APPLICATION OF GGE BIPLOT ANALYSIS TO EVALUATE GENOTYPE (G), ENVIRONMENT (E), AND G×E INTERACTION ON *PINUS RADIATA*: A CASE STUDY*

M. DING†, B. TIER,
Animal Genetics and Breeding Unit, University of New England,
Armidale, NSW 2351, Australia

W. YAN,
Eastern Cereal and Oilseed Research Centre, Agriculture and Agri-Food Canada,
960 Carling Ave, Ottawa, ON, Canada, K1A 0C6

H. X. WU,
CSIRO Forest Biosciences**,
P O Box E4008, Kingston ACT 2604, Australia

M. B. POWELL and T. A. McRAE
Southern Tree Breeding Association Inc.
P O Box 1811, Mount Gambier SA 5290, Australia

(Received for publication 20 December 2007; revision 5 February 2008)

ABSTRACT

Genetics, genetics×environment (GGE) biplot analysis is an effective method, based on principal component analysis, to fully explore multi-environment trial data. It allows visual examination of the relationships among the test environments, genotypes, and the genotype × environment (G×E) interactions. Data from multi-environment trials of *P. radiata* D. Don containing 165 to 216 families in five environments were used to demonstrate the results and application of GGE biplot analysis. There were non-overlapping clusters of two and three sites, which indicated two distinct environments. The best family for both of the distinct environments was also identified. Genetic correlations among sites ranged from 0.98 to –0.50, indicating that there were large G×E interactions among the test environments.

Keywords: GGE biplot; principal component analysis; G×E interaction; *Pinus radiata*.

---

* Paper originally presented at the inaugural Australasian Forest Genetics Conference "Breeding for Wood Quality", Hobart, 11–14 April 2007.

† Corresponding author: mding@une.edu.au

**Research carried out under the Ensis UJV

INTRODUCTION

The phenotypic expression of a genotype is a mixture of genotype (G) and environment (E) components, and interactions (G×E) between them. G×E interactions complicate the process of selection of genotypes with superior performance. Multi-environment trials are widely used by plant breeders to evaluate the relative performance of genotypes for environments (Delacy *et al.* 1996). Numerous methods have been developed to reveal patterns of G×E interaction, such as joint regression (Finlay & Wilkinson 1963; Eberhart & Russell 1966; Perkins & Jinks 1968), additive main effects and multiplicative interaction (Gauch 1992), and Type B genetic correlation (Burdon 1977; Yamada 1962). These methods are commonly used to analyse multi-environment trial data and have also been used to study G×E interaction in *P. radiata* (Ades & Garnier 1997; Johnson & Burdon 1990; Wu & Matheson 2005).

GGE biplot analysis was recently developed to apply some of the functions of these methods jointly. As a proportion of total phenotypic variation, environment explains most of the variation and genetics and genetics×environment are usually small (Yan & Rajcan 2002). However, only the genetics and G×E interaction are relevant to cultivar evaluation, particularly when G×E interaction is identified as repeatable (Cooper & Hammer 1996). Hence, Yan *et al.* (2000) deliberately put these 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 and referred to as a GGE biplot (Yan 2002; Yan *et al.* 2000).

A GGE biplot is a data visualisation tool, which graphically displays a G×E interaction in a two-way table (Yan *et al.* 2000). A 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;
(2) Genotype evaluation (the mean performance and stability), and
(3) Test-environmental evaluation.

GGE biplot analysis is increasingly being used in G×E interaction data analysis in agriculture (Butrón *et al.* 2004; Crossa *et al.* 2002; Dehghani *et al.* 2006; Kaya *et al.* 2006; Ma *et al.* 2004; Yan & Hunt 2001). However, there has been no report of its application to forestry so far. As a case study, this paper applies the technique to reveal the patterns of G×E interaction on *P. radiata* and then compares biplots with other methods.

Materials and Method

The genetic material originated from Australia-wide diallel mating experiments; the details were described by Wu & Matheson (2005). The five sites chosen for this study were 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; the longitudinal span is 30°36´ from 115°58´E to 146°34´E. There were 165 to 216 full-sib families per trial. The seedlings were planted in two site types, which were second-rotation *P. radiata* (2nd PR) and previous pasture crop site (pasture). Each trial was a randomised incomplete block design with three replicates. The diameter at breast height was assessed at 10.5 years of age (for details see Table 1). Least square means (LSmeans) of families in each site were calculated using SAS Mixed Procedure (SAS 1999).

TABLE 1–Summary information from five sites and average of the growth traits, with standard deviation for diameter at breast height (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
Annual 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 ±3 5

* 2nd PR = second-rotation of *Pinus radiata* crop; Pasture = previous crop was pasture.

Models for a GGE Biplot

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

$$Y_{ij} - ì - â_j = ë_1 î_{i1} ç_{j1} + ë_2 î_{i2} ç_{j2} + ε_{ij} \quad [1]$$

where Y_{ij} is the measured mean (dbh) of genotype i in environment j,

ì is the grand mean,

$â_j$ is the main effect of environment j,

$ì + â_j$ is the mean yield across all genotypes in environment j,

$ë_1$ and $ë_2$ are the singular values for the first and second principal components, respectively,

$î_{i1}$ and $î_{i2}$ are eigenvectors of genotype i for the first and second principal components, respectively,

$ç_{1j}$ and $ç_{2j}$ are eigenvectors of environment j for the first and second principal components, respectively,

$å_{ij}$ is the residual associated with genotype i in environment j.

First and second principal component 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:

$$g_{il} = \ddot{e}_l^{fl}\, \hat{\imath}_{il} \text{ and } e_{1j} = \ddot{e}^{1}{}_{1}^{fl}\, \varsigma_{1j} \quad [2]$$

where f_1 is the partition factor for the first principal component. Theoretically, f_1 can be a value between 0 and 1, but 0.5 is most commonly used (Yan 2002). To generate the GGE biplot, the formula [1] was presented as:

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

The data can be standardised to remove any heterogeneity of variances among the environments. The formula for GGE biplot is reorganised as follows:

$$(Y_{ij} - u\ \beta_j) / s_j = \sum_{i=1}^{k} g_{i1}\, e_{1j} + \varepsilon_{ij} \quad [4]$$

where s_j is the standard deviation in environment j, l=1, 2,…,k, g_{il} and are first principal component scores for genotype i and environment j, respectively. This model was used to generate a biplot of "which-won-where". For the analysis of the relationship between the trials, genotype, and environment evaluation, the model [4] was used. 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 analyses with joint regression and Type B genetic correlation.

RESULTS

The first two principal components explain 54.5% (first = 30.6%, second = 23.9%) of the total GGE variation using the standardised model [4]; similarly, using the unstandardised model [3] explains 55.3% (first = 30.4%, second = 24.9%) of total GGE variation.

The results are presented in 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 visualises 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 (Fig. 1) indicates the best genotype(s) in each environment and group of environments (Yan & 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

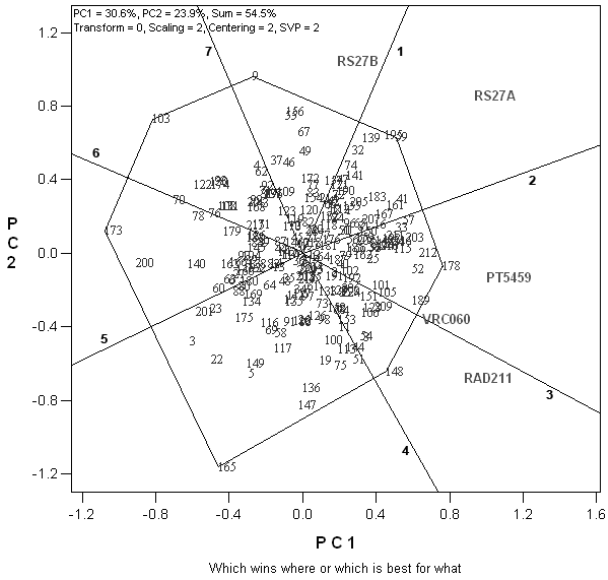

FIG. 1–Polygon view of the GGE biplot showing the "which-won-where" using standardised data including 216 families

(Yan 2002). 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, and so on. These seven rays divide the biplot into seven sections, and the five sites fall into four of these seven sections. The vertex families for each quadrant are the ones 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 family 148, in RS27A is family 59, in RS27B is family 9. The other vertex families 165, 173, and 103 are poorest in all five sites.

2. Interrelationship among environments

The summary of the interrelationships among the environments is given in Fig. 2. The lines that connect the biplot origin and the markers for the environments are environment vectors, and 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 separated into two groups. Group one includes PT5459, RAD211, and VRC060. Group two involves RS27A and RS27B. This grouping coincides with both geographic distance (West Australian vs South Australian and Victorian sites) and site type (pasture vs second rotation *P. radiata*).

A comparison of the correlation coefficients (cosine angles) and Type B genetic correlations from previous analysis is shown in Table 2. The smallest angle is

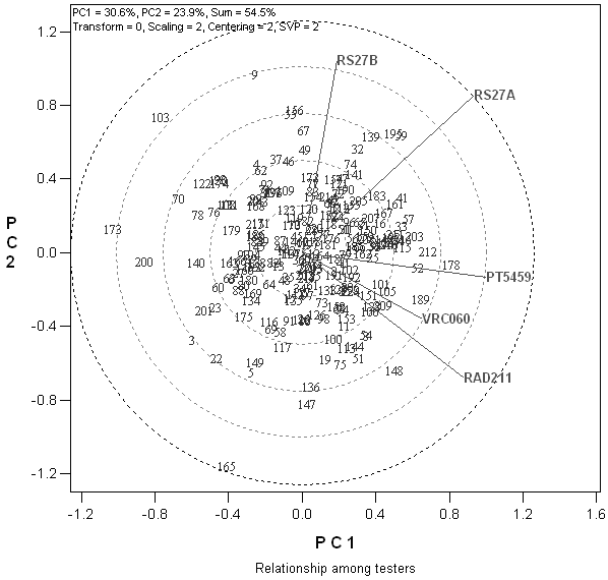

FIG. 2–The relationship between five sites using standardised data including 216 families

TABLE 2–Type B genetic correlations (lower triangle, from previous analysis) and correlation coefficients (cosine angle) between the sites (upper triangle)

Site	PT5459	RAD211	VRC060	RS27A	RS27B
PT5459		0.85	0.92	0.60	0
RAD211	0.35		0.98	0.09	–0.5
VRC060	0.72	0.90		0.21	–0.34
RS27A	0.36	0.23	0.38		0.81
RS27B	0.01	0	0	0.51	

between RAD211 and VRC060, implying that the highest correlation is between them. The approximate correlation coefficient is 0.98. The next smallest angle occurs between VRC060 and PT5459. The angles between RAD211 and RS27B, and VRC060 and RS27B are greater than 90°, showing the negative correlations between them.

3. Family mean yields and their stability

The ranking of 165 families by mean yield and stability is indicated in Fig. 3. The line passing through the biplot origin from upper left to lower right is called the average environment axis, which is defined by the average first and second principal component scores of all environments. Closeness to the circle indicates higher mean yield. The line which passes through the origin and is perpendicular to the average environment axis with a double arrow represents the stability of

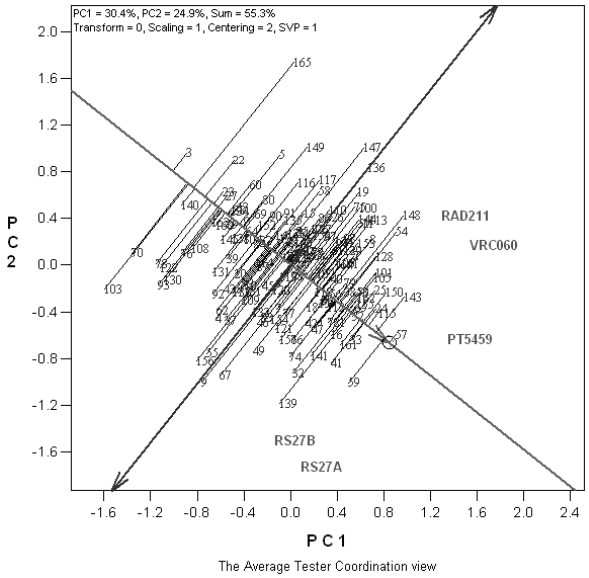

FIG. 3–Mean performance and stability of 165 families using unstandardised data

genotypes. Either direction away from the biplot origin, on this axis, indicates greater G×E interaction and reduced stability (Yan & Hunt 2002). For broad selection, the ideal genotypes are those 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 average environment axis. Families 57, 15, and 24 belong to this group for environment group one. On the other hand, for specific selection, the ideal genotypes are those that have high mean yield but low stability and respond best to particular environments. For environment in group (PT5459, RAD211, and VRC060), the mean yield of families is as followings: the highest are families 57 and 143 and so on, with the worst family shown as 19. For group two (RS27A and RS27B), the mean yield of families is as followings: the highest is family 59, then family 41, so on; the worst is family 9.

4. Examining the genotypes and environments

The performance of different genotypes in PT5459 is shown in Fig 4a. The line labelled PT5459 that passes through the biplot origin is the PT5459 axis. The genotypes are ranked according to their projections on to the PT5459 axis. The second 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. Genotypes ranking above the mean are the families on the right side of the second line, e.g., families 57, 143, and 148. Families 149, 165, and 3 on the left side of the second line are below the mean genotype ranking.

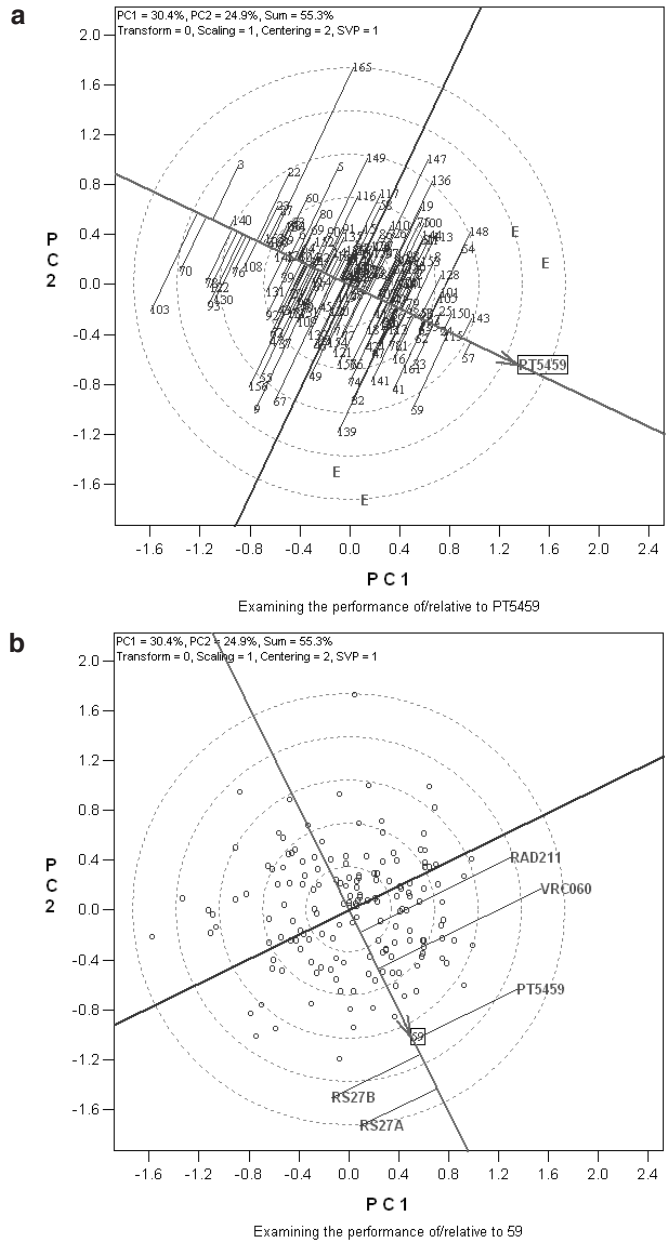

FIG. 4 a–Different genotypes in a given environment (PT5459)
b–A given genotype (family 59) in five sites

Results for family 59 on five sites are shown in Fig. 4b. The line labelled family passing through the biplot origin is the family 59 axis. The environments are ranked along the family 59 axis towards the label family 59. Thus, the relative performance of family 59 in different environments is as follows: RS27A>RS27B>

PT5459>VRC060>RAD211. The line 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 on all five sites.

DISCUSSION AND CONCLUSION

The GGE biplot analysis integrates some features from the methods of additive main effects and multiplicative interaction (Gauch 1992), joint regression, and Type B genetic correlation and allows visual interpretation of G×E interaction.

The first feature is the ability to visualise the interaction between genotypes and environments ("which-won-where"). GGE biplot stands for genotype main effect plus GxE interaction (Ma *et al*. 2004). these are both based on the statistical model of principal component analysis. GGE biplot analysis is based on environment-centred principal component analysis, whereas additive main effects and multiplicative interaction analysis is referred to as double centered principal component analysis (Kroonenberg 1995). However, if the purpose is for "which-won-where", additive main effects and multiplicative interaction could be misleading (Yan & Ma 2006). The GGE biplot has many visual interpretations that additive main effects and multiplicative interaction does not have; particularly it allows visualisation of any crossover GxE interaction. This part of GxE interaction is usually essential to a breeding programme. In addition, in comparison with different additive main effects and multiplicative interaction family models (Dias & Krznowskib 2003; Zobel *et al*. 1988), GGE biplot is close to the best additive main effects and multiplicative interaction model in most cases (Yan & Ma 2006). Moreover, GGE biplot is more logical for biological objectives in terms of explaining the first principal component score, which represents genotypic level rather than additive level (Yan *et al*. 2000).

The second feature is to show the interrelationship between the environments which is similar to the Type B genetic correlation between trials. The graphic of interrelationships between environments displays the correlation between the trials. 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, and negative or zero occurs between RS27B and the other trials in group one. Inconsistencies can occur because biplot does not explain 100% of the GGE variation.

The third feature is to visualise the interrelationship among genotypes based on both mean performance and stability. The visualising graphic of genotype means and their stability shows different genotype groups classified into four groups. Group one is highly desirable with high yield and high stability. The group with high yield but low stability is desirable for specific selection, whereas low yield and low stability may be for special breeding purposes, e.g., selection for

drought resistance. The most undesirable group is low yield but high stability. The classification is similar to the previous work on family behaviour plots using joint regression (Finlay & Wilkinson 1963). Moreover, GGE biplot not only shows different genotype groups, but also shows their favourite environments. If one assumes that the classical method of joint regression is informative, the results of GGE biplot analysis are similarly informative.

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 G×E interaction, and when the G×E interaction pattern is complex. In such cases, the GGE biplot consisting of first and second principal components may be insufficient to explain the GGE, even though the most important patterns of the multi-environment trials are already displayed (Yan & Rajcan 2002). However, Yan & Ma (2006) suggested three strategies to achieve a better understanding of the data.

Unlike conventional approaches, which allow hypothesis testing, the 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 & Rajcan 2002).

REFERENCES

ADES, P.K.; GARNIER-GÉRÉ, P.-H. 1997: Making sense of provenance × environment interaction in *Pinus radiata*. Pp. 113–119 *in* Burdon, R.D.; Moore, J.M. (Ed.) "IUFRO '97 Genetics of Radiata Pine". Proceedings of NZ FRI-IUFRO Conference 1–4 December and Workshop 5 December, Rotorua, New Zealand. *FRI Bulletin No.203*.

BURDON, R.D. 1977: Genetic correlation as a concept for studying genotype-environment interaction in forest tree breeding. *Silvae Genetica 26(5–6)*: 168–175.

BUTRÓN, A.; VELASCO, P.; ORDÁS, A.; MALVAR, R.A. 2004: Yield evaluation of maize cultivars across environments with different levels of pink stem borer infestation. *Crop Science 44*: 741–747.

COOPER, M.; HAMMER, G.L. 1996: "Plant Adaptation and Crop Improvement". CAB International, Wallingford.

CROSSA, J.; CORNELIUS, P.L.; YAN, W. 2002: Biplots of linear-bilinear models for studying crossover genotype × environment interaction. *Crop Science 42*: 136–144.

DEHGHANI, H.A.; EBADI; YOUSEFI, A. 2006: Biplot analysis of genotype by environment interaction for barley yield in Iran. *Agronomy Journal 98*: 388–393.

DELACY, I.H.; BASFORD, K.E.; COOPER, M.; BULL, J.K. 1996: Analysis of multi-environment trials — an historical perspective. *In* Cooper, M.; Hammer, G.L. (Ed.) "Plant Adaptation and Crop Improvement". CAB Iinternational.

DIAS, C.T. DOS S.; KRZANOWSKIB, W.J. 2003: Model selection and cross validation in additive main effect and multiplicative interaction models. *Crop Science 43*: 865–873.

EBERHART, S.A.; RUSSEL, W.A. 1966: Stability parameters for comparing varieties. *Crop Science 6*: 36–40.

FINLAY, K.W.; WILKINSON, G.N. 1963: The analysis of adaptation in a plant breeding programme. *Australian Journal of Agricultural Research 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.

JOHNSON, G.R.; BURDON, R.D. 1990: 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; TANER, S. 2006: GGE-biplot analysis of multi-environment yield trials in bread wheat. *Turkish Journal of Agriculture & Forestry 30*: 325–337.

KROONENBERG, P.M. 1995: "Introduction to Biplots for G×E Tables". University of Queensland, Brisbane.

MA, B.L.; YAN, W.; DWYER, L.M.; FRÉGEAU-REID, J.; VOLDENG, H.D.; DION, Y.; NASS, H. 2004: Graphic analysis of genotype, environment, nitrogen fertilizer, and their interaction on spring wheat yield. *Agronomy Journal 96*: 169–180.

PERKINS, J.M.; JINKS, J.L. 1968: Environmental and genotype-environmental interactions and physical measures of the environment. *Heredity 25*: 29–40.

SAS INSTITUTE INC. 1999: "SAS Language and Procedures". Usage 2, Version 6.

WU, H.X.; MATHESON, A.C. 2005: Genotype by environment interactions in an Australia-wide radiata pine diallel mating experiment: implications for regionalized breeding. *Forest Science 51*: 29–40.

YAMADA, Y. 1962: Genotype × environment interaction and genetic correlation of the same trait under different environments. *Japanese Journal of Genetics 37*: 498–509.

YAN, W. 2002: Singular-value partitioning in biplot analysis of multienvironment trial data. *Agronomy Journal 94*: 990–996.

YAN, W.; 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.; MA, B.L. 2006: "Model Diagnosis and GGE Biplot Analysis". Eastern Cereal and Oilseed Research Centre, Agriculture and Agri-food Canada. 39 p.

YAN, W.; RAJCAN, I.R. 2002: Biplot analysis of test sites and trait relations of soybean in Ontario. *Canadian Journal of Plant Science 42*: 11–20.

YAN, W.; TINKER, N.A. 2005: An integrated biplot analysis system for displaying, interpreting, and exploring genotype × environment interaction. *Crop Science 45*: 1004–1016.

YAN, W.; HUNT, L.A.; SHENG, Q.; SZLAVNICSA, Z. 2000: Cultivar evaluation and mega-environment investigation based on the GGE biplot. Crop Science 40: 597–605.

ZOBEL, R.W.; WRIGHT, M.J.; GAUCH, H.G. 1988: Statistical analysis of a yield trial. *Agronomy Journal 80*: 388–393.