Theor Appl Genet (1993) 86:731–736

THEORETICAL AND APPLIED GENETICS

Genotype × environment interactions in a core collection of French perennial ryegrass populations

G. Charmet[1], F. Balfourier[1], C. Ravel[1], J.-B. Denis[2]

[1] INRA, Station d' amélioration des plantes, 63039 Clermont Ferrand, France
[2] INRA, Station de biométrie, 78000 Versailles, France

Received: 11 November 1992 / Accepted: 4 January 1993

Abstract. A representative sample (core collection) of natural populations of perennial ryegrass (*Lolium perenne* L.) from France was evaluated for agronomic traits at seven locations. This sample exhibits a high level of genotype-environment interaction for most traits. The interactions for summer-growth (a key-factor of adaptation in most French regions) were studied by means of regression using climatic factors of the evaluation sites and the sites of population origin as covariates. This method succeeded in explaining most of the interaction term and also part of the main effects. It appears that populations from either warm or dry sites generally have a positive interaction when evaluated in a site with similar characteristics, as expected as a consequence of natural selection. A population component of regression on environmental covariates, however, was significant and could be exploited through breeding to improve adaptation of perennial ryegrass to either drier or warmer regions.

Key words: Genotype-environment interaction – Factor regression – Genetic resources – Plant breeding

Introduction

The concept of a core collection of genetic resources (Frankel 1984; Brown 1989a, b) was introduced to improve management facilities and the availability of genetic resources for plant breeders. A representative sample from a larger collection allows the breeder to evaluate the sample more extensively and to identify the most promising accessions for further study or for re-sampling. In most crop species, agronomic traits generally show significant genotype × environment interactions, which must be taken into account when using genetic resources in breeding programmes. Again the manageable size of a core collection permits the use of multisite evaluation trials to estimate these interactions.

Several methods have been proposed to analyse genotype × environment interactions (Freeman 1973; Denis and Vincourt 1982), but joint regression on environmental means is probably the most popular (Yates and Cochran 1938; Finlay and Wilkinson 1963; Hardwick and Wood 1972; Perkins 1972). Wood (1976) suggested the use of environmental variables such as climate and soil type, instead of location means, as independent measures of the environment. Denis (1980, 1988) generalized this method through the use of covariates associated with both environments and genotypes, calling it factor regression analysis.

These methods may be particularly useful for studying the adaptation of natural populations to different environments, which can be characterized by environmental variables related to their sites of origin. One can also anticipate some ecological interpretation.

Perennial ryegrass is one of the most commonly used grasses in temperate countries and is appreciated for its high nutritive value to herbivores. In many regions, however, its yield is limited by its poor adaptation to summer heat and drought. An assessment of genetic resources from various origins for their ability to grow in summer is therefore a key factory in breeding programmes aimed at improving the adaptation of perennial ryegrass to more continental regions. Since ryegrass is native in Europe, a large sample of wild populations has been collected in France and

Communicated by A. R. Hallauer
Correspondence to: G. Charmet

evaluated at seven locations (Charmet et al. 1990). Our objective is to apply factor regression to analyse genotype × environment interactions (i.e., population × evaluation site) in two subsets of this core collection using climatic covariates of both collection sites and evaluation sites.

Materials and methods

An extensive sample of natural populations of perennial ryegrass was collected in France and evaluated for agronomic traits in a multisite network. High levels of genotype × environment interaction were reported and used to classify this collection of French populations (Charmet et al. 1990). A representative subset was formed by sampling populations in each cluster and by covering the full range of ecogeographic factors of their sites of origin, such as climate and habitat (Balfourier and Charmet 1991).

The populations of the core are mapped in Fig. 1. The core was divided into two subsets corresponding to their evaluation dates: subset 1 includes 54 populations evaluated in 1985 and subset 2 includes 58 populations evaluated in 1986. The evaluation procedures were fully described in Charmet et al. (1990). At each of the seven evaluation sites (Fig. 2), a complete design with three replications and with ten plants from every population in each block was used. Data from spaced plant nurseries were averaged by blocks within location, the population × block term being used as error in ANOVA.

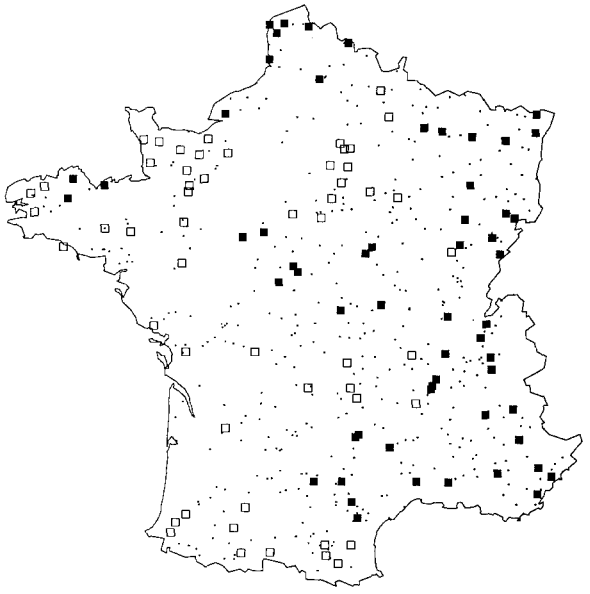

Fig. 1. Map of the two subsets of the core collection of *Lolium perenne* populations: *small points* indicate the origin of the 550 populations initially collected, while *squares* show the location of the 112 populations from the core

Fig. 2. Map of the seven evaluation sites for the two subsets

Population × location interaction was firstly tested using model 1:

$$X_{ijk} = \mu + l_i + l.b_{ik} + g_j + \theta_{ij} + \varepsilon_{ijk},$$

where l_i is the main location effect, $l.b_{ik}$ is the hierachised block effect, g_j is the main genotype effect, and θ_{ij} is the genotype × location (i.e., population × evaluation site) interaction. All terms in ANOVA were considered as fixed effects.

Summer growth is obviously influenced by summer conditions at the evaluation site, mostly by temperature and water supply. Summer conditions at the sites of population origin may also act as a selection pressure, presumably favouring adaptation to similar conditions. Therefore, we used the following two climatic factors as covariates, either for the populations (50 years average at the closest climatic station) or for the evaluation site (data from the actual evaluation year: 1985 for subset 1 and 1986 for subset 2):

Tmax, average maximum daily temperature of the warmest month (July).
HB, hydric balance (rainfall minus potential evapotranspiration) over the June–August period; HB is thus negative for dry sites.

Computations were carried out using the INTERA package on a PC (Decoux and Denis 1990).

The full model using all covariates is model 2:

$$\begin{aligned} E(X_{ij}) = {} & \mu + \rho_1 HB_i + \rho_2 Tmax_i + \alpha'_i \\ & \text{(Corresponds to } l_i \text{ of model 1)} \\ & + \rho_3 HB_j + \rho_4 Tmax_j + \beta'_j \\ & \text{(Corresponds to } g_j \text{ of model 1)} \\ & + HB_i \nu_1 HB_j + HB_i \nu_2 Tmax_j \\ & + Tmax_i \nu_3 HB_j + Tmax_i \nu_4 Tmax_j \\ & + \tau_{1j} HB_i + \tau_{2j} Tmax_i + \tau_{3i} HB_j + \tau_{4i} Tmax_j, \end{aligned}$$

where ρ_k is the regression coefficient of main effects onto the covariates, ν_k is the coefficient of the 2-by-2 products of covariates, and τ_k is the regression coefficient of the rest of the interaction term not accounted for by the products of covariates.

Table 1. Terms of the factor regression model with two covariates for genotype and environment

Location covariates	Population covariates			
	0	HB	Tmax	Rest
0	μ	$\rho_1 HB_j$	$\rho_2 Tmax_j$	β'_j
HB	$\rho_3 HB_i$	$HB_i v_1 HB_j$	$HB_i v_2 Tmax_j$	$\tau_{1j} HB_i$
Tmax	$\rho_4 Tmax_i$	$Tmax_i v_3 HB_j$	$Tmax_i v_4 Tmax_j$	$\tau_{2j} Tmax_i$
Rest	α'_i	$\tau_{3i} Tmax_j$	$\tau_{4i} HB_j$	R_{ij}

Model 2 includes two parts: part 1 corresponds to the regression of the main effects l_i and g_j on the covariates, and part 2 to the regression of the interaction term q_{ij}. Part 2 includes several multiplicative terms: i.e., 2-by-2 products of covariates. They are not similar to the multiplicative terms proposed by Mandel (1971), which do not use external information from covariates. This is better understood using the presentation of Table 1. Each square of the table represents one term of model 2, with its corresponding degrees of freedom. The results of ANOVA will be presented using such recapitulative tables, which are clearer than the classical presentation (Denis 1991).

Although the experimental design was initially balanced, the introduction of several covariates makes the model non-orthogonal. This implies that the sum of squares for each term is dependent on its rank of introduction. Several rounds of ANOVA using different ordering of covariates were conducted to determine the best subset: i.e., that minimizing the interaction sum of squares not accounted for by factor regression. Moreover, a single covariate may be sufficient to explain most of the interaction and could lead to a simpler model. Therefore, we used the procedure described by Baril (1992) to avoid introducing new covariates. One multiplicative term is computed from the residual interaction matrix (i.e., after factor regression). A covariate is declared significant at a given probability level if its mean square exceeds the mean square of the sum of the multiplicative and the residual term by the appropriate F-value. Given the analogy between the multiplicative modelisation of interaction and principal component analysis, this means that the set of interaction terms remaining after factor regression has no more peculiar direction and can therefore be considered as isotropic noise.

Results

The results of ANOVA using model 1 on the two subsets are presented in classical form in Table 2. Population × location interaction is statistically significant in both analyses. F values from ANOVA of factor regression on subset 1 and subset 2 are presented in Table 3. The two covariates are used for both subsets to make the multiplicative term non-significant. In both subsets the interaction not accounted for by factor regression is not significant. Thus the method succeeded in explaining most population × location interactions by the two-by-two covariate products and their remaining linear regression. Regression on either population or location covariates also explains a part of the corresponding main effects, but the rest of the main effect remains significant. Although the regression of location main effect on both covariates is significant, the regression term on hydric balance has the highest mean squares, indicating that average summer growth is mostly dependent on water availability at each site. The regressions of population main effect on the two covariates have mean squares of the same magnitude. Particular attention should be given to the significance of covariate products, since they could be interpreted in the light of ecological genetics.

Table 2. ANOVA tables of model 1 for the two subsets of populations

Item	SS	*df*	MS	F
Subset 1				
Location	228.10	6	38.02	152.1**
Block in LOC	38.88	14	2.77	11.1**
Population	43.14	53	0.81	3.3*
Interaction	128.79	318	0.40	1.6*
Residual	185.75	743	0.25	
Subset 2				
Location	178.60	6	29.76	119.1**
Block in LOC	35.88	14	2.56	10.2**
Population	116.31	57	2.04	8.2**
Interaction	173.44	342	0.50	2.0**
Residual	203.00	799	0.25	

Table 3. F values on subset 1 and subset 2 and their significance (*, 0.05; **, 0.01% level of probability) for model 2

Location covariate	Population covariate			
	0	HB	Tmax	Rest
Subset 1				
0		7.2**	10.8**	1.7*
HB	337**	6.2**	0.1	1.5*
Tmax	9.7*	1.0	11.4**	1.4*
Rest	55.7**	3.0*	1.7*	0.6NS
Subset 2				
0		53.3**	30.8**	4.3*
HB	138**	0.7	19.8**	1.8*
Tmax	22.8*	3.4	0.2	1.6*
Rest	83.8**	1.5	5.3**	0.8NS

The parameters v_k, and ρ_k of model 2 are given in Table 4. The signs of the parameters are of particular interest. In the first line and column, they are the regression coefficients of main effects, whereas in the rest of the table they are cofactors of covariate products. It is remarkable that the same coefficients for the two subsets are very similar. This allows us to draw similar conclusions from the two subsets.

The regression coefficient for hydric balance at the evaluation site of the main effects is positive, which means that environmental means are higher in the wetter locations. The coefficient of regression of main effect on population HB is negative in subset 1 and positive in subset 2, indicating that average population yield is not univocally associated with hydric balance at the relevant site of origin. The regression coefficient for the temperature of both main effects is negative, leading to the conclusion that evaluation sites with high maximum temperature yield less in summer than cooler sites, as do populations originating from warm sites.

In the analysis of subset 1 (Table 3) two covariate products are statistically significant: HB × HB and Tmax × Tmax, both with positive coefficients. Thus, populations originating from either dryer sites (HB high and negative) or warmer sites show positive interactions (i.e., partially compensate their negative main effect) when evaluated in drier (negative HB) and/or warmer locations.

Only one product is significant in the ANOVA of subset 2 (Table 3): HB (location) × Tmax (population). This can be explained by the particular conditions encountered during summer 1986, which was very hot at all locations but with very different amounts of rainfall due to localized storms. Hydric balance, therefore, is the most significant location covariate, and it interacts mostly with the Tmax of the site of population origin, with a negative coefficient. The Tmax term could be high enough to give an advantage to populations originating from warm sites when evaluated in a dry location (HB high and negative): this is expected as a consequence of natural selection. After regression on covariate products, the regression of the rest of the interaction onto the location covariates remains significant. This represents a part of genotypic variability not explained by the population covariates, and thus may illustrate some interesting source of adaptation to limiting conditions.

A useful way to present the population coefficient of regression is to map them for HB (Fig. 3) and for Tmax (Fig. 4). Figures 3 and 4 enable us to locate the populations which are expected to have the highest positive (or negative) interaction in the drier or warmer evaluation sites (filled or empty symbols, respectively). For instance, most populations from south-east France have positive regression coefficients on HB, leading to a negative interaction term in drier locations. This could be attributed to a genetically inherited summer dormancy favoured by natural selection under a sub-mediterranean climate. The regression coefficient on Tmax is rather different (the correlation between these two coefficients is only 0.66). The populations from Normandy and Massif central now have low negative values, indicating that their yield in summer is probably limited by higher temperature rather than by water deficit, while some populations in Britanny or in the south-west appear to be more tolerant to higher temperatures.

Table 4. ρ and τ parameters of model 2 for the two subsets

Location component	Population component		
	Main	HB	Tmax
Subset 1			
Main	19.03	−0.038	−0.626
HB	+0.178	0.0024	0.0146
Tmax	−0.701	0.0035	0.035
Subset 2			
Main	16.80	+0.025	−0.422
HB	0.487	0.0001	−0.017
Tmax	−0.111	0.0015	0.0026

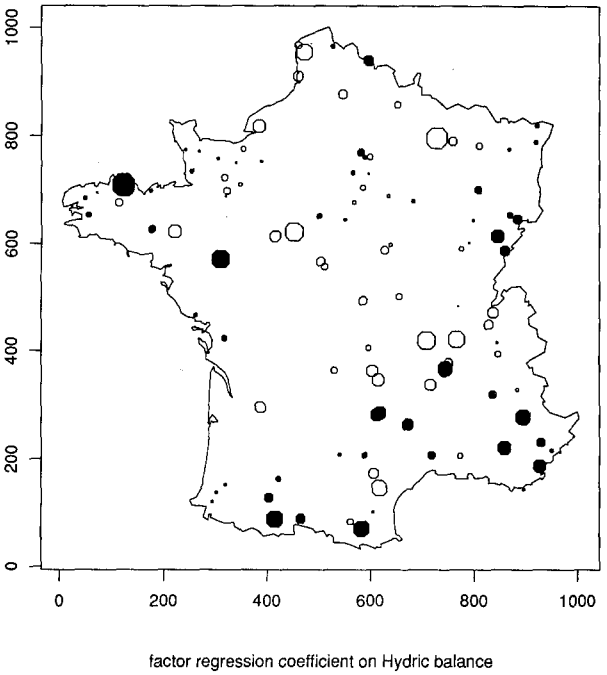

Fig. 3. Map of the τ_j regression coefficient on HB: symbol size is proportional to the absolute value of tj. *Empty symbols* represent negative coefficients, *filled symbols* the positive ones

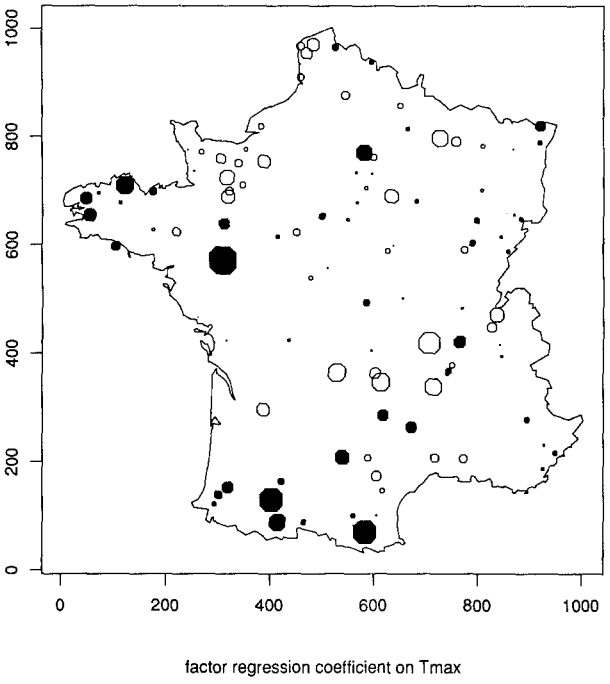

Fig. 4. Map of the τ_j regression parameters of populations on Tmax

Discussion and conclusion

Factor regression analysis succeeded in reducing the unexplained part of genotype (population) × environment (evaluation site) interaction to a non-significant level. The use of only two covariates allowed us to explain 54% and 46% of the interaction sum of squares for subsets 1 and 2 respectively: i.e., more than the use of a single multiplicative term (34% and 37% respectively) of Mandel's (1971) analysis as suggested by Crossa et al. (1990).

This method, when applied to natural populations, permitted some ecological or evolutionary inferences to be drawn. For instance, it allowed us to confirm that populations originating from drier or warmer sites show positive interactions when evaluated in locations with a similar climate, as would be expected from the influence of natural selection.

Mapping factor regression coefficients permitted the location of the origins where populations show a good "residual" (i.e., after taking into account their own climatic covariate) tolerance to either high temperature or water deficit. The usefulness of such observations for using genetic resources in breeding programmes is obvious.

Similar results have already been reported in forage grasses. Breese (1969) used joint regression analysis on cocksfoot (*Dactylis glomerata*) populations. He found distinct differences in the slope of regression against environmental means between origins. The populations from Portugal (warm) had less slope than populations from the French Massif central (cool). This method was further developed by Wright (1971) who used joint regression on both environment and genotype means, and also calculated heritabilities of the regression components. Hayward et al. (1982) used clustering procedures for the recognition of patterns of adaptation in a collection of Italian ryegrass (*Lolium multiflorum*) populations. They succeeded in relating the observed pattern of adaptation to the ecological situations from where the populations derived.

Obviously the climatic covariates used in this study may seem to be somewhat arbitrarily chosen, and also to be correlated with each other. Perkins (1972, 1974) suggested the use of principal component analysis on environmental variables, then the use of the first principal components as covariates for the regression. This method may not be very useful for only two covariates. It did not improve the percentage of interaction variance accounted for by the regression and made the biological interpretation more confusing. Wood (1976) also stated that the direct use of environmental variables would be preferable.

In the present study, the use of the same environmental covariates for both the sites of population origin and the evaluation sites makes the interpretation of regression coefficient easier in terms of natural selection history. Mapping of regression coefficients allows plant breeders to identify the most tolerant populations to specific limiting factors of the environment in relation to their breeding objectives. A similar presentation on maps was proposed by Kempton (1984) in the case of a biplot analysis of aphid distribution in the UK. Obviously, further studies would be needed to measure the inheritance of these regression parameters (Hill and Samuel 1971). It would be particularly useful to improve the tolerance to the limiting factors, and more especially to recombine the populations identified in this core collection as possessing tolerance to either high temperature or to water deficit.

References

Balfourier F, Charmet G (1991) Relationships between agronomic characters and ecogeographical factors in a collection of French perennial ryegrass populations. Agronomie 11:645–657

Baril CP (1992) Factor regression for interpreting genotype-environment interaction in bread-wheat trials. Theor Appl Genet 83:1022–1026

Breese EL (1969) The measurement and significance of genotype-environment interactions in grasses. Heredity 24:27–44

Brown AHD (1989a) The case for core collections. In: Brown AHD, Frankel OH, Marshall DR, Williams JT (eds) The use of plant genetic resources 3. Cambridge University Press, Cambridge, pp 136–156

Brown AHD (1989b) Core collections: a practical approach to genetic resources management. Genome 31:818–824
Charmet G, Balfourier F, Bion A (1990) Agronomic evaluation of a collection of French perennial ryegrass populations: multivariate classification using genotype × environment interactions. Agronomie 10:807–823
Crossa J, Gauch HG, Zobel RW (1990) Additive main effects and multiplicative interaction analysis of two international maize cultivar trials. Crop Sci 30:493–500
Decoux G, Denis JB (1990) INTERA, manuel d'utilisation, laboratoire de biométrie. INRA, Versailles, France
Denis JB (1980) Analyse de regression factorielle. Biom Praxim 20:1–34
Denis JB (1988) Two-way analysis using covariates. Statistics 19:123–132
Denis JB (1991) Ajustement de modèles linéaires et bilinéaires sous contraintes linéaires avec données manquantes. Rev Stat appliquées 34:5–24
Denis JB, Vincourt P (1982) Panorama des methodes statistiques d' analyse des interactions genotype × milieu. Agronomie 2:219–230
Finlay KW, Wilkinson GN (1963) The analysis of adaptation in plant-breeding programmes. Aust J Agric Res 14: 742–754
Frankel OH (1984) Genetic perspectives of germplasm conservation. In: Arber WK, Illimense K, Peacock WJ, Starlinger P (eds) Genetic manipulation: impact of man and society. Cambridge University Press, Cambridge, pp 161–170
Freeman GH (1973) Statistical methods for the analysis of genotype-environment interactions. Heredity 31:339–354
Hardwick RC, Wood JT (1972) Regression methods for studying genotype-environment interactions. Heredity 28:209–222
Hayward MD, De Lacey IH, Tyler BF, Drake DW (1982) The application of pattern analysis for the recognition of adaptation in a collection of *Lolium Multiflorum* populations. Euphytica 31:383–396
Hill J, Samuel CJA (1971) Measurement and inheritance of environmental response amongst selected material of *Lolium perenne*. Heredity 27:265–276
Kempton RA (1984) The use of biplots in interpreting variety by environment interactions. J Agric Sci 103:123–135
Mandel J (1971) A new analysis of variance model for non-additive data. Technometrics 13:1–18
Perkins JM (1972) The principal component analysis of genotype-environmental interactions and physical measures of the environment. Heredity 29:51–71
Perkins JM (1974) Orthogonal and principal component analysis of genotype-environmental interactions for multiple traits. Heredity 32:189–209
Wood JT (1976) The use of environmental variables in the interpretation of genotype-environment interaction. Heredity 37:1–7
Wright AJ (1971) The analysis and prediction of some two-factor interactions in grass breeding. J Agric Sci 76:301–306
Yates F, Cochran WG (1938) The analysis of groups of experiments. J Agric Sci 28:556–580