Selection Is No More Efficient in Haploid than in Diploid Life Stages of an Angiosperm and a Moss

Péter Szövényi,*[1,2,3,4] Mariana Ricca,[1,3] Zsófia Hock,[2] Jonathan A. Shaw,[5] Kentaro K. Shimizu,[1] and Andreas Wagner[1,3,6,7]

[1]Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland

[2]Institute of Systematic Botany, University of Zurich, Zurich, Switzerland

[3]Swiss Institute of Bioinformatics, Quartier Sorge-Batiment Genopode, Lausanne, Switzerland

[4]MTA-ELTE-MTM Ecology Research Group, ELTE, Biological Institute, Budapest, Hungary

[5]Department of Biology, Duke University

[6]Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Singapore

[7]The Santa Fe Institute, Santa Fe, New Mexico

***Corresponding authors:** E-mail: peter.szoevenyi@uzh.ch, pis@duke.edu.

Associate editor: Michael Purugganan

Abstract

The masking hypothesis predicts that selection is more efficient in haploids than in diploids, because dominant alleles can mask the deleterious effects of recessive alleles in diploids. However, gene expression breadth and noise can potentially counteract the effect of masking on the rate at which genes evolve. Land plants are ideal to ask whether masking, expression breadth, or expression noise dominate in their influence on the rate of molecular evolution, because they have a biphasic life cycle in which the duration and complexity of the haploid and diploid phase varies among organisms. Here, we generate and compile genome-wide gene expression, sequence divergence, and polymorphism data for *Arabidopsis thaliana* and for the moss *Funaria hygrometrica* to show that the evolutionary rates of haploid- and diploid-specific genes contradict the masking hypothesis. Haploid-specific genes do not evolve more slowly than diploid-specific genes in either organism. Our data suggest that gene expression breadth influence the evolutionary rate of phase-specific genes more strongly than masking. Our observations have implications for the role of haploid life stages in the purging of deleterious mutations, as well as for the evolution of ploidy.

Key words: **high throughput sequencing, haploid, diploid, biphasic life cycle, masking, expression breadth, expression noise, evolutionary rate.**

Introduction

The masking hypothesis suggests that the number of chromosomal copies present in a cell affects the efficacy of selection (Kondrashov and Crow 1991). According to this hypothesis, selection is more efficient in haploids than in diploids, because recessive mutations are directly exposed to selection in haploids, whereas their phenotypic effect can be masked in heterozygote diploids through dominant alleles. As a consequence, evolutionary rates in haploids and diploids should differ, provided that the majority of mutations are recessive or partially recessive (Orr and Otto 1994; Otto and Gerstein 2008). Predictions of the masking hypothesis have been experimentally confirmed in unicellular organisms (primarily yeast). Pertinent experiments showed that diploid and tetraploid yeast strains are less sensitive to mutagens, but they experience slightly slower fitness recovery than haploid strains after mutagen treatment (Mable and Otto 2001). Furthermore, adaptation is faster in haploid than in diploid strains when the rate of adaptation is limited by selection (Zeyl et al. 2003; Gerstein et al. 2011). Observations on simple plant tissues, such as protoplasts, also led to similar conclusions (Krumbiegel 1979; Destombe et al. 1993).

The masking hypothesis also predicts that in organisms with life cycles that alternate between a haploid and a diploid phase, genes with phase-specific expression should differ in their evolutionary rate (Bell 2008). Recessive deleterious or beneficial mutations will be more effectively purged or go to fixation in genes whose expression is restricted to the haploid phase. In contrast, genes silenced in the haploid and only expressed in the diploid phase may be hidden from haploid selection, and thus may accumulate recessive deleterious or beneficial mutations without affecting haploid fitness (Shaw and Beer 1997; Otto 2004). In other words, diploid-specific genes should evolve more rapidly. This applies especially to biphasic life cycles, in which both diploid and haploid phases build a multicellular body, and exhibit extensive phase-specific gene expression on which natural selection can act (Charlesworth D and Charlesworth B 1992; Seoighe et al. 2005; Bell 2008). Yet, whether predictions of the masking hypothesis hold for complex multicellular organisms with biphasic life cycles is not clear.

Mol. Biol. Evol. 30(8):1929–1939 doi:10.1093/molbev/mst095 Advance Access publication May 17, 2013

Two major factors could counteract the effect of masking on the evolutionary rate of genes, especially in multicellular organisms with biphasic life cycles. The first is gene expression breadth, which strongly affects the rate of evolution, so much so that it is one of the best predictors of evolutionary rate (Slotte et al. 2011; Woody and Shoemaker 2011; Yang and Gaut 2011). Broadly expressed genes evolve more slowly than genes with restricted expression (Park and Choi 2010). Genes with phase-specific expression are by definition narrowly expressed, and should thus evolve fast, in contrast to what the masking hypothesis would predict.

The second factor is gene expression noise, which has effects akin to those of decreasing effective population size, and thus results in decreased efficacy of selection (Wang and Zhang 2011). Because haploid-specific genes may be noisier than diploid-specific genes (Cook et al. 1998; Yin et al. 2009), they would experience relaxed selection and evolve faster, again in contradiction to the masking hypothesis (Wang and Zhang 2011). In sum, both expression breadth and noise could decrease or even reverse the effect of masking on the efficacy of selection. It is not clear whether masking or these two factors dominate in their effect on evolutionary rate.

Land plants are ideal to study the effect of masking, expression breadth, and noise on the evolutionary rate of genes in multicellular organisms with biphasic life cycles. All land plants possess a biphasic life cycle with multicellular haploid and diploid phases showing variable morphological complexity and extensive phase-specific gene expression (Honys and Twell 2003; Pina et al. 2005; Ma et al. 2008; Haerizadeh et al. 2009; Wuest et al. 2010; Hafidh et al. 2012; Russell et al. 2012). Here, we use *Arabidopsis thaliana*, an angiosperm, and *Funaria hygrometrica*, a moss, to study the molecular evolutionary rate of genes specific to haploid and diploid phases, because these organisms are two end points of the relative morphological complexity continuum. *Arabidopsis thaliana* has a dominant diploid phase, the leafy shoot, and a highly reduced haploid phase consisting of a few cells (Wuest et al. 2010). In contrast, the relative dominance of these phases is reversed in *F. hygrometrica*, and a dominant haploid phase (leafy shoot) alternates with a reduced diploid phase (Shaw et al. 2011). Given that most mutations are recessive and deleterious (Wright and Andolfatto 2008; Gossmann et al. 2010; Slotte et al. 2011; Yang and Gaut 2011), these plants provide a unique opportunity to distinguish the effect of masking, expression breadth and noise on the evolutionary rate of proteins (fig. 1). If evolutionary rates are primarily determined by masking in the diploid phase, haploid-specific genes should evolve more slowly in both species than diploid-specific or unspecific genes—genes expressed in both phases. In contrast, if expression breadth predominates and genes specific to the dominant phase are more broadly expressed haploid-specific genes should evolve faster than diploid-specific genes in *A. thaliana*, whereas they should evolve more slowly in *F. hygrometrica*. Alternatively, if genes specific to the dominant phase are less broadly expressed (e.g., increasing complexity is associated with greater expression specialization), we should observe the opposite pattern. Nevertheless, genes with unspecific expression should evolve more slowly than genes with specific expression in both species. Finally, if expression noise predominates, we would expect haploid-specific genes to evolve more rapidly than diploid-specific or unspecific genes in both species, regardless of the complexity of the phases.

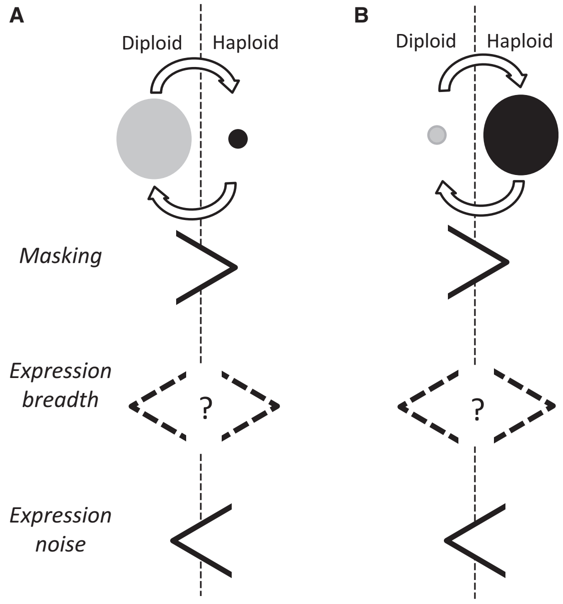

Fig. 1. The predicted effect of masking, expression breadth and expression noise on the evolutionary rate of genes in organisms with biphasic life cycles. Diameters of solid circles refer to the relative complexity of phases: (*A*) diploid-dominant and (*B*) haploid-dominant biphasic life cycle. Greater and smaller signs show the relative rate of evolution (d*N*/d*S*) expected in the two phases when the effect of masking, expression breadth and expression noise is individually considered. Dashed greater or smaller signs with question marks indicate the ambiguous association between expression breadth and relative complexity of phases.

To distinguish the factors contributing to evolutionary rate, we generated and compiled genome-wide data on the expression of genes that are preferentially expressed in the haploid (gametophyte) and diploid (sporophyte) phase in each of the two species. By integrating these data with sequence divergence and polymorphism data, we asked the following questions: First, do evolutionary rates differ in genes with haploid-specific, diploid-specific, and unspecific expression? Second, is selection more efficient in haploid-specific genes? Third, can factors other than ploidy account for the difference in evolutionary rates between phases? Finally, are evolutionary rate differences among phases consistent with the masking hypothesis, or do expression breadth and expression noise exert a dominant influence on them?

Results

Relative Dominance of Phases Is Mirrored by the Number of Genes Showing Phase-Specific Expression

We first identified genes with haploid-specific, diploid-specific, and unspecific expression in two *A. thaliana* data sets (see Materials and Methods), which differed in the numbers of genes in these three categories. Specifically, the first data set contained 64 genes with haploid-specific, 2,598 genes with diploid-specific, and 8,806 genes with unspecific expression. In the second data set, 425 genes showed haploid-specific, 2,699 genes showed diploid-specific, and 8,246 genes showed unspecific expression. Haploid-specific genes were clearly outnumbered by diploid-specific and unspecific genes, which is not surprising given the highly reduced haploid phase in angiosperms. In *A. thaliana*, genes specific to the reduced haploid phase were less broadly expressed than diploid-specific and unspecific genes in both data sets investigated (mean and [95% confidence interval (CI)], data set 1: $\tau_{\text{haploid-specific}} = 0.3487$ [0.3295–0.3680], $\tau_{\text{diploid-specific}} = 0.26078$ [0.2588–0.2628], $\tau_{\text{unspecific}} = 0.2066$ [0.2035–0.2100]; data set 2: $\tau_{\text{haploid-specific}} = 0.3738$ [0.3633–0.3843], $\tau_{\text{diploid-specific}} = 0.2822$ [0.2788–0.2855], $\tau_{\text{unspecific}} = 0.2285$ [0.2267–0.2304]).

Similarly to *Arabidopsis*, different definitions of expression preference affected the number of genes falling into each of the three categories in the moss *F. hygrometrica*. When we used a fold-change [$\log_2$(fold-change)] threshold of two, 1,049 genes in the moss showed haploid-specific, 1,309 genes showed diploid-specific, and 7,608 genes showed unspecific expression. Applying a fold-change threshold of four, these numbers changed to 542, 400, and 9,096 genes, whereas a threshold of six resulted in 343, 243, and 9,461 genes in each of the respective categories. In the last two cases, haploid-specific genes clearly dominated diploid-specific genes, but this difference was less pronounced than the opposite difference in *A. thaliana*. In contrast to *A. thaliana*, genes specific to the reduced (diploid) phase of the life cycle were more broadly expressed than genes specific to the dominant haploid phase at all 3-fold-change thresholds (mean and [95% CI], fold-change 2: $\tau_{\text{haploid-specific}} = 0.2854$ [0.2670–0.3040], $\tau_{\text{diploid-specific}} = 0.1803$ [0.1694–0.1911]; fold-change 4: $\tau_{\text{haploid-specific}} = 0.4335$ [0.4060–0.4610], $\tau_{\text{diploid-specific}} = 0.3065$ [0.2838–0.3292]; fold-change 6: $\tau_{\text{haploid-specific}} = 0.5191$ [0.4836–0.5546], $\tau_{\text{diploid-specific}} = 0.3476$ [0.3157–0.3795]). Nevertheless, genes with unspecific expression had the greatest expression breadth similarly to what we observed in *A. thaliana* (fold-change 2: $\tau_{\text{unspecific}} = 0.0330$ [0.0318–0.0361]; fold-change 4: $\tau_{\text{unspecific}} = 0.0482$ [0.0458–0.0506]; fold-change 6: $\tau_{\text{unspecific}} = 0.05675$ [0.0542–0.0593]).

Haploid-Specific Genes Evolve Faster or at a Similar Rate than Diploid-Specific Genes

Evolutionary rates in haploid- and diploid-specific genes showed a similar trend in both *A. thaliana* and *F. hygrometrica* (fig. 2; tables 1 and 2). Specifically, when we defined phase specificity as a binary variable (see Materials and Methods), haploid-specific genes evolved faster (higher d*N*/d*S*) than diploid-specific genes in *A. thaliana*, and they did so in both expression data sets we analyzed. *F. hygrometrica* showed a similar trend but the difference in evolutionary rates among haploid- and diploid-specific genes was less pronounced (fig. 2 and table 2). On average, haploid-specific genes evolved slightly faster than diploid-specific genes at all fold-change thresholds investigated. Nevertheless, at a threshold of fold-change 6 the significance of the difference diminished. Genes with unspecific expression always showed slower evolutionary rates than haploid-specific or diploid-specific genes both in *A. thaliana* and *F. hygrometrica*. Furthermore, this difference was independent on the data set investigated or the fold-change threshold used (fig. 2; tables 1 and 2).

Defining phase specificity using strict threshold values is subjective and might bias our conclusions. Therefore, we repeated our analysis taking into account the continuous nature of phase specificity. Treating phase specificity as a continuous variable (fold-change = $\log_2$[gene expression in the haploid/gene expression in the diploid phase]) and analyzing its effect on the evolutionary rate of proteins (d*N*/d*S*) led to very similar conclusions. In *A. thaliana* phase, specificity of genes was significantly positively correlated with evolutionary rates ($\rho = 0.2587$, $P < 2.2 \times 10^{-16}$). In contrast, expression specificity showed only a very weak positive correlation with evolutionary rates in *F. hygrometrica* ($\rho = 0.0242$, $P = 0.01513$).

Haploid-Specific Genes Evolve under Relaxed Selective Constraints

Faster evolutionary rates (higher d*N*/d*S*) in the haploid phase may be either caused by rapid fixation of beneficial mutations or by reduced selective constraints. To distinguish between these two competing hypotheses, we first conducted a direct test using genome-wide polymorphism data available for *A. thaliana* (the same test was not possible in the moss, because no genome-wide polymorphism data are available there). Beneficial mutations are expected to rapidly reach fixation within species, whereas slightly deleterious mutations should remain polymorphic for a longer period of time. Therefore, if elevated among-species d*N*/d*S* ratios were caused by the fixation of beneficial mutations, ratio of nonsynonymous to synonymous polymorphisms within-species should be lower in haploid-specific than in diploid-specific genes. In contrast, if elevated among-species d*N*/d*S* ratios were mainly caused by relaxed selection, the ratio of nonsynonymous to synonymous polymorphisms should be higher for haploid-specific than for diploid-specific genes. To distinguish between these alternative hypotheses, we estimated the number of nonsynonymous to synonymous polymorphisms for haploid-specific, diploid-specific, and unspecific genes using the 19 *A. thaliana* genome data set with *A. lyrata* as an outgroup. The average ratio of nonsynonymous to synonymous polymorphisms was significantly higher in haploid-specific than in diploid-specific genes (table 3). Furthermore, genes with phase-specific expression showed a higher ratio than genes with unspecific expression (table 3).

We also conducted an indirect test using only among-species divergence data. As an indicator of the efficacy of selection, we used the strength of correlation between expression level and evolutionary rate. In *A. thaliana*, gene expression level and evolutionary rates (d*N*/d*S*) of genes were more weakly correlated in haploid-specific than in diploid-specific genes, at a statistical significance that depended on the data set used (table 4). Similarly, gene expression and evolutionary rates were more weakly correlated (if at all) in haploid-specific

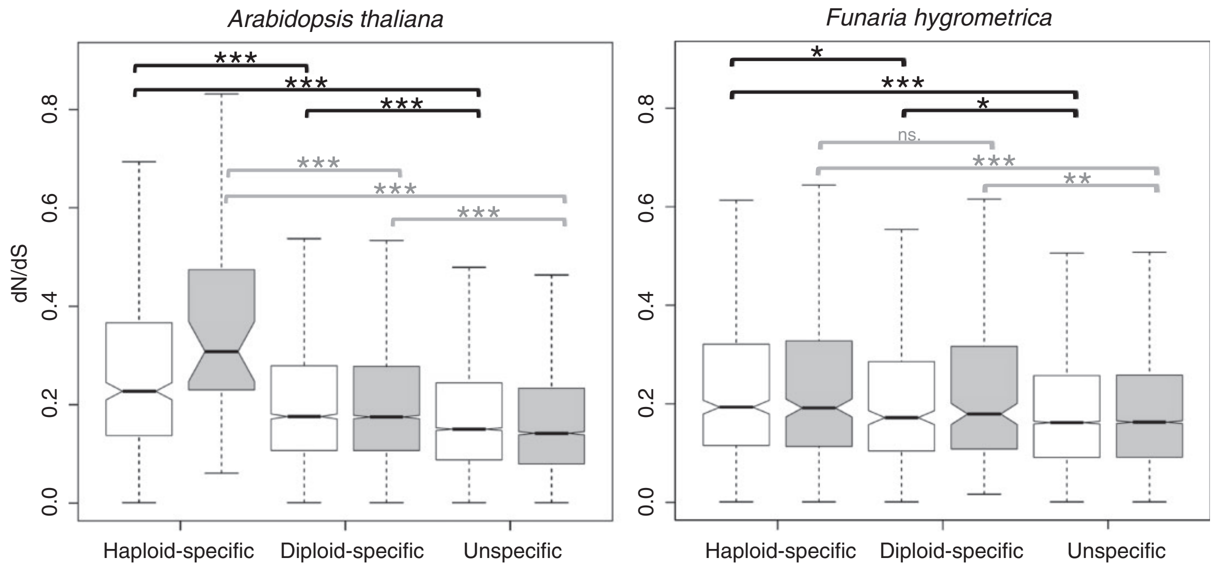

Fig. 2. Evolutionary rates of genes (d*N*/d*S*) with haploid-, diploid-specific or unspecific expression in *Arabidopsis thaliana* and in the moss (*Funaria hygrometrica*). In *A. thaliana*, white boxes refer to data set 1 and gray boxes to data set 2. Similarly, in *F. hygrometrica* white boxes refer to a fold-change threshold of four and gray boxes to a threshold of six ($\log_2$[fold-change]). Asterisks indicate whether medians are significantly different according to a Wilcoxon rank sum test after Bonferroni correction ($\alpha' = \alpha/m$) for multiple testing (n.s.: $P > 0.05$; *$P \leq 0.05$; **$P \leq 0.01$; ***$P \leq 0.001$).

Table 1. Evolutionary Rate (Median and Mean d*N*/d*S* Values) of Haploid-Specific, Diploid-Specific, and Unspecific Genes in *Arabidopsis thaliana*.

Expression Preference	dN/dS [Median (Mean)]			
	Data Set 1	IQR Data Set 1	Data Set 2	IQR Data Set 2
Haploid specific	0.2277 (0.2976)^a	0.1375–0.3665	0.3311 (0.3914)^a	0.2334–0.4740
Diploid specific	0.1760 (0.2142)^b	0.1069–0.2791	0.1750 (0.2157)^b	0.1069–0.2777
Unspecific	0.1504 (0.2004)^c	0.0879–0.2444	0.1421 (0.1779)^c	0.0798–0.2334

Note.—IQR, interquartile range. Values marked with different superscript letters are significantly different within columns after Bonferroni correction ($\alpha' = \alpha/m$) for multiple testing ($P < 0.05$, Wilcoxon rank sum test).

Table 2. Evolutionary Rate (Median and Mean d*N*/d*S* Values) of Haploid-Specific, Diploid-Specific, and Unspecific Genes in *Funaria hygrometrica* at 3-Fold-Change Thresholds.

Expression Preference	dN/dS [Median (Mean)]					
	Fold-Change > 2	IQR	Fold-Change > 4	IQR	Fold-Change > 6	IQR
Haploid specific	0.1813 (0.2262)^a	0.1076–0.3040	0.1932 (0.2343)^a	0.1156–0.3210	0.1916 (0.2359)^a	0.1135–0.3278
Diploid specific	0.1714 (0.2076)^b	0.1021–0.2724	0.1720 (0.2103)^b	0.1047–0.2854	0.1794 (0.2270)^a	0.1085–0.3140
Unspecific	0.1600 (0.1895)^c	0.0892–0.2547	0.1619 (0.1928)^c	0.0913–0.2573	0.1629 (0.1935)^c	0.0917–0.2583

Note.—IQR, interquartile range. Values marked with different superscript letters are significantly different within columns after Bonferroni correction ($\alpha' = \alpha/m$) for multiple testing ($P < 0.05$, Wilcoxon rank sum test).

genes than in diploid-specific genes especially at higher fold-change thresholds in the moss (table 5). Finally, expression level and evolutionary rates were most strongly correlated for genes with unspecific expression, both in *A. thaliana* and *F. hygrometrica* (tables 4 and 5).

Variables Other than Ploidy Cannot Account for the Evolutionary Rate Difference among Phase-Specific Genes

Evolutionary rate differences between haploid-specific, diploid-specific, and unspecific genes might also be caused by other factors known to be correlated with evolutionary rates (Slotte et al. 2011; Yang and Gaut 2011). In the previous paragraph, we showed that the correlation between level of expression and evolutionary rates differ among haploid-specific, diploid-specific, and unspecific genes. Therefore, level of expression cannot account for differential evolutionary rates in the three categories of genes. Here, we explored this question further and asked whether the differences in evolutionary rates among the three gene categories (expression preference treated as binary variable; see Materials and Methods) persist if we account for the influence of three prominent such factors, namely gene length, GC content, and average intron

Table 3. Nonsynonymous Polymorphism/Synonymous Polymorphism Ratio (Mean and Median Values, $\log_{10}$ Transfromed) in Haploid-Specific, Diploid-Specific, and Unspecific Genes Using the 19 *Arabidopsis* Genome Data Set.

Expression Preference	Log(Nonsynonymous Polymorphism/Synonymous Polymorphism)			
	Data Set 1	95% CI (Data Set 1)	Data Set 2	95% CI (Data Set 2)
Haploid specific	0.0949 (0.1290)^{a}	−0.0052 to 0.2631	0.3172 (0.5281)^{a}	0.1730 to 0.8832
Diploid specific	−0.2183 (−0.2973)^{b}	−0.3501 to −0.2445	−0.1318 (−0.2245)^{b}	−0.2634 to −0.1856
Unspecific	−0.2796 (−0.4300)^{c}	−0.4630 to −0.3970	−0.3309 (−0.4809)^{c}	−0.5183 to −0.4435

NOTE.—Values marked with different superscript letters are significantly different within columns after Bonferroni correction ($\alpha' = \alpha/m$) for multiple testing ($P < 0.05$, Wilcoxon rank sum test).

Table 4. Strength and Significance of Nonparametric Correlation among Evolutionary Rates (d*N*/d*S* Values) and Average Expression Levels of Genes for Haploid-Specific, Diploid-Specific, and Unspecific Genes in *Arabidopsis thaliana*.

Expression Preference	Spearman's Rho (Its Significance)	
	Data Set 1	Data Set 2
Haploid specific	−0.0440 (NS)^{NA}	−0.2500 ($P < 0.001$)^{a,*}
Diploid specific	−0.3300 ($P < 0.001$)^{a}	−0.3270 ($P < 0.001$)^{a}
Unspecific	−0.3600 ($P < 0.001$)^{b}	−0.3900 ($P < 0.001$)^{b}

NOTE.—NS, not significant; NA, not analyzed (Spearman's rho was not significant). Values marked with different superscript letters are significantly different within columns after Bonferroni correction ($\alpha' = \alpha/m$) for multiple testing ($P < 0.05$, Z-test). *Significance between haploid- and diploid-specific genes is $P = 0.077$.

length. Specifically, we conducted an analysis of covariance (ANCOVA) with expression preference as the main variate and these properties as continuous covariates. The result was that none of the three factors affects the general observations we had made for *A. thaliana*. In contrast, all covariates could marginally account for the evolutionary rate difference between diploid-specific and unspecific genes in *F. hygrometrica*. Nevertheless, evolutionary rates of haploid- and diploid-specific genes remained significantly different even after accounting for the effect of all covariates (supplementary tables S1 and S2, Supplementary Material online).

We repeated this analysis treating phase specificity as a continuous variable ($\log_2$[expression in the haploid/expression in the diploid phase]) that led to similar conclusions. In *A. thaliana*, nonparametric partial correlation between phase specificity and evolutionary rates (d*N*/d*S*) remained highly significant and positive even after accounting for the effect of average gene expression, gene length, GC content, and average intron length ($\rho_{\text{partial}} = 0.2520$, $P < 1.1556 \times 10^{-170}$). We obtained very similar results for *F. hygrometrica*. The partial nonparametric correlation between phase-specificity and evolutionary rate remained very weak even after controlling for the effect of expression level, gene length, GC content, and average intron length ($\rho_{\text{partial}} = 0.0238$, $P = 2.0782 \times 10^{-6}$). Altogether, these results further confirm that the covariates investigated do not considerably affect the general conclusions we made in *A. thaliana* but they might slightly modify the results obtained for *F. hygrometrica*.

Finally, we asked whether biased distribution of molecular functions among haploid-specific, diploid-specific, and unspecific genes could explain the evolutionary rate difference we observed. We compared molecular function of haploid-specific, diploid-specific, and unspecific gene groups using the molecular function ontology of the GO data base (Gene Ontology Consortium 2005). GO annotation for phase-specific genes was sparse in both species. In *A. thaliana*, 201 (of 425) haploid specific, 1,371 (of 2,699) diploid-specific, and 4,306 (of 8,246) unspecific genes had GO annotations. In *F. hygrometrica*, 189 (of 542) haploid-specific, 120 (of 400) diploid-specific, and 3,184 (of 9,096) unspecific genes had annotations. We found that relative frequency of genes associated to a particular GO term was very similar in the group of haploid-specific, diploid-specific, and unspecific genes (supplementary fig. S1, Supplementary Material online). On level two of the molecular function ontology, no GO term showed significantly different gene abundances (Fisher's exact test) among the three groups either in *A. thaliana* or in *F. hygrometrica*. Therefore, biased distribution of molecular functions among the three groups of genes is unlikely to be the primary determinant of the evolutionary rate difference observed.

Discussion

Our analysis provides three important observations on the evolutionary rate of genes with phase-specific expression. First, genes with haploid-specific expression do not evolve more slowly than genes with diploid-specific expression. Second, selection is no more efficient for haploid- than for diploid-specific genes. Third, genes with unspecific expression always evolve more slowly than genes with haploid-specific or diploid-specific expression. In the following paragraphs, we first confront our observations with available experimental evidences. After that we discuss our findings in the light of the three main candidate explanations—masking, expression breadth and expression noise—for the evolutionary rate patterns we observed.

Experimental Evidence on Complex Organisms Support Our Finding

Our observation that genes with haploid-specific expression evolve faster or with a similar rate than diploid-specific genes, and the observation that this difference is due to relaxed rather than to positive selection differs from previous observations made primarily in yeast (Mable and Otto 2001; Zeyl et al. 2003; Otto and Gerstein 2008; Gerstein et al. 2011). The likely reason is that yeast is a single-celled organism, because other observations on multicellular plants with biphasic life

Table 5. Strength and Significance of Nonparametric Correlation among Evolutionary Rates (d*N*/d*S* Values) and Average Expression for Haploid-Specific, Diploid-Specific, and Unspecific Genes in *F. hygrometrica* at 3-Fold-Change [$\log_2$(fold-change)] Thresholds.

Expression Preference	Spearman's Rho (Its Significance)		
	Fold-Change > 2	Fold-Change > 4	Fold-Change > 6
Haploid specific	-0.1371 ($P < 0.001$)[a]	-0.0579 (NS)[NA]	0.0102 (NS)[NA]
Diploid specific	-0.1852 ($P < 0.001$)[a]	-0.1663 ($P < 0.001$)[a]	-0.1158 ($P < 0.05$)[a]
Unspecific	-0.2488 ($P < 0.001$)[b]	-0.2488 ($P < 0.001$)[b]	-0.2489 ($P < 0.001$)[b]

Note.—NS, not significant; NA, not analyzed (Spearman's rho was not significant). Values marked with different superscript letters are significantly different within columns after Bonferroni correction ($\alpha' = \alpha/m$) for multiple testing ($P < 0.05$, Z test).

cycles support our observations. First, imprinted genes (which by definition have haploid expression) evolve faster than nonimprinted genes in *A. thaliana* (Wolff et al. 2011). Second, some gene families with pollen-specific, and thus haploid-specific expression evolve rapidly in the same species (Schein et al. 2004). Third, haploid-specific genes can experience less stringent selective constraints than diploid-specific genes. Specifically, some deleterious mutations with a severe impact on the diploid phase have only a slight effect on the haploid phase and thus can be successfully inherited in *A. thaliana* (Whittle and Johnston 2003; Onodera et al. 2008; Muralla et al. 2011). In sum, other observations in plants are consistent with our observations.

Findings Contradict Predictions of the Masking Hypothesis

According to the masking hypothesis, haploid-specific genes should evolve more slowly than diploid-specific genes, because purifying selection is more efficient in haploids (Kondrashov and Crow 1991; Orr and Otto 1994; Mable and Otto 2001). Our finding that haploid-specific genes evolve faster or with a similar rate than diploid-specific genes in both species clearly contradicts this expectation. Therefore, our data suggest that the masking hypothesis alone is insufficient to explain evolutionary rate difference of proteins with haploid-specific, diploid-specific, or unspecific expression in multicellular plants.

Neither the Expression Noise Nor the Expression Breadth Hypothesis Alone Can Explain Our Findings

Haploid-expressed genes are predicted to suffer more expression noise than their diploid counterparts (Cook et al. 1998; Yin et al. 2009). Increased expression noise has the same effect as reducing effective population size, and thus makes selection less effective on genes with haploid-specific expression (Wang and Zhang 2011). Thus, if expression noise were the predominant factor governing evolutionary rate, then haploid-specific genes would evolve more rapidly than genes with diploid-specific expression. Furthermore, genes with unspecific expression should evolve faster than diploid-specific genes, because they spend a part of each generation in the haploid phase where selection is less effective. These predictions hold regardless of the relative complexity of the phases, and should thus be correct in both *A. thaliana* and the moss. However, our data contradict them, because genes with unspecific expression evolve most slowly in both organisms. Furthermore, haploid-specific and diploid-specific genes evolve at a similar rate in the moss, but not in *A. thaliana*. Therefore, our observations cannot be fully explained by expression noise either.

Gene expression breadth is one of the strongest predictor of evolutionary rate in multicellular organisms (Koonin 2011; Slotte et al. 2011; Yang and Gaut 2011). In particular, broadly expressed genes evolve more slowly than genes whose expression is specific to one tissue or life-cycle stage. Assuming that phase dominance is a proxy of expression breadth, genes specific to the morphologically and structurally less complex life-cycle phase would evolve most rapidly, regardless of ploidy. We found that this assumption is valid for *A. thaliana* but not for the moss where genes specific to the dominant haploid phase appears to be less broadly expressed than those specific to the reduced diploid phase. Therefore, haploid-specific genes should evolve fastest in both *A. thaliana* and the moss if expression breadth were the primary determinant of evolutionary rates. In contrast, we found that haploid-specific genes evolve more rapidly than diploid-specific genes in *A. thaliana* but this difference was very weak or nonexistent in the moss. This contradicts our expectation because we recorded similar expression breadth difference among the three groups of genes in both species that should alter evolutionary rates in a similar extent. In sum, none of the three factors (masking, expression noise, and breadth) alone can explain our findings in full.

Combined Effect of Gene Expression Breadth and Masking Can Explain Our Data

We found that haploid-specific genes are less broadly expressed than diploid-specific genes in both species investigated. Therefore, the effect of gene expression breadth and noise is expected to point in the same direction and will increase evolutionary rates of haploid- compared with diploid-specific genes. In contrast, masking is expected to have an opposing effect and will decrease evolutionary rates in haploid- compared with diploid-specific genes. Consequently, evolutionary rate difference between haploid-specific, diploid-specific, and unspecific genes must arise by the combined effect of these forces.

We found that haploid-specific genes are less broadly expressed than diploid-specific and unspecific genes in both organisms. Nevertheless, evolutionary rate difference between haploid- and diploid-specific genes was only significant in the species with a reduced haploid phase (*A. thaliana*). This observation could be best explained by the combined effect of expression breadth and a dominance dependent effect of masking on the evolutionary rate of genes. We hypothesize that the relative life span of the haploid phase modulates the strength of purging in the haploid phase. Purging of deleterious alleles can be more efficient in the long-lived haploid phase of the moss (Szövényi et al. 2011) than in the relatively short-lived haploid phase of *A. thaliana* (Wuest et al. 2010). Therefore, intensified haploid purging in the moss can balance the effect of expression breadth leading to similar evolutionary rates in haploid- and diploid-specific genes. This is in sharp contrast to *A. thaliana* in which the effect of expression breadth dominates over purging in the haploid phase. As a result, we argue that the combined effect of gene expression breadth and masking (in a dominance dependent manner) can best explain the broad patterns of molecular evolution we see in our two study species. Finally, the observation that evolutionary rates of haploid- and diploid-specific genes are similar in the moss suggests a negligible effect of expression noise on the evolutionary rate of genes.

General Evolutionary Implications

As we discuss next, our observations have profound implications on the purging of the deleterious mutations that a population harbors—the population's genetic load—and on life cycle evolution in general.

It has been argued that the exposure and purging of deleterious mutations during the haploid phase of a life cycle can be crucial to decrease a population's genetic load (Charlesworth and Willis 2009), and especially so in organisms with an extensive haploid phase. However, multiple observations suggest that the genetic load can only be partially purged, even in species with an extensive haploid phase. For instance, in rotifers with haploid and diploid phases of similar complexities, haploid exposure to selection cannot eliminate the genetic load completely (Tortajada et al. 2009). Similarly, in insects with haplodiploidy, deleterious mutations are only partially purged (Henter 2003). Moreover, a substantial genetic load accumulates in plant species in spite of their complex haploid phase (Byers and Waller 1999).

Our observations can help explain these patterns, because they show that haploid-specific genes are not necessarily under greater selective constraints than diploid-specific genes. The reason is the important influence of gene expression breadth that can potentially counteract the dominance-dependent effect of masking (e.g., purging in the haploid phase). As a result, when the haploid phase is highly reduced, the effect of expression breadth will override the effect of masking and haploid-specific genes will experience less efficient selection than diploid-specific genes. In contrast, when dominance of the haploid phase rises, haploid purging will become more efficient and expression breadth of haploid-specific genes is likely to increase (due to expression specialization). This may lead to either similar or lower evolutionary rates in haploid-specific compared with diploid-specific genes depending on the expression breadth of haploid-specific genes. Our study shows that in the moss, the opposing effect of gene expression breadth and haploid purging balance each other out because haploid- and diploid-specific genes evolve with a similar rate. This is likely due to the highly specialized haploid developmental stages of the moss life cycle (protonemata and leafy shoot) increasing expression specialization and thus evolutionary rate of genes. Therefore, when differentiation among haploid developmental stages is considerable, evolutionary rate of haploid- and diploid-specific genes is expected to be similar even in organisms with an extensive haploid phase. Nevertheless, haploid-specific genes should experience more efficient selection in species which have highly reduced diploid and an extensive haploid phase that lacks highly specialized developmental stages as it is in some red algae (Blouin et al. 2011).

Greater efficacy of selection on haploid-specific than on diploid-specific genes is a central assumption made by theoretical models of life cycle evolution in eukaryotes (Bell 1997; Thornber 2006). In such models, the mutational robustness of the diploid phase is important for the prevalence of diploid-dominant life cycles in nature (Otto and Gerstein 2008). The diploid phase is expected to accumulate a greater number of recessive deleterious mutations than the haploid phase, because such mutations can be masked by dominant alleles in a heterozygous state. In contrast to these predictions, we show that the accumulation of such mutations need not be driven by masking. This suggests that future models of life cycle evolution need to take additional other factors, such as gene expression breadth into account. The role of masking may be less important for the evolution of biphasic life cycles than commonly thought.

Conclusions

Both theoretical and empirical evidence show that the effects of ploidy on the evolutionary dynamics of DNA in unicellular organisms can be best explained by the masking hypothesis (Otto and Gerstein 2008): Selection is more efficient in haploids and leads to slower rates of evolution in haploid-specific genes. Our study shows that this is unlikely to be true in complex multicellular organisms. That is, the rate at which phase-specific genes evolve in complex multicellular organisms with biphasic life cycles contradicts the masking hypothesis. The effect of masking becomes small in such organisms and is overridden by the opposing effects of gene expression breadth.

Materials and Methods

Gene Expression Data Set

We compiled two gene expression data sets for *A. thaliana* that contain information about the phase specificity of genes (haploid-specific, diploid-specific, or unspecific expression), to find out whether our conclusions are sensitive to variation in the data. The first data set combines the AtGenExpress expression data (Schmid et al. 2005) and a data set published by Wuest at al. (2010). We obtained the MAS5 normalized AtGenExpress data from the database plexdb (http://www.plexdb.org, last accessed June 1, 2012), which contains comprehensive microarray data for almost all sporophytic developmental stages, and for the male gametophyte. The data set of Wuest et al. (2010) catalogs the male and female gametophyte transcriptomes using laser-dissected tissues samples. We used presence and absence calls to identify whether a gene is expressed in a particular developmental stage. For the AtGenExpress data set, we derived these calls from the MAS5 normalized version of the data, and combined these data with presence/absence calls from supplementary table 1 in Wuest et al. (2010) ("PANP" calls). We applied a majority consensus rule wherever replicates of the same developmental stage provided conflicting presence/absence values. In addition, wherever two or more replicate measurements did not show an unambiguous presence/absence call (referred to as M after MAS5 normalization), we removed the corresponding microarray probe from the analysis. Furthermore, we discarded probe sets mapping to more than one *Arabidopsis* Genome Initiative identifier. We will refer to the compilation of AtGenExpress data and the data by Wuest et al. (2010) as our data set 1. This data set likely misses many stage-specific genes and thus represents a minimum estimate of stage-specific expression (personal communication by S. Wuest). For this reason, we also compiled a data set 2 from published microarray data investigating haploid gene expression in *A. thaliana*. Specifically, we combined data describing gene expression in the male gametophyte (supplementary table 1 in Becker et al. 2003; supplementary table 1 in Honys and Twell 2003; supplementary table 1 in Pina et al. 2005; supplementary table 1 in Borges et al. 2008; supplementary table 1 in Wang et al. 2008) and in the female gametophyte (supplementary table 1 in Johnston et al. 2007). We retrieved normalized gene expression values from the original publications, and merged gene expression data sets by their probe set identifiers. We then discarded probe sets mapping to more than one gene identifier, and declared genes with a zero expression intensity value as lacking expression in a particular developmental stage.

We experimentally generated a generation-biased genome-wide gene expression data set for the moss *F. hygrometrica*, which is a close relative of the model moss *Physcomitrella patens* (Szövényi et al. 2011). To this end, we first collected samples of three important haploid developmental stages, specifically germinating spores, protonemata, and young gametophores (four biological replicates each). We also collected three developmental stages of the diploid phase (sporophyte), specifically sporophytes shorter than 5 mm, elongated needle-like sporophytes, and sporophytes with swollen capsules (four biological replicates each). After extraction, we pooled the four biological replicates in equimolar ratios for sequencing, and subjected the resulting six samples to single-end RNA sequencing (75 bp). Each sample was run on a single lane of an Illumina GAIIx flow cell. Sequence data obtained are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress, last accessed June 1, 2012) under accession number E-MTAB-1664.

Prior to assembly, we filtered the raw sequence data by removing all reads containing low quality (phred quality value < 20) or ambiguous ("N") base calls, using the FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/, last accessed June 1, 2012). After this quality filtering, we checked reads for possible adaptor contamination using Tagdust (Lassmann et al. 2009) by specifying a false discovery rate of 0.01, and discarded all reads with a significant match against the adaptor data base. We then assembled reads into virtual transcripts using the assembler Trinity (Grabherr et al. 2011), which produces the most contiguous transcriptome assembly (e.g., highest number of full-length transcripts with high sensitivity) for nonmodel species (Grabherr et al. 2011). We then obtained normalized gene expression estimates τ (as in Li and Dewey 2011) for each putative gene (and not for each single transcript), using the expectation maximization algorithm implemented in RNA-Seq by Expectation-Maximization (RSEM) (Li and Dewey 2011).

Estimating Evolutionary Rates

We estimated evolutionary rates of *A. thaliana* proteins by assessing their divergence from the closely-related *A. lyrata* genome. To this end, we retrieved *A. thaliana* and *A. lyrata* coding sequences and protein sequences from TAIR8 and from the draft *A. lyrata* genome, respectively (Hu et al. 2011; http://genome.jgi-psf.org/Araly1/Araly1.download.html/Araly1_GeneModels_FilteredModels6_nt.fasta, Araly1_GeneModels_FilteredModels6_aa.fasta, last accessed June 10, 2012). We identified one-to-one orthologous proteins in these genomes with a reciprocal mutual best hit strategy using BlastP (Altschul et al. 1997), and kept protein pairs showing at least 30% identity along 150 aligned amino acids for further analysis (Rost 1999). To generate codon-based alignments, we first pre-aligned protein sequences of *A. thaliana*–*A. lyrata* using MUSCLE (with default options, Edgar 2004), and then mapped these alignments onto nucleotide alignments using pal2nal (Suyama et al. 2006). We estimated evolutionary rates of proteins by computing the number of nonsynonymous substitutions per nonsynonymous site (d*N*), and the number of synonymous substitutions per synonymous site (d*S*). To estimate d*N*, d*S*, and their ratio (d*N*/d*S*), we used PAML (Yang 2007) applying the F3 × 4 model with the pairwise option (runmode = −2). We subjected the output of PAML to an extra filtering step that retained only alignments with $dS < 2$ and $dN < 2$.

In the moss *F. hygrometrica*, we estimated evolutionary rates of predicted proteins through their divergence from the *P. patens* v1.6 genome (Rensing et al. 2008; http://www.phytozome.net/). We identified one-to-one orthologs between *P. patens* proteins and *F. hygrometrica* virtual transcripts using blastx, applying the very same threshold as in *Arabidopsis*. After that, we obtained protein translations of virtual transcripts based on the orthologous *P. patens* gene models, using Wise2 (ftp://ftp.ebi.ac.uk/pub/software/unix/wise2/, last accessed June 1, 2012), and discarded virtual transcripts containing internal stop codons. We computed alignments, d*N*/d*S* estimates, and filtered the data as described for *A. thaliana* in the previous paragraph.

Expression Specificity and Its Effect on the Evolutionary Rates of Genes

In the first set of analyses, we defined phase specificity of genes as a binary variable. Using the *A. thaliana* data sets 1 and 2, we assigned genes to the three categories of haploid-specific, diploid-specific, and unspecific expression as follows. In *A. thaliana*, we called a gene haploid-specific if it was expressed in at least one haploid developmental stage, but in none of the diploid stages. We called a gene diploid-specific if it was expressed in at least one diploid but in none of the haploid developmental stages. Finally, we called a gene unspecific if it was expressed in at least one haploid and one diploid developmental stage.

In *F. hygrometrica*, we distinguished genes with diploid- from those with haploid-specific or unspecific expression by their average fold-change value between haploid and diploid tissues [$\log_2$(fold-change = gametophyte/sporophyte)]. We calculated this ratio using the average gene expression estimates obtained by RSEM (τ according to Li and Dewey 2011) for each putative gene model, and used three fold-change thresholds ($\log_2$(haploid/diploid) = 2, 4, and 6) in our analysis. We called genes with a fold-change greater than 2, 4, or 6 haploid-specific whereas genes with a fold-change threshold smaller than −2, −4, or −6 were called diploid-specific at the 3-fold-change thresholds (2, 4, and 6), respectively. The rest of the genes we assigned to the unspecific category. We applied pairwise two-tailed Wilcoxon rank sum tests (Sokal and Rohlf 2012) to compare d*N*/d*S* values among the three categories of genes (haploid-specific, diploid-specific, or unspecific) in both species.

Defining gene expression specificity as a binary variable is subjective and might introduce biases in the analysis. Therefore, we performed a second set of analyses to investigate the effect of phase-specific expression on evolutionary rates, taking into account the continuous nature of gene expression. Using gene expression as a continuous variable allowed us to apply the very same statistical methodology for both the microarray and RNA seq data sets that is expected to better account for the inherent technical differences between the two expression measurement techniques. We defined phase specificity as the average fold-change genes experience between the two phases ($\log_2$[expression in the haploid phase/expression in the diploid phase]) using the normalized gene expression values of the *A. thaliana* (data set 1) and the *F. hygrometrica* data sets. Then, we investigated the relationship between gene expression specificity and evolutionary rates of genes (d*N*/d*S*) using nonparametric correlation analysis (Sokal and Rohlf 2012).

We quantified expression breadth of genes with the index τ (Yanai et al. 2005) to investigate the relationship between expression breadth and relative complexity of the phases:

$$\tau = \frac{\sum_{j=1}^{n}\left[1-\log_2 S(i,j)/\log_2 S(i,\max)\right]}{n-1}$$

where n refers to the number of tissues, $S(i,\max)$ is the highest expression of gene i across all tissues and $S(i,j)$ is the expression of gene i in the jth tissue. τ approaches 1 when the gene is exclusively expressed in one tissue and 0 if it is equally expressed across all tissue types investigated. We used gene expression estimates of data set 1 (for *A. thaliana*) and expression estimates obtained by RSEM (τ according to Li and Dewey 2011 for *F. hygrometrica*) to quantify the expression breadth (τ) of genes.

The Causes of Evolutionary Rate Difference between Specific and Unspecific Genes

To assess the causes of evolutionary rate difference among haploid-specific, diploid-specific, or unspecifically expressed genes, we performed a direct test using polymorphism data and an indirect test using only divergence data. We performed the direct test only for *A. thaliana*, using publicly available genome-wide polymorphism data. To distinguish positive from relaxed selection, we used the 19 *Arabidopsis* genome data set, obtaining consolidated protein coding DNA sequences and protein sequences from the 19 *Arabidopsis* genome site (Gan et al. 2011; http://mus.well.ox.ac.uk/19genomes/, last accessed May 3, 2012). With this data in hand, we paired genes occurring in all 19 *Arabidopsis* accessions with their previously identified *A. lyrata* ortholog, and aligned protein sequences using muscle (Edgar 2004) with default parameters. We then used the resulting protein alignments to guide nucleotide alignments with pal2nal (Mikita et al. 2006).

To investigate whether differences in the d*N*/d*S* ratio are due to relaxed selection or to the fixation of beneficial mutations, we estimated the ratio of nonsynonymous to synonymous polymorphisms for each gene using MK.pl (Holloway et al. 2007). Beneficial mutations are expected to rapidly reach fixation within species whereas slightly deleterious mutations should segregate within species. Therefore, if elevated among-species d*N*/d*S* ratios are mainly due to relaxed selection, a higher nonsynonymous to synonymous polymorphisms ratio is expected for haploid-specific than for diploid-specific genes. In contrast, when elevated among-species d*N*/d*S* ratios are caused by the fixation of beneficial mutations, species-wide nonsynonymous to synonymous polymorphism ratios should be lower in haploid-specific than in diploid-specific genes. Finally, we compared $\log_{10}$-transformed values of these ratios among the three major gene categories (haploid-specific, diploid-specific, or unspecific) using Wilcoxon rank sum tests (Sokal and Rohlf 2012).

We performed the second, indirect test for both species. This test only relies on divergence data, as no polymorphism data are available for the moss *F. hygrometrica*. It uses the strength of correlation between gene expression and evolutionary rate (d*N*/d*S*) as an indicator of the efficacy of selection (Koonin 2011). We calculated nonparametric Spearman rank correlations between mean gene expression levels (averaged over all developmental stages) and evolutionary rates d*N*/d*S* for the three main categories of genes (haploid-specific, diploid-specific, or unspecific), and compared these nonparametric correlation coefficients among the three categories using a *Z*-test (Sokal and Rohlf 2012).

Controlling for the Influence of Confounding Factors

First, we investigated this question defining phase specificity as a binary variable. We used ANCOVA to investigate whether evolutionary rate difference among genes with haploid-specific, diploid-specific, or unspecific expression can be accounted for by other factors not accounted for in our major hypotheses (Slotte et al. 2011; Yang and Gaut 2011). To this end, we retrieved structural properties of genes that are known to correlate with evolutionary rate from the general feature files (gff) of the *A. thaliana* (TAIR8) and *P. patens* (v6) genomes. Then, we assessed the differences in the ratio d*N*/d*S* among the three gene categories by including total gene length, GC content, and average intron length as covariates in the ANCOVA (Sokal and Rohlf 2012). To fulfill assumptions of the test, we $\log_{10}$ transformed d*N*/d*S* values. We conducted the ANCOVA on both *A. thaliana* data sets (data set 1 and 2) and on the *F. hygrometrica* data set with a fold-change [$\log_2$(fold-change)] threshold of 4.

We also investigated the effect of confounding factors not accounted for in our primary analyses, treating gene expression as a continuous variable and defining phase specificity as the average fold-change genes experience between the two phases ($\log_2$[expression in the haploid phase/expression in the diploid phase]). We performed partial nonparametric correlation analysis (Sokal and Rohlf 2012) between expression specificity and evolutionary rates (d*N*/d*S*), whereas controlling for the effect of average gene expression intensity, total gene length, GC content, and average intron length. Properties of genes were retrieved from the appropriate general feature files (gff) (discussed earlier).

Molecular function is known to affect evolutionary rates of genes. Therefore, we asked whether genes with haploid-specific, diploid-specific, or unspecific expression show considerably different molecular functions. This analysis was only conducted for data set 2 of *A. thaliana* because the number of phase-specific genes with GO annotation prohibited a meaningful statistical analysis in data set 1. In *F. hygrometrica*, a threshold value of $\log_2$(fold-change) = 4 was used to define genes with phase-specific and unspecific expression. GO annotation for each gene was retrieved from publicly available annotation files (*A. thaliana*: ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Athaliana/annotation/Athaliana_167_annotation_info.txt.gz, last accessed June 1, 2012; *P. patens*: ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Ppatens/annotation/Ppatens_152_annotation_info.txt.gz, last accessed June 1, 2012). After that, we compared the functional annotation of haploid-specific, diploid-specific, and unspecific gene groups using Fisher's exact tests (applying Bonferroni correction) for each GO term separately. In this analysis, we only used GO terms on level two of the molecular function ontology. We performed all statistical analyses using R (R Development Core Team 2011).

Supplementary Material

Supplementary tables S1 and S2 and figure S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the Swiss National Science Foundation Ambizione grant PZ00P3_131726 to P.S., Swiss National Science Foundation grants 315230-129708, the YeastX project of SystemsX.ch, and the University Priority Research Program in Systems Biology at the University of Zurich.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389–3402.

Becker JD, Boavida LC, Carneiro J, Haury M, Feijo JA. 2003. Transcriptional profiling of *Arabidopsis* tissues reveals the unique characteristics of the pollen transcriptome. *Plant Physiol.* 133: 713–725.

Bell G. 1997. The evolution of the life cycle of brown seaweeds. *Biol J Linn Soc.* 60:21–38.

Bell G. 2008. Selection: the mechanism of evolution, 2nd edition. Oxford: Oxford University Press.

Blouin NA, Brodie JA, Grossman AC, Xu P, Brawley SH. 2011. *Porphyra*: a marine crop shaped by stress. *Trends Plant Sci.* 16:29–37.

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.

Byers DL, Waller DM. 1999. Do plant populations purge their genetic load? Effects of population size and mating history on inbreeding depression. *Annu Rev Ecol Evol Syst.* 30:479–513.

Charlesworth D, Charlesworth B. 1992. The effects of selection in the gametophyte stage on mutational load. *Evolution* 46:703–720.

Charlesworth D, Willis J. 2009. The genetics of inbreeding depression. *Nat Rev Genet.* 10:783–796.

Cook DL, Gerber AN, Tapscott SJ. 1998. Modeling stochastic gene expression: implications for haploinsufficiency. *Proc Natl Acad Sci U S A.* 95:15641–15646.

Destombe C, Godin J, Nocher M, Richerd S, Valero M. 1993. Differences in response between haploid and diploid isomorphic phases of *Gracilaria verrucosa* (Rhodophyta: Gigartinales) exposed to artificial environmental conditions. *Hydrobiologia* 261:131–137.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.

Gan X, Stegle O, Behr J, et al. (23 co-authors). 2011. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* 477: 419–423.

Gene Ontology Consortium. 2005. Gene ontology: tool for the unification of biology. *Nat Genet.* 25:25–29.

Gerstein AC, Cleathero LA, Mandegar MA, Otto SP. 2011. Haploids adapt faster than diploids across a range of environments. *J Evol Biol.* 24:531–540.

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.

Grabherr MG, Haas BJ, Yassour M, et al. (21 co-authors). 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29:644–652.

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.

Hafidh S, Breznenová K, Růžička P, Feciková J, Capková V, Honys D. 2012. Comprehensive analysis of tobacco pollen transcriptome unveils common pathways in polar cell expansion and underlying heterochronic shift during spermatogenesis. *BMC Plant Biol.* 12:24.

Henter HJ. 2003. Inbreeding depression and haplodiploidy: experimental measures in a parasitoid and comparisons across diploid and haplodiploid insect taxa. *Evolution* 57:1793–1803.

Holloway AK, Lawniczak MKN, Mezey JG, Begun DJ, Jones CD. 2007. Adaptive gene expression divergence inferred from population genomics. *PLoS Genet.* 3:2007–2013.

Honys D, Twell D. 2003. Comparative analysis of the *Arabidopsis* pollen transcriptome. *Plant Physiol.* 132:640–652.

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.

Johnston AJ, Meier P, Gheyselinck J, Wuest SE, Federer M, Schlagenhauf E, Becker JD, Grossniklaus U. 2007. Genetic subtraction profiling identifies genes essential for *Arabidopsis* reproduction and reveals interaction between the female gametophyte and the maternal sporophyte. *Genome Biol.* 8:R204.

Kondrashov A, Crow J. 1991. Haploidy or diploidy—which is better. *Nature* 351:314–315.

Koonin EV. 2011. Are there laws of genome evolution? *PLoS Comput Biol.* 7:e1002173.

Krumbiegel R. 1979. Response of haploid and diploid protoplasts from *Datura innoxia* Mill. and *Petunia hybrida* L. to treatment with X-rays and a chemical mutagen. *Environ Exp Bot.* 19:99–103.

Lassmann T, Hayashizaki Y, Daub CO. 2009. TagDust—a program to eliminate artifacts from next generation sequencing data. *Bioinformatics* 25:2839–2840.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.

Ma J, Skibbe DS, Fernandes J, Walbot V. 2008. Male reproductive development: gene expression profiling of maize anther and pollen ontogeny. *Genome Biol.* 9:R181.

Mable BK, Otto SP. 2001. Masking and purging mutations following EMS treatment in haploid, diploid and tetraploid yeast (*Saccharomyces cerevisiae*). *Genet Res.* 77:9–26.

Muralla R, Lloyd J, Meinke D. 2011. Molecular foundations of reproductive lethality in *Arabidopsis thaliana*. *PLoS One* 6:e28398.

Onodera Y, Nakagawa K, Haag JR, Pikaard D, Mikami T, Ream T, Ito Y, Pikaars CS. 2008. Sex-biased lethality or transmission of defective transcription machinery in *Arabidopsis*. *Genetics* 180:207–218.

Orr HA, Otto SP. 1994. Does diploidy increase the rate of adaptation? *Genetics* 136:1475–1480.

Otto SP. 2004. Two steps forward, one step back: the pleiotropic effects of favoured alleles. *Proc Biol Sci.* 271:705–714.

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

Park SG, Choi SS. 2010. Expression breadth and expression abundance behave differently in correlations with evolutionary rates. *BMC Evol Biol.* 10:241.

Pina C, Pinto F, Feijo 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.

R Development Core Team. 2011. R: a language and environment for statistical computing [Internet]. Vienna (Austria): R Foundation for Statistical Computing [cited 2013 May 3]. Available from: http://www.R-project.org/

Rensing SA, Lang D, Zimmer AD, et al. (70 co-authors). 2008. The genome of the moss *Physcomitrella patens* reveals evolutionary insights into the conquest of land by plants. *Science* 319:64–69.

Rost B. 1999. Twilight zone of protein sequence alignments. *Protein Eng.* 12:85–94.

Russell SD, Gou X, Wong CE, Wang X, Yuan T, Wei X, Bhalla PL, Singh MB. 2012. Genomic profiling of rice sperm cell transcripts reveals conserved and distinct elements in the flowering plant male germ lineage. *New Phytol.* 195:560–573.

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 J. 2005. A gene expression map of *Arabidopsis* development. *Nat Genet.* 37:501–506.

Seoighe C, Gehring C, Hurst LD. 2005. Gametophytic selection in *Arabidopsis thaliana* supports the selective model of intron length reduction. *PLoS Genet.* 1:e13.

Shaw AJ, Beer SC. 1997. Gametophyte-sporophyte variation and covariation in mosses. *Adv Bryol.* 6:35–63.

Shaw J, Szövényi P, Shaw B. 2011. Bryophyte diversity and evolution: windows into the early evolution of land plants. *Am J Bot.* 98:1–18.

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

Sokal RR, Rohlf FJ. 2012. Biometry: the principles and practice of statistics in biological research, 4th ed. New York: W. H. Freeman and Co.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34:W609–W612.

Szövényi P, Rensing SA, Lang D, Shaw AJ, Wray G. 2011. Generation-biased gene expression in a bryophyte model system. *Mol Biol Evol.* 28:803–812.

Thornber CS. 2006. Functional properties of the isomorphic biphasic algal life cycle. *Integr Comp Biol.* 46:605–614.

Tortajada AM, Carmona MJ, Serra M. 2009. Does haplodiploidy purge inbreeding depression in rotifer populations? *PLoS One* 4:e8195.

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.

Wang Z, Zhang J. 2011. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *Proc Natl Acad Sci U S A.* 108:E67–E76.

Whittle CA, Johnston MO. 2003. Male-biased transmission of deleterious mutations in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A.* 100: 4055–4059.

Wolff P, Weinhofer I, Seguin J, Roszak P, Beisel C, Donoghue MT, Spillane C, Nordborg M, Rehmsmeier M, Köhler C. 2011. High-resolution analysis of parent-of-origin allelic expression in the *Arabidopsis* endosperm. *PLoS Genet.* 6:e102126.

Woody JL, Shoemaker RC. 2011. Gene expression: sizing it all up. *Front Genet.* 2:70.

Wright SI, Andolfatto P. 2008. The impact of natural selection on the genome: emerging patterns in *Drosophila* and *Arabidopsis*. *Annu Rev Ecol Evol Systemat.* 39:193–213.

Wuest SE, Vijverberg K, Schmidt A, Weiss M, Gheyselinck J, Lohr M, Wellmer F, Rahnenfuhrer 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.

Yanai I, Benjamin H, Shmoish M, et al. (12 co-authors). 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* 21:650–659.

Yang L, Gaut BS. 2011. Factors that contribute to variation in evolutionary rate among *Arabidopsis* genes. *Mol Biol Evol.* 28:2359–2369.

Yang Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.

Yin S, Wang P, Deng W, Zheng H, Hu L, Hurst LD, Kong X. 2009. Dosage compensation on the active X chromosome minimizes transcriptional noise of X-linked genes in mammals. *Genome Biol.* 10:R74.

Zeyl C, Vanderford T, Carter M. 2003. An evolutionary advantage of haploidy in large yeast populations. *Science* 299:555–558.