

# New Genes Interacted With Recent Whole-Genome Duplicates in the Fast Stem Growth of Bamboos

Guihua Jin,<sup>†,1</sup> Peng-Fei Ma<sup>1</sup>, Xiaopei Wu,<sup>†,1</sup> Lianfeng Gu,<sup>2</sup> Manyuan Long,<sup>\*,3</sup> Chengjun Zhang,<sup>\*,1</sup> and De-Zhu Li<sup>\*,1</sup>

<sup>1</sup>Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, China

<sup>2</sup>Basic Forestry and Proteomics Research Center, College of Forestry, Fujian Agriculture and Forestry University, Fuzhou, Fujian, China

<sup>3</sup>Department of Ecology and Evolution, The University of Chicago, Chicago, IL, USA

<sup>†</sup>These authors contributed equally to this work.

\*Corresponding authors: E-mails: dzl@mail.kib.ac.cn; zhangchengjun@mail.kib.ac.cn; mlong@uchicago.edu.

Associate editor: Michael Purugganan

## Abstract

As drivers of evolutionary innovations, new genes allow organisms to explore new niches. However, clear examples of this process remain scarce. Bamboos, the unique grass lineage diversifying into the forest, have evolved with a key innovation of fast growth of woody stem, reaching up to 1 m/day. Here, we identify 1,622 bamboo-specific orphan genes that appeared in recent 46 million years, and 19 of them evolved from noncoding ancestral sequences with entire de novo origination process reconstructed. The new genes evolved gradually in exon–intron structure, protein length, expression specificity, and evolutionary constraint. These new genes, whether or not from de novo origination, are dominantly expressed in the rapidly developing shoots, and make transcriptomes of shoots the youngest among various bamboo tissues, rather than reproductive tissue in other plants. Additionally, the particularity of bamboo shoots has also been shaped by recent whole-genome duplicates (WGDs), which evolved divergent expression patterns from ancestral states. New genes and WGDs have been evolutionarily recruited into coexpression networks to underline fast-growing trait of bamboo shoot. Our study highlights the importance of interactions between new genes and genome duplicates in generating morphological innovation.

**Key words:** fast stem growth, orphan genes, de novo genes, WGD, woody bamboos, evolutionary innovation.

## Introduction

Evolutionary innovations are throughout the Tree of Life and contribute to organismal diversification (Pigliucci 2008; Wagner and Lynch 2010). The morphological and physiological innovations allow organisms to explore new niches to generate biological diversity. Examples of major innovations have been known from flowers in angiosperms, lung in tetrapods, feathers in birds, and wings in insects (Liem 1988; Averof and Cohen 1997; Albert et al. 2002; Prum and Brush 2002). How novel traits originate becomes a fundamental question in evolutionary biology. However, the majority of studies have focused on external environmental induce for innovations. We know little about the genetic basis underlying the appearance of phenotypic novelties, despite a link between the innovations and genomic novelty in general and new genes in particular that has long been hypothesized (Chen et al. 2013; Erwin 2021). Recent studies provided direct evidence in support of this hypothesis by attributing phenotypic innovation to the evolution of new genes (Kaessmann 2010; Chen et al. 2011; Long et al. 2013; Erwin 2015; Santos et al. 2017). Evolutionary new genes, also called as lineage-specific genes that emerged recently in a given lineage, can be derived from existing genetic elements like the process of

divergence of duplicated genes (Long et al. 2013). A dramatic genetic novelty is the orphan genes created through various molecular evolutionary processes, especially de novo genes from noncoding sequences with several well-documented organisms (Carvunis et al. 2012; McLysaght and Gueronzi 2015; Xie et al. 2019; Zhang et al. 2019; Heames et al. 2020). New genes in animals and plants tend to be preferentially expressed in the male reproductive tissues (Dai et al. 2006), and thus an “out of testis” hypothesis for the emergence of new genes has been proposed (Dai et al. 2006; Vinckenbosch et al. 2006).

Recent advances in genome sequencing of multiple related species enable to explore the relationship between evolutionary innovations and origin of new genes. The species distribution of a gene suggests its possible age (Long et al. 2013). Computationally, by studying genome evolution within a phylogenetic content, phylostratigraphy can detect evolutionary origin of orphan genes through sequence similarity searches in genomes across the Tree of Life (Domazet-Loso et al. 2007). This approach identifies specifically the origin of new genes which lack traceable homologs to existing genes in other lineages. These orphan genes may have been created in several alternative mechanisms from rapid sequence

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

evolution of homologous copies to lateral transfer to de novo origination (Moyers and Zhang 2016; Domazet-Loso et al. 2017; Zhang et al. 2019; Vakirlis et al. 2020; Jin et al. 2021). It assigns every gene within a genome to a given phylogenetic rank designated as phylostratum (PS)—describing the age of gene in a phylogenetic context, parallel to chronostratigraphic age in geology. The genes underlying innovation are thus hypothesized to be enriched in the corresponding phylostratum when innovation first emerged (Sestak et al. 2013; Sestak and Domazet-Loso 2015; Trigós et al. 2017; Shi et al. 2020). To distinguish the orphan genes that are created de novo from other alternative mechanisms, the closely related species is searched for their noncoding ancestral sequences and reconstruction of origination processes (Murphy and McLysaght 2012; Zhang et al. 2019). In combination with gene expression analyses, further insights can be gained into the birth process of new genes and their potential functions. Two similar transcriptome indices for gene, the transcriptome age index (TAI) and transcriptome divergence index (TDI), have been developed with higher values indicating younger and more divergent transcriptome (Domazet-Loso and Tautz 2010; Quint et al. 2012). Moreover, new genes need to be recruited into the genetic network to be functional and the interaction of genes can be identified by the gene coexpression network based on common expression profiles (Shao et al. 2019).

Here we use the bamboos (Poaceae, Bambusoideae) as a model system to study this question. Bambusoideae is the only major grass lineage diversifying into the forest habit with more than 1,600 species worldwide (Soreng et al. 2017). In contrast to other grasses, the majority of bamboo species have woody, tall, and lignified stems (fig. 1A), reaching up to approximately 20 m in the widely cultivated moso bamboo (*Phyllostachys edulis*) and even more than 30 m in a few species such as *Dendrocalamus sinicus* (Chang and Wu 2000; Song et al. 2016), being the tallest grasses in the world. Furthermore, the growth and development of these tall stems can be rapidly completed within 2–3 months, showing a “slow-fast-slow” pattern (Zhou 1983). For example, the shoot of moso bamboo can grow 1 m in height within 24 h (Ueda 1960; Liese and Kohl 2015; Song et al. 2016), which is hundreds of times faster than other woody trees (Li et al. 2020). This trait of fast growth of woody stem could be considered as a key innovation in bamboos, likely facilitating their adaptation to the forest habit with access to light and thus vast species diversification. Another unique trait of bamboos is the infrequent flowering with intervals as long as 20–60 years while a high ability of propagation by clone (Janzen 1976).

Previous works have mainly focused on morphology, anatomy, and physiology to study the trait of fast growth of woody stem in bamboos (Ueda 1960; He et al. 2002; Wang, Ren, et al. 2012; Song et al. 2016). Only a few genes related to plant hormones network, cell cycle regulation and cell wall metabolism have been investigated for the evolution of this unique trait (Cui et al. 2012; Peng, Lu, et al. 2013; Li et al. 2018; Wei et al. 2018). Recent genomic studies revealed the polyploidization history of woody bamboos (Guo et al. 2019), which was suggested to be related to the origin and evolution

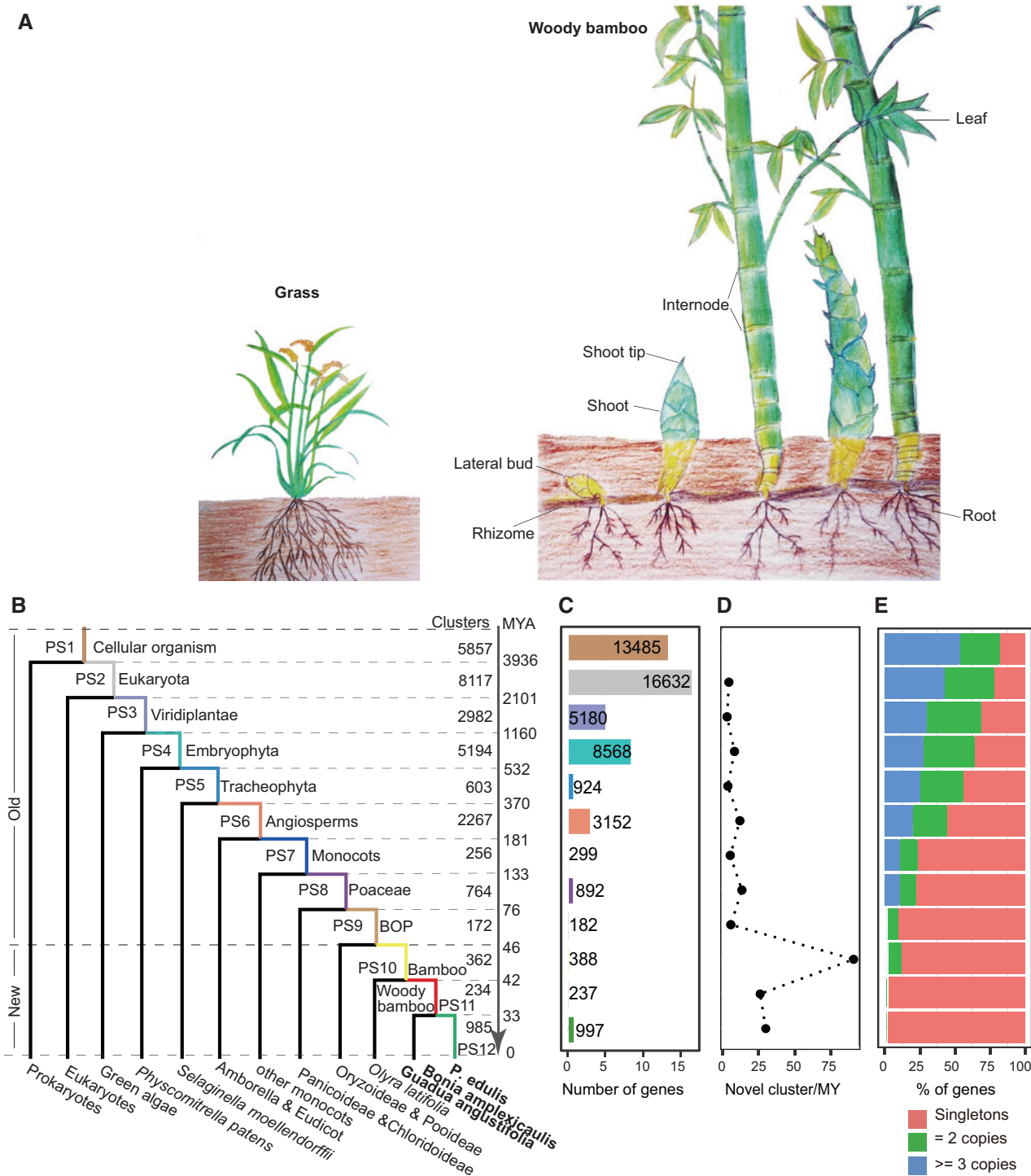
of this trait. By taking advantage of bamboo genomes recently sequenced, we combined phylostratigraphy with a noncoding ancestral search for orphan and de novo gene candidates (Zhang et al. 2019), to determine the phylogenetic age of genes in the genomes of woody bamboos and found that new genes were highly and specifically expressed in the shoot tissue. The noncoding ancestral sequences, as core evidence of de novo genes, are rarely detected in metazoans from mammals to invertebrates. The *Oryza* genomes in the grass family provided the first genomic evidence for existence of de novo genes (Zhang et al. 2019). We inferred that genomes of bamboos, as a grass subfamily Bambusoideae, likely contain de novo genes with detectable noncoding ancestral sequences. Our study highlights a central role for orphan and de novo genes, through interaction with whole-genome duplicates (WGDs), in the origination and evolution of morphological innovations in bamboos, pinpointing a general correlation between phenotypic innovations and genomic novelty.

## Results

### Phylogenetic Origination of Woody Bamboo Genes and Their Features

To trace the phylogenetic origins of woody bamboo genes, we selected one herbaceous bamboo *Olyra latifolia* and three woody bamboos with sequenced genomes (Zhao et al. 2018; Guo et al. 2019) for phylostratigraphic analysis, together with other 65 representative genomes across the Tree of Life (supplementary table 1, Supplementary Material online). The three woody species, *Bonia amplexicaulis*, *Guadua angustifolia*, and *P. edulis*, represent all of the three major lineages of woody bamboos in the Bambusoideae (Guo et al. 2019). We separately adopted the genes of *Bo. amplexicaulis* and *P. edulis*, for which the quality of genome assemblies is much higher than that of *G. angustifolia*, as query to conduct analyses. In considering the polyploid nature of woody bamboos, the paralogous genes within genome were gathered to generate nonredundant queries (supplementary fig. 1, Supplementary Material online). In all, we defined 12 phylostrata (PS) ranks ranging from the oldest PS1 (cellular organisms) to the youngest PS12 (species specific) based on the ladder-like phylogenetic tree and nearly identical results were obtained in using *P. edulis* and *Bo. amplexicaulis* (fig. 1B–E; supplementary fig. 2, Supplementary Material online). We thus focused on the results below using the genes of *P. edulis* as query.

A total of 50,936 genes from *P. edulis* were assigned to 12 PSs (fig. 1B and C) and the number of genes per rank was positively correlated with gene ages (Pearson correlation coefficient  $r = 0.7742$ ,  $P = 0.00312$ ). There were five peaks of appearance of genes during the evolutionary history of bamboos, which were associated with the emergence of cellular organisms (PS1), Eukaryota (PS2), Viridiplantae (PS3), Embryophyta (PS4), and Angiosperms (PS6). This pattern of distribution for gene ages was also observed in the previous studies of *Arabidopsis* and rice (*Oryza sativa*) (Cui et al. 2015). An overwhelming proportion of 96.82% genes could be traced back to the old gene ranks from PS1 to PS9 before



**FIG. 1.** Morphological character and phylostratigraphic age of *Phyllostachys edulis*. (A) Comparison of the growth pattern of grass and woody bamboo. (B) Phylostratigraphic ages distribution of *P. edulis*. Numbers denote the number of nonredundant gene clusters per phylostratum (PS1–PS12). The tree stratifies species by major evolutionary innovations, from the emergence of simple unicellular organisms up to *P. edulis*. The dating of phylogenetic tree is from TimeTree (Kumar et al. 2017). (C) Number of genes per phylostratum. (D) Gene fixation rate per PS. (E) Gene copies distribution of each cluster.

the origin of bamboos. As such, there were only 1,622 genes emerging in the bamboo lineage, including 388 of PS10, 237 of PS11, and 997 of PS12 (supplementary table 2, Supplementary Material online), which could be considered as bamboo-lineage specific. We called them evolutionary new genes, or orphan genes, thereafter. Splitting from the subfamily

Pooideae approximately 46 million years (Mys) ago, the bamboo lineage has a rate of origination of approximately 35 orphan genes/My. We also calculated the fixation rate of gene families for different PSs and found that the internode leading to bamboos corresponding to PS10 had the highest rate of 90.5 clusters/My (fig. 1D; supplementary table 3,



[Supplementary Material online](#)). Interestingly, the opportunity for genes to being singleton decreased with their ages, and 41.15% of the PS1 genes had multiple copies whereas 98.78% of the PS12 genes were all single copy ([fig. 1E](#)).

### New Genes Preferentially Expressed in the Fast-Growing Phase of Shoot

To identify the expression pattern of each phylostratum, we used transcriptome data from seven tissues of *P. edulis*, including root, leaf, rhizome tip, later bud, shoot tip, inflorescence, and shoot ([supplementary fig. 3](#) and [table 4](#), [Supplementary Material online](#)). Among them, the shoot tissue contained eight different development stages according to the height of shoot ([fig. 2A](#)). Expression was observed for 81.10% of the total genes (41,307 genes) having a value of fragments per kilobase million (FPKM)  $\geq 1$  at least in one of the seven tissues ([supplementary tables 5 and 6](#), [Supplementary Material online](#)). Surprisingly, the new genes tended to be more expressed than the old ones with 71.91% of PS10, 86.92% of PS11, and 83.65% of PS12 genes expressed whereas a range of 66.03–83.42% of PS1–PS9 genes expressed.

Phylotranscriptomic analyses combining gene age and expression information revealed generally higher TAI values for the shoot than the other tissues ([fig. 2A](#)). This value rose to the highest point in the shoots of 0.5–2.0 m in height, corresponding to the turning point from the slow to fast growth of the *P. edulis* shoot ([Song et al. 2016](#)). A similar trend was also observed with TDI ([fig. 2B](#)). These results suggested higher expression level of new genes and thus divergent transcriptomes in the shoot, particularly at the height of 0.5–2.0 m. Similarly, at the individual level of phylostratum, highly elevated expression level of new genes of PS11, originating at the common ancestor of woody bamboos, was found in the shoots and also peaking at the 2.0 m stage ([fig. 2C](#)). All the genes from the remaining phylostrata with the exception of PS5–PS7 and PS9, whose genes had a much lower degree of altered expression than that of PS11, showed no significant changes of expression across the tissues. To avoid the impact of potential spurious new genes with low expression levels, we further analyzed gene ages and expression levels for the top 1,000 highly expressed genes per tissue. An index of expression preference (EP = expression frequency/gene frequency) was calculated for each phylostratum ([supplementary tables 7 and 8](#), [Supplementary Material online](#)) with greater value pointing to preferred expression of genes in a given tissue. Results showed that the genes of PS11 had the highest EP value in the shoot tissue ([fig. 2D](#)), further proving new genes highly and specifically expressed in the shoot. In addition, for the average gene expression level, the genes of PS1 had the highest expression level and significantly higher than other PSs in all the tissues excluding shoot ( $P < 0.05$  in the Wilcoxon rank sum test, [fig. 2E](#); [supplementary fig. 4](#), [Supplementary Material online](#)). In contrast, the genes of PS11 showed the highest expression level in the shoot, indicating a special role of the PS11 genes within the new genes in the rapid growth of *P. edulis* shoot.

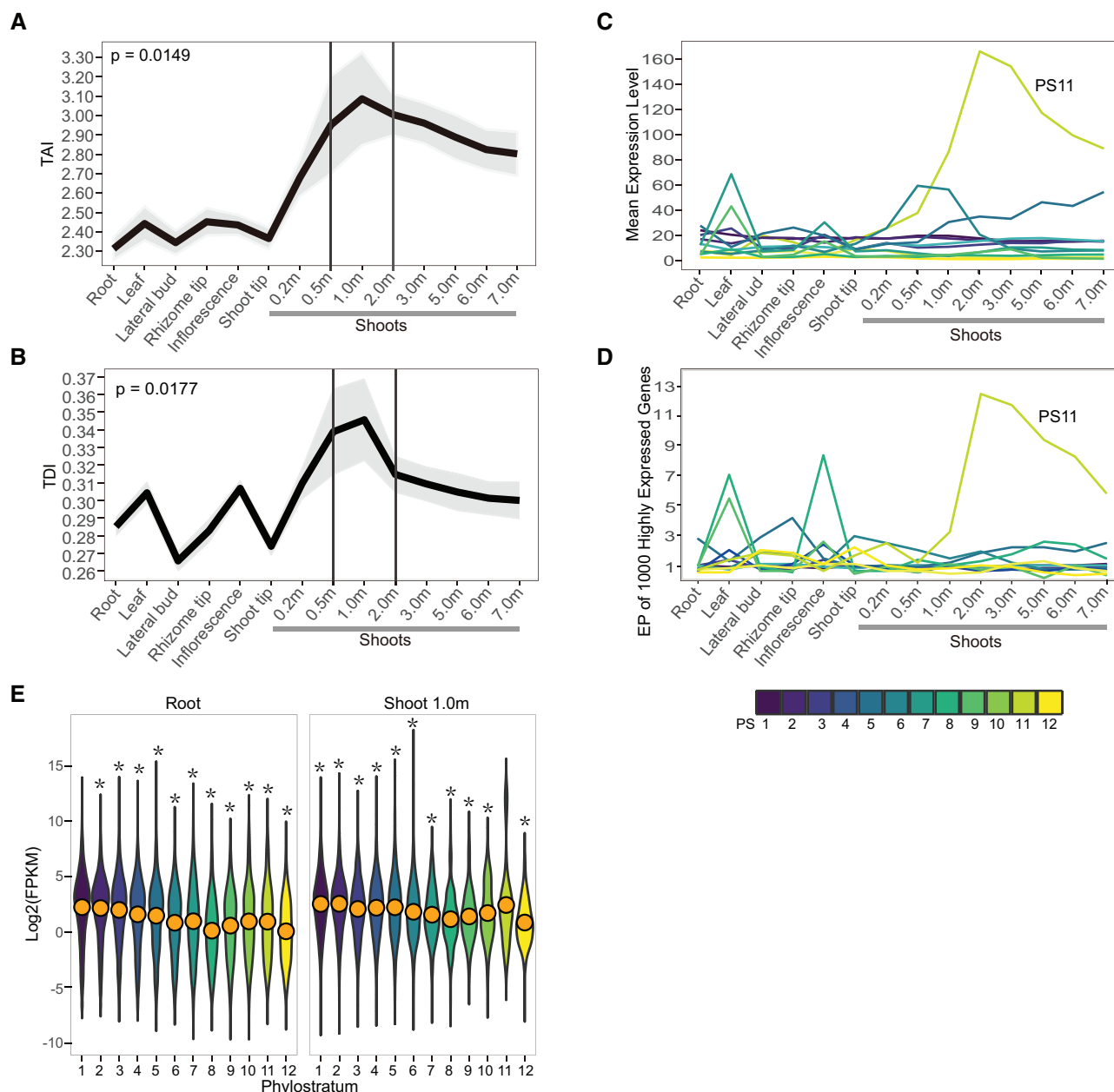
To check whether these findings were applicable to other woody bamboos, we conducted a similar analysis for the *D. sinicus* with available transcriptome data ([Chen et al. 2018](#)). In agreement with the observations in *P. edulis*, the TAI value reached a peak at the fast growth stage of the *D. sinicus* shoot and new genes were also highly expressed in this stage ([supplementary fig. 5](#), [Supplementary Material online](#)). Together, these results suggested that the findings above could hold true for the whole lineages of woody bamboos. In sum, the expression data above reveal that the new genes especially those from PS11 have played an important role in the rapid growth of shoot for woody bamboos.

### Nineteen PS11 Genes Identified as De Novo Genes and Their Step-Wise Evolution of New Genes

To gain insight into the origination process of new genes of PS11, we performed a pipeline as in [Zhang et al. \(2019\)](#) to detect those of being de novo origin, which could represent veritable leaps of evolutionary innovation ([Blevins et al. 2021](#)). Among the 237 PS11 genes, we identified 19 de novo genes with high confidence ([supplementary table 9](#), [Supplementary Material online](#)). All of these de novo genes had complete coding frames in the woody bamboos *P. edulis* and *Bo. amplexicaulis*, whereas their orthologous sequences were noncoding in at least one of the four outgroup species (*Ol. latifolia*, *Raddia distichophylla*, *Brachypodium distachyon*, and rice). Three examples of identified genes were shown for the de novo origination processes in [figure 3A–C](#).

For the de novo gene *PH02Gene41079*, it became a gene through one nucleotide substitution at 19 bp from “TAG” to “GAG” to remove premature stop codon ([supplementary fig. 6](#), [Supplementary Material online](#)). For *PH02Gene36478*, it originated by an 10 bp deletion to resolve frameshift and the premature stop codon ([fig. 3B](#); [supplementary fig. 7](#), [Supplementary Material online](#)); For *PH02Gene38072*, it transformed through multiple steps including four substitution and one indel mutation ([fig. 3C](#); [supplementary fig. 8](#), [Supplementary Material online](#)). Totally, there were seven de novo genes formed by substitution like that and the remaining 12 genes by indel mutation in a similar way ([supplementary table 9](#), [Supplementary Material online](#)). Furthermore, all the de novo genes had transcriptional evidence supported by full-length transcriptome data in *P. edulis* and/or *Bo. amplexicaulis*. Among them, 11 genes also had transcriptional evidence in *Ol. latifolia*, suggesting that the noncoding RNA transcription had emerged earlier than the open reading frames (ORFs) in evolution, similar to *Oryza* ([Zhang et al. 2019](#)).

We further examined the 19 de novo genes for their potential functionality and particularly from the translational evidence. Firstly, we calculated the substitutions at synonymous sites (Ks), nonsynonymous sites (Ka) and Ka/Ks value of orthologs between *P. edulis* and *Bo. amplexicaulis*. Among them, we detected nine genes with Ka/Ks value significantly less than one and two genes significantly larger than 1 ( $P < 0.05$ ,  $\chi^2$  test) ([supplementary table 10](#), [Supplementary Material online](#)), indicating purifying and positive selection, respectively. These results suggest they are undergoing

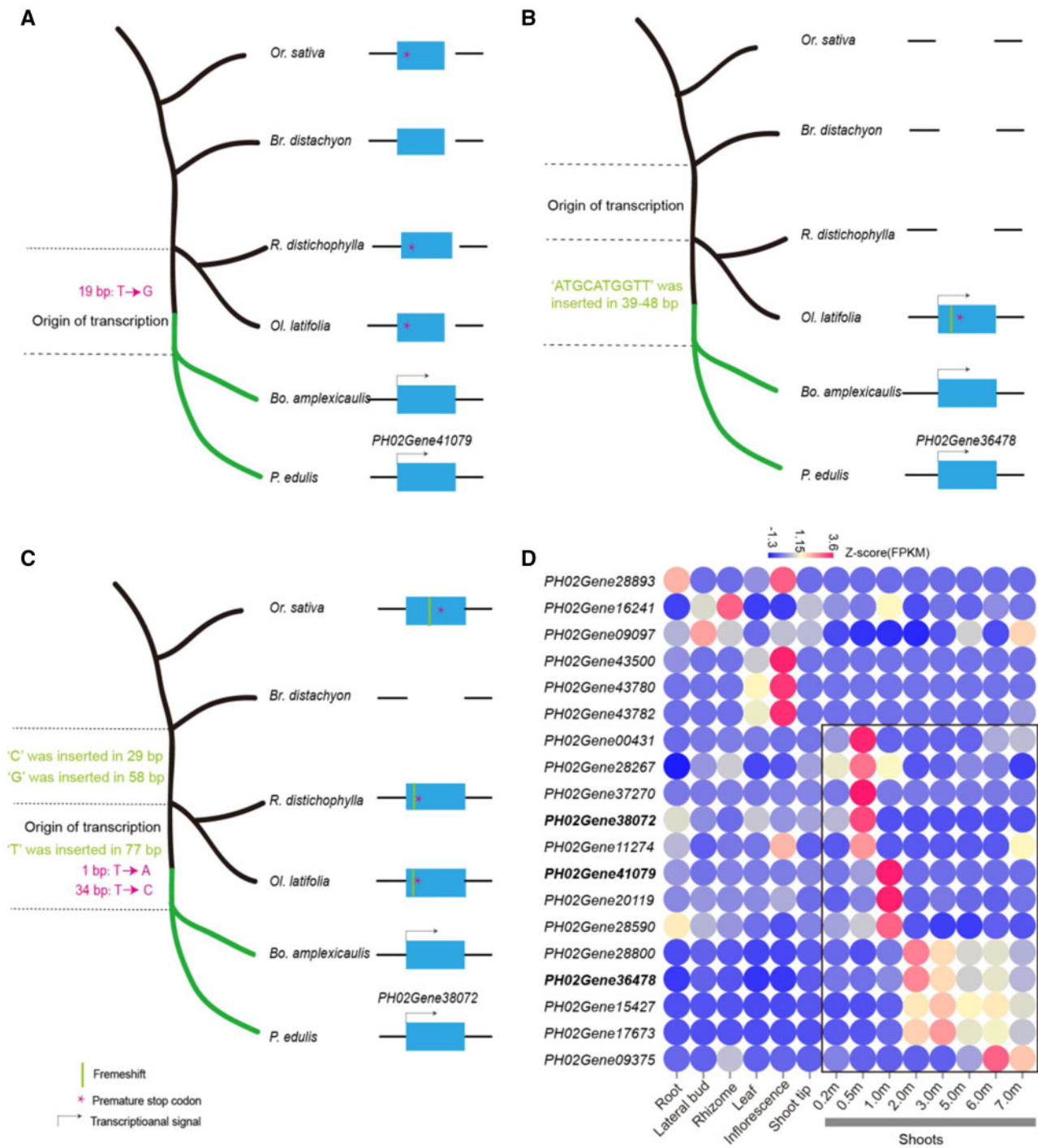


**FIG. 2.** Evolutionary transcriptome profiles of *Phyllostachys edulis*. (A) TAI profiles. A high TAI value indicates that the sample expresses a high number of younger genes. (B) TDI profile. The gray shade represents the standard deviation by permutation analysis. The  $P$  value was derived by application of flat line test. (C) Mean expression level of genes from each phylostratum (PS). (D) EP of the top 1,000 highly expressed genes. (E) The expression level of each PS in root and in shoot 1.0 m. The Wilcoxon rank sum test was used for comparison of expression difference between PS1 and other PSs ( $*P < 0.05$ ) in root, and between PS11 and other PSs in shoot 1.0 m.

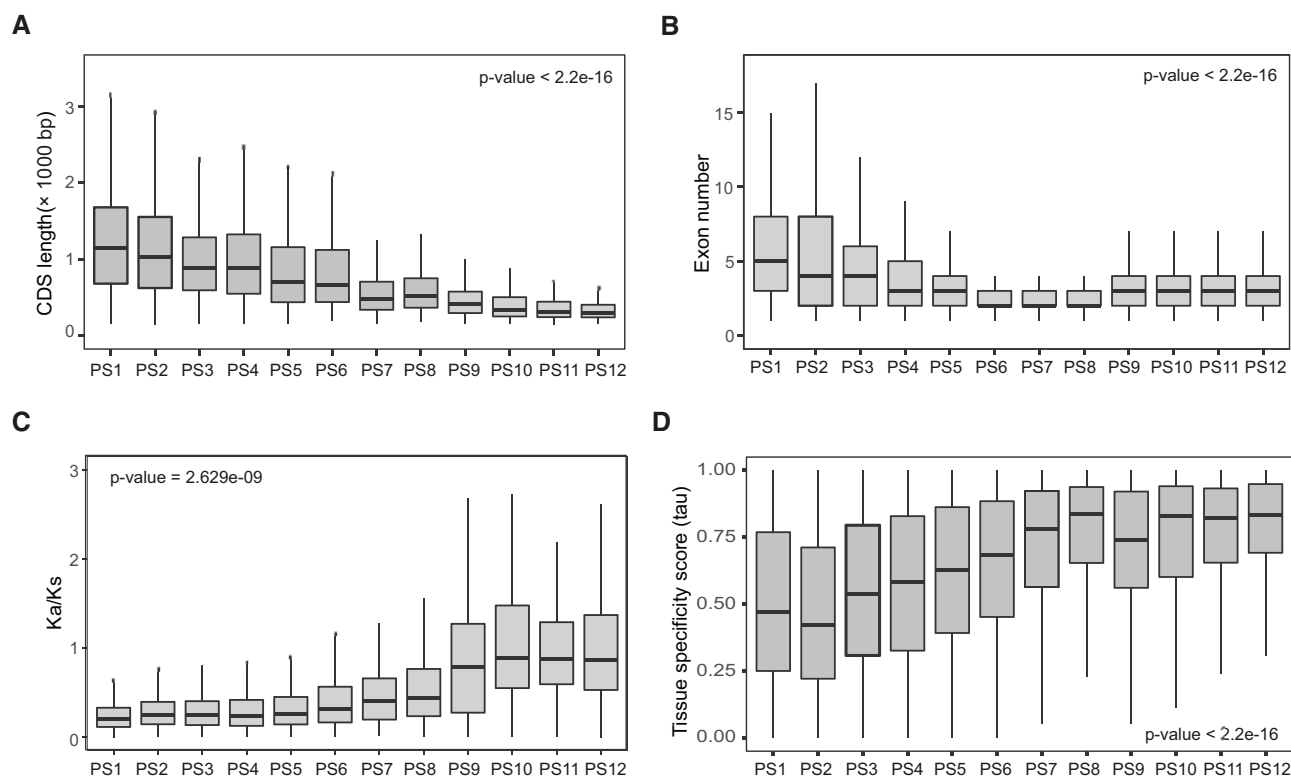
evolutionary constraints and thus support the coding potential of these de novo genes. We subsequently analyzed the expression patterns of all de novo genes in the seven *P. edulis* tissues and found that all of them were tissue-specific expressed (fig. 3D) like the expression pattern of de novo genes in rice (Zhang et al. 2019). The tissue with the largest number of specifically expressed genes (SEGs) was the shoot with 13 genes and the second one was the inflorescence with four genes. Moreover, the shoot-specific genes PH02Gene15427, PH02Gene17673, PH02Gene28800, and PH02Gene36478 were highly expressed with FPKM > 1,000 at the stage of 1.0–7.0 m and had the mass spectrometry

(MS) proteomics peptides (supplementary tables 11 and 12, Supplementary Material online). PH02Gene36478 (fig. 3B) is an example with six peptides, supported by liquid chromatography-tandem MS (LC–MS/MS) in the shoots (supplementary fig. 9, Supplementary Material online). These results lend further support for that the new genes particularly de novo genes could be related to the fast growth of woody bamboo shoots.

In general, we found that these de novo genes and orphan genes experienced gradual evolutionary processes with their ages in terms of their exon–intron structure, protein length, and expression tissue specificity. The



**FIG. 3.** De novo gene birth of PS11 genes and their expression. (A) Origination process for the de novo gene *PH02Gene41079*. De novo creation of ORF can be one "T → G" substitute at 19 bp to remove the premature stop codon in the ancestor of woody bamboos. Its expression is detected in two woody bamboo species. The sequence alignments of *PH02Gene41079* were exhibited in [supplementary figure 6, Supplementary Material online](#). (B) Origination process for the de novo gene *PH02Gene36478*. One 8-bp deletion resolved frameshift and the following premature stop codon. The noncoding ancestral sequence in *Ol. latifolia* acquired expression. The sequence alignments of *PH02Gene36478* were exhibited in [supplementary figure 7, Supplementary Material online](#). The gap in the species represents the absence of homologous sequence from that species. (C) Origination process for the de novo gene *PH02Gene38072* via two-step process, as follows. 1) Two 1-bp frameshifts resolved the frameshift in *Or. sativa* and premature stop codon. 2) One "T to A" substitution created start codon, and "T to C" substitution resolved the premature stop codon, and one "T" insert resolved frameshift in *R. distichophylla* and *Ol. latifolia*. The sequence alignments of *PH02Gene38072* were exhibited in [supplementary figure 8, Supplementary Material online](#). (D) Expression heatmap for 19 de novo genes. Black box highlighted shoot-biased expression pattern. The expressions of the three exemplified de novo genes were shown in bold.



**Fig. 4.** Patterns of new genes in evolution, expression and gene structures. (A) Younger genes are shorter than older genes. (B) Younger genes have fewer exons than older genes. (C) Evolutionary rates of young genes are higher than old genes. (D) Younger genes are more tissue specific expression. A Kruskal–Wallis rank sum test was used to determine significance.

features of gene length, exon number, evolutionary constraint, and tissue-specific expression all showed clear age-dependent trends ( $P < 0.01$  in the Kruskal–Wallis rank sum test) (fig. 4; [supplementary table 13, Supplementary Material online](#)). In contrast to old genes, new genes tended to possess shorter CDS sequences (a median value of 303 bp for PS12 vs. 1,179 bp for PS1) (fig. 4A) and less exons (three exons of PS12 vs. five ones of PS1) (fig. 4B) and evolved with less evolutionary constraint (Ka/Ks) (0.92 vs. 0.26) (fig. 4C). The new genes also showed higher expression specificity (0.83 vs. 0.47) (fig. 4D). These observations reveal well the stepwise evolution of novel gene structures: New genes gradually recruited more exons, expanded their exon lengths, enhanced evolutionary constraint, and acquired the ability of more broad expression.

### Divergent Expression of WGD Duplicated Genes in Shoot

As the observation of shoot-specific expression of de novo genes, we further examined the SEGs at the whole genome level across the *P. edulis* tissues, which may provide clues about tissue-specific and gene-specific biological functions (Favery et al. 2001; Wagner and Lynch 2010; Borg et al. 2011). Totally, we identified 7,013 SEGs, including 574 in roots, 43 in later buds, 356 in leaves, 22 in rhizome tip, 194 in inflorescence, 42 in shoot tips, and 5,782 in shoots (table 1; [supplementary table 14, Supplementary Material online](#)). A majority of 82.45% of detected SEGs was found in shoots and concentrated in the stage of 0.5–1.0 m height with

4,227 out of 5,782 shoot-SEGs (fig. 5A). The large number of SEGs identified for shoots might be caused by the more samplings of this tissue with eight developmental stages than those of the other six tissues with each just having one stage. To assess this possibility, we performed eight simulations for identifying SEGs with each one only selecting one developmental stage of shoots ([supplementary table 15, Supplementary Material online](#)). Similar results were obtained when selecting the samples of shoots at 0.5 m (65.61% of the total 6,305 SEGs identified in shoot) and at 1.0 m (64.43% of 6,121 SEGs) height whereas there were still higher proportion of SEGs (20.35–36.23%) identified in shoots for the remaining stages despite not significantly for all of them. Therefore, we can conclude that the shoots indeed have more tissue-specific expressed genes, not only for the 19 de novo genes but also for all the genes in the genome, especially at the stages of 0.5 and 1.0 m. The shoot SEGs were enriched in functions involving the plant-type cell wall organization, biogenesis and mitochondrial mRNA processing ([supplementary fig. 10, Supplementary Material online](#)).

We also identified SEGs for rice in four tissues ([supplementary table 16, Supplementary Material online](#)) for comparison and found that the tissue having the most SEGs was the inflorescence (54.45% of a total 3,242 SEGs) ([supplementary table 17, Supplementary Material online](#)). This was common for reproductive organs to have more tissue-specific genes as found in many species of animals and plants (Wu et al. 2014; Gossmann et al. 2016; Zhang et al. 2018; Fang et al. 2020), in sharply contrast to the situation found in the woody



**Table 1.** The Number of SEGs in Seven *Phyllostachys edulis* Tissues.

PS	Root	Lateral bud	Leaf	Inflorescence	Shoot tip	Rhizome tip	Shoot	Total
PS1	163	10	143	65	10	3	1,389	1,783
PS2	135	11	88	48	15	3	1,793	2,093
PS3	81	9	71	23	3	4	609	800
PS4	120	10	47	45	8	7	1,049	1,286
PS5	16	2	0	0	1	1	138	158
PS6	36	1	0	0	1	4	389	431
PS7	6	0	0	0	0	0	43	49
PS8	7	0	0	0	1	0	123	131
PS9	2	0	1	0	0	0	24	27
PS10	2	0	1	2	0	0	35	40
PS11	2	0	1	3	2	0	36	44
PS12	4	0	4	8	1	0	154	171
Total	574	43	356	194	42	22	5,782	7,013

bamboos. Moreover, the identified 5,782 shoot-SEGs above included 5,557 old genes out of a total 49,314 (PS1–PS9) and 225 new genes out of a total 1,622 (PS10–PS12). The possibility of being specifically expressed in the shoots was slightly higher for new genes than for old genes (14.06% vs. 11.34%,  $\chi^2 = 10.318$ ,  $P = 0.0013$ ).

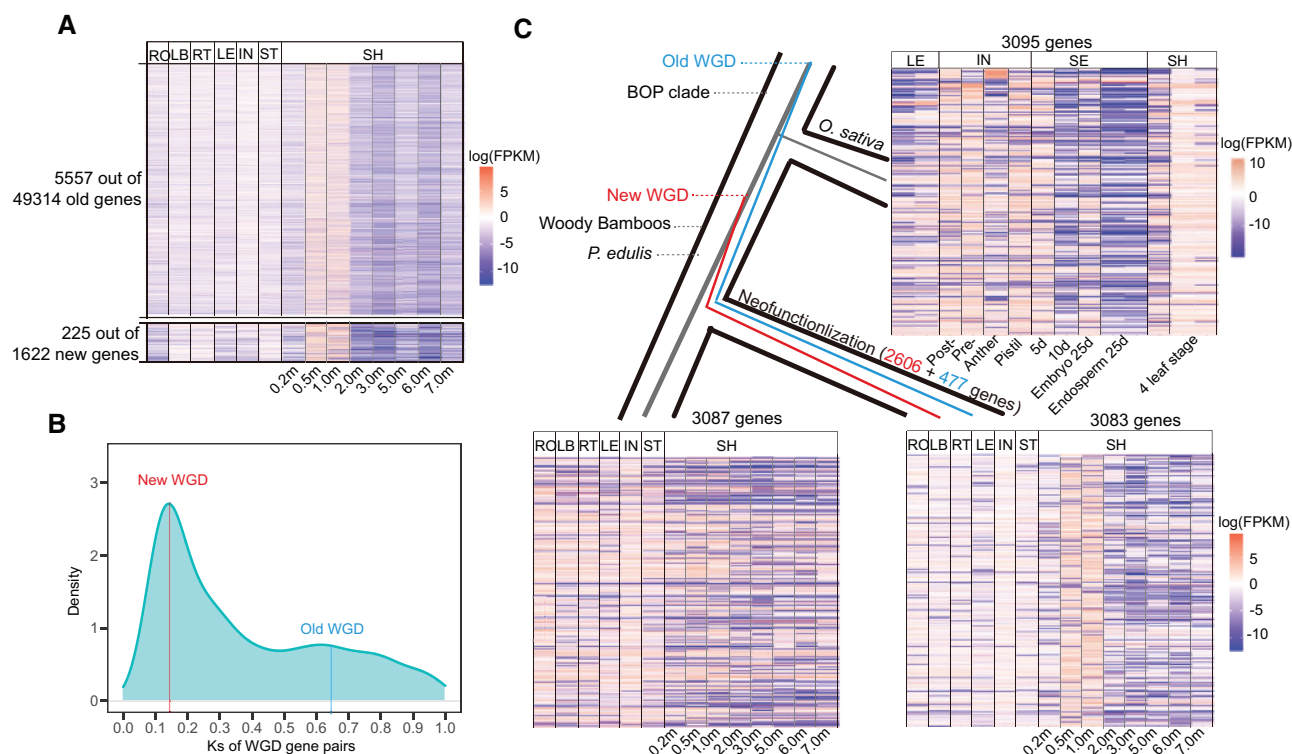
In considering the WGD history of woody bamboos (Peng, Lu, et al. 2013; Guo et al. 2019), we implemented the Dup\_GenFinder pipeline (Qiao et al. 2019) to infer the origin of shoot-SEGs. We found that 3,083 genes (53.32% out of 5,782 shoot SEGs) were derived from the WGD events with paralogs in the genome, and 28.09% of them had all copies whereas the remaining had only one copy specifically expressed. The WGD genes were significantly enriched in the shoot-SEGs ( $P < 2.2e-16$  of Fisher's exact test). Among them, 2,606 genes could be attributed to the recent independent WGD event of woody bamboos at 22 Ma (Peng, Lu, et al. 2013; Guo et al. 2019) and the remaining 477 ones to the old  $\rho$  WGD event shared by the grass family (fig. 5B) (Ma et al. 2021). We further investigated the evolution of gene expression for these WGD-duplicated genes by assuming that the expression patterns in the rice without recent WGD represented the ancestral state. Based on the collinearity between the genomes of *P. edulis* and rice, 3,095 orthologous genes were identified in the rice, which corresponded to shoot-biased WGDs. There were only 457 tissue-specific genes whereas 2,638 genes were broadly expressed across the different tissues (fig. 5C). Moreover, 322 of these 457 tissue-specific genes came from the inflorescence tissues as expected. On the other hand, nearly all of the 3,087 paralogous genes to the shoot-SEGs within *P. edulis* were also broadly expressed with only 49 of them being tissue-specific. For the 3,083 shoot-biased expression WGD duplicates, we only detected 10 shoot-SEGs with Ka/Ks value more than 1 ( $(\text{Ka/Ks} \geq 1) = 10/3,083 = 0.0032$ ), suggesting that obtaining a duplicate pseudogene is an event with a significantly small probability (supplementary fig. 11 and table 18, Supplementary Material online). This suggests that most gene pairs were underlying purifying selection, therefore their translated products are not functionless. Together, these results suggested that the divergence of gene expressions for the WGD duplicated genes in woody bamboos were accompanied by one of duplicated copies being specifically expressed in the shoot.

Evolution of New Genes by Coexpression with Old Genes

To acquire their new functional roles, new genes need to be integrated into the gene interaction network, which can be inferred from gene expression data (Zhang et al. 2015). Using the weighted gene coexpression network analysis for the samples of shoots at eight different developmental stages and the other six tissues (Langfelder and Horvath 2008), we estimated correlations between genes across transcriptome samples and clusters genes with similar profiles into modules. We clustered 12,728 genes into 13 modules (black, blue, green, tan, pink, red, purple, yellow, turquoise, greenyellow, brown, magenta, and gray modules) ranging in size from 39 to 4,041 genes (fig. 6A; supplementary table 19 and fig. 12, Supplementary Material online). Among these modules, six of which (black, blue, green, tan, purple, and yellow modules) were found to be related to the shoots (supplementary figs 13 and 14, Supplementary Material online). These shoot-specific networks included from 169 to 2,244 genes and were generally enriched for genes belonging to different synthetic and catabolic pathways, including RNA biosynthetic and metabolic process and response to stimulus at 0.2 m, chromatin organization at 0.5 m, transport and translational elongation at 1.0 m, cell wall organization at 2.0 and 3.0 m, and biogenesis from 5.0 to 7.0 m (supplementary fig. 15, Supplementary Material online).

All the 13 modules except gray module harbored both old genes and new genes (fig. 6A), whereas the five modules (black, blue, green, purple, and yellow) with enrichment of WGD-duplicated genes with divergent expression patterns were all shoot-specific ( $P$  value of the Fisher's Exact Test  $< 0.01$ ). Even the black module at the stages of 2.0 and 3.0 m in height was meanwhile enriched for de novo genes ( $P < 0.01$  in Fisher's Exact Test). These results suggested that new genes were coexpressed with a large number of old genes in the shoot, presumably recruited into the networks through WGD-duplicated genes showing shift of expression toward shoots. Taking the module enriched for de novo genes as an example for more details, we found that *PH02Gene28800* was one of the three de novo genes in the black module originating in the common ancestor of woody bamboos. According to the weighted values connected to this gene, we selected 50





**FIG. 5.** Shoot SEGs and their evolutionary model. (A) SEGs of shoots in *P. edulis*. RO, root; LB, leaf; RT, rhizome tip; LB, lateral bud; IN, inflorescence; ST, shoot tip; SH, shoot. (B) Ks distribution of shoot biased related WGD-pairs in *P. edulis*. (C) Evolution of expression pattern of SBW. SBW, shoot biased WGDs.

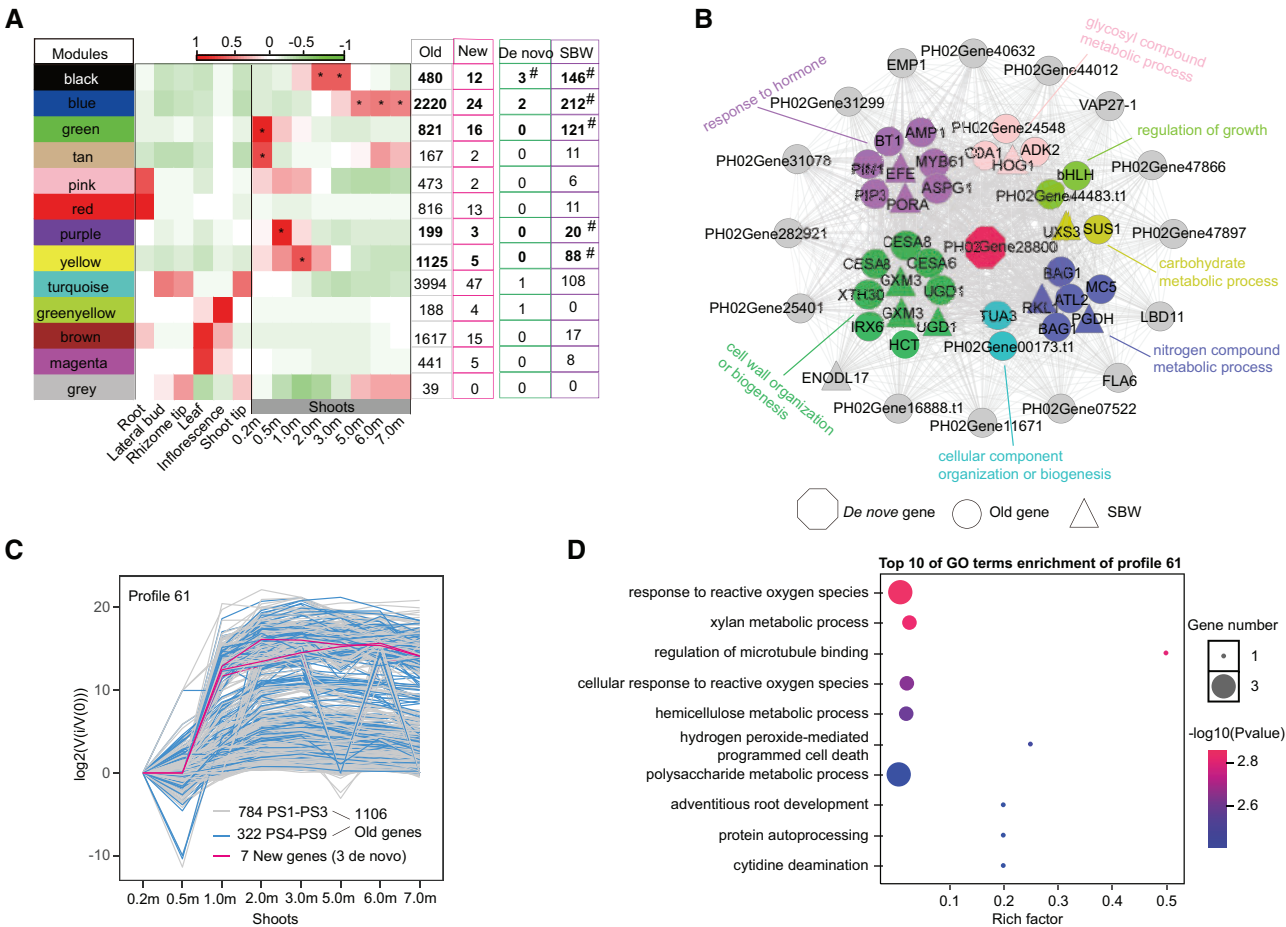
genes to display the network (fig. 6B). Among these genes, ten genes were WGD-duplicated genes specifically expressed in the shoot and the remaining 40 ones were evolutionarily conserved old genes. The functional enrichments of these genes were cell wall organization or biogenesis (ten genes), response to hormone (eight genes), glycosyl compound metabolic process (four genes), and regulation of growth (two genes) (supplementary table 20, Supplementary Material online). These pathways were all previously reported to be involved in the fast growth of woody bamboo shoots (Paque et al. 2014; Li et al. 2018), underlying how the new genes to function in the fast growth of bamboo shoot by interacting with the existed genes.

To further check whether the eight different developmental stages of shoots have similar trends of coexpression for new and old genes, we clustered expression profiles of shoots by Short Time Series Expression Miner, which was designed to analyze time series with three to eight points (Ernst and Bar-Joseph 2006). In total, 13,001 genes were clustered in to 19 expression patterns ( $P < 0.05$ ), and 12 of which were comprised of both old and new ones (supplementary fig. 16, Supplementary Material online), showing a general similar coexpression pattern of new genes and old genes among different stages of shoots. Among them, Profile 61 harbored 7 new genes and 1,106 old genes (784 from PS1–PS3 and 322 from PS4–PS9) with overexpression at the 2.0 and 3.0 m stages (fig. 6C). Within the seven new genes of Profile 61, there were three de novo genes (PH02Gene15427, PH02Gene17673, and PH02Gene28800). GO function enrichments of Profile 61 were mainly involving the xylan metabolic

process and hemicellulose metabolic process, both of which were cell wall related (fig. 6D). The cell wall metabolism was involved in the fast growth of woody bamboo shoots (Cui et al. 2012; He et al. 2013; Peng, Zhang, et al. 2013; Wang et al. 2019), further suggesting the important roles of the evolution of coexpression for new genes and old genes in the emergence of this unique trait of woody bamboos.

## Discussion

Evolutionary new genes have been seen as one of major drivers in phenotypic evolution (Khalturin et al. 2009; Kaessmann 2010; Tautz and Domazet-Loso 2011; Chen et al. 2013). The fast growth of woody shoot and consequent tall stems in bamboos, a key innovation and distinguishing trait within the grass family (Song et al. 2011; Lima et al. 2012; Buckingham et al. 2014), provides an opportunity to explore the role of new genes in the morphological diversification of the grass species. A novel cell type called long parenchyma cells was even evolved in the woody bamboo shoot (He et al. 2002; Gritsch and Murphy 2005; Wei et al. 2018). With definite phenotypic innovations and recent WGD events (Guo et al. 2019), which provide major sources of new genes (McLysaght et al. 2002; Long et al. 2003; Kellis et al. 2004; Edger et al. 2015; Clark and Donoghue 2018), the bamboos can be an ideal system to study the connection between genomic novelty and evolutionary innovation. Nevertheless, to date, a comprehensive investigation of the genomic basis underlying the evolution of fast growth of woody shoot in bamboos is still lacking with existing studies mostly focusing on a few genes and their associated gene families (Peng, Lu, et



**FIG. 6.** Coexpression of new genes and old genes. (A) Gene coexpression network identified by WGCNA. Thirteen coexpression modules were detected, including black, blue, green, tan, pink, red, purple, yellow, turquoise, greenyellow, brown, magenta, and gray modules. The “\*” represent modules correlated with shoots when the correlation value is more than 0.50 and  $P$  value  $< 0.01$ . The “#” represent de novo genes and SBW genes significant enrichment in module when the  $P$  value of Fisher’s exact test  $< 0.01$ . (B) A de novo gene *PH02Gene28800* centric coexpression networks. The top 50 correlated genes were shown in the network. Colors refer to GO enrichment terms. Gray refers to genes without enriched in any GO terms. The shapes refer to gene compositions. (C) Profile 61 showing high overexpression at 1.0–3.0 m stages of shoot by STEM analysis. Colors refer to gene ages. (D) The GO enrichment of the Profile 61.

al. 2013; Peng, Zhang, et al. 2013; Wei et al. 2018, 2019). By collecting genomic and transcriptome data of bamboos, our phylostratigraphic and transcriptomic analyses here have enabled us to acquire a systematic view of bamboo shoot evolution at the genome level.

With sequenced genomes representing the four major lineages of Bambusoideae (Guo et al. 2019), a total of 1,622 orphan genes were identified in bamboos by using phylostratigraphy, with the highest birth rate for new genes coinciding with the origin of this subfamily. By linking gene ages to expression, both higher TAI and TDI values suggest that the transcriptomes of shoots are young and divergent in evolution among the various bamboo tissues. These include vegetative organs such as leaf and root, as well as reproductive tissue of inflorescence. Moreover, the new genes originating in the common ancestor of woody bamboos (PS11) show the highest expression level among genes of different phylostrata and in the shoot at the developmental stage of 2.0 m. This is in sharp contrast to that the new genes are generally lower expressed than old genes (Donoghue et al. 2011; Palmieri et al.

2014; Zhao et al. 2014), which contains many housekeeping genes, although the expression level of old genes is higher than that of new genes in other tissues of bamboos. In addition to young transcriptome, a large number of genes are specifically and highly expressed in the shoots, particularly at the stages of 0.5 and 1.0 m. The majority of these shoot biased expressed genes are derived from the duplicated genes of WGD with one copy experiencing the divergent expression patterns from the ancestral state in the shoot.

Furthermore, within the origin mechanisms of new genes the de novo origin is probably the most exciting and these genes may play a key role in the evolution of innovative trait (Van Oss and Carvunis 2019; Blevins et al. 2021). To avoid potential sources of orphan genes alternative to de novo origination in the phylostratigraphic analysis (Moyers and Zhang 2016; Domazet-Lošo et al. 2017; Vakirlis et al. 2020), we used a strict procedure of de novo gene identification that requires reconstruction of noncoding ancestors and processes of de novo origination (Zhang et al. 2019). We identified 19 candidate de novo genes in the woody bamboos with

significantly similar ancestral noncoding sequences and reconstructed the origination processes of these de novo genes. As expected, these genes are all specifically expressed in individual tissues as suggested for de novo genes previously (Levine et al. 2006; Toll-Riera et al. 2009; Schlotterer 2015). Interestingly, most of them and 13 genes were found to be expressed in the shoots at developmental stages from 0.5 to 2.0 m and the following tissue was the inflorescence but with only four de novo genes expressed. Altogether, we can conclude that the evolution of woody bamboo shoot closely correlated with genomic novelties from highly expressed orphan genes (including de novo genes) to altered expression patterns for WGD duplicated genes. Moreover, these events mainly occur in the stages of 0.5–2.0 m, a key transition point for *P. edulis* from slow to fast growth (Xu et al. 2011; Song et al. 2016). As such, the development of woody bamboo shoot shows an “inverse hourglass” model for the evolutionary age of the transcriptome with old genes expressed at early and late stages. This is in contrast to the common hourglass model as found firstly for embryogenesis (Domazet-Lošo and Tautz 2010; Quint et al. 2012; Levin et al. 2016), and afterwards many postembryonic phases of plant development, with a phylotypic stage expressing the oldest transcriptome set (Drost et al. 2016; Leiboff and Hake 2019; Tribble and Kronauer 2021). The transition stage from slow to fast growth may reflect the most unique features of shoot and thus the morphological and molecular patterns are coupled in the woody bamboos.

The observation of shoot serving as a major source of genomic novelties in woody bamboos is intriguing. A theory of “out-of-testis” for the emergence of new genes was firstly proposed in animals (Levine et al. 2006; Vinckenbosch et al. 2006; Kaessmann 2010). Biased expression of new genes in male reproductive tissues has also been identified afterwards in the model plants of *Arabidopsis* and rice (Cui et al. 2015), as well as recently in nematodes (Rodelsperger et al. 2021). Several hypotheses have been put forward to explain this phenomenon and may be driven by a common evolutionary force of male gametophyte competition (Cui et al. 2015; Wang et al. 2016). However, in the woody bamboos, the shoot rather than the reproductive tissues acts as an “innovation incubator” for the evolution of new genes. A potential explanation is that, as polyploid plants, the woody bamboos often reproduce vegetatively with the complex system of rhizomes and shoots playing an important role in propagation, and use sexual reproduction occasionally with long flowering cycles of 20–60 years in general (Janzen 1976).

With an enrichment of genomic novelties in the growth of woody bamboo shoot revealed, these novelties represent a mix of products of new genes as well as WGD duplicated genes undergoing divergent expression. Gene duplication through WGD is a major source for generating new genes and function (Conant and Wolfe 2008; Innan and Kondrashov 2010; Zhen et al. 2012; Sandve et al. 2018), but other process such as de novo formation can also generate new genes (Zhang et al. 2019; Blevins et al. 2021). And our identified new genes here are all resulted from other processes rather than WGD duplication. In particular, there are 19 genes

of de novo origin, the first case described in bamboos, and the majority of them are related to the evolution of woody bamboo shoot. On the other hand, the WGD duplicated genes are also involved in this innovation of woody bamboos with rapid divergence of expression in the shoot for one copy, pointing to potential functional novelty (Taylor and Raes 2004; Braasch et al. 2016; Sandve et al. 2018). In all, the combination of new genes and WGD duplicated genes forms the molecular basis of the innovation of fast woody shoot growth in bamboos. This extends our understanding of novel gene and function formation in bamboos and will provide an important resource for future studies of gene function.

Furthermore, we demonstrated the gene interaction between new genes and old genes by the WGCNA analysis, and many of the identified coexpression modules are enriched in genes involving the process of cell wall metabolism, cell cycle regulation, and plant hormones network. All of these processes are closely related to the rapid growth of woody bamboo shoot (Cui et al. 2012; He et al. 2013; Peng, Zhang, et al. 2013; Wang et al. 2019). These results further validate the role of new genes through recruitment into the existing genetic networks in the evolution of woody bamboo shoot. This is somewhat expected as previous studies have shown that new genes should be integrated into and reshape ancestral genetic interaction network to acquire their corresponding biological functions (Zu et al. 2019). In future, more detailed analysis and functional studies of identified new genes and their interacted genes are needed to comprehensively characterize the roles of genomic novelties in the woody bamboo shoots.

## Materials and Methods

### RNA Extraction and Full-Length RNA Sequencing

The samples (leaves and shoots) of *Bo. amplexicaulis* and *Ol. latifolia* were collected from the same plant for genome sequencing in Guo et al. (2019). Two total RNA samples (leaves and shoots) were extracted using the miRcute Plant miRNA Isolation Kit (DP504) (TIANGEN BIOTECH Corporation, Beijing, China). Total RNA samples were treated with Dnase I to remove DNA contaminant. RNA quantity was determined using Nanodrop, gel electrophoresis, and further by the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA). The first-strand cDNA was synthesized using the Clontech SMARTer PCR cDNA synthesis Kit (Clontech). The library was prepared according to the Isoform Sequencing (Iso-Seq) protocol, as described by Pacific Biosciences (PacBio). A total of two single molecular real-time (SMRT) cells were sequenced on the PacBio Sequel platform (Biomarker Tech, Beijing, China). The cDNA library (1–6 kb) was constructed and sequenced using the PacBio Sequel System. Two SMRT cells generated 46.9 Gb raw data.

According to the standard protocol of ISO-seq (SMRT Analysis 5.1.0), raw Polymerase reads that have full passes  $\geq 1$  and the predicted consensus accuracy  $>0.9$  were selected (minPasses = 1) (<https://github.com/PacificBiosciences/pbbioconda>, last accessed October 1, 2020). Then, 941,750 circular consensus sequences (CCSs) were classified into full-length (FL) and non-FL (NFL) CCS according to whether



included 5'/3' cDNA primers and poly (A) tail at the same time. At last, full-length nonchimeric (FLNC) reads were subjected to isoform-level clustering by iterative isoform-clustering for error correction algorithm and herein similar sequences assigned to a cluster. Each cluster was identified as a uniform isoform. NFL cDNA reads were then applied to polish each cluster to produce high-quality isoforms (accuracy > 99%). The difference of 5' end was not considered when collapsing redundant transcripts. The FLNC sequences were mapped into the genome by Gmap (–cross-species, –allow-close-indels 0) (Wu and Watanabe 2005). Mapped reads were further collapsed by cDNA\_Cupcake package ([https://github.com/Magdoll/cDNA\\_Cupcake/wiki](https://github.com/Magdoll/cDNA_Cupcake/wiki), last accessed June 15, 2021) with >85% alignment coverage and 90% alignment identify. The nonredundant FLNC reads were listed in [supplementary table 21, Supplementary Material online](#) and the raw data are available at the NCBI SRA archive (PRJNA764002).

### Phylostratigraphic Analysis of *P. edulis* and *Bo. amplexicaulis*

Gene age was estimated using the genomic phylostratigraphic approach as described previously (Domazet-Loso et al. 2007). Considering the polyploid nature of woody bamboos (*P. edulis*:  $2n = 4\times$ ; *Bo. amplexicaulis*:  $2n = 6\times$ ) (Zhao et al. 2018; Guo et al. 2019), the paralogous genes of *P. edulis* and *Bo. amplexicaulis* were gathered into clusters using OrthoMCL software (Li et al. 2003) with the parameter of *l* at 1.5, respectively ([supplementary fig. 1, Supplementary Material online](#)). Only the longest protein of each cluster was picked up as its representative. Combined representatives of gene clusters with singleton (without paralogs), we generated nonredundant queries for *P. edulis* and *Bo. amplexicaulis*, respectively. Additionally, we download 69 genomes across the Tree of Life to represent 12 evolutionary levels or phylostrata (PS), starting from the origin of cellular organisms (PS1) and ending at the origin of *P. edulis* (PS12) ([supplementary table 1, Supplementary Material online](#)).

To evaluate the impact of full-length transcriptome data on gene age analysis, we set up two reference data sets: 1) 69 genomes (data set 1); 2) 69 genomes combined with full-length isoforms of *Bo. amplexicaulis* and *Ol. latifolia* ([supplementary table 22, Supplementary Material online](#)). The addition of full-length transcriptome data mainly affected the number of young genes. Finally, we compared redundant query sequences of *P. edulis* and *Bo. amplexicaulis* with other 68 genomes and full-length transcriptome data via BLASTp and TBlastN algorithm with an e-value  $10^{-4}$  threshold, and then assigned all protein-coding genes into 12 phylostrata. The methods and parameters were consistent with previous studies (Domazet-Loso et al. 2007). According to phylostratigraphic method, if a putative homolog of one gene was first identified in one phylostratum, it was assumed as the age of that gene. If no BLAST hit was detected in other phylostrata, the corresponding protein was assigned to the youngest PS (PS12). Gene cluster fixed rates were calculated according to the formula,  $r = N/T$ : *r* representing the number of clusters per My; *N* representing the number of clusters in the

phylostratum; *T* being the duration of the interval My in the phylostratum. Based on the cluster results of OrthoMCL, gene copies of each cluster were classified into one of the following three groups by its family size in the clustering according to Guo (2013): 1) A singleton is a single-copy gene; 2) a two-gene family has two copies; and 3) a multigene family has  $\geq 3$  copies.

### Gene Expression Quantification and Proteomic Peptides Identification

We obtained published RNA-seq data for *P. edulis* from NCBI SRA databases (Gao et al. 2014; Zhao et al. 2016; Wang et al. 2017, 2019) ([supplementary table 4, Supplementary Material online](#)). Briefly, we trimmed reads using the Trimmomatic (Bolger et al. 2014) after analysis in FASTQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, last accessed July 14, 2021), and then computed expression level for each gene using RSEM v1.2.16 (Li and Dewey 2011) with Bowtie2 aligner (Langmead and Salzberg 2012). Read counts of gene expression data were removed batch effect by the removeBatchEffect function in Limma package (Ritchie et al. 2015).

We also downloaded transcriptome data of *D. sinicus*, another tropical woody bamboo species, under the accession of PRJNA418355 (Chen et al. 2018). Because without reference genome of *D. sinicus*, the clean reads were assembled into contigs using the program Trinity v2.4.0 (Grabherr et al. 2011). The transcript abundance was quantified using software RSEM v1.2.16 software (Li and Dewey 2011) with Bowtie2 aligner (Langmead and Salzberg 2012). The protein of *D. sinicus* transcripts was predicted using TransDecoder (<http://transdecoder.sourceforge.net/>, last accessed July 30, 2021).

The available MS raw data of leaf and seeding (Yu et al. 2019), and shoot (Tao et al. 2020) were download. MaxQuant 1.6.17.0 (Tyanova et al. 2016) was used to perform a database search against the moso bamboo protein database to identify peptides.

### Transcriptome Index Calculation

TAI and TDI of each tissue/developmental stage were calculated via myTAI package (Drost et al. 2018). The TAI is a measure that reflects the evolutionary age of a transcriptome at a given ontogenetic stage, where higher values correspond to younger transcriptomes (Domazet-Loso and Tautz 2010). TAI value of a sample is defined as the weighted mean of phylostratum rank  $ps_i$  of gene *i* by the expression value  $e_{is}$  in the transcriptome of sample *s*,

$$TAIs = \frac{\sum_{i=1}^n e_{is} \cdot ps_i}{\sum_{i=1}^n e_{is}}$$

where *n* is the total number of genes in the analysis. High TAI indicates that the transcriptome is evolutionarily young, whereas low TAI value suggests the transcriptome is evolutionarily ancient.

The TDI is a weighted mean of Ka/Ks ratios of gene *i* by the expression value  $e_{is}$  in the transcriptome of samples (Quint et al. 2012),



$$\text{TDIs} = \frac{\sum_{i=1}^n \left( \frac{K_{a_i}}{K_{s_i}} \right) \cdot e_{is}}{\sum_{i=1}^n e_{is}}$$

High TDI indicates that the transcriptome is more divergent, whereas low TDI value represents a more conserved transcriptome.

Relative expression (RE) of genes for a given phylostratum (ps) and developmental stage (s) was computed using methods described by Domazet-Loso and Tautz (2010):

$$\text{RE}(\text{ps})_s = \frac{\bar{f} - \bar{f}_{\min}}{\bar{f}_{\max} - \bar{f}_{\min}}$$

where  $\bar{f}$  is the average partial concentration of RNAs from phylostratum for a given stage and  $\bar{f}_{\max}$ ,  $\bar{f}_{\min}$  are the maximal and minimal average partial concentration from phylostratum across all considered stages, respectively.

### EP of Highly Expressed Genes

In each tissue, we extracted the 1,000 highest expressed gene of each tissue as the highly expression set (supplementary table 8, Supplementary Material online). Gene age composition of the highly expressed set was counted. The expression ratio (Exp) of a given phylostratum (PS) was calculated in each tissue,

$$\text{Exp}(\text{ps})_s = \frac{\sum_{i=1}^n f}{n(\text{ps})} * 100$$

where  $f$  represented the sum of expression value for a given PS, and  $n$  was the gene number of a given PS. To assess the expression pattern of genes from different ages across different tissues, we calculated the EP of PS in each tissue,

$$\text{EP}(\text{ps})_s = \frac{\text{Exp}(\text{ps})}{\left( \frac{n}{1000} \right)}$$

### Detection of Woody Bamboo De Novo ORFs

Following de novo detective pipeline from Zhang et al. (2019), we identified de novo genes from PS11 genes in *P. edulis* which lack similarity protein sequences out of bamboos. We used nucleotide-to-nucleotide BLAT (minIdentity = 80) to align these 237 *P. edulis* PS11 orphan ORFs to other five related genomes, including other woody bamboo *Bo. amplexicaulis* (Guo et al. 2019), and four outgroup species *Ol. latifolia* (Guo et al. 2019), *R. distichophylla* (Li et al. 2021), *Br. distachyon*, and rice. If exact nucleotide matches covered at least 20% of the corresponding *P. edulis* ORFs, one effective hit was accepted. Only *P. edulis* ORFs had no more than one effective hit in each outgroup species were retained for subsequent analyses. Moreover, we further used BLAT to align these orthologous sequences to *P. edulis* ORFs, to retrieve highly similar orthologous sequences. Some *P. edulis* ORFs were identified as woody bamboo de novo ORFs that had orthologous coding sequences in the *P. edulis* and *Bo. amplexicaulis*, and had noncoding sequences in outgroup species. We used PAML (Yang 2007) to detect the signals ( $\omega = K_a/K_s$ ) of natural selection in de novo orthologous gene pairs between

*P. edulis* and *Bo. amplexicaulis* with runmode = -2 for ML pairwise comparison. To further test whether the  $\omega$  ratio of a model significantly deviated from neutral evolution ( $\omega = 1$ ), we incorporated the neutral model, which estimates model parameters by fixing  $\omega = 1$  (Yang 1997). The statistical significances between the estimated  $\omega$  and  $\omega = 1$  models were estimated by calculating twice the log-likelihood difference following a  $\chi^2$  distribution.

To better understand whether the genes across different phylostrata possess various generic features, we analyzed four important genic characteristics. Using custom Perl scripts, we calculated protein length, intron number, and GC content at the CDS level. The  $K_a/K_s$  ratio for *P. edulis* versus *Bo. amplexicaulis* was calculated using ParaAT v.2.0 (Zhang et al. 2012) and KAKS\_Calculator v.2.0 (Zhang et al. 2006). We calculated the tissue specificity score (tau) of a gene (Yanai et al. 2005; Kryuchkova-Mostacci and Robinson-Rechavi 2017) using the roonysgalbi/tispec package in R package (https://rdr.io/github/roonysgalbi/tispec, last accessed March 30, 2021).

### Identification Shoot Biased WGDs in *P. edulis*

We used R package “SEGtool” to identify SEGs in each tissue (default parameters,  $P$  value < 0.05) (Zhang et al. 2018). Although SEGtool can detect genes with specifically high and low expression, we only used the SEGs with high expression in this study for further investigation. The number of developmental stages in a certain tissue may affect SEG number. Because the shoot stage contains eight developmental stages, and the other tissues only contain one developmental stage, for fair comparison, we only selected one shoot developmental stage when identifying SEGs for each tissue. We also download rice expression data from http://rice.plantbiology.msu.edu/expression.shtml (last accessed January 15, 2021) to compare SEG patterns with *P. edulis*. Mcscanx (Wang, Tang, et al. 2012) was used to identify syntenic gene pairs between *P. edulis* and rice.

Gene duplications were classified within the *P. edulis* genome using the Dup\_GenFinder pipeline (Qiao et al. 2019). This pipeline used the results of an all-to-all BLASTp of the *P. edulis* gene set to itself, and BLASTp results comparing the *P. edulis* gene set to a closely related species rice. Dup\_GenFinder synthesizes the outputs of these analyses to classify gene duplications into one of five categories, including whole-genome duplication (WGD), tandem duplications, proximal duplications, transposed duplications, and dispersed duplications. Based on the Dup\_GenFinder results, the shoot SEGs that match the WGD gene pair were called as shoot biased WGDs (SBW). Then, we calculated the  $K_a/K_s$  value for all duplication gene pairs, by using ParaAT v.2.0 (Zhang et al. 2012) and KAKS\_Calculator v.2.0 (Zhang et al. 2006).

### Gene Coexpression Analysis by WGCNA and STEM

The R package, WGCNA, was used to perform the weighted correlation network analysis for all 14 ontogenetic stages in R 3.5.3 with the parameters of “softPower = 10, and minModuleSize = 100” (Langfelder and Horvath 2008). Additionally, we extracted the corresponding gene

information for each module for further analysis. After the modules were identified, the module eigengene (ME) was summarized by the first principal component of the module expression levels. To construct the developmental stage-specific co-expression network, we regarded sampled tissues and developmental stages as trait. Module-trait relationships were estimated using the correlation between MEs and traits, which allowed efficient identification of the relevant modules. To evaluate the correlation strength, we calculated the module significance (MS), which is defined as the average absolute gene significance (GS) of all the genes involved in the module. The GS is measured as the  $\log_{10}$  transformation of the  $P$  value ( $\log P$ ) in the linear regression between gene expression and trait information. In the WGCNA, modules with the highest MS score among all modules are usually defined as the key module and selected for further analysis (Langfelder and Horvath 2008; Wright et al. 2015). Cytoscape v3.6.1 was used to visualize coexpression network (Shannon et al. 2003).

Genes were clustered into co-expression profiles based on their expression dynamics by implementing clustering method STEM v. 1.3.11 (Maximum number of model profiles = 90) (Ernst et al. 2005; Ernst and Bar-Joseph 2006). STEM software uses a nonparametric clustering method to assign genes to predefined expression profiles. It considers expression profiles to be significant if the number of genes assigned to a cluster departs from random. GO enrichment analysis was performed using the clusterProfiler R package (version 3.18.1) (Yu et al. 2012).

## Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

## Acknowledgments

We thank Liangzhong Niu and Jingxia Liu for assistance for field sampling, and thank Rong Leng for assistance for collective published data, and thank Zhenhua Guo and Yupeng Cun for helpful discussions, and thank Jinwen Luo for the drawings of grass and woody bamboo. This work was supported by the CAS Strategic Priority Research Program (Grant Number 31000000 to D.-Z.L.), and the National Natural Science Foundation of China (Grant Number 31571311 to C.Z.), and the National Science Foundation (NSF 2020667 to M.L.), and Youth Innovation Promotion Association of Chinese Academy of Sciences (No. Y201972 to P.-F.M.).

## Author Contributions

D.-Z.L., C.Z., and M.L. designed the research. G.-H. J. and P.-F. M. planned and carried out analysis with the help of X.-P.W. and L.-F.G., G.-H.J., and P.-F. M wrote the manuscript with the help of M.L., D.-Z.L., and C.Z.

## Data Availability

Raw data for *Bo. amplexicaulis* and *Ol. latifolia* are available at NCBI sequencing read archive (BioProject ID: PRJNA764002).

## References

- Albert VA, Oppenheimer DG, Lindqvist C. 2002. Pleiotropy, redundancy and the evolution of flowers. *Trends Plant Sci.* 7(7):297–301.
- Averof M, Cohen SM. 1997. Evolutionary origin of insect wings from ancestral gills. *Nature* 385(6617):627–630.
- Blevins WR, Ruiz-Orera J, Messegueur X, Blasco-Moreno B, Villanueva-Canas JL, Espinar L, Diez J, Carey LB, Alba MM. 2021. Uncovering *de novo* gene birth in yeast using deep transcriptomics. *Nat Commun.* 12(1):604.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.
- Borg M, Brownfield L, Khatib H, Sidorova A, Lingaya M, Twell D. 2011. The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in *Arabidopsis*. *Plant Cell* 23(2):534–549.
- Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores A, Desvignes T, Batzel P, Catchen J, et al. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet.* 48(4):427–437.
- Buckingham KC, Wu LR, Lou YP. 2014. Can't see the (bamboo) forest for the trees: examining bamboo's fit within international forestry institutions. *Ambio* 43(6):770–778.
- Carvunis AR, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, Simonis N, Charleaux B, Hidalgo CA, Barrette J, Santhanam B, et al. 2012. Proto-genes and *de novo* gene birth. *Nature* 487(7407):370–374.
- Chang ST, Wu JH. 2000. Green-color conservation of ma bamboo (*Dendrocalamus latiflorus*) treated with chromium-based reagents. *J Wood Sci.* 46(1):40–44.
- Chen LN, Guo XJ, Cui YZ, Zheng XG, Yang HQ. 2018. Comparative transcriptome analysis reveals hormone signaling genes involved in the launch of culm-shape differentiation in *Dendrocalamus sinicus*. *Genes* 9:4.
- Chen SD, Krinsky BH, Long MY. 2013. New genes as drivers of phenotypic evolution. *Nat Rev Genet.* 14(9):645–660.
- Chen SD, Yang HW, Krinsky BH, Zhang A, Long MY. 2011. Roles of young serine-endopeptidase genes in survival and reproduction revealed rapid evolution of phenotypic effects at adult stages. *Fly (Austin)* 5(4):345–351.
- Clark JW, Donoghue PCJ. 2018. Whole-genome duplication and plant macroevolution. *Trends Plant Sci.* 23(10):933–945.
- Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet.* 9(12):938–950.
- Cui K, He CY, Zhang JG, Duan AG, Zeng YF. 2012. Temporal and spatial profiling of internode elongation-associated protein expression in rapidly growing culms of bamboo. *J Proteome Res.* 11(4):2492–2507.
- Cui X, Lv Y, Chen ML, Nikoloski Z, Twell D, Zhang DB. 2015. Young genes out of the male: an insight from evolutionary age analysis of the pollen transcriptome. *Mol Plant.* 8(6):935–945.
- Dai HZ, Yoshimatsu TF, Long MY. 2006. Retrogene movement within- and between-chromosomes in the evolution of *Drosophila* genomes. *Gene* 385:96–102.
- Domazet-Loso T, Brajkovic J, Tautz D. 2007. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet* 23(11):533–539.
- Domazet-Loso T, Carvunis AR, Alba MM, Sestak MS, Bakaric R, Neme R, Tautz D. 2017. No evidence for phylostratigraphic bias impacting inferences on patterns of gene emergence and evolution. *Mol Biol Evol* 34(4):843–856.
- Domazet-Loso T, Tautz D. 2010. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* 468(7325):815–818.
- Donoghue MTA, Keshavaiah C, Swamidatta SH, Spillane C. 2011. Evolutionary origins of Brassicaceae specific genes in *Arabidopsis thaliana*. *BMC Evol Biol.* 11:47.
- Drost HG, Bellstadt J, O'Maileidigh DS, Silva AT, Gabel A, Weinholdt C, Ryan PT, Dekkers BJ, Bentsink L, Hilhorst HW, et al. 2016. Post-

- embryonic hourglass patterns mark ontogenetic transitions in plant development. *Mol Biol Evol.* 33(5):1158–1163.
- Drost HG, Gabel A, Liu JL, Quint M, Grosse I. 2018. myTAI: evolutionary transcriptomics with R. *Bioinformatics* 34(9):1589–1590.
- Edger PP, Heide-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, Heckel DG, Der JP, Wafula EK, Tang M, et al. 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proc Natl Acad Sci U S A.* 112(27):8362–8366.
- Ernst J, Bar-Joseph Z. 2006. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* 7:191.
- Ernst J, Nau GJ, Bar-Joseph Z. 2005. Clustering short time series gene expression data. *Bioinformatics* 21(Suppl 1):i159–i168.
- Erwin DH. 2021. A conceptual framework of evolutionary novelty and innovation. *Biol Rev Camb Philos Soc.* 96(1):1–15.
- Erwin DH. 2015. Novelty and innovation in the history of life. *Curr Biol.* 25(19):R930–R940.
- Fang LZ, Cai WT, Liu SL, Canela-Xandri O, Gao YH, Jiang JC, Rawlik K, Li BJ, Schroeder SG, Rosen BD, et al. 2020. Comprehensive analyses of 723 transcriptomes enhance genetic and biological interpretations for complex traits in cattle. *Genome Res.* 30(5):790–801.
- Favery B, Ryan E, Foreman J, Linstead P, Boudonck K, Steer M, Shaw P, Dolan L. 2001. KOJAK encodes a cellulose synthase-like protein required for root hair cell morphogenesis in Arabidopsis. *Genes Dev.* 15(1):79–89.
- Gao J, Zhang Y, Zhang CL, Qi FY, Li XP, Mu SH, Peng ZH. 2014. Characterization of the floral transcriptome of moso bamboo (*Phyllostachys edulis*) at different flowering developmental stages by transcriptome sequencing and RNA-Seq analysis. *Plos One* 9(6):e98910.
- Gossmann TI, Saleh D, Schmid MW, Spence MA, Schmid KJ. 2016. Transcriptomes of plant gametophytes have a higher proportion of rapidly evolving and young genes than sporophytes. *Mol Biol Evol.* 33(7):1669–1678.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29(7):644–652.
- Gritsch CS, Murphy RJ. 2005. Ultrastructure of fibre and parenchyma cell walls during early stages of culm development in *Dendrocalamus asper*. *Ann Bot.* 95(4):619–629.
- Guo YL. 2013. Gene family evolution in green plants with emphasis on the origination and evolution of *Arabidopsis thaliana* genes. *Plant J.* 73(6):941–951.
- Guo ZH, Ma PF, Yang GQ, Hu JY, Liu YL, Xia EH, Zhong MC, Zhao L, Sun GL, Xu YX, et al. 2019. Genome sequences provide insights into the reticulate origin and unique traits of woody bamboos. *Mol Plant.* 12(10):1353–1365.
- He CY, Cui K, Zhang JG, Duan AG, Zeng YF. 2013. Next-generation sequencing-based mRNA and microRNA expression profiling analysis revealed pathways involved in the rapid growth of developing culms in moso bamboo. *BMC Plant Biol.* 13:119–114.
- He XQ, Suzuki K, Kitamura S, Lin JX, Cui KM, Itoh T. 2002. Toward understanding the different function of two types of parenchyma cells in bamboo culms. *Plant Cell Physiol.* 43(2):186–195.
- Heames B, Schmitz J, Bornberg-Bauer E. 2020. A continuum of evolving de novo genes drives protein-coding novelty in *Drosophila*. *J Mol Evol.* 88(4):382–398.
- Hu HY, Uesaka M, Guo S, Shimai K, Lu TM, Li F, Fujimoto S, Ishikawa M, Liu SP, Sasagawa Y, et al.; EXPANDE Consortium. 2017. Constrained vertebrate evolution by pleiotropic genes. *Nat Ecol Evol.* 1(11):1722–1730.
- Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet.* 11(2):97–108.
- Janzen DH. 1976. Why bamboos wait so long to flower. *Annu Rev Ecol Syst.* 7(1):347–391.
- Jin GH, Zhou YL, Yang H, Hu YT, Shi Y, Li L, Siddique A, Liu CN, Zhu AD, Zhang CJ, et al. 2021. Genetic innovations: transposable element recruitment and de novo formation lead to the birth of orphan genes in the rice genome. *J Syst Evol.* 59(2):341–351.
- Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* 20(10):1313–1326.
- Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428(6983):617–624.
- Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TCG. 2009. More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet.* 25(9):404–413.
- Kryuchkova-Mostacci N, Robinson-Rechavi M. 2017. A benchmark of gene expression tissue-specificity metrics. *Brief Bioinform.* 18:205–214.
- Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.
- Leiboff S, Hake S. 2019. Reconstructing the transcriptional ontogeny of maize and sorghum supports an inverse hourglass model of inflorescence development. *Curr Biol.* 29(20):3410–3419.
- Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, Senderovich N, Kovalev E, Silver DH, Feder M, et al. 2016. The mid-developmental transition and the evolution of animal body plans. *Nature* 531(7596):637–641.
- Levine MT, Jones CD, Kern AD, Lindfors HA, Begun DJ. 2006. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proc Natl Acad Sci U S A.* 103(26):9935–9939.
- Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.
- Li L, Cheng ZC, Ma YJ, Bai QS, Li XY, Cao ZH, Wu ZN, Gao J. 2018. The association of hormone signalling genes, transcription and changes in shoot anatomy during moso bamboo growth. *Plant Biotechnol J.* 16(1):72–85.
- Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13(9):2178–2189.
- Li W, Shi C, Li K, Zhang QJ, Tong Y, Zhang Y, Wang J, Clark L, Gao LZ. 2021. Draft genome of the herbaceous bamboo *Raddia distichophylla*. *G3 (Bethesda)* 11(2):jkaa049.
- Li ZH, Chen CJ, Mi RY, Gan WT, Dai JQ, Jiao ML, Xie H, Yao YG, Xiao SL, Hu LB. 2020. A strong, tough, and scalable structural material from fast-growing bamboo. *Adv Mater.* 32(10):1906308.
- Liem KF. 1988. Form and function of lungs - the evolution of air breathing mechanisms. *Am Zool.* 28(2):739–759.
- Liese W, Kohl M. 2015. Bamboo. The plant and its uses. Switzerland: Springer International Publishing.
- Lima RAF, Rother DC, Muler AE, Lepesch IF, Rodrigues RR. 2012. Bamboo overabundance alters forest structure and dynamics in the Atlantic Forest hotspot. *Biol Conserv.* 147(1):32–39.
- Long MY, Betran E, Thornton K, Wang W. 2003. The origin of new genes: glimpses from the young and old. *Nat Rev Genet.* 4(11):865–875.
- Long MY, VanKuren NW, Chen SD, Vibranovski MD. 2013. New gene evolution: little did we know. *Annu Rev Genet.* 47:307–333.
- Ma PF, Liu YL, Jin GH, Liu JX, Wu H, He J, Guo ZH, Li DZ. 2021. The *Pharus latifolius* genome bridges the gap of early grass evolution. *Plant Cell* 33(4):846–864.
- McLysaght A, Guerzoni D. 2015. New genes from non-coding sequence: the role of *de novo* protein-coding genes in eukaryotic evolutionary innovation. *Philos Trans R Soc Lond B Biol Sci.* 370(1678):20140332.
- McLysaght A, Hokamp K, Wolfe KH. 2002. Extensive genomic duplication during early chordate evolution. *Nat Genet.* 31(2):200–204.
- Moyers BA, Zhang JZ. 2016. Evaluating phylostratigraphic evidence for widespread *de novo* gene birth in genome evolution. *Mol Biol Evol.* 33(5):1245–1256.
- Murphy DN, McLysaght A. 2012. *De novo* origin of protein-coding genes in murine rodents. *PLoS One* 7(11):e48650.
- Palmieri N, Kosiol C, Schlötterer C. 2014. The life cycle of *Drosophila* orphan genes. *Elife* 3:e01311.



- Paque S, Mouille G, Grandont L, Alabadi D, Gaertner C, Goyallon A, Muller P, Primard-Brisset C, Sormani R, Blazquez MA, et al. 2014. AUXIN BINDING PROTEIN1 links cell wall remodeling, auxin signaling, and cell expansion in *Arabidopsis*. *Plant Cell* 26(1):280–295.
- Peng Z, Lu Y, Li L, Zhao Q, Feng Q, Gao Z, Lu H, Hu T, Yao N, Liu K, et al. 2013. The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*). *Nat Genet* 45(4):456–461.
- Peng ZH, Zhang CL, Zhang Y, Hu T, Mu SH, Li XP, Gao J. 2013. Transcriptome sequencing and analysis of the fast growing shoots of moso bamboo (*Phyllostachys edulis*). *PLoS One* 8(11):e78944.
- Pigliucci M. 2008. What, if anything, is an evolutionary novelty? *Philos Sci* 75(5):887–898.
- Prum RO, Brush AH. 2002. The evolutionary origin and diversification of feathers. *Q Rev Biol* 77(3):261–295.
- Qiao X, Li QH, Yin H, Qi KJ, Li LT, Wang RZ, Zhang SL, Paterson AH. 2019. Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biol* 20(1):38–23.
- Quint M, Drost HG, Gabel A, Ullrich KK, Bonn M, Grosse I. 2012. A transcriptomic hourglass in plant embryogenesis. *Nature* 490(7418):98–101.
- Ritchie ME, Phipson B, Wu D, Hu YF, Law CW, Shi W, Smyth GK. 2015. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43(7):e47.
- Rodolger C, Ebbing A, Sharma DR, Okumura M, Sommer RJ, Korswagen HC. 2021. Spatial transcriptomics of nematodes identifies sperm cells as a source of genomic novelty and rapid evolution. *Mol Biol Evol* 38(1):229–243.
- Sandve SR, Rohlfs RV, Hvidsten TR. 2018. Subfunctionalization versus neofunctionalization after whole-genome duplication. *Nat Genet* 50(7):908–909.
- Santos ME, Le Bouquin A, Crumiere AJ, Khila A. 2017. Taxon-restricted genes at the origin of a novel trait allowing access to a new environment. *Science* 358(6361):386–389.
- Schlotterer C. 2015. Genes from scratch - the evolutionary fate of *de novo* genes. *Trends Genet* 31:215–219.
- Sestak MS, Bozicevic V, Bakaric R, Dunjko V, Domazet-Loso T. 2013. Phylostratigraphic profiles reveal a deep evolutionary history of the vertebrate head sensory systems. *Front Zool* 10(1):18.
- Sestak MS, Domazet-Loso T. 2015. Phylostratigraphic profiles in zebrafish uncover chordate origins of the vertebrate brain. *Mol Biol Evol* 32:299–312.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498–2504.
- Shao Y, Chen C, Shen H, He BZ, Yu D, Jiang S, Zhao S, Gao Z, Zhu Z, Chen X, et al. 2019. GenTree, an integrated resource for analyzing the evolution and function of primate-specific coding genes. *Genome Res* 29(4):682–696.
- Shi L, Derouiche A, Pandit S, Rahimi S, Kalantari A, Futo M, Ravikumar V, Jers C, Mokkapat V, Vlahovick K, et al. 2020. Evolutionary analysis of the *Bacillus subtilis* genome reveals new genes involved in sporulation. *Mol Biol Evol* 37(6):1667–1678.
- Song X, Peng C, Zhou G, Gu H, Li Q, Zhang C. 2016. Dynamic allocation and transfer of non-structural carbohydrates, a possible mechanism for the explosive growth of moso bamboo (*Phyllostachys heterocycla*). *Sci Rep* 6:25908.
- Song XZ, Zhou GM, Jiang H, Yu SQ, Fu JH, Li WZ, Wang WF, Ma ZH, Peng CH. 2011. Carbon sequestration by Chinese bamboo forests and their ecological benefits: assessment of potential, problems, and future challenges. *Environ Rev* 19(NA):418–428.
- Soreng RJ, Peterson PM, Romaschenko K, Davidse G, Teisher JK, Clark LG, Barbera P, Gillespie LJ, Zuloaga FO. 2017. A worldwide phylogenetic classification of the Poaceae (Gramineae) II: an update and a comparison of two 2015 classifications. *J Syst Evol* 55(4):259–290.
- Tao GY, Ramakrishnan M, Vinod KK, Yrjala K, Satheesh V, Cho J, Fu Y, Zhou MB. 2020. Multi-omics analysis of cellular pathways involved in different rapid growth stages of moso bamboo. *Tree Physiol* 40(11):1487–1508.
- Tautz D, Domazet-Loso T. 2011. The evolutionary origin of orphan genes. *Nat Rev Genet* 12(10):692–702.
- Taylor JS, Raes J. 2004. Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet* 38:615–643.
- Toll-Riera M, Bosch N, Bellora N, Castelo R, Armengol L, Estivill X, Alba MM. 2009. Origin of primate orphan genes: a comparative genomics approach. *Mol Biol Evol* 26(3):603–612.
- Tribble W, Kronauer DJC. 2021. Hourglass model for developmental evolution of ant castes. *Trends Ecol Evol* 36(2):100–103.
- Trigos AS, Pearson RB, Papenfuss AT, Goode DL. 2017. Altered interactions between unicellular and multicellular genes drive hallmarks of transformation in a diverse range of solid tumors. *Proc Natl Acad Sci U S A* 114(24):6406–6411.
- Tyanova S, Temu T, Cox J. 2016. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* 11(12):2301–2319.
- Ueda K. 1960. Studies on the physiology of bamboo, with reference to practical application. Kyoto (Japan): Kyoto University Forests Press.
- Vakirlis N, Carvunis AR, McLysaght A. 2020. Synteny-based analyses indicate that sequence divergence is not the main source of orphan genes. *Elife* 9:e53500.
- Van Oss SB, Carvunis AR. 2019. *De novo* gene birth. *PLoS Genet* 15(5):e1008160.
- Vinckenbosch N, Dupanloup I, Kaessmann H. 2006. Evolutionary fate of retroposed gene copies in the human genome. *Proc Natl Acad Sci U S A* 103(9):3220–3225.
- Wagner GP, Lynch VJ. 2010. Evolutionary novelties. *Curr Biol* 20(2):R48–R52.
- Wang J, Tao F, Marowsky NC, Fan C. 2016. Evolutionary fates and dynamic functionalization of young duplicate genes in *Arabidopsis* genomes. *Plant Physiol* 172(1):427–440.
- Wang TT, Wang HY, Cai DW, Gao YB, Zhang HX, Wang YS, Lin CT, Ma LY, Gu LF. 2017. Comprehensive profiling of rhizome-associated alternative splicing and alternative polyadenylation in moso bamboo (*Phyllostachys edulis*). *Plant J* 91(4):684–699.
- Wang XQ, Ren HQ, Zhang B, Fei BH, Burgert I. 2012. Cell wall structure and formation of maturing fibres of moso bamboo (*Phyllostachys pubescens*) increase buckling resistance. *J R Soc Interface* 9(70):988–996.
- Wang Y, Gao Y, Zhang H, Wang H, Liu X, Xu X, Zhang Z, Kohnen MV, Hu K, Wang H, et al. 2019. Genome-wide profiling of circular RNAs in the rapidly growing shoots of moso bamboo (*Phyllostachys edulis*). *Plant Cell Physiol* 60(6):1354–1373.
- Wang YP, Tang HB, DeBarry JD, Tan X, Li JP, Wang XY, Lee TH, Jin HZ, Marler B, Guo H, et al. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* 40(7):e49.
- Wei Q, Guo L, Jiao C, Fei ZJ, Chen M, Cao JJ, Ding YL, Yuan QS. 2019. Characterization of the developmental dynamics of the elongation of a bamboo internode during the fast growth stage. *Tree Physiol* 39(7):1201–1214.
- Wei Q, Jiao C, Ding YL, Gao S, Guo L, Chen M, Hu P, Xia SJ, Ren GD, Fei ZJ. 2018. Cellular and molecular characterizations of a slow-growth variant provide insights into the fast growth of bamboo. *Tree Physiol* 38(4):641–654.
- Wright RM, Aglyamova GV, Meyer E, Matz MV. 2015. Gene expression associated with white syndromes in a reef building coral, *Acropora hyacinthus*. *BMC Genomics* 16(1):371.
- Wu DD, Wang X, Li Y, Zeng L, Irwin DM, Zhang YP. 2014. “Out of pollen” hypothesis for origin of new genes in flowering plants: study from *Arabidopsis thaliana*. *Genome Biol Evol* 6(10):2822–2829.
- Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21(9):1859–1875.
- Xie C, Bekpen C, Kunzel S, Keshavarz M, Krebs-Wheaton R, Skrabar N, Ullrich KK, Tautz D. 2019. A *de novo* evolved gene in the house mouse regulates female pregnancy cycles. *Elife* 8:e44392.



- Xu Y, Wong MH, Yang JL, Ye ZQ, Jiang PK, Zheng SJ. 2011. Dynamics of carbon accumulation during the fast growth period of bamboo plant. *Bot Rev.* 77(3):287–295.
- Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* 21(5):650–659.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci.* 13(5):555–556.
- Yu GC, Wang LG, Han YY, He QY. 2012. ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16(5):284–287.
- Yu XL, Wang YS, Kohnen MV, Piao MX, Tu M, Gao YB, Lin CT, Zuo ZC, Gu LF. 2019. Large scale profiling of protein isoforms using label-free quantitative proteomics revealed the regulation of nonsense-mediated decay in moso bamboo (*Phyllostachys edulis*). *Cells-Basel* 8:744.
- Zhang L, Ren Y, Yang T, Li GW, Chen JH, Gschwend AR, Yu Y, Hou GX, Zi J, Zhou R, et al. 2019. Rapid evolution of protein diversity by *de novo* origination in *Oryza*. *Nat Ecol Evol.* 3(4):679–690.
- Zhang Q, Liu W, Liu CJ, Lin SY, Guo AY. 2018. SEGtool: a specifically expressed gene detection tool and applications in human tissue and single-cell sequencing data. *Brief Bioinform.* 19(6):1325–1336.
- Zhang W, Landback P, Gschwend AR, Shen B, Long M. 2015. New genes drive the evolution of gene interaction networks in the human and mouse genomes. *Genome Biol.* 16:202.
- Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, Yu J. 2006. KaKs\_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* 4(4):259–263.
- Zhang Z, Xiao J, Wu J, Zhang H, Liu G, Wang X, Dai L. 2012. ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. *Biochem Biophys Res Commun.* 419(4):779–781.
- Zhao H, Gao Z, Wang L, Wang J, Wang S, Fei B, Chen C, Shi C, Liu X, Zhang H, et al. 2018. Chromosome-level reference genome and alternative splicing atlas of moso bamboo (*Phyllostachys edulis*). *Gigascience* 7(10):giy115.
- Zhao H, Lou Y, Sun H, Li L, Wang L, Dong L, Gao Z. 2016. Transcriptome and comparative gene expression analysis of *Phyllostachys edulis* in response to high light. *BMC Plant Biol.* 16:34.
- Zhao L, Saelao P, Jones CD, Begun DJ. 2014. Origin and spread of *de novo* genes in *Drosophila melanogaster* populations. *Science* 343(6172):769–772.
- Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P. 2012. Parallel molecular evolution in an herbivore community. *Science* 337(6102):1634–1637.
- Zhou F. 1983. Studies on bamboo shoot culm growth. *Bamboo Res.* 2:17–39.
- Zu J, Gu YX, Li Y, Li CT, Zhang WY, Zhang YE, Lee U, Zhang L, Long MY. 2019. Topological evolution of coexpression networks by new gene integration maintains the hierarchical and modular structures in human ancestors. *Sci China Life Sci.* 62(4):594–608.