Rapid report

Evidence of neutral transcriptome evolution in plants

Author for correspondence:
Martin Broadley
Tel: +44 115 9516382
Fax: +44 115 9516334
Email: martin.broadley@nottingham.ac.uk

Received: *26 June 2008*
Accepted: *29 July 2008*

M. R. Broadley[1]*, P. J. White[2]*, J. P. Hammond[3]*, N. S. Graham[1]*, H. C. Bowen[3], Z. F. Emmerson[1], R. G. Fray[1], P. P. M. Iannetta[2], J. W. McNicol[2] and S. T. May[1]

[1]School of Biosciences, University of Nottingham, Sutton Bonington, Loughborough LE12 5RD, UK; [2]SCRI, Invergowrie, Dundee DD2 5DA, UK; [3]University of Warwick, Wellesbourne, Warwick CV35 9EF, UK

Key words: Affymetrix, *Alyssum*, *Arabidopsis halleri*, *Arabidopsis lyrata*, *Brassica*, *Capsella*, hyperaccumulator, *Thlaspi*.

Summary

• The transcriptome of an organism is its set of gene transcripts (mRNAs) at a defined spatial and temporal locus. Because gene expression is affected markedly by environmental and developmental perturbations, it is widely assumed that transcriptome divergence among taxa represents adaptive phenotypic selection. This assumption has been challenged by neutral theories which propose that stochastic processes drive transcriptome evolution.

• To test for evidence of neutral transcriptome evolution in plants, we quantified 18 494 gene transcripts in nonsenescent leaves of 14 taxa of Brassicaceae using robust cross-species transcriptomics which includes a two-step physical and *in silico*-based normalization procedure based on DNA similarity among taxa.

• Transcriptome divergence correlates positively with evolutionary distance between taxa and with variation in gene expression among samples. Results are similar for pseudogenes and chloroplast genes evolving at different rates. Remarkably, variation in transcript abundance among root-cell samples correlates positively with transcriptome divergence among root tissues and among taxa.

• Because neutral processes affect transcriptome evolution in plants, many differences in gene expression among or within taxa may be nonfunctional, reflecting ancestral plasticity and founder effects. Appropriate null models are required when comparing transcriptomes in space and time.

New Phytologist (2008) **doi**: 10.1111/j.1469-8137.2008.02640.x

Introduction

The transcriptome is the set of transcripts (mRNAs) of an organism at a defined spatial and temporal locus. The abundance of thousands of transcripts varies markedly in response to environmental and developmental perturbations, affecting protein translation and activity, and thus organism phenotypes. Consequently, it is widely assumed that variation in the abundance of transcripts among individuals within a population will ultimately lead to divergence in transcript abundance among populations and taxa through adaptive

*These authors contributed equally to this work.

phenotypic selection. Pääbo and colleagues recently proposed a neutral theory of transcriptome evolution which challenges this belief (Khaitovich *et al.*, 2004, 2005, 2006). In their theory, divergence in transcript abundance among taxa is driven not by adaptive selection, but by stochastic processes in which, by analogy to neutral genome evolution (Kimura, 1968; Gould, 2002), most differences in transcript abundance among individuals are likely to be selectively neutral or nearly neutral. Thus, within stabilizing constraints, variation in the expression of a transcript among individuals will drive expression divergence among populations and taxa as a consequence of drift. Empirical evidence of neutral transcriptome evolution is strongest among primate taxa of small population size (Khaitovich *et al.*, 2004, 2005, 2006). In these studies, transcriptome divergence among taxa accumulates monotonically over time, correlating positively with evolutionary distance among taxa and with variation in gene transcript abundance among individuals. However, a neutralist interpretation of these data has been challenged (Gilad *et al.*, 2006), based on inherent uncertainties in quantifying and normalizing the transcriptomes of taxa whose genome sequences are polymorphic, and because of sampling constraints.

The aim of this study was to determine if there is evidence of neutral transcriptome evolution among plants. We chose to study taxa of the Brassicaceae family, which has a basal age of *c.* 40 Myr (Bailey *et al.*, 2006). Despite rapid speciation in several clades and a complex genome, which has undergone several duplications, the Brassicaceae phylogeny is well resolved based on chloroplast gene and internal transcribed spacer (ITS) sequences (Koch *et al.*, 2005; Bailey *et al.*, 2006; Beilstein *et al.*, 2006; Schranz *et al.*, 2007). Furthermore, Brassicaceae genome sizes are small among angiosperms (Johnston *et al.*, 2005) and several are currently being fully sequenced: those of *Arabidopsis lyrata*, *Capsella rubella*, *Brassica rapa* and *Eutrema salsuginea* (previously *Thellungiella halophila*). These efforts are hastening our understanding of complex genome evolution, making the Brassicaceae a tractable model family for evolutionary studies (Schranz *et al.*, 2007). However, we chose this family specifically because it contains *Arabidopsis thaliana*, for which there are > 4500 transcriptome data sets obtained on a single array type (ATH1-121501 GeneChip, ATH1; Affymetrix, Santa Clara, CA, USA), curated in the public domain (Craigon *et al.*, 2004). This unique public resource supports the meta-analyses of transcriptomes from tissue-specific, developmental, mutant vs wild type, and environmental-response studies (Schmid *et al.*, 2005; Brady *et al.*, 2007). Furthermore, ATH1 – which comprises 11 perfect-match (PM)/mismatch (MM) pairs of short (25-mer) oligonucleotide probes per probe-set – has proven an effective platform for robust cross-taxa transcriptome analyses, using genomic DNA-based masking strategies and normalizations to remove the effects of DNA polymorphisms among species on RNA expression estimates (Hammond *et al.*, 2005, 2006; Graham *et al.*, 2007). These techniques have been applied to the study of gene regulation in *Brassica* crops of economic consequence (Hammond *et al.*, 2005) and in extremophile species adapted to metalliferous and saline habitats (Broadley *et al.*, 2007). In this study, we use robust cross-taxa transcriptomics and meta-analyses to show that neutral processes profoundly affect transcriptome divergence among plant taxa.

Materials and Methods

Seeds were obtained of *Arabidopsis thaliana* (L.) Heynh. (Columbia (Col-0) and Landsberg *erecta* (L*er*-0); donor: Nottingham Arabidopsis Stock Centre, Nottingham, UK), *Arabidopsis lyrata* ssp. *petraea* (L.) O'Kane & Al-Shehbaz and *Arabidopsis halleri* (L.) O'Kane & Al-Shehbaz (donor: Mark Macnair, University of Exeter, Exeter, UK; Filatov *et al.*, 2006), *Capsella bursa-pastoris* (L.) Medik. (720-81 and 798 accessions) and *Capsella rubella* Reut. (donor: Michael Lenhard, John Innes Centre, Norwich, UK), *Alyssum murale* Waldst. & Kit. and *Alyssum lesbiacum* (Candargy) Rech.f. (donor: Alan Baker, University of Melbourne, Melbourne, Australia), *Brassica cretica* Lam. and *Brassica oleracea* L. (donor: Genetic Resources Unit, University of Warwick, Warwick, UK), and *Thlaspi arvense* L. and *Noccaea caerulescens* (J. Presl & C. Presl) F.K. Mey. (previously *Thlaspi caerulescens*) (from Ganges, southern France (Hammond *et al.*, 2006) and from Youlgrave, Derbyshire, UK (collected by Martin Broadley and Steven Whiting)). *Cleome pinnata* Pursh. (Cleomeaceae; seed from B and T World Seeds, Olonzac, France) was used as an outgroup.

For cross-taxa normalizations, a two-step strategy before transcriptome analyses was adopted using established methods (Hammond *et al.*, 2005, 2006; Graham *et al.*, 2007). For the first physical step, 250 ng of genomic DNA was extracted from the 15 taxa and hybridized to separate ATH1 GeneChips. *Noccaea caerulescens* (Viviez) genomic DNA hybridizations were reported previously (Hammond *et al.*, 2006). To estimate evolutionary distance among taxa, for each of the 16 taxa, a custom 'chip definition file' (taxon1.cdf ... taxon16.cdf) was produced using Xpecies V2.0 scripts (available from www.affymetrix.arabidopsis.info/xpecies; Hammond *et al.*, 2005, 2006; Graham *et al.*, 2007). For each taxon, the custom *.cdf file was based on *A. thaliana*, but with 100 000 of the least informative PM probes excluded, that is, PM probes with the lowest DNA hybridization signal intensities for each taxon. Subsequently, a similarity matrix of 'least-informative' probe pairs common to each taxa pair was constructed. Branch lengths (based on proportional units) and topology for the 16-taxon phylogeny were calculated using Neighbor-Joining in PHYLIP (Version 3.67; Felsenstein, 1989). For the second *in silico* step, we developed a universal custom chip definition file ('universal.cdf'), retaining only those PM/MM probe pairs whose corresponding PM probe had DNA hybridization signal intensities > 100 (Fig. 1b), as described previously (Broadley *et al.*, 2007). In the present study, the universal.cdf defines 41 370 probe pairs (16.3% of the total), representing 18 494

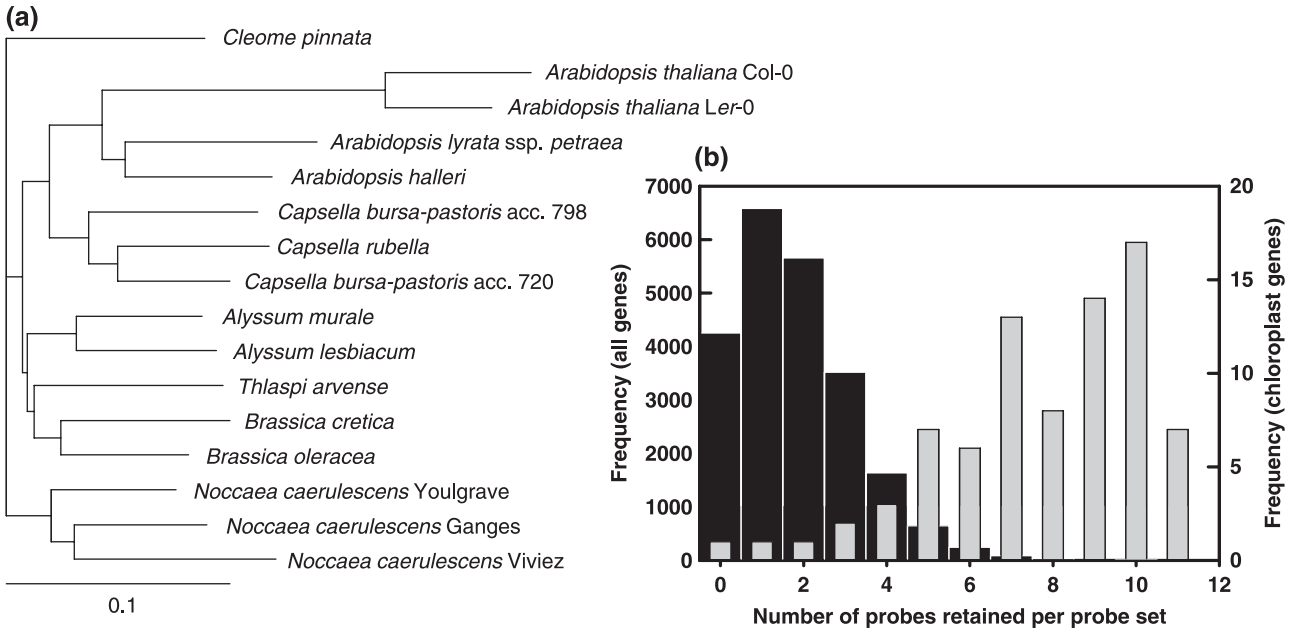

Fig. 1 DNA hybridizations to ATH1 GeneChips resolve a consensus Brassicaceae phylogeny. (a) Unrooted topology of 16 taxa of Brassicaceae based on DNA similarity (i.e. the branch-length scale is a proportional estimate of dissimilarity). (b) The distribution of perfect-match (PM) probe pairs retained on the universal chip definition file (universal.cdf) used for subsequent transcriptome analyses. Black bars represent the distribution of 41 370 probe pairs among 18 494 ATH1 probe sets (out of 22 746), retained at a DNA hybridization signal intensity of 100 across the 14 taxa. Chloroplast genes (grey bars) represent 79 retained ATH1 probe sets, annotated as chloroplast genes, out of a total of 80 on the ATH1 array.

ATH1 probe sets (81.3% of the total), at 2.2 probe pairs per probe set (Fig. 1b).

For transcriptome analyses, all plants were grown simultaneously on 0.8% agar with 0.25× Murashige and Skoog (MS) salts, adjusted to pH 5.6 with NaOH, under conditions described previously (Hampton *et al.*, 2004). At 19 d after sowing, nonsenescent, fully expanded rosette leaves were pooled from eight to 22 plants and snap-frozen at −70 C. The experiment was performed in triplicate in three sequential blocks. RNA was extracted, and 2–4 μg was reverse-transcribed, labelled in an *in vitro* transcription assay, fragmented and hybridized to ATH1 as described previously (Hammond *et al.*, 2005). Thus, 42 RNA .cel files were analysed in GeneSpring (Version GX7.3; Agilent Technologies, Santa Clara, CA, USA) using the universal.cdf and a Robust Multi-array Average (RMA) prenormalization routine. Raw probe-set (gene) expression values were subsequently normalized to the median gene expression value across all 42 samples, and $\log_e$-transformed. The arithmetic mean and variance of the normalized, $\log_e$-transformed gene expression was calculated for each taxon ($n = 3$). Transcriptome divergence was calculated for 91 pair-wise comparisons of taxa. Transcriptome divergence between members of a pair of taxa is the square-root of the sum of the squared differences in mean expression value for all genes. Expression diversity for each taxon was the variance among triplicate samples; genes were ranked by their arithmetic mean variances across all 14 taxa.

For meta-analyses, three independent data sets were used, all of which were re-analysed in GeneSpring (Agilent) after normalizing the data using our universal.cdf and RMA algorithms. First, we used 34 RNA .cel files generated by Benfey and colleagues (Brady *et al.*, 2007). These data comprised 12 cell types ($n = 3$ unless stated), sampled by cell sorting or dissection: atrichoblasts; columella; cortex; endodermis-cortex-quiescent centre (QC); endodermis beyond mature hair zone QC; epidermis; epidermis-lateral root cap; pericycle; phloem ($n = 2$); QC ($n = 2$); root stele to elongation zone; xylem. Transcriptome divergence among cell types was calculated as the square-root of the sum of the squared differences in mean expression value for 66 pair-wise comparisons. Transcriptome divergence was calculated for all genes, and for genes with the highest and lowest expression diversities. Expression diversity for each cell type was the variance among samples; genes were ranked subsequently by their arithmetic mean variances across all 12 cell types. The second and third data sets were combined with our 42 cross-taxa RNA .cel files. The second data set comprised 126 RNA .cel files representing a 'developmental-baseline' of *A. thaliana* Col-0, generated by the AtGenExpress Consortium (Schmid *et al.*, 2005). These data include shoot, root and floral material from a variety of tissue ages and growth conditions (Supporting Information Fig. S1). The third data set comprised 24 RNA .cel files generated by Macnair and colleagues, representing shoot and root tissue from *A. lyrata* ssp. *petraea* and *A. halleri* grown at high and low zinc

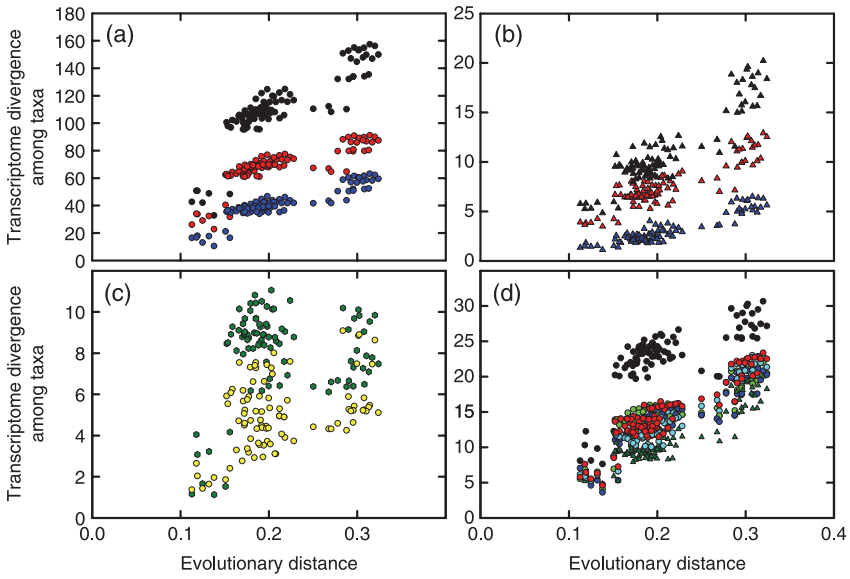

Fig. 2 Transcriptome divergence among Brassicaceae increases monotonically with evolutionary distance and correlates positively with expression diversity. Transcriptome divergence is shown for (a) all genes (n = 18 494; black circles), including those with the highest (n = 4624; red) and lowest (n = 4624 blue) expression diversities, respectively; (b) pseudogenes (n = 511), with labels as for (a); (c) 39 randomly selected chloroplast genes (n = 79; green), and all genes with eight probes per probe set (n = 39; yellow) among 14 taxa; and (d) 500 randomly selected genes represented by one (black), two (green), three (red), four (blue), and five (cyan) probe pairs per probe set, and for pseudogenes (green triangles).

supplies (Filatov *et al.*, 2006). Thus, we analysed 192 RNA .cel files simultaneously. Principal components analysis (PCA) was performed on normalized, $\log_e$-transformed gene expression values, using GenStat (Version 10.1.0.72; Lawes Agricultural Trust, VSN International, Hemel Hempstead, UK).

Results and Discussion

We have found evidence that stochastic variation in gene expression has a role in plant transcriptome evolution. Our study required the use of robust cross-taxa transcriptomics, which involves a two-step physical and *in silico*-based normalization procedure. For the physical step, we hybridized genomic DNA from each taxon to the ATH1 array. This provided a robust estimate of evolutionary divergence among taxa. Thus, a similarity-based topology of ATH1 PM probes hybridizing to DNA from Brassicaceae taxa approximates the Brassicaceae phylogeny based on chloroplast and ITS sequences (Fig. 1a; Koch *et al.*, 2005; Bailey *et al.*, 2006; Beilstein *et al.*, 2006). The populations of *N. caerulescens* from southern France – which included an in-group control from Viviez – resolve as expected. *Capsella* does not resolve according to taxonomic characters which are notoriously difficult to interpret within this group (Ceplitis *et al.*, 2005; Hawes *et al.*, 2005). Thus, it may be possible to use genomic DNA hybridizations to microarrays to support phylogenetic reconstruction within certain groups of organisms. For the *in silico* step, we developed a universal chip definition file ('universal.cdf'), retaining only those PM/MM probe pairs whose corresponding PM probe had raw DNA hybridization signal intensities > 100 for all 14 taxa (Fig. 1b). In this study, the universal.cdf defines 41 370 probe pairs (16.3% of the total), representing 18 494 ATH1 probe sets (81.3% of the total), at 2.2 probe pairs per probe set (Fig. 1b). The 80 chloroplast genes on ATH1 are represented by 79 genes on the universal.cdf at an average of 7.8 (SEM ± 0.28) probe pairs per probe set, consistent with sequence conservation among chloroplast genes. Of 3889 *A. thaliana* genes annotated as pseudogenes (TAIR7.0), 744 have probe sets on ATH1, of which 511 are retained in the universal.cdf with one probe per probe set, consistent with greater sequence divergence among pseudogenes.

We quantified gene transcripts in nonsenescent leaves sampled from 14 Brassicaceae taxa cultivated *in vitro*. The mean expression and variance of a transcript were estimated using $\log_e$-normalized data, derived from three biological replicates. Expression diversity for each gene was calculated as the mean variance across all 14 taxa; this value represents variation which can be attributable to genetic, environmental or technical error components. Transcriptome divergence between members of each pair of taxa (n = 91) was represented by a single numeric descriptor, calculated as the square-root of the sum of the squared differences in mean expression value across (1) all 18 494 genes, (2) 511 pseudogenes, and (3) 79 chloroplast genes. Transcriptome divergence accumulated monotonically as a function of evolutionary time for all genes (Fig. 2a), for pseudogenes (Fig. 2b) and for chloroplast genes (Fig. 2c). The significance of associations between transcriptome divergence and evolutionary distance was determined using Mantel tests, based on 10 000 random permutations for all genes, pseudogenes and chloroplast genes (GenStat Version 10.1.0.72). For all genes and pseudogenes, no permutation produced a stronger association between transcriptome divergence and evolutionary distance than was observed in this study. For chloroplast genes, only 1% of permutations produced a stronger association between transcriptome divergence and evolutionary distance. Transcriptome divergence was greater for the 25% of genes with the highest ranked expression diversity than for the 25% of genes with the

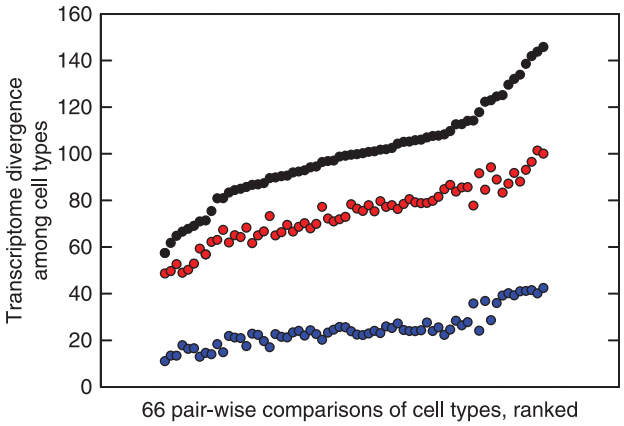

Fig. 3 Genes whose expression varies the most among individual samples have greater expression divergence among *Arabidopsis thaliana* root cell types. Data are from 34 source data sets representing 12 root cell types (Brady *et al*., 2007). Transcriptome divergence was calculated for 66 pair-wise comparisons of cell types and ranked for all genes. Transcriptome divergence between members of a pair of cell types is the square-root of the sum of the squared differences in mean expression value for all genes (n = 18 494; black circles), and genes with the highest (n = 4624; red) and lowest (n = 4624; blue) expression diversities, respectively.

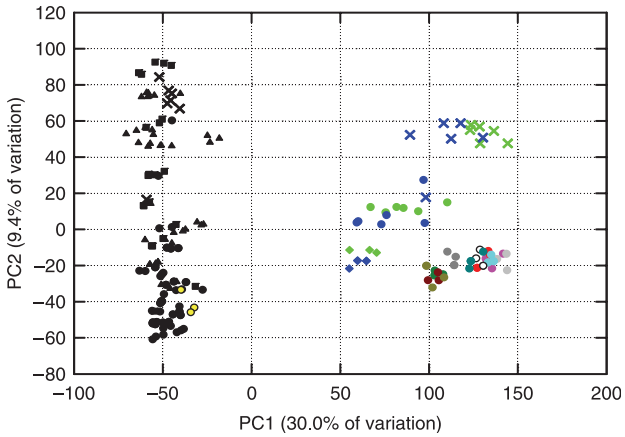

Fig. 4 Transcriptome divergence among functionally homologous tissues sampled from congenerics is greater than divergence among functionally specialized tissues from conspecifics. Data are principal components (PC) 1 and 2 from a PC analysis of transcriptome data from (1) 42 ATH1 data sets reported in this study, (2) 126 ATH1 data sets of 'developmental-baseline' tissue for *Arabidopsis thaliana* Col-0 from the AtGenExpress Project (Schmid *et al*., 2005); (3) 24 ATH1 data sets of shoot and root tissue from *Arabidopsis lyrata* ssp. *petraea* and *Arabidopsis halleri* grown at high and low zinc supplies (Filatov *et al*., 2006). Black symbols, *A. thaliana* Col-0; yellow, *A. thaliana* Ler-0; light green, *A. halleri*; blue, *A. lyrata* ssp. *petraea*; dark red, *Capsella bursa-pastoris* 798; dark green, *C. bursa-pastoris* 720; dark yellow, *Capsella rubella*; red, *Alyssum murale*; pink, *Alyssum lesbiacum*; cyan, *Brassica oleracea*; light grey, *Brassica cretica*; dark grey, *Thlaspi arvense*; dark cyan, *Noccaea caerulescens* Youlgrave; white, *N. caerulescens* Ganges. Circles, leaf/shoot material; cross symbols, roots (colours as before); triangles, floral tissues, including flowers, carpels, petals, sepals, stamens and pedicels; squares, stems, hypocotyls and shoot apex tissues. PC1 and 2 are highlighted. A full 45-panel figure representing PCs 1–10 is available as Supporting Information Fig. S1.

lowest ranked expression diversity, both for all genes (Fig. 2a) and for pseudogenes (Fig. 2b). These observations are consistent with a neutralist interpretation (Khaitovich *et al*., 2004).

We tested if the physical properties of mRNAs affected estimates of transcriptome divergence using information for 13 012 *A. thaliana* Ler-0 genes (Narsai *et al*., 2007), 10 454 of which were represented on the universal.cdf. These genes were ranked, in turn, according to mRNA stability, length, number of introns, or GC content. Within each category, genes were grouped into two sets representing the 1000 highest and 1000 lowest values. There were no significant differences in the rate of transcriptome divergence among taxa in terms of stability ($P = 0.96$), length ($P = 0.38$), number of introns ($P = 0.40$), or GC content ($P = 0.19$) of the mRNA. We also tested whether hybridization artefacts affected estimates of transcriptome divergence. When two or more probe pairs were retained in a probe set, estimates of transcriptome divergence as a function of evolutionary time were unaffected (Fig. 2d).

If stochastic variation drives transcriptome divergence among taxa, then it will also drive transcriptome divergence among tissues (Khaitovich *et al*., 2004, 2006), albeit subject to evolutionary developmental and stabilizing constraints operating at several selection loci (Gould, 2002). As the *A. thaliana* root transcriptome has been mapped at high resolution, we sought evidence of neutral transcriptome properties among 18 454 genes from 34 individual ATH1 arrays, representing cell-sorted or dissected cell types from 12 root zones (Brady *et al*., 2007). There is strong evidence that stochastic variation is an important component of the transcriptome divergence among root cell types. Thus, genes with the highest expression diversity among samples diverged among cell types at a much faster rate than genes with the lowest expression diversity, based on 66 pair-wise comparisons (Fig. 3). Remarkably, there was a significant positive correlation between expression diversity in the 12 root cell types of Brady *et al*. (2007) and leaf-transcriptome divergence among the 14 Brassicaceae taxa from the present study ($r = 0.28$, $\mathrm{df} = 18\,492$, $t = 29.3$, $P = 1.6 \times 10^{-181}$). Tissue-specific data from several taxa are now required to disentangle cell type-specific functional insights from evolutionary and developmental constraints.

A profound consequence of neutral transcriptome evolution is that a difference in the expression level of a gene between two taxa may reflect an ancestral plasticity and/or a founder effect of a small population size, rather than a functional adaptation (Khaitovich *et al*., 2004, 2006). Here we observed that functionally distinct tissues (e.g. leaves, roots and floral organs) sampled from *A. thaliana* at different tissue ages and under different growth conditions have transcriptomes that are more similar to each other than to the transcriptomes of corresponding, functionally homologous tissues sampled from different species under identical growth conditions (Fig. 4).

This observation is consistent with a neutralist interpretation of transcriptome divergence among taxa, and with evidence that relatively small numbers of genes are likely to control organ identity (Soltis *et al*., 2007). Thus, principal component (PC) 1 separates the transcriptomes of *A. thaliana* (Col-0 and L*er*-0) from all other taxa, irrespective of growth conditions or plant part. In PC1, leaf transcriptomes of *A. thaliana* Col-0 and L*er*-0 accessions from our current study were more similar to green *and* nongreen tissue (roots and floral organs) transcriptomes of *A. thaliana* Col-0 from previous studies (Schmid *et al*., 2005) than to the leaf transcriptomes of close relatives from the *Arabidopsis* genus or tribe Camelineae (which contains both *Capsella* and *Arabidopsis*; Schranz *et al*., 2007), despite the obvious conservation of leaf function and identical growth conditions. Similarly, in PC1, the leaf transcriptomes of *A. lyrata* ssp. *petraea* and *A. halleri* from our current study were more similar to shoot *and* root transcriptomes of *A. lyrata* ssp. *petraea* and *A. halleri* measured previously (Filatov *et al*., 2006) than to the leaf transcriptomes of other close relatives. *Arabidopsis lyrata* ssp. *petraea* and *A. halleri* ($2n = 16$) are more closely related to each other than either species is to *A. thaliana*, and are inter-fertile. PC2 separates green and nongreen tissue transcriptomes and higher resolution phylogenetic and anatomical separation of transcriptomes occurs in PC3–PC10 (Fig. S1).

In conclusion, we have found compelling evidence for neutral transcriptome evolution, from new cross-taxa studies and from examination of *Arabidopsis* transcriptomes mapped previously. Transcriptome divergence among taxa and tissues is clearly heritable. It will be intriguing to discover how much of this is attributable to simple sequence polymorphisms, genome rearrangements, transposon insertions or epigenetic modifications. As sequencing costs decline and more genomes and transcriptomes are sequenced, it will be possible to further explore the causes and effects of neutral transcriptome evolution. Whilst many differences in transcript abundance among tissues and among taxa will have functional consequences, integrated 'omics and phenotypic data are required for several species before these can be interpreted correctly in space and time.

Acknowledgements

All .cel files and normalization scripts are available from http://affymetrix.arabidopsis.info/xpecies. Funding for this work was provided in part by the UK Biotechnology and Biological Sciences Research Council (BBSRC), and UK Engineering and Physical Sciences Research Council (EPSRC) through the UK Centre for Plant Integrative Biology (CPIB; MRB, NJG, RGF and STM). Funding was also provided by the Scottish Executive's Environment and Rural Affairs Department through the SCRI Innovation Fund (SEERAD; PJW, JWM and PPMI), and the UK Department for Environment, Food, and Rural Affairs (Defra; JPH and HCB). We thank Malcolm Bennett (University of Nottingham) for comments on the manuscript.

References

Bailey CD, Koch MA, Mayer M, Mummenhoff K, O'Kane SL, Warwick SI, Windham MD, Al-Shehbaz IA. 2006. Toward a global phylogeny of the Brassicaceae. *Molecular Biology and Evolution* **23**: 2142–2160.

Beilstein MA, Al-Shehbaz IA, Kellogg EA. 2006. Brassicaceae phylogeny and trichome evolution. *American Journal of Botany* **93**: 607–619.

Brady SM, Orlando DA, Lee J-Y, Wang JY, Koch J, Dinneny JR, Mace D, Ohler U, Benfey PN. 2007. A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* **318**: 801–806.

Broadley MR, White PJ, Hammond JP, Zelko I, Lux A. 2007. Zinc in plants. *New Phytologist* **173**: 677–702.

Ceplitis A, Su Y, Lascoux M. 2005. Bayesian inference of evolutionary history from chloroplast microsatellites in the cosmopolitan weed *Capsella bursa-pastoris* (Brassicaceae). *Molecular Ecology* **14**: 4221–4233.

Craigon DJ, James N, Okyere J, Higgins J, Jotham J, May S. 2004. NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. *Nucleic Acids Research* **32**: D575–D577.

Felsenstein J. 1989. PHYLIP-Phylogeny Inference Package (Version 3.2). *Cladistics* **5**: 164–166.

Filatov V, Dowdle J, Smirnoff N, Ford-Lloyd B, Newbury HJ, Macnair MR. 2006. Comparison of gene expression in segregating families identifies genes and genomic regions involved in a novel adaptation, zinc hyperaccumulation. *Molecular Ecology* **15**: 3045–3059.

Gilad Y, Oshlack A, Rifkin SA. 2006. Natural selection on gene expression. *Trends in Genetics* **22**: 456–461.

Gould SJ. 2002. *The structure of evolutionary theory*. Cambridge, MA, USA: Harvard University Press.

Graham NS, Broadley MR, Hammond JP, White PJ, May ST. 2007. Optimising the analysis of transcript data using high density oligonucleotide arrays and genomic DNA-based probe selection. *BMC Genomics* **8**: 344.

Hammond JP, Bowen HC, White PJ, Mills V, Pyke KA, Baker AJM, Whiting SN, May ST, Broadley MR. 2006. A comparison of the *Thlaspi caerulescens* and *T. arvense* shoot transcriptomes. *New Phytologist* **170**: 239–260.

Hammond JP, Broadley MR, Craigon DJ, Higgins J, Emmerson Z, Townsend H, White PJ, May ST. 2005. Using genomic DNA-based probe-selection to improve the sensitivity of high-density oligonucleotide arrays when applied to heterologous species. *Plant Methods* **1**: 10.

Hampton CR, Bowen HC, Broadley MR, Hammond JP, Mead A, Payne KA, Pritchard J, White PJ. 2004. Cesium toxicity in Arabidopsis. *Plant Physiology* **136**: 3824–3837.

Hawes C, Begg GS, Squire GR, Iannetta PPM. 2005. Individuals as the basic accounting unit in studies of ecosystem function: functional diversity in shepherd's purse, *Capsella*. *Oikos* **109**: 521–534.

Johnston JS, Pepper AE, Hall AE, Chen ZJ, Hodnett E, Drabek J, Lopez R, Price HJ. 2005. Evolution of genome size in Brassicaceae. *Annals of Botany* **95**: 229–235.

Khaitovich P, Enard W, Lachmann M, Pääbo S. 2006. Evolution of primate gene expression. *Nature Reviews Genetics* **7**: 693–702.

Khaitovich P, Pääbo S, Weiss G. 2005. Toward a neutral evolutionary model of gene expression. *Genetics* **170**: 929–939.

Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, Muetzel B, Wirkner U, Ansorge W, Pääbo S. 2004. A neutral model of transcriptome evolution. *PloS Biology* **2**: 0682–0689, doi: 10.1371/journal.pbio.0020132.

Kimura M. 1968. Evolutionary rate at the molecular level. *Nature* **217**: 624–626.

Koch MA, Dobeš C, Matschinger M, Bleeker W, Vogel J, Kiefer M, Mitchell-Olds T. 2005. Evolution of the *trnF*(GAA) gene in *Arabidopsis* relatives and the Brassicaceae family: monophyletic origin and subsequent diversification of a plastidic pseudogene. *Molecular Biology and Evolution* **22**: 1032–1043.

Narsai R, Howell KA, Millar AH, O'Toole N, Small I, Whelan J. 2007. Genome-wide analysis of mRNA decay rates and their determinants in *Arabidopsis thaliana*. *Plant Cell* **19**: 3418–3436.

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU. 2005. A gene expression map of *Arabidopsis thaliana*. *Nature Genetics* **37**: 501–506.

Schranz ME, Song B-H, Windsor AJ, Mitchell-Olds T. 2007. Comparative genomics in the Brassicaceae: a family-wide perspective. *Current Opinion in Plant Biology* **10**: 168–175.

Soltis DE, Ma H, Frohlich MW, Soltis PS, Albert VA, Oppenheimer DG, Altman NS, dePamphilis C, Leebens-Mack J. 2007. The floral genome: an evolutionary history of gene duplication and shifting patterns of gene expression. *Trends in Plant Science* **12**: 358–367.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Principal components analysis (PCA) of transcriptome data from (1) 42 ATH1 data sets reported in this study, (2) 24 ATH1 datasets of shoot and root tissue from *Arabidopsis lyrata* ssp. *petraea* and *Arabidopsis halleri* grown at high and low zinc supplies (Filatov *et al.*, 2006), and (3) 126 ATH1 data sets of 'developmental-baseline' tissue for *Arabidopsis thaliana* Col-0 from the AtGenExpress Project (Schmid *et al.*, 2005). Legend as for Fig. 4.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than about missing material) should be directed to the *New Phytologist* Central Office.

