		M	E	C	2	7	7	4	**Operator:** Zhang Yexia	**Dispatch:** 14.10.05	**PE:** Ann Cowie
	Journal Name				Manuscript No.				**Proofreader:** Chen Xiaoming	**No. of Pages:** 14	**Copy-editor:** Apple Rosales

Molecular Ecology (2005) doi: 10.1111/j.1365-294X.2005.02774.x

Natural genetic variation in whole-genome expression in *Arabidopsis thaliana*: the impact of physiological QTL introgression

THOMAS E. JUENGER,* TIERNEY WAYNE,* SANDRA BOLES,* V. VAUGHAN SYMONDS, * JOHN MCKAY† and SEAN J. COUGHLAN‡

**The University of Texas at Austin, Section of Integrative Biology and Institute for Cellular and Molecular Biology, 1 University Station C0930, Austin, TX 78712, †University of California Davis, Center for Population Biology, One Shields Avenue, Davis, CA 95616, ‡Agilent Technologies Inc., Little Falls Site, 2859 Centerville Road, Wilmington, DE 19808-1644, USA*

Abstract

A long-standing and fundamental question in biology is how genes influence complex phenotypes. Combining near-isogenic line mapping with genome expression profiling offers a unique opportunity for exploring the functional relationship between genotype and phenotype and for generating candidate genes for future study. We used a whole-genome microarray produced with ink-jet technology to measure the relative expression level of over 21 500 genes from an *Arabidopsis thaliana* near-isogenic line (NIL) and its recurrent parent. The NIL material contained two introgressions (bottom of chromosome II and top of chromosome III) of the Cvi-1 ecotype in a L*er*-2 ecotype genome background. Each introgression 'captures' a Cvi allele of a physiological quantitative trait loci (QTL) that our previous studies have shown increases transpiration and reduces water-use efficiency at the whole-plant level. We used a mixed model anova framework for assessing sources of expression variability and for evaluating statistical significance in our array experiment. We discovered 25 differentially expressed genes in the introgression at a false-discovery rate (FDR) cut-off of 0.20 and identified new candidate genes for both QTL regions. Several differentially expressed genes were confirmed with QRT–PCR (quantitative reverse transcription–polymerase chain reaction) assays. In contrast, we found no statistically significant differentially expressed genes outside of the QTL introgressions after controlling for multiple tests. We discuss these results in the context of candidate genes, cloning QTL, and phenotypic evolution.

Keywords: *Arabidopsis thaliana*, eQTL, gene expression, MAANOVA, oligonucleotide array

Received 30 May 2005; revision accepted 12 September 2005

Introduction

A major goal of modern genetics is to understand the complex mapping of genotype to phenotype. This is an enormous challenge as nearly all phenotypes in natural populations are affected by many loci, the environment, and their interactions (Falconer & Mackay 1989; Lynch & Walsh 1998). Classic approaches for studying polygenic traits have largely ignored the molecular, physiological, and developmental processes that link genotype to organism-level phenotype and instead have focused on statistical aspects of variability (Barton & Keightley 2002). Genetic mapping techniques provide one way to localize effects to particular genome locations (Lander & Schork 1994; Lynch & Walsh 1998; Flint & Mott 2001), but these studies only rarely provide genuine biological insight into how genotypes translate into phenotypes (Schlichting & Pigliucci 1998). Similarly, molecular and functional genetic approaches are seldom applied to phenotypes at the whole-organism scale. Technological and molecular advances of the last decade are now allowing additional

Correspondence: Thomas E. Juenger, Fax: 512/471 3878; E-mail: tjuenger@mail.utexas.edu

information to be collected at intermediate levels of organization (e.g. transcription, translation, protein modification/degradation, developmental interactions). These studies are likely to provide a better understanding of the biological processes important in the translation of DNA to phenotype across ontogeny and environments (Jin *et al*. 2001; Gibson & Mackay 2002).

An important step in this regard has been the development of tools for screening mRNA transcript abundance at a whole-genome level (Brown & Botstein 1999). The abundance of mRNA within specific cells, tissues, or organs is both genetically and environmentally determined and as such can be considered a quantitative trait (Eves *et al*. 2002; Schadt *et al*. 2003; Townsend *et al*. 2003; Brem & Krugylak 2005). Expression profiling and genetic mapping have documented heritable patterns of gene expression and have provided evidence for both *cis*- and *trans*-acting regulation of natural expression variability. In several cases, genetic studies of natural variation coupled with gene expression analyses have been used to identify candidate genes for complex diseases (Aitman *et al*. 1999; Karp *et al*. 2000; Eves *et al*. 2002; Schadt *et al*. 2003), physiological processes (Townsend *et al*. 2003; Fay *et al*. 2004; Baxter *et al*. 2005), and developmental characteristics (Wayne & McIntyre 2002; van Driessche *et al*. 2005) as well as to provide the first partitioning of genetic and environmental components of transcriptional variance (Jin *et al*. 2001; Brem *et al*. 2002; Oleksiak *et al*. 2002; Gibson *et al*. 2004; Morley *et al*. 2004; Wayne *et al*. 2004; Brem & Krugylak 2005; Chesler *et al*. 2005; Manley *et al*. 2005; Story *et al*. 2005). These data are of interest to ecological and evolutionary biologists, in part, as they provide some of the first data concerning the importance of gene expression changes in phenotypic variability and evolution (Townsend *et al*. 2003; Fay *et al*. 2004).

We have recently identified quantitative trait loci (QTL) controlling whole-plant physiology in *Arabidopsis thaliana* using a set of recombinant inbred lines derived from genetically diverse parental ecotypes (L*er*-2 × Cvi-1; L*er*-2 × Col-4) (Hausmann *et al*. 2005; Juenger *et al*. 2005). These studies focused on integrating quantitative genetic analyses with functional studies of physiological traits, plant performance, and plant responses to soil-moisture deficit and drought. We captured two $\delta^{13}C$ (a lifetime integrated physiological trait) QTL as near-isogenic lines (NIL) to further characterize their physiological basis and confirm their genomic location (Juenger *et al*. 2005). NIL containing Cvi parental alleles at these two $\delta^{13}C$ QTL increase stomatal conductance (g_s), decrease transpiration efficiency, and alter the rate of water loss from detached leaves. QTL effects were constitutive and did not respond to soil moisture variability. The gene(s) underlying these QTL may therefore play important roles in controlling stomatal closure during leaf dehydration as well as during gas exchange under well-watered conditions. We are currently using a number of molecular approaches to identify the causal genetic polymorphisms underlying these QTL.

Here, we report on gene expression studies comparing mRNA abundance in rosette leaf tissue from a near-isogenic line containing Cvi alleles at two physiological QTL to its recurrent parent (L*er*-2). These experiments are used to locate expression level polymorphisms within the introgression elements that may be causally related to the mapped physiological QTL. Furthermore, these data allow an exploration of the transcriptional impact of QTL introgression at a whole-genome level (outside of the introgression elements) and may therefore be helpful in generating hypotheses of either regulatory or downstream functional impacts of sequence physiological QTL. Together, these data provide new insight into the role of gene expression in natural quantitative genetic variability.

Materials and methods

Whole-genome expression analyses

We used a randomized design with three replicates each of a single NIL and its recurrent parent (L*er*-2) in a study of whole-rosette gene expression. The NIL material was developed by backcrossing a recombinant inbred line (RIL) capturing the two target QTL to L*er*-2, along with selfing and marker-assisted selection, to obtain a homozygous line containing the two Cvi introgressions in a L*er*-2 background. The starting RIL in this case was derived from the original L*er*-2 × Cvi mapping population that was densely genotyped with 322 molecular markers with an average intermarker distance of 1.5 centimorgans (cM) (Alonso-Blanco *et al*. 1998). This dense mapping provides an ideal starting point for NIL construction as it clearly indicates the location of 'unwanted' Cvi blocks. After initial NIL construction, plant material was again genotyped in the targeted intervals using 17 and 21 markers near the chromosome II and III introgressions, respectively. The chromosome II introgression occurs from ~17.2 to ~19.2 Mbp and the chromosome III introgression occurs from the left telomere to ~0.70 Mbp. It is important to emphasize, nevertheless, that the particular breakpoints for these introgression have not been fine-mapped and therefore the exact size of the introgressions is unknown. In total, the introgression contains at least 933 genes (chromosome II, 654; chromosome III, 279 genes). Seven hundred eighty-nine (789) of these genes are represented with 60-mer oligos on the Agilent *Arabidopsis* 2 array.

Plants were grown in small (5 cm) pots filled with a peat-based potting mix (Promix BT®). Seeds were planted into moist soil and then cold stratified in the dark for 7 days at 4 °C. The plants germinated and grew in a growth chamber (~150 µmol m^{-2} s^{-1} PPFD). Photoperiod was 14 h light/10 h dark and the temperature cycled 21/20 °C light/dark

with a constant RH of 50%. The position of experimental plants within the growth chamber was randomly shuffled every other day over the course of the experiment. After 20 days, the rosette of each plant was harvested and flash frozen on liquid N_2. Total RNA was extracted from each using a Plant RNeasy® Plant kit (QIAGEN). The tissue used for RNA extraction excluded root material, but included apical meristem, rosette leaves, and remaining cotyledons. RNA extractions were checked for quality and quantified using a 2100 Bioanalyzer instrument (Agilent Technologies).

cRNA was synthesized from 400 ng of each total RNA sample with Cy3 or Cy5-CTP using the Agilent Technologies Low input Linear Amplification kit according to the manufacturer's recommendations. Our experiments utilize the Agilent *Arabidopsis*-2 Oligo Array representing 21 500 genes and covering approximately 80% of the *Arabidopsis thaliana* genome. These arrays are based on the ink-jet 'printing' of 60-base oligo probes developed from the ATH1 version 3 database of The Institute for Genomic Research (TIGR) (Coughlan *et al*. 2004). Samples were hybridized to microarrays in a series of replicated dye-swaps involving 27 arrays. The design was unbalanced due to the number of arrays available for the experiment as well as the amount of resource expended on particular biological samples. In standard microarray notation, our design was as follows: N17⇄L1, N1⇄L1, N1→L1, N1⇄L2, N1⇄L2, N2⇄L1, N2→L2, N2⇄L2, N2⇄L2, N3⇄L1, N3⇄L1, N3⇄L2, N3→L2, L2⇄L3, L1⇄L2, where N1–N3 correspond to three biological replicates of the NIL and L1–L3 correspond to three biological replicates of the recurrent parent (L*er*-2). One-microgram cRNA samples were fragmented and hybridizations were performed using the formamide-based buffer of Hughes *et al*. (2001) with the exception that an Agilent Surehyb microarray hybridization chamber and accessories were used (final hybridization volume of 500 μL) for 18 h at 42 °C. Microarray images were obtained by scanning each slide with an Agilent DNA microarray scanner. Mean intensity values (nonbackground subtracted) for each spot were used in all subsequent analyses.

QRT–PCR confirmation of differentially expressed genes

Quantitative reverse transcription–polymerase chain reactions (QRT–PCR) were used to confirm differential expression of five candidate genes identified through the whole-genome array analysis. Here, we report results for AT2G42500 (serine-threonine protein phosphatase PP2A-3 catalytic subunit), AT2G42590 [14-3-3 protein (GF14 mu)], AT3G01480 (peptidyl-prolyl *cis-trans* isomerase), AT3G01460 (PHD finger family protein/methyl-CpG binding domain-containing protein), and AT3G02220 (expressed protein). Fifteen replicate plants each of L*er*-2 and the NIL were grown in a second experiment as described above. Total RNA was extracted from each replicate plant using a QIAGEN RNeasy Plant kit and subsequently treated with QIAGEN RNase-free DNase. Each sample was quantitatively analysed using an ABgene Absolute™ one-step QRT–PCR kit (ABgene) and the TaqMan® fluorogenic 5′-nuclease PCR assay. Under this protocol, both the RT and PCR are performed in the same reaction mix in a single assay tube. Gene-specific PCR products were measured in real-time with an Applied Biosystems 7900 Sequence Detection System (Applied Biosystems). Primers and probes for each gene were designed using Primer Express® (Applied Biosystems) with FAM and nonfluorescent quencher probe labelling. Primers sequences used for AT2G42500 were: forward, 5′ TGCCCTGATACCAATTACCTGTTT; reverse, 5′ GACTAACAGCGTAACAGTTTCAACA; and probe, CCACGGTCCACATAGTC. Primer sequences used for AT2G42590 were: forward, 5′ GGTTTGGCTTTGAATTTCTCTGTCT; reverse, 5′ CCTGCTTAGCAAGGTGACATG; and probe, CCTTTCAGGTGCGTTCAT. Primer sequences for AT3G1480 were: forward, 5′GCGTTAAGGCTCTAGATTCTGTTGA; reverse, 5′ CCCTTGCTGCAATGTTCTACTTG'; and probe, CCTGCCTCACATTCCGT. Primer sequences used for AT2G42590 were: forward, 5′ GGTTTGGCTTTGAATTTCTCTGTCT; reverse, 5′ CCTGCTTAGCAAGGTGACATG; and probe, CCTTTCAGGTGCGTTCAT. Primer sequences for AT3G1460 were: forward, 5′GACGGATCAACTGAAGCTTTTCTTG; reverse, 5′ GGCTGATCTGCATACATAACACGTA; and probe, TCCACAGCTCAAGAACA. Primer sequences for AT3G02220 were: forward, 5′ GTCAGAAGTGTACAAAGCGAAATGTT; reverse, 5′ GCACACACCTTTTGCTCCTT; and probe, ACAACCAGGGCATAACT.

The relative amount of mRNA generated from each assay was determined on the basis of the threshold cycle (C_T) value for each reaction reflecting the number of PCR cycles needed to identify a target gene. The *ACTIN2* gene was used as an endogenous control to normalize the quantity of RNA used in particular reactions. The primers for *ACTIN2* were: forward, TCCGTTTTGAATCTTCCTCAATCTCA; reverse, TTGAATATCATCAGCCTCAGCCATT; probe, CTTTCTTTCCAAGCTCATAAAA. The C_T value for *ACTIN2* was subtracted from that of each target gene to obtain ΔC_T values, which were used in subsequent statistical analyses. The relative quantitative value of mRNA was expressed as $2^{-\Delta\Delta CT}$, where $\Delta\Delta C_T$ is the difference between the C_T values of the genetic lines (using the line with lower transcript abundance as the calibrator). We also evaluated the eukaryotic 18S rRNA gene as an endogenous control in experiments — these analyses gave identical results to the *ACTIN2* gene. We therefore present only the analyses based on *ACTIN2* standardization.

Confirming probe sequences in introgression element

Our study makes use of the *Arabidopsis*-2 array developed from Col-4 sequence and relies on hybridizing L*er*-2 and

Cvi-1 cRNA to the array probes. It is possible that some of the differential expression observed in the introgression element, where cRNA sequences may differ, is driven by sequence/probe mismatch in Cvi-1 or L*er*-2 that alters hybridization dynamics. To explore this possibility, we developed primers to amplify and sequence the array probes for a set of chromosome III genes within the introgression. Genomic amplified products spanning each probe were cleaned using an ExoSAP-IT® (Amersham Biosciences) procedure and subsequently single pass cycle-sequenced using BigDye® chemistry and an Applied Biosystems 3730 capillary sequencer. Col-4, L*er*-2, and Cvi-1 sequences were aligned using SEQUENCHER™ (Gene Codes Corp.) and mismatches were scored relative to the Col-4 reference.

Statistical analysis

Our experiment generated expression data for two classes of genes, those occurring within or those occurring outside of the NIL introgression elements. Genes occurring within the element may exhibit differential expression as a result of (i) sequence polymorphisms between L*er*-2 and Cvi that generate true difference in mRNA abundance or (ii) mismatches in the probe sequences and subsequent differential hybridization. In contrast, genes occurring outside the introgression element should only differ in transcript abundance as a secondary (*trans*- or downstream) response of polymorphisms within the introgression element. Given this structure, we split our expression analyses into the corresponding two classes of genes and analysed each separately. The data for each set were imported into the R software environment (www.R-project.org) and analysed using an ANOVA-based framework with the R/MAANOVA package (Kerr *et al*. 2000; Kerr & Churchill 2001; Wolfinger *et al*. 2001; Churchill 2002). The data exhibited intensity-dependent biases and therefore log(2) transformed data were further corrected with a global lowess transformation. The overall expression variability was decomposed using the general ANOVA:

$$y_{ij} = \mu + A_i + D_j + (AD)_{ij} + G_{g(ij)} + (DG)_{jg} + (LG)_{kg} + P_{m(ij)} + \varepsilon_{ijkgm}$$

where y_{ij} is the logarithmic intensity on array i in dye channel j. The data for each gene were decomposed into an overall mean, μ; the 'global' effects of A_i, D_j, and $(AD)_{ij}$ which were used to normalize for variation in arrays (A) and dyes (D); the gene effect G_g accounting for the average signal for gene g across arrays, dyes, and treatments; gene-specific dye effects $(DG)_{jg}$; 'line'-specific expression $(LG)_{kg}$ of gene g; a replicate plant effect $P_{m(ij)}$; and measurement error, ε_{ijkgm}. In this case, 'line' corresponds to either the NIL or recurrent parent genetic material. Array and dye are indexed by i and j. The genetic line and plant indices, k, m are determined by values of i and j. The plant term is nested within genetic line. Terms for array, plant, and error were considered random effects and variance components were estimated separately for each gene using methods described in Little *et al*. 1996). We calculated three F-statistics (F_1, F_3, F_S) corresponding to contrast of the $(LG)_{kg}$ term for each gene as implemented in R/MAANOVA. Each of these tests makes a different assumption about the error variance for each gene (F_1 — unique gene error; F_3 — common error across genes; F_S — shrinkage-based estimate of each error term). We focus our attention on F_S as it has been shown to be a relatively robust test statistic (Cui *et al*. 2005). We evaluated statistical significance of the genetic line term with distribution free significance levels obtained through random permutations of the data. We shuffled biological samples to generate the distribution of the test statistic under the null hypothesis of no differential expression between the NIL and recurrent parent. In the case of the genes occurring within the introgression, we performed 10 000 permutations of the data. Shuffling of the larger data set was restricted to 500 permutations due to computational limitations. False-discovery rates (FDR) were calculated from permutation-based P values using the q-VALUE software implemented in the R statistical package (Story & Tibshirani 2003). Normalized mRNA data (ΔC_T values) from the QRT–PCR experiment were analysed using a mixed-model ANOVA with Proc Mixed in SAS (Little *et al*. 1996). For QRT–PCR analyses, we fit models incorporating biological replicates and one-step reaction replication as random effects and genetic line as a fixed effect. In this case, the 'biological replicate' term is comprised of both true environmental variation in transcript abundance among plants as well as variation generated in the RNA extraction process.

Results

Microarray analyses

The three F-statistics were computed for each gene and their null distributions were generated with permutations. At a nominal permutation significance level of 0.05, the F_1, F_3, and F_S tests detected 84, 65, and 85 genes, respectively, for genes occurring within the Cvi introgression elements. The volcano plot of these F-tests is shown in Fig. 1. The distribution of F_S-test P-values for these genes is presented in Fig. 2 and exhibits a relatively uniform distribution with a small peak of low P-values. We used an FDR method to obtain a gene list controlling for multiple tests. In this case, we detected 25 differentially expressed genes at an FDR of 0.20 (Table 1). We found no genes differentially expressed at a standard Bonferroni cut-off (corrected $\alpha = 0.05/789 = 0.00006$).

Color image

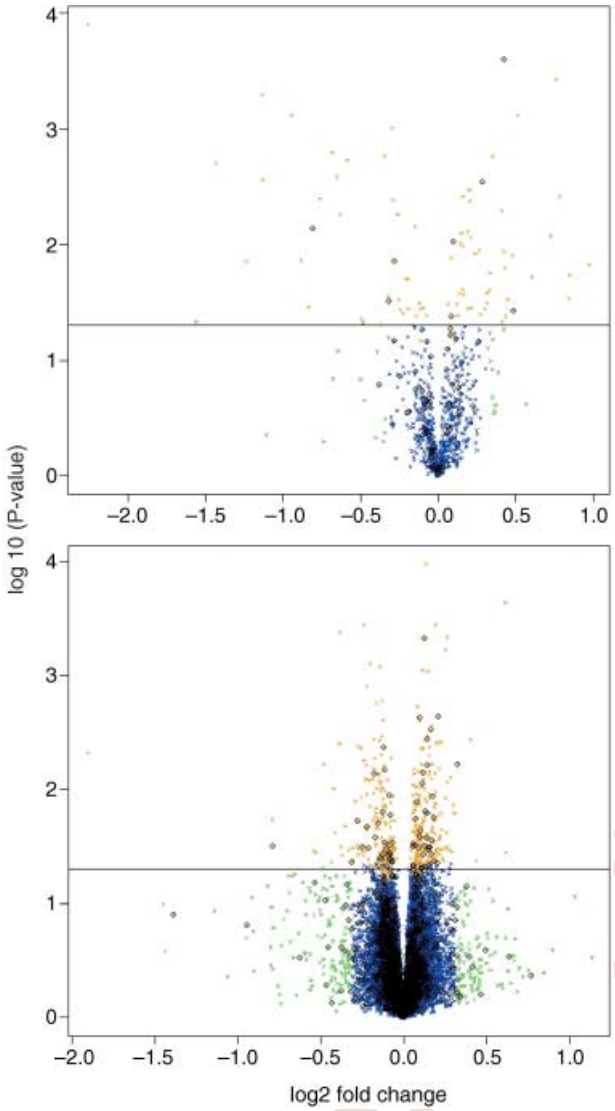

Fig. 1 Volcano plots representing the results from *F*-tests for introgression (top) and remaining genome (bottom) region analyses. Each scatterplot is the –log10-transformed *P*-values (nominal permutation) from the gene specific *F*-tests vs. the log2 fold change. The horizontal line represents the significance threshold of the F_1 test. The green dots are the genes selected by the F_3 test and the orange dots by the F_S test. Flagged genes are indicated by a black circle.

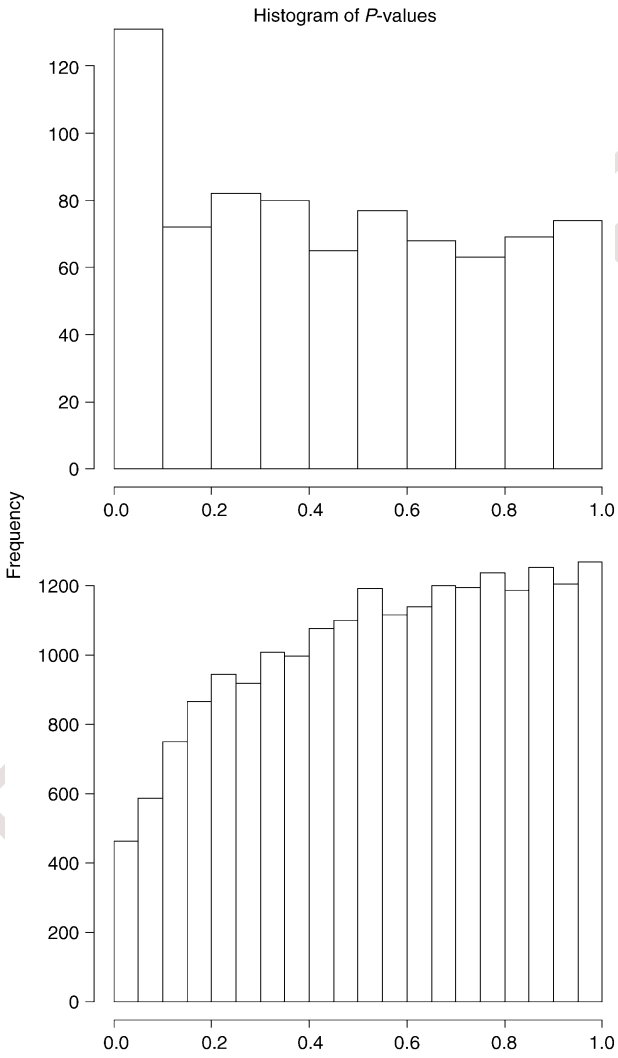

Fig. 2 Nominal permutation *P*-value distributions for introgression element genes (top panel) and genes from the remaining genomic locations (bottom). A uniform *P*-value distribution is expected under the null hypothesis of no differentially expressed genes. The peak of low *P*-values in the introgression analysis is indicated of a number of differentially expressed genes.

For genes occurring outside of the introgression, a nominal permutation significance level of $P < 0.05$ identified 454, 88, and 464 genes for the F_1, F_3, and F_S tests, respectively. The volcano plot of these *F*-tests is shown in Fig. 1. The distribution of *P*-values for these genes is also presented in Fig. 2. In this case, the *P*-value distribution is skewed with an excess of high *P*-values and a deficiency of low *P*-values. This is an unexpected *P*-value distribution and FDR methods for controlling multiple tests will perform poorly in this case (Yang 2004). We found no significant differentially expressed genes from this panel when using any reasonable FDR threshold or a simple Bonferroni correction method (corrected $\alpha = 0.05/20\,715 = 0.000002$). Nonetheless, we present a list of genes occurring outside the introgression with 'suggestive' expression differences (Table 2).

We detected no consistent pattern of either 'up' or 'down' regulation of gene expression between the NIL or L*er*-2 (Fig. 3). In general, expression differences were larger for genes within (*cis*) rather than outside (*trans*) of the introgression element. The genomic location of the set of differentially expressed genes is presented in Fig. 4.

Table 1 The top differentially expressed genes deemed significant at the nominal permutation *P*-value < 0.05 for genes occurring within the introgression element. *Q*-values from false-discovery rate tests provide a direct means of assessing confidence in the gene list in the face of multiple testing

Gene	Description	*P*-value	*Q*-value
AT2G44200	expressed protein	0	0
AT2G43970	La domain-containing protein	0.00025	0.06832
AT2G44850	expressed protein	0.00038	0.06832
AT3G01720	expressed protein	0.00038	0.06832
AT2G41760	expressed protein	0.00063	0.07591
AT2G42500	protein phosphatase 2A (PP2A)	0.00063	0.07591
AT2G44530	putative phosphoribosyl pyrophosphate synthetase	0.00114	0.11685
AT2G45830	expressed protein	0.0015	0.13007
AT2G46530	transcriptional factor B3 family protein/auxin-responsive factor AUX/IAA-related	0.00175	0.13007
AT2G43240	nucleotide-sugar transporter family protein	0.00199	0.13007
AT2G46440	cyclic nucleotide-regulated ion channel, putative	0.00199	0.13007
AT3G02180	expressed protein, twisted leaves mutant	0.00225	0.1348
AT2G43950	expressed protein	0.00275	0.14101
AT2G41710	ovule development protein, putative, similar to ovule development protein AINTEGUMENTA	0.00301	0.14101
AT2G44230	expressed protein	0.00314	0.14101
AT3G02740	putative aspartyl protease	0.00314	0.14101
AT3G01430	hypothetical protein	0.00389	0.14846
AT2G42270	U5 small nuclear ribonucleoprotein helicase, putative	0.00402	0.14846
AT2G42590	14-3-3 protein GF14 mu (grf9)	0.00413	0.14846
AT2G46180	intracellular protein transport protein USO1-related, similar to Intracellular protein transport protein	0.00413	0.14846
AT2G44100	GDP dissociation inhibitor	0.0049	0.16781
AT2G43130	Ras-related GTP-binding protein (ARA-4)	0.00604	0.1916
AT2G43920	thiol methyltransferase, putative	0.00629	0.1916
AT3G02470	S-adenosylmethionine decarboxylase	0.00654	0.1916
AT2G43850	putative protein kinase	0.00667	0.1916
AT2G45370	hypothetical protein	0.00753	0.20315
AT2G44065	ribosomal protein L2 family protein	0.00766	0.20315
AT2G44680	casein kinase II beta chain	0.00792	0.20315
AT2G47950	expressed protein	0.00864	0.21407
AT2G44650	chloroplast chaperonin 10	0.00913	0.21868
AT2G42210	mitochondrial import inner membrane translocase subunit	0.00952	0.22059
AT2G43210	UBX domain-containing protein	0.01099	0.2349
AT2G47960	expressed protein	0.01099	0.2349
AT2G42670	expressed protein	0.01111	0.2349
AT2G45300	5-enolpyruvylshikimate-3-phosphate (EPSP) synthase	0.01209	0.24827
AT2G41290	strictosidine synthase-related	0.01257	0.24925
AT3G02780	isopentenyl diphosphate:dimethylallyl diphosphate isomerase (IPP2)	0.01283	0.24925
AT2G45620	expressed protein	0.01461	0.26674
AT2G42320	unknown protein	0.01497	0.26674
AT2G46430	cyclic nucleotide-regulated ion channel (CNGC3)	0.01497	0.26674
AT3G02220	unknown protein	0.01522	0.26674
AT3G02790	zinc finger (C2H2 type) family protein	0.01601	0.27384
AT2G44050	6,7-dimethyl-8-ribityllumazine synthase precursor	0.01686	0.28183
AT2G43370	U1 small nuclear ribonucleoprotein 70 kDa, putative	0.01784	0.29138
AT3G02410	hypothetical protein, weak similarity to kynurenine formamidase	0.01941	0.3051
AT2G43900	putative inositol polyphosphate 5-phosphatase	0.01953	0.3051
AT2G46110	3-methyl-2-oxobutanoate hydroxy-methyl-transferase	0.02156	0.32965
AT2G45930	expressed protein	0.02243	0.33574
AT3G01850	putative D-ribulose-5-phosphate 3-epimerase	0.02346	0.34408
AT3G02250	putative auxin-independent growth promoter	0.02409	0.34624
AT2G42890	Meiosis protein mei2, putative	0.02833	0.39632
AT2G45530	expressed protein	0.02868	0.39632
AT3G02710	nuclear associated protein-related/NAP-related	0.03022	0.40142

Table 1 *Continued*

Gene	Description	*P*-value	*Q*-value
AT3G02400	hypothetical protein	0.03033	0.40142
AT2G44210	expressed protein	0.03072	0.40142
AT2G45120	putative C2H2-type zinc finger protein	0.03135	0.40224
AT2G44450	glycosyl hydrolase family 1	0.03243	0.40283
AT3G01500	carbonic anhydrase, chloroplast precursor	0.03307	0.40283
AT2G46410	myb-related protein CAPRICE (CPC)	0.03393	0.40283
AT3G02540	ubiquitin family protein	0.03457	0.40283
AT3G02070	OTU-like cysteine protease family protein	0.03516	0.40283
AT3G02460	putative plant adhesion molecule	0.03555	0.40283
AT2G44820	hypothetical protein	0.03591	0.40283
AT2G41900	putative CCCH-type zinc finger protein	0.03616	0.40283
AT3G01190	peroxidase 27 (PER27) (P27)	0.03745	0.40283
AT3G02090	putative mitochondrial processing peptidase	0.03794	0.40283
AT2G46650	putative cytochrome b5	0.03855	0.40283
AT2G44260	expressed protein	0.03868	0.40283
AT2G48090	expressed protein	0.03868	0.40283
AT2G43700	lectin protein kinase family protein	0.04057	0.41319
AT3G02300	regulator of chromosome condensation (RCC1) protein	0.04082	0.41319
AT3G01090	Snf1-related protein kinase KIN10 (AKIN10)	0.0427	0.42023
AT2G46330	arabinogalactan-protein (AGP16)	0.04297	0.42023
AT2G41620	nucleoporin interacting component family protein	0.04347	0.42023
AT2G45510	cytochrome p450, putative	0.04386	0.42023
AT2G43780	hypothetical protein	0.04559	0.42146
AT3G03150	expressed protein	0.0461	0.42146
AT3G01330	transcription factor, putative/E2F-like repressor E2L2	0.04672	0.42146
AT3G01460	PHD finger family protein/methyl-CpG binding domain-containing protein	0.04672	0.42146
AT3G02340	RING zinc-finger protein, putative	0.04725	0.42146
AT2G45960	plasma membrane intrinsic protein 1B (PIP1B)/aquaporin PIP1.2 (PIP1.2)	0.04815	0.42146
AT2G45290	putative transketolase precursor	0.04852	0.42146
AT2G43340	expressed protein	0.04901	0.42146
AT3G01920	yrdC family protein	0.04927	0.42146

QRT–PCR confirmation

We used TaqMan-based QRT–PCR to confirm the differential expression of a small set of candidate genes. In each case, 15 replicate plants and three replicate one-step reactions per RNA were analysed to yield highly accurate estimates of transcript abundance. Overall, four of five candidate genes were confirmed as differentially expressed (Table 3). Interestingly, the false positive (AT3G01460) exhibited unusually strong dye–gene interaction in the microarray experiment. These complexities may have led to poor estimation of its expression pattern in the array analysis (Kerr & Churchill 2001). We found little among-reaction but considerable among-biological replicate differences in transcript abundance. On average, replicate plants explained approximately 63% of the total random variance in transcript abundance. In contrast, replicate reactions on average explained less than 10% of the random variance. The TaqMan assay allowed highly reliable estimation of transcript abundance given adequate biological replication (Fig. 5). Expression differences ranged from approximately three- to fivefold (Fig. 6) between L*er*-2 and Cvi. In each case, transcript abundances were higher in L*er*-2 and in the expected direction as determined through microarray analysis. We found that the expression differences estimated through QRT–PCR were generally larger (~38%) than those estimated directly from our microarray studies.

Probe sequencing

We sequenced amplicons covering 24 array probe locations on chromosome 3 in Cvi and L*er*-2 to explore the potential of probe/target mismatches. Overall, we found 0.42%, 1.11%, and 1.29% sequence divergence between Col-4 and L*er*-2, Col-4 and Cvi, and Cvi and L*er*-2, respectively (estimated through 11 934 bases sequenced). Table 4 lists the number of mismatches detected between Cvi or L*er*-2 sequences and the microarray probe. No probe polymorphisms were detected in 13 of 24 sequenced probes. In nine probes, only a small number of SNP mismatches or single base pair deletions were detected. A major lesion involving a large (53 bp) insertion was detected in the L*er*-2 sequence of the AT3G01720 probe. This deletion likely results in differential hybridization of target to array probe.

Table 2 The top differentially expressed genes significant at the nominal permutation *P*-value < 0.005 for genes occurring outside of the introgression element. These genes did not meet a significance threshold associated with control for multiple testing. As such, these genes should only be considered suggestive hypotheses until further confirmation

Gene	Descriptions	*P*-value
AT3G26720	glycosyl hydrolase family 38 protein	0.00011
AT1G70850	Bet v I allergen family protein, similar to Csf-2 (Cucumis sativus)	0.00024
AT1G66860	expressed protein, similar to Hypothetical protein RP404	0.00028
AT1G07530	scarecrow-like transcription factor 14 (SCL14)	0.00031
AT4G04630	expressed protein	0.00034
AT3G51600	lipid transfer protein 5 mRNA, complete cds	0.00049
AT5G05550	expressed protein	0.00050
AT1G77060	mutase family protein, similar to carboxyvinyl-carboxyphosphonate phosphorylmutase GB:O49290 from (Arabidopsis thaliana)	0.00053
AT2G30210	laccase, putative/diphenol oxidase	0.00058
AT4G20020	expressed protein	0.00077
AT3G58030	zinc finger (C3HC4-type RING finger) family protein	0.00092
AT1G08040	expressed protein	0.00121
AT5G17200	glycoside hydrolase family 28 protein/polygalacturonase (pectinase) family protein	0.00140
AT4G00860	putative pathogenesis-related protein whose transcript level is induced in response to ozone and pathogenic Pseudomonas	0.00147
AT5G67590	Mutant leaves have a reduced capacity for cold acclimation, appear water-soaked, leak electrolytes, and accumulate reactive oxygen species constitutively	0.00159
AT5G62090	expressed protein	0.00171
AT4G10100	molybdopterin synthase small subunit (cnx7) mRNA, complete	0.00227
AT1G55030	F-box family protein	0.00244
AT1G66590	cox19 family protein	0.00245
AT1G12050	fumarylacetoacetase, putative	0.00253
AT5G57270	expressed protein	0.00265
AT3G52030	F-box family protein/WD-40 repeat family protein	0.00307
AT4G22990	SPX (SYG1/Pho81/XPR1) domain-containing protein	0.00311
AT5G15920	structural maintenance of chromosomes (SMC) family protein	0.00317
AT1G05890	zinc finger protein-related	0.00325
AT5G55200	co-chaperone grpE protein	0.00361
AT1G77145	expressed protein	0.00370
AT1G34770	MAGE-8 antigen-related	0.00377
AT1G64690	expressed protein	0.00382
AT5G56170	expressed protein, contains similarity to GPI-anchored protein	0.00394
AT3G04090	major intrinsic family protein/MIP family protein	0.00395
AT1G13450	DNA binding protein GT-1	0.00417
AT1G70880	Bet v I allergen family protein	0.00417
AT1G68500	expressed protein	0.00426
AT2G31440	expressed protein, identical to cDNA endonuclease III homologue (nth1 gene)	0.00441
AT3G49840	proline-rich family protein	0.00444
AT3G55770	LIM domain-containing protein	0.00457
AT4G28760	expressed protein	0.00465
AT5G53930	expressed protein	0.00482
AT1G69545	leucine-rich repeat family protein	0.00484
AT2G26070	expressed protein	0.00484
AT1G22410	2-dehydro-3-deoxyphosphoheptonate aldolase, putative/3-deoxy-D-arabino-heptulosonate 7-phosphate synthase, putative/DAHP synthetase, putative	0.00491

Discussion

We combined whole-genome gene expression analyses with near-isogenic line mapping to further our studies of two physiological QTL in *Arabidopsis thaliana*. We compared mRNA abundance of a NIL and its recurrent parent at over 21 500 genes with an ink-jet printed platform. We discovered 25 differentially expressed genes within the introgressions at a false-discovery rate (FDR) cut-off of 0.20 and identified new candidate genes for both QTL regions. Under a relatively strict FDR threshold, 3.2% of the introgression genes were differentially expressed between Cvi and L*er*-2. Extrapolating to the entire genome, we anticipate at least 816 'genetically' based differentially

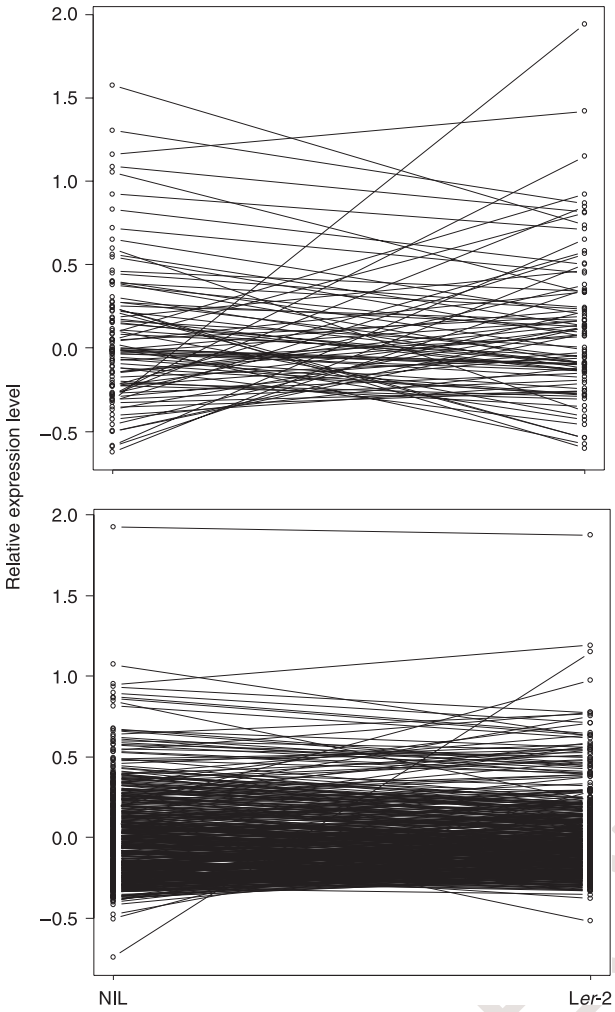

Fig. 3 Relative expression plots for genes significant at the nominal permutation threshold of $P < 0.05$ for the introgression element (top panel) and remaining genomic locations (bottom panel) for the microarray study.

expressed genes between these two ecotypes. This is surely an underestimate given the relatively low power of our microarray experiment and the limited scope of our survey (e.g. single tissue type, controlled environmental conditions, two ecotype accessions). These results suggest considerable standing genetic variation in gene expression in *A. thaliana*.

Microarray resources are exciting new tools for model systems and functional genomics studies. Most array resources have been developed from the sequences of single 'wild-type' or reference strains. However, researchers are generally interested in utilizing these resources more broadly in studies incorporating diverse genetic backgrounds (e.g. natural genetic variability, mutagenized lines, alternative wild-type strains) or in some cases across ecotype or species boundaries (Jackson *et al*. 2002; Borevitz & Nordborg 2003). Unfortunately, it may be difficult to distinguish between true gene expression differences and hybridization polymorphisms when applying genetically diverse samples to the existing model system arrays (Borevitz & Nordborg 2003; Ronald *et al*. 2005). We used the Agilent *Arabidopsis*-2 array developed from Col-4 sequence and hybridized L*er*-2 and Cvi cRNA to the array probes. We sequenced 24 probes to explore the impact of sequence diversity on hybridization characteristics and detected very few polymorphisms within the sampled probes. Detailed experimental validation studies by Hughes *et al*. (2001) have shown this level of heterogeneity to be inconsequential for target/probe hybridization dynamics in 60-mer inkjet oligo arrays. On the other hand, we did detect one major insertion polymorphism (53-bp insertion) (AT3G01720) that likely resulted in a false detection in our microarray study. Borevitz *et al*. (2003) found that ~4% of Affymetrix GeneChip features exhibited detectable sequence polymorphisms and that ~1–2% of genes were deleted between two *Arabidopsis* accessions (L*er* and Col). These experiments focused on the hybridization of ecotype genomic DNA to RNA expression chip 25-mer probes to specifically detect hybridization polymorphisms. Importantly, the detection of a 'single feature' polymorphism does not necessarily translate into substantially altered hybridization dynamics. This will depend on the degree of probe/target mismatch, hybridization dynamics, and the magnitude of true differential expression. In particular, gene expression false-positives due to hybridization polymorphisms may be less important when using long oligo probes. Given the known sequence divergence of *Arabidopsis* ecotypes (see Nordborg *et al*. 2005) 60-mer oligo arrays, when used with care, should be valuable for further studies of natural variation in gene expression.

A major challenge for all full genome analyses is the large number of expected false-positives resulting from extensive multiple testing (Rice 1989; Benjamini & Hochberg 1995; Story & Tibshirani 2003). We utilized an array with ~21 500 elements and an expectation of ~1000 false-positives when setting $\alpha = 0.05$. One method to control for multiple-testing is the false-discovery rate (FDR). Instead of controlling the chance of *any* false positives (as in family wise error rate methods), the FDR controls the expected *proportion* of false positives. A FDR threshold is determined from the observed *P*-values and the expectation of a uniform distribution under the null hypothesis of no differential expression. One limitation of the general approach, however, is the assumption that *P*-values of null tests will be uniformly distributed in the range of [0, 1]. In particular, this expectation relies on the assumption that each test is independent (Benjamini & Hochberg 1995). The degree of expression dependence is, of course, unknown and likely to vary with organism, stage of development,

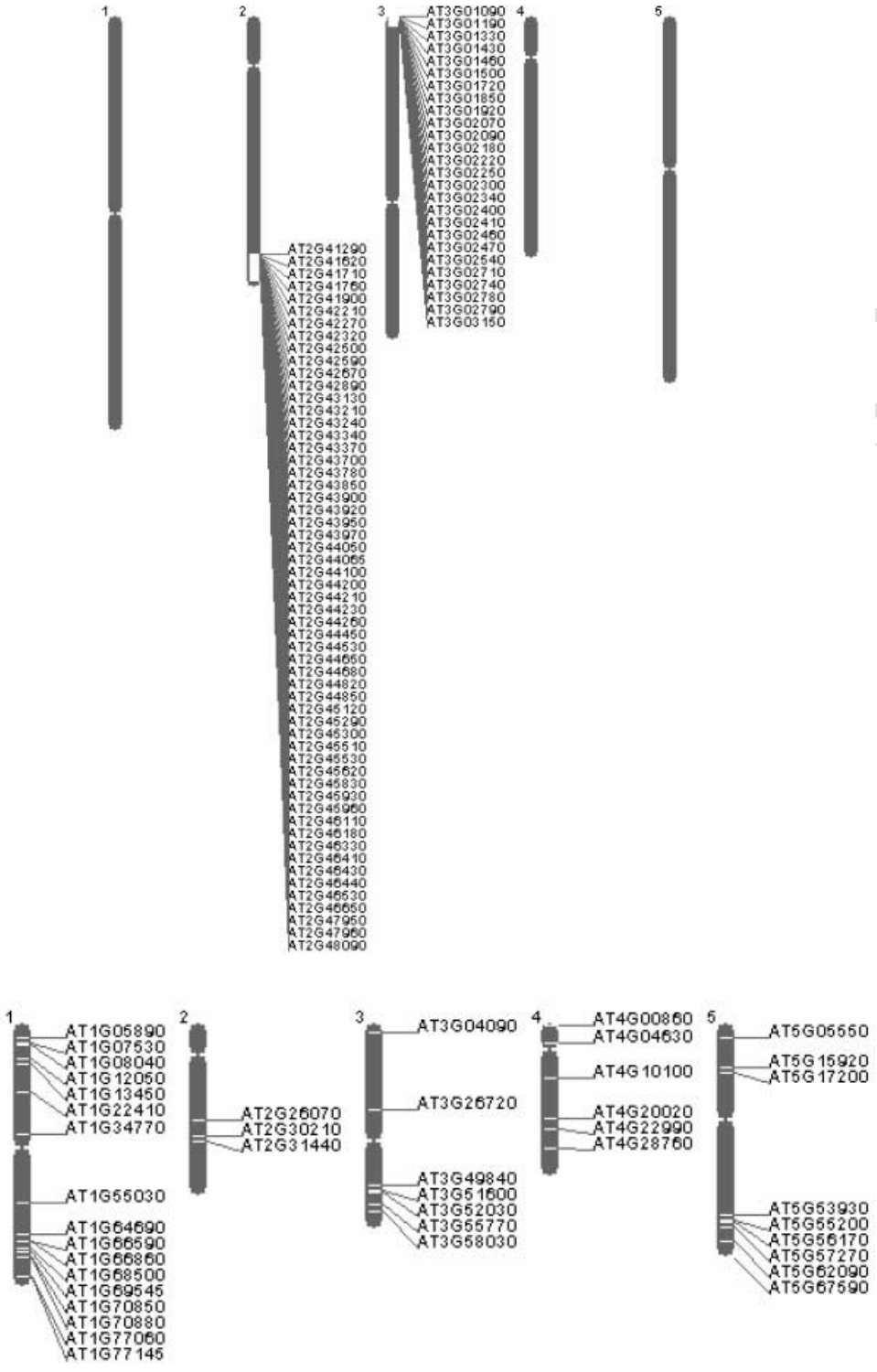

Fig. 4 The genomic location of the differentially expressed genes presented in Tables 1 and 2.

and tissue type sampled (e.g. Schmid *et al*. 2005). In particular, some dependence is expected if *trans* differences are regulated by one or few transcription factors within the introgressed region in our study.

The FDR approach is likely to be inappropriate in analyses with unusual *P*-value distributions. For example, in our experiment the *P*-value distribution for genes occurring outside of the QTL introgression is skewed with too

	Variance components				
Random effects	At3g02220	At3g01460	At3g01480	At2g42500	At2g42590
Plant (genetic line)	0.4565**	0.4304**	0.6419**	0.1381**	0.1291**
Reactions (plant)	0.0000^{NS}	0.0004^{NS}	0.0000^{NS}	0.06211*	0.0512*
Error	0.2266***	0.2180***	0.2425***	0.0603***	0.0476***
Fixed effect	*F*-value	*F*-value	*F*-value	*F*-value	*F*-value
Genetic line	74.38***	1.37^{NS}	29.88***	130.39***	140.07***

*corresponds to $P < 0.01$; **corresponds to $P < 0.005$; ***corresponds to $P < 0.0001$; NS, not significant.

Table 3 Mixed-model ANOVA testing for the effect of biological replicates nested within genetic line (L*er*-2 or NIL), replicate reactions nested within RNA extracted from each biological replicate, and genetic line for each of five tested mRNA

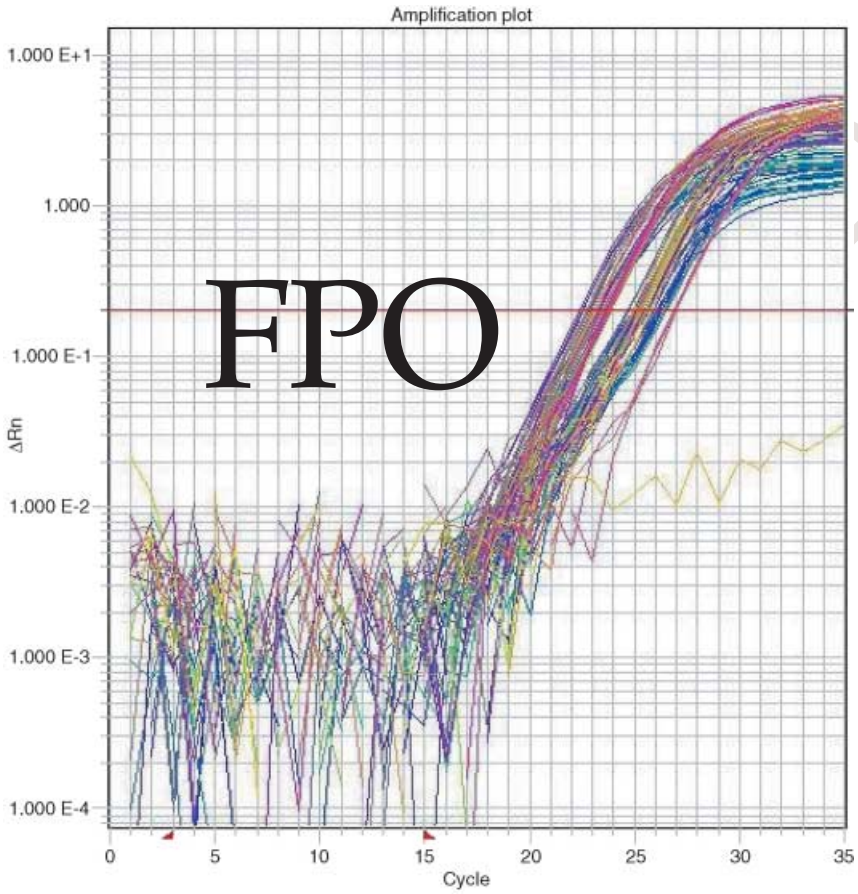

Fig. 5 Quantitative RT–PCR amplification plots for replicate L*er*-2 and NIL samples for AT3G02220 showing clear separation of C_T values for L*er*-2 and Cvi.

14

Color image

few low and too many high *P*-values (Fig. 3). A number of mechanisms could generate this distribution including dependence among the test statistics (correlated patterns of gene expression), an inadequate statistical model (hidden structure not modelled in the analysis), substantial cross hybridization on the array, or experimental errors (e.g. sample or array mislabelling). It is virtually impossible to distinguish whether the observed pattern is an inherent result of the biological data or an artefact of the experiment or analysis. Regardless, it is clear that the signal of differential expression is weak for genes occurring outside of the introgression element compared to those occurring within the QTL interval. Interestingly, other studies using NIL or congenic mapping and expression studies have also detected few differentially expressed genes outside of introgression elements (Aitman *et al*. 1999;

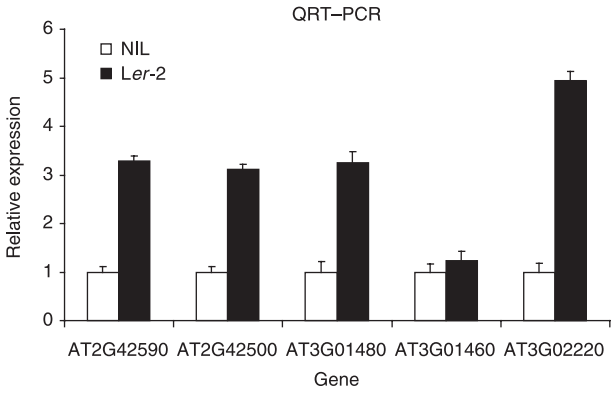

Fig. 6 $\Delta\Delta C_T$ values for gene expression of candidates identified through microarrays studies. Each bar represents the relative amount of mRNA for each gene relative to the line with lower transcript abundance ± 1 SE (N = 15, L*er*-2; N = 16, NIL).

Table 4 Results of probe sequencing and mismatches between L*er*-2, Cvi, and the Col-4 60-mer probe. Genes notated by * are identified in Table 1 as significantly differentially expressed as determined by a nominal P-value threshold of $P < 0.05$

	Mismatches from Col-4 probe	
Gene	L*er*-2	Cvi
AT3G01100	1 bp	1 bp
AT3G01220	1 bp	none
AT3G01330*	none	none
AT3G01430*	none	none
AT3G01440	none	none
AT3G01470	none	none
AT3G01480	none	2 bp
AT3G01500*	none	none
AT3G01720*	53-bp insertion	none
AT3G02070*	none	none
AT3G02090*	none	none
AT3G02180*	none	none
AT3G02220*	none	1 bp, 2 single base deletions
AT3G02250*	none	2 bp
AT3G02280	none	1 bp
AT3G02400*	none	none
AT3G02410*	none	2 bp
AT3G02460*	none	1 bp
AT3G02470*	none	none
AT3G02540*	none	no data
AT3G02660	none	none
AT3G02740*	none	none
AT3G02790*	no data	1 bp
AT3G03000	none	none

15 Eaves *et al*. 2000; Karp 2000; Baxter *et al*. 2005). Several QTL
16 studies exploring gene expression variability have often detected *cis*-acting QTL as transcription polymorphisms (Brem *et al*. 2002; Schadt *et al*. 2003). These data suggest *trans*-acting expression variability may be rarer in natural populations (although see Cavalieri *et al*. 2000; Morley *et al*. 2004; Wayne *et al*. 2004; Chesler *et al*. 2005; Yvert *et al*. 2005).

The QTL regions studied here affect patterns of carbon isotope discrimination (an outcome of varying stomatal regulation of photosynthesis), instantaneous stomatal conductance, whole-plant transpiration efficiency, and
17 patterns of water loss (Hausmann *et al*. 2005; Juenger *et al*. in press). These results suggest that genes underlying the observed physiological effects may play important roles in controlling stomate closure. Our gene expression experiments identify a number of candidate genes that may be involved in guard cell signal transduction or physiological processes related to gas exchange and carbon fixation. For example, a protein phosphatase (PP2A) catalytic subunit was confirmed as differentially expressed with QRT–PCR (~3.5-fold under-expressed in Cvi compared to L*er*-2). Several studies suggest PP2As act as negative regulators of ABA signalling and play an important role in the regulation of anion channels, guard cell signalling, and stomata
18 closure (Pei *et al*. 1997; Kwak *et al*. 2002). Similarly, QRT–
19 PCR confirmed differential expression of a 14-3-3 protein-signalling regulator. It has been shown that 14-3-3 proteins bind directly to and activate the plasma membrane ATPase pump which drives stomatal opening in blue light (Kinoshita & Shimazaki 1999). Other studies suggest a role of 14-3-3 proteins in regulating a diversity of ion channels and pumps (Boer 2002). Carbonic anhydrase (CA) exhibited an approximately twofold increase in expression compared to L*er*-2. CA is involved in the conversion of atmospheric CO_2 to HCO_3- and may play an important role in the internal conductance of CO_2 from the stomata interior to chloroplasts and sites of carbon fixation (Badger & Price 1994). As the study QTL impact the diffusion of CO_2 and water through stomata, it is possible that the induction of CA expression in the NIL is related to the observed QTL effects. Finally, we identified a number of differentially expressed metabolic enzymes (e.g. S-adenosylmethionine decarboxylase), transcription factors (e.g. transcriptional factor B3 family protein/auxin-responsive factor AUX/IAA-related), ion channel proteins (e.g. cyclic nucleotide-regulated ion channel), and unknown 'expressed proteins' (four of the top five differentially expressed genes). None of the FDR significant genes were found to be commonly regulated by water-deficit treatments or highly expressed in guard cells as determined in previous microarray experiments with *Arabidopsis* (Bray 2004; Leonhardt *et al*. 2004). Additional QRT–PCR confirmation and functional studies are needed to more fully understand the biological significance of these candidate genes and to explore additional loci from our gene lists.

20

In summary, we have used a microarray approach to detect expression polymorphisms associated with two introgressions sharing the common trait of increased whole-plant transpiration. These data provide a direct means for assessing the degree of natural genetic variation in gene expression as well as offer an independent means for identifying genes causally associated with the quantitative traits of interest. Several of the differentially expressed genes detected in our study are involved in physiological processes related to guard cell function and are therefore excellent candidates for future study. In addition, a small number of novel candidate genes lacking genome annotation were identified. These are exciting genes for additional study as they would have been unlikely candidates for follow-up in traditional fine-mapping efforts and yet may help determine the function of unknown genes. Together, genetic mapping and gene expression studies hold great promise for elucidating the molecular basis of ecologically important traits.

Acknowledgements

A special thanks to S. Sommerville, D. Galbraith, and S. Sen for their early support of T.J.'s interest in gene expression studies. The statistical analysis of these data was greatly improved by advice and suggestions from S. Sen, G. Churchill, H. Wu, K. Kerr, K. Shockley, and M. Lachmann — any errors remain our own. Initial plant material for NIL construction was generously provided by M. Koornneef. This work was supported by NSF DEB 042011 to T.J. and NSF DEB 0419969 to J.M.

References

Aitman TJ, Glazier AM, Wallace CA *et al.* (1999) Identification of Cd36 (Fat) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. *Nature Genetics*, **21**, 76–83.

Alonso-Blanco C, Peeters AJM, Koornneef M *et al.* (1998) Development of an AFLP based linkage map of L*er*, Col, and Cvi *Arabidopsis thaliana* ecotypes and construction of a L*er*/Cvi recombinant inbred line population. *The Plant Journal*, **14**, 259–271.

Barton NH, Keightley PD (2002) Understanding quantitative genetic variation. *Nature Reviews Genetics*, **3**, 11–21.

Baxter CJ, Sabar S, Quic WP *et al.* (2005) Comparison of changes in fruit gene expression in tomato introgression lines provides evidence of genome-wide transcriptional changes and reveals links to mapped QTLs and described traits. *Journal of Experimental Botany*, **56**, 1591–1604.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*, **57**, 289–300.

Boer AH (2002) Plant 14-3-3 proteins assist ion channels and pumps. *Biochemical Society Transactions*, **30**, 416–421.

Borevitz JO, Nordborg M (2003) The impact of genomics on the study of natural variation in *Arabidopsis*. *Plant Physiology*, **132**, 718–725.

Borevitz JO, Liang D, Plouffe D *et al.* (2003) Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Research*, **13**, 513–523.

Bray EA (2004) Gene commonly regulated by water-deficit stress in *Arabidopsis thaliana*. *Journal of Experimental Botany*, **55**, 2331–2341.

Brem RB, Kruglyak L (2005) The landscape of genetic complexity across 5700 gene expression traits in yeast. *Proceedings of the National Academy of Sciences, USA*, **102**, 1572–1577.

Brem B, Yvert G, Clinton R *et al.* (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science*, **296**, 752–755.

Brown PO, Botstein D (1999) Exploring the world of the genome with DNA microarrays. *Nature Genetics*, **21**, 33–37.

Cavalieri D, Townsend JP, Hartl DL (2000) Manifold anomalies in gene expression in a vineyard isolate of *Saccharomyces cervisiae* revealed by DNA microarray analysis. *Proceedings of the National Academy of Sciences, USA*, **97**, 12369–12374.

Chesler EJ, Lu L, Shou S *et al.* (2005) Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nature Genetics*, **37**, 233–242.

Churchill GA (2002) Fundamentals of experimental design for cDNA microarrays. *Nature Genetics*, **32** (4s), S490–S495.

Coughlan SJ, Agrawal V, Meyers B (2004) A comparison of global gene expression measurement technologies in *Arabidopsis thaliana*. *Comparative and Functional Genomics*, **5**, 245–252.

Cui X, Kerr MK, Churchill GA (2003) Transformations for cDNA microarray data. *Statistical Applications in Genetics and Molecular Biology*, **2**, 4.

Cui X, Hwang JT, Qiu J *et al.* (2005) Improved statistical tests for differential gene expression by shrinkage variance components. *Biostatistics*, **6**, 59–75.

Falconer DJ, Mackay TFC (1989) *Introduction to Quantitative Genetics*, 3rd edn. Longman Sci. and Tech., Harlow, UK

Fay JC, McCullough HL, Sniegowski PD *et al.* (2004) Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cervisiae*. *Genome Biology*, **5**, R26.

Flint J, Mott R (2001) Finding the molecular basis of quantitative traits: successes and pitfalls. *Nature Reviews Genetics*, **2**, 437–445.

Gibson G, Riley-Berger R, Harshmann L *et al.* (2004) Extensive sex-specific nonadditivity of gene expression in *Drosophila melanogaster*. *Genetics*, **167**, 1791–1799.

Hartl DL, Meiklejohn CD, Castillo-Davis CI *et al.* (2004) Expression profiling in evolutionary genetics. In: *The Evolution of Population Biology* (eds Singh R, Uyenoyama M). Cambridge University Press.

Hausmann N, Juenger T, Sen S *et al.* (2005) Quantitative trait loci affecting $\delta^{13}C$ and response to differential water availability in *Arabidopsis thaliana*. *Evolution*, **59**, 81–96.

Hughes RR, Mao M, Jones AR *et al.* (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesize. *Nature Biotechnology*, **19**, 342–347.

Hyeon-Se L, Want J, Tian L *et al.* (2004) Sensitivity of 70-mer oligonucleotides and cDNAs for microarray analysis of gene expression in *Arabidopsis* and its related species. *Plant Biotechnology Journal*, **2**, 45–57.

Jackson RB, Linder CR, Lynch M *et al.* (2002) Linking molecular insights and ecological research. *Trends in Ecology & Evolution*, **17**, 409–414.

Jin W, Riley R, Wolfinger RD *et al.* (2001) Contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nature Genetics*, **29**, 389–395.

Juenger T, Mckay JK, Hausmann N *et al.* (2005a) Identification and characterization of QTL underlying whole-plant physiology in

Arabidopsis thaliana: $\delta^{13}C$, stomatal conductance and transpiration efficiency. *Plant, Cell and Environment*, **28**, 697–708.

Juenger T, Sen S, Stowe K *et al.* (2005b) Epistasis and genotype-environment interaction for quantitative trait loci affecting flowering time in *Arabidopsis thaliana*. *Genetica*, **123**, 87–105.

Kerr MK, Churchill GA (2001) Statistical design and the analysis of gene expression microarray data. *Genetical Research*, **77**, 123–128.

Kerr MK, Martin M, Churchill GA (2000) ANOVA for gene expression microarray data. *Journal of Computational Biology*, **7**, 819.

Kinoshita T, Shimazaki K (1999) Blue light activates the plasma membrane by phosphorylation of the C-terminus in stomatal guard cells. *EMBO Journal*, **18**, 5548–5558.

Kwak JM, Moon J-H, Murata Y *et al.* (2002) Disruption of a guard cell-expressed protein phosphatase 2A regulatory subunit, RCN1, confers abscisic acid insensitivity in *Arabidopsis*. *The Plant Cell*, **14**, 2849–2861.

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. *Science*, **265**, 2037–2048.

Leonhardt N, Kwak JM, Robert N *et al.* (2004) Microarray expression analyses of *Arabidopsis* guard cells and isolation of a recessive abscisic acid hypersensitive protein phosphatase 2C mutant. *The Plant Cell*, **16**, 596–615.

Little RC, Milliken GA, Stroup WW *et al.* (1996) *SAS® System for Mixed Models*. SAS Institute Inc, Cary, North Carolina.

Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, Massachusetts.

Manley KF, Wang J, Willams RW (2005) Weighting by heritability for detection of quantitative trait loci with microarray estimates of gene expression. *Genome Biology*, **6**, R27.

Morley M, Molony CM, Weber R *et al.* (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature*, **430**, 743–747.

Nordborg M, Hu TT, Ishino Y *et al.* (2005) The pattern of polymorphism in *Arabidopsis thaliana*. *Public Library of Science Biology*, **3** (7), 196–205.

Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. *Nature Genetics*, **32**, 261–266.

Pei ZM, Kuchitsu K, Ward J *et al.* (1997) Differential abscissic acid regulation of guard cell slow anion channels in *Arabidopsis* wild-type and *abi1* and *abi2* mutants. *The Plant Cell*, **9**, 409–423.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

Ronald R, Akey JM, Whittle J (2005) Simultaneous genotyping, gene-expression measurement, and detection of allele-specific expression with oligonucleotide arrays. *Genome Research*, **15**, 284–291.

Schadt EE, Monks SA, Drake TA *et al.* (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature*, **422**, 297–302.

Schlichting CE, Pigliucci M (1998) *Phenotypic Evolution: A Reaction Norm Perspective*. Sinauer Associates, Sunderland, Massachusetts.

Schmid M, Davison TS, Henz SR *et al.* (2005) A gene expression map of *Arabidopsis thaliana* development. *Nature Genetics*, **37**, 501–506.

Schroeder JI, Allen GJ, Hugoouvieux V *et al.* (2001) Guard cell signal transduction. *Annual Review of Plant Physiology and Plant Molecular Biology*, **52**, 627–658.

Seki M, Ishida J, Narusaa M *et al.* (2002) Monitoring the expression pattern of around 7000 *Arabidopsis* genes under ABA treatments using a full-length cDNA microarray. *Functional and Integrative Genomics*, **2**, 282–291.

Story JD, Akey JM, Kruglyak L (2005) Multiple locus linkage analysis of genomewise expression in yeast. *Public Library of Science*, **3**: e267.

Story JD, Tibshirani R (2003) Statistical significance for genome-wide studies. *Proceedings of the National Academy of Science, USA*, **100**, 9440–9445.

Townsend JP, Cavalieri D, Hartl DL (2003) Population genetic variation in genome-wide gene expression in wine yeasts. *Molecular Biology and Evolution*, **20** (6), 955–963.

van Driessche NV, Demsar J, Booth EO *et al.* (2005) Epistasis analysis and global transcriptional phenotypes. *Nature Genetics*, **37**, 471–477.

Wayne ML, Mcintyre LM (2002) Candidate gene discovery: combining mapping and microarrays in the genetic analysis of an adaptive trait in *Drosophila*. *Proceedings of the National Academy of Science, USA*, **99**, 14903–14906.

Wayne ML, Pan Y-J, Nuxhdin SV *et al.* (2004) Additivity and trans-acting effects on gene expression in male *Drosophila simulans*. *Genetics*, **168**, 1413–1420.

Wolfinger RD, Gibson G, Wolfinger R *et al.* (2001) Assessing gene significance from cDNA microarray expression data via mixed models. *Journal of Computational Biology*, **8**, 625–637.

Woo Y, Affourtit J, Daigle S *et al.* (2004) A comparison of cDNA, Oligonucleotide, and Affymetrix GeneChip gene expression microarray platforms. *Journal of Biomolecular Techniques*, **15**, 276–284.

Wu H, Kerr K, Churchill GA (2003) MAANOVA: A software package for the analysis of spotted cDNA microarray experiments. In: *The Analysis of Gene Expression Data: An Overview of Methods and Software* (eds Parmigiani G, Garrett ES, Irizarry RA, Zeger SL), pp. 313–431. Springer, New York.

Yang X (2004) Q-value methods may not always control false discovery rate in genomic applications. *Proceedings of the 2004 IEEE Computational Systems Bioinformatics Conference*.

Yvert G, Brem RB, Whittle J *et al.* (2005) Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nature Genetics*, **35**, 57–64.

Author Query Form

Journal: Molecular Ecology

Article: mec_2774.fm

Dear Author,

During the copy-editing of your paper, the following queries arose. Please respond to these by marking up your proofs with the necessary changes/additions. Please write your answers on the query sheet if there is insufficient space on the page proofs. Please write clearly and follow the conventions shown on the attached corrections sheet. If returning the proof by fax do not write too close to the paper's edge. Please remember that illegible mark-ups may delay publication.

Many thanks for your assistance.

No.	Query	Remarks
1	Falconer & Mackay 1996 has been changed to Falconer & Mackay 1989 so that this citation matches the one in the list.	
2	Gibson & Mackay 2002 has not been found in the list.	
3	Eves et al. 2002 has not been included in the list	
4	Townsend 2003 has been changed to Townsend et al. 2003 (here and in the rest of the paper) so that this citation matches the one in the list.	
5	Altman et al. 1999 has been changed to Aitman *et al*. 1999 so that this citation matches the one in the list.	
6	Karp et al. 2000 has not been included in the list	
7	Brem 2002 has been changed to Brem et al. 2002 so that this citation matches the one in the list.	
8	Morely et al. 2004 has been changed to Morley et al. 2004 so that this citation matches the one in the list.	
9	Storey et al. 2005 has been changed to Story et al. 2005 so that this citation matches the one in the list.	
10	Alonso-Blanco et al. 1989 has been changed to Alonso-Blanco et al. 1998 so that this citation matches the one in the list. Further, 'b' from 1998b has been deleted as there is only one Alonso-Blanco et al. 1998 in the reference list.	
11	Littell et al. 1996 has been changed to Little et al. 1996 so that this citation matches the one in the list.	

No.	Query	Remarks
12	Storey & Tibshirani 2003 has been changed to Story & Tibshirani 2003 so that this citation matches the one in the list.	
13	Figure 4 is of low resolution. Please supply another one with a higher resolution. Further, for more information about supplying electronic artwork, please see the journal webpage or Blackwell's artwork electronic guidelines at http://www.blackwellpublishing.com/authors/digill.asp	
14	Figure 5 is of low resolution. Please supply another one with a higher resolution. Further, for more information about supplying electronic artwork, please see our the journal webpage or our electronic artwork guidelines at http://www.blackwellpublishing.com/authors/digill.asp.	
15	Eaves et al. 2000 has not been included in the list	
16	Karp 2000 has not been included in the list	
17	Please confirm if Juenger et al. in press here refers to Juenger, Mckay, Hausmann, et al. 2005(a) or to Juenger, Sen, Stow, et al. 2005(b).	
18	Pei et al. 1977 has been changed to Pei et al. 1997 so that this citation matches the one in the list.	
19	Please confirm if rewording of this sentence (Original: 14-3-3 proteins have been shown ... /Changed to: It has been shown that 14-3-3 proteins ...) is OK to avoid starting the sentence with an arabic numeral.	
20	Badger & Price 1994 has not been included in the list	
21	2005 has been added to reference entry Chesler et al. in accordence with the one found in the text. Please confirm if this is correct.	
22	Cui et al. 2003 has not been found in the text. Further, please check expanded journal title.	
23	Please supply complete name of publisher for Falconer & Mackay 1989 (Longman Sci. and Tech.).	
24	Hartl et al. 2004 has not been found in the text.	
25	Hyeon-Se et al. 2004 has not been found in the text.	
26	'a' and 'b' have been added to the two Juenger et al. 2005 in the reference list. Please assign 'a' or 'b' to the citations for Juenger et al. 2005 made in the text.	

No.	Query	Remarks
27	Oleksa has been changed to Oleksiak so that this list matches the one in the text.	
28	Schroeder et al. 2001 has not been found in the text.	
29	Seki et al. 2002 has not been found in the text.	
30	'van' has been added to the surname of first author Driessche in Drissche et al. so that this citation matches the one in the text. Further, publication year, 2005, has been added. Please confirm if this is correct.	
31	Woo et al. 2004 has not been found in the text.	
32	Wu & Kerr 2003 has not been found in the text.	
33	Please supply a paragraph about the reserach interests of the authors.	

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	*Textual mark*	*Marginal mark*
Leave unchanged	··· under matter to remain	Stet
Insert in text the matter indicated in the margin		New matter followed by
Delete	through matter to be deleted	
Delete and close up	through matter to be deleted	
Substitute character or substitute part of one or more word(s)	/ through letter or through word	New letter or new word
Change to italics	— under matter to be changed	
Change to capitals	under matter to be changed	
Change to small capitals	under matter to be changed	
Change to bold type	under matter to be changed	
Change to bold italic	under matter to be changed	
Change to lower case	Encircle matter to be changed	
Change italic to upright type	(As above)	
Insert 'superior' character	/ through character or where required	under character e.g.
Insert 'inferior' character	(As above)	over character e.g.
Insert full stop	(As above)	
Insert comma	(As above)	
Insert single quotation marks	(As above)	and/or
Insert double quotation marks	(As above)	and/or
Insert hyphen	(As above)	
Start new paragraph		
No new paragraph		
Transpose		
Close up	linking letters	
Insert space between letters	between letters affected	
Insert space between words	between words affected	
Reduce space between letters	between letters affected	
Reduce space between words	between words affected	