Development of an Affymetrix dual species *(Neotyphodium lolii/ Lolium perenne)* Symbiosis GeneChip®

C.R. VOISEY[1*], A.K. KHAN[2], Z.A. PARK[1], L.J. JOHNSON[1], R.D. JOHNSON[1], M. RAMAKRISHNA[1], M. CAO[1], S. BASSETT[1], C.N. GABORIT[1], A.F. McCULLOCH[2], D. SIMON[1], S. RASMUSSEN[1], M.J. CHRISTENSEN[1] and G.T. BRYAN[1].
[1] *AgResearch, Grasslands Research Centre, Tennent Drive, Private Bag 11008, Palmerston North 4442, New Zealand.* [2] *AgResearch, Invermay Agricultural Centre, Puddle Alley, Private Bag 50034, Mosgiel, New Zealand.*
*christine.voisey@agresearch.co.nz

Abstract

The aim of this project was to undertake large scale transcript profiling of endophyte and plant genes during symbiosis, and to determine the impact of targeted endophyte gene deletions on expression of plant and endophyte genes. We have designed and developed an Affymetrix NimbleExpress™ GeneChip® representing expressed sequence tags (ESTs) of the fungal endophyte *Neotyphodium lolii* Lp19 and its ryegrass host, *Lolium perenne*. In total, 8511 genes were represented on the microarrays with approximately eleven 25 base pair oligonucleotides per gene. Experiments were conducted to analyse differential expression of genes from endophyte-infected and endophyte-free plant material, and from endophytes grown in culture. In some symbioses, endophytes had targeted mutations in genes involved in signalling, synthesis of secondary metabolites or in genes of unknown function. Here we describe the processes which guided design of the GeneChip®, the results of quality control assessments of hybridised arrays and considerations concerning statistical analyses of gene expression.

Keywords: Affymetrix, GeneChip®, NimbleExpress, *Neotyphodium lolii*, *Epichloë festucae*, ryegrass, *Lolium perenne*, endophyte, symbiosis.

Introduction

DNA microarrays enable changes in the transcriptome of an organism to be determined under defined environmental conditions. The dataset created is comprised of gene expression profiles for thousands of genes, permitting a global assessment (at the RNA level) of the response of an organism to the conditions imposed.

Recently, a number of high throughput EST sequencing projects in Australia and New Zealand have provided the first large datasets of transcripts from *Neotyphodium* (Felitti *et al.* 2006; Johnson *et al.* 2007d) and *Epichloë* (Felitti *et al.* 2006) fungal endophytes and their grass hosts. These sequences have facilitated the development of microarrays for a comprehensive analysis of transcripts involved in symbiosis between these species, and permit questions concerning mechanisms of symbiosis to be addressed. Felitti *et al*. (2006) have constructed cDNA arrays of ESTs from *N. coenophialum*, *N. lolii* and *E. festucae*, in culture and in association with ryegrass or tall fescue, and used the arrays to identify genes expressed in *N. lolii* and *N. coenophialum* during saprophytic growth in culture. Here we present the development of a NimbleExpress™ Affymetrix GeneChip® representing a subset of sequences from both endophytes and their grass host. The objective of developing this GeneChip® was to identify genes differentially expressed during symbiosis between *E. festucae* Fl1, *N. lolii* Lp19 and *L. perenne*. A further aim was to identify genes involved in specific processes or pathways by comparing expression profiles of genes from wild-type symbioses with ryegrass infected with endophytes containing targeted deletions in individual genes.

We describe the methodology used to select and prepare sequences for GeneChip® design, the outcomes of data quality assessment and preliminary statistical treatments of data.

Table 1 Ten samples (each represented by three biological replicates) were subjected to microarray analysis in GeneChip® hybridisation experiments. Material was harvested from ryegrass leaf sheath (l.s.) or pseudostem (p.) and from endophytes grown in potato dextrose broth (c.). Endophytes with targeted gene deletions included Δ*SidF* (disrupted siderophore gene, iron acquisition), ΔNRPS1 (disrupted NRPS1 gene, a non-ribosomal peptide synthetase), Δ*acyA* (disrupted adenylate cyclase gene, cAMP synthesis) and ΔNC25 (disrupted NC25 gene, function unknown).

	Tissue	Endophyte species	Endophyte genotype
1	Ryegrass (l.s.)	*N. lolii* Lp19	Wild-type
2	Ryegrass (l.s.)	*E. festucae* Fl1	Wild-type
3	Ryegrass (l.s.)		endophyte-free
4	Ryegrass (l.s.)	*E. festucae* Fl1	ΔSidF
5	Ryegrass (l.s.)	*E. festucae* Fl1	ΔNRPS1
6	Ryegrass (p.)	*E. festucae* Fl1	Wild-type
7	Ryegrass (p.)	*E. festucae* Fl1	Δ*acyA*
8	Ryegrass (p.)	*E. festucae* Fl1	ΔNC25
9	Endophyte (c.)	*N. lolii* Lp19	Wild-type
10	Endophyte (c.)	*E. festucae* Fl1	Wild-type

Material and Methods

GeneChip® design

The gene sequences tiled on the NimbleExpress™ microarray were obtained from a number of sources. ESTs were derived from *N. lolii* Lp19, either growing in liquid potato dextrose broth or in minimal media, and also from suppressive subtractive hybridisation (SSH) libraries of *N. lolii* Lp19 growing in association with *L. perenne* (Johnson *et al.* 2007d). The sequences were trimmed and assembled into contiguous sequences (hereafter referred to as contigs) as described (Johnson *et al.* 2007d). In addition, open reading frames (ORFs) were identified from two sequenced clones (L3 and L23) isolated from a bacterial artificial chromosome (BAC) library of *N. lolii* (Lp19). ORFs were predicted using Genezilla (Majoros *et al.* 2004) and SNAP (Korf 2004) based on *Fusarium graminearum* parameter models (Jason Stajich, http://fungal.genome.duke.edu/), and by aligning endophyte EST contigs (described above) against the BAC genomic contigs using Exonerate (Slater & Birney 2005). Further ORFs were extracted from additional sources of endophyte genomic DNA (gDNA), including the secondary metabolite gene clusters for peramine (Tanaka *et al.* 2005), ergovaline (Fleetwood *et al.* 2007) , lolitrem B (Young *et al.* 2006) and loline (Spiering *et al.* 2005), plus degenerate PCR fragments enriched for non-ribosomal peptide synthetases (NRPS), polyketide synthases (PKS) and ATP-binding cassette (ABC) transporters. These sequences will be described in greater detail elsewhere. Finally, endophyte genes were also sourced from the public domain (GenBank). Ryegrass EST contigs (Sawbridge *et al.* 2003) were selected from libraries representing genes of interest in the context of endophyte/grass symbiosis based on homology to characterised genes in other plant species. Functional categories of ryegrass genes selected for representation on the GeneChip® included synthesis of plant hormones, primary and secondary metabolites, transcription factors, transporters, plus genes implicated in responses to abiotic stress, wounding, signalling, UV, and disease resistance. During GeneChip® design, priority was given to tiling all endophyte genes, plus the ryegrass genes from SSH libraries. The balance of GeneChip® capacity was then apportioned to *L. perenne* EST contigs and to control sequences from both species. Selection of control genes will be described elsewhere.

Preparation of biological material

For details of the growing conditions for endophytes and plants, and the extraction and purification of RNA, refer to Johnson *et al.* (2007b). Analysis of RNA quality, labelling, chip hybridisations, washes and GeneChip® scanning followed recommended procedures and were conducted at the Centre for Genomics and Proteomics, School of Biological Sciences, University of Auckland, New Zealand.

Results

Proportions of species represented on the GeneChip®

Approximately 32% of the probe sets were derived from endophytes, 55% from ryegrass and a further 13% remained uncharacterised after BLAST analysis and empirical methods (to be described later) to identify species of origin.

Quality assessment of hybridisation data

Thirty arrays were hybridised with three biological replicates each for 10 experimental samples (Table 1). The data were subjected to the following quality assessment metrics prior to Robust Multi-array Average (RMA) normalisation (Irizarry *et al.* 2003).

Scale factors were calculated for each GeneChip® to determine uniformity of scale between hybridisations. Scale factors were similar over all 21 slides hybridised with cRNA from the symbiota, indicating that RNA quality, labelling and hybridisation was consistent across these slides. Higher than average scale factors were expected for endophyte-only, and ryegrass-only samples as the proportion of expressed sequences, and so overall mean intensity, on these chips was expected to be lower. This was, however, not the case for the slides to which cRNA from *N. lolii* Fl1 (growing in culture) or endophyte-free *L. perenne* was hybridised. Notably, the variance in scale factor for the three slides of *N. lolii* Lp19 growing in culture was considerably higher than that for slides from any other treatment suggesting that the quantity or quality of the RNA, or the labelling reaction, was not optimal in one or more of the replicates. This was later confirmed (see below).

'Present', 'Marginal' or 'Absent' (PMA) calls were generated using the "mas5calls" procedure in the "affy" BioConductor package (Gautier *et al.* 2004). This procedure uses the Wilcoxon signed rank test to classify the probe sets as 'Present' or otherwise. The percentage of 'Present' calls for slides of the same treatment was consistent, with slides varying by only 0.2-4.3%, indicating that the quality of the GeneChip® hybridisations was high. As expected, the proportion of 'Present' calls was considerably lower (compared with slides hybridised with cRNA of endophyte-infected plants) on slides hybridised with *E. festucae* or *N. lolii* in culture and marginally lower on slides to which only ryegrass cRNA was hybridised. The maximum proportion of 'Present' probe sets detected on a slide was 59.4%.

Standard Affymetrix bacterial spike-in controls, BioB, BioC, BioD and CreX were detected on all 30 slides, indicating that labelling and hybridisation of target cRNA and scanning processes performed according to specifications.

A further 13 bacterial genes were tiled on the GeneChip® as cRNA quality controls. Separate probe sets targeting the 3′ and 5′ ends of the gene were tiled for each bacterial gene, and the 3′ to 5′ ratio calculated to give a measure of the quality of the cRNA. Sixteen of the 17 bacterial genes tiled on the GeneChips® had 3′ to 5′ ratios of approximately 1 on most arrays, indicating that RNA quality was high and that *in vitro* transcription during labelling had been efficient. However, for one slide (replicate 1 of *N. lolii* Lp19 in culture), 3′ to 5′ ratios were significantly increased in several bacterial control genes indicating that the quality of cRNA in this sample was compromised.

Identification of differentially expressed (DE) genes

The comparisons between gene expression profiles of biological samples are currently being analysed (Johnson *et al.* 2007a; Johnson *et al.* 2007b; Johnson *et al.* 2007c; Voisey *et al.* 2007). Differential expression was established by ranking probe sets based on the results of performing a modified t-test on the normalised intensity values, using the "limma" package in Bioconductor (Smyth 2005). This approach will be described in more detail elsewhere.

Discussion

We have designed a NimbleExpress Affymetrix GeneChip® representing a subset of genes from two species, *N. lolii* and *L. perenne,* to identify genes and processes that are important in symbiosis. Overall, quality assessment metrics indicated that the quality of the microarray data was good, with acceptable

results for 29 of the 30 slides. A single slide (representing *N. lolii* Lp19 in culture) failed quality assessment, possibly due to poor labelling or RNA degradation during storage.

The microarrays have enabled identification of differentially expressed genes involved in symbiosis between endophytes and ryegrass, and from ryegrass infected with endophytes with deletions in genes involved in cAMP signalling (Voisey *et al.* 2007), iron acquisition (Johnson *et al.* 2007a), secondary metabolite biosynthesis (Johnson *et al.* 2007b) and other genes of unknown function (Johnson *et al.* 2007c). The microarray data from these studies will be logged in the public domain. Functional annotation of genes and investigation of their role in symbiosis will be a priority for this team in the future and we will explore further data analysis methods to utilise this microarray dataset fully, such as clustering of expression profiles and promoter analysis of co-expressed genes. The annotation of hypothetical and orphan endophyte sequences will also be greatly assisted by the endophyte genome sequencing initiative currently underway (C. Schardl, pers. comm.).

ACKNOWLEDGEMENTS

We would like to acknowledge Liam Williams at the University of Auckland Microarray Facility at the Centre for Genomics and Proteomics (New Zealand) for analysis of RNA quality, labelling of cRNA, GeneChip® hybridisation and scanning. Sophie Borchert, Catherine Tootill and Jennifer Pratt (Grasslands Research Centre, Palmerston North, AgResearch) provided technical assistance. This project was supported by the Foundation for Research Science and Technology (FRST), New Zealand, through contracts C10X0203, AGRX0204 and PROJ-10333-ECOS-AGR, and AgResearch AR&C projects 5598/1 and 9959/1.

REFERENCES

Felitti, S.; Shields, K.; Ramsperger, M.; Tian, P.; Sawbridge, T.; Webster, T.; Logan, E.; Erwin, T.; Forster, J.; Edwards, D.; Spangenberg, G. 2006. Transcriptome analysis of *Neotyphodium* and *Epichloë* grass endophytes. *Fungal Genetics and Biology 43*: 465.

Fleetwood, D.J.; Scott, B.; Lane, A.; Tanaka, A.; Johnson, R.D. 2007. A complex ergovaline gene cluster in Epichloë endophytes of grasses. *Applied and Environmental Microbiology.* in press.

Gautier, L.; Cope, L.; Bolstad, B.M.; Irizarry, R.A. 2004. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics 20*: 307-315.

Irizarry, R.A.; Hobbs, B.; Collin, F.; Beazer-Barclay, Y.D.; Antonellis, K.J.; Scherf, U.; Speed, T.P. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics 4*: 249-264.

Johnson, L.; Steringa, M.; Koulman, A.; Christensen, M.; Johnson, R.; Voisey, C.; Bryan, G.T.; Lamont, I.; Rasmussen, S. 2007a. Biosynthesis of an extracellular siderophore is essential for maintenance of mutualistic endophyte-grass symbioses. pp. 177 *In*: Proceedings of the 6th International Symposium on Fungal Endophytes of Grasses. Eds. Popay, A.J.: Thom, E.R. Grassland Research and Practice Series No. 13. New Zealand Grassland Association.

Johnson, L.J.; Voisey, C.R.; Johnson, R.D.; Khan, A.K.; Park, Z.A.; Ramakrishna, M.; Cao, M.; Simon, D.; Christensen, M.; Bryan, G.T.; Rasmussen, S. 2007b. Dual Affymetrix GeneChip® analysis of the perennial ryegrass-endophyte symbiosis. pp. 509 *In*: Proceedings of the 6th International Symposium on Fungal Endophytes of Grasses. Eds. Popay, A.J.: Thom, E.R. Grassland Research and Practice Series No. 13. New Zealand Grassland Association.

Johnson, R.D.; Borchert, S.; Christensen, M.; Johnson, L.J.; Koulman, A.; van Gils, M.J.; Bryan, G. 2007c. A gene identified from *Neotyphodium lolii* is expressed only *in planta* and regulates the biosynthesis of a putative oligopeptide secondary metabolite. pp. 485 *In*: Proceedings of the 6th International Symposium on Fungal Endophytes of Grasses. Eds. Popay, A.J.: Thom, E.R. Grassland Research and Practice Series No. 13. New Zealand Grassland Association.

Johnson, R.D.; Khan, A.K.; Voisey, C.R.; Bassett, S.; C., G.; Johnson, L.J.; Christensen, M.; McCulloch, A.; Bryan, G.T. 2007d. Analysis of expressed sequence tags derived from the endophytic fungus *Neotyphodium lolii* grown *in vitro* and in association with its host plant perennial ryegrass. pp. 501 *In*: Proceedings of the 6th International Symposium on Fungal Endophytes of Grasses. Eds. Popay, A.J.: Thom, E.R. Grassland Research & Practice Series 13. New Zealand Grassland Association.

Korf, I. 2004. Gene finding in novel genomes. *BMC Bioinformatics 5*: 59.

Majoros, W.H.; Pertea, M.; Salzberg, S.L. 2004. TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. *Bioinformatics 20*: 2878-2879.

Sawbridge, T.; Ong, E.-K.; Binnion, C.; Emmerling, M.; McInnes, R.; Meath, K.; Nguyen, N.; Nunan, K.; O'Neill, M.; O'Toole, F. 2003. Generation and analysis of expressed sequence tags in perennial ryegrass (*Lolium perenne* L.). *Plant Science 165*: 1089-1100.

Slater, G.; Birney, E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics 6*: 31.

Smyth, G.K. 2005. Limma: linear models for microarray data. pp. 397-420. In: Bioinformatics and Computational Biology Solutions using R and Bioconductor. Eds. Gentleman, R.; Carey, V.; Dudoit, S.; Irizarry, R.; Huber, W. Springer, New York.

Spiering, M.J.; Moon, C.D.; Wilkinson, H.H.; Schardl, C.L. 2005. Gene clusters for insecticidal loline alkaloids in the grass-endophytic fungus *Neotyphodium uncinatum*. *Genetics 169*: 1403-1414.

Tanaka, A.; Tapper, B.A.; Popay, A.; Parker, E.J.; Scott, B. 2005. A symbiosis expressed non-ribosomal peptide synthetase from a mutualistic fungal endophyte of perennial ryegrass confers protection to the symbiotum from insect herbivory. *Molecular Microbiology 57*: 1036-1050.

Voisey, C.R.; Christensen, M.T.; Johnson, R.D.; Johnson, L.J.; Khan, A.K.; Park, Z.A.; Ramakrishna, M.; Cao, M.; Koolaard, J.P.; Koulman, A.; Borchert, S.; Gaborit, C.N.; Pratt, J.; Bryan, G.T. 2007. The role of cAMP signalling in the symbiosis between *Epichloë festucae* and *Lolium perenne*. pp. 457 *In*: Proceedings of the 6th International Symposium on Fungal Endophytes of Grasses. Eds. Popay, A.J.: Thom, E.R. Grassland Research and Practice Series No. 13. New Zealand Grassland Association.

Young, C.A.; Felitti, S.; Shields, K.; Spangenberg, G.; Johnson, R.D.; Bryan, G.T.; Saikia, S.; Scott, B. 2006. A complex gene cluster for indole-diterpene biosynthesis in the grass endophyte *Neotyphodium lolii*. *Fungal Genetics and Biology 43*: 679-693.