Plant reproductive development is characterised by a transcriptomic evolutionary bulge

Toni I. Gossmann[1,2,§,*], Dounia Saleh[1,§], Marc W. Schmid[3],
Michael A. Spence[2] and Karl J. Schmid[1,*]
[1]University of Hohenheim, Germany [2]University of Sheffield, UK
[3]University of Zurich, Switzerland
[§] Co-First Authors
*Correspondence: toni.gossmann@gmail.com, karl.schmid@uni-hohenheim.de

July 21, 2015

Abstract

Reproductive traits in plants tend to evolve rapidly due to various causes that include plant-pollinator coevolution and pollen competition, but the genomic basis of reproductive trait evolution is still largely unknown. To characterise evolutionary patterns of genome wide gene expression in reproductive tissues and to compare them to developmental stages of the sporophyte, we analysed evolutionary conservation and genetic diversity of protein-coding genes using microarray-based transcriptome data from three plant species, *Arabidopsis thaliana*, rice (*Oryza sativa*) and soybean (*Glycine max*). In all three species a significant shift in gene expression occurs during gametogenesis in which genes of younger evolutionary age and higher genetic diversity contribute significantly more to the transcriptome than in other stages. We refer to this phenomenon as "evolutionary bulge" during plant reproductive development because it differentiates the gametophyte from the sporophyte. The extent of the bulge pattern is much stronger than the transcriptomic hourglass, which postulates that during early embryo development an increased proportion of ancient and conserved genes contribute to the total transcriptome. In the three plant species, we observed an hourglass pattern only in *A. thaliana* but not in rice or soybean, which suggests that unlike the evolutionary bulge of reproductive genes the transcriptomic hourglass is not a general pattern of plant embryogenesis, which is consistent with the absence of a morphologically defined phylotypic stage in plant development.

Introduction

Reproductive traits in plants and animals tend to be highly diverse and rapidly evolving within and between closely related species (Swanson and Vacquier, 2002; Barrett, 2002; Parsch and Elle-

gren, 2013). Their diversity may be influenced by the coevolution with pollinators or pathogens that infect reproductive tissues, the mating system (i.e. selection for the maintenance of self-incompatibility), the rapid evolutionary dynamics of sex chromosomes, genomic conflicts between parents and offspring, or from sexual selection (Baack et al., 2015). Some genes and proteins expressed in reproductive tissues exhibit high rates of evolution (Swanson and Vacquier, 2002; Parsch and Ellegren, 2013). In plants, they include genes encoding the self-incompatibility system (Nasrallah et al., 2002; Tang et al., 2007), pollen-coat proteins (Schein et al., 2004) and imprinted genes controlling resource allocation to offspring (Spillane et al., 2007). The rapid evolution of reproductive traits and their underlying genes is in contrast to other tissues and developmental stages that appear to be more conserved. In particular, the phylotypic stage in animals (a stage during embryonic development common to a broad range of animal taxa where a very similar morphology is observed) represents the archetype of morphological evolutionary conservation within a phylum (Duboule, 1994).

Although reproductive traits appear to evolve rapidly in animals, plants and other organisms with anisogamic sexual reproduction (Lipinska et al., 2015), there is a fundamental difference between these groups. In animals, a group of cells are set aside during early development, which forms the germ line. Plants do not have a germ line, but are characterized by alternating sporophytic and haploid gametophytic stages (Schmidt et al., 2011; Grossniklaus, 2011). The function of genes expressed in the sporophyte and gametophyte differ from each other due to their roles in development and reproduction. Hence the type and strength of selection acting during the two stages should also differ. Furthermore, the haploid stage immediately exposes recessive mutations to selection which causes different evolutionary dynamics of genes expressed in the gametophyte compared to genes only expressed in a diploid stage (Gossmann et al., 2014b).

Currently it is little understood which processes drive the rapid evolution of plant reproductive genes on a genome-wide scale. During plant gametogenesis, the transcription profile changes dramatically, and genes involved in reproduction are enriched in this phase (Schmid et al., 2005; Fujita et al., 2010; Xiao et al., 2011; O'Donoghue et al., 2013). However, a focus on genes whose expression is enriched in a specific tissue introduces a bias for genes with specific expression patterns that ignores the contribution of other genes to the total diversity of expression patterns (Arunkumar et al., 2013; Gossmann et al., 2014b; Harrison et al., 2015). To characterise the evolutionary dynamics of transcriptomic profiles it is therefore necessary to combine the genome-wide expression intensity of all genes expressed in a given tissue and stage with evolutionary parameters quantifying the level of polymorphism, rate of molecular evolution or long-term evolutionary conservation (Slotte et al., 2011). For this purpose, evolutionary indices such as the transcriptome age index (TAI), which measures the long-term conservation of expressed genes weighted by the relative expression of the gene, or the divergence index (TDI), which compares the rate of non-synonymous to synonymous substitutions in a protein-coding gene between closely related species (Domazet-Lošo and Tautz, 2010; Kalinka et al., 2010; Quint et al., 2012) were developed to test whether the phylotypic stage as defined by Haeckel has a molecular equivalent. Studies in vertebrates (zebrafish) and insects

(*Drosophila melanogaster*) confirmed this hypothesis because genes expressed during the phylotypic stage were more conserved and less rapidly evolving than genes expressed in other stages of development (Domazet-Lošo and Tautz, 2010; Kalinka et al., 2010). Although plants do not have a clear morphologically defined phylotypic stage, a transcriptomic hourglass was also postulated for the model plant *Arabidopsis thaliana* because old and slowly evolving genes contribute disproportionally to the overall transcriptome during early stages of embryo development (Quint et al., 2012; Drost et al., 2015) (but see (Piasecka et al., 2013)).

Based on the above considerations, we reasoned that the morphologically and developmentally diverse reproductive stages of plants, in particular the gametophyte, should be characterized by a high proportion of expressed genes with a lower degree of long-term evolutionary conservation and a higher rate of divergence between closely related species. We tested this hypothesis by comparing the transcriptome-based indices of gametogenesis and reproductive stages to other stages of plant development, such as the putative phylotypic stage, based on the magnitude and direction of transcriptomic indices. We based our analysis on three different evolutionary time scales and used gene expression and genome sequence data from three flowering plant species and one moss. Our results show that the rate of evolution of genes expressed in reproductive stages is much higher relative to the extent of conservation of the putative phylotypic or other sporophytic stages. For this reason, we termed this pattern 'evolutionary bulge' which describes the difference of transcriptome indices of reproductive developmental stages to other stages.

Results

Transcriptome indices throughout development

We conducted an analysis of global expression during gamete development, as well as developmental stages preceeding and following gametogenesis (e.g. flower development and embryogenesis). Our study differs from previous studies by including more developmental stages. The analysis of global expression accounts for the evolutionary contribution of the whole transcriptome and is less biased than a restriction on genes that are specifically expressed or significantly overexpressed in certain stages and tissues. To this end, we combined microarray expression levels with measures of evolutionary conservation and polymorphism of genes to calculate evolutionary transcriptome indices of developmental stages (Kalinka et al., 2010; Domazet-Lošo and Tautz, 2010). We analysed data from the three plant species *Arabidopsis thaliana*, rice (*Oryza sativa*) and soybean (*Glycine max*) to test whether observed patterns are of a general nature. In addition to two previously defined transcriptome evolutionary indices (TAI and TDI (Kalinka et al., 2010; Domazet-Lošo and Tautz, 2010; Quint et al., 2012)), we also calculated a transcriptome polymorphism index (TPI; see Materials and Methods) which is a measure of current evolutionary constraint. In brief, the three indices reflect an association of an evolutionary parameter with genome-wide gene expression on three different time scales (long, medium and short term) and are the average (over all genes) of an evolutionary parameter weighted by transcription intensity, namely gene age (TAI), diver-

gence (TDI) and diversity (TPI). We obtained gene expression data for 19 developmental stages for *A. thaliana*, 15 stages for rice and 7 stages for soybean from publicly available sources (Table 1 and S1 File). The datasets for *Arabidopsis* included two stages from the female and six from the male gametophyte, five male stages from rice and one male stage from soybean.

Table 1: **Summary of microarray-based expression data from different developmental stages used in this study.** Further details about the individual datasets are provided in Supporting File S1.

Species	Developmental Stage	References
A. thaliana	*Pre-Reproductive stage:* Shoot apex 7 days (SA7D), Shoot apex 14 days (SA14D), Shoot after bolting (SAB), Flower stage 9 (FS9), Flower stage 12 (FS12), Flower stage 15 (FS15)	(Schmid et al., 2005)
	Reproductive stage: Megaspore mother cell (MMC), Egg cell (EC), Unicellular pollen (UCP), Bicellular pollen (BCP), Tricellular pollen (TCP), Pollen mature (MP), Sperm (S), Pollen-tube (PT)	(Honys and Twell, 2004; Borges et al., 2008; Wang et al., 2008; Wuest et al., 2010; Schmidt et al., 2011; Schmid et al., 2012)
	Post-reproductive stage: Quadrant (Q), Globular (G), Heart (H), Torpedo (T), Mature (M)	(Le et al., 2010; Zuber et al., 2010)
Rice	*Pre-Reproductive stage* Shoot 4 weeks (S4W)	(Fujita et al., 2010)
	Reproductive stage: Unicellular pollen (UCP), Bicellular pollen (BCP), Tricellular pollen (TCP), Mature pollen (MP), Germinated pollen (GP)	(Wei et al., 2010)
	Post-Reproductive stage: Fertilisation (F), Zygote formation (Z), 0 Days After Pollination (0DAP), 1 Days After Pollination (1DAP), 2DAP, 3DAP, 4DAP, 9DAP, 12DAP	(Fujita et al., 2010; Gao and Xue, 2012)
Soybean	*Pre-Reproductive stage:* Sporophyte (S)	(Haerizadeh et al., 2009)
	Reproductive stage: Mature pollen (MP)	(Haerizadeh et al., 2009)
	Post-Reproductive stage: Globular (G), Heart (H), Cotyledon (C), Seed parenchym (SP), Seed meristem (SSM)	(Le et al., 2007)

Evolutionary transcriptome indices increase during reproductive stages

In all three species we observed the highest values for the three evolutionary indices during reproductive stages (Fig. 1) and their values differ significantly from the sporophytic developmental stages. In *A. thaliana*, the highest TAI value is found in the pollen tube stage (t-test; $P < 6.5 \times 10^{-34}$ for all pairwise comparisons with sporophytic stages) and the highest TDI and TPI values in sperm cells ($P < 2.2 \times 10^{-15}$). For rice, the highest values for the TAI, TDI and TPI were the mature pollen and germinated pollen ($P < 6 \times 10^{-27}$ for all pairwise comparisons, no significant difference between mature and germinated pollen) and for soybean the germinated pollen ($P < 7.3 \times 10^{-6}$). The *A. thaliana* and rice expression data cover consecutive reproductive stages in which the evo-

lutionary indices increase during gamete maturation and peak at a final reproductive stage (e.g., sperm, pollen, pollen tube). A similar trend is observed in the female gametophytic tissue in *A. thaliana*. By comparing the transition to and from reproductive stages, there is a strong difference between gametophytic and sporophytic phases, suggesting a distinct evolutionary dynamics of reproductive compared to sporophytic stages. A larger dataset from the *A. thaliana* gene expression developmental atlas confirms this observation because gametophytic stages consistently show high evolutionary transcriptome indices if only genes with significant expression in each tissue or stage are taken into account (S1 Fig). The comparison of evolutionary indices between pre- and postgametic developmental stages reveal that the lowest values of these indices are not consistently the lowest during embryogenesis, as suggested by the hourglass hypothesis. Except for *A. thaliana*, there is no particular stage during embryogenesis that has the lowest TAI, TDI and TPI values. Taken together, these results indicate a shift in gene expression during gamete maturation where rapidly evolving genes predominantly contribute to the transcriptome. In contrast, there is little evidence for a transcriptomic hourglass during embryo development as a general pattern of plant development.

The evolutionary bulge is not caused by few genes

We further investigated the possibility that the choice of the test statistic or statistical artifacts may cause the evolutionary bulge but we present only the results for *A. thaliana* because they were very similar for the two other species. To test whether low quality aligments of few genes inflate diversity and polymorphism indices, we calculated the TDI and TPI based on the weighted median (see Material and Methods). The median estimates are generally lower than means indicating that only few genes have highly elevated divergence and diversity estimates, but the results obtained with the weighted mean and weighted median differ only marginally (S2 Fig). Therefore, the evolutionary bulge is not caused by few outlier genes with inflated divergence or diversity estimates. Previous studies of the transcriptome hourglass used raw expression values (Domazet-Lošo and Tautz, 2010; Quint et al., 2012), although large absolute differences in the expression level of genes with a high and low expression level may allow a few genes to dominate the overall transcriptome index (Piasecka et al., 2013). We therefore conducted the above analysis with log_2 transformed data, but additionally verified the bulge pattern with raw and log_{10}-transformed expression data (S3 Fig). In both cases, the relative pattern of transcriptome indices are very similar to the log_2 transformed expression data suggesting that the evolutionary bulge is not driven by few genes with very high expression levels.

Fewer genes are expressed during reproductive stages

All transcriptome data for a given species were generated with the same Affymetrix array, but hybridisations were conducted in independent experiments by different research groups with more than one genotype (Table 1). Therefore, the expression intensities may be confounded by experimental conditions. The above results were based on the joint pre-processing (e.g., normalisation) of all

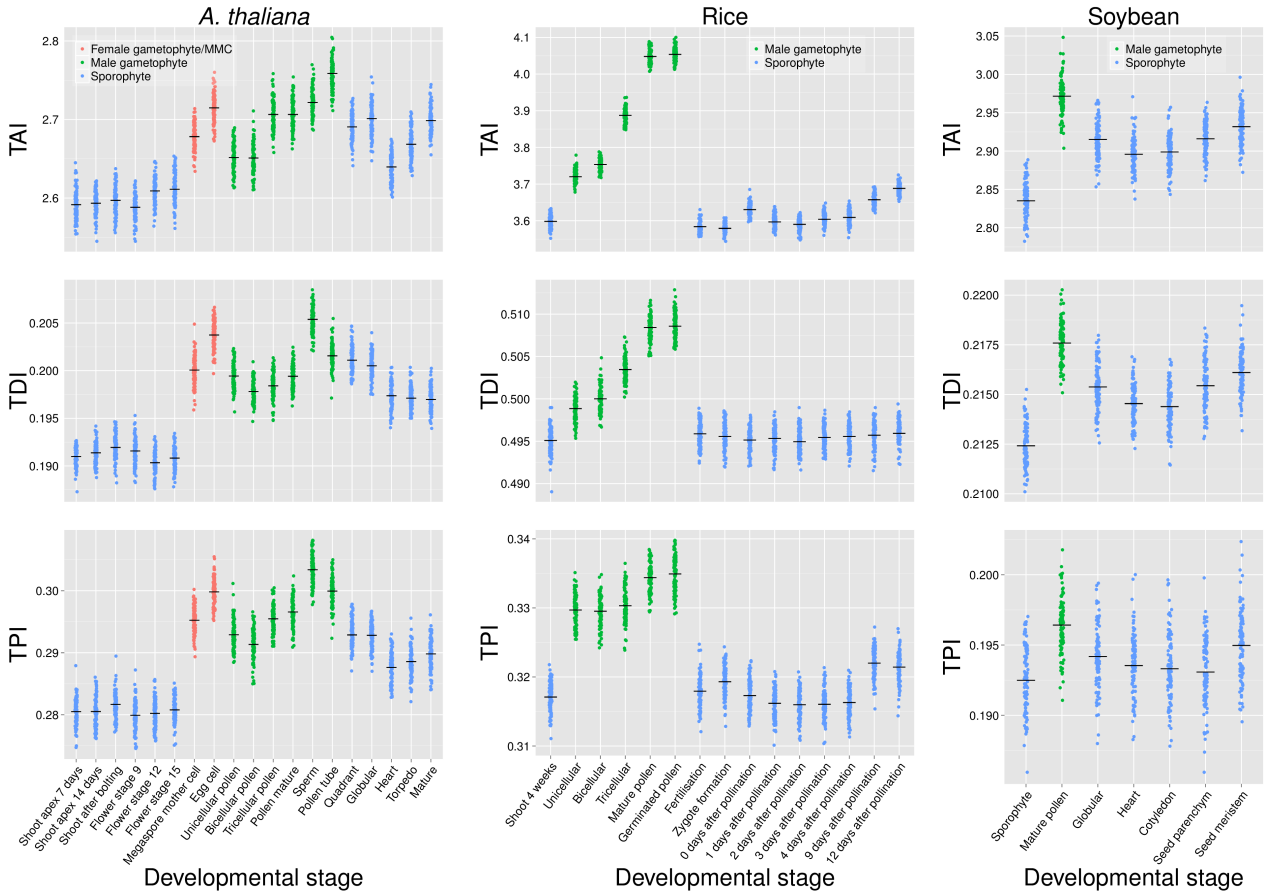


Figure 1: **Evolutionary transcriptome indices for *A. thaliana*, rice and soybean.** Plot of transcriptome age index (TAI), transcriptome divergence index (TDI) transcriptome polymorphism index (TPI) for available data from *Arabidopsis thaliana*, rice and soybean for different developmental stages and tissues. Black bars indicate the transcriptome index and the coloured dots are the indices calculated from random samples (with replacement) of genes to obtain a confidence interval of the index. Blue dots indicate non-reproductive tissues, green and red dots indicate male and female reproductive tissues, respectively.

data. To test for confounding effects we also calculated the transcriptome indices by pre-processing datasets independently of each other (S4 Fig). This led to a relative shift of transcriptome indices between pre- and postgametophytic developmental stages, but the evolutionary bulge during gametogenesis remained as a robust pattern. In *A. thaliana*, the number of significantly expressed genes is reduced in the pollen transcriptome (Pina et al., 2005), and the number of expressed genes may vary between different developmental stages and influence the indices. Unfortunately, the available data make it difficult to calculate P-values of the expression signal along with the expression intensity because of the heterogeneity of experiments. However, using P-values associated with gene expression from a larger dataset for *A. thaliana* compromising 102 different experiments of various developmental stages (S1 Table) we calculated modified transcriptome indices (see Materials and Methods) by including only genes that are significantly expressed in a given stage with an FDR < 0.1 (S1 Fig). With few exceptions, reproductive tissues have higher evolutionary indices, and the number of significantly expressed genes differs between reproductive and vegetative phase

($P = 2 \times 10^{-12}$, U-test of the median number of genes significantly expressed in reproductive versus sporophytic tissues).

The three evolutionary indices are largely independent

Our observation of an evolutionary bulge pattern with all three indices may be confounded by a non-independence of the three indices (Cai and Petrov, 2010). Quint et al.(Quint et al., 2012) concluded from a weak but significant correlation between gene age and d_N/d_S that these two evolutionary parameters are largely independent. To investigate possible correlations between TAI, TDI and TPI, we conducted a partial correlation analysis that accounted for additional, potentially co-varying factors (Gossmann et al., 2014a). We first established a mathematical relationship between the evolutionary transcriptome indices and Pearson's correlation coefficient, r (Supplementary Text S2). Then, we calculated the correlation coefficients between absolute gene expression and evolutionary parameter (e.g., gene age, d_N/d_S and p_N/p_S). By assuming that expression variation between samples is similar and the same genes are analysed across stages, the evolutionary index is proportional to the correlation coefficient, r. The analysis of correlation supports the evolutionary bulge pattern because we observed the lowest (negative) correlation between indices in the reproductive stages which is consistent with a high transcriptome index (Table 2). The only exception was the polymorphism index (TPI) of the two domesticated species (rice and soybean) which was influenced in the reproductive stage by differences in expression variance between reproductive and sporophytic stages (S6 Fig). We then compared these correlation coefficients to partial correlations by taking the other two evolutionary parameters, as well as gene length and d_S (a proxy for mutation rate) as co-variates (Table 2). In this comparison, the evolutionary bulge based on partial correlations is qualitatively very similar to the pairwise correlations. The only exception is the TAI of rice because gene age and d_S are strongly correlated in this species (Spearman $r = 0.34$). In summary, this analysis illustrates that the evolutionary indices are largely independent.

Differences between species in up- and down-regulation of young and old genes during reproduction

We further determined whether the different expression patterns during gamete development resulted from the up- respective down-regulation of young or old genes. We used the absolute expression values of each stage and performed linear regression of mean expression values over the phylostratum values of each stage (Fig. 2). For each pair of stages we then calculated the difference in the slopes of the regression lines to infer how strongly the correlation varied between stages. To further determine whether a change in the slope was caused by a change in the expression level of older or younger genes, we compared the expression level inferred from the regression between each pair of stages for the oldest (first phylostratum) and youngest genes (last phylostratum), respectively. In all three species, the relative expression of both old and young genes differed between developmental stages, but the extent of change varied between stages and species. In *A. thaliana*, the differences were mainly caused by a change in the expression level of young genes (Fig. 2b and

Table 2: **Correlation of gene expression with three evolutionary indices.** The analysis was based on Pearson's correlation and partial correlation for selected development stages. For the partial correlations, the other two evolutionary parameters as well as gene length and d_s were used as co-variates.

Correlation of gene expression intensity with	Gene age		d_N/d_S		p_N/p_S	
	r	r (partial)	r	r (partial)	r	r (partial)
A. thaliana						
Flower stage 9	-0.24***	-0.11***	-0.34***	-0.22***	-0.26***	-0.13***
Egg cell	-0.18***	-0.11***	-0.20***	-0.11***	-0.15***	-0.07***
Sperm	-0.14***	-0.08***	-0.13***	-0.07***	-0.09***	-0.04***
Pollen tube	-0.07***	$0.01^{n.s.}$	-0.19***	-0.16***	-0.12***	-0.04***
Heart	-0.21***	-0.09***	-0.26***	-0.16***	-0.21***	-0.11***
Rice						
Shoot 4 weeks	-0.15***	$0.01^{n.s.}$	-0.25***	-0.04***	-0.06***	$0.01^{n.s.}$
Mature pollen	-0.05***	$-0.01^{n.s.}$	-0.08***	$-0.01^{n.s.}$	-0.06***	-0.03***
Zygote formation	-0.17***	-0.02*	-0.25***	-0.04***	-0.04***	0.03**
Soybean						
Sporophyt	-0.10***	-0.06***	-0.22***	-0.18***	-0.10***	-0.04***
Mature pollen	$-0.01^{n.s.}$	$0.00^{n.s.}$	-0.11***	-0.09***	-0.06***	-0.03**
Heart	-0.07***	-0.03***	-0.16***	-0.14***	-0.07***	-0.03**

c) and in rice by a higher expression of young and a lower expression of older genes (Fig. 2f and g). In soybean, the change in expression was mainly caused by the lower expression level of old genes (Fig. 2j and k).

We also compared the expression levels between stages by grouping genes by their average values of d_N/d_S and p_N/p_S (Fig.) to test whether expression levels differ between slow and rapidly evolving genes. In *A. thaliana*, conserved genes (low d_N/d_S and p_N/p_S) showed a lower expression level and divergent genes (high d_N/d_S and p_N/p_S) a higher expression level in reproductive stages, especially in pollen and pollen tubes. In rice, genes with low d_N/d_S and p_N/p_S values showed strongly decreased mean expression levels in reproductive stages, whereas in soybean, mean expression levels decreased independently from d_N/d_S and p_N/p_S during reproduction. To summarize, although the slopes of correlation differed significantly between gametophytic and sporophytic tissues in all three species, they seem to be caused by different evolutionary processes in each species.

Strong purifying selection in *A. thaliana*

To investigate the extent of advantageous and deleterious mutations contributing to the rapid evolution of reproductive traits we calculated the neutrality index (NI) per gene for *A. thaliana* (see Material and Methods). NI<1 indicates an increased role of positive selection while NI>1 indicates an excess of nonsynonymous polymorphisms that do not contribute to nonsynonymous divergence

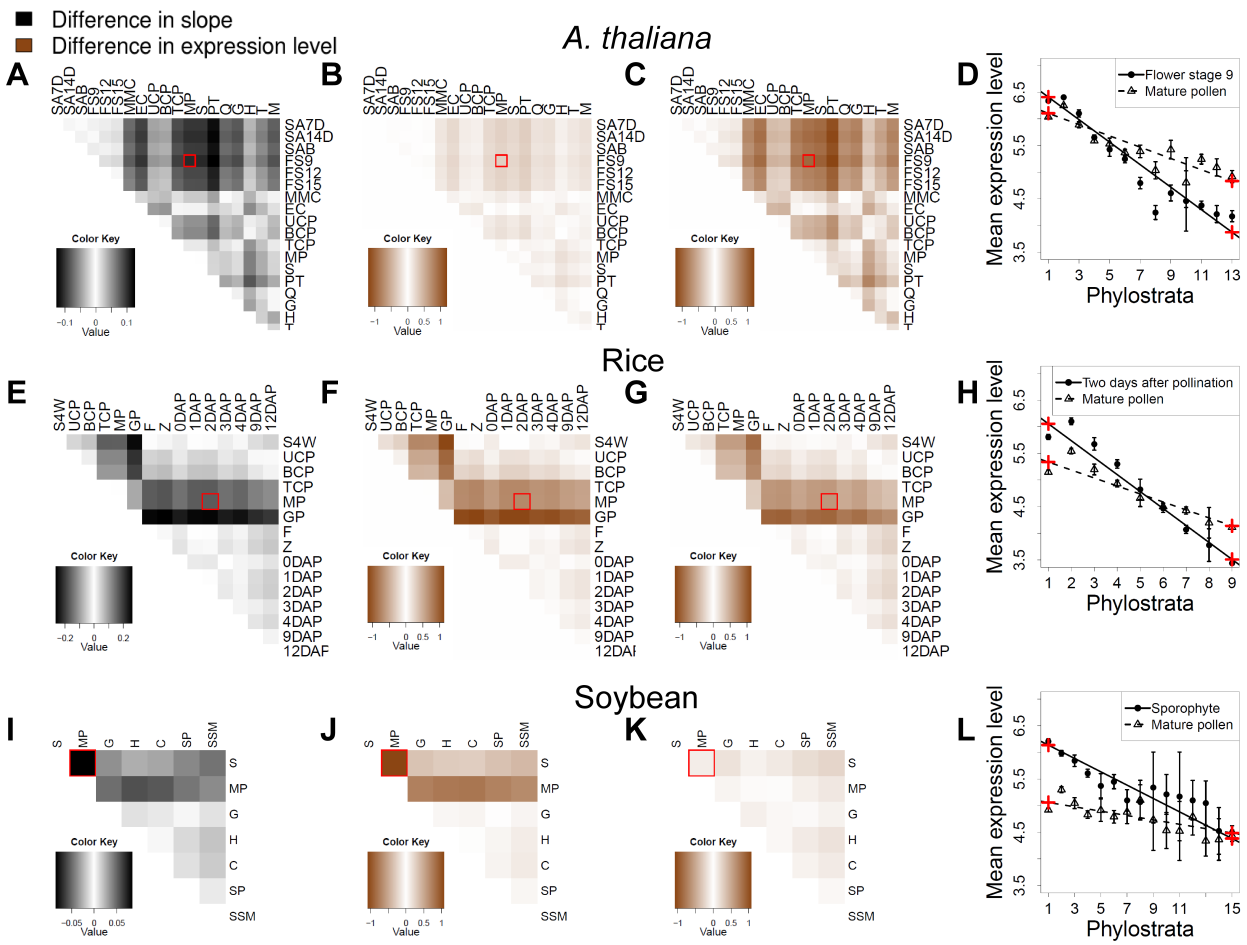


Figure 2: **Difference in expression level between young and old genes and between developmental stages.** (a-d) *A. thaliana* (e-h) rice (i-l) soybean (a, e, i) Heatmaps of differences in linear regression slopes between pairs of developmental stages included in the analysis. (b, f, j) Heatmaps of differences in expression level inferred from linear regressions between pairs of developmental stages for the first phylostratum (PS= 1). (c, g, k) Heatmaps of differences in expression level inferred from linear regressions between pair of developmental stages for the youngest phylotratum (PS= 13 in *A. thaliana*; PS= 9 in rice; and PS= 15 in soybean). (d, h, l) Mean, confidence interval and linear regression of expression level for several phylostrata at two stages: Flower stage 9 and mature pollen in *A. thaliana*, 2DAP and mature pollen in rice, sporophyte and mature pollen in soybean. Red crosses represent the expression level inferred from the linear regressions for PS=1 and PS=13/9/15, respectively. For abbreviations of developmental stages, see Table 1.

and indicate the action of purifying selection. Generally there appears a limited role of adaptive substitutions in plants (Gossmann et al., 2010), but there are exceptions for highly expressed pollen and pollen tube genes (Arunkumar et al., 2013; Gossmann et al., 2014b). To investigate the relative role of purifying and positive selection during developmental stages we calculated the median NI of the per gene NI weighted by its expression intensity. NI is increased in later pollen stages (Fig. 3), which is a significant difference in a comparison of tricellular/mature pollen and pollentube to sporophytic tissues ($P < 9.9 \times 10^{-28}$), suggesting that purifying selection strongly contributes to the evolutionary bulge in male reproduction. It is noteworthy that median NI for sperm genes is reduced, indicating either a greater role of positive or of relaxed purifying selection.

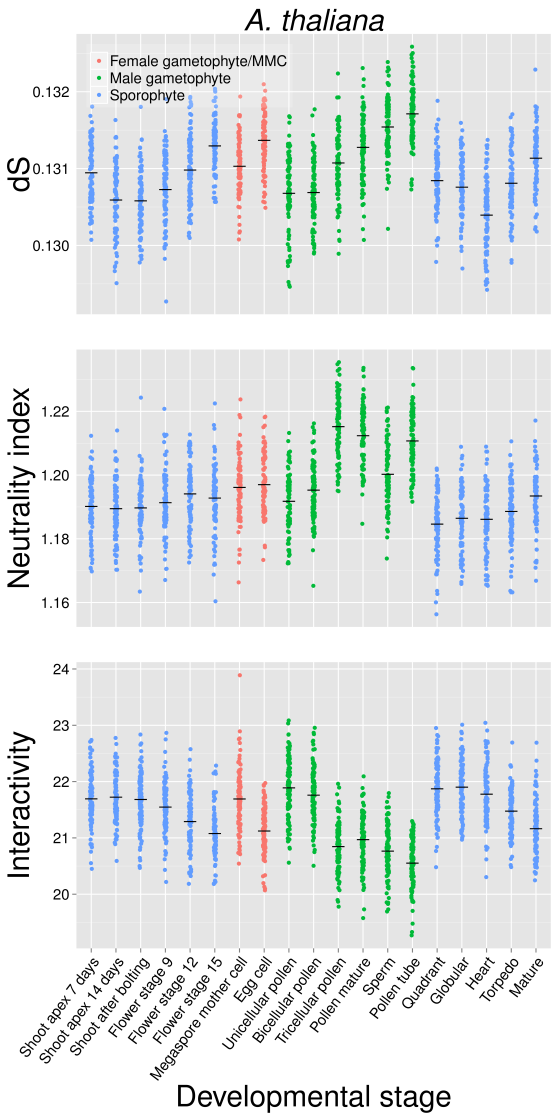


Figure 3: **Transcriptome indices for d_s, neutrality index and gene interactions for *A. thaliana*.** Upper panel: Median per gene d_s (synonymous per site substitution rate, a proxy for the neutral mutation rate) weighted by gene expression. Middle panel: Median per gene neutrality index (NI, a measurement of the departure from neutrality, with NI$\approx$ 1 indicating neutrality) weighted by gene expression. Lower panel: Average number of gene interaction partners weighted by gene expression.

Fewer protein-protein interactions in male but not female reproductive tissues in *A. thaliana*

During reproductive development the tissue complexity of the gametophyte in higher plants is reduced to single cells or a few cells suggesting a reduced interaction between cells and cell types compared to other stages. Highly connected genes tend to evolve slower as a consequence of their functional importance (Alvarez-Ponce and Fares, 2012). Such genes, however, may be less expressed in the gametophytic stage and therefore contribute less to the bulge pattern. This hypothesis is supported by a reduced expression level of old genes in *A. thaliana* (Fig. 2b) as well as rice and

soybean (Fig. 2f and j). Using data from the *Arabidopsis* interactome database (see Materials and Methods) we indeed find that during the late male stages the level of interactivity is reduced with the lowest value occurring in pollen tube (Mean number of interaction partners per gene weighted by the expression intensity, Fig. 3, $P < 0.03$). Interestingly, for the female gametophyte, which is a tissue of much higher complexity, such a reduction in the extent of protein-protein interactions is not observed.

Discussion

According to our analysis, rapidly evolving genes comprise a large proportion of the transcriptome during plant gametophytic development and this observation is consistent for three different angiosperm plants when combining widely independent evolutionary data with transcriptomic information. Although the rapid evolution of reproductive genes seems to be a robust pattern, it is still unclear which biological processes are mainly responsible for this phenomenon. We propose several possible explanations for evolutionary bulge and compare this pattern to the transcriptomic hourglass (Quint et al., 2012).

Influence of the haploid gametophyte on the transcriptome

A major difference between haploid reproductive and diploid sporophytic stages is that recessive mutations are always exposed in the haploid gametophyte. As a consequence, natural selection acts efficiently on such mutations (Otto and Gerstein, 2008). This effect is less important in self-fertilizing species, like the three species in this study, because they tend to be highly homozygous (Wright et al., 2013) and deleterious recessive mutations are also rapidly removed in diploid sporophytic tissues (Szövényi et al., 2014). Therefore, selection acts with the same efficiency on genes expressed haploid and diploid tissues of self-fertilizing species, which contributes little to the evolutionary bulge pattern. On the other hand, the evolution of male reproductive genes is very similar between *A. thaliana* (Gossmann et al., 2014b) and the outcrossing relative *Capsella grandiflora* (Arunkumar et al., 2013). Possible explanations are (1) that low but sufficient levels of outcrossing in self-fertilizing species to remove differences in selection (Bomblies et al., 2010), (2) that most mutations are dominant and therefore exposed to selection in outcrossers, or (3) that the genetic diversity of gametophytic genes is influenced by higher rates of *de novo* mutations during gametogenesis. The latter explanation is based on the observation that replication or transcription during reproductive development induces mutations in yeast (Rattray et al., 2015). Although it is unknown whether this is also the case in plants, silent divergence d_S, which is a proxy for mutation rate, is increased for genes predominantly expressed in sperm and pollen tube stages (Fig. 3; pairwise comparison to other stages with a t-test, $P < 1.7 \times 10^{-4}$). However, a higher d_S caused by an increased mutation rate reduces the d_N/d_S ratio and therefore the TDI index of reproductive stages which is contrast to the observed bulge pattern. Taken together, it seems unlikely that evolutionary bulge in the three selfing species used in this study is caused by their mating system. Future analyses of the

effect of haploidy vs. diploidy on transcriptome indices should include comparisons of closely related outcrossing and self-fertilizing species.

In haploid gametophytes gene dosage needs to be adjusted, which may influence the expression and selection of gametophytic genes. Gene dosage compensation occurs in plants with sex chromosomes such as *Silene latifolia* (Muyle et al., 2012), but dosage compensation can also be achieved by differential methylation (Köhler et al., 2012; Lafon-Placette and Köhler, 2015) and therefore is not exclusively associated with sex chromosomes. Consistent with incomplete dosage compensation, particularly in rice and soybean, absolute expression levels are reduced during gametogenesis (S6 Fig). Plant species such as mosses, which have an extended generation of multicellular haploid gametophytes that differentiate into early vegetative and later reproductive stages allow to investigate the effects of haploidy and dosage compensation on transcriptome indices. Although only limited expression data for gametophytic stages of mosses are available, we tested whether young genes contribute to the gene age of gametophytic transcriptome using expression and genome data from *Physcomitrella patens*. The transcriptomic age index based on one gametophytic and two sporophytic stages (O'Donoghue et al., 2013) shows an increase during the haploid stage, which is consistent with the evolutionary bulge pattern (Fig. 4; Pairwise t-test; $P < 3.2 \times 10^{-10}$) and the wide distribution of this pattern in plants.

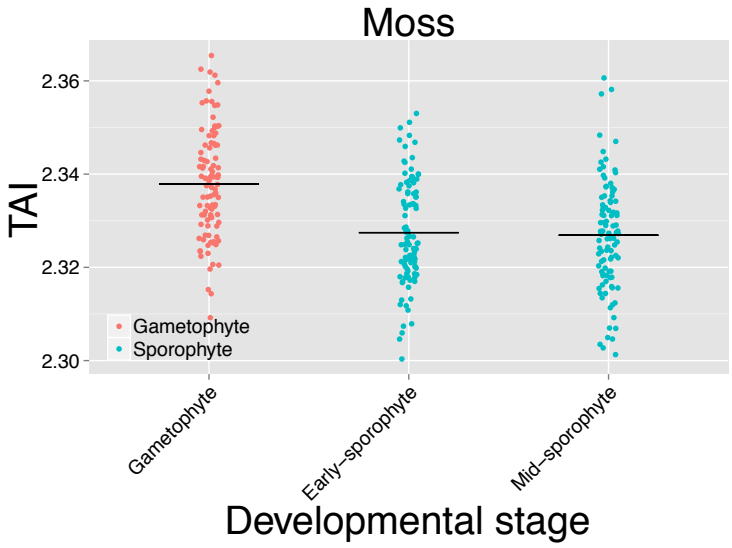


Figure 4: **Estimates of the transcriptomic age index (TAI) for three different developmental stages in the moss *Physcomitrella patens*.**

Influence of genomic conflicts

Evolutionary conflicts between female and male genes can drive rapid evolution as a consequence of female mate choice, sexual selection or differential expression of maternal and paternal genomes (León et al., 2014). Previously, we did not find evidence for co-evolutionary patterns of male and female reproductive tissues (Gossmann et al., 2014b), which may be influenced by a limited power

to identify female-specific genes because we focussed on sex-biased gene expression, i.e., transcripts enriched in male or female tissues, which led to a small sample of genes for the female tissue. The present global transcriptome analysis reduced this bias and we found an evolutionary bulge in both male and female tissues of *A. thaliana*. A rapid evolution of reproductive traits by selection on (possibly antagonistic) male-female interactions is consistent with this observation, although a direct effect is difficult to show with our data. Based on current knowledge, we doubt that such a selection-driven co-evolution is exclusive to gametophytic genes as pollen interactions also occur with sporophytic tissues because, for example, anther tissues contribute to protein composition of the pollen surface (Yang et al., 2007; Huang et al., 2011; Ling et al., 2015). Furthermore, the higher rate of evolution may be more influenced by a reduced level of protein-protein or other network interactions in the gametophyte compared to other tissues (Fig. 3), which imposes fewer constraints on rates of evolution. However, further molecular analyses of such interactions are required to test whether molecular interactions between male and female tissues for example during pollen tube outgrowth or seed development are involved in antagonistic coevolution and contribute to a bulge pattern.

Consequences of domestication on the evolutionary bulge

Two of the three species in our study are domesticated plants (rice and soybean). Patterns of polymorphism in domesticated species are affected by past domestication bottlenecks (Gossmann et al., 2010). Although domestication is often thought to be associated with a few 'key domestication' genes (Meyer and Purugganan, 2013), the global expression pattern of domesticated species may be substantially altered compared to wild relatives (e.g. (Rapp et al., 2010; Yoo and Wendel, 2014)). Moreover, the extent of the differences in global expression due to domestication may vary for different developmental stages. The (partial) correlations of expression intensity and polymorphism, p_N/p_S, did not support a bulge pattern, which suggest different types and strengths of selection in the recent evolution of rice and soybean (Table 2). However, the lack of support may represent an experimental bias because the polymorphism data of the two domesticated species are obtained from SNP arrays that are designed to genotype common variants, and likely miss rare variants, which leads to a downward bias of p_N/p_S. The impact of domestication on the transcriptomic profile during development in rice and soybean is currently unknown. Since the evolutionary bulge pattern is influenced by different processes in the three species (Fig. 2 and), domestication may explain some differences between the wild and the two crop species.

Variation of transcriptome indices in reproductive tissues of *A. thaliana*

Recent studies investigated the evolutionary rates of genes expressed in the male gametophyte of *A. thaliana* (Szövényi et al., 2013; Cui et al., 2015; Gossmann et al., 2014b; Harrison et al., 2015). Although interpretations of observed patterns differ, genes specifically expressed in the haploid stages tend to have higher rates of evolution as a consequence of relaxed purifying selection, if taking gene expression intensity and diversity are taken into account (Szövényi et al., 2013; Harrison et al.,

2015). Of note, sperm and pollen-specific genes evolve very differently (Arunkumar et al., 2013; Gossmann et al., 2014b). In our analysis, the pollen tube of *A. thaliana* showed lower TDI and TPI, but higher TAI values than the sperm cell (Fig. 1; see also (Cui et al., 2015)), which indicates that tissue-specific effects additionally influence the evolutionary bulge pattern. The expression weighted neutrality index (NI) differs between sperm and late pollen stages (Fig. 3) which shows a shift in the relative contribution of positive and negative selection and supports tissue-specific effects. A possible explanation is an enrichment of slightly deleterious mutations that are more effectively removed due to purifying selection. Unfortunately it is difficult to disentangle the extent of the different selective forces on a gene-by-gene basis with our data. As noted before, a focus on tissue-specific enriched genes represents a bias because these genes tend to show a narrow expression pattern and a high expression level. In plants, both factors correlate with the rate of molecular evolution, but in opposite directions (Slotte et al., 2011).

Male and female tissues differ in their degree of protein-protein interactions. The late pollen and the sperm cell stages are less interactive than other stages, which likely results from the single cell state of these tissues compared to the female gametophyte or sporophyte. In contrast to the pollen and sperm cell they consist of several cell types, which contribute to a higher rate of evolution in male tissue.

Hybridisation-based microarrays likely do not include the full set of expressed genes and tend to be biased towards better characterized and more conserved genes, which may lead to lower estimates of transcriptome indices. Potentially important genes such as lineage-specific or *de novo* genes are likely not fully covered by this analysis (Cui et al., 2015; Wu et al., 2014). Currently it is unclear to what extent the contribution of de novo genes differs between gametophytic and sporophytic stages, but RNAseq analysis of single cells will allow a much higher resolution of evolutionary patterns to disentangle possible causes of the bulge.

The evolutionary bulge is a stronger pattern than the transcriptomic hourglass

In comparison to the evolutionary bulge, the embryogenic hourglass seems to be a less robust pattern. We reproduced the hourglass in *A. thaliana*, but found little support for it in rice or soybean (Fig. 1). This may be caused by an incomplete sampling of embryonic stages in the latter two species, or from a technical issue in rice because the dissection of embryonic from surrounding tissue is difficult and prone to contamination. The latter issue may be reflected in the low variability of transcriptome indices during rice embryo development (S6 Fig). Further research is required to verify the generality of the hourglass model during plant embryogenesis using approaches such as single cell transcriptomics and RNASeq. The transcriptome indices during embryogenesis are not consistently lower than in other sporophytic tissues (Fig. 1 and S1 Fig), which indicates that embryogenic genes do not consistently evolve slower than in other tissues. The indices showed substantial variation among reproductive tissues but less variation in flower development or embryogenesis, which further illustrates that the evolutionary bulge is a stronger pattern than the hourglass in plants.

Material and Methods

Sequence data and software

We obtained the genome sequences of *A. thaliana* (Arabidopsis Genome Initiative, 2000), rice (*Oryza sativa* (International Rice Genome Sequencing Project, 2005)) and soybean (*Glycine max* (Schmutz et al., 2010)) from the plant genome database (Duvick et al., 2008) and the plant duplication database (Lee et al., 2013) along with their outgroups *Arabidopsis lyrata* (Hu et al., 2011), *Sorghum bicolor* (Paterson et al., 2009) and *Phaseolus vulgaris* (Schmutz et al., 2014), respectively. Polymorphism data were obtained from 80 *Arabidopsis thaliana* accessions (Cao et al., 2011). To identify coding SNP information for rice we used the Rice Haplotype Map Project Database (2nd Generation, `http://www.ncgr.ac.cn/RiceHap2/index.html`) and soybean we used SNP information deposited in SNPdb (Sherry et al., 2001) and extracted coding SNPs from the soybean genome annotation. We used R and Python scripts to conduct statistical analyses.

Gene expression data

Gene expression data were obtained for the three plants species from the PlexDB database (Dash et al., 2012) and GEO databases (Barrett et al., 2013). In particular, we focused on development stages preceding gametogenesis, during gametogenesis and embryogenic developments (Table 1 and S1 File). For each species, Robust Multi-array Analysis (RMA (Irizarry et al., 2003)) and invariant set (IS) methods were performed with the `affy` Bioconductor package to normalize all datasets simultaneously. Scatterplots of expression between replicates showed better results for RMA normalization (data not shown). Therefore, unless stated otherwise, expression data shown in this study are based on a normalisation across experiments using RMA with log_2 transformation. Since different laboratory conditions can affect expression patterns (Massonnet et al., 2010), we controlled for these effects in the *A. thaliana* data (Schmid et al., 2005) by removing datasets that were obtained from plants with different growth conditions before RNA extraction (S1 File). To check whether the differences in expression between experimental conditions were negligible compared to the differences between stages, we generated scatterplots for the mature pollen stage (S7 Fig) that was common to different experiments (Honys and Twell, 2004; Schmid et al., 2005; Borges et al., 2008; Wang et al., 2008). Scatterplots showed an expression profile that was similar between experiments with RMA normalization over all experiments and when normalized independently (Fig.s S7 Fig b and c) and also showed more variation between expression levels when compared to non-normalized and IS normalized expression (Fig.s S7 Fig a, d and e). Scatterplots between non-normalized experiments and between IS normalized experiments showed less variation in expression levels, but in general, the correlations between expression levels from different experiments were highly independent from the normalization method. For rice and soybean, all experiments were kept for normalization. Gene expression data for *Physcomitrella patens* for mature gaemtophyte, early and mid sporophyte (O'Donoghue et al., 2013) were downloaded from GEO (GSE32928) and the array and genome annotation (V1.6) was obtained from `www.cosmoss.org/physcome_project/wiki/Downloads`. In

this dataset, two samples per chip are hybridized, each with a different fluorescent dye (green Cy3 and red Cy5). Expression values were averaged across samples.

Evolutionary parameters

We obtained estimates for TAI (transcriptome age index), TDI (transcriptome divergence index) and TPI (transcriptome polymorphism index) for each developmental stage. A transcriptome index is the average of an evolutionary parameter like gene age (TAI), divergence (TDI) and diversity (TPI) that is weighted by the expression level of each gene. Confidence intervals were obtained by bootstrapping, using 100 sets of genes for each experimental stage. For estimates of gene age we followed the procedure of Quint et al.(Quint et al., 2012) which is based on the construction of a phylostratigraphic map. We used one-way BLAST (default parameters) hits against a sets of genomes that are assigned to a certain phylostrata and the BLAST hit to the most distant phylostratum defines the gene age (Albà and Castresana, 2007). The oldest genes have a gene age value of 1 and the highest gene age value was assigned to genes that are specific to a given species. Details about the hierarchical order, the genomes assigned to each phylostratum and number of genes with assigned gene age can be found in S8 Fig.

To calculate a per gene estimate of divergence we calculated d_N/d_S using pairwise alignments of homologous genes identified by INPARANOID from the whole genome comparison with its respective outgroup (Remm et al., 2001; Ostlund et al., 2010). We obtained per gene estimates of d_N/d_S $(= K_a/K_s)$ estimates for genes specific to species pairs with the KaKs_calculator (Zhang et al., 2006). We also introduce a new test statistic, the transcriptomic polymorphism index (TPI).

$$\mathrm{TPI}_s = \frac{\sum_{i=1}^{n} (P_N/N/((P_S+1)/S))e_{is}}{\sum_{i=1}^{n} e_{is}},$$

where s is the developmental stage, n the number of genes, e_{is} the expression intensity of gene i in developmental stage s, P_N and P_S the numbers of nonsynonymous and synonymous polymorphisms, respectively, and N and S are the numbers of nonsynonymous and synonymous sites, respectively. We used the ratio of nonsynonymous per site polymorphisms to synonymous per site polymorphism to estimate the distribution of fitness effects. Higher values of p_N/p_S reflect an excess of slightly deleterious mutations (Keightley and Eyre-Walker, 2007). For technical reasons we used $P_S + 1$ rather than P_S as suggested by Stoletzki and Eyre-Walker(Stoletzki and Eyre-Walker, 2011) because some genes have no synonymous polymorphisms and therefore would need to be excluded from the analysis which is biased (Stoletzki and Eyre-Walker, 2011). For compactness we refer to the term $P_N/N/((P_S + 1)/S)$ as p_N/p_S throughout the manuscript.

We tested whether transcriptome indices are different between stages by bootstrapping 100 samples of each index per stage and then performing a two-sample t-test to test for the differences in the means of bootstrapped values. If not noted otherwise, only the highest P-value in the comparison of stages is reported.

Modified variants of the transcriptome index

We calculated the weighted median transcriptome index of an evolutionary parameter x and assumed that $\sum_{i=1}^{n} e_i = 1$. The weighted median of the evolutionary index is then x_f with f such that

$$\sum_{i<f} e_i < 1/2 \text{ and } \sum_{i>f} e_i \leq 1/2.$$

The standardized transcriptome index that does not consider genes with a non-significant expression (S1 Fig) was calculated as follows:

$$T(x)I'_s = \frac{\sum_{i=1}^{n} x_i e_{is}}{\sum_{i=1}^{n} e_{is}} - \bar{x},$$

where $\bar{x}$ is the arithmetic mean of $x_1, \ldots, x_n$ and n the number of significantly expressed genes. We further obtained per gene neutrality index (NI) for *A. thaliana* as follows:

$$NI = \frac{d_S p_N}{d_N p_S}$$

where $p_S = (P_S + 1)/S$. The number of protein interactions for *A. thaliana* were obtained from the *Arabidopsis* interactome database (`ftp://ftp.arabidopsis.org/home/tair/Proteins/Protein_interaction_data/Interactome2.0/`).

Supporting Information

S1 Fig

Transcriptome indices from available tissue-specific expression datasets of *A. thaliana* calculated for significantly expressed genes. The transcriptome index was corrected for differences in gene number (see Materials and Methods). The number of significantly expressed genes is given in green, significant expression was addressed using P-values associated with the expression and corrected for multiple testing (FDR < 0.1).

S2 Fig

Transcriptome index using the weighted median for *A. thaliana*. TDI (upper panel) and TPI (lower panel) for *A. thaliana* based on a weighted median approach.

S3 Fig

TAI, TDI and TPI for *A. thaliana* based on alternative transformation of expressions signals. (a) log_{10} transformed expression data and (b) raw expression signal.

S4 Fig

Transcriptome indices for *A. thaliana* using independent normalisation. TAI (left panel), TDI (middle panel) and TPI (right panel) for *A. thaliana* based on independent normalisation of gene expression in reproductive and sporophytic tissues.

S5 Fig

Mean expression level, K_a/K_s (=d_N/d_S) and p_N/p_S at different developmental stages. (a, c, e) Mean expression level and confidence intervals calculated on sets of genes classified according to K_a/K_s values. (b, d, f) Mean expression level and confidence intervals calculated on sets of genes classified according to K_a/K_s values. Thresholds are defined as quartiles(Q).

S6 Fig

Expression means and variances for the three plant species and each developmental stage. Expression intensities are based on log_2 RMA normalized expression. Left panels: Mean expression. Right panels: Expression variance.

S7 Fig

Comparison of normalisation methods for the mature pollen stage from two different experiments for *A. thaliana*. (a) without normalization. (b) after RMA normalization over all experiments. (c) after RMA normalization on each experiment separately. (d) after invariant set normalization over all experiments. (e) after invariant set normalization on each experiment separately. Experiments (used as background for normalisation): Schmid et al. (COL-0), Honys et al. (Le), Borges et al. (transgenic), Wang et al. (COL-0, desiccated), Wang et al. (COL-0, hydrated)

S8 Fig

Phylostratum assignment of available genomes for gene age estimates. Phylogenetic tree of plant species used for the gene age estimates. Based on the phylogenetic tree each plant species can be assigned to a phylostratum which is defined as the last common ancestor with the reference species. Some species have the same last common ancestor with the reference species and are therefore assigned to the same phylostratum. The gene age of a node can be calculated as the distance to the root of the tree. BLAST hits based on default parameters with a genome in the most distant phylostratum define the gene age. Different approaches can be used to obtain the gene age, but have little influence on the estimate (Albà and Castresana, 2007).

S1 Table

Extended list of microarray datasets from *A. thaliana*.

S1 Text

Relationship between weighted mean, slope and Pearson's r.

S1 File

Detailed information about the gene expression datasets used for normalisation for *A. thaliana*, rice and soybean.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (EVOREP project SCHM1354/7-1) to KJS. The authors are grateful to Arne Jahn and Jonna Kulmuni for critical comments on the manuscript.

Author Contribution

TIG and KJS designed the study. TIG, DS and MWS analyzed the data. MAS contributed to analyze the data. TIG and KJS wrote the manuscript. All authors contributed to writing, editing and revising the manuscript.

References

Albà MM, Castresana J. 2007. On homology searches by protein Blast and the characterization of the age of genes. *BMC Evol Biol.* 7:53.

Alvarez-Ponce D, Fares MA. 2012. Evolutionary rate and duplicability in the *Arabidopsis thaliana* protein-protein interaction network. *Genome Biol Evol.* 4:1263–1274.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature.* 408:796–815.

Arunkumar R, Josephs EB, Williamson RJ, Wright SI. 2013. Pollen-specific, but not sperm-specific, genes show stronger purifying selection and higher rates of positive selection than sporophytic genes in *Capsella grandiflora*. *Mol Biol Evol.* 30:2475–2486.

Baack E, Melo MC, Rieseberg LH, Ortiz-Barrientos D. 2015. The origins of reproductive isolation in plants. *New Phytol.* .

Barrett SC. 2002. The evolution of plant sexual diversity. *Nature Reviews Genetics.* 3:274–284.

Barrett T, Wilhite SE, Ledoux P, et al. (17 co-authors). 2013. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.* 41:D991–D995.

Bomblies K, Yant L, Laitinen RA, Kim ST, Hollister JD, Warthmann N, Fitz J, Weigel D. 2010. Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. *PLoS Genet*. 6:e1000890.

Borges F, Gomes G, Gardner R, Moreno N, McCormick S, Feijó JA, Becker JD. 2008. Comparative transcriptomics of *Arabidopsis* sperm cells. *Plant Physiol*. 148:1168–1181.

Cai JJ, Petrov DA. 2010. Relaxed purifying selection and possibly high rate of adaptation in primate lineage-specific genes. *Genome Biol Evol*. 2:393–409.

Cao J, Schneeberger K, Ossowski S, et al. (17 co-authors). 2011. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet*. 43:956–963.

Cui X, Lv Y, Chen M, Nikoloski Z, Twell D, Zhang D. 2015. Young Genes out of the Male: An Insight from Evolutionary Age Analysis of the Pollen Transcriptome. *Mol Plant*. 8:935–945.

Dash S, Hemert JV, Hong L, Wise RP, Dickerson JA. 2012. PLEXdb: gene expression resources for plants and plant pathogens. *Nucleic Acids Res*. 40:D1194–D1201.

Domazet-Lošo T, Tautz D. 2010. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature*. 468:815–818.

Drost HG, Gabel A, Grosse I, Quint M. 2015. Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. *Mol Biol Evol*. 32:1221–1231.

Duboule D. 1994. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Dev Suppl*. pp. 135–142.

Duvick J, Fu A, Muppirala U, Sabharwal M, Wilkerson MD, Lawrence CJ, Lushbough C, Brendel V. 2008. PlantGDB: a resource for comparative plant genomics. *Nucleic Acids Res*. 36:D959–D965.

Fujita M, Horiuchi Y, Ueda Y, et al. (18 co-authors). 2010. Rice expression atlas in reproductive development. *Plant Cell Physiol*. 51:2060–2081.

Gao LL, Xue HW. 2012. Global analysis of expression profiles of rice receptor-like kinase genes. *Mol Plant*. 5:143–153.

Gossmann TI, Santure AW, Sheldon BC, Slate J, Zeng K. 2014a. Highly variable recombinational landscape modulates efficacy of natural selection in birds. *Genome Biol Evol*. 6:2061–2075.

Gossmann TI, Schmid MW, Grossniklaus U, Schmid KJ. 2014b. Selection-driven evolution of sex-biased genes is consistent with sexual selection in *Arabidopsis thaliana*. *Mol Biol Evol*. 31:574–583.

Gossmann TI, Song BH, Windsor AJ, Mitchell-Olds T, Dixon CJ, Kapralov MV, Filatov DA, Eyre-Walker A. 2010. Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Mol Biol Evol*. 27:1822–1832.

Grossniklaus U. 2011. Plant germline development: a tale of cross-talk, signaling, and cellular interactions. *Sex Plant Reprod.* 24:91–95.

Haerizadeh F, Wong CE, Bhalla PL, Gresshoff PM, Singh MB. 2009. Genomic expression profiling of mature soybean (*Glycine max*) pollen. *BMC Plant Biol.* 9:25.

Harrison MC, Mallon EB, Twell D, Hammond RL. 2015. Pollen-specific genes accumulate more deleterious mutations than sporophytic genes under relaxed purifying selection in *Arabidopsis thaliana*. *bioRxiv*. .

Honys D, Twell D. 2004. Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. *Genome Biol.* 5:R85.

Hu TT, Pattyn P, Bakker EG, et al. (30 co-authors). 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat Genet.* 43:476–481.

Huang MD, Hsing YIC, Huang AHC. 2011. Transcriptomes of the anther sporophyte: availability and uses. *Plant Cell Physiol.* 52:1459–1466.

International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. *Nature.* 436:793–800.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics.* 4:249–264.

Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, Ohler U, Bergman CM, Tomancak P. 2010. Gene expression divergence recapitulates the developmental hourglass model. *Nature.* 468:811–814.

Keightley PD, Eyre-Walker A. 2007. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics.* 177:2251–2261.

Köhler C, Wolff P, Spillane C. 2012. Epigenetic mechanisms underlying genomic imprinting in plants. *Annu Rev Plant Biol.* 63:331–352.

Lafon-Placette C, Köhler C. 2015. Epigenetic mechanisms of postzygotic reproductive isolation in plants. *Curr Opin Plant Biol.* 23:39–44.

Le BH, Cheng C, Bui AQ, et al. (15 co-authors). 2010. Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. *Proc Natl Acad Sci U S A.* 107:8063–8070.

Le BH, Wagmaister JA, Kawashima T, Bui AQ, Harada JJ, Goldberg RB. 2007. Using genomics to study legume seed development. *Plant Physiol.* 144:562–574.

Lee TH, Tang H, Wang X, Paterson AH. 2013. PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res.* 41:D1152–D1158.

León GDTD, García-Aguilar M, Gillmor CS. 2014. Non-equivalent contributions of maternal and paternal genomes to early plant embryogenesis. *Nature.* 514:624–627.

Ling S, Chen C, Wang Y, Sun X, Lu Z, Ouyang Y, Yao J. 2015. The mature anther-preferentially expressed genes are associated with pollen fertility, pollen germination and anther dehiscence in rice. *BMC Genomics.* 16:101.

Lipinska A, Cormier A, Luthringer R, Peters AF, Corre E, Gachon CMM, Cock JM, Coelho SM. 2015. Sexual dimorphism and the evolution of sex-biased gene expression in the brown alga *ectocarpus. Mol Biol Evol.* 32:1581–1597.

Massonnet C, Vile D, Fabre J, et al. (27 co-authors). 2010. Probing the reproducibility of leaf growth and molecular phenotypes: a comparison of three *Arabidopsis* accessions cultivated in ten laboratories. *Plant Physiol.* 152:2142–2157.

Meyer RS, Purugganan MD. 2013. Evolution of crop species: genetics of domestication and diversification. *Nat Rev Genet.* 14:840–852.

Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GAB. 2012. Rapid de novo evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLoS Biol.* 10:e1001308.

Nasrallah ME, Liu P, Nasrallah JB. 2002. Generation of self-incompatible *Arabidopsis thaliana* by transfer of two S locus genes from *A. lyrata. Science.* 297:247–249.

O'Donoghue MT, Chater C, Wallace S, Gray JE, Beerling DJ, Fleming AJ. 2013. Genome-wide transcriptomic analysis of the sporophyte of the moss *Physcomitrella patens. J Exp Bot.* 64:3567–3581.

Ostlund G, Schmitt T, Forslund K, Köstler T, Messina DN, Roopra S, Frings O, Sonnhammer ELL. 2010. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.* 38:D196–D203.

Otto SP, Gerstein AC. 2008. The evolution of haploidy and diploidy. *Curr Biol.* 18:R1121–R1124.

Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet.* 14:83–87.

Paterson AH, Bowers JE, Bruggmann R, et al. (45 co-authors). 2009. The Sorghum bicolor genome and the diversification of grasses. *Nature.* 457:551–556.

Piasecka B, Lichocki P, Moretti S, Bergmann S, Robinson-Rechavi M. 2013. The hourglass and the early conservation models–co-existing patterns of developmental constraints in vertebrates. *PLoS Genet.* 9:e1003476.

Pina C, Pinto F, Feijó JA, Becker JD. 2005. Gene family analysis of the *Arabidopsis* pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. *Plant Physiol.* 138:744–756.

Quint M, Drost HG, Gabel A, Ullrich KK, Bönn M, Grosse I. 2012. A transcriptomic hourglass in plant embryogenesis. *Nature.* 490:98–101.

Rapp RA, Haigler CH, Flagel L, Hovav RH, Udall JA, Wendel JF. 2010. Gene expression in developing fibres of Upland cotton (*Gossypium hirsutum L.*) was massively altered by domestication. *BMC Biol.* 8:139.

Rattray A, Santoyo G, Shafer B, Strathern JN. 2015. Elevated mutation rate during meiosis in *Saccharomyces cerevisiae*. *PLoS Genet.* 11:e1004910.

Remm M, Storm CE, Sonnhammer EL. 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol.* 314:1041–1052.

Schein M, Yang Z, Mitchell-Olds T, Schmid KJ. 2004. Rapid evolution of a pollen-specific oleosin-like gene family from *Arabidopsis thaliana* and closely related species. *Mol Biol Evol.* 21:659–669.

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

Schmid MW, Schmidt A, Klostermeier UC, Barann M, Rosenstiel P, Grossniklaus U. 2012. A powerful method for transcriptional profiling of specific cell types in eukaryotes: laser-assisted microdissection and RNA sequencing. *PLoS One.* 7:e29685.

Schmidt A, Wuest SE, Vijverberg K, Baroux C, Kleen D, Grossniklaus U. 2011. Transcriptome analysis of the *Arabidopsis* megaspore mother cell uncovers the importance of RNA helicases for plant germline development. *PLoS Biol.* 9:e1001155.

Schmutz J, Cannon SB, Schlueter J, et al. (45 co-authors). 2010. Genome sequence of the palaeopolyploid soybean. *Nature.* 463:178–183.

Schmutz J, McClean PE, Mamidi S, et al. (41 co-authors). 2014. A reference genome for common bean and genome-wide analysis of dual domestications. *Nat Genet.* 46:707–713.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29:308–311.

Slotte T, Bataillon T, Hansen TT, Onge KS, Wright SI, Schierup MH. 2011. Genomic determinants of protein evolution and polymorphism in *Arabidopsis*. *Genome Biol Evol.* 3:1210–1219.

Spillane C, Schmid KJ, Laoueillé-Duprat S, Pien S, Escobar-Restrepo JM, Baroux C, Gagliardini V, Page DR, Wolfe KH, Grossniklaus U. 2007. Positive darwinian selection at the imprinted MEDEA locus in plants. *Nature.* 448:349–352.

Stoletzki N, Eyre-Walker A. 2011. Estimation of the neutrality index. *Mol Biol Evol.* 28:63–70.

Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. *Nat Rev Genet.* 3:137–144.

Szövényi P, Devos N, Weston DJ, Yang X, Hock Z, Shaw JA, Shimizu KK, McDaniel SF, Wagner A. 2014. Efficient purging of deleterious mutations in plants with haploid selfing. *Genome Biol Evol.* 6:1238–1252.

Szövényi P, Ricca M, Hock Z, Shaw JA, Shimizu KK, Wagner A. 2013. Selection is no more efficient in haploid than in diploid life stages of an angiosperm and a moss. *Mol Biol Evol.* 30:1929–1939.

Tang C, Toomajian C, Sherman-Broyles S, Plagnol V, Guo YL, Hu TT, Clark RM, Nasrallah JB, Weigel D, Nordborg M. 2007. The evolution of selfing in *Arabidopsis thaliana*. *Science.* 317:1070–1072.

Wang Y, Zhang WZ, Song LF, Zou JJ, Su Z, Wu WH. 2008. Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis*. *Plant Physiol.* 148:1201–1211.

Wei LQ, Xu WY, Deng ZY, Su Z, Xue Y, Wang T. 2010. Genome-scale analysis and comparison of gene expression profiles in developing and germinated pollen in *Oryza sativa*. *BMC Genomics.* 11:338.

Wright SI, Kalisz S, Slotte T. 2013. Evolutionary consequences of self-fertilization in plants. *Proc Biol Sci.* 280:20130133.

Wu DD, Wang X, Li Y, Zeng L, Irwin DM, Zhang YP. 2014. "Out of pollen" hypothesis for origin of new genes in flowering plants: study from *Arabidopsis thaliana*. *Genome Biol Evol.* 6:2822–2829.

Wuest SE, Vijverberg K, Schmidt A, Weiss M, Gheyselinck J, Lohr M, Wellmer F, Rahnenführer J, von Mering C, Grossniklaus U. 2010. *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. *Curr Biol.* 20:506–512.

Xiao L, Wang H, Wan P, Kuang T, He Y. 2011. Genome-wide transcriptome analysis of gametophyte development in *Physcomitrella patens*. *BMC Plant Biol.* 11:177.

Yang C, Vizcay-Barrena G, Conner K, Wilson ZA. 2007. MALE STERILITY1 is required for tapetal development and pollen wall biosynthesis. *Plant Cell.* 19:3530–3548.

Yoo MJ, Wendel JF. 2014. Comparative evolutionary and developmental dynamics of the cotton (*Gossypium hirsutum*) fiber transcriptome. *PLoS Genet.* 10:e1004073.

Zhang Z, Li J, Zhao XQ, Wang J, Wong GKS, Yu J. 2006. KaKs Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics.* 4:259–263.

Zuber H, Davidian JC, Aubert G, et al. (11 co-authors). 2010. The seed composition of *Arabidopsis* mutants for the group 3 sulfate transporters indicates a role in sulfate translocation within developing seeds. *Plant Physiol.* 154:913–926.

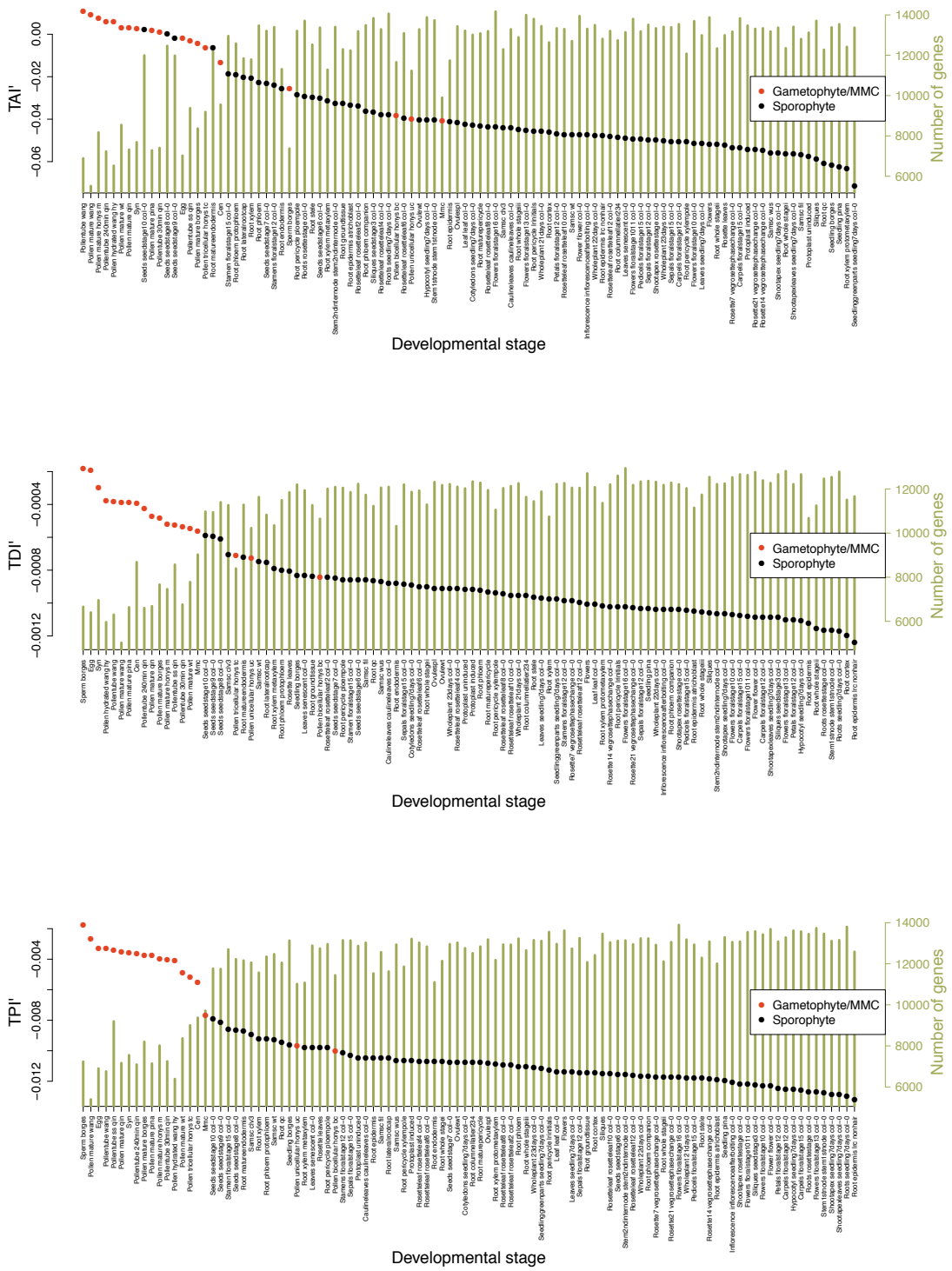


Figure S1: **Transcriptome indices from available tissue specific expression datasets of *A. thaliana* calculated for significantly expressed genes.** Note, the transcriptome index was corrected for differences in gene number (see materials and methods). The number of significantly expressed genes is given in green, significant expression was addressed using P-values associated with the expression and corrected for multiple testing (FDR < 0.1).

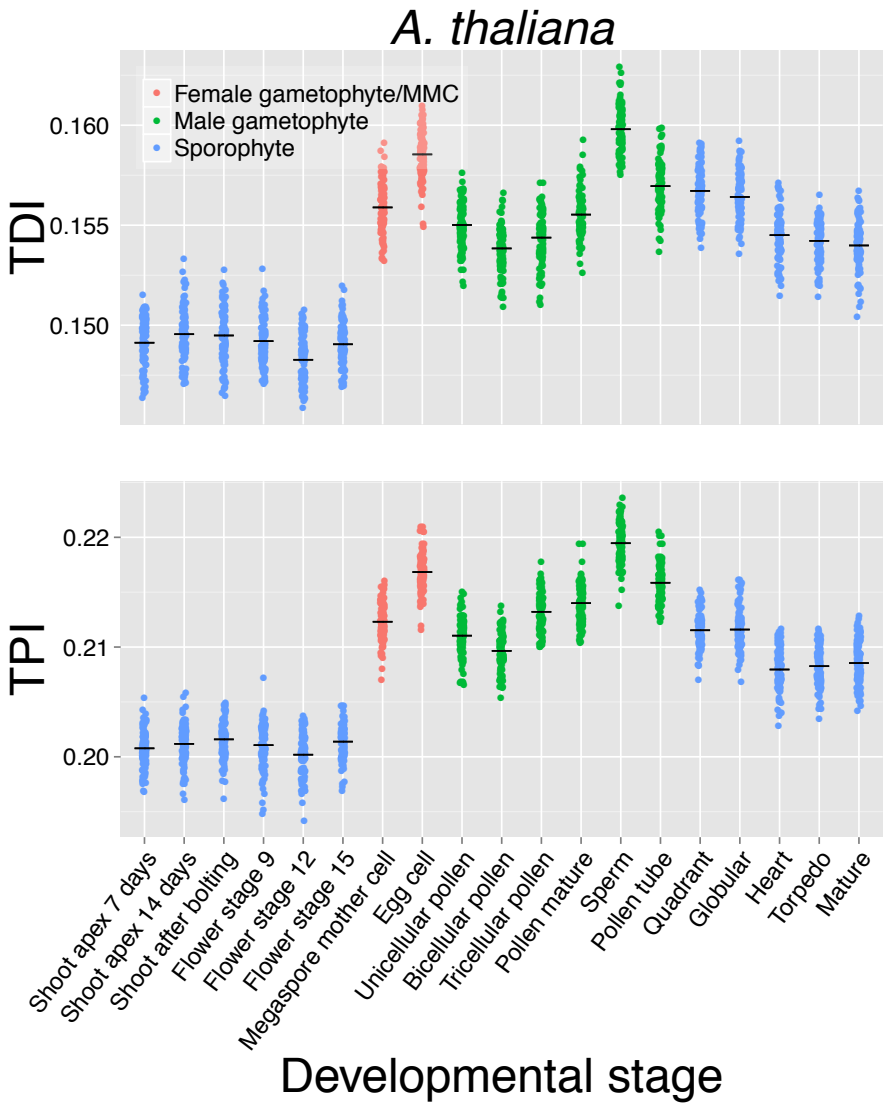
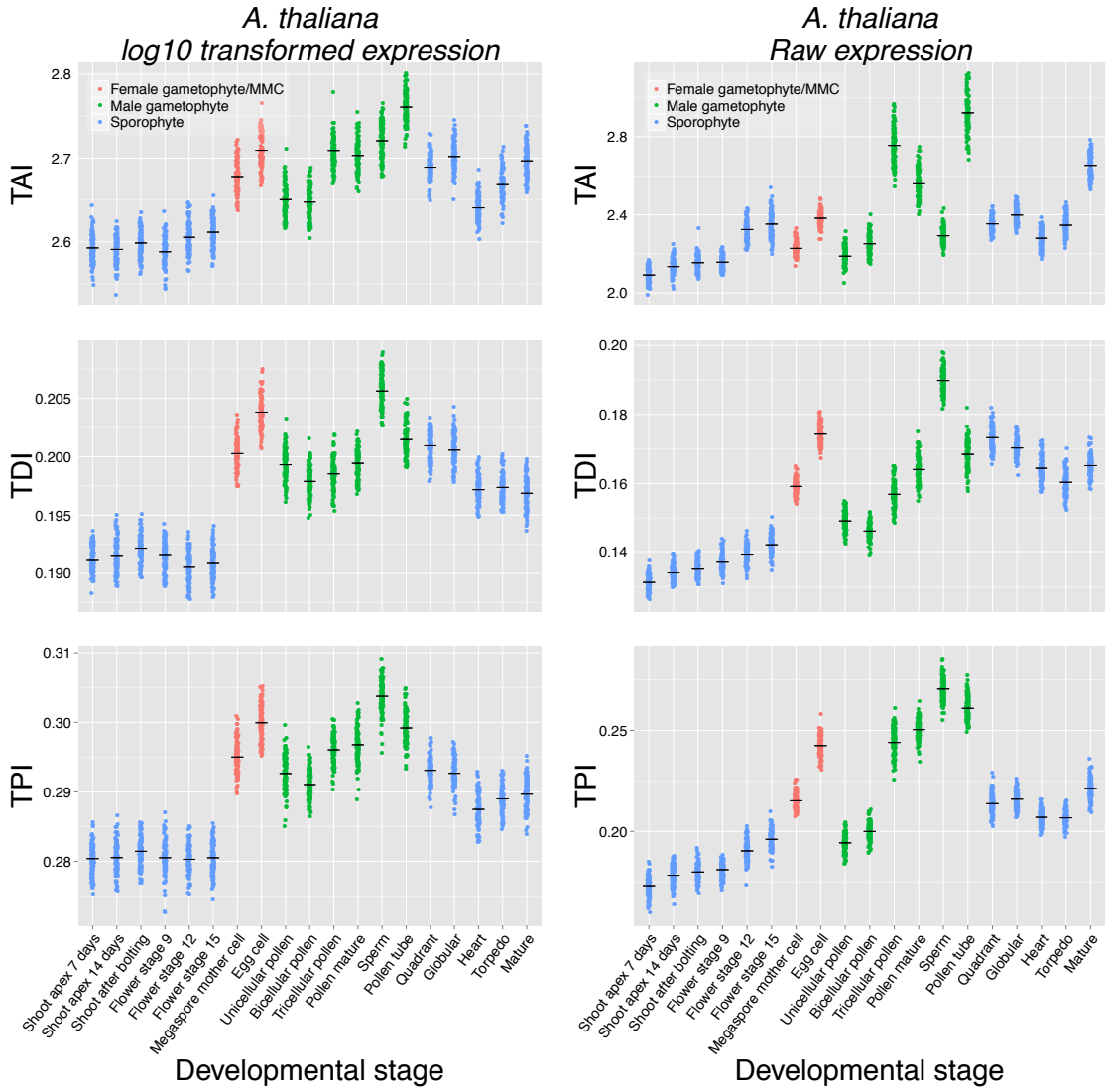


Figure S2: **Transcriptome index using the weighted median for *A. thaliana*.** TDI (upper panel) and TPI (lower panel) for *A. thaliana* based on a weighted median approach.



(a) (b)

Figure S3: **TAI, TDI and TPI (right panel) for *A. thaliana* based on alternative transformation of expressions signals.** (a) log_{10} transformed expression data and (b) raw expression signal.

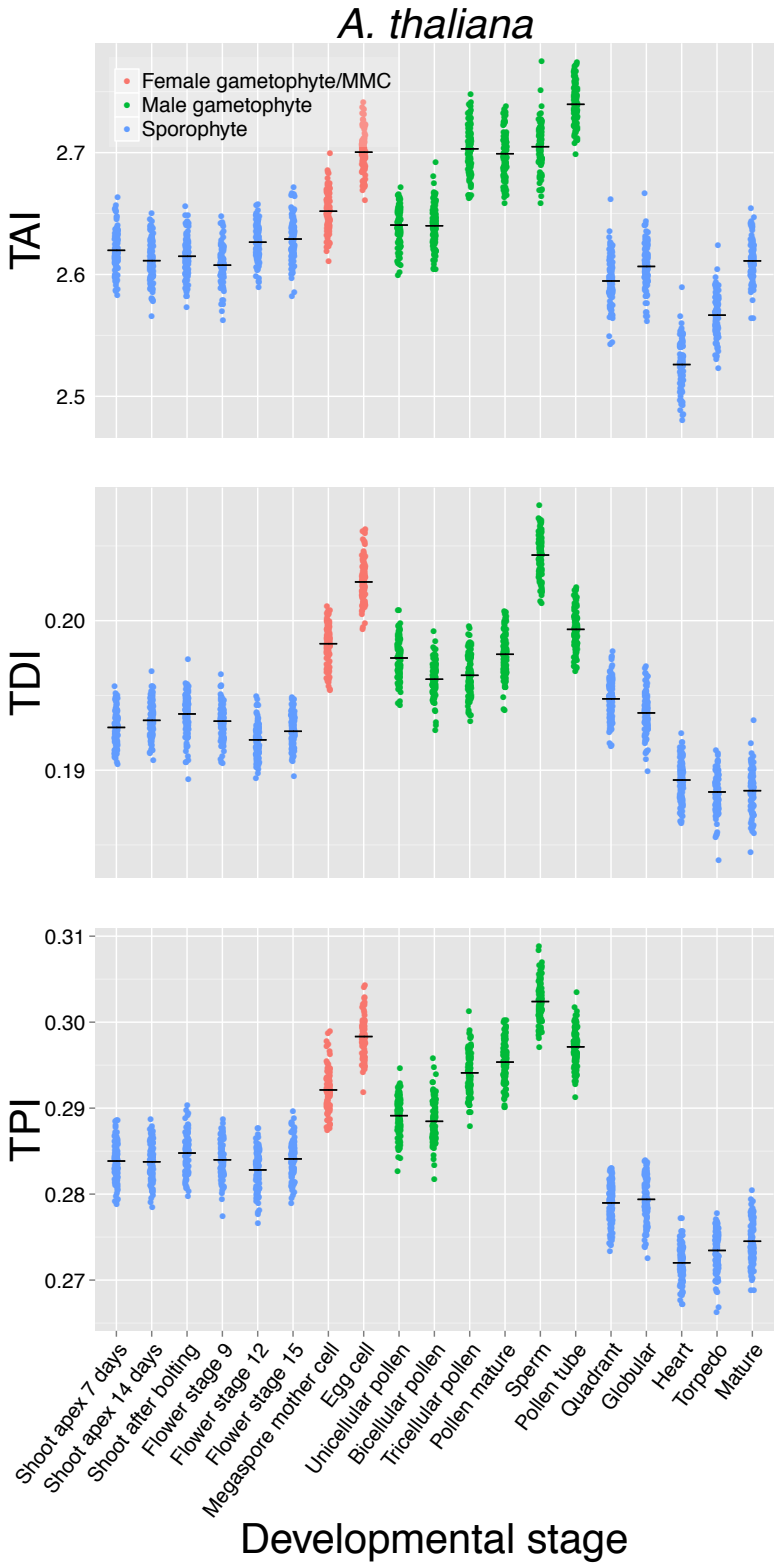


Figure S4: **Transcriptome indices for *A. thaliana* using independent normalisation.** TAI (left panel), TDI (middle panel) and TPI (right panel) for *A. thaliana* based on independent normalisation of gene expression in reproductive and sporophytic tissues.

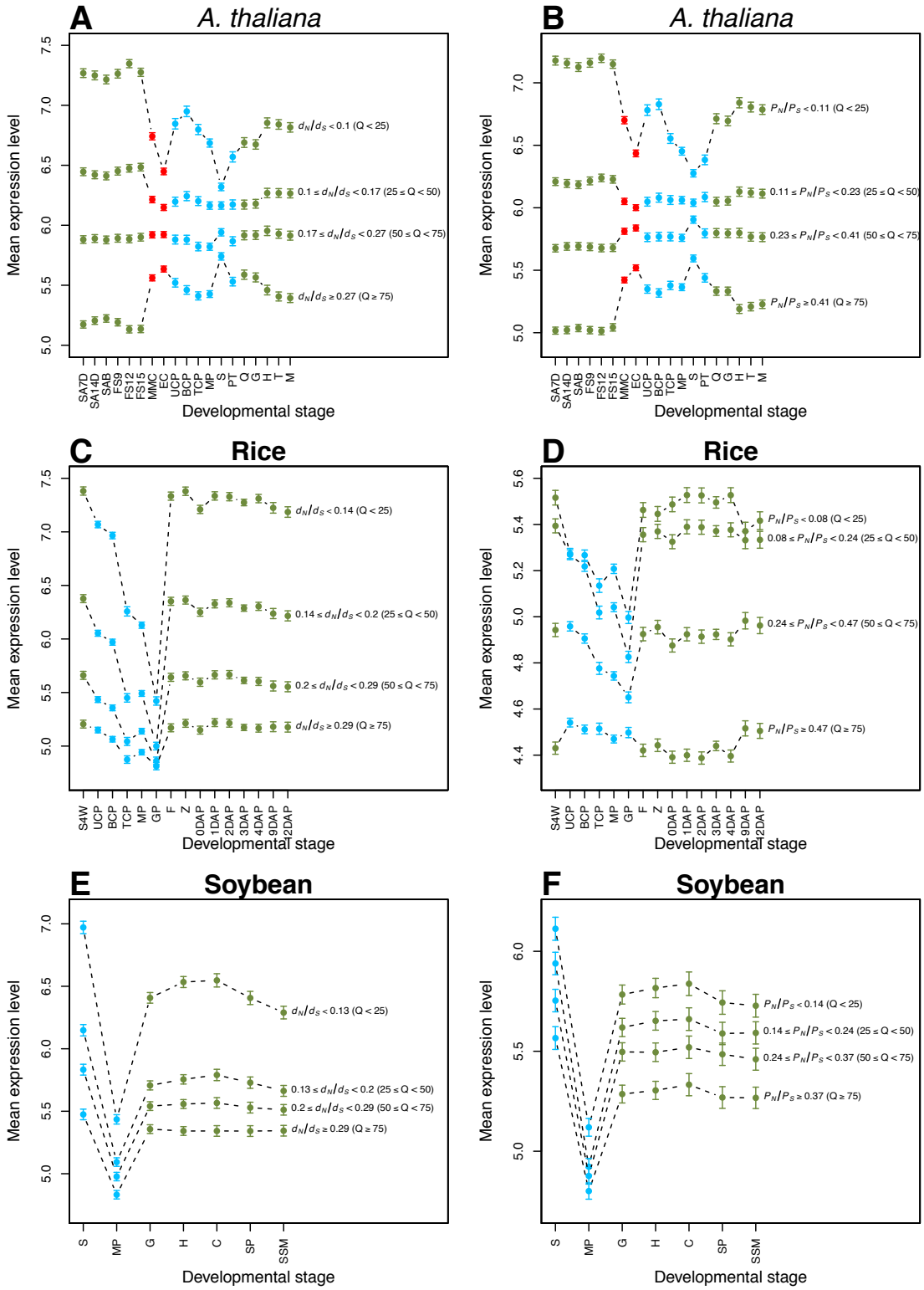


Figure S5: **Mean expression level, K_a/K_s (=d_N/d_S) and p_N/p_S at different developmental stages.** (a, c, e) Mean expression level and confidence intervals calculated on sets of genes classified according to K_a/K_s values. (b, d, f) Mean expression level and confidence intervals calculated on sets of genes classified according to Ka/Ks values. Thresholds are defined as quartiles(Q).

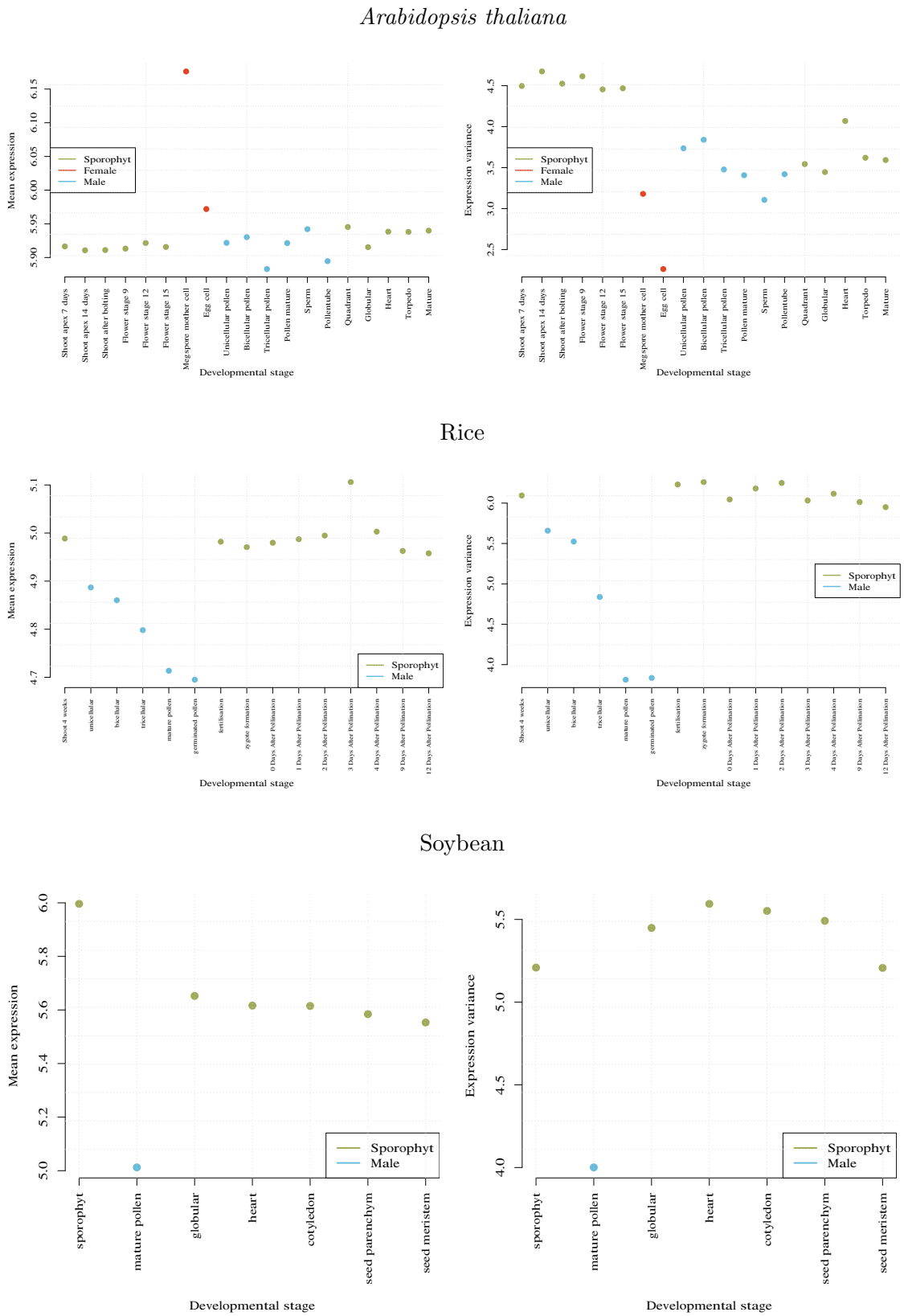


Figure S6: **Expression means and variances for the three plant species and each developmental stage.** Expression intensities are based on log_2 RMA normalized expression. Left panels: Mean expression. Right panels: Expression variance.

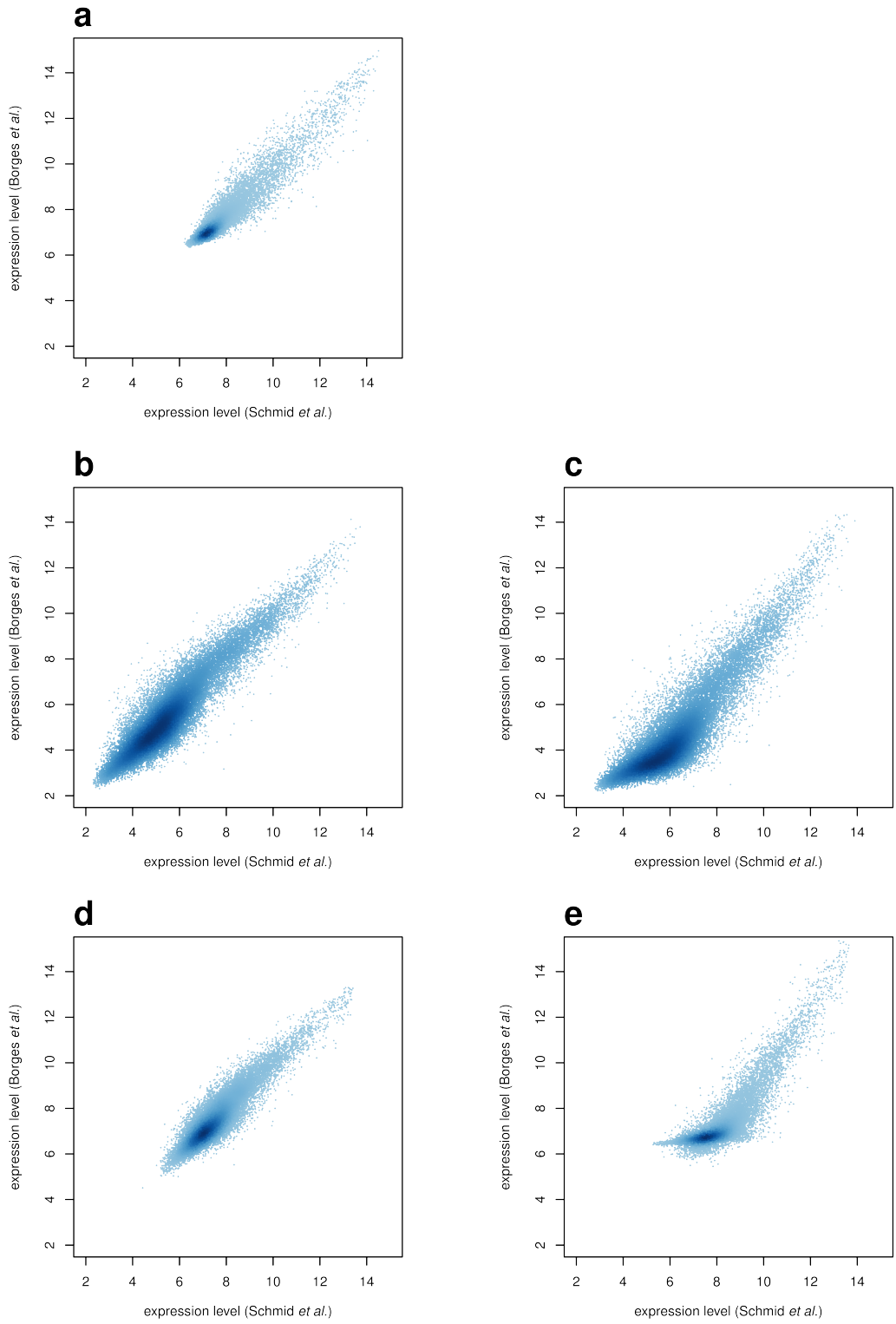
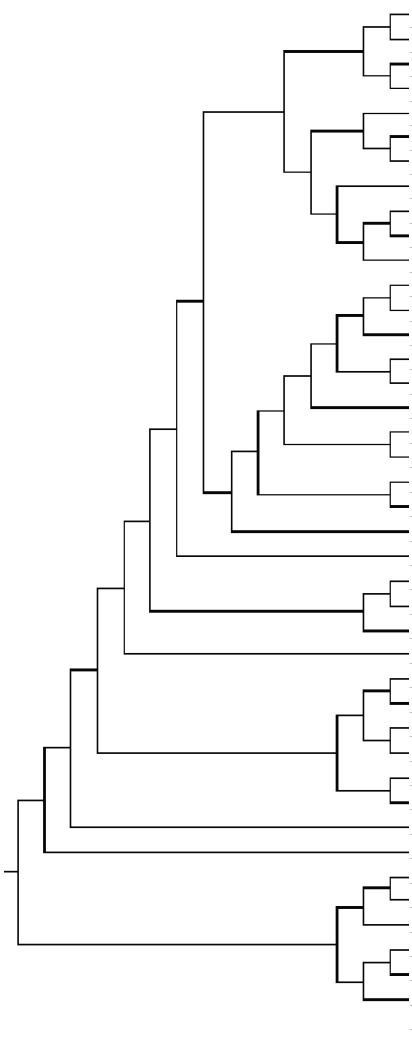


Figure S7: **Comparison of normalisation methods for the mature pollen stage from two different experiments for *A. thaliana*.** (a) without normalization. (b) after RMA normalization over all experiments. (c) after RMA normalization on each experiment separately. (d) after invariant set normalization over all experiments. (e) after invariant set normalization on each experiment separately. Experiments (used as background for normalisation): Schmid et al. (COL-0), Honys et al. (Le), Borges et al. (transgenic), Wang et al. (COL-0, desiccated), Wang et al. (COL-0, hydrated)

Assignment of genomes for phylostrata (gene age) analysis



	At	Os	Gm	Pp
Manihot esculenta	9	6	11	4
Ricinus communis	9	6	11	4
Linum usitatissimum	9	6	11	4
Populus trichocarpa	9	6	11	4
Medicago truncatula	9	6	13	4
Phaseolus vulgaris	9	6	14	4
Glycine max	9	6	15	4
Cucumis sativus	9	6	12	4
Prunus persica	9	6	12	4
Malus domestica	9	6	12	4
Fragaria vesca	9	6	12	4
Arabidopsis thaliana	13	6	10	4
Arabidopsis lyrata	12	6	10	4
Capsella rubella	11	6	10	4
Brassica rapa	11	6	10	4
Thellungiella halophila	11	6	10	4
Carica papaya	11	6	10	4
Gossypium raimondii	11	6	10	4
Theobroma cacao	11	6	10	4
Citrus sinensis	10	6	10	4
Citrus clementina	10	6	10	4
Eucalyptus grandis	9	6	10	4
Vitis vinifera	9	6	9	4
Solanum tuberosum	8	6	8	4
Solanom lycopersicum	8	6	8	4
Mimulus guttatus	8	6	8	4
Aquilegia coerulea	7	6	7	4
Sorghum bicolor	6	7	6	4
Zea mays	6	7	6	4
Setaria italica	6	7	6	4
Panicum virgatum	6	7	6	4
Oryza sativa	6	9	6	4
Brachypodium distachyon	6	8	6	4
Selaginella moellendorffii	5	5	5	4
Physcomitrella patens	4	4	4	5
Chlamydomonas reinhardtii	3	3	3	3
Volvox carteri	3	3	3	3
Coccomyxa subellipsoidea	3	3	3	3
Micromonas pusilla CCMP1545	3	3	3	3
Micromonas pusilla RCC299	3	3	3	3
Ostreococcus lucimarinus	3	3	3	3
Other Eukaryota	2	2	2	2
Archaea and Bacteria	1	1	1	1

Number of genes per phylostratum

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Arabidopsis	10146	5939	1441	3351	483	1611	635	332	311	31	1370	288	1235		
Rice	15295	8661	2263	4765	762	3406	3236	170	10318						
Soybean	21446	14251	4340	10414	1629	5015	2072	819	292	512	175	576	409	224	4036
Moss	12955	7015	1939	3916	12532										

Note: The assignment of phylostrata for *A. thaliana* was done according to Quint et al (2012) to keep consistency

Abbreviations: At – *Arabidopsis thaliana*; Os – *Oryza sativa*; Gm – *Glycine max* ; Pp - *Physcomitrella patens*

Figure S8: **Phylostratum assignment of available genomes for gene age estimates.** Phylogenetic tree of plant species used for the gene age estimates. Based on the phylogenetic tree each plant species can be assigned to a phylostratum which is defined as the last common ancestor with the reference species. Some species have the same last common ancestor with the reference species and are therefore assigned to the same phylostratum. The gene age of a node can be calculated as the distance to the root of the tree. BLAST hits based on default parameters with a genome in the most distant phylostratum define the gene age. Different approaches can be used to obtain the gene age, but have little influence on the estimate (Albà and Castresana, 2007).

Supporting Table S1: Extended microarray dataset for *A. thaliana*

Data used for microarray analysis.

accession/file	sample
EarlyHE1ATH1 [1]	OvuleWt1
EarlyHE2ATH1 [1]	OvuleWt2
EarlyHE3ATH1 [1]	OvuleWt3
EarlysplHO1ATH1 [1]	OvuleSpl1
EarlysplHO2ATH1 [1]	OvuleSpl2
EarlysplHO3ATH1 [1]	OvuleSpl3
Eggs1 [2]	Egg1
Eggs4 [2]	Egg4
Eggs5 [2]	Egg5
CC1 [2]	Cen1
CC2 [2]	Cen2
CC3 [2]	Cen3
Synergids1 [2]	Syn1
Synergids2 [2]	Syn2
Synergids3 [2]	Syn3
MMC1 [3]	MMC1
MMC2 [3]	MMC2
MMC3 [3]	MMC3
MMC4 [3]	MMC4
At_flowers_I4_ATH1_IGC1_JDB [4]	flowers1
At_flowers_I5_ATH1_IGC1_JDB [4]	flowers2
At_leaves_H5_ATH1_IGC1_JDB [4]	rosette_leaves1
At_leaves_H6_ATH1_IGC1_JDB [4]	rosette_leaves2
At_pollenGrain_J8_ATH1_IGC1_JDB [4]	pollen_mature_pina1
At_pollenGrain_J9_new_ATH1_IGC1_JDB [4]	pollen_mature_pina2
At_seedling_K2_ATH1_IGC1_JDB [4]	seedling_pina1
At_seedling_K3_ATH1_IGC1_JDB [4]	seedling_pina2
At_siliques_I10_ATH1_IGC1_JDB [4]	siliques1
At_siliques_I9_ATH1_IGC1_JDB [4]	siliques2
GSM142734 [5]	pollen_unicellular_honys_uc1
GSM142735 [5]	pollen_bicellular_honys_bc1
GSM142736 [5]	pollen_tricellular_honys_tc1
GSM142737 [5]	pollen_unicellular_honys_uc2
GSM142738 [5]	pollen_bicellular_honys_bc2
GSM142739 [5]	pollen_tricellular_honys_tc2
GSM142740 [5]	pollen_mature_honys_m1
At_Pollen_Rep1_ATH1_IGC_FB [6]	pollen_mature_borges_1
At_Pollen_Rep2_ATH1_IGC_FB [6]	pollen_mature_borges_2
At_Pollen_Rep3_ATH1_IGC_FB [6]	pollen_mature_borges_3
At_Sperm_Rep1_ATH1_IGC_FB [6]	sperm_borges_1
At_Sperm_Rep2_ATH1_IGC_FB [6]	sperm_borges_2
At_Sperm_Rep3_ATH1_IGC_FB [6]	sperm_borges_3
At_Seedling_Rep1_ATH1_IGC_FB [6]	seedling_borges_1
At_Seedling_Rep2_ATH1_IGC_FB [6]	seedling_borges_2
At_Seedling_Rep3_ATH1_IGC_FB [6]	seedling_borges_3
GSM154503 [7]	pollen_mature_wang_1
GSM154504 [7]	pollen_mature_wang_2
GSM154505 [7]	pollen_hydrated_wang_hy1
GSM154506 [7]	pollen_hydrated_wang_hy2
GSM154507 [7]	pollentube_wang_1
GSM154508 [7]	pollentube_wang_2
GSM433634 [8]	pollen_mature_qin_1
GSM433635 [8]	pollen_mature_qin_2
GSM433636 [8]	pollen_mature_qin_3
GSM433637 [8]	pollen_mature_qin_4

GSM433638 [8]	pollentube_30min_qin_1
GSM433639 [8]	pollentube_30min_qin_2
GSM433640 [8]	pollentube_30min_qin_3
GSM433641 [8]	pollentube_30min_qin_4
GSM433642 [8]	pollentube_240min_qin_1
GSM433643 [8]	pollentube_240min_qin_2
GSM433644 [8]	pollentube_240min_qin_3
GSM433645 [8]	pollentube_240min_qin_4
GSM433646 [8]	pollentube_SS_qin_1
GSM433647 [8]	pollentube_SS_qin_2
GSM433648 [8]	pollentube_SS_qin_3
GSM342135 [9]	SAMSC_wt_1
GSM342136 [9]	SAMSC_wt_2
GSM342137 [9]	SAMSC_wt_3
GSM342138 [9]	SAMSC_clv3_1
GSM342139 [9]	SAMSC_clv3_2
GSM342140 [9]	SAMSC_clv3_3
GSM342141 [9]	SAMSC_fil_1
GSM342142 [9]	SAMSC_fil_2
GSM342143 [9]	SAMSC_fil_3
GSM342144 [9]	protoplast_induced_1
GSM342145 [9]	protoplast_induced_2
GSM342146 [9]	protoplast_uninduced_1
GSM342147 [9]	protoplast_uninduced_2
GSM342148 [9]	SAMSC_wus_1
GSM342149 [9]	SAMSC_wus_2
GSM284384 [10, 11]	globularEmbryo1
GSM284385 [10, 11]	globularEmbryo2
GSM284386 [10, 11]	globularSuspensor1
GSM284387 [10, 11]	globularSuspensor2
GSM284388 [10, 11]	globularMicropylarEndosperm1
GSM284389 [10, 11]	globularMicropylarEndosperm2
GSM284390 [10, 11]	globularPeripheralEndosperm1
GSM284391 [10, 11]	globularPeripheralEndosperm2
GSM284392 [10, 11]	globularChalazalEndosperm1
GSM284393 [10, 11]	globularChalazalEndosperm2
GSM284394 [10, 11]	globularChalazalEndosperm3
GSM284395 [10, 11]	globularChalazalSeedCoat1
GSM284396 [10, 11]	globularChalazalSeedCoat2
GSM284397 [10, 11]	globularGeneralSeedCoat1
GSM284398 [10, 11]	globularGeneralSeedCoat2
GSM311273 [11]	pre-globularEmbryo1
GSM311274 [11]	pre-globularEmbryo2
GSM311275 [11]	pre-globularMicropylarEndosperm1
GSM311276 [11]	pre-globularMicropylarEndosperm2
GSM311277 [11]	pre-globularPeripheralEndosperm1
GSM311278 [11]	pre-globularPeripheralEndosperm2
GSM311279 [11]	pre-globularChalazalEndosperm1
GSM311280 [11]	pre-globularChalazalEndosperm2
GSM311281 [11]	pre-globularChalazalSeedCoat1
GSM311282 [11]	pre-globularChalazalSeedCoat2
GSM311283 [11]	pre-globularGeneralSeedCoat1
GSM311284 [11]	pre-globularGeneralSeedCoat2
GSM311285 [11]	pre-globularWholeSeed1
GSM311286 [11]	pre-globularWholeSeed2
GSM311287 [11]	linear-cotyledonEmbryo1
GSM311288 [11]	linear-cotyledonEmbryo2
GSM311289 [11]	linear-cotyledonCellularizedEndosperm1
GSM311290 [11]	linear-cotyledonCellularizedEndosperm2
GSM311291 [11]	linear-cotyledonChalazalEndosperm1
GSM311292 [11]	linear-cotyledonChalazalEndosperm2

GSM311293 [11]	linear-cotyledonChalazalSeedCoat1
GSM311294 [11]	linear-cotyledonChalazalSeedCoat2
GSM311295 [11]	linear-cotyledonGeneralSeedCoat1
GSM311296 [11]	linear-cotyledonGeneralSeedCoat2
GSM499418 [11]	linear-cotyledonWholeSeed1
GSM499419 [11]	linear-cotyledonWholeSeed2
GSM499420 [11]	linear-cotyledonMicropylarEndosperm1
GSM499421 [11]	linear-cotyledonMicropylarEndosperm2
GSM378645 [11]	heartEmbryo1
GSM378646 [11]	heartEmbryo2
GSM378647 [11]	heartMicropylarEndosperm1
GSM378648 [11]	heartMicropylarEndosperm2
GSM378649 [11]	heartPeripheralEndosperm1
GSM378650 [11]	heartPeripheralEndosperm2
GSM378651 [11]	heartChalazalEndosperm1
GSM378652 [11]	heartChalazalEndosperm2
GSM378653 [11]	heartChalazalEndosperm3
GSM378654 [11]	heartChalazalSeedCoat1
GSM378655 [11]	heartChalazalSeedCoat2
GSM378656 [11]	heartChalazalSeedCoat3
GSM378657 [11]	heartSeedCoat1
GSM378658 [11]	heartSeedCoat2
GSM378659 [11]	heartWholeSeed1
GSM378660 [11]	heartWholeSeed2
GSM378733 [11]	matureEmbryo1
GSM378734 [11]	matureEmbryo2
GSM378735 [11]	matureMicropylarEndosperm1
GSM378736 [11]	matureMicropylarEndosperm2
GSM378737 [11]	maturePeripheralEndosperm1
GSM378738 [11]	maturePeripheralEndosperm2
GSM378739 [11]	matureChalazalEndosperm1
GSM378740 [11]	matureChalazalEndosperm2
GSM378741 [11]	matureChalazalSeedCoat1
GSM378742 [11]	matureChalazalSeedCoat2
GSM378743 [11]	matureSeedCoat1
GSM378744 [11]	matureSeedCoat2
GSM378745 [11]	matureWholeSeed1
GSM378746 [11]	matureWholeSeed2
ATGE_12_A [12]	rosetteLeaf_rosetteLeaf2_Col-0_A
ATGE_12_B [12]	rosetteLeaf_rosetteLeaf2_Col-0_B
ATGE_12_C [12]	rosetteLeaf_rosetteLeaf2_Col-0_C
ATGE_13_A [12]	rosetteLeaf_rosetteLeaf4_Col-0_A
ATGE_13_B [12]	rosetteLeaf_rosetteLeaf4_Col-0_B
ATGE_13_C [12]	rosetteLeaf_rosetteLeaf4_Col-0_C
ATGE_14_A [12]	rosetteLeaf_rosetteLeaf6_Col-0_A
ATGE_14_B [12]	rosetteLeaf_rosetteLeaf6_Col-0_B
ATGE_14_C [12]	rosetteLeaf_rosetteLeaf6_Col-0_C
ATGE_15_A [12]	rosetteLeaf_rosetteLeaf8_Col-0_A
ATGE_15_B [12]	rosetteLeaf_rosetteLeaf8_Col-0_B
ATGE_15_C [12]	rosetteLeaf_rosetteLeaf8_Col-0_C
ATGE_16_A [12]	rosetteLeaf_rosetteLeaf10_Col-0_A
ATGE_16_B [12]	rosetteLeaf_rosetteLeaf10_Col-0_B
ATGE_16_C [12]	rosetteLeaf_rosetteLeaf10_Col-0_C
ATGE_17_A [12]	rosetteLeaf_rosetteLeaf12_Col-0_A
ATGE_17_B [12]	rosetteLeaf_rosetteLeaf12_Col-0_B
ATGE_17_C [12]	rosetteLeaf_rosetteLeaf12_Col-0_C
ATGE_1_A [12]	cotyledons_seedling7days_Col-0_A
ATGE_1_B [12]	cotyledons_seedling7days_Col-0_B
ATGE_1_C [12]	cotyledons_seedling7days_Col-0_C
ATGE_22_A [12]	wholePlant_21days_Col-0_A
ATGE_22_B [12]	wholePlant_21days_Col-0_B

ATGE_22_C [12]	wholePlant_21days_Col-0_C
ATGE_23_A [12]	wholePlant_22days_Col-0_A
ATGE_23_B [12]	wholePlant_22days_Col-0_B
ATGE_23_C [12]	wholePlant_22days_Col-0_C
ATGE_24_A [12]	wholePlant_23days_Col-0_A
ATGE_24_B [12]	wholePlant_23days_Col-0_B
ATGE_24_C [12]	wholePlant_23days_Col-0_C
ATGE_25_A [12]	leaves_senescent_Col-0_A
ATGE_25_B [12]	leaves_senescent_Col-0_B
ATGE_25_C [12]	leaves_senescent_Col-0_C
ATGE_26_A [12]	caulineLeaves_caulineLeaves_Col-0_A
ATGE_26_B [12]	caulineLeaves_caulineLeaves_Col-0_B
ATGE_26_C [12]	caulineLeaves_caulineLeaves_Col-0_C
ATGE_27_A [12]	stem2ndInternode_stem2ndInternode_Col-0_A
ATGE_27_B [12]	stem2ndInternode_stem2ndInternode_Col-0_B
ATGE_27_C [12]	stem2ndInternode_stem2ndInternode_Col-0_C
ATGE_28_A2 [12]	stem1stNode_stem1stNode_Col-0_A
ATGE_28_B2 [12]	stem1stNode_stem1stNode_Col-0_B
ATGE_28_C2 [12]	stem1stNode_stem1stNode_Col-0_C
ATGE_29_A2 [12]	inflorescence_inflorescenceAfterBolting_Col-0_A
ATGE_29_B2 [12]	inflorescence_inflorescenceAfterBolting_Col-0_B
ATGE_29_C2 [12]	inflorescence_inflorescenceAfterBolting_Col-0_C
ATGE_2_A [12]	hypocotyl_seedling7days_Col-0_A
ATGE_2_B [12]	hypocotyl_seedling7days_Col-0_B
ATGE_2_C [12]	hypocotyl_seedling7days_Col-0_C
ATGE_31_A2 [12]	flowers_floralStage10_Col-0_A
ATGE_31_B2 [12]	flowers_floralStage10_Col-0_B
ATGE_31_C2 [12]	flowers_floralStage10_Col-0_C
ATGE_32_A2 [12]	flowers_floralStage1011_Col-0_A
ATGE_32_B2 [12]	flowers_floralStage1011_Col-0_B
ATGE_32_C2 [12]	flowers_floralStage1011_Col-0_C
ATGE_33_A [12]	flowers_floralStage13_Col-0_A
ATGE_33_B [12]	flowers_floralStage13_Col-0_B
ATGE_33_C [12]	flowers_floralStage13_Col-0_C
ATGE_34_A [12]	sepals_floralStage12_Col-0_A
ATGE_34_B [12]	sepals_floralStage12_Col-0_B
ATGE_34_C [12]	sepals_floralStage12_Col-0_C
ATGE_35_A [12]	petals_floralStage12_Col-0_A
ATGE_35_B [12]	petals_floralStage12_Col-0_B
ATGE_35_C [12]	petals_floralStage12_Col-0_C
ATGE_36_A [12]	stamens_floralStage12_Col-0_A
ATGE_36_B [12]	stamens_floralStage12_Col-0_B
ATGE_36_C [12]	stamens_floralStage12_Col-0_C
ATGE_37_A [12]	carpels_floralStage12_Col-0_A
ATGE_37_B [12]	carpels_floralStage12_Col-0_B
ATGE_37_C [12]	carpels_floralStage12_Col-0_C
ATGE_39_A [12]	flowers_floralStage16_Col-0_A
ATGE_39_B [12]	flowers_floralStage16_Col-0_B
ATGE_39_C [12]	flowers_floralStage16_Col-0_C
ATGE_3_A [12]	roots_seedling7days_Col-0_A
ATGE_3_B [12]	roots_seedling7days_Col-0_B
ATGE_3_C [12]	roots_seedling7days_Col-0_C
ATGE_40_A [12]	pedicels_floralStage15_Col-0_A
ATGE_40_B [12]	pedicels_floralStage15_Col-0_B
ATGE_40_C [12]	pedicels_floralStage15_Col-0_C
ATGE_41_A [12]	sepals_floralStage15_Col-0_A
ATGE_41_B [12]	sepals_floralStage15_Col-0_B
ATGE_41_C [12]	sepals_floralStage15_Col-0_C
ATGE_42_B [12]	petals_floralStage15_Col-0_B
ATGE_42_C [12]	petals_floralStage15_Col-0_C
ATGE_42_D [12]	petals_floralStage15_Col-0_D

ATGE_43_A [12]	stamen_floralStage15_Col-0_A
ATGE_43_B [12]	stamen_floralStage15_Col-0_B
ATGE_43_C [12]	stamen_floralStage15_Col-0_C
ATGE_45_A [12]	carpels_floralStage15_Col-0_A
ATGE_45_B [12]	carpels_floralStage15_Col-0_B
ATGE_45_C [12]	carpels_floralStage15_Col-0_C
ATGE_4_A [12]	shootApexLeaves_seedling7days_Col-0_A
ATGE_4_B [12]	shootApexLeaves_seedling7days_Col-0_B
ATGE_4_C [12]	shootApexLeaves_seedling7days_Col-0_C
ATGE_5_A [12]	leaves_seedling7days_Col-0_A
ATGE_5_B [12]	leaves_seedling7days_Col-0_B
ATGE_5_C [12]	leaves_seedling7days_Col-0_C
ATGE_6_A [12]	shootApex_seedling7days_Col-0_A
ATGE_6_B [12]	shootApex_seedling7days_Col-0_B
ATGE_6_C [12]	shootApex_seedling7days_Col-0_C
ATGE_73_A [12]	pollen_mature_wt_A
ATGE_73_B [12]	pollen_mature_wt_B
ATGE_73_C [12]	pollen_mature_wt_C
ATGE_76_A [12]	siliques_seedStage3_Col-0_A
ATGE_76_B [12]	siliques_seedStage3_Col-0_B
ATGE_76_C [12]	siliques_seedStage3_Col-0_C
ATGE_77_D [12]	siliques_seedStage4_Col-0_D
ATGE_77_E [12]	siliques_seedStage4_Col-0_E
ATGE_77_F [12]	siliques_seedStage4_Col-0_F
ATGE_78_D [12]	siliques_seedStage5_Col-0_D
ATGE_78_E [12]	siliques_seedStage5_Col-0_E
ATGE_78_F [12]	siliques_seedStage5_Col-0_F
ATGE_79_A [12]	seeds_seedStage6_Col-0_A
ATGE_79_B [12]	seeds_seedStage6_Col-0_B
ATGE_79_C [12]	seeds_seedStage6_Col-0_C
ATGE_81_A [12]	seeds_seedStage7_Col-0_A
ATGE_81_B [12]	seeds_seedStage7_Col-0_B
ATGE_81_C [12]	seeds_seedStage7_Col-0_C
ATGE_82_A [12]	seeds_seedStage8_Col-0_A
ATGE_82_B [12]	seeds_seedStage8_Col-0_B
ATGE_82_C [12]	seeds_seedStage8_Col-0_C
ATGE_83_A [12]	seeds_seedStage9_Col-0_A
ATGE_83_B [12]	seeds_seedStage9_Col-0_B
ATGE_83_C [12]	seeds_seedStage9_Col-0_C
ATGE_84_A [12]	seeds_seedStage10_Col-0_A
ATGE_84_B [12]	seeds_seedStage10_Col-0_B
ATGE_84_D [12]	seeds_seedStage10_Col-0_D
ATGE_7_A2 [12]	seedlingGreenParts_seedling7days_Col-0_A
ATGE_7_B2 [12]	seedlingGreenParts_seedling7days_Col-0_B
ATGE_7_C2 [12]	seedlingGreenParts_seedling7days_Col-0_C
ATGE_87_A [12]	rosette7_vegRosettePhaseChange_Col-0_A
ATGE_87_B [12]	rosette7_vegRosettePhaseChange_Col-0_B
ATGE_87_C [12]	rosette7_vegRosettePhaseChange_Col-0_C
ATGE_89_A [12]	rosette14_vegRosettePhaseChange_Col-0_A
ATGE_89_B [12]	rosette14_vegRosettePhaseChange_Col-0_B
ATGE_89_C [12]	rosette14_vegRosettePhaseChange_Col-0_C
ATGE_90_A [12]	rosette21_vegRosettePhaseChange_Col-0_A
ATGE_90_B [12]	rosette21_vegRosettePhaseChange_Col-0_B
ATGE_90_D [12]	rosette21_vegRosettePhaseChange_Col-0_C
ATGE_8_A [12]	shootApex_rosetteStage_Col-0_A
ATGE_8_B [12]	shootApex_rosetteStage_Col-0_B
ATGE_8_C [12]	shootApex_rosetteStage_Col-0_C
ATGE_91_A [12]	leaf_leaf_Col-0_A
ATGE_91_B [12]	leaf_leaf_Col-0_B
ATGE_91_C [12]	leaf_leaf_Col-0_C
ATGE_92_A [12]	flower_flower_Col-0_A

ATGE_92_B [12]	flower_flower_Col-0_B
ATGE_92_C [12]	flower_flower_Col-0_C
ATGE_9_A [12]	roots_rosetteStage_Col-0_A
ATGE_9_B [12]	roots_rosetteStage_Col-0_B
ATGE_9_C [12]	roots_rosetteStage_Col-0_C
AGL42_1 [13]	root_qc_1
AGL42_2 [13]	root_qc_2
pet111_1 [13]	root_columnellatier234_1
pet111_2 [13]	root_columnellatier234_2
pet111_3 [13]	root_columnellatier234_3
APL_2 [14]	root_phloem_1
APL_3 [14]	root_phloem_2
S18_1 [14]	root_xylem_1
S18_2 [14]	root_xylem_2
S18_3 [14]	root_xylem_3
CORTEX_1 [14]	root_cortex_1
CORTEX_2 [14]	root_cortex_2
CORTEX_3 [14]	root_cortex_3
COBL9_1 [15]	root_epidermis_1
COBL9_2 [15]	root_epidermis_2
J0121_1 [15]	root_pericycle_xylempole_1
J0121_2 [15]	root_pericycle_xylempole_2
J0121_3 [15]	root_pericycle_xylempole_3
rm1000_1 [15]	root_pericycle_LRinitials_1
rm1000_2 [15]	root_pericycle_LRinitials_2
S17_1 [15]	root_pericycle_ploempole_1
S17_2 [15]	root_pericycle_ploempole_2
S17_3 [15]	root_pericycle_ploempole_3
S32_1 [15]	root_phloem_protophloem_1
S32_2 [15]	root_phloem_protophloem_2
S32_3 [15]	root_phloem_protophloem_3
SUC2_1 [15]	root_phloem_companion_1
SUC2_2 [15]	root_phloem_companion_2
SUC2_3 [15]	root_phloem_companion_3
S4_1 [15]	root_xylem_protometaxylem_1
S4_2 [15]	root_xylem_protometaxylem_2
S4_3 [15]	root_xylem_protometaxylem_3
xylem_2501_1 [15]	root_xylem_metaxylem_1
xylem_2501_2 [15]	root_xylem_metaxylem_2
xylem_2501_3 [15]	root_xylem_metaxylem_3
E30_1 arexdb.org	root_matureendodermis_1
E30_2 arexdb.org	root_matureendodermis_2
E30_3 arexdb.org	root_matureendodermis_3
J2661_1 [16]	root_maturepericycle_1
J2661_2 [16]	root_maturepericycle_2
J2661_3 [16]	root_maturepericycle_3
WER_1 [16]	root_epidermis_LRC_nonhair_1
WER_2 [16]	root_epidermis_LRC_nonhair_2
WER_3 [16]	root_epidermis_LRC_nonhair_3
gl2_1 [17]	root_epidermis_atrichoblast_1
gl2_2 [17]	root_epidermis_atrichoblast_2
gl2_3 [17]	root_epidermis_atrichoblast_3
J0571_1 [17]	root_groundtissue_1
J0571_2 [17]	root_groundtissue_2
J0571_3 [17]	root_groundtissue_3
LRC_1 [17]	root_lateralrootcap_1
LRC_2 [17]	root_lateralrootcap_2
LRC_3 [17]	root_lateralrootcap_3
scr5_1 [17]	root_endodermis_1
scr5_2 [17]	root_endodermis_2
scr5_3 [17]	root_endodermis_3

wol_1 [17]	root_stele_1
wol_2 [17]	root_stele_2
wol_3 [17]	root_stele_3
StageI_1 [15]	root_whole_StageI_1
StageI_2 [15]	root_whole_StageI_2
StageI_3 [15]	root_whole_StageI_3
StageI_4 [15]	root_whole_StageI_4
StageII_1 [15]	root_whole_StageII_1
StageII_2 [15]	root_whole_StageII_2
StageII_3 [15]	root_whole_StageII_3
StageII_4 [15]	root_whole_StageII_4
StageIII_1 [15]	root_whole_StageIII_1
StageIII_2 [15]	root_whole_StageIII_2
StageIII_3 [15]	root_whole_StageIII_3
StageIII_4 [15]	root_whole_StageIII_4
GSM184531 [17]	root_protoplast_1
GSM184532 [17]	root_protoplast_2
GSM184533 [17]	root_protoplast_3
GSM184534 [17]	root_protoplast_KNO3_1
GSM184535 [17]	root_protoplast_KNO3_2
GSM184536 [17]	root_protoplast_KNO3_3
GSM205156 [18]	mesophyll_protoplast_1
GSM205160 [18]	mesophyll_protoplast_2
GSM413936 [19]	mesophyll_protoplast_CPK_T0_1
GSM413937 [19]	mesophyll_protoplast_CPK_T1_1
GSM413938 [19]	mesophyll_protoplast_CPK_T2_1
GSM413939 [19]	mesophyll_protoplast_CPK_T0_2
GSM413940 [19]	mesophyll_protoplast_CPK_T1_2
GSM413941 [19]	mesophyll_protoplast_CPK_T2_2
GSM413942 [19]	mesophyll_protoplast_flg22_T0_1
GSM413943 [19]	mesophyll_protoplast_flg22_T1_1
GSM413944 [19]	mesophyll_protoplast_flg22_T2_1
GSM413945 [19]	mesophyll_protoplast_flg22_T0_2
GSM413946 [19]	mesophyll_protoplast_flg22_T1_2
GSM413947 [19]	mesophyll_protoplast_flg22_T2_2

Bibliography

[1] Yu H, Hogan P, Sundaresan V (2005) Analysis of the female gametophyte transcriptome of *Arabidopsis* by comparative expression profiling. Plant Physiology 139: 1853-1869.

[2] Wuest S, Vijverberg K, Schmidt A, Weiss M, Gheyselinck J, et al. (2010) *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. Current Biology 20: 1-7.

[3] Schmidt A, Wuest SE, Vijverberg K, Baroux C, Kleen D, et al. (2011) Transcriptome analysis of the *Arabidopsis* megaspore mother cell uncovers the importance of RNA helicases for plant germ line development. PLOS Biology 9: e1001155.

[4] Pina C, Pinto F, Feijó J, Becker J (2005) Gene family analysis of the *Arabidopsis* pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. Plant Physiology 138: 744-756.

[5] Honys D, Twell D (2004) Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. Genome Biology 5: R85.

[6] Borges F, Gomes G, Gardner R, Moreno N, McCormick S, et al. (2008) Comparative transcriptomics of *Arabidopsis* sperm cells. Plant Physiology 148: 1168-1181.

[7] Wang Y, Zhang W, Song L, Zou J, Su Z, et al. (2008) Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis*. Plant Physiology 148: 1201-1211.

[8] Qin Y, Leydon A, Manziello A, Pandey R, Mount D, et al. (2009) Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. PLOS Genetics 5: e1000621.

[9] Yadav R, Girke T, Pasala S, Xie M, Reddy G (2009) Gene expression map of the *Arabidopsis* shoot apical meristem stem cell niche. PNAS 106: 4941-4946.

[10] Le B, Cheng C, Bui A, Wagmaister J, Henry K, et al. (2010) Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. PNAS 107: 8063-8070.

[11] Zuber H, Davidian J, Aubert G, Aimé D, Belghazi M, et al. (2010) The seed composition of *Arabidopsis* mutants for the group 3 sulfate transporters indicates a role in sulfate translocation within developing seeds. Plant Physiology 154: 913-926.

[12] Schmid M, Davison T, Henz S, Pape U, Demar M, et al. (2005) A gene expression map of *Arabidopsis thaliana* development. Nature Genetics 37: 501-506.

[13] Nawy T, Lee J, Colinas J, Wang J, Thongrod S, et al. (2005) Transcriptional profile of the *Arabidopsis* root quiescent center. The Plant Cell 17: 1908-1925.

[14] Lee J, Colinas J, Wang J, Mace D, Ohler U, et al. (2006) Transcriptional and posttranscriptional regulation of transcription factor expression in *Arabidopsis* roots. PNAS 103: 6055-6060.

[15] Brady S, Orlando D, Lee J, Wang J, Koch J, et al. (2007) A high-resolution root spatiotemporal map reveals dominant expression patterns. Science 318: 801-806.

[16] Levesque M, Vernoux T, Busch W, Cui H, Wang J, et al. (2006) Whole-genome analysis of the SHORT-ROOT developmental pathway in *Arabidopsis*. PLOS Biology 4: e143.

[17] Birnbaum K, Shasha D, Wang J, Jung J, Lambert G, et al. (2003) A gene expression map of the *Arabidopsis* root. Science 302: 1956-1960.

[18] Baena-González E, Rollang F, Thevelein J, Sheen J (2007) A central integrator of transcription networks in plant stress and energy signalling. Nature 448: 938-942.

[19] Boudsocq M, Willmann M, McCormack M, Lee H, Shan L, et al. (2010) Differential innate immune signalling via Ca(2+) sensor protein kinases. Nature 464: 418-422.

S1 Text: Relationship between weighted mean, slope and Pearson's r

The correlation coefficient r (Pearson's r) of the weights w (expression intensities) and the measured values x (evolutionary parameter, e.g. gene age, d_N/d_S or p_N/p_S) is

$$\rho_{x,w} = \frac{E\left((X-\mu_x)(W-\mu_w)\right)}{\sigma_x \sigma_w}$$

where μ_x is the mean of x and μ_w is the mean of w. This can be expanded to

$$\rho_{x,w} = \frac{E\left(XW - \mu_x W - \mu_w X + \mu_x \mu_w\right)}{\sigma_x \sigma_w}$$

and then to

$$\begin{aligned} \rho_{x,w} &= \frac{E(XW) - \mu_x\mu_w - \mu_x\mu_w + \mu_x\mu_w}{\sigma_x\sigma_w} \\ &= \frac{\mu_w(\mu_{xw} - \mu_x)}{\sigma_x\sigma_w} \end{aligned}$$

For the weighted mean μ_{xw} (transcriptome evolutionary index)

$$\begin{aligned} \mu_{xw} &= \frac{\sum xw}{\sum w} \\ &= \frac{nE(XW)}{n\mu_w} \\ &= \frac{E(XW)}{\mu_w}. \end{aligned}$$

Consequently,

$$\frac{\rho_{x,w}\sigma_w\sigma_x}{\mu_w} = \mu_{xw} - \mu_x.$$

Note, that μ_x and σ_x are constant across different stages when the same genes are considered for each stage.
Also, for linear regression, the coefficient (slope) β is

$$\beta = \frac{Cov(x,w)}{\sigma_x^2}$$

which can be rearranged to

$$\beta = \frac{\mu_w(\mu_{xw} - \mu x)}{\sigma_x^2}$$

and therefore

$$\mu_{xw} - \mu_x = \frac{\rho_{x,w}\sigma_w\sigma_x}{\mu_w} = \frac{\beta\sigma_x^2}{\mu_w}.$$

For the special case of $\sum_{i=1}^{n} w_i = 1$ (where n is the number of genes)

$$\mu_{xw} - \mu_x = n\rho_{x,w}\sigma_w\sigma_x = n\beta\sigma_x^2.$$

Note, that n and σ_x are constant across different stages when the same genes are considered for each stage