 2005; Tinker, 2005; Yan, 2000; Yan, 2002; Yan, 2003; Yan, 2006). However, there has been no report of its application to forestry so far. As a case study, the purposes of this paper are attempted to apply the technique to reveal the patterns of

GxE interaction on *P. radiata*, and also to compare the method with previous analysis using other methods.

## Materials and method

The genetic materials originated from Australia-wide diallel mating experiments. The details have been described by Wu et al. (2005). Five sites were chosen for this study. They distributed in four regions of Australia: Busselton (RS27A and RS27B), Myrtleford (RAD211), Traralgon (VRC060) and Mount Gambier (PT5459) which represent the broad range of commercial environments. The span of latitude is 4°22' from 33° 52'S to 38°14'S, longitude is 30°36' from 115°58'E to 146° 34'E. In each site, there were the same 165 full-sib families. The seedlings were planted in two site types, which were second-rotation radiata pine (2nd PR) and previous pasture crop site (pasture). Each trial was a randomized incomplete block design with 3 replicates. The growth trait of diameter at breast height (DBH) was accessed at 10.5 years of age. The details are shown in Table 1. Least square means (LSmeans) of families in each site were calculated using SAS Mixed procedure (SAS, 1991).

## The models for a GGE biplot

The model for a GGE biplot (Yan, 2002) based on singular value decomposition (SVD) of first two principal components is:

$$Y_{ij} - \mu - \beta_j = \lambda_1 \xi_{i1} \eta_{j1} + \lambda_2 \xi_{i2} \eta_{j2} + \varepsilon_{ij} \quad [1]$$

where  $Y_{ij}$  is the measured mean (DBH) of genotype  $i$  in environment  $j$ ,  $\mu$  is the grand mean,  $\beta_j$  is the main effect of environment  $j$ ,  $\mu + \beta_j$  being the mean yield across all genotypes in environment  $j$ ,  $\lambda_1$  and  $\lambda_2$  are the singular values (SV) for the first and second principal component (PC1 and PC2), respectively,  $\xi_{i1}$  and  $\xi_{i2}$  are eigenvectors of genotype  $i$  for PC1 and PC2, respectively,  $\eta_{1j}$  and  $\eta_{2j}$  are eigenvectors of environment  $j$  for PC1 and PC2, respectively,  $\varepsilon_{ij}$  is the residual associated with genotype  $i$  in environment  $j$ .

PC1 and PC2 eigenvectors cannot be plotted directly to construct a meaningful biplot before the singular values are partitioned into the genotype and environment eigenvectors. Singular-value partitioning is implemented by,

$$g_{il} = \lambda_l^{f_l} \xi_{il} \text{ and } e_{lj} = \lambda_l^{1-f_l} \eta_{lj} \quad [2]$$

where  $f_l$  is the partition factor for  $PC_l$ . Theoretically,  $f_l$  can be a value between 0 and 1, but 0.5 is most commonly used.

To generate the GGE biplot, the formulae [1] was presented as:

$$Y_{ij} - \mu - \beta_j = g_{i1}e_{1j} + g_{i2}e_{2j} + \varepsilon_{ij} \quad [3]$$

If the data was environment-standardized, the common formulae for GGE biplot was reorganized as follows:

$$(Y_{ij} - \mu - \beta_j) / s_j = \sum_{l=1}^k g_{il}e_{lj} + \varepsilon_{ij} \quad [4]$$

where  $s_j$  is the standard deviation in environment  $j$ ,  $l=1, 2, \dots, k$ ,  $g_{il}$  and  $e_{lj}$  are  $PC_l$  scores for genotype  $i$  and environment  $j$ , respectively.

We used environment standardized model [4] to generate biplot of “which-won-where”. For the analysis of relationship between the trials, genotype and environment evaluation, we used unstandardized model [3]. The analyses were conducted and biplots generated using the “GGEbiplot” software (Yan, 2005).

The comparison was made between the results of this GGE biplot analysis and the results from previous analysis with joint regression and type B genetic correlation.

## Results

The first two PCs explain 54.5% (PC1=30.6%, PC2=23.9%) of the total GGE variation using environment-standardized model, similar as using unstandardized model, which explains 55.3% (PC1=30.4%, PC2=24.9%) of total GGE variation.

The results are presented as four sections: section one represents the results of “which-won-where” to identify the best genotypes for each environment; section two shows the relationship between the sites and the groups of environments; section three gives the results of family performance and their stability; section four visualizes the performance

of different genotypes in one environment (PT5459) and the relative adaptation of one genotype (family 59) to different environments;

#### 1. The “which-won-where” patterns

The polygon view of the GGE biplot (shown in figure 1) indicates the best genotype(s) in each environment and groups of environments (Hunt, 2002). The polygon is formed by connecting the markers of the genotypes that are furthest away from the biplot origin such that all other genotypes are contained in the polygon. The rays are lines that are perpendicular to the sides of the polygon or their extension (Yan, 2002). In figure 1, ray 1 is perpendicular to the side that connects family 9 and family 59, ray 2 is perpendicular to the side that connects family 59 and family 178, so on. These seven rays divide the biplot into seven sections, and five sites fall into four of them. The vertex families for each quadrant are the one that gave the highest yield for the environments that fall within that quadrant. The highest yield in environment PT5459 is family 178, in RAD211 and VRC060 are family 148, in RS27A is family 59, in RS27B is family 9, respectively. The other vertex families - 165, 173 and 103 are poorest in all five sites.

#### 2. Interrelationship among environments

Figure 2 provides the summary of the interrelationships among the environments. The lines that connect the biplot origin and the markers for the environments are called environment vectors. The angle between the vectors of two environments is related to the correlation coefficient between them. The cosine of the angle between the vectors of two environments approximates the correlation coefficient between them (Kroonenberg, 1995; Yan, 2002). Based on the angles of environment vectors, the five sites are grouped into two groups. One group includes PT5459, RAD211 and VRC060. Group two involves RS27A and RS27B. This result is coincident with the geographic pattern which belongs to different site type, 2nd PR and ex-pasture, respectively.

Table 2 shows the comparison of the correlation coefficients (cosine angles) and type B genetic correlations from previous analysis. The smallest angle is between RAD211 and VRC060, implying there are the highest correlation between them. The approximate correlation coefficient is 0.98. Subsequently, the smaller angle occurs between VRC060 and PT5459. The angles between RAD211 and RS27B, and VRC060 and RS27B are

greater than  $90^\circ$ , showing the negative correlation between them. The results approximately reflect the relationship among the five sites as a previous analysis.

### 3. Family mean yield and their stability

Figure 3 shows the ranking of 165 families of their mean yield and stability. The line passing through the biplot origin is called the average environment axis (AEC), which is defined by the average PC1 and PC2 scores of all environments. More close to concentric circles indicates higher mean yield. The line which passes through the origin and is perpendicular to the AEC with double arrows represents the stability of genotypes (defined as A line). Either direction away from the biplot origin, on this axis, indicates greater GxE interaction and reduced stability (Yan, 2002). For broad selection, the ideal genotypes are that have both high mean yield and high stability (defined as genotype group one). In the biplot, they are close to the origin and have the shortest vector from the AEC. Family 57, 15, 24 belong to this group for environment group one. On the other hand, for specific selection, the ideal genotypes are that have high mean yield but low stability and respond best to particular environments. For group one environment (PT5459, RAD211 and VRC060), the mean yield of family is in the following order: the highest is family 57, then 143 and so on. The worst is family 19. For group two (RS27A and RS27B), the mean yield of family is in the following order: the highest is family 59, then family 41, so on. The worst is family 9.

### 4. Examining the genotypes and environments

Figure 4a shows the performance of different genotypes in PT5459. A line that passes through the biplot origin and labelled PT5459 is the PT5459 axis. The genotypes are ranked according to their projections on to the PT5459 axis. The line passing through the biplot origin and perpendicular to the PT5459 axis separates genotypes that yield below the mean and above the mean in PT5459. Above the mean genotypes ranking are the families in right side of line A, e.g., families 57, 143 and 148. Families 149, 165 and 3 in the left side of line A are below the mean genotypes ranking.

Figure 4b shows the results for family 59 in five sites. The line passes through the biplot origin and labelled family is the family 59 axis. The environments are ranked along the family 59 axis in the direction towards the labeled for the family 59. Thus, the relative performance of family 59 in different environments follows the order as,

RS27A>RS27B>PT5459>VRC060>RAD211. The perpendicular to the family 59 axis separates environments in which family 59 is below and above the mean. However, family 59 is above the mean in all five sites.

### **Discussion and conclusion**

Compared to the methods of AMMI, joint regression and type B genetic correlation, GGE biplot analysis integrates some features from all of them. It allows visual interpretation of GxE interaction.

The first is to visualize the interaction between genotypes and environments (which-won-where). Two biplots can be used to visualize “which-won-where”, AMMI and GGE biplot which are both based on the statistical model of principal component analysis (PCA). The differences of the two methods, GGE biplot analysis is based on environment-centred PCA, whereas AMMI analysis is referred to double centred PCA (Kroonenberg, 1995). AMMI stands for the additive main effect and multiplicative interaction (Gauch, 1992), and GGE biplot stands for genotype main effect plus GxE interaction (Ma, 2004). However, if the purpose for “which-won-where”, AMMI could be misleading (Yan, 2006). Superior to AMMI, the GGE biplot has many visual interpretations that an AMMI does not have, particularly allows visualization of any crossover GxE interaction. This part of GxE interaction is usually essential to breeding program. In addition, comparing with different AMMI family models (AMMI0 to AMMIk, Dias, 2003; Zobel, 1988), GGE biplot is always close to the best AMMI models in most cases (Ma, 2004). Moreover, GGE biplot is more logical and biological for practice than AMMI in terms of explanation of PC1 score, which represents genotypic effect rather than additive main effect (Yan, 2000).

The second is to show the interrelationship among the environments which is approximately equal to the type B genetic correlation between the trials. The graphic of interrelationships among environments displays the correlation between the trials accurately. In terms of the relative trend of relationship between the trials, the GGE biplot shows the same pattern as the estimates using type B genetic correlation. The highest correlation occurs between RAD211 and VRC060, the negative or none occur between

RS27B and the other trials in group one. Inconsistencies can occur because biplot did not explain 100% of the GGE variation.

The third is to visualize the interrelationship among genotypes which the ranking of families based on both mean performance and stability. The visualizing graphic of genotype means and their stability shows different genotype groups which were classified four groups. Group one is highly desirable which is high yield and high stability. The group with high yield but low stability is desirable for specific selection, whereas low yield and low stability is possible for special breeding purposes, e.g. drought resistance selection. The most undesirable group is low yield but high stability. The classification is similar to the previous work of family behaviour plots using joint regression (Finlay and Wilkinson, 1963). However, GGE biplot not only shows different genotype groups, but also show their favourite environments. If one assumes the classical method of joint regression is informative, the comparison of the results indicates the repeatability is approximately 95% using these two methods.

The limitations of the GGE biplot are that it may explain only a small proportion of the total GGE. This can happen when the genotype main effect is considerably smaller than the GxE interaction and when the GxE interaction pattern is complex. In such cases, the GGE biplot consisting of PC1 and PC2 may be insufficient to explain the GGE, even though the most important patterns of the METs is already displayed (Yan, 2002). However, the author suggested three strategies to achieve a fully understanding of the data (Yan, 2006).

Unlike conventional approaches, which allow testing hypothesis, GGE biplot approach does not have a serious statistical test. Therefore, the GGE biplot is better used as a hypothesis-generator rather than as a decision-maker (Yan, 2002). However, Yan et al. (2006) has tried to give the theory fundamental for appropriately using GGE biplot in unpublished work. If being backed by formal statistical analysis with integrating the elegant GGE biplot, METs data should be fully explored.

**Acknowledgements:** Meimei Ding is grateful to STBA and FWPRDC for funding support for her study. We do thank Mike Powell for supplying the data and additional support.

## References

- Ades, P. K., and P. -H. Garnier-Géré. 1997. Making sense of provenance X environment interaction in *Pinus radiata*. Proceeding of IUFRO' 97 Genetics of radiate pine, Rotorua, New Zealand, 1-4 December 1997.
- Burdon, R.D. 1977. Genetic correlation as a concept for studying genotype-environment interaction in forest tree breeding
- Butrón, A., P. Velasco, A. Ordás, and R. A. Malvar. 2004. Yield evaluation of maize cultivars across environments with different levels of pink stem borer infestation. *Crop Science* 44:741-747.
- Crossa, J., Cornelius, and W. Yan. 2002. Biplots of linear -bilinear models for studying crossover genotype X environment interaction. *Crop Science* 42:136-144.
- Dehghani, H. A. Ebadi, and A. Yousefi. 2006. Biplot analysis of genotype by environment interaction for Barley yield in Iran. *Agron. J.* 98: 388-393.
- Delacy, I. H., K. E. Basford, and M. Cooper, J. K. Bull. 1996 Analysis of multi-environment trials- an historical perspective. *Plant Adaptation and Crop Improvement*. Eds. M. Cooper and G. L. Hammer. CAB INTERNATIONAL.
- Dias, C. T. dos S., and W. J. Krzanowskib. 2003. Model selection and cross validation in additive main effect and multiplicative interaction models. *Crop Science*. 43: 865-873.
- Eberhart, S.A., and W. A. Russel. 1966. Stability parameters for comparing varieties. *Crop Science*. 6:36-40.
- Finlay, K.W., and G. N. Wilkinson. 1963. The analysis of adaptation in a plant breeding programme. *Aust. J. Agri. Res.* 14:742-754.
- Gabriel, K. R. 1971. The biplot graphic of matrices with application to principal component analysis. *Biometrics* 58: 453-467.
- Gauch, H.G. 1992. *Statistical analysis of regional yield trials: AMMI analysis of factorial designs*. Elsevier, Amsterdam, Netherlands.
- Hammer, G. L., and M. Cooper. 1996. *Plant Adaptation and Crop Improvement*. CAB INTERNATIONAL, Wallingford.

- Johnson, G. R. and R. D. Burdon. 1989. Family-site interaction in *Pinus radiata*: implications for progeny testing strategy and regionalized breeding in New Zealand. *Silvae Genetica* 39: 55-62.
- Kaya, Y. M., Akcura, and S. Taner. 2006. GGE-biplot analysis of multi-environment yield trials in bread wheat. *Turk J. Agric. For.* 30: 325-337.
- Kroonenberg, P.M. 1995. Introduction to biplots for GXE tables, University of Queensland, Brisbane.
- Ma, B. L., W. Yan, and L. M. Dwyer, J. Frégeau-Reid, H. D. Voldeng, Y. Dion, and H. Nass. 2004. Graphic analysis of genotype, environment, Nitrogen fertilizer, and their interaction on Spring Wheat yield. *Agron. J.* 96: 169-180.
- Perkins, J.M., and J.L. Jinks. 1968. Environmental and genotype-environmental interactions and physical measures of the environment. *Heredity* 25:29-40.2006.
- Samonte S. O. P. B., L. T. Wilson, A. M. McClung, and J. C. Medley. 2005. Targeting cultivars onto rice growing environments using ammi and sreg gge biplot analysis. *Crop Science* 45: 2414-2424
- SAS Institute Inc., 1991. SAS Language and Procedures: usage 2, Version 6.
- Wu, H. X. and A. Colin Matheson. 2005. Genotype by environment interactions in an Australia-wide *radiata* pine diallel mating experiment: implications for regionalized breeding. *Forest Science* 51(1): 29-40.
- Yamada, Y. 1962. Genotype X environment interaction and genetic correlation of the same trait under different environments. *Jap. J. Genetics* 37:498-509.
- Yan, W. 2002. Singular-value partitioning in biplot analysis of multienvironment trial data. *Agron. J.* 94:990-996.
- Yan, W. and L. A. Hunt 2002. Biplot analysis of multi-environment trial data, *In* M. S. Kang, ed. *Quantitative Genetics, Genomics and Plant Breeding*. CAB International, Willingford.
- Yan, W., and B. L., Ma. 2006. Model diagnosis and GGE biplot analysis, pp. 39. Eastern Cereal and Oilseed Research Centre, Agriculture and Agri-food Canada.
- Yan, W., and Hunt, L. A. 2001. Genetic and environmental causes of genotype by environment interaction for winter wheat yield in Ontario. *Crop Science* 41:19-25.

- Yan, W., and I. R. Rajcan. 2002b. Biplot analysis of test sites and trait relations of soybean in Ontario. *Can. J. Plant Sci.* 42:11-20.
- Yan, W., and M. S. Kang. 2003. GGE biplot analysis: a graphical tool for breeders, *In* M. S. Kang, ed. *Geneticists, and Agronomist*. CRC Press, Boca Raton, FL.
- Yan, W., and N. A. Tinker. 2005. An integrated biplot analysis system for displaying, interpreting, and exploring genotype X environment interaction. *Crop Science* 45:1004-1016.
- Yan, W., L.A. Hunt, Q. Sheng, and Z. Szlavnic. 2000. Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Science* 40:597-605.
- Zobel, R. W., M. J. Wright, and H. G. Gauch. 1988. Statistical analysis of a yield trial. *Agron. J.* 80: 388-393.

Table 1. Summary information of five sites and average of the growth trait- DBH

Trial	PT5459	RAD211	VRC060	RS27A	RS27B
Region	Mount Gambier	Myrtleford	Traralgon	Busselton	Busselton
State	SA	VIC	VIC	WA	WA
Latitude	37° 33′	36° 41′	38° 14′	33° 52′	33° 52′
Longitude	140° 53′	146° 34′	146° 29′	115° 58′	115° 58′
Elevation (m)	70	370	68	120	120
Annul rainfall (mm)	680	1100	790	1100	1100
Soil type	Sandy	Sandy loam	Sandy loam	Clay loam	Clay loam
Site type	2nd PR	2nd PR	2nd PR	Pasture	Pasture
Mean DBH(mm)	174 ± 22	158 ± 34	201 ± 44	233 ± 41	242 ± 35

2nd PR is second-rotation of radiata pine crop; Pasture is previous pasture crop.

Table 2. Type B genetic correlations (lower left, comes from previous analysis) and correlation coefficients (cosine (angle)) between the sites (upper right)

Site	PT5459	RAD211	VRC060	RS27A	RS27B
PT5459		0.85	0.92	0.60	0
RAD211	0.43		0.98	0.09	-0.5
VRC060	0.71	0.77		0.21	-0.34
RS27A	0.53	0.30	0.34		0.81
RS27B	0.11	-0.02	-0.08	0.51	

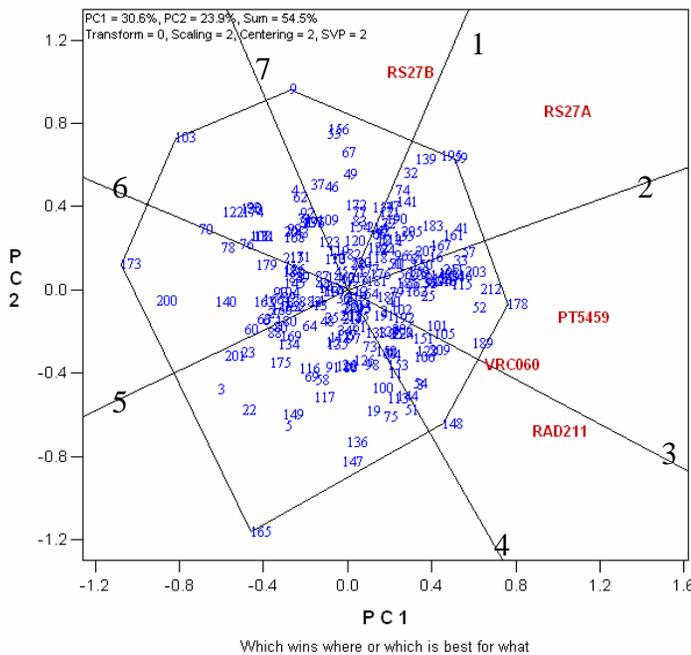


Figure 1. Polygon view of the GGE biplot show the “which-won-where”

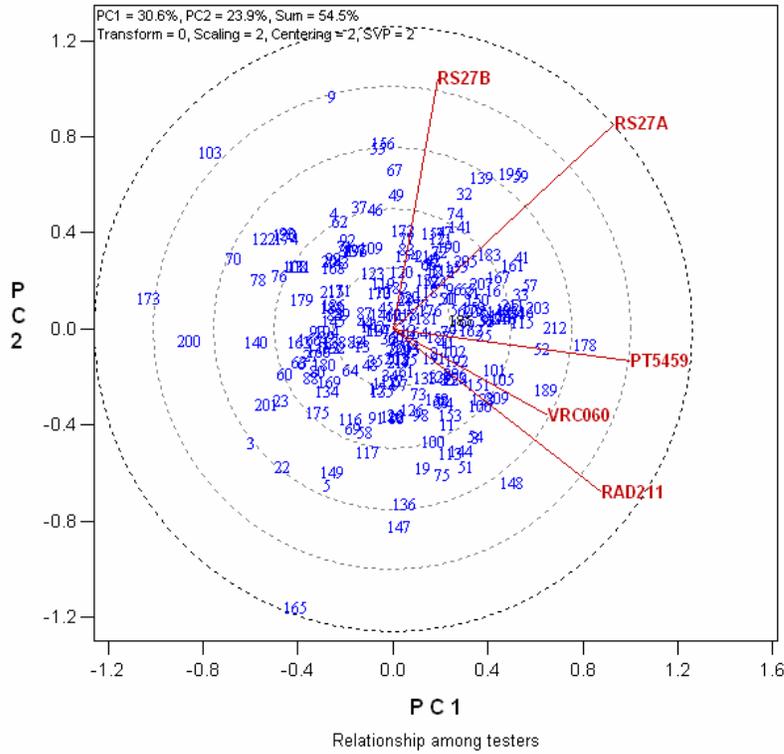


Figure 2. The relationship between five sites

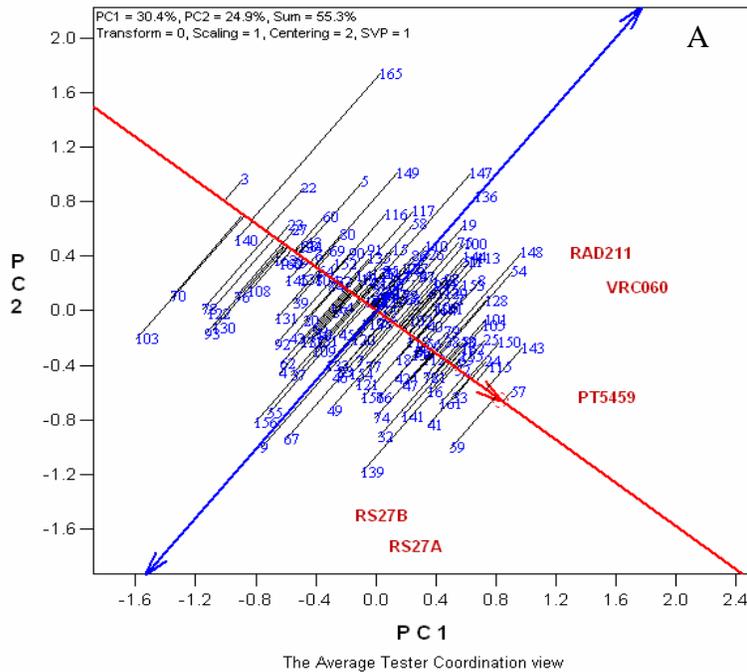


Figure 3. Ranking of 165 families based on both mean and stability

