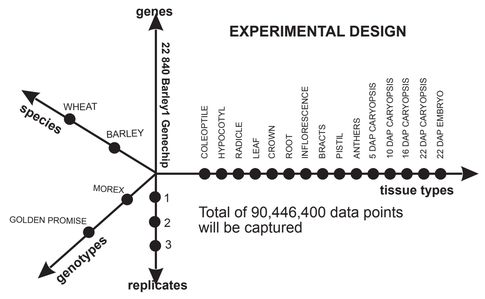
GENE EXPRESSION PATTERNS IN BARLEY DEVELOPMENT

Arnis Druka (1), Gary Muehlbauer (2), Ilze Druka (1), Rico Caldo (3), Ute Baumann (8), Andreas Schreiber (8), Roger Wise (3), Timothy Close (9), Andris Kleinhofs (5), Andreas Graner (6), Alan Schulman (7), Peter Langridge (8), Kazuhiro Sato (10) , Patrick Hayes (4), David Marshall (1) and Robbie Waugh (1)

1 Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, SCOTLAND
2 Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, 55108, USA
3 Department of Plant Pathology, Iowa State University, Ames, Iowa 50110-1120 USA
4 Department of Crop and Soil Science, Oregon State University, Corvallis, OR 97331 USA
5 Departments of Crop and Soil Sciences & Genetics and Cell Biology, Washington State University, Pullman, WA 99164-6420, USA
6 Institut für Pflanzengenetik und Kulturpflanzenforschung, Correnstraße 2, D-06466 Gatersleben, GERMANY
7 MTT/BI Plant Genomics Laboratory, University of Helsinki and MTT Agrifood Research Finland, P.O. Box 56, FIN-00014 Helsinki, FINLAND
8 University of Adelaide, Plant Science, Waite Campus, PMB 1, Glen Osmond SA 5064, AUSTRALIA
9 Department of Botany and Plant Sciences, University of California, Riverside, CA, 92521-0124, USA
10 Research Institute for Bioresources, Okayama University, Kurashiki, 710-0046, JAPAN

Barley, *Hordeum vulgare* L. is particularly well suited for investigating patterns in monocot development. The recent release of the 22K Affymetrix Barley1 GeneChip probe array provides the opportunity to examine gene expression throughout the life cycle of a major cereal crop. In order to provide a reference data set for future investigations and hypothesis testing, transcriptional profiles of ca 22,700 barley genes were examined for 8 developmental stages and 15 tissue types in three independent replications. Here we present the results of initial data analysis from barley and outline the potential of this dataset for immediate genetic target identification and use for tissue-specific studies.



SPECIFIC OBJECTIVES AND METHODS

1) Evaluate Barley1 GeneChip

Affymetrix internal controls and measures from the MAS 5.0 EXP files were used to assess data quality and to perform analysis of variance (ANOVA). The total number of statistically significant positive signals and the level of variability across 21 condition was estimated.

2) What are the genes specifically expressed in each of 15 tissue types?

Diagonal linear discriminant analysis (DLDA) (Dudoit, S. & Fridlyand, J. 2003) was used to determine tissue specific genes. Probe sets with permutation test p value >0.005 were selected as a tissue specific.

3) What are the genes differentially expressed between two barley cultivars?

ANOVA using Welch t-test (p>0.05) and Benjamini & Hochberg False Discovery Rate detection algorithmn as a multiple testing correction was used to determine differentially expressed genes between 6 tissue types of cvs Morex and Golden Promise.

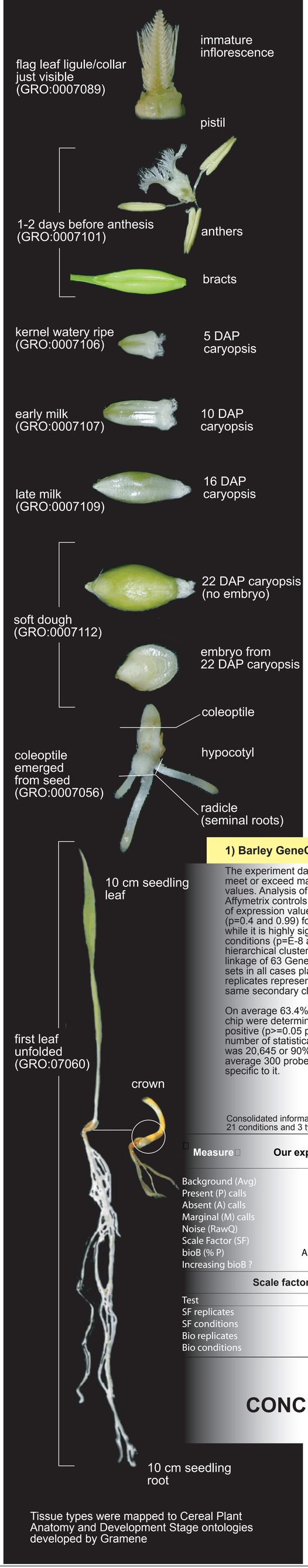
4) Exploration of different complexity reduction methods

To identify naturally occurring clusters within the data set, QT clustering (Heyer, L. et al 1999) and a template matching method (Pavlidis, P. & Noble, S. 2001) was used on three different experiment interpretations; tissue types (15 conditions), genotype-tissue types (12 conditions) and a seed development time course (6 conditions).

5) Assessment of the hypothesis building potential

Gene function can be inferred based on clustering of expression patterns with known function genes. Hypothetical genes have to be present in the classes containing informative known function genes. Assesment of distribution of the hypotetical genes was done by calculating proportion of them in gene classes from 2); 3) and 4)

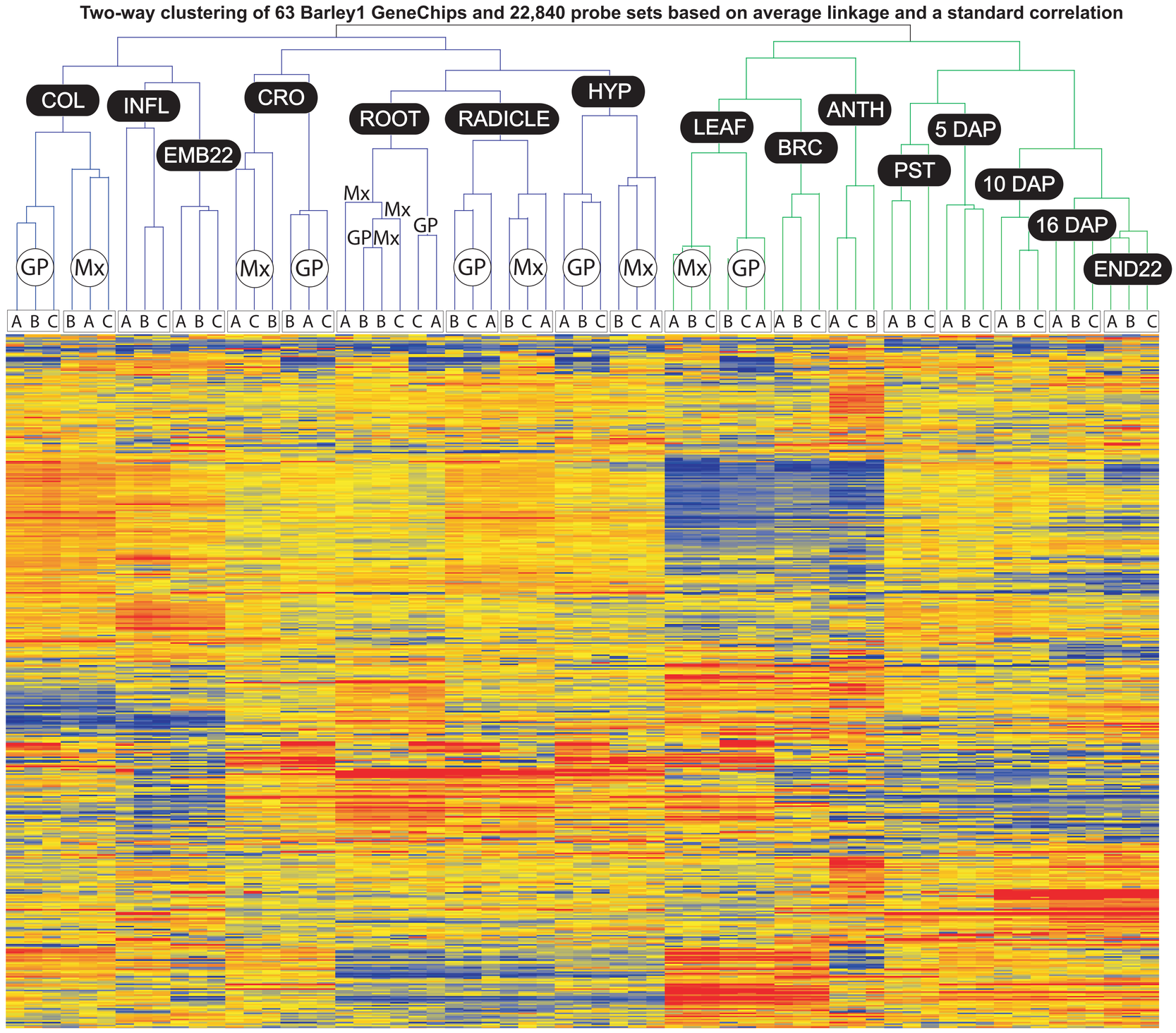
The data set will be publicly available from:
BarleyBase at http://barleypop.vrac.iastate.edu/BarleyBase/
ArrayExpress at http://www.ebi.ac.uk/arrayexpress/



Tissue types were mapped to Cereal Plant Anatomy and Development Stage ontologies developed by Gramene

RESULTS

Two-way clustering of 63 Barley1 GeneChips and 22,840 probe sets based on average linkage and a standard correlation



COL - coleoptile
INFL - immature inflorescence
EMB22 - embryo from 22DAP caryopsis
CRO - crown
HYP - hypocotyl
BRC - bracts
ANTH - anthers
PST - pistil
5, 10, 16 DAP - caryopsis, days after pollination
END22 - 22 DAP caryopsis without embryo
GP - Golden Promise
Mx - Morex

1) Barley GeneChip evaluation

The experiment data quality measures meet or exceed manufacturers suggested values. Analysis of variance using Affymetrix controls indicated that variance of expression values is not significant (p=0.4 and 0.99) for biological replicates, while it is highly significant between conditions (p=E-8 and E-23). The 2-way hierarchical clustering based on average linkage of 63 GeneChips and 22,840 probe sets in all cases placed corresponding replicates representing a tissue type on the same secondary clade.

On average 63.4% +/- 3.4 % probe sets per chip were determined as statistically positive (p>=0.05 per probe set). The total number of statistically significant probe sets was 20,645 or 90%, meaning that on average 300 probesets per condition are specific to it.

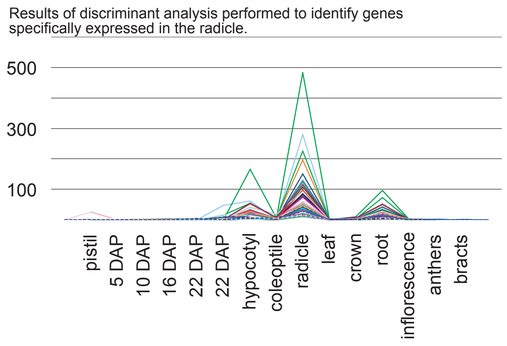
Consolidated information from 63 GeneChips representing 21 conditions and 3 type-2 biological replicates.

Measure	Our experiment (mean)	STDV	Affymetrix suggested
Background (Avg)	57.3	9	20-100
Present (P) calls	63.40%	3.40%	
Absent (A) calls	35.00%	3.30%	
Marginal (M) calls	1.60%	0.20%	
Noise (RawQ)	2.6	0.3	replicates
Scale Factor (SF)	1.1	0.3	replicates
bioB (% P)	A - 3%, M - 1%	na	P>50%
Increasing bioB ?	yes	na	increase
Scale factor (SF) and Bio control analysis of variance			
Test	F	P	F critical
SF replicates	1	0.4	3.2
SF conditions	7.5	3.90E-08	1.8
Bio replicates	0.3	0.99	1.7
Bio conditions	9.8	1.77E-23	1.6

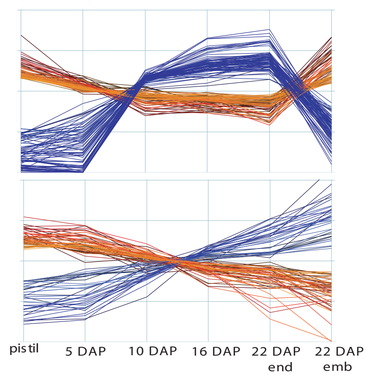
2) Finding tissue type classes

The 15,936 statistically significant differentially expressed genes between at least two conditions were found. Using slightly modified linear diagonal discriminant analysis (d>0.5) on average 85 probe sets per tissue type were identified .

Results of discriminant analysis performed to identify genes specifically expressed in the radicle.



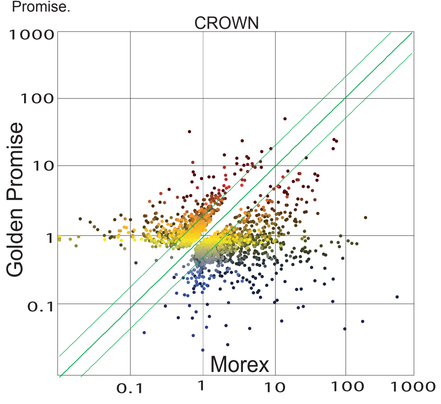
Seed development time course analysis using template matching method with smooth correlation >0.9



3) Genotype-dependent gene expression

A total of 1,236 probe sets were identified as differentially expressed (>2.5 fold) between two barley cultivars, Golden Promise and Morex. 446 (about 30%) probe sets were common for at least two different tissue types. Crown and leaf were the tissue types where most of the differentially expressed genes were found, while in the root none was detected.

Results of 2-way ANOVA to identify differentially expressed genes in the crown tissue between cvs Morex and Golden Promise.



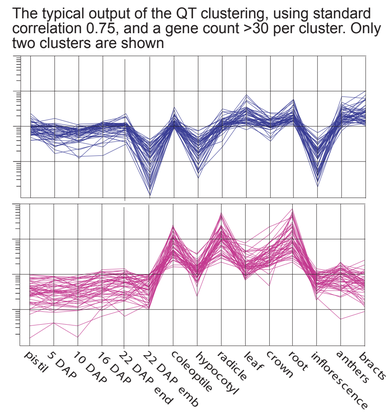
Expression profiles during seed development

Besides QT clustering we used template matching method for finding genes differentially expressed during seed development as shown in the graphs on the left. The overlap between seed development QT clusters and a supervised partitioning was 34% suggesting that by using several alternative partitioning methods it is possible to achieve significant increase in efficiency of complexity reduction.

4) Complexity reduction

Using QT clustering we found 4,774 probe sets in 84 clusters from tissue type interpretation (r>0.75, min 25), 5,824 probe sets in 40 clusters from genotype-tissue interpretation (r>0.75, min 50) and 3788 probe sets grouped in 69 clusters (r>0.9, min 50) from seed development time course. The 55% of total probe sets or 78% of informative probe sets can be assigned to the clusters. The overlap between three groups was only 17%.

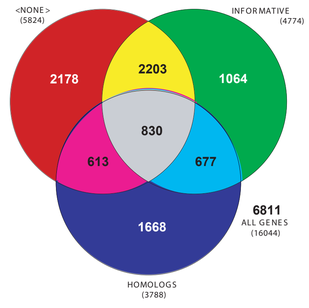
The typical output of the QT clustering, using standard correlation 0.75, and a gene count >30 per cluster. Only two clusters are shown



5) Partitioning of the hypothetical genes

There are 9,529 hypothetical gene probe sets on the on the Barley1 GeneChip of which 5,096 of them belong to different expression classes. There are total of 13,397 explained informative probe sets, 62% of which have some functional assignment. The individual classes contain on average about 30% of the hypothetical genes, indicating unbiased expression profile-based partitioning of the hypothetical genes.

Venn diagram representing different groups of hypothetical genes.



<NONE>
best BlastX hits

HOMOLOGS
best BlastX hits to hypothetical rice, arabidopsis and other proteins

INFORMATIVE
group of hypothetical genes distributed among previously established groups.

CONCLUSIONS

By performing this experiment we have captured highly informative data set which can be used for expression pattern - based gene annotation, immediate genetic target identification, hypothesis building, expression data validation and the future experiment planning.
Sequence - based annotation and confirmation of the selected probe sets is our current task. Analysis and integration of wheat expression data is planned for the near future.