Title

Increased evolutionary rates and conserved transcriptional response following allopolyploidization in brown algae.

Journal

Evolution; international journal of organic evolution, 73(1)

ISSN

0014-3820

Authors

Sousa, Filipe
Neiva, João
Martins, Neusa
et al.

Publication Date

2019

DOI

10.1111/evo.13645

doi:10.1111/evo.13645

Increased evolutionary rates and conserved transcriptional response following allopolyploidization in brown algae

Filipe Sousa,[1] João Neiva,[1] Neusa Martins,[1] Rita Jacinto,[1] Laura Anderson,[2] Peter T. Raimondi,[2] Ester A. Serrão,[1] and Gareth A. Pearson[1,3]

[1]*CCMAR—Centro de Ciências do Mar da Universidade do Algarve, Edifício 7, Gambelas, Faro 8005–139, Portugal*

[2]*Long Marine Laboratory, University of California, Santa Cruz, California 95064*

[3]*E-mail: gpearson@ualg.pt*

Received January 29, 2018

Accepted October 24, 2018

Genome mergers between independently evolving lineages, via allopolyploidy, can potentially lead to instantaneous sympatric speciation. However, little is known about the consequences of allopolyploidy and the resultant "genome shock" on genome evolution and expression beyond the plant and fungal branches of the Tree of Life. The aim of this study was to compare substitution rates and gene expression patterns in two allopolyploid brown algae (Phaeophyceae and Heterokonta) and their progenitors in the genus *Pelvetiopsis* N. L. Gardner in the north-east Pacific, and to date their relationships. We used RNA-seq data, all potential orthologues, and putative single-copy loci for phylogenomic, divergence, and gene expression analyses. The multispecies coalescent placed the origin of allopolyploids in the late Pleistocene (0.35–0.05 Ma). Homoeologues displayed increased nonsynonymous divergence compared with parental orthologues, consistent with relaxed selective constraint following allopolyploidization, including for genes with no evidence of pseudogenization or neofunctionalization. Patterns of homoeologue-orthologue expression conservation and expression level dominance were largely shared with both natural plant and fungal allopolyploids. Our results provide further support for common cross-Kingdom patterns of allopolyploid genome evolution and transcriptional responses, here in the evolutionarily distinct marine heterokont brown algae.

KEY WORDS: **Allopolyploidy, gene expression, hybridization, marine diversity, substitution rates, transcriptional response.**

Polyploidy or whole genome duplication induces significant changes in genome organization, expression patterns, and phenotypes of organisms (Hegarty and Hiscock 2008; Buggs et al. 2014). It is the most common mechanism of sympatric speciation (Otto and Whitton 2000) and several polyploidization events are known to be at the origin of major evolutionary lineages, in both plants and animals (Masterson 1994; Spring 1997; Adams and Wendel 2005; Soltis et al. 2009; Jiao et al. 2011). There is also evidence of increased evolutionary diversification in polyploids (Otto and Whitton 2000; Soltis et al. 2009) and of higher genetic diversity within polyploids than within their diploid relatives (Brochmann et al. 1992, Soltis et al. 1993) although opposing patterns have also been observed (Mayrose et al. 2011).

Two very different types of polyploidy can be distinguished: autopolyploidy, which results from the duplication of the same genome, and allopolyploidy, in which the genomes of two different species merge in a single organism as a consequence of hybridization. Allopolyploidization causes genomic changes known to generate adaptive leaps and even abrupt speciation events (Mallet 2007; Soltis et al. 2009). Gene duplication via allopolyploidization may also have evolutionary consequences, as duplicate genes are generally thought to evolve neutrally after an early period of relaxed selective constraint (Lynch and Conery 2000). The retention of gene duplicates can be a consequence of sub- and neofunctionalization, caused by the accumulation of fixed mutations (Force et al. 1999), by selection for high

expression (Seoighe and Wolfe 1999), or by differential expression (DE) in response to environmental stress (Dong and Adams 2011). It is estimated that 20–50% of duplicate genes are preserved for tens to hundreds of millions of years (Lynch and Force 2000). Genetic redundancy can impact the evolution of coding sequences by relaxed selection and increased substitution rates, which may favor the accumulation of replacement mutations that influence gene function (Akhunov et al. 2013). Studies in allopolyploid cotton (Senchina et al. 2003; Flagel et al. 2012), *Arabidopsis* (Barrier et al. 2001), and wheat (Akhunov et al. 2013) have all shown evidence of increased substitution rates in allopolyploids following genome duplication, compared to their diploid progenitors. Fixed heterozygosity and the increased number of alleles available for selection in allopolyploids are also thought to be an evolutionary advantage (Hegarty and Hiscock 2008), as the resulting genetic and phenotypic changes can grant species the ability to expand to new ecological niches and even outcompete progenitor species (Leitch and Leitch 2008; Madlung 2013). In fact, certain conditions, such as sexual reproduction, small population sizes, and codominance of beneficial alleles, may promote the adaptation of polyploid populations to changing environments (Otto and Whitton 2000). Although novel functions are known to arise via retention of gene duplicates (Rensing 2014), the evolution of new functions is rare, and the majority of duplicate genes eventually become silent (Lynch and Conery 2000). However, reciprocal gene loss or silencing in independent polyploid populations may lead to reproductive isolation and, thus, promote speciation (Adams and Wendel 2005).

In allopolyploids, genome doubling has also been shown to have a stabilizing effect on gene expression changes (''transcriptome shock'') following hybridization (Hegarty et al. 2006; Zhang et al. 2016), and transcriptional responses in recently formed allopolyploids are thought to follow a conserved cross-kingdom pattern (Cox et al. 2014). The retention of DE present in parental diploid species accounts in large part for the differences in expression observed between homoeologues, although a small proportion of differences may be acquired following allopolyploidization (Buggs et al. 2014). Phenotypic effects caused by novel or trangressive gene expression patterns (Yoo et al. 2013) and by tissue-specific gene expression (Buggs et al. 2011), following allopolyploidization are thought to provide the basis for natural selection in newly formed species.

In marine systems, allopolyploidy and its role in diversification are still poorly studied. Genome size differences have been attributed to palaeopolyploidization events in brown algae (Phillips et al. 2011), but very few examples of allopolyploids are actually known (e.g., Niwa and Sakamoto 2010). However, two putative allopolyploid lineages were recently identified in the intertidal brown algal genus *Pelvetiopsis* N. L. Gardner (Fucaceae and Heterokontophyta), one of which appears to have undergone range and habitat expansion (Neiva et al. 2017). The polyploid nature of the two taxa was initially inferred by the fixed heterozygosity observed at nuclear microsatellite loci and the occasional presence of three alleles (Neiva et al. 2017). Both allopolyploids (*Pelvetiopsis hybrida* and *Pelvetiopsis limitata*) share a paternal ancestor corresponding to *Pelvetiopsis arborescens*. The maternal ancestor of *P. hybrida* corresponds to *Pelvetiopsis californica* (*Hesperophycus californicus* P. C. Silva), while the maternal ancestor of *P. limitata* is either unsampled or (likely) extinct. The genus is distributed along temperate rocky shores in the north-east Pacific, across an area spanning several marine ecoregions that is known for its species richness and genetic diversity (Johansson et al. 2015). The phylogeny of brown algae has been mostly concerned with ordinal to generic relationships (Silberfeld et al. 2010; Cánovas et al. 2011), while few studies have dealt with infrageneric phylogenies and reticulation patterns (Neiva et al. 2017). The discovery of new, well-established lineages of allopolyploid origin in this genus raised several questions at the genomic level: are the changes to substitution rates observed in allopolyploid model plants common to brown algae? How does allopolyploidization affect gene expression? To address these questions, we used a phylogenomic dataset derived from RNA-Seq to: (1) infer the evolutionary history of *Pelvetiopsis*, including the timing of allopolyploid speciation events; (2) estimate rates of synonymous and nonsynonymous substitutions within allopolyploids and between parental genomes; and (3) compare gene expression patterns in allopolyploid versus diploid parental genomes. The very limited genetic differentiation between allopolyploid *Pelvetiopsis* species and their extant ancestors (as assessed with both mtDNA and microsatellites, Neiva et al. 2017) makes this genus particularly well suited to investigate how parental genomes are transferred and functionally integrated in new species and how it might contributes to adaptation to new niches. We provide evidence for significant changes in evolutionary rates, particularly nonsynonymous substitutions, in the most recently formed allopolyploid *Pelvetiopsis* species, and show that the transcriptional response in this allopolyploid is mostly conserved.

Methods

SAMPLING AND PREPARATION OF RNA-SEQ DATA FOR PHYLOGENOMICS

Individuals of *P. californica* (P. C. Silva) Neiva, P. T. Raimondi, G. A. Pearson and Serrão (*H. californicus* P. C. Silva), *P. hybrida* Neiva, P. T. Raimondi, G. A. Pearson and Serrão, and *P. arborescens* N. L. Gardner were collected from the field on 29 May 2014 and 31 May 2014 during receding daytime low tides between Point Conception and Monterey Bay, California (between latitudes 35.289620° and 36.637970°). Individuals of

P. limitata (Setchell) N. L. Gardner were collected from the same area on 8 February 2013 and 1 June 2014. Tissues were immediately flash frozen in liquid nitrogen for transport to the laboratory, where they were stored at −80 C before being lyophilized and transported on silica crystals to Portugal. Total RNA was extracted as described in Pearson et al. 2006 from reproductive tissue (receptacles) obtained from either a single individual or from pooled individuals (Table 1). In addition to the reference (see below), the outgroup species for phylogenomic analyses were single individuals of *Ascophyllum nodosum* (L.) Le Jolis, *Fucus gardneri* P. C. Silva, *Fucus ceranoides* L., *Pelvetia canaliculata* (L.) Decaisne and Thuret, and *Silvetia compressa* (J. Agardh) Serrão, T. O. Cho, S. M. Boo, and Brawley. For all species and samples, total RNA was treated with DNAse I to remove contaminant DNA, and RNA-Seq library preparation and 100 bp paired-end (PE) sequencing on the Illumina HiSeq 2000 platform were performed by a service provider (BGI, Shenzhen, Guangdong, China).

SAMPLING AND PREPARATION OF RNA-SEQ DATA FOR GENE EXPRESSION ANALYSIS

Gene expression analyses of *P. hybrida*, *P. arborescens*, and *P. californica* were performed on additional samples collected on 17 September 2016. For each species, three replicate samples consisting of three pooled individuals were collected (Table 1). Sample collection, RNA extraction, and library preparation were performed as described above. Sequencing was performed by the same provider (BGI) using the Illumina HiSeq 4000 platform.

REFERENCE TRANSCRIPTOMES

Read mapping was performed against a reference transcriptome from *F. ceranoides* (EMBL-EBI European Nucleotide Archive Accession PRJEB11969), a member of the genus sister to *Pelvetiopsis* (Cánovas et al. 2011). The reference was prepared from Illumina 100 bp PE reads of male and female reproductive tissues. A total of 35.1 M high-quality read pairs (FastQC, Trimmomatic) were digitally normalized to reduce coverage variation and low abundance k-mers (khmer package, Brown et al. 2012), before de novo assembly with Velvet-Oases (version 0.2.09, Schultz et al. 2012). Individual k-mer assemblies (k = 21–63, steps of 2) were merged using the custom script (oases_pipeline.py) provided. Representative transcripts for each locus were selected from the merged assembly using a second script, following the recommendations from the authors (OasesVersion 0.2.04OutputToCsvDataBase.py; https://sites.google.com/a/brown.edu/bioinformatics-in-biomed/velvet-and-oases-transcriptome), and further filtered for redundancy using CD-HIT-EST with cut-off identity of 95% (Li and Godzik 2006; Fu et al. 2012). Open reading frames were predicted using FragGeneScan (Rho et al., 2010), which were subsequently analyzed using BLASTx against a custom Stramenopile database (E-value cut off $\leq 10^{-10}$). Top hits against the brown alga *Ectocarpus* strain Ec32 were retained, resulting in a final reference transcriptome containing 17,200 transcripts.

A second reference transcriptome from *Fucus vesiculosus* was prepared in the same way (328.6 M 100 bp PE reads; unpublished data), resulting in 10,308 transcripts with top BLASTx hits against *Ectocarpus* strain Ec32.

IDENTIFICATION OF PUTATIVE SINGLE-COPY LOCI FOR PHYLOGENOMICS

Reciprocal and all against self- BLASTn searches using the reference transcriptomes for *F. ceranoides* and *F. vesiculosus* were performed. Transcripts that gave unique hits (E-value $\leq 10^{-4}$) in both self- and reciprocal BLAST analyses were retained as a conservative set of 2027 putatively single-copy (SC) orthologous loci.

READ MAPPING, PHASING, AND ALIGNMENT

The PE reads of each sample were mapped against the *F. ceranoides* reference transcriptome using Bowtie2 (Langmead and Salzberg 2012). Mapped .bam files were phased (Samtools phase), and consensus sequences (.fasta) were generated using Samtools mpileup and the SeqIO module in BioPython (Cock et al. 2009). For each sample, a .bed file with SC loci was generated from each .bam file using the "bamToBed" option in bedtools (Quinlan and Hall 2010). A final .bam file containing only SC loci was obtained with samtools (Li et al. 2009). All .bam files with SC loci were visualized with the graphical viewer Tablet (Milne et al. 2010). Three samples (*P. hybrida*, *P. limitata*_205, and *P. limitata*_213) showed evidence of genome-wide heterozygosity (i.e., two character states supported by a similar number of reads). This, together with the lack of such a pattern of heterozygosity in *P. arborescens* and *P. californica*, was considered indicative of putative allopolyploid origin as suggested previously (Neiva et al. 2017).

Files containing sequences, at each SC locus, for all species were obtained with a custom Python script, and aligned (MUSCLE; Edgar 2004) in Translator X (Abascal et al. 2010). All alignments were optimized for phylogenetic analyses by using TrimAl (Capella-Gutiérrez et al. 2009) to remove sites with gaps, resulting in 1771 alignments with no missing data, of which 1431 were at least 500 bp long and were used for further analyses.

PHYLOGENETIC TREE-BASED HOMOEOLOGUE SORTING

Phased homoeologues from each of the putative allopolyploid species were assigned to their parental lineage using a tree distance approach. For each allopolyploid sample, the best-scoring gene trees were estimated from alignments of outgroup, parental, and the two-phased homoeologue sequences using RaxML

Table 1. List of samples used in this study.

Sample name	Species	Sample composition	Place of collection	Coordinates	ENA accessionNo.
P. hybrida	*Pelvetiopsis hybrida*	Five individuals (pooled)	Rancho Marino, CA	35°32′25″N, 121°05′34″W	ERS1987720
P. limitata 205	*Pelvetiopsis limitata*	Five individuals (pooled)	Andrew Molera, CA	36°16′50″N, 121°51′47″W	ERS1987721
P. limitata 213	*P. limitata*	Two individuals (pooled)	Terrace Point, CA	36°56′54″N, 122°03′55″W	ERS1987722
P. californica 1	*Pelvetiopsis californica*	Single individual	Cayucos, CA	35°26′51″N, 120°57′00″W	ERS1987723
P. californica 5	*P. californica*	Single individual	Cayucos, CA	35°26′51″N, 120°57′00″W	ERS1987724
P. californica 6	*P. californica*	Single individual	Cayucos, CA,	35°26′51″N, 120°57′00″W	ERS1987725
P. arborescens	*Pelvetiopsis arborescens*	Five individuals (pooled)	Garrapata, CA	36°28′08″N, 121°56′02″W	ERS1987726
Fucus gardneri	*Fucus gardneri*	Single individual	Berkeley Bay, CA	37°51′35″N, 122°18′56″W	ERS1987727
Fucus ceranoides	*Fucus ceranoides*	Male and female individual	Rio Minho, Portugal	41°53′27″N, 08°49′30″W	ERS1987728
Ascophyllum nodosum	*Ascophyllum nodosum*	Male and female individual	Viana do Castelo, Portugal	41°41′57″N, 08°51′11″W	ERS1987729
Silvetia compressa	*Silvetia compressa*	Single individual	Davenport Landing, CA	37°00′40″N, 122°11′52″W	ERS1987730
Pelvetia canaliculata	*Pelvetia canaliculata*	Single individual	Viana do Castelo, Portugal	41°41′57″N, 08°51′11″W	ERS2041716, ERS2041717
*Parb_4	*P. arborescens*	Three individuals (pooled)	Garrapata, CA	36°28′08″N, 121°56′02″W	ERS2597673
*Parb_5	*P. arborescens*	Three individuals (pooled)	Garrapata, CA	36°28′08″N, 121°56′02″W	ERS2597674
*Parb_6	*P. arborescens*	Three individuals (pooled)	Garrapata, CA	36°28′08″N, 121°56′02″W	ERS2597675
*Pcal_1	*P. californica*	Three individuals (pooled)	Rancho Marino, CA	35°32′25″N, 121°05′34″W	ERS2597676
*Pcal_4	*P. californica*	Three individuals (pooled)	Rancho Marino, CA	35°32′25″N, 121°05′34″W	ERS2597677
*Pcal_5	*P. californica*	Three individuals (pooled)	Rancho Marino, CA	35°32′25″N, 121°05′34″W	ERS2597678
*Phyb_4	*P. hybrida*	Three individuals (pooled)	Rancho Marino, CA	35°32′25″N, 121°05′34″W	ERS2597679
*Phyb_5	*P. hybrida*	Three individuals (pooled)	Rancho Marino, CA	35°32′25″N, 121°05′34″W	ERS2597680
*Phyb_6	*P. hybrida*	Three individuals (pooled)	Rancho Marino, CA	35°32′25″N, 121°05′34″W	ERS2597681

Species names, number of individuals pooled in each sample, place of collection, and accession numbers in the European Nucleotide Archive are given for each sample. CA, California.

***Samples collected for the genetic expression study.**

version 8.1 (Stamatakis 2014) under the General Time Reversible (GTR) + Gamma model. Six fully resolved model trees were chosen from observed gene trees, each with a different placement of the two homoeologue sequences and with at least one of the homoeologue sequences sister to a putative parental lineage. The six model trees represent the three most common topologies recovered from the alignments, for each allopolyploid sample. The Robinson–Foulds distance (Robinson

and Foulds 1981) from each gene tree to each model tree was scored using the Dendropy library (Sukumaran and Holder 2010). Trees with a Robinson–Foulds distance of zero to each of the model trees were listed, and the name of the homoeologue sequences was modified accordingly in the corresponding complete alignments. Homoeologues recovered as sister to *P. arborescens* were named "A" and the alternatives were named "B." Homoeologues of all three putative allopolyploid samples (*P. limitata* 205 and 213 and *P. hybrida*) were identified and renamed in 989 alignments (files available on Dryad).

SPECIES TREE ANALYSIS

The 989 loci were analyzed as a concatenated dataset in RaxML version 8.1. Sequences of *P. canaliculata* were excluded from species tree analyses to avoid the effect of long-branch attraction on node age estimates. A full bootstrap analysis with 200 replicates was conducted using the GTR + Gamma model and *A. nodosum* + *S. compressa* as outgroups (Fig. S1). The numbers of parsimony-informative sites for each of the 989 loci were calculated using PAUP* version 4.0 (Swofford 2003), and 50 loci with the most parsimony-informative sites were chosen for coalescent-based species tree inference. The 50 loci were tested for recombination using the program RDP version 4.7 (Martin et al. 2015). An initial search with the RDP, GENECONV, MaxChi, and Chimaera methods and a *P*-value of 0.05 was performed, and each event found was confirmed with all methods. Recombination was accepted if the event was found by at least two methods with a *P*-value smaller or equal to 0.01, with both parental donors identified and a clear supporting tree. Two loci showed evidence of putative recombination affecting a single sequence (*P. limitata* 205 A in one locus, *P. hybrida* B in the other locus) and were excluded from further analysis. Recombination affecting both *P. hybrida* sequences was detected in four other loci, albeit with *P*-values greater than 0.01, but these genes were also excluded from further analyses. The best-fitting model of nucleotide substitution was inferred for each of the remaining 44 alignments using jModelTest version 2.1 (Darriba et al. 2012). Of the 44 loci, those with 5–7% parsimony-informative sites (25 loci) were chosen for further analysis under coalescent models, in order to avoid both loci with poor phylogenetic signal and fast-evolving loci. A *BEAST (Heled and Drummond 2009) analysis was performed on the 25 loci using the corresponding substitution models and a relaxed uncorrelated clock model. A normal prior with mean $= 3.6E^{-9}$ was chosen for clock rates, an estimate based on rates of SC nuclear loci in plants (Sousa et al. 2014). A normal prior with mean $= 12.5E^{6}$ was defined for the root age, based on age estimates of the split between *Ascophyllum* + *Silvetia* and the remainder of the Fucaceae (Cánovas et al. 2011). Analyses ran for 200×10^{6} mcmc generations in BEAST version 1.8 (http://beast.bio.ed.ac.uk/) at the CIPRES gateway (Miller et al., 2010), and runs were validated using Tracer version 1.6. Trees were summarized with TreeAnnotator version 1.7 with the appropriate burnin.

dN/*dS* Ratios

For dN/dS calculations, all alignments obtained from read mapping to the reference transcriptome were used. The original alignments were optimized for *dN*/*dS* calculations by using a custom Python script ("Locus-pipe.py"), which aligns sequences in the correct frame using Translator X and trims all codons with gap positions using the Gblocks method (Castresana 2000), resulting in 5448 inframe and gap-free alignments. To confirm that the sequences corresponded to functional protein-coding genes rather than pseudogenes, the amino acid alignments generated by TranslatorX were scanned for the presence of stop codons in the middle of sequences. Alignments with stop codons were excluded from *dN*/*dS* calculations. The codeml algorithm in PAML4 (Yang 2007) was used to estimate *dN*, *dS*, and *dN*/*dS* between pairs of samples. To avoid the potential for low-synonymous divergence to amplify *dN*/*dS*, a cut off of $dS \geq 0.01$ for all comparisons was implemented.

HOMOEOLOGUE ASSIGNMENT AND GENE EXPRESSION ANALYSES

A pipeline consisting of read mapping with HyLiTE (Duchemin et al. 2015) and expression analysis using EBSeq (Leng et al. 2013) allowed for a comparison of per locus evolutionary rates for gene transcript sequences within allopolyploid *P. hybrida*, with diploid copies of the same genomes (*P. californica* and *P. arborescens*). A total of 5515 transcripts were mapped using this pipeline. Homoeologue expression levels in the polyploid samples were estimated using HyLiTE, which assigns reads to parental genomes using the mapped alignments of the polyploid and two parental diploid samples as input. For the analysis of *P. hybrida*, the parental samples used were *P. arborescens* and *P. californica.* To normalize the number of expressed reads by locus size and sequencing depth, transcripts per million (TPM) were calculated for each locus from the number of reads assigned to each parent. To correct for cases where TPM = 0 for one of the parents and enable ratio calculations, a small pseudocount was added to all TPM values (TPM + 0.1).

DE ANALYSIS OF *P. HYBRIDA* HOMOEOLOGUES AND PARENTAL ORTHOLOGUES

DE was analyzed using RSEM-EBSeq (Li and Dewey 2011; Leng et al. 2013). Conservatively, DEs between homoeologues and between diploid parental orthologues were assigned to three categories based on posterior probabilities of differential (PPDE) or posterior probabilities of equal expression (PPEE) given by EBSeq: (1) equally expressed (PPDE ≤ 0.05), (2) differentially

expressed (PPEE $\leq$ 0.05), and (3) others (PPDE and PPEE between 0.05 and 0.95). Transcripts were considered only if both homoeologues and orthologues satisfied conditions (1) or (2); those belonging to condition (3) were disregarded. This resulted in 4415 transcripts with statistical support (of 5515, or 80%). Of these, 872 were members of the SC set (83.2% of mapped SC loci). Estimated expression patterns were assigned to four groups and nine combinations of orthologue-homoeologue relationships following the general outline developed in the recent literature (Yoo et al. 2013; Cox et al. 2014); (1) parental expression inheritance, where either equal expression (EE) or DE relationships between orthologues are maintained in the homoeologues (2) homoeologue expression blending, where DE orthologues are EE between homoeologues, (3) homoeologue expression bias, in which EE orthologues are DE in the allopolyploid, and (4) homoeologue expression reversal, where DE orthologues are DE in the opposite direction in the allopolyploid. Expression patterns derived from EBSeq analysis of four groups (*P. arborescens, P. hybrida_A, P. hybrida_B, and P. californica*) were further used to investigate patterns of expression level dominance in genes with DE between parental progenitors.

ANNOTATION

Selected gene sets were functionally annotated by comparing translated amino acid sequences against *Ectocarpus* strain Ec32 at the genome site using Blastp (http://bioinformatics.psb.ugent.be/blast/moderated/?project=orcae_EctsiV2) and against public protein databases (InterPro, Pfam).

Results and Discussion

EVIDENCE FOR ALLOPOLYPLOIDY, PARENTAL ASSIGNMENT, AND SPECIES DATING

From our initial pool of 2027 putative SC loci identified from *Fucus* reference transcriptomes, 989 had homoeologues that were unambiguously assigned to parental genomes in all three allopolyploid samples. The species trees generated from both the concatenated dataset and the 25 loci coalescent analysis fully resolved phylogenetic relationships among the taxa, revealing that the earliest split within the genus *Pelvetiopsis* was between the *P. limitata_B* genome and the sister clade containing *P. arborescens* and *P. californica* (Fig. 1, S1). The position of homoeologues on the tree confirmed: 1) the allopolyploid origin of *P. hybrida* as a hybrid between *P. californica* and *P. arborescens* (previously inferred from multilocus genotyping and mitochondrial DNA sequences by Neiva et al. (2017), and (2) that *P. limitata* originated from a second allopolyploidization event.

Homoeologues from *P. limitata_A* were sister to *P. arborescens*, while the diploid sister to *P. limitata_B* was unsampled. The parental role of *P. arborescens* in both allopolyploid lineages is remarkable, while maternally inherited mitochondrial data (Coyer et al. 2002; Neiva et al. 2017) implies a paternal (sperm) contribution in both cases. The genetic structure of the genus has been extensively sampled and analyzed over the entire biogeographic range (Neiva et al. 2017), and it appears that the species corresponding to the ancestral *P. limitata B* genome is extinct or has at best an extremely limited and/or cryptic extant distribution. While our conclusions rest on genomic features reconstructed from RNA-Seq data, and conclusive cytological evidence of chromosomal duplication are still lacking, the alternative hypothesis of clonally propagating F1 hybrids can be discounted on either ecological or biogeographical grounds. First, although clonal propagation in fucoid seaweeds (with or without hybridization) occurs rarely in marginal, sheltered habitats (Brinkhuis 1976; Tatarenkov et al. 2005; Coyer et al. 2006; Mathieson et al. 2006; Neiva et al. 2012), it is unfeasible in high energy environments such as the intertidal zone. Second, F1 hybrid swarms are not supported either by the very narrow endemic (and nonoverlapping, in the case of *P. arborescens* and *P. hybrida*) range of the paternal ancestor *P. arborescens*, or by the apparent absence/rarity of a diploid counterpart of *P. limitata B* (Neiva et al. 2017). Finally, phylogenetic analysis suggests that allopolyploid and diploid entities are independently evolving and diverging (Fig. 1).

The multispecies coalescent analysis of divergence time for 25 nuclear loci using *BEAST indicated that both allopolyploidization events occurred recently, within the last 400,000 years (late Pleistocene) (Fig. 1). The hybridization event giving rise to *P. limitata* is estimated to have occurred first, between 292,000 and 353,000 years ago, with *P. hybrida* arising subsequently between 50,000 to 162,000 years ago (node height 95% highest posterior densities). The age of the split between *Fucus and Pelvetiopsis* and *Ascophyllum and Silvetia* was inferred at about 9 Ma, which is more recent than (but within the range of) previous estimates (12.5 Ma; Cánovas et al. 2011). The separation between *Fucus* and *Pelvetiopsis* was estimated at 5.6 Ma (end of the Miocene), whereas the common ancestor of the four known extant *Pelvetiopsis* species is estimated at about 1.5 Ma. The gene trees coestimated in *BEAST show topological variation likely to be, at least in part, due to coalescent stochasticity, which justifies the choice of a coalescence-based method for species tree inference and dating.

Considering the broad distribution of allopolyploid *P. limitata* from Point Conception (California) to Vancouver Island (British Columbia), we speculate that it may have replaced its unsampled maternal ancestor across most of its range. Even if this parental lineage is eventually found, it is unlikely to have an equally broad biogeographical distribution or abundance. Likewise, its male ancestor *P. arborescens* is a narrow-endemic climatic relict, present only in a short stretch of coastline just south of

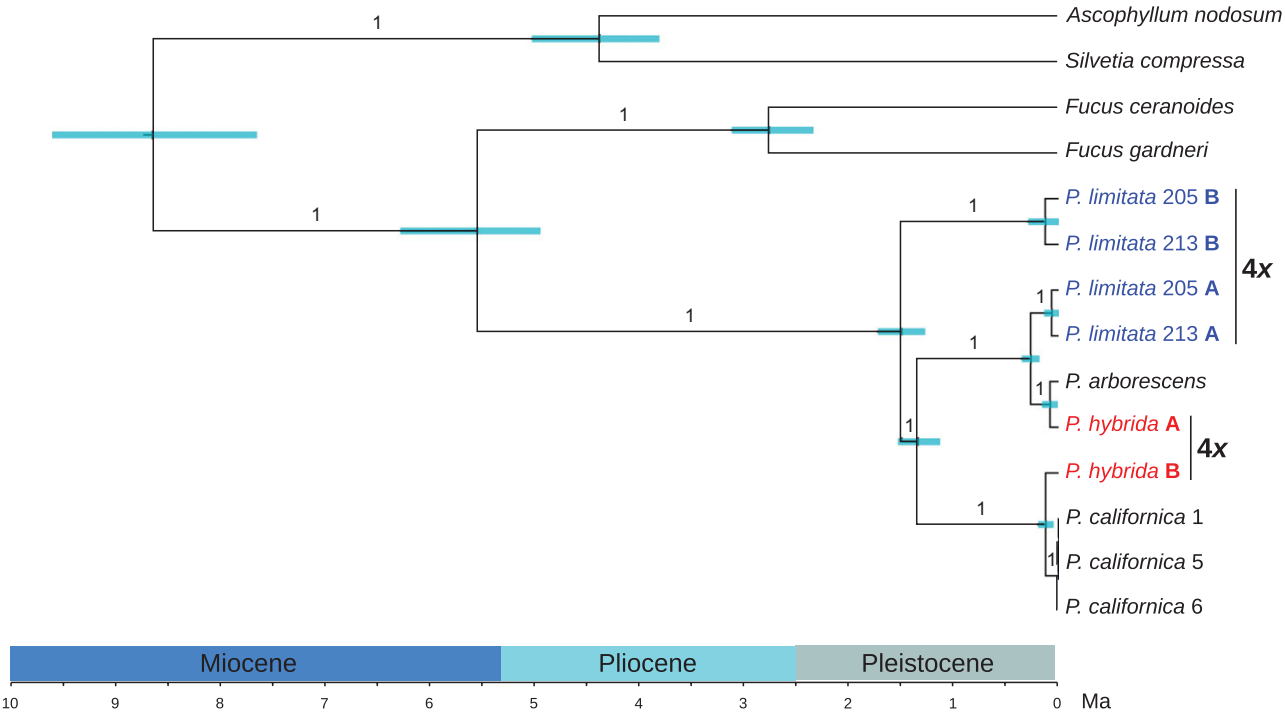

Figure 1. **Phylogenetic relationships in *Pelvetiopsis* inferred from 25 low-copy nuclear loci. The species tree inferred in *Beast is scaled to geological time in units of million years (Ma). Samples corresponding to allopolyploid species are represented by two separate labels, one for each of the subgenomes A and B. Branch support values represent posterior probabilities. Node bars represent the 95% height highest posterior density (HPD) for node ages.**

Monterey Bay (see Neiva et al. 2017). In other words, widespread *P. limitata* is a genomic merger of two relict and extinct ancestral diploid genomes. *P. hybrida* is also a well-established species, albeit within a narrower biogeographical range than *P. limitata* (Neiva et al. 2017). It may be the case, however, that this species is still expanding its range, as it is rather young. The successful establishment of *P. hybrida* may be related to the occupation of a slightly different niche than its diploid progenitors, as the species occupies a higher vertical range in the intertidal than *P. californica* when in sympatry, suggesting greater tolerance to emersion stressor(s), for example, desiccation or thermal extremes (E. Serrão, G. Pearson, J. Neiva, personal observation). Taken together, these observations emphasize the significant role played by allopolyploidy in shaping the genetic diversity and biogeographic range of *Pelvetiopsis* in the north-east Pacific, and in contributing to preserve its evolutionary legacy (as retained in extant genomes) in contrast with the narrow endemism (or extinction) of diploid progenitors.

ACCELERATED EVOLUTIONARY RATES IN HOMOEOLOGUES COMPARED WITH PARENTAL ORTHOLOGUES

Evolutionary rates (*dN*, *dS*, and *dN/dS*) between homoeologues in *P. hybrida* and between the diploid parental orthologues of *P. arborescens* and *P. californica* for 880 SC loci (with $dS \geq 0.01$ and alignment length > 500 bp) are shown in Figure 2A–C. Overall means for all three metrics were greater for homoeologues than orthologues (Table S1). Paired *t*-tests on homoeologue-orthologue comparisons indicated greater *dN/dS* between homoeoelogues than between parental orthologues ($P = 0.0018$). This was a result of increased rates of nonsynonymous substitution *(dN)* between homoeologues in allopolyploid *P. hybrida*, compared to its diploid progenitors (Fig. 2B; $P = 0.0038$). In contrast, *dS* between homoeologues did not differ significantly overall from that between parental gene copies (Fig. 2C; $P = 0.225$), although we observed number of homoeologue outliers with high *dS*. At a genome-wide level, such a pattern is consistent with relaxed selective constraint following genome duplication, due to the presence of homoeologous gene copies, as predicted by theory (Ohno 1970).

To the best of our knowledge, this is the first time that such a pattern has been shown for a brown alga or any marine organism, adding in terms of phylogenetic diversity to previous studies performed using plant model systems. The number of genes used here was sufficiently high to obtain statistical robustness, and comparable to recent studies of similar scope (e.g., Akhunov et al. 2013). Our results are unaffected by possible homoeologue misassignment, as substitution rate estimation compares homoeologue gene pairs independently of taxonomic identity. These results also represent one of the few direct comparisons between a very recently formed allopolyploid species and its diploid ancestors. They

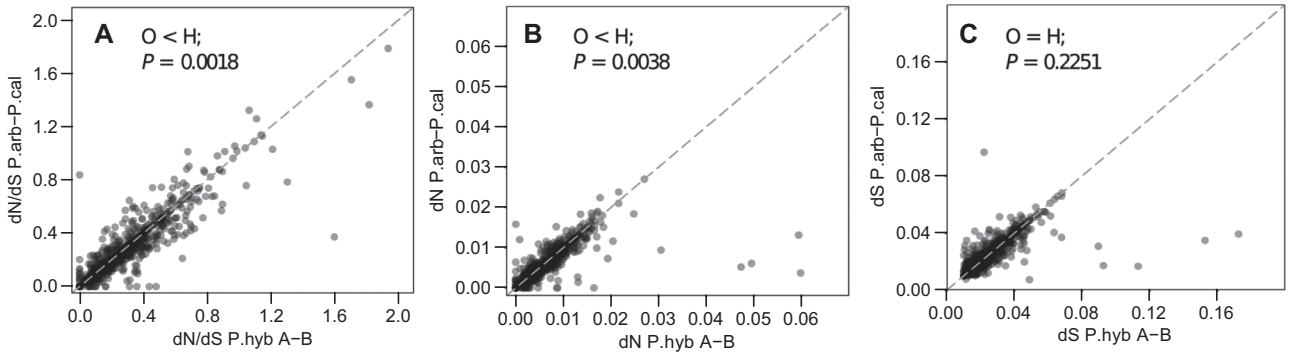

Figure 2. Comparison of evolutionary rates in orthologs and homoeologues. Scatterplots showing values of *dN*, *dS*, and *dN*/*dS* for A and B homoeologues of *Pelvetiopsis hybrida* (*x*-axis) and for parental orthologues (*y*-axis). *P*-values below 0.05 indicate a significant difference in rates (paired *t*-tests), values below the diagonal indicate rates greater in the allopolyploid, and those above indicate rates greater in the parental comparison.

support the idea that relaxed evolutionary constraint acts very early following genome duplication, providing a brief potential for the generation of evolutionary novelty, in contrast to more ancient duplication events where gene silencing and/or degeneration may prevail (Lynch and Conery 2000). In the case of *P. limitata* it was not possible to make the same comparison between parental and daughter genomes, as one of the parental genomes is apparently extinct. Due to the limitations and uncertainties caused by the absence of one of the parental genomes of *P. limitata*, further analyses and discussion focus on the patterns observed in *P. hybrida* only.

LOW LEVELS OF HOMOEOLOGUE SILENCING AND CONSERVED TRANSCRIPTIONAL RESPONSES IN PELVETIOPSIS HYBRIDA

Based on inferred transcription, gene loss in *P. hybrida* was seen because allopolyploidization was not observed in SC genes. Of the 2,027 SC loci identified from *Fucus* sp. reference transcriptomes, expression of 1048 was detected in both the *P. arborescens* and the *P. californica* parental genomes (at least one mapped read), and all were expressed in both the *P. hybrida* A and B homoeologous genomes. Overall, of 5515 transcripts for which we obtained read mapping in both parental species, only one failed to map with *P. hybrida* reads (i.e., <0.02 %), indicating a similar quality of expression data (depth and mapping success) across samples. Homoeologue silencing (i.e., no expression for one homoeologue) was found in 113 genes in the full dataset (2.05%), and in 23 SC genes (1.14%). Such low levels of homoeologue silencing in young allopolyploid lineages appear to be a conserved evolutionary phenomenon, since it has been reported both in natural cotton (*Gossypium hirsutum*, 1–2 Ma) and fungal (about 0.3 Ma) allopolyploids (Yoo et al. 2013; Cox et al. 2014). Indeed, the half-life of duplicated genes has been estimated to be a few million years (Lynch and Conery 2000), an order of magnitude longer than the estimated age of *P. hybrida*. In both the full and SC datasets, homoeologue expression silencing (defined as a complete absence of read mapping for one homoeologue) was approximately threefold more frequent in the *P. hybrida B* (maternal) genome (3.2:1 and 2.8:1 for the full and SC data, respectively: Table S2 and Fig. S2). In addition, reads mapping to two plastid-encoded genes/operons were uniquely assigned to the *P. hybrida-B* homoeologue. Given that organellar inheritance is maternal in fucoid algae (Coyer et al. 2002), this result confirmed the expected identity of the maternal parent (*P. californica*; Neiva et al., 2017), and validated the HyLite pipeline for homoeologue assignment (Table S2).

We next compared homoeologue-orthologue expression patterns in allopolyploid *P. hybrida* and parental lineages. Expression conservation dominated (65.6%), in which relative homoeologue expression matched that of parental orthologues (Fig. 3). Within this group, EE between homoeologues and orthologues dominated. The second most common pattern was acquired homoeologue bias (EE between parental gene copies, but bias between homoeologues; 20%), while the least common categories were homoeologue expression blending (8.5%) and reversed expression bias (6.0%) (Fig. 3). We repeated this analysis using the restricted set of SC genes and obtained similar results (Fig. S3). The fate of gene expression in *P. hybrida* homoeologues is broadly similar to the patterns uncovered in recent analyses of both plant and fungal allopolyploids (Yoo et al. 2013; Cox et al. 2014), particularly with regard to expression conservation. However, they differ in one aspect, that is, acquired homoeologue expression bias was more frequent than expression blending in *P. hybrida*. As a result, there are more genes with divergent expression in the allopolyploid (57.3%) than between parental taxa (43.6%; Table S2; Fig. S4), which is in contrast to patterns recently observed in fungi (Cox et al. 2014). Despite this, our results provide overall support for the generality of gene expression patterns in allopolyploids, here seen for the first time in marine Phaeophyceae, an independently evolving multicellular eukaryotic lineage on the

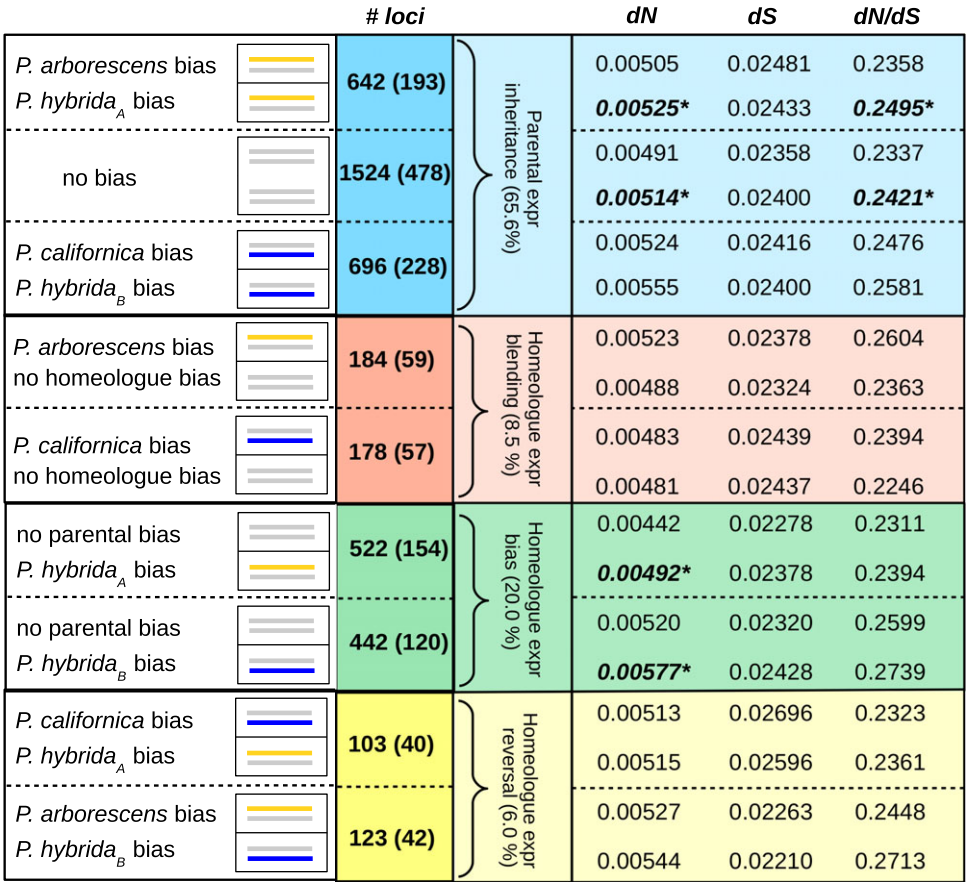

Figure 3. **Comparative expression patterns and evolutionary rates in parental and allopolyploid genomes. Relative expressions of *Pelvetiopsis arborescens*- and *Pelvetiopsis californica*-derived gene copies were calculated for the parental diploid genomes and the *P. hybrida* allopolyploid genome (RSEM-EBSeq pipeline; Li & Dewey 2011; Leng et al. 2013). The nine possible combinations of biased/nonbiased expression (Yoo et al. 2013; Cox et al 2014) are shown at the left, and overexpressed genomes are indicated in color in the boxes (red for *P. arborescens*, blue for *P. californica*) for parental species (upper) and allopolyploid (lower). Equal or underexpression is indicated by grey lines. The numbers of transcript loci with statistical support in each category are shown to the right of the boxes; numbers in parentheses are transcripts for which evolutionary rates are also available. Evolutionary rates are shown to the right (*dN*, *dS*, and *dN*/*dS*); pairs of rates in bold type are significantly higher in the allopolyploid (paired *t*-tests; $P < 0.05$ [*]). Expression patterns are assigned to four categories: expression sharing between parental orthologues and homoeologues (blue shading); bias in parentals, equal expression in homoeologues ("blending", orange shading); equal expression in parentals, bias in homoeologues (green shading); and expression reversal (bias in parentals, opposite bias in homoeologues; yellow shading).**

Tree of Life (Yoon et al. 2004; Cock et al. 2010). All expressed genes included in this analysis were functional, based on the absence of premature stop codons, in contrast with earlier studies in allopolyploid plants (Akhunov et al. 2013).

HOMOEOLOGUE DIVERGENCE INCREASES WITHOUT LOSS OF FUNCTION AND/OR NEOFUNCTIONALIZATION

This study provided an opportunity to compare estimates of evolutionary rate variation (*dN*, *dS*, and *dN/dS*) between homoeologues and orthologues, with estimates of gene expression, to evaluate evolutionary rate variation in the nine expression categories outlined above. Consistent with the analysis of the whole dataset, *dS* was not significantly different within any expression category (Fig. 3). In contrast, *dN* was significantly higher in homoeologues in four categories: expression inheritance with no bias (paired *t*-test, $P < 0.05$; 478 genes); expression inheritance with *P. arborescens and P. hybrida-A* bias ($P < 0.05$; 193 genes); homoeologue expression bias in *P. hybrida_A* ($P < 0.05$; 154 genes), and *P. hybrida_B* ($P < 0.05$; 120 genes). Overall, *dN/dS* tended to be higher in homoeologue than orthologue comparisons (seven out of the nine groups, Fig. 3), but was statistically significant only for the two expression inheritance categories for which *dN* was also significant. These results show that even genes with inherited parental expression (with either no expression bias or bias toward *P. californica* and *P. hybrida_B*) are diverging more rapidly in the allopolyploid than between the parental genomes. This was not a consequence of either loss of function (based

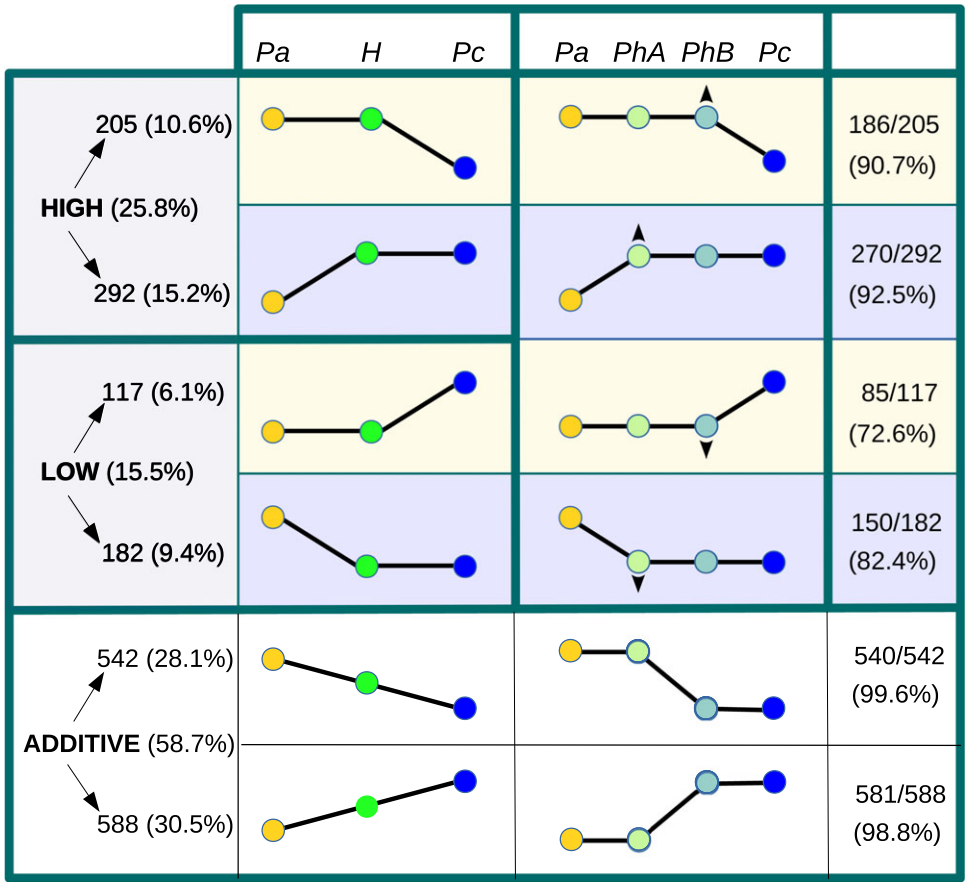

Figure 4. Expression level dominance in parental (*P. arborescens* and *P. californica*) and allopolyploid (*P. hybrida*_A and B) genomes. Data are for genes showing differential expression between parental progenitors. At the left are shown the number and percentage of genes showing high- and low-expression level dominance, and those with additive expression (genes with transgressive expression in *P. hybrida* were not included). In the second column from the left, expression in parental and allopolyploid genomes is depicted (Pa = *P. arborescens*, H = *P. hybrida*, Pc = *P. californica*), with pale yellow and blue backgrounds for expression level dominance involving *P. arborescens* and *P. californica*, respectively. To the right of this, patterns of homoeologue-specific expression are shown (PhA and PhB = *P. hybrida*_A and _B subgenomes, respectively): arrowheads indicate over- or underexpression relative to the respective parental orthologue, together with the proportion and percentage of genes showing that pattern (far right column).

on the absence of inframe stop codons) or neofunctionalization (based on expression patterns being shared with parents). Therefore, other processes contribute to relaxed selective constraint driving the observed divergence, for example, increased tolerance to mildly deleterious substitutions buffered by the greater number of available alleles. Higher rates of *dN* in the allopolyploid were also found in genes showing acquired bias in *P. hybrida*-A. Here, expression bias could result from conflicting regulatory mechanisms, resembling classical dominance, in which factor(s) modulating expression of a gene in the parents preferentially affects one homoeologue over the other in the allopolyploid genetic background, resembling classical dominance (see Fig. 8 in Cox et al. 2014). This smaller number of genes represent candidates for non- or neofunctionalization through unconstrained evolution of the transcriptionally suppressed homoeologue copy.

EXPRESSION LEVEL DOMINANCE IN *P. HYBRIDA* IS DRIVEN BY REGULATORY CHANGES IN NONDOMINANT HOMOEOLOGUES

Expression level dominance refers to the nonadditive patterns of allopolyploid gene expression observed in recent studies (Rapp et al. 2009; Bardil et al. 2011; Yoo et al. 2013; Cox et al. 2014; Wang et al. 2016), in which allopolyploid gene expression tends to resemble one of the parents, rather than being the average of the orthologues. We asked whether this phenomenon exists in *P. hybrida* by analyzing homoeologue-orthologue relationships for 1926 genes that displayed DE between parental orthologues. Expression level dominance was found in 41.3%, and additive expression in 58.7% of the genes, and a clear bias toward high over low expression level dominance was seen (25.8 vs. 15.5%) (Fig. 4).

Our results agree with recent studies (Yoo et al. 2013; Cox et al. 2014) showing that the principal underlying driver of expression level dominance was altered regulation of the nondominant homoeologue (Fig. 4). This was the case for the large majority of genes showing expression level dominance (86.8% overall). We further observed that nondominant homoeologue regulation was more frequent in high compared with low expression level dominance involving either allopolyploid subgenome (Fig. 4; 91.8% vs. 78.6%, respectively), suggesting some regulatory bias toward factors enhancing transcription over those repressing it. In contrast, the vast majority of genes with additive expression (>99% overall) showed no change in homoeologue expression relative to either of the parental species (Fig. 4). Thus, while additivity in *P. hybrida* is largely achieved through transcriptional conservation in the allopolyploid subgenomes (i.e., no change in homoeologue expression relative to parental genomes), the major pattern of expression level dominance implies regulatory changes specifically affecting the nondominant homoeologue. Although the nature of the regulatory mechanisms underlying "transcriptome shock" (Buggs et al. 2011) remains uncertain, and may include both cis- and transacting regulatory elements, as well as epigenetic factors (Ha et al. 2009; He et al. 2010), the acquired regulation of nondominant homoeologues suggests a role for their transactivation or repression (Yoo et al. 2013). Furthermore, it is clear that expression level dominance involves both subgenomes, lending allopolyploids the potential for considerable transcriptomic diversification relative to their diploid progenitors. This potential for phenotypic flexibility and novelty could be a key factor in broadening or even extending the niche occupancy of allopolyploid brown algal lineages (Neiva et al. 2017).

Conclusions

Allopolyploidy was accompanied by diversification and ecogeographic expansion in *Pelvetiopsis*. We present the first study of global gene expression and evolutionary rates in allopolyploid brown algae. This work extends the analysis of allopolyploids to a novel branch of the Tree of Life, and provides support for earlier suggestions that allopolyploid genome evolution follows common trajectories across disparate phylogenetic groups (Yoo et al. 2013; Cox et al. 2014). Levels of gene retention were very high in the allopolyploid *P. hybrida*, which was shown to be of recent evolutionary origin. We provide evidence of evolutionary rate increases in allopolyploid genomic backgrounds. In general, increased divergence (*dN*) was not restricted to homoeologues with strong expression bias (where neofunctionalization or pseudogenization could be suspected), but occurred across several expression categories, an observation consistent with neutral processes associated with genome duplication, and which may be common to allopolyploids more broadly. In common with plant and fungal systems, homoeologue gene expression patterns most frequently reflected those found in the parental lineages, while reduced expression bias (expression blending) in homoeologues was the second most common pattern. Relatively few genes showed either acquired or reversed expression bias in allopolyploids. The prevalence of expression level dominance involving regulation of the nondominant homoeologue is also shared with plant and fungal systems. Thus our data support conserved cross-kingdom transcriptional responses to allopolyploidy, mainly involving a conservative retention of parental gene expression levels.

AUTHOR CONTRIBUTIONS

F. S., J. N., E. A. S., and G. A. P. planned and designed the research. J. N., N. M.,R. J., L. A., P. T. R., and E. A. S. performed experiments and conducted fieldwork. F. S. and G. A. P. analysed data and wrote the manuscript.

ACKNOWLEDGMENTS

Funding by FCT/MEC to CCMAR – UID/Multi/04326/2013; the David and Lucile Packard Foundation and Bureau of Ocean Energy Management, Pew Foundation; Foundation for Science and Technology (FCT) of Portugal PTDC/MAR- EST/6053/2014, EXPL/BIA-EVF/2263/2013, Biodiversa/0004/2015 and post-doctoral fellowships SFRH/BPD/88935/2012 and SFRH/BPD/122567/2016. We thank Dan Reed (MSI, UCSB) for help in the processing of the samples.

DATA ARCHIVING

EMBL Nucleotide Archive (sequences; archived); Dryad (alignments; upon acceptance). The doi of our data is https://doi.org/10.5061/dryad.70955d1.

LITERATURE CITED

Abascal, F., R. Zardoya, and M. J. Telford. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 38:W7–W13.

Adams, K. L., and J. F. Wendel. 2005. Polyploidy and genome evolution in plants. Curr. Opin. Plant Biol. 8:135–141.

Akhunov, E. D., S. Sehgal, H. Liang, S. Wang, A. R. Akhunova, G. Kaur, W. Li, K. L. Forrest, D. See, H. Šimková, et al. 2013. Comparative analysis of syntenic genes in grass genomes reveals accelerated rates of gene structure and coding sequence evolution in polyploid wheat. Plant Physiol. 161:252–265.

Bardil, A., J. D. de Almeid>a, M. C. Combes, P. Lashermes, and B. Bertrand. 2011. Genomic expression dominance in the natural allopolyploid *Coffea arabica* is massively affected by growth temperature. New Phytol. 192:760–774.

Barrier, M., R. H. Robichaux, and M. D. Purugganan. 2001. Accelerated regulatory gene evolution in an adaptive radiation. Proc. Natl. Acad. Sci. 98:10208–10213.

Brinkhuis, B. H. 1976. The ecology of temperate salt-marsh fucoids. I. Occurrence and distribution of *Ascophyllum nodosum* ecads. Mar. Biol. 34:325–338.

Brochmann, C., D. E. Soltis, and P. S. Soltis. 1992. Electrophoretic relationships and phylogeny of Nordic polyploids in *Draba* (Brassicaceae). Plant Syst. Evol. 182:35–70.

Brown, C. T., A. Howe, Q. Zhang, A. B. Pyrkosz, and T. H. Brom. 2012. A reference-free algorithm for computational normalization of shotgun sequencing data. *arXiv:1203.4802v2*.

Buggs, R. J. A., J. F. Wendel, J. J. Doyle, D. E. Soltis, P. S. Soltis, and J. E. Coate. 2014. The legacy of diploid progenitors in allopolyploid gene expression patterns. Philos. Trans. R. Soc. B. Biol. Sci. 369:20130354. https://doi.org/10.1098/rstb.2013.0354.

Buggs, R. J. A., L. Zhang, N. Miles, J. A. Tate, L. Gao, W. Wei, P. S. Schnable, W. B. Barbazuk, P. S. Soltis, and D. E. Soltis. 2011. Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. Curr. Biol. 21:551–556.

Cánovas, F. G., C. F. Mota, E. A. Serrão, and G. A. Pearson. 2011. Driving south: a multi-gene phylogeny of the brown algal family Fucaceae reveals relationships and recent drivers of a marine radiation. BMC Evol. Biol. 11:371. https://doi.org/10.1186/1471-2148-11-371.

Capella-Gutiérrez, S., J. M. Silla-Martínez, and T. Gabaldón. 2009. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17: 540–552.

Cock, J. M., L. Sterck, P. Rouzé, D. Scornet, A. E. Allen, G. Amoutzias, V. Anthouard, F. Artiguenave, J. M. Aury, J. H. Badger, et al. 2010. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. Nature 465:617–621.

Cock, P. J. A., T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, and B. Wilczynski. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25:1422–1423.

Cox, M. P., T. Dong, G. G. Shen, Y. Dalvi, D. B. Scott, and A. R. D. Ganley. 2014. An interspecific fungal hybrid reveals cross-kingdom rules for allopolyploid gene expression patterns. PLoS Genet. 10:e1004180. https://doi.org/10.1371/journal.pgen.

Coyer, J. A., G. Hoarau, G. A. Pearson, E. A. Serrão, W. T. Stam, and J. L. Olsen. 2006. Convergent adaptation to a marginal habitat by homoploid hybrids and polyploid ecads in the seaweed genus Fucus. Biol. Lett. 2:405–408.

Coyer, J. A., A. F. Peters, G. Hoarau, W. T. Stam, and J. L. Olsen. 2002. Inheritance patterns of ITS1, chloroplasts and mitochondria in artificial hybrids of the seaweeds *Fucus serratus* and *F. evanescens* (Phaeophyceae). Eur. J. Phycol. 37:173–178.

Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9:772. https://doi.org/10.1038/nmeth.2109.

Dong, S., and K. L. Adams. 2011. Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids. New Phytol. 190:1045–1057.

Duchemin, W., P.-Y. Dupont, M. A. Campbell, A. R. D. Ganley, and M. P. Cox. 2015. HyLiTE: accurate and flexible analysis of gene expression in hybrid and allopolyploid species. BMC Bioinformatics 16:8. https://doi.org/10.1186/s12859-014-0433-8.

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

Flagel, L. E., J. F. Wendel, and J. A. Udall. 2012. Duplicate gene evolution, homoeologous recombination, and transcriptome characterization in allopolyploid cotton. BMC Genomics 13:302. https://doi.org/10.1186/1471-2164-13-302.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545.

Fu, L., B. Niu, Z. Zhu, S. Wu, and W. Li. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152.

Ha, M., J. Lu, L. Tian, V. Ramachandran, K. D. Kasschau, E. J. Chapman, J. C. Carringtone, X. Chend, X.-J. Wangf, and Z. J. Chen. 2009. Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. Proc. Natl. Acad. Sci. 106:17835–17840.

He, G., X. Zhu, A. A. Elling, L. Chen, X. Wang, L. Guo, M. Liang, H. He, H. Zhang, F. Chen, et al. 2010. Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. Plant Cell 22:17–33.

Hegarty, M. J., G. L. Barker, I. D. Wilson, R. J. Abbott, K. J. Edwards, and S. J. Hiscock. 2006. Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. Curr. Biol. 16:1652–1659.

Hegarty, M. J., and S. J. Hiscock. 2008. Genomic clues to the evolutionary success of polyploid plants. Curr. Biol. 18:435–444.

Heled, J., and A. J. Drummond. 2009. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27:570–580.

Jiao, Y., N. J. Wickett, S. Ayyampalayam, A. S. Chanderbali, L. Landherr, P. E. Ralph, L. P. Tomsho, Y. Hu, H. Liang, P. S. Soltis, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. Nature 473:97–100.

Johansson, M. L., F. Alberto, D. C. Reed, P. T. Raimondi, N. C. Coelho, M. A. Young, P. T. Drake, C. A. Edwards, K. Cavanaugh, J. Assis, et al. 2015. Seascape drivers of *Macrocystis pyrifera* population genetic structure in the northeast Pacific. Mol. Ecol. 24:4866–4885.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–359.

Leitch, A. R., and I. J. Leitch. 2008. Genomic plasticity and the diversity of polyploid plants. Science 320:481–483.

Leng, N., J. A. Dawson, J. A. Thomson, V. Ruotti, A. I. Rissman, B. M. G. Smits, J. D. Haag, M. N. Gould, R. M. Stewart, and C. Kendziorski. 2013. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics 29:1035–1043.

Li, B., and C. N. Dewey. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics 12:323. https://doi.org/10.1186/1471-2105-12-323.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis and R. Durbin. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079.

Li, W., and A. Godzik. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659.

Lynch, M., and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155.

Lynch, M., and A. Force. 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics 154:459–473.

Madlung, A. 2013. Polyploidy and its effect on evolutionary success: old questions revisited with new tools. Heredity 110:99–104.

Mallet, J. 2007. Hybrid speciation. Nature 446:279–283.

Martin, D. P., B. Murrell, M. Golden, A. Khoosal, and B. Muhire. 2015. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol. 1:1–5.

Masterson, J. 1994. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Sci. Pap. Ed. Guid. to Sci. Inf. 264:421–423.

Mathieson, A. C., C. J. Dawes, A. L. Wallace, and A. S. Klein. 2006. Distribution, morphology, and genetic affinities of dwarf embedded *Fucus* populations from the Northwest Atlantic Ocean. Bot. Mar. 49:283–303.

Mayrose, I., S. H. Zhan, C. J. Rothfels, K. Magnuson-Ford, M. S. Barker, L. H. Rieseberg, and S. P. Otto. 2011. Recently formed polyploid plants diversify at lower rates. Science 333:1257.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Pp. 45–52. In 2010 Gateway computing environments workshop (GCE 2010). Institute of Electrical and Electronics Engineers (IEEE), New Orleans, LA.

Milne, I., M. Bayer, L. Cardle, P. Shaw, G. Stephen, F. Wright, and D. Marshall. 2010. Tablet—next generation sequence assembly visualization. Bioinformatics 26:401–402.

Neiva, J., G. A. Pearson, M. Valero, and E. A. Serrão. 2012. Drifting fronds and drifting alleles: range dynamics, local dispersal and habitat isolation shape the population structure of the estuarine seaweed *Fucus ceranoides*. J. Biogeogr. 39:1167–1178.

Neiva, J., E. A. Serrão, L. Anderson, P. T. Raimondi, N. Martins, L. Gouveia, C. Paulino, N. C. Coelho, K. A. Miller, D. C. Reed, et al. 2017. Cryptic diversity, geographical endemism and allopolyploidy in NE Pacific seaweeds. BMC Evol. Biol. 17:1–12.

Niwa, K., and T. Sakamoto. 2010. Allopolyploidy in natural and cultivated populations of *Porphyra* (bangiales, rhodophyta). J. Phycol. 46:1097–1105.

Ohno, S. 1970. Evolution by gene duplication. Springer-Verlag, Berlin, Germany.

Otto, S. P., and J. Whitton. 2000. Polyploid incidence and evolution. Annu. Rev. Genet. 34:401–437.

Pearson, G. A., A. Lago-Leston, M. Valente, and E. A. Serrão. 2006. Simple and rapid RNA extraction from freeze-dried tissue of brown algae and seagrasses. Eur. J. Phycol. 41:97–104.

Phillips, N., D. F. Kapraun, A. G. Garreta, M. A. R. Siguan, J. L. Rull, N. S. Soler, R. Lewis, and H. Kawai. 2011. Estimates of nuclear DNA content in 98 species of brown algae (*Phaeophyta*). AoB Plants 11:1–8.

Quinlan, A. R., and I. M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

Rapp, R. A., J. A. Udall, and J. F. Wendel. 2009. Genomic expression dominance in allopolyploids. BMC Biol. 7:18. https://doi.org/10.1186/1741-7007-7-18.

Rensing, S. A. 2014. Gene duplication as a driver of plant morphogenetic evolution. Curr. Opin. Plant Biol. 17:43–48.

Rho, M., H. Tang, and Y. Ye. 2010. FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Res. 38:e191. https://doi.org/10.1093/nar/gkq747.

Robinson, D. F., and L. R. Foulds. 1981. Comparison of phylogenetic trees. Mathematical Biosciences 53:131–147.

Schulz, M. H., D. R. Zerbino, M. Vingron, and E. Birney. 2012. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28:1086–1092.

Senchina, D. S., I. Alvarez, R. C. Cronn, B. Liu, R. D. Noyes, and J. F. Wendel. 2003. Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. Mol. Biol. Evol. 20:633–643.

Seoighe, C., and K. H. Wolfe. 1999. Yeast genome evolution in the post-genome era. Curr. Opin. Microbiol. 2:548–554.

Silberfeld, T., J. W. Leigh, H. Verbruggen, C. Cruaud, B. deReviers, and F. Rousseau. 2010. A multi-locus time-calibrated phylogeny of the brown algae (Heterokonta, Ochrophyta, Phaeophyceae): Investigating the evolutionary nature of the "brown algal crown radiation." Mol. Phylogenet. Evol. 56:659–674.

Soltis, D. E., V. A. Albert, J. Leebens-Mack, C. D. Bell, A. H. Paterson, C. Zheng, D. Sankoff, W. Claude, C. W. dePamphilis, P. K. Wall., and et al. 2009. Polyploidy and angiosperm diversification. Am. J. Bot. 96:336–348.

Soltis, D. E., P. S. Soltis, and L. H. Rieseberg. 1993. Molecular data and the dynamic nature of polyploidy. CRC Crit. Rev. Plant Sci. 12:243–273.

Sousa, F., Y. J. K. Bertrand, S. Nylinder, B. Oxelman, J. S. Eriksson, and B. E. Pfeil. 2014. Phylogenetic properties of 50 nuclear loci in *Medicago* (Leguminosae) generated using multiplexed sequence capture and next-generation sequencing. PLoS ONE 9:e109704.

Spring, J. 1997. Vertebrate evolution by interspecific hybridisation–are we polyploid? FEBS Lett. 400:2–8.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Sukumaran, J., and M. T. Holder. 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26:1569–1571.

Swofford, D. L. 2003. PAUP*: phylogenetic analysis using parsimony, version 4.0 b10.

Tatarenkov, A., L. Bergström, R. B. Jönsson, E. A. Serrão, L. Kautsky, and K. Johannesson. 2005. Intriguing asexual life in marginal populations of the brown seaweed *Fucus vesiculosus*. Mol. Ecol. 14:647–651.

Wang, X., H. Zhang, Y. Li, Z. Zhang, L. Li, and B. Liu. 2016. Transcriptome asymmetry in synthetic and natural allotetraploid wheats, revealed by RNA-sequencing. New Phytol. 209:1264–1277.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.

Yoo, M. J., E. Szadkowski, and J. F. Wendel. 2013. Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity 110:171–180.

Yoon, H. S., J. D. Hackett, C. Ciniglia, G. Pinto, and D. Bhattacharya. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Mol. Biol. Evol. 21:809–818.

Zhang, H., X. Gou, A. Zhang, X. Wang, N. Zhao, Y. Dong, L. Li, and B. Liu. 2016. Transcriptome shock invokes disruption of parental expression-conserved genes in tetraploid wheat. Sci. Rep. 6:26363.

Associate Editor: S. Smith
Handling Editor: M. Servedio

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table 1. Evolutionary rates (*dN*, *dS*, and *dN*/*dS*) for single copy genes ($dS \geq 0.01$) between *Pelvetiopsis hybrida* homoeologues and the parental taxa *Pelvetiopsis californica* and *Pelvetiopsis arborescens*.

Table 2. Gene expression and annotation data for 1747 putative single-copy genes in homoeologues of *P. hybrida* (A and B genomes) and orthologues of the parental species *P. arborescens* and *P. californica*.

Figure S1. Maximum-likelihood summary tree inferred from a concatenated dataset of 989 single-copy nuclear loci using the GTR + Gamma model.

Figure S2. Scatterplots of comparative expression (as Log10 transcripts per million [TPM]) between 1747 single-copy orthologues from parental taxa *P. arborescens* and *P. californica*, and homoeologues identified from *P. hybrida* (genomes A and B).

Figure S3. Relative expression of *P. arborescens*- and *P. californica*-derived gene copies was calculated for the parental diploid genomes and the *P. hybrida* allopolyploid genome (RSEM-EBSeq pipeline; Li & Dewey 2011; Leng et al. 2013) using the single-copy (SC) dataset.

Figure S4. Fold-change (FC) ratios (as Log_2) of expression data between homoelogues (*x*-axis) plotted against parental orthologs (*y*-axis) for the full dataset of 1048 SC gene transcripts.