BMC Genomics

Open Access

CrossMark

Transcriptome sequencing reveals genome-wide variation in molecular evolutionary rate among ferns

Amanda L. Grusz[1,3*], Carl J. Rothfels[2] and Eric Schuettpelz[1]

Abstract

Background: Transcriptomics in non-model plant systems has recently reached a point where the examination of nuclear genome-wide patterns in understudied groups is an achievable reality. This progress is especially notable in evolutionary studies of ferns, for which molecular resources to date have been derived primarily from the plastid genome. Here, we utilize transcriptome data in the first genome-wide comparative study of molecular evolutionary rate in ferns. We focus on the ecologically diverse family Pteridaceae, which comprises about 10 % of fern diversity and includes the enigmatic vittarioid ferns—an epiphytic, tropical lineage known for dramatically reduced morphologies and radically elongated phylogenetic branch lengths. Using expressed sequence data for 2091 loci, we perform pairwise comparisons of molecular evolutionary rate among 12 species spanning the three largest clades in the family and ask whether previously documented heterogeneity in plastid substitution rates is reflected in their nuclear genomes. We then inquire whether variation in evolutionary rate is being shaped by genes belonging to specific functional categories and test for differential patterns of selection.

Results: We find significant, genome-wide differences in evolutionary rate for vittarioid ferns relative to all other lineages within the Pteridaceae, but we recover few significant correlations between faster/slower vittarioid loci and known functional gene categories. We demonstrate that the faster rates characteristic of the vittarioid ferns are likely not driven by positive selection, nor are they unique to any particular type of nucleotide substitution.

Conclusions: Our results reinforce recently reviewed mechanisms hypothesized to shape molecular evolutionary rates in vittarioid ferns and provide novel insight into substitution rate variation both within and among fern nuclear genomes.

Keywords: *Adiantum*, Molecular evolution, Nucleotide substitution rate, Vittarioid ferns

Background

Improvements in technology and subsequent, rapidly diminishing costs have revolutionized molecular analyses of non-model lineages, such as ferns, and studies of these groups are finally moving into the "next generation" sequencing era [1–7]. Yet, even with the noted advances, researchers working on ferns have been somewhat limited in their ability to obtain and use genome-scale data, due primarily to the absence of an annotated genome sequence and difficulties posed by the large, highly repetitive genomes of most fern species [3, 7–9]. Fortunately, RNA sequencing of expressed genes can offer reasonable first insights into the evolution of fern genomes through the comparative lens of transcriptomics [8–19].

Among the many outstanding aspects of fern evolutionary biology that can be explored with transcriptome data are inconsistencies in molecular evolutionary rate across the fern tree of life [20–26]. Although especially pronounced in ferns, this phenomenon, of course, is not unique to the group. Unequal rates of sequence evolution are well documented throughout eukaryotes, both among lineages [27–34] and within individual genomes [35–38]. Many explanations for unequal rates of molecular evolution have been proposed, ranging from variation in generation time and fluctuations in mass-specific

* Correspondence: grusza@d.umn.edu
[1]Department of Botany, Smithsonian Institution, MRC 166 PO Box 37012, Washington, DC 20013-7012, USA
[3]Department of Biology, University of Minnesota Duluth, 1035 Kirby Drive, Duluth, MN 55812, USA
Full list of author information is available at the end of the article

Grusz *et al. BMC Genomics* (2016) 17:692

Page 2 of 11

metabolic rate to disparities in latitude and environmentally available energy among lineages [27, 39–50]. But, the environmental and life history traits that are most commonly correlated with molecular evolutionary rate vary among branches of the eukaryotic phylogeny. In animals, evidence indicates that generation time or, more accurately, the rate of germline DNA replication [51] may be a strong driver of evolutionary rate [52, 53]. Other aspects, such as disparity in body size and the presence of damaging by-products of cellular metabolism, have also been implicated [40, 52].

Meanwhile, factors thought to influence molecular evolutionary rate in plants, particularly angiosperms, include environmental variables with potential mutagenic effects, like ultraviolet radiation and temperature [42, 44, 54, 55], as well as damaging metabolic byproducts and inefficient DNA repair [49]. Demographic effects involving population size, life history strategy (e.g., annual vs. perennial), habit (e.g., herbaceous vs. woody), and rates of mitosis have also been put forth to explain unequal evolutionary rates across plants [26, 33, 41, 56]. Although a consensus has not been reached regarding which of these factors most consistently influences substitution rates in plants, patterns of rate variation in many groups studied to date appear to be shared across cellular compartments (e.g., in bladderworts [32, 49] and *Veronica* [33]; but see [57]).

Within ferns, multiple lineages exhibit extreme heterogeneity in molecular evolutionary rate, most notably the filmy ferns [21], tree ferns [24], and vittarioid ferns [25]. But, in the studies to date, significant differences in lineage-specific evolutionary rate have been detected almost exclusively through comparisons of plastid gene regions. The one investigation to incorporate non-plastid loci examined just one gene from the nuclear genome and two from the mitochondrial genome [25], leaving significant shifts in evolutionary rates within and among the nuclear genomes of ferns effectively unexplored.

Here, we employ comparative transcriptomics to test whether previous observations of unequal molecular evolutionary rates among species at plastid loci are mirrored across the nuclear genome. We focus on the Pteridaceae, a cosmopolitan fern family comprising about 50 genera and more than 1000 species (roughly 10% of documented fern diversity) with ecological specializations to rupestral, aquatic, epiphytic, and terrestrial habitats [22, 23]. Taking a pairwise approach, we look for significant differences in rates of molecular evolution among the genomes of 12 species distributed broadly across the 3 largest clades in the family [22, 23] and evaluate life history characteristics potentially affecting these patterns. We focus particular attention on the *Adiantum* and vittarioid sister lineages, in an attempt to uncover genomic clues that may help to explain their disparate morphologies, ecologies, and life histories [25].

By examining 2091 nuclear gene regions, this study offers insight into the evolutionary drivers of divergence between *Adiantum* and the vittarioid ferns and sheds light on the mechanisms influencing evolutionary rates in the cosmopolitan fern family, Pteridaceae.

Results
Comparing relative substitution rates
Relative rates of molecular evolution were examined at 2091 loci for 12 taxa from across the Pteridaceae (Additional file 1) [23]. We found a multitude of significant differences in overall molecular substitution rate among sampled taxa (Fig. 1; Additional file 2). For most (but not all) loci examined, the vittarioid ferns (v) consistently had the fastest relative substitution rates (Fig. 1). Representatives of the pteridoid clade (P, Fig. 1), *Pityrogramma* in particular, also showed fast substitution rates relative to other taxa (excluding the vittarioid ferns). The xeric-adapted cheilanthoid ferns (C), especially *Myriopteris*, exhibited consistently lower rates than other members of the Pteridaceae sampled. Separate pairwise comparisons of rate for individual substitution types (Additional file 3), nonsynonymous substitution rates (dN, Fig. 2; Additional file 4), and synonymous substitution rates (dS, Fig. 2; Additional file 5) each presented patterns of rate variation that were consistent with our pairwise comparisons of overall substitution rates (Fig. 1; Additional file 2).

GO enrichment analyses
To detect whether any gene families contribute disproportionately to substitution rate variation for vittarioid ferns, we conducted gene ontology (GO) annotations (Additional files 6 and 7) and GO enrichment analyses. Specifically, we tested whether any particular GO categories were over- or underrepresented among the loci for which vittarioids had significantly faster or slower substitution rates. Because there is no annotated fern genome available, many orthogroups in our study did not map to a known GO category. For this reason, we carried out two separate tests of enrichment/purification for the faster loci and two for the slower loci: one that included all significantly faster or slower orthogroups, regardless of annotation status, and a second that included only those significantly faster or slower loci with known GO annotations. In our analyses blind to annotation status, we used a reference set (RS) including all 2091 loci (RS1; Additional file 8); in our analyses considering only annotated genes, we used a RS including only the subset of total loci with known annotations (RS2; Additional file 8). Thus we conducted four separate analyses, one for each test set (TS): all significantly faster loci (TS1; Additional file 8) versus RS1; significantly faster loci with annotations (TS2; Additional file 8) versus RS2; all significantly slower loci (TS3) versus RS1; and significantly slower loci with annotations
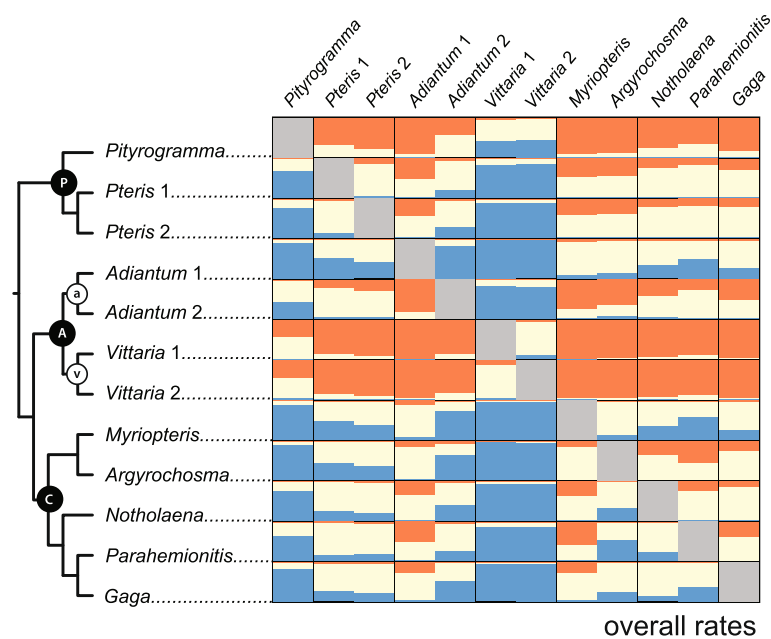
Grusz *et al. BMC Genomics* (2016) 17:692

Page 3 of 11

Fig. 1 Summary of pairwise comparisons of overall substitution rates in the Pteridaceae for 2091 loci. Each box is colored based on the proportion of loci for which the row taxon is significantly slower (blue), faster (orange), or not significantly different (cream) than the corresponding column taxon. P = pteridoid clade, A = adiantoid clade, C = cheilanthoid clade, a = *Adiantum* clade, and v = vittarioid clade [22, 23]. For numerical values, see Additional file 2

(TS4) versus RS2 (see Additional file 8). When we restricted comparisons to those with GO annotations (TS2 vs. RS2 and TS4 vs. RS2), no GO categories showed evidence of being significantly enriched or purified among faster/slower vittarioid loci. However, when all loci were included in the reference set, whether annotated or not (RS1), one GO category was significantly purified among those annotated loci for which vittarioid ferns were significantly slower than other taxa (TS3): *GO:0050896: response to stimulus: physiological response to stimulus [exact]*.

Tests for selection

In an effort to distinguish whether selective forces could be driving the fast substitution rates observed for vittarioid ferns, we also tested for significant differences in the ratio of nonsynonymous to synonymous substitutions (i.e., dN/dS or ω) for each of the 2091 orthogroups in our study. We used model comparisons to determine if dN/dS for the vittarioid crown group (ω_V) and/or its stem branch (ω_B) was statistically distinct from the rest of the phylogeny (ω_T). For a plurality of loci (42.4 %), the best fitting model applied a single (shared) dN/dS across the three partitions ($\omega_T = \omega_V = \omega_B$; Fig. 3; Table 1), with vittarioids apparently not experiencing unique selective pressures. For an additional 22.3 % of loci analyzed, a two-rate model allowing ω_V to vary relative to the rest of the phylogeny ($\omega_T = \omega_B \neq \omega_V$; Fig. 3; Table 1) best explained the data. Rates for roughly equal proportions of orthogroups (15.2 and 14.0 %; Fig. 3; Table 1) were best explained by a two-

rate model with either a unique dN/dS for the vittarioid stem branch ($\omega_T = \omega_V \neq \omega_B$) or the remainder of the tree ($\omega_T \neq \omega_V = \omega_B$), respectively. The three-rate model ($\omega_T \neq \omega_V \neq \omega_B$) best explained the data for the fewest orthogroups (5.9 %; Fig. 3; Table 1).

Regardless of which dN/dS model best explained the data for a given orthogroup and its corresponding partitions, the vast majority of regions analyzed had estimated ω values that were significantly less than 1 across all partitions, indicating that purifying selection has most likely been acting (Table 1). Neutrality was identified as the second most common evolutionary regime, with dN/dS estimates for ca. 3 % of loci failing to reject neutrality for at least one partition (i.e., with ω values not significantly different from 1; Table 1). Evidence for positive selection (i.e., an ω significantly greater than 1) was detected in only a handful (5; <1 %) of the loci surveyed (Table 1). Among the loci for which evidence of positive selection was detected, most had no known function based on existing GO annotations. Only a single locus determined to be under positive selection (ω_B partition, model: $\omega_T \neq \omega_V \neq \omega_B$; Table 1) mapped to a known GO category: *GO:0009987: cellular process: cell growth and/or maintenance [narrow]; cell physiology [exact]; cellular physiological process [exact]*.

Discussion

Here, we examine variation in molecular evolutionary rate among taxa spanning the three largest clades within the ecologically diverse fern family Pteridaceae [23].
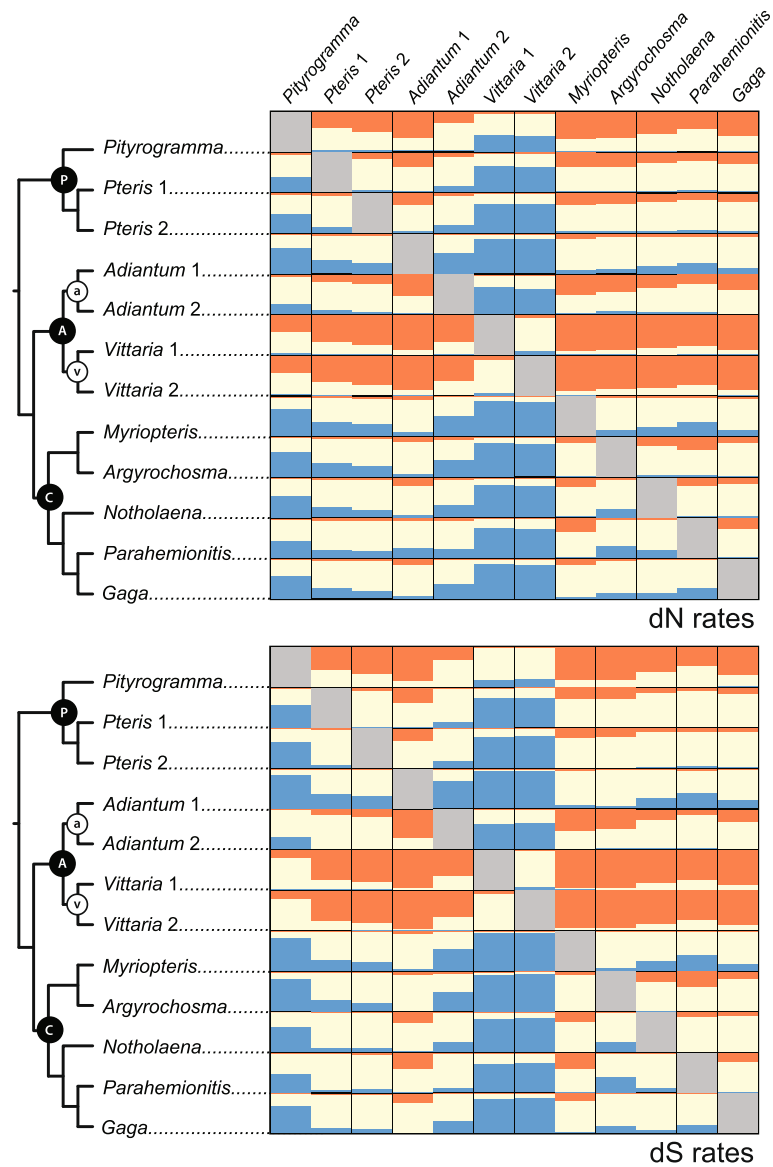
Grusz *et al. BMC Genomics* (2016) 17:692

Page 4 of 11

Fig. 2 Summary of pairwise comparisons of nonsynonymous (dN) and synonymous (dS) substitution rates in the Pteridaceae for 2091 nuclear loci. Each box is colored based on the proportion for which the row taxon is significantly slower (blue), faster (orange), or not significantly different (cream) than the corresponding column taxon. P = pteridoid clade, A = adiantoid clade, C = cheilanthoid clade, a = *Adiantum* clade, and v = vittarioid clade [22, 23]. For numerical values, see Additional files 4 and 5

Previous studies based primarily on plastid data uncovered considerable rate heterogeneity within this family, especially with respect to the morphologically simplified, epiphytic vittarioid lineage [23, 25]. However, it was unclear whether such patterns extended to the nuclear genomes in these ferns [58]. Using pairwise rate comparisons for 2091 nuclear loci, we ask whether the significant variation in molecular evolutionary rates previously reported for plastid loci is mirrored in fern nuclear genomes. Given the extremely fast plastid gene substitution rates that characterize vittarioid ferns, we focus particular attention on the nuclear gene substitution rates for members of this lineage.

Our results show that the variation in molecular evolutionary rate previously detected among members of the Pteridaceae (vittarioid ferns in particular) at plastid loci is reflected in their nuclear genomes (Figs. 1 and 2). Furthermore, we find that the vittarioid ferns are not alone in having significantly faster substitution rates (relative to other members of the family). Pteridoid ferns (*Pityrogramma*, especially) and at least one member of the *Adiantum* clade (*Adiantum* 2) are also characterized
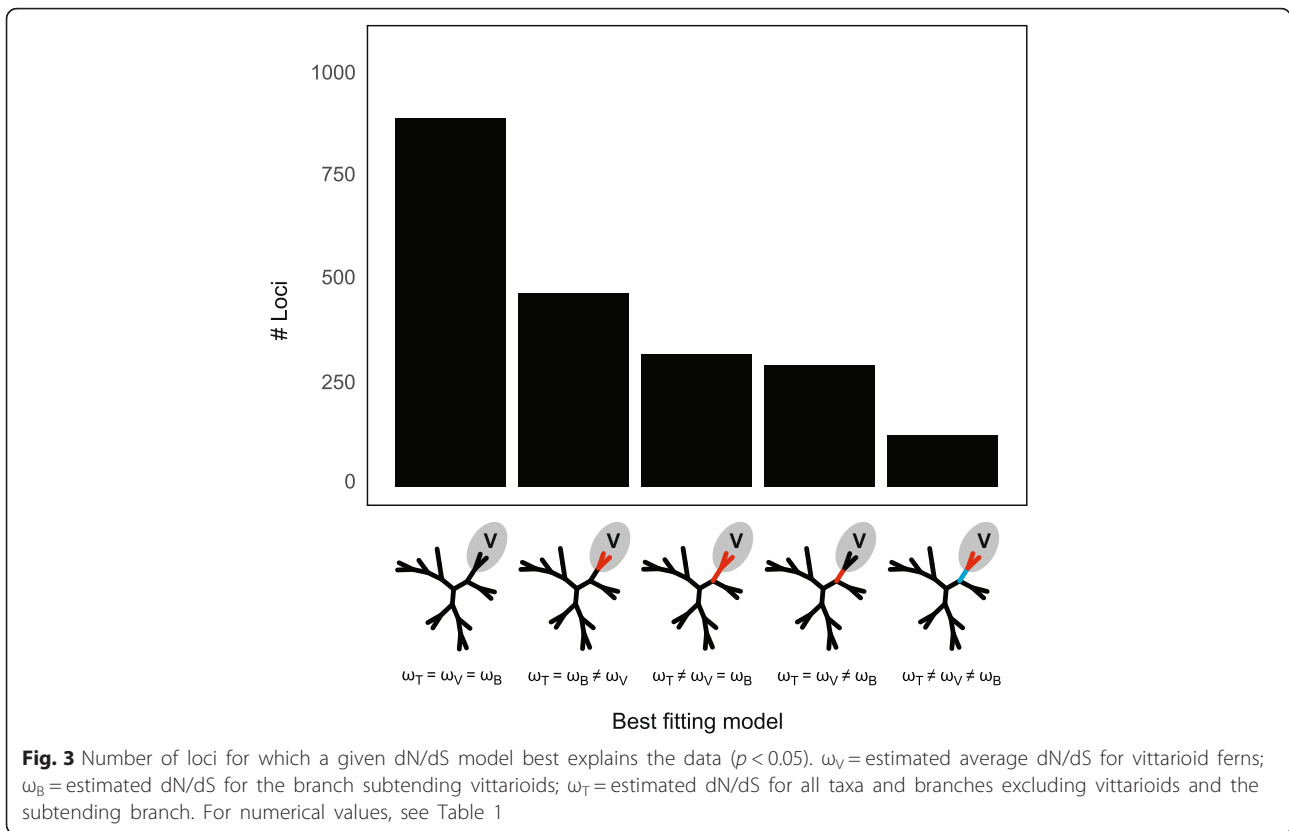
Grusz *et al. BMC Genomics* (2016) 17:692

Page 5 of 11

Fig. 3 Number of loci for which a given dN/dS model best explains the data ($p < 0.05$). ω_V = estimated average dN/dS for vittarioid ferns; ω_B = estimated dN/dS for the branch subtending vittarioids; ω_T = estimated dN/dS for all taxa and branches excluding vittarioids and the subtending branch. For numerical values, see Table 1

Table 1 Best-fitting models of dN/dS (ω) across 2091 loci analyzed. For each model, the absolute number (and proportion) of loci for which that model best explained the data (according to AIC) is provided[a]

Model	# Loci	Positive selection $\omega > 1$	Neutral $\omega = 1$	Purifying selection $\omega < 1$
$\omega_T = \omega_V = \omega_B$	888 (42.4 %)	1 (0.11 %)	0	887 (99.9 %)
$\omega_T = \omega_B \neq \omega_V$	467 (22.3 %)			
$\omega_T = \omega_B$		0	0	467 (100 %)
ω_V		3 (0.64 %)	49 (10.5 %)	415 (88.9 %)
$\omega_T \neq \omega_V = \omega_B$	318 (15.2 %)			
ω_T		0	0	318 (100 %)
$\omega_V = \omega_B$		0	2 (0.62 %)	316 (99.4 %)
$\omega_T = \omega_V \neq \omega_B$	294 (14.0 %)			
$\omega_T = \omega_V$		0	0	294 (100 %)
ω_B		1 (0.34 %)	2 (0.68 %)	291 (98.9 %)
$\omega_T \neq \omega_V \neq \omega_B$	124 (5.9 %)			
ω_T		0	0	124 (100 %)
ω_V		0	8 (6.45 %)	116 (93.5 %)
ω_B		1 (0.81 %)	1 (0.81 %)	122 (98.4 %)

[a]ω_T = estimated average dN/dS across all sampled taxa with the exclusion of vittarioid ferns and their subtending branch; ω_V = estimated average dN/dS for vittarioid ferns; ω_B = estimated dN/dS for branch subtending vittarioid ferns (see Fig. 3)

by relatively fast substitution rates (Figs. 1 and 2). Results from our model comparisons (ω) also indicate that unique selective pressures are probably not driving a rate increase at a plurality of loci in vittarioid ferns (Fig. 3).

While variation in molecular evolutionary rate has been documented across the tree of life, factors influencing substitution rates (e.g., body size, reproductive strategies, and life history) differ among lineages [40, 41, 48].

Metabolism and evolutionary rates

Mutations are the fundamental fuel for molecular evolution, but the efficiency of DNA repair and the rate of fixation ultimately drive rates of molecular substitution across organismal genomes [35, 59]. Furthermore, many aspects of an organism's biology that are critical to its survival can play an important role in generating mutations. One prominent example is the production of reactive oxygen species as a byproduct of cellular metabolism [40]. Because mitochondria are the sites of metabolism, they are expected to be a prime location for oxidative DNA damage [52, 60]. Given this, if the production of reactive oxygen species via metabolism is a major factor influencing substitution rates, we should expect increased substitution rates, especially with respect to $CC \rightarrow TT$ mutations, in mitochondrial genes [48, 61]. Some studies have found a strong signal for metabolic effects, or correlates thereof [62, 63], on mitochondrial substitution rates (but see [53, 64]), but doubt remains as to whether such effects extend to the nuclear genome [65–67].

Contrary to expectations based on the mitochondrial hypothesis of elevated evolutionary rates, Rothfels and Schuettpelz [25] found that increased substitution rates observed in plastid genes of vittarioid ferns were mostly not reflected in the two mitochondrial genes examined. Complementing this finding, our own comparisons of rate variation for individual nucleotide substitution types across nuclear genes in vittarioid ferns (and the Pteridaceae as a whole) do not show an overabundance of $CC \rightarrow TT$ mutations (Additional file 3). These results, in combination, indicate that mutagenic metabolic byproducts are probably not driving increased evolutionary rates in our focal group.

Among the most important outcomes of our study is the observation that patterns of rate heterogeneity among adiantoid (A) plastid genomes [25] are also detected in nuclear genome comparisons (Figs. 1 and 2, Additional file 3). Such parallel rate variation across cellular compartments is in conflict with the notion of DNA polymerase fidelity being a primary driver of molecular evolutionary rates in this group, owing to the fact that different DNA polymerases function in nuclei and plastids [41]. Therefore, causes of molecular evolutionary rate heterogeneity that result in asymmetrical differences among cellular compartments can be reasonably disregarded.

Influence of generation time and DNA replication on rates

Like metabolism, DNA transcription, replication, and repair are essential yet potentially mutagenic biological processes [48, 68, 69]. Because animals retain a sequestered germ line, DNA replication has been frequently invoked to explain the strong correlation between their generation times and evolutionary rates [52, 53]. Plants, however, lack an early-sequestered germ line, so rates of mitotic division in the apical meristem are considered a better proxy for DNA replication frequency than actual time to reproduction [56]. Unfortunately, estimates for rates of mitosis in plants are lacking and the situation is further complicated in ferns, with time to reproduction also encompassing the period prior to fertilization when these plants exist in their independent haploid (gametophyte) life phase.

Some authors have hypothesized that ferns with an especially prolonged gametophyte phase may have an increased probability of fixation of somatic mutations in the progeny of meristematic cells [70, 71]. While estimates of overall generation time or rate of mitotic division in ferns are lacking, some estimates exist for the relative duration of the gametophyte life history stage [72]. Interestingly, vittarioid ferns, which we find to have the highest substitution rates within the Pteridaceae, are known for their long-lived, vegetatively reproducing gametophytes [72]. It is reasonable, then, to further consider the roles played by longevity and/or mitotic frequency in guiding evolutionary rates within ferns.

Genome duplication and evolutionary rates

Not surprisingly, many facets of genome biology correlate with evolutionary rates, such as levels of gene expression (negatively correlated [73]), degree of promoter region methylation (positively correlated [74]), and extent of gene duplication (positively correlated [38, 75]). Additionally, genome structure itself (e.g., the position of plastid inverted repeats) has recently been shown to influence molecular substitution rates [76]. Our data indicate that genome duplication, in particular, may be an important characteristic influencing evolutionary rates in ferns. When viewed in light of published chromosome counts [77], our results reveal that taxa with the highest relative substitution rates within the Pteridaceae—*Vittaria, Pityrogramma, Pteris,* and *Adiantum* 2—tend to exhibit high ploidy levels compared to most other species sampled.

Environmental influences on substitution rate in ferns

In this study, sampled taxa are specialized to a variety of environmental conditions, from the xeric-adapted cheilanthoid ferns (*Argyrochosma, Gaga, Myriopteris,*

Grusz *et al. BMC Genomics* (2016) 17:692

Page 7 of 11

Notholaena, *Parahemionitis*), to terrestrial/rupestral members of the *Adiantum* and pteridoid clades, and even epiphytic, tropical lineages comprising vittarioid ferns. Abiotic and/or environmental variables thought to affect mutation rates, as well as rates of subsequent fixation, are numerous. Among them, the influences of latitude [44] (but see [78]), habitat [79, 80], environmentally available energy [45], temperature [42–44], ultraviolet radiation [44, 81], and species interactions [54] have gained the most traction as possible explanations for evolutionary rate discrepancies. In particular, De Vries [82] speculated that water stress could decrease the rate of metabolic activity in plants via reduced respiration. For this reason, if metabolism were a significant source of DNA substitutions, we might expect that ferns growing in drought-prone habitats would have reduced substitution rates. Our results do reveal decreased evolutionary rates across the xeric-adapted cheilanthoid ferns sampled, in agreement with De Vries' hypothesis. However, epiphytes (like the fast evolving vittarioids) are also known for their adaptations to drought, even in tropical environments [72]. This inconsistency, combined with a general lack of support for a metabolic explanation (as noted above) only reinforces the conclusion that oxidative damage is probably not a major factor contributing to rate variation in ferns.

Another important environmental engine potentially driving an increase in DNA mutations is the presence of ultraviolet radiation. It has been shown that such radiation can leave the same genomic signature as mutagenic metabolic byproducts ($CC \rightarrow TT$ mutations) [83]. But, as discussed above, our survey of rate variation across different nucleotide substitution types (Additional file 3) does not uncover a notable increase in $CC \rightarrow TT$ mutations (relative to other substitution types) in taxa with faster overall substitution rates.

Conclusions

This study reveals genome-wide variation in molecular evolutionary rate across the fern family Pteridaceae. Our data indicate that damaging metabolic byproducts and environmental influences such as ultraviolet radiation are likely not driving evolutionary rate differences in this group. Other previously proposed mechanisms thought to result in rate heterogeneity may be more plausible. Among them, the roles played by life history traits (including the rate of mitosis and duration of the haploid gametophytic life stage) and/or gene/genome duplication are especially compelling. Expanded sampling and additional studies of cytology and population genetics in ferns (including estimates of effective population sizes), as well as the assembly and annotation of nuclear and mitochondrial genomes, could further refine our understanding of the timing (e.g., crown versus stem),

causes, and consequences of molecular evolutionary rate shifts in this diverse lineage of vascular plants.

Methods

Taxon sampling and transcriptome sequencing

Thirteen taxa from the fern family Pteridaceae were sampled for our analyses. Twelve represented our focal ingroup: *Adiantum aleuticum* (Rupr.) C. A. Paris (= *Adiantum* 1), *Adiantum raddianum* C. Presl (= *Adiantum* 2), *Argyrochosma nivea* (Poir.) Windham, *Gaga arizonica* (Maxon) Fay W. Li & Windham, *Parahemionitis cordata* (Hook. & Grev.) Fraser-Jenk., *Myriopteris rufa* Fée, *Notholaena montieliae* Yatsk. & Arbeláez, *Pityrogramma trifoliata* (L.) R. M. Tryon, *Pteris ensiformis* Burm. f. (= *Pteris* 1), *Pteris vittata* L. (= *Pteris* 2), *Vittaria appalachiana* Farrar & Mickel (= *Vittaria* 1), and *Vittaria lineata* (L.) Sm. (= *Vittaria* 2). One additional taxon, *Cryptogramma acrostichoides* R. Br., was chosen as the outgroup. *Cryptogramma* has been shown based on rigorous phylogenetic studies to be sister to the remainder of Pteridaceae (including all 12 ingroup taxa), making it an appropriate outgroup for our analyses [22, 23]. These 13 taxa were selected based on the availability of transcriptome data shared by members of the 1KP consortium [84]. Collaborators of 1KP contributed leaf material from which RNA was extracted and sequenced by the consortium, using Illumina paired-end technology (Additional file 1) [5, 15, 85].

Transcriptome assembly and assessment of orthology

Orthology assessment largely followed the phylogenetic approach described in Yang and Smith [86] (for a visual summary of our particular approach to orthology assessment, as described here, see Additional file 9). First, transcriptomes were assembled by the 1KP consortium using SOAPdenovo-trans [87] as described in Johnson et al. [5] and Rothfels et al. [7]. Assembled nucleotide sequences were then translated into peptides using TransDecoder, implemented within the Trinity package [88, 89]. Next, the program ProteinOrtho was used to identify putatively orthologous regions among the translated amino acid sequences [90]. Orthologous regions (hereafter referred to as "orthogroups") containing at least one transcript per species for each of our 13 sampled taxa were identified and retained for further analyses (Additional file 10: http://dx.doi.org/10.5061/dryad.rg22j).

Amino acid sequences of the orthogroups identified by ProteinOrtho were aligned automatically using MUSCLE [91, 92]. Phylogenetic analyses of the resulting alignments were conducted in RAxML version 8 [93] using a JTT + Γ model of protein evolution (-m = PROTGAMMAJTT). Resulting gene trees were rooted using Newick Utilities (nw_reroot) [94] with *Cryptogramma* specified as the outgroup. Because of the potential for short read lengths to limit the power of likelihood

Grusz *et al. BMC Genomics* (2016) 17:692

Page 8 of 11

reconstructions, we assessed trees using a test of topology in the program TreeFix version 1.1.10 (likelihood model = treefix.models.raxmlmodel.RAxMLModel0.2.4; reconciliation model = treefix.models.duplossmodel.DupLossModel1.0.1) [95]. TreeFix optimizes gene tree topologies by maximizing a known species tree topology, and assumes that any discordance between a gene tree and the presumed species tree results from gene duplication and/or gene loss. Using the output trees from TreeFix (with ≥13 tips), any subtrees that did not disagree with our species tree hypothesis were pruned (prune_paralogs_MO.py) [86] and retained. Gene trees were then compared to the established species tree hypothesis [22] to determine whether a single sequence for each of our thirteen taxa was still present using the program OptRoot (-d = rf) [96]. Trees with 13 taxa and a topology consistent with the expected species tree (rf = 0) were retained for subsequent analyses.

Protein alignments for all retained orthogroups were then reverse-translated to their corresponding in-frame nucleotide sequence alignments using the back_translate.pl script. To remove plastid sequences, orthogroups were compared to the two published plastid genomes from the Pteridaceae, *Myriopteris lindheimeri* (Hook.) J. Sm. [4] and *Adiantum capillus-veneris* L. [97]. BLAST databases were created for each of the plastid genomes and a fasta-formatted file containing a single nucleotide sequence (from *Adiantum* 2) for each orthologous region was used to query against each of the plastid genome databases. Any orthogroups with a significant BLAST hit to either (or both) of the plastid genome sequences were discarded prior to analysis (e-value threshold = 1e^{-5}).

Testing for rate heterogeneity and signatures of selection

We executed a series of pairwise relative rate comparisons among our 12 focal taxa (*Cryptogramma* was used as the outgroup for all pairwise comparisons) using a likelihood ratio test approach implemented in the program HyPhy (PairwiseRelativeRate.bf) [98]. We first tested for variation in overall substitution rate, using a GTR model of sequence evolution, by comparing the likelihood scores obtained using a single, shared substitution rate for both taxa to the likelihood score obtained when the substitution rate for each taxon was estimated separately. In these analyses, we estimated branch lengths independently, with all other model parameters shared (i.e., using the "global" option); rate parameters were inferred by maximum likelihood and observed nucleotide frequencies were used as equilibrium frequencies ("observed" option). We also performed separate pairwise relative rate tests constraining (or not) each transition ("t_ct", "t_ag") and transversion ("t_ac", "t_at", t_gt", "t_cg") type. For these analyses, substitution rates were again estimated using a GTR model of sequence evolution; all parameters were estimated independently

for each branch (i.e., using the "local" option) and equilibrium frequencies were based on the observed nucleotide frequencies (i.e., using the "observed" option).

To look for signatures of selection, we performed two additional sets of pairwise relative rate comparisons (PairwiseRelativeRate.bf) [98]. One set of analyses compared the likelihood value estimated when the rate of nonsynonymous substitution (dN; "nonSynRate") was constrained to be equal between the two focal taxa to the likelihood estimate obtained when dN was allowed to vary. A second set of analyses was performed in which synonymous substitution rate (dS; "synRate") was examined in the same manner. Both sets of analyses (dN and dS) were implemented using the MG4 model of codon evolution and the universal genetic code ("universal" option); all parameters were estimated independently for each branch ("local" option).

An additional set of analyses comparing alternative models of dN/dS (= ω) across the tree for each locus (SelectionLRT.bf) [98] were run to determine the degree to which selection may be driving rate differences in vittarioid ferns. For these analyses, the tree was partitioned into three components (each with an associated dN/dS): the clade of interest (i.e., vittarioid ferns; ω_V), the branch subtending the clade of interest (ω_B), and the remainder of the tree (ω_T). Likelihood model comparisons (AIC) were used to determine which of the following null hypotheses best explained the data available for each locus: (1) $\omega_T = \omega_V = \omega_B$; (2) $\omega_T = \omega_V \neq \omega_B$; (3) $\omega_T = \omega_B \neq \omega_V$; (4) $\omega_T \neq \omega_V = \omega_B$; (5) $\omega_T \neq \omega_V \neq \omega_B$.

GO enrichment analyses

Blast2GO [99–102], implemented within the CLC Main Workbench version 7.6.2 (Qiagen, Aarhus, Denmark), was used to BLAST, map, and annotate gene ontology (GO) categories corresponding to each of our 2091 analyzed orthogroups [103–106]. Following annotation, we performed GO enrichment analyses using goatools (find_enrichment.py, –fdr option, pval = 0.05) to determine whether any GO categories were significantly over- or underrepresented (i.e., enriched or purified) among the loci for which substitution rates in vittarioid ferns, in particular, were found to be significantly faster or significantly slower than at least one other focal taxon (for test sets analyzed, see Additional file 8). Significance was assessed using a Bonferroni Correction to account for multiple comparisons. Because a closely related, annotated genome was not available, many of our 2091 loci did not map to a known GO category. For this reason, we performed two separate GO enrichment analyses for each vittarioid rate category (i.e., significantly faster/slower loci), for a total of four separate comparisons. In the first comparison, the test set (TS1) included all loci for which vittarioid ferns were significantly faster than at

Grusz *et al. BMC Genomics* (2016) 17:692

Page 9 of 11

least one other sampled taxon, and the reference set (RS1) included the entire population of loci, regardless of their annotation status. In the second comparison, the test set (TS2) included only annotated loci for which vittarioid ferns were significantly faster than at least one other sampled taxon, and the reference set (RS2) included only the population of loci with known GO annotations. In the third comparison, the test set (TS3) included all loci for which vittarioid ferns were significantly slower than at least one other sampled taxon, and the reference set (RS1) included the entire population of loci, regardless of their annotation status. In the final comparison, the test set (TS4) included only annotated loci for which vittarioid ferns were significantly slower than at least one other sampled taxon, and the reference set (RS2) included only the population of loci with known GO annotations (for all test and reference sets, see Additional file 8).

Additional files

Additional file 1: Taxon sampling. Taxon sampling and corresponding voucher information for samples included in our analyses, as provided by contributors to the 1KP Project [84]. (PDF 76 kb)

Additional file 2: Pairwise overall rate comparisons. Comparisons of pairwise relative substitution rate for 2091 orthogroups across sampled taxa from the Pteridaceae. (PDF 89 kb)

Additional file 3: Pairwise rate comparisons per substitution type. Summary of pairwise relative rate comparisons for individual substitution types across sampled taxa from the Pteridaceae for 2091 loci. (PDF 580 kb)

Additional file 4: Pairwise nonsynonymous rate comparisons. Comparisons of pairwise relative nonsynonymous rate (dN) for 2091 orthogroups across the Pteridaceae. (PDF 90 kb)

Additional file 5: Pairwise synonymous rate comparisons. Comparisons of pairwise relative synonymous rate (dS) for 2091 orthogroups across sampled taxa from the Pteridaceae. (PDF 90 kb)

Additional file 6: Orthogroup summary statistics. Summary statistics for the 2091 orthogroups analyzed in this study. (PDF 288 kb)

Additional file 7: Summary of annotated GO categories. GO annotation distribution for loci corresponding to molecular function, biological processes, and cellular compartments. (PDF 403 kb)

Additional file 8: GO analyses data files. Zipped folder containing 'test set' and 'reference set' loci for all GO analyses performed herein, along with a corresponding readme document describing individual analyses. (ZIP 3916 kb)

Additional file 9: Flowchart for determining orthologous regions. Flow chart illustrating steps taken to identify single copy orthologous regions across our 13 sampled taxa. (PDF 344 kb)

Additional file 10: Orthogroup data file. Zipped folder containing fasta-formatted reads identified by ProteinOrtho, used for all downstream orthogroup determination and analysis, along with readme document and relevant project-specific scripts (also available online via Dryad: http://dx.doi.org/10.5061/dryad.rg22j). (ZIP 12021 kb)

Authors' contributions
ALG and ES together conceived and designed the study. CJR provided samples from which data were analysed. ALG performed data analyses and drafted the manuscript. All authors contributed to and approved content comprising the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Author details
[1]Department of Botany, Smithsonian Institution, MRC 166 PO Box 37012, Washington, DC 20013-7012, USA. [2]Department of Integrative Biology, University of California Berkeley, 1001 Valley Life Sciences Building, Berkeley, CA 94720-2466, USA. [3]Department of Biology, University of Minnesota Duluth, 1035 Kirby Drive, Duluth, MN 55812, USA.

References
1. Barker MS, Wolf PG. Unfurling fern biology in the genomics age. BioScience. 2010;60:177–85.
2. McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT. Applications of next-generation sequencing to phylogeography and phylogenetics. Mol Phylogenet Evol. 2013;66:526–38.
3. Wolf PG, Sessa EB, Merchant DB, Li FW, Rothfels CJ, Sigel EM, et al. An exploration into fern genome space. Genome Biol Evol. 2015;7:2533–44.
4. Wolf PG, Der JP, Duffy AM, Davidson JB, Grusz AL, Pryer KM. The evolution of chloroplast genes and genomes in ferns. Plant Mol Biol. 2011;76:251–61.
5. Johnson MTJ, Carpenter EJ, Tian Z, Bruskiewich R, Burris JN, Carrigan CT, et al. Evaluating methods for isolating total RNA and predicting success of sequencing phylogenetically diverse plant transcriptomes. Plos One. 2012;7:e50226.
6. Gao L, Wang B, Wang ZW, Zhou Y, Su YJ, Wang T. Plastome sequences of Lygodium japonicum and Marsilea crenata reveal the genome organization transformation from basal ferns to core leptosporangiates. Genome Biol Evol. 2013;5:1403–7.
7. Rothfels CJ, Larsson A, Li FW, Sigel EM, Huiet L, Burge DO, Ruhsam M, Graham SW, Stevenson D, Wong GKS, Korall P, Pryer KM. Transcriptome-mining for single-copy nuclear markers in ferns. Plos One. 2013;8:e76957.
8. Li FW, Pryer KM. Crowdfunding the Azolla fern genome project: a grassroots approach. GigaScience. 2014;3:16.
9. Sessa EB, Banks JA, Barker MS, Der JP, Duffy AM, Graham SW, et al. Between two fern genomes. GigaScience. 2014;3:15.
10. Bégu D, Castandet B, Araya A. RNA editing restores critical domains of a group I intron in fern mitochondria. Curr Genet. 2011;57:317–25.
11. Der JP, Barker MS, Wickett NJ, dePamphlis CW, Wolf PG. De novo characterization of the gametophyte transcriptome in bracken fern, Pteridium aquilinum. BMC Genomics. 2011;12:99.
12. Boothby TC, Zipper RS, van der Weele CM, Wolniak SM. Removal of retained introns regulates translation in the rapidly developing gametophyte of Marsilea vestita. Dev Cell. 2013;24:517–29.
13. Bushart TJ, Cannon AE, ul Hague A, San Miguel P, Mostajeran K, Clark GB, et al. RNA-seq analysis identifies potential modulators of gravity response in spores of Ceratopteris (Pteridaceae): Evidence for modulation by calcium pumps and apyrase activity. Am J Bot. 2013;100:161–74.

Grusz *et al. BMC Genomics* (2016) 17:692

Page 10 of 11

14. Li FW, Villareal JC, Kelly S, Rothfels CJ, Melkonian M, Frangedakis E, et al. Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. Proc Natl Acad Sci U S A. 2014;111:6672–7.

15. Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, et al. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc Natl Acad Sci U S A. 2014;111:E4859–68.

16. Yin Y, Johns MA, Cao H, Rupani M. A survey of plant and algal genomes and transcriptomes reveals new insights into the evolution and function of cellulose synthase superfamily. BMC Genomics. 2014;15:260.

17. Aya K, Kobayashi M, Tanaka J, Ohyanagi H, Suzuki T, Yano K, et al. De novo transcriptome assembly of a fern, Lygodium japonicum, and a web resource database, Ljtrans DB. Plant Cell Physiol. 2015;56:e5(1-14).

18. Guo W, Grewe F, Mower JP. Variable frequency of plastid RNA editing among ferns and repeated loss of uridine-to-cytidine editing from vascular plants. Plos One. 2015. doi:10.1371/journal.pone.0117075.

19. Vanneste KL, Sterck A, Myburg A, Van de Peer Y, Mizrachi E. Horsetails are ancient polyploids: Evidence from Equisetum giganteum. Plant Cell. 2015;27:1567–78.

20. Soltis PS, Soltis DE, Savolainen V, Crane PR, Barraclough TG. Rate heterogeneity among lineages of tracheophytes: Integration of molecular and fossil data and evidence for molecular living fossils. Proc Natl Acad Sci U S A. 2002;99:4430–5.

21. Schuettpelz E, Pryer KM. Reconciling extreme branch length differences: Decoupling time and rate through the evolutionary history of filmy ferns. Syst Biol. 2006;55:485–502.

22. Schuettpelz E, Pryer KM. Fern phylogeny inferred from 400 leptosporangiate species and three plastid genes. Taxon. 2007;56:1037–50.

23. Schuettpelz E, Schneider H, Huiet L, Windham MD, Pryer KM. A molecular phylogeny of the fern family Pteridaceae: Assessing overall relationships and the affinities of previously unsampled genera. Mol Phylogenet Evol. 2007;44:1172–85.

24. Korall PE, Schuettpelz E, Pryer KM. Abrupt deceleration of molecular evolution linked to the origin of arborescence in ferns. Syst Biol. 2010;64:2786–92.

25. Rothfels CJ, Schuettpelz E. Accelerated rate of molecular evolution for vittarioid ferns is strong and not driven by selection. Syst Biol. 2014;63:31–54.

26. Yang Y, Moore MJ, Brockington SF, Soltis DE, Wong GKS, Carpenter EJ, et al. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. Mol Biol Evol. 2015;32:2001–14.

27. Britten RJ. Rates of DNA evolution differ between taxonomic groups. Science. 1986;231:1393–8.

28. Schubart CD, Diesel R, Hodges SB. Rapid evolution to terrestrial life in Jamaican crabs. Nature. 1998;393:363–5.

29. Johnson KP, Seger J. Elevated rates of nonsynonymous substitution in island birds. Mol Biol Evol. 2001;18:874–81.

30. Lumbsch HT, Hipp AL, Divakar PK, Blanco O, Crespo A. Accelerated evolutionary rates in tropical and oceanic parmelioid lichens (Ascomycota). BMC Evol Biol. 2008;8:257.

31. Otálora MAG, Aragón G, Martínez I, Wedin M. Cardinal characters on a slippery slope—a re-evaluation of phylogeny, character evolution, and evolutionary rates in the jelly lichens (Collemataceae s. str.). Mol Phylogenet Evol. 2013;68:185–98.

32. Jobson RW, Albert VA. Molecular rates parallel diversification contrasts between carnivorous plant sister lineages. Cladistics. 2002;18:127–36.

33. Müller K, Albach DC. Evolutionary rates in Veronica L. (Plantaginaceae): Disentangling the influence of life history and breeding system. J Mol Evol. 2010;70:44–56.

34. Wicke S, Schäferhoff B, dePamphlis CW, Müller KF. Disproportionate plastome-wide increase of substitution rates and relaxed purifying selection in genes of carnivorous Lentibulariaceae. Mol Biol Evol. 2013;31:529–45.

35. Kimura M. DNA and the neutral theory. Philos Trans R Soc B. 1986;312:43–354.

36. Müller K, Borsch T. Phylogenetics of Utricularia (Lentibulariaceae) and molecular evolution of the trnK intron in a lineage with high substitutional rates. Plant Syst Evol. 2005;250:39–67.

37. Shen B, Fang T, Yang T, Jones G, Irwin DM, Zhang S. Relaxed evolution in the tyrosine aminotransferase gene Tat in Old World fruit bats (Chiroptera: Pteropodidae). Plos One. 2014. doi:10.1371/journal.pone.009748.

38. De La Torre AR, Lin YC, Van de Peer Y, Ingvarsson PK. Genome-wide analysis reveals diverged patterns of codon bias, gene expression and rates of sequence evolution in Picea gene families. Genome Biol Evol. 2015. doi:10.1093/gbe/evv044.

39. Laird CD, McConaughy BL, McCarthy BJ. Rate of fixation of nucleotide substitutions in evolution. Nature. 1969;224:149–54.

40. Martin AP, Palumbi SR. Body size, metabolic rate, generation time, and the molecular clock. Proc Natl Acad Sci U S A. 1993;90:4087–91.

41. Gaut BS, Morton BR, McCaig BC, Clegg MT. Substitution rate comparisons between grasses and palms: Synonymous rate differences at the nuclear gene Adh parallel rate differences at the plastid gene rbcL. Proc Natl Acad Sci U S A. 1996;93:10274–9.

42. Allen AP, Brown JH, Gillooly JF. Global biodiversity, biochemical kinetics, and the energy-equivalence rule. Science. 2002;297:1545–8.

43. Wright SD, Gray RD, Gardner RC. Energy and the rate of evolution: inferences from plant rDNA substitution rates in the western Pacific. Evolution. 2003;57:2893–8.

44. Davies TJ, Savolainen V, Chase MW, Moat J, Barraclough TG. Environmental energy and evolutionary rates in flowering plants. Philos Trans R Soc B. 2004;271:2195–200.

45. Evans KL, Gaston KJ. Can the evolutionary-rates hypothesis explain species-energy relationships? Funct Ecol. 2005;19:899–915.

46. Evans KL, Warren PH, Gaston KJ. Species-energy relationships at the macroecological scale: a review of mechanisms. Biol Rev. 2005;80:1–25.

47. Wright S, Keeling J, Gillman L. The road from Santa Rosalia: A faster tempo of evolution in tropical climates. Proc Natl Acad Sci U S A. 2009;20:7718–22.

48. Bromham L. Why do species vary in their rate of molecular evolution? Biol Lett. 2009;5:401–4.

49. Ibarra-Laclette E, Albert VA, Pérez-Torres CA, Zamudio-Hernández F, Ortega-Estrada M, Herrera-Estrella A, Herrera-Estrella L. Transcriptomics and molecular evolutionary rate of the bladderwort (Utricularia), a carnivorous plant with a minimal genome. BMC Plant Biol. 2011;11:101.

50. Lourenço JM, Glémin S, Chiari Y, Galtier N. The determinants of molecular substitution process in turtles. J Evol Biol. 2013;26:38–50.

51. Chao L, Carr DE. The molecular clock and the relationship between population size and generation time. Evolution. 1993;47:688–90.

52. Welch JJ, Bininda-Emonds ORP, Bromham L. Correlates of rate variation in mammalian protein-coding sequences. BMC Evol Biol. 2008;8:53.

53. Mooers AØ, Harvey PH. Metabolic rate, generation time, and the rate of molecular evolution in birds. Mol Phylogenet Evol. 1994;3:344–50.

54. Lutzoni F, Pagel M. Accelerated evolution as a consequence of transitions to mutualism. Proc Natl Acad Sci U S A. 1997;94:11422–7.

55. Whittle CA. The influence of environmental factors, the pollen : ovule ratio and seed bank persistence on molecular evolutionary rates in plants. J Evol Biol. 2006;19:302–8.

56. Lanfear R, Ho SYW, Davies TJ, Moles AT, Aarssen L, Swenson NG, Warman L, Zanne AE, Allen AP. Taller plants have lower rates of molecular evolution. Nat Commun. 2013;4:1879.

57. Parkinson CL, Mower JP, Qiu YL, Shirk AJ, Song K, Young ND, et al. Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. BMC Evol Biol. 2005;5:73.

58. Rothfels CJ, Li FW, Sigel EM, Huiet L, Larsson A, Burge DO, et al. The evolutionary history of ferns inferred from 25 low-copy nuclear genes. Am J Bot. 2015;102:1089–107.

59. Kimura M. Evolutionary rate at the molecular level. Nature. 1968;217:624–6.

60. Cooke MS, Evans MD, Dizdaroglu M, Lunec J. Oxidative DNA damage: mechanisms, mutation, and disease. FASEB J. 2003;17:1195–214.

61. Newcomb TG, Allen KJ, Tkeshelashvili L, Loeb LA. Detection of tandem CC→TT mutations induced by oxygen radicals using mutation-specific PCR. Mutat Res. 1999;417:21–30.

62. Gillooly JF, Brown JH, West GB, Savage VM, Charnov EL. Effects of size and temperature on metabolic rate. Science. 2001;293:2248–51.

63. Gillooly JF, Allen AP, West GB, Brown JH. The rate of DNA evolution: Effects of body size and temperature on the molecular clock. Proc Natl Acad Sci U S A. 2005;102:140–5.

64. Fontanillas E, Welch JJ, Thomas JA, Bromham L. The influence of body size and net diversification rate on molecular evolution during the radiation of animal phyla. BMC Evol Biol. 2007;l7:95.

65. Martin AP. Substitution rates of organelle and nuclear genes in sharks: Implicating metabolic rate (again). Mol Biol Evol. 1999;16:996–1002.

66. Hoffmann S, Spitkovsky D, Radicella JP, Epe B, Wiesner RJ. Reactive oxygen species derived from the mitochondrial respiratory chain are not responsible for the basal levels of oxidative base modifications observed in nuclear DNA of mammalian cells. Free Radical Bio Med. 2004;36:765–73.

67. Lanfear R, Thomas JA, Welch JJ, Brey T, Bromham L. Metabolic rate does not calibrate the molecular clock. Proc Natl Acad Sci U S A. 2007;104:15388–93.

Grusz *et al. BMC Genomics* (2016) 17:692

Page 11 of 11

68. Park C, Qian W, Zhang J. Genomic evidence for elevated mutation rates in highly expressed genes. EMBO Rep. 2012;13:1123–9.

69. Chen J, Furano AV. Breaking bad: The mutagenic effect of DNA repair. DNA Repair. 2015;32:43–51.

70. Klekowski EJ. Mutational load in clonal plants: a study of two fern species. Evolution. 1984;38:417–26.

71. Schneider H, Smith AR, Cranfill R, Hildebrand TJ, Haufler CH, Ranker TA. Unraveling the phylogeny of polygrammoid ferns (Polypodiaceae and Grammitidaceae): exploring aspects of the diversification of epiphytic plants. Mol Phylogenet Evol. 2003;31:1041–63.

72. Watkins JEJ, Cardelús CL. Ferns in an angiosperm world: Cretaceous radiation into the epiphytic niche and diversification on the forest floor. Int J Plant Sci. 2012;173:695–710.

73. Cherry JL. Expression level, evolutionary rate, and the cost of expression. Genome Biol Evol. 2010;2:757–69.

74. Chuang TJ, Chiang TW. Impacts of pretranscriptional DNA methylation, transcriptional transcription factor, and posttranscriptional microRNA regulations on protein evolutionary rate. Genome Biol Evol. 2014;6:1530–41.

75. Liu HJ, Tang ZX, Han XM, Yang ZL, Zhang FM, Yang HL, Lui YJ, Zeng QY. Divergence in enzymatic activities in the soybean GST supergene family provides new insight into the evolutionary dynamics of whole genome duplications. Mol Biol Evol. 2015;32:2844–59.

76. Li FW, Kuo LY, Pryer KM, Rothfels CJ. Genes translocated into the plastid inverted repeat show decelerated substitution rates and elevated GC content. Genome Biol Evol. 2016. In press.

77. Rice A, Glick L, Abadi S, Einhorn M, Kopelman NM, Salman-Minkov A, Mayzel J, Chay O, Mayrose I. The Chromosome Counts Database (CCDB) - a community resource of plant chromosome numbers. New Phytol. 2015;206:19–26.

78. Bromham L, Cardillo M. Testing the latitudinal gradient in species richness and rates of molecular evolution. J Evol Biol. 2003;16:200–7.

79. Hebert PDN, Remigio EA, Colbourne JK, Taylor DJ, Wilson CC. Accelerated molecular evolution in halophilic crustaceans. Evolution. 2002;56:090–926.

80. Colbourne JK, Wilson CC, Hebert PDN. The systematics of Australian Daphnia and Daphniopsis (Crustacea: Cladocera): a shared phylogenetic history transformed by habitat-specific rates of evolution. Biol J Linn Soc. 2006;89:469–88.

81. Ossowski S, Schneeberger K, Lucas-Lledó KI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M. The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. Science. 2009;327:92–4.

82. De Vries FWTP. The cost of maintenance processes in plant cells. Ann Bot-London. 1975;39:77–92.

83. Nakazawa H, English E, Randell PL, Nakazawa K, Martel N, Armstrong BK, Yamasaki H. UV and skin cancer: Specific p53 gene mutation in normal skin as a biologically relevant exposure measurement. Proc Natl Acad Sci U S A. 1994;91:360–4.

84. oneKP. 2016. http://www.onekp.com. Accessed 1 Sept 2014.

85. Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc Natl Acad Sci U S A. 2010;107:4623–8.

86. Yang Y, Smith SA. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy for phylogenomics. Mol Biol Evol. 2014;31:3081–92.

87. Short Oligonucleotide Analysis Package. 2013. http://soap.genomics.org.cn/SOAPdenovo-Trans.html. Accessed 1 Sept 2014.

88. Garbherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol. 2013;29:644–52.

89. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nat Protoc. 2013;8:1494–512.

90. Lerchner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. ProteinOrtho: Detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics. 2011;12:124.

91. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.

92. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:113.

93. Stamatakis A. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.

94. Junier T, Zdobnov EM. The Newick Utilities: High-throughput phylogenetic tree processing in the UNIX shell. Bioinformatics. 2010;26:1669–70.

95. Wu YC, Rasmussen MD, Bansal MS, Kellis M. TreeFix: Statistically informed gene tree error correction using species trees. Syst Biol. 2013;62:110–20.

96. Wehe A. OptRoot. 2012. http://www.wehe.us/optroot.html. Accessed 1 Sept 2014.

97. Wolf PG, Rowe CA, Sinclair RB, Hasebe M. Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern, Adiantum capillus-veneris L. DNA Res. 2003;10:59–65.

98. Pond SLK, Frost SDQ, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 2005;21:676–9.

99. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21:3674–6.

100. Conesa A, Götz S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics. 2008;2008:1–12.

101. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 2008;36:3420–35.

102. Götz S, Arnold R, Sebastián-León P, Martín-Rodríguez S, Tischler P, Jehl MA, et al. B2G-FAR, a species-centered GO annotation repository. Bioinformatics. 2011;27:919–24.

103. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9.

104. Blake JA. Ten quick tips for using the gene ontology. PLoS Comput Biol. 2013;9:E1003343. doi:10.1371/journal.pcbi.1003343.

105. The Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res. 2015;43(Database issue):D1049–56.

106. Haibao T, Klopfenstein D, Pedersen B, Flick P, Sato K, Ramirez F, et al. GOATOOLS: Tools for Gene Ontology. Zenodo. 2015.