Discriminating Stay-Green Grasses Using Hyperspectral Imaging and Chemometrics

J. Taylor
Institute of Grassland and Environmental Research, UK

B. Moore
Institute of Grassland and Environmental Research, UK

J. J. Rowland
University of Wales, UK

H. Thomas
Institute of Grassland and Environmental Research, UK

H. J. Ougham
Institute of Grassland and Environmental Research, UK

The XX International Grassland Congress took place in Ireland and the UK in June-July 2005. The main congress took place in Dublin from 26 June to 1 July and was followed by post congress satellite workshops in Aberystwyth, Belfast, Cork, Glasgow and Oxford. The meeting was hosted by the Irish Grassland Association and the British Grassland Society.

Proceedings Editor: D. A. McGilloway

Publisher: Wageningen Academic Publishers, The Netherlands

Discriminating stay-green grasses using hyperspectral imaging and chemometrics

J. Taylor, B. Moore, J.J. Rowland*, H. Thomas and H. Ougham
*Institute of Grassland and Environmental Research, Plas Gogerddan, Aberystwyth SY23 3EB and *Department of Computer Science, University of Wales, Aberystwyth, SY23 3EB, UK. Email: janet.taylor@bbsrc.ac.uk*

Keywords: heterozygote, homozygote, machine learning, principal components analysis, trait screening

Introduction Screening of plant collections for traits can be expensive, in terms of the number of plants to be screened, the duration of the plant lifecycle and the required observations. This study describes the application of a non-invasive method, hyperspectral imaging, combined with multivariate analysis, to distinguish between homozygous wild-type (YY) *Lolium multiflorum* and *Lolium multiflorum* F2 back cross plants heterozygous for y, a recessive *Festuca pratensis* stay-green gene (Thomas *et al*., 1997).

Materials and methods Plants were maintained and propagated in a frost-free greenhouse as vegetative tillers grown in soil-based compost. Four comparable leaf segments were imaged from 4 biological replicates of homozygous and 4 heterozygous plants, on the same day and under identical conditions. Images were acquired in the range 450nm to 720nm at a resolution of 2nm. The hyperspectral imaging system comprised a SenSys CCD camera coupled to a CRI VariSpec LCTF fitted with a hot mirror, with a standard F-mount 35mm Nikon lens. Images were processed using Matlab 6.5.1. Spectra were extracted from individual pixels in each leaf image and corrected for background reflectance using a Spectralon™ reflectance standard. Individual pixel spectra were used in the modelling, to utilise spatial and well as spectral information. One biological replicate of each of the homozygous and heterozygous samples were removed from the dataset to form an independent test set (13500 spectra). Remaining replicates were further combined into a training data set (40740 spectra).

Results Unsupervised learning was carried out using Principal Components Analysis (Jackson, 1991) (Matlab 6.5.1, The Mathworks, (www.mathworks.com), running under Windows XP). The training data set was used to form the PCA model. Data were centred and scaled prior to analysis. The first four PCs accounted for 87.3%, 3.1%, 0.9% and 0.8% of the total variance, respectively. The remaining independent test samples were projected into the original PC space; these are shown in Figure 1. While there are two distinct groups in PC space, there is some overlap of the groups using unsupervised learning. A java implementation (WEKA, www.cs.waikato.ac.nz/ml/weka/) of a supervised decision tree algorithm (C4.5, Quinlan, 1993) was applied to the training set, using 3 fold cross validation. The model produced only misclassified 90 out of 40740 spectra, an accuracy of over 99%. The resultant confusion matrix is shown in Table 1.

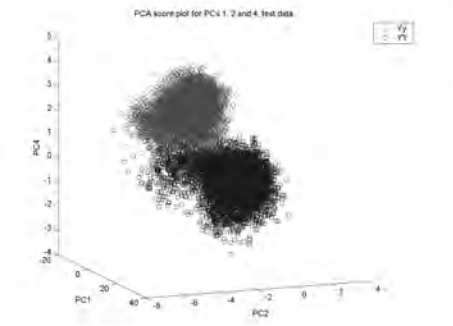

Figure 1 PCA Scores for components 1, 2 and 4

Table 1 Confusion Matrix for C4.5 decision tree

Training set (cross validated)		
Yy	YY	← classified as
17221	34	Yy
45	23440	YY

Conclusions Determination of the allele status of the F2 backcross population is currently verified by a further test cross, which can add up to a year to a breeding programme or genetic study. The combination of hyperspectral imaging, together with chemometric methods indicate that a rapid screening for this trait may be developed. The next stage step is to screen a wider number of plants to further validate the models described.

References

Thomas, H *et al*. (1997) Introgression, tagging and expression of a leaf senescence gene in Festulolium. New Phytologist 137: 29-34.
Jackson, J.E. (1991). A Users Guide to Principal Components. John Wiley and Sons, New York.
Quinlan, J.Ross (1993). C4.5 Programs for Machine Learning. Morgan Kaufmann, San Mateo, CA