

RESEARCH ARTICLE

Open Access



# Divergence and adaptive evolution of the gibberellin oxidase genes in plants

Yuan Huang<sup>1</sup>, Xi Wang<sup>1</sup>, Song Ge<sup>2</sup> and Guang-Yuan Rao<sup>1\*</sup>

## Abstract

**Background:** The important phytohormone gibberellins (GAs) play key roles in various developmental processes. GA oxidases (GAoxs) are critical enzymes in GA synthesis pathway, but their classification, evolutionary history and the forces driving the evolution of plant GAox genes remain poorly understood.

**Results:** This study provides the first large-scale evolutionary analysis of GAox genes in plants by using an extensive whole-genome dataset of 41 species, representing green algae, bryophytes, pteridophyte, and seed plants. We defined eight subfamilies under the GAox family, namely C<sub>19</sub>-GA2ox, C<sub>20</sub>-GA2ox, GA20ox, GA3ox, GAox-A, GAox-B, GAox-C and GAox-D. Of these, subfamilies GAox-A, GAox-B, GAox-C and GAox-D are described for the first time. On the basis of phylogenetic analyses and characteristic motifs of GAox genes, we demonstrated a rapid expansion and functional divergence of the GAox genes during the diversification of land plants. We also detected the subfamily-specific motifs and potential sites of some GAox genes, which might have evolved under positive selection.

**Conclusions:** GAox genes originated very early—before the divergence of bryophytes and the vascular plants and the diversification of GAox genes is associated with the functional divergence and could be driven by positive selection. Our study not only provides information on the classification of GAox genes, but also facilitates the further functional characterization and analysis of GA oxidases.

**Keywords:** Characteristic motif, Functional divergence, Gibberellin, Phytohormones, Positive selection, 2-oxoglutarate-dependent dioxygenase

## Background

The gibberellins (GAs), one of important phytohormones, form a large family of diterpene hormones that are involved in various growth and developmental processes in plants [1–6]. Among hundreds of plant GAs, only a few, such as GA1 and GA4 are bioactive [6]. The function of bioactive GAs generally depends on their concentration in a given tissue, and this is mainly affected by their biosynthesis and/or deactivation [7]. In *Arabidopsis*, the synthesis pathway of GAs has been elucidated, and most genes encoding enzymes to catalyze GA biosynthesis have been identified [7–11]. In the GA synthesis pathway, three classes of enzyme (i.e., Terpene synthases (TPSs), P450s and GA oxidases (GAoxs)) are required for the biosynthesis of bioactive GAs from geranylgeranyl diphosphate (GGDP, a common precursor of terpenes in plants), and the pathway can be divided into

two main steps: early and late. The early steps are catalyzed by a series of enzymes, i.e., CPS, KS, KO, and KAO, which are all encoded by single genes. The enzymes catalyzing later steps (i.e., GA2 oxidase, GA20 oxidase, and GA3 oxidase) are encoded by small gene families, and these oxidases are all soluble 2-oxoglutarate-dependent dioxygenases (2OGDs) [7, 10, 12, 13]. In comparison with the biosynthesis enzymes in the early steps, those in the later steps are differentially regulated by developmental and environmental cues, and these enzymes play key roles in the regulation of bioactive GA levels. For example, the loss-of-function in GA20 oxidase (GA20ox) and GA3 oxidase (GA3ox) can generate dwarf phenotypes, such as the well-known Green Revolution *sd-1* [10, 14–17]. Interestingly, the function of GA2 oxidase (GA2ox) is to decrease the levels of active GAs rather than increasing the GA level as do GA20ox and GA3ox [18]. So far, most studies have focused on the functions of GA2ox, GA20ox and GA3ox [10, 12, 18–21]. However, little is known about the evolutionary history of the GAox genes [13, 22–24].

\* Correspondence: rao@pku.edu.cn

<sup>1</sup>College of Life Sciences, Peking University, Beijing 100871, China  
Full list of author information is available at the end of the article

GAoxs are ubiquitous in vascular plants, but no GAs has ever been isolated and chemically identified from bryophytes [25–28]. Those authors hypothesized that GAs may have first appeared in ancient pteridophytes and that the hormonal signaling pathway developed later during the evolution of land plants. However, several putative GA biosynthetic genes have been found in the *Physcomitrella patens* genome, suggesting the GAs could appear in the earlier lineages of land plants [29, 30]. Although the pathways leading to production of bioactive GAs have been studied in plants [31–33], the evolutionary history of the GA synthetic pathway and its catalyzing enzyme genes, especially of GAox genes remains unclear.

The GAox family has been evolutionarily analyzed, but only focusing on the GA20ox, GA3ox and GA2ox subfamilies of a few angiosperm species [13, 23, 24]. Han and Zhu (2011) identified 61 GA oxidase genes from rice (*Oryza sativa*), Arabidopsis and soybean (*Glycine max*), and performed the phylogenetic analysis of these GAox genes. They were grouped into four clades (GA20ox, GA3ox, C<sub>19</sub>-GA2ox, and C<sub>20</sub>-GA2ox.), but their relationship remains unclear, especially of new putative GA20ox homologs. For example, four new genes were described under the GA20ox subfamily (i.e., *OsGA20ox5*, *OsGA20ox6*, *OsGA20ox7* and *OsGA20ox8*), but three of them (*OsGA20ox5*, *OsGA20ox6*, and *OsGA20ox8*) did not reside within the GA20ox subgroup according to their phylogenetic tree. Meanwhile, Han and Zhu (2011) showed that the GA2ox subfamily is not monophyletic [23], which was supported by Giacomelli and her colleagues' study [24]. In addition, there is controversy over the relationship between GA20ox, GA3ox, C<sub>19</sub>-GA2ox and C<sub>20</sub>-GA2ox [23, 24]. Giacomelli et al. (2013) found the sister relationship between C<sub>19</sub>-GA2ox and GA3ox, while the close relationship was found between GA3ox and C<sub>20</sub>-GA2ox in Han and Zhu's study. Not only the relationship between GAox genes but the evolutionary history and expansion mechanisms of the GAox genes are also poorly understood.

The GAox family belongs to the 2OGDs superfamily. The catalytic core of GAoxs has a double-stranded  $\beta$ -helix (DSBH) fold containing a HX[DE] dyad (where X could be any amino acid) and a conserved carboxy-terminal histidine which together chelate a single iron atom. Specifically, the DSBH region has seven conserved strands that are common to all these kinds of proteins and is arranged in two sheets in a jelly-roll topology [34]. To date, the function of several GAox genes has been determined but their gene structure not studied systematically [10, 20, 35–40]. Of GA oxidases, GA 20ox was first isolated from cotyledons of immature pumpkin seeds by Lange et al. (1994) [41]. Structurally this gene has characteristic motifs that are highly conserved in 2OGDs. Later on, other GAox

genes were isolated and characterized in plants, such as Carrizo citrange [42], poplar [43], *Phaseolus coccineus* and Arabidopsis [19], rice [38, 44, 45], spinach [37, 46], watermelon [39], pea [47, 48]. On the characteristic motifs of different GA oxidases, no consensus has come out. Meanwhile, neither diagnostic sequences nor gene structures have been established for identification of GAox subfamilies. For example, four GAox homologs were found in the *Physcomitrella patens* genome, but it is hard to conclude what they are and which gene subfamily they belong to [30]. Thus, characterizing the motif specific to every gene subfamily would be of particular value. In brief, the study of gene structures and conserved motifs will not only provide an insight into the classification of GAox genes, but also their evolutionary history and functional diversification.

Almost all aspects of the functional divergence of genes are in some way linked to gene duplications, which occur ubiquitously [49–54]. What was the evolutionary force driving the functional divergence of genes or gene families? Generally, gene duplications are thought to be an important precursor to the functional divergence of genes, and selection is the main evolutionary force driving gene function diversification [22, 52, 55–59]. We investigated whether the functional divergence of GAox genes was driven by positive selection, in particular for four newly described ones. We first analyzed the distribution of positively selected amino acids in the characteristic motifs of genes. Then, we used the programs FunDi and GroupSim to validate the hypothesis that the conserved domains were associated with divergence (change) in function among GAs, and performed the GroupSim to identify the specificity determining positions (SDPs) [60–62].

With the availability of more whole genome sequence data and new genomics tools developed in model plants, the study of GA biosynthetic pathways, GAox genes and their evolution is now feasible. To reach the above objectives, we performed a comprehensive analysis of the GAox gene homologs in 41 species, using whole genome sequencing data, covering the main lineages of green plants from green algae, bryophytes, pteridophytes and gymnosperms to angiosperms. First, a genome-wide search was conducted and GAox homologs identified. Then, we performed phylogenetic analysis to explore the evolutionary history of GAox genes and established the diagnostic features of conserved motifs for every GAox subfamily. Next, the expansion mechanism and evolutionary forces associated with the GAox genes were investigated by analyzing the evolutionary history, structure and functional divergence of these genes. Finally, we compared the GA biosynthesis pathways in bryophytes, pteridophytes and seed plants, to explore the evolution of pathways and their relationships with the adaptation of plants to terrestrial environments.

## Methods

### Sequence acquisition and compilation

In order to investigate the evolutionary history of the GAox family in plants, all possible sequences of GAox homologs need to be sourced from whole genome sequencing data. GAox belongs to the 2OGDs superfamily, which has a very conserved and characteristic 2OGDs motif (Pfam id, PF03171) [7, 34, 63]. Functional GAox genes of *Arabidopsis* and rice, and four newly found GAox genes (*OsGA20ox5*, *OsGA20ox6*, *OsGA20ox7*, and *OsGA20ox8*; Han and Zhu, 2011) were used as queries to BLAST against the whole genome sequences of 41 species. The sequences with the active site core of 2OGD, and the motifs such as NYYPXCXXP, LPWKET, LSWSEA were recognized as putative GAox genes. GAox homologs of each species were downloaded from the following databases: <http://www.phytozome.net>; <http://congenie.org>; <http://www.Arabidopsis.org/>; <http://rice.plantbiology.msu.edu/>; <http://www.amborella.org/>. Then, an extensive screening of databases (dbEST, GenBank) was performed using different BLAST algorithms to obtain all GAox homologs [64]. Finally, we manually checked all the search results to reduce hits with partial conserved-domains and other false positives, using BioEdit (v7.0.5) [65]. Sequences were accepted from BLAST results as long as they shared at least 35 % identity, and had an expected threshold lower than  $1e-40$  with the conserved domain in the 2OGDs family. Meanwhile, the gene prediction and annotation analysis was considered for the GAox homologs acquisition. Then, 854 GAox homologous were obtained from 41 species; these covered all the main lineages of plants including unicellular green algae, bryophytes, pteridophytes, gymnosperms and angiosperms. Additionally, two ACOs (1-aminocyclopropane-1-carboxylate) of rice were included in phylogenetic analyses considering both ACOs and GA oxidases belong to the 2OGD superfamily.

### Multiple sequence alignment and phylogenetic analysis

A protein multiple sequence alignment (MSA) was generated using MUSCLE (Multiple Sequence Comparison by Log-Expectation) software [66] with the default settings. Experimentally characterized GAox genes of *Arabidopsis thaliana* and *Oryza sativa* were used as criteria to detect potential problems such as frame shift mutations. Protein-coding DNA sequences (CDS) were aligned based on the protein alignments in the DAMBE with the default parameters, then converted the CDS alignments into PAML format for further analyses [67, 68].

RAxML 7.2.6 (Randomized Axelerated Maximum Likelihood) [69] was used to construct maximum-likelihood (ML) phylogenetic trees under the JTT amino acid substitution model which was derived from

ProtTest 2.4 [70], with an estimated gamma distribution parameter, and optimized starting with a BIONJ tree. Statistical support for the nodes on the maximum-likelihood tree was evaluated by bootstrap analysis with 500 iterations.

### Gene structure, motif identification and homology modeling analysis

To study the evolution of the GAox gene structure, the DNA sequences corresponding to each predicted gene and the gene structure of the GAox gene family were all retrieved from <http://www.phytozome.net>.

To obtain the characteristic motif of each GAox subfamily, the following steps were taken: 1) the web-based multiple expectation maximization for motif elicitation (MEME) analysis (<http://meme-suite.org/>, [71]) was employed with the following parameters: maximum number of motifs = 15, and optimum motif widths constrained to between 6 and 30 residues. 2) The NCBI-CDD (Conserved Domain Database) (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) and the SMART (Simple Modular Architecture Research Tool, <http://smart.embl-heidelberg.de/>) databases were used for identification and function explanation of those putative motifs [72]. 3) Multiple sequence alignment of the GAox genes was achieved using MUSCLE software, to find the characteristic motifs of every GAox subfamily. 4) To examine the relationship between the characteristic motif of every GAox subfamily and the putative functional region, the 3D structures of eight OsGAox genes, i.e., *OsGA20ox7*, *OsGA20ox5*, *OsGA20ox4*, *OsGA3ox2*, *OsGA20ox5*, *OsGA20ox6*, *OsGA20ox7* and *OsGA20ox8*, were predicted following the homology modeling using the Swiss-Model (<http://swissmodel.expasy.org/>) with the default parameters, and the higher reliability PDBs were selected as templates according to the global quality estimation (GMQE) and qualitative model energy analysis (QMEAN) scores [73–75].

Three-dimensional model images with the characteristic motifs and the putative positive selection and functional divergence sites were manipulated and rendered in PyMOL (<http://pymol.sourceforge.net/>) [76]. All the sequence logos were generated using the online Weblogo platform (<http://weblogo.berkeley.edu/>).

### Testing positive selection and functional divergence

We used the programs PAML 4.8, FunDi and GroupSim to detect sites involved in the functional divergence of GAox sequences and the adaptive evolution of GAox subfamilies. Sites identified as being functionally divergent or positively selective by all or two of these three programs were highlighted in the 3D protein structure of GA oxidases, which could be tested with direct experimental techniques in the future.

In order to reduce computation time and the potential impact of synonymous site saturation due to sequence divergence [77], the positive selection test was conducted on the site and branch-site models with 46 less divergent GAox homologs from six grass species of monocots. Two ACOs of rice were added in phylogenetic analysis of GAox genes. We performed our analysis using alignments inferred by phylogeny-aware multiple sequence alignment program Probabilistic Alignment Kit (PRANK) [78, 79], which has been reported to outperform other aligners in simulations [80]. In addition, by GUIDANCE filtering [81], the PRANK alignment has a higher average GUIDANCE score of 0.70, while the MAFFT, MUSCLE and CLUSTALW alignments is 0.65, 0.63 and 0.66 respectively, though the GUIDANCE cannot tell which alignment is better.

Selective pressure was tested based on the phylogeny of GAox genes by comparing the nonsynonymous/synonymous substitution ratios ( $\omega = d_N/d_S$ ) with  $\omega = 1$ ,  $< 1$ , and  $> 1$ , which indicate neutral evolution, purifying selection, and positive selection, respectively. To test the functional divergence of different GAox subfamilies, the codeml program in the PAML 4.8 package was used to investigate possible selection acting on different gene subfamilies [82]. We selected F3X4 as codon frequency model, as it accounts for the most important feature of the mutation process, the unequal base frequencies, and the correction for transition/transversion bias is obtained by ML estimation of the kappa parameter [82–85]. The equilibrium frequencies of codons and nucleotides among subfamilies are measured by the DAMBE. Considering the fact that positive selection often operates episodically on a few amino acid sites in a small number of lineages of a phylogenetic tree [86, 87], the modified branch-site model A was run for each gene in each lineage across the phylogenetic tree of the monocots. The LRT (likelihood ratio test) is a general method for testing assumptions (model parameters) through comparison of two competing hypotheses. We used branch-site model A to construct branch-site test 2, which is also called the branch-site test of positive selection. The null and alternative hypotheses are as follows: Null hypothesis (branch site model A, with  $\omega_2 = 1$  fixed): model = 2; NSsites = 2; fix\_omega = 1; omega = 1. Alternative hypothesis (branch site model A, with  $\omega_2$  estimated): model = 2; NSsites = 2; fix\_omega = 0; omega = 1.5 (or any value  $> 1$ ). Because of very high sequence divergence, and the likelihood of saturation of synonymous changes, we treated the branch-site test as an exploratory analysis. Additional analyses at the codon level (sites models) and amino acid level (described below) were employed to more formally assess the hypothesis that GAoxs have been subject to functional divergence following the major gene duplication events. Therefore, we analyzed separately each of eight subfamilies (less divergent)

under the sites models, only with GAox homologs of angiosperms. Unlike branch-site model, site-specific models allowing  $\omega$  to vary among sites [55, 88, 89] were used to determine whether particular amino-acid residues within GAox families have been subject to positive selection. We use an LRT comparing M0 (one-ratio) with M3 (discrete) ( $\kappa = 3$ ) to test for variable selective pressure among sites, and two LRTs to test for sites evolving by positive selection, comparing: M1a (Nearly Neutral) against M2a (Positive Selection), and M7 (beta) ( $\kappa = 10$ ) against M8 (beta &  $\omega$ ) ( $\kappa = 10$ ). The PP (posterior probability) for the sites under positive selection was calculated by the Bayes empirical Bayes (BEB) method [90].

Default options were used in the GroupSim and FunDi using QmmRAXML [91]. In order to evaluate the predictions made by FunDi, we analyzed all the comparisons of eight GAox subfamilies, except for the C<sub>19</sub>-GAox/C<sub>20</sub>-GA2ox and C<sub>20</sub>-GA2ox/GAox-C, due to their large datasets (gene homologs and the alignments) and computation cost (Additional file 1: Table S3). The sites of each subfamily motif were highlighted with red if they occurred in more than five comparisons (Fig. 1). Generally, sites with a functional divergence score above 0.5 were considered to be functionally divergent in order to identify the maximum number of possible candidate sites for consideration [60, 62]. In addition, we analyzed the subfamily characteristic motifs, and mapped those sites detected by two or three programs (PAML 4.8, FunDi and GroupSim) on the 3D protein structures (Figs. 1 and 2). We found that these sites are those positively selected sites detected by the branch-site model of PAML 4.8 with PP  $\geq 95\%$  and those under functional divergence predicted by FunDi and GroupSim with PP  $\geq 99\%$ .

## Results

### Identification and distribution of GAox genes in plants

Our BLAST search results revealed 854 sequences of GAox homologs from the 41 species with whole genome sequencing data; these represent the main lineages of green plants (Additional file 2: Table S2). In algae, 6 putative GAox homologs were found based on the catalytic core amino acid sequences of 2OGDs in the carboxyl terminus (Fig. 3). As shown in Additional file 2: Table S2, the GAox genes varied across the different species and plant taxonomic groups, and there exist different distribution patterns for GAox subfamilies, especially for GAox-A, GAox-B, GAox-C and GAox-D. For instance, all 37 land plants have GAox gene homologs from the subfamilies C<sub>19</sub>-GA2ox, C<sub>20</sub>-GA2ox, GA20ox and GA3ox, while other subfamily genes are not present in one or more taxa, especially in some basal lineages of the land plants. In *Physcomitrella patens* genome, putative GAox homologs were only found in C<sub>19</sub>-GA2ox, GA3ox and GAox-C. There were no homologs of C<sub>20</sub>-GA2ox and GAox-B in a

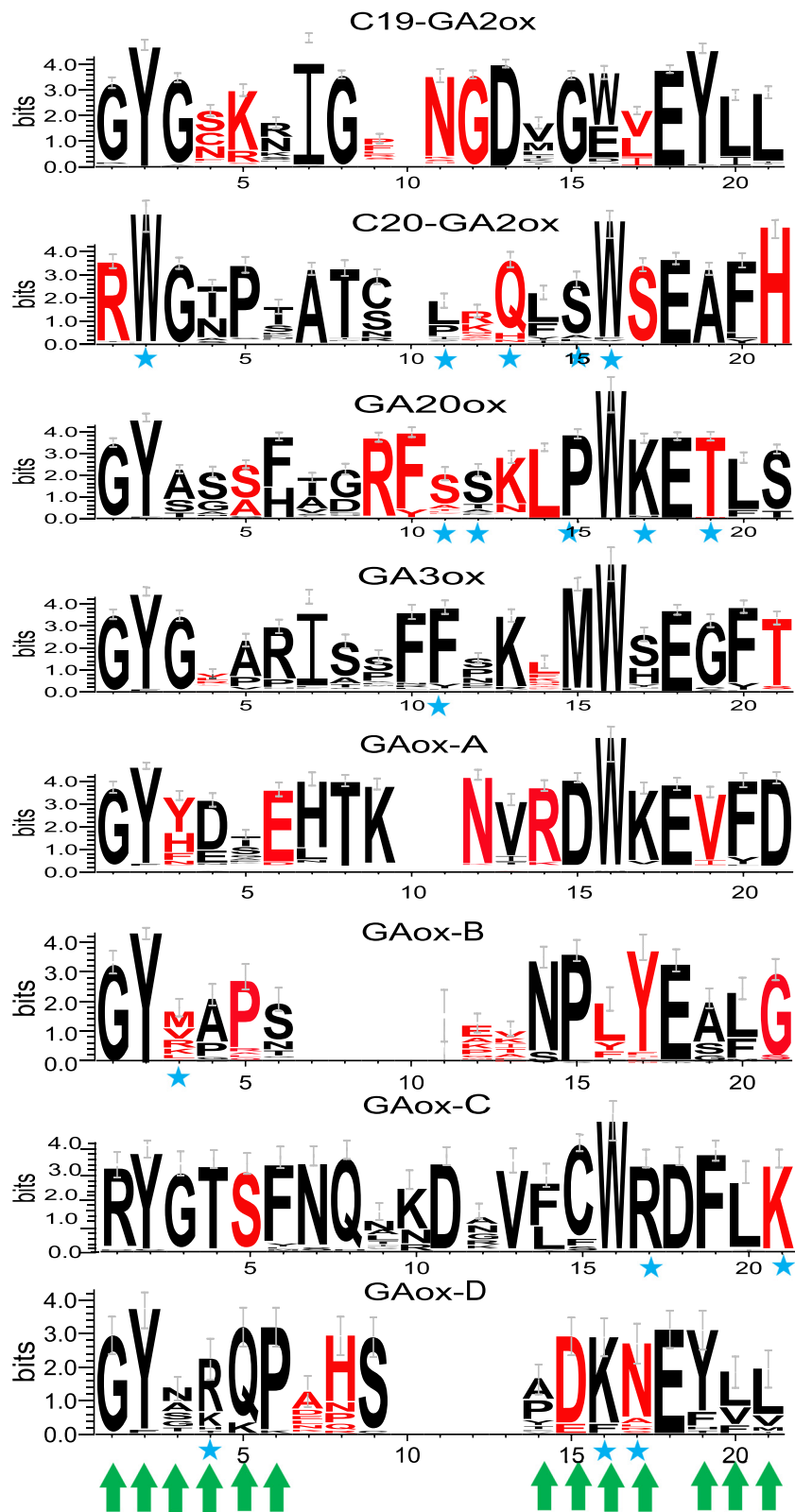


Fig. 1 (See legend on next page.)

(See figure on previous page.)

**Fig. 1** Specific conserved domains of the eight GA oxidase subfamilies and sites predicted to be functionally divergent by at least two of the three programs (PAML 4.8, FunDi, and GroupSim). The overall height of each stack indicates the sequence conservation at that position, whereas the height of symbols within each stack reflects the relative frequency of the corresponding amino acid. Green arrows mean that the sites in these columns are under functional divergence detected by GroupSim with PP  $\geq 99\%$ ; the sites highlighted in red are under functional divergence detected by FunDi with PP  $\geq 99\%$ ; the sites highlighted with blue star are positive selected sites detected by PAML 4.8 with PP  $\geq 95\%$

pteridophytes *Selaginella moellendorffii*. Homologs of C<sub>20</sub>-GA2ox and GA20ox were not found in a gymnosperm *Picea abies*. Of angiosperms, monocotyledons have all eight GAox subfamily homologs, while basal angiosperm *Amborella trichopoda* and some eudicotyledons have no GAox-D homologs in their genome. In addition, not only the distribution pattern but the gene sequence and structure were also varied between GAox genes of algae and land plants (Additional file 3: Data S1 and Additional file 4: Figure S1).

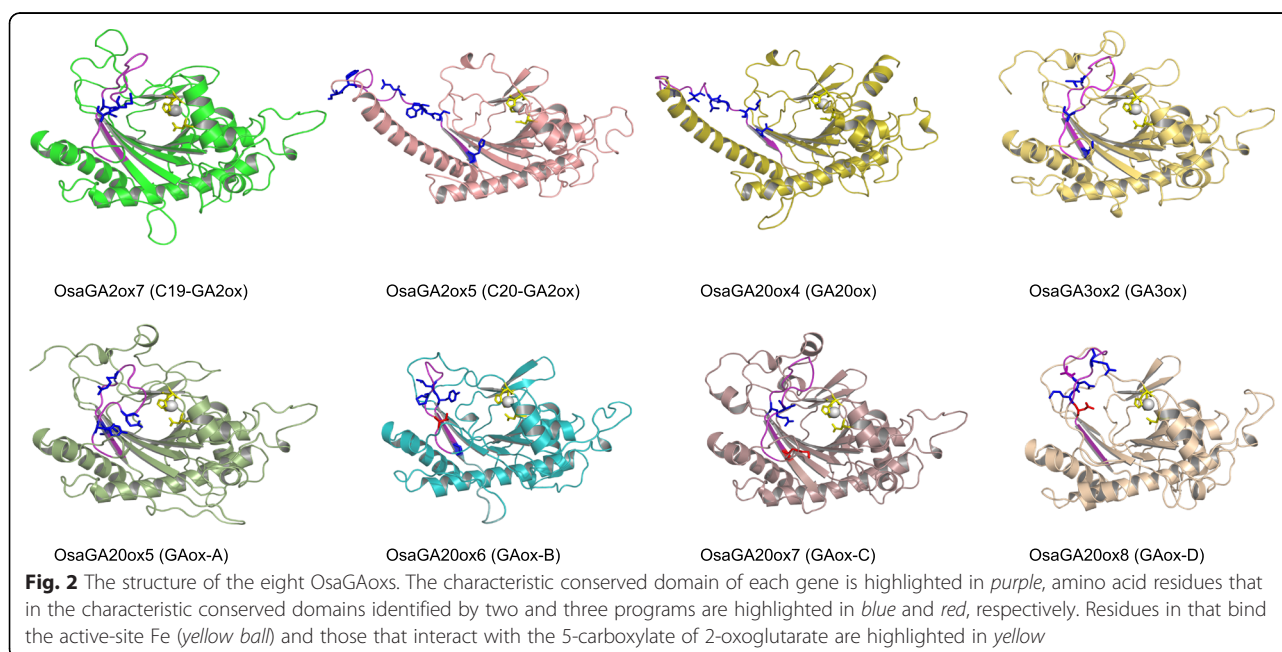
#### Phylogenetic analysis of plant GAox genes

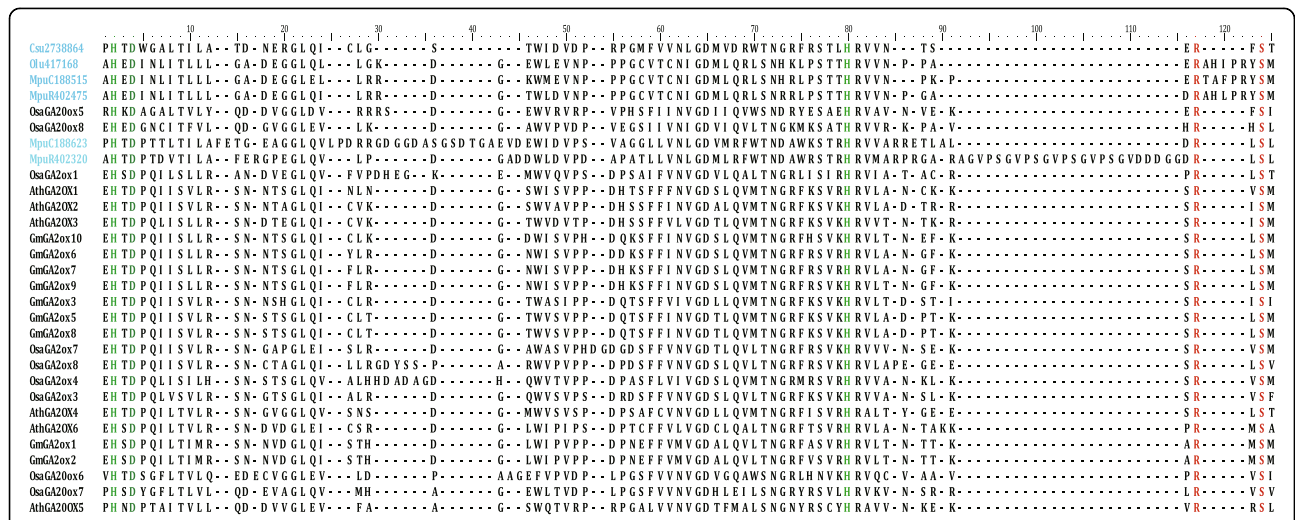
According to the phylogenetic analysis of 854 GAox homologs, the maximum-likelihood (ML) tree showed that all GAox homologs from green algae were clustered into a single clade (Alga clade), while those from land plants formed another main clade (Embryophyta clade), which was further divided into eight subclades with high statistical support, corresponding to the subfamilies C<sub>19</sub>-GA2ox, C<sub>20</sub>-GA2ox, GA20ox, GA3ox, GAox-A, GAox-B, GAox-C and GAox-D, respectively (Fig. 4a). Out of those, the latter four subfamilies corresponding to *OsGA20ox5*, *OsGA20ox6*, *OsGA20ox7* and *OsGA20ox8* reported in Han and Zhu's study (2011) were characterized for the first time. Each of these new subfamilies

(i.e., GAox-A, GAox-B, GAox-C and GAox-D) formed a monophyletic group, respectively (Fig. 4a), and they are phylogenetically distant from the GA20ox subfamily. GAox homologs from *Physcomitrella patens*, *Selaginella moellendorffii*, *Picea abies* and *Amborella trichopoda* resided in the basal lineages of the tree, which is consistent with previous studies on the phylogenetic position of these species. In addition, for angiosperm species, subfamilies C<sub>19</sub>-GA2ox, C<sub>20</sub>-GA2ox, and GA20ox exhibited a distinct binary branching pattern in the phylogenetic tree, which suggests that large scale genome duplications may have played a critical role in the evolution of these gene subfamilies.

#### Characteristic motifs of GAox subfamilies

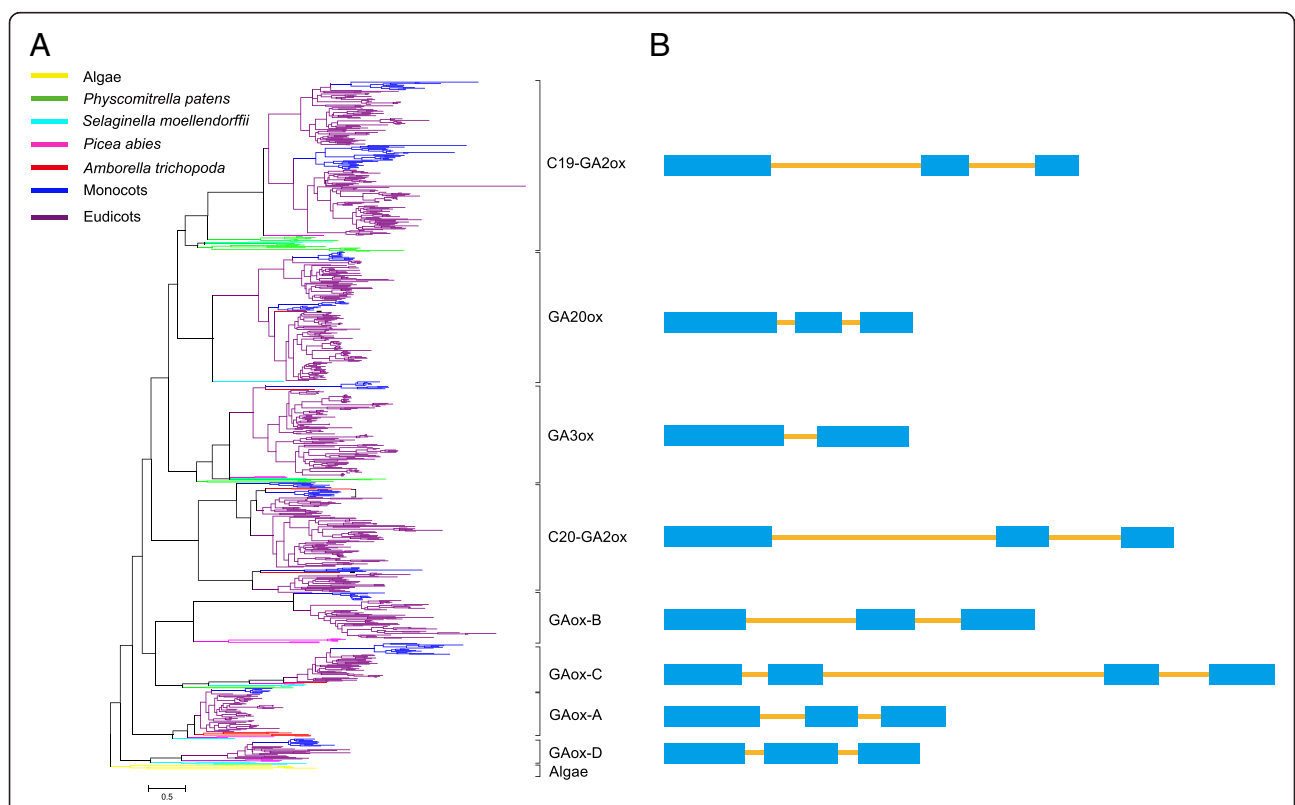
When we used MEME to identify and examine the conserved motifs of GAoxs, we found a total of 15 conserved motifs (Table 1). Among these 15 motifs, 13 were shared by most GAoxs, and motifs 1 and 2 represent the catalytic core of 2OGDs. Motifs 7, 9, 11 and 12, were found in different GAoxs. For example, motif 7 was absent from all the GA20ox and GA3ox genes, whereas motifs 9 and 11 were not observed in any of the GA2ox genes. Interestingly, motifs 7, 9 and 12 were all localized in N-terminals of the amino acid sequences. The





alignment of N-terminal amino acid sequences of all GAox gene homologs revealed eight types of gene sequence (Fig. 1 and Additional file 3: Data S1), which correspond to the C<sub>19</sub>-GA2ox, C<sub>20</sub>-GA2ox, GA20ox, GA3ox, GAox-A, GAox-B, GAox-C and GAox-D

subfamilies, respectively. Figure 1 shows that there are 16 to 21 amino acids in the characteristic motifs specific to the eight subfamilies of the GAox family, and the motifs of subfamilies GAox-A, GAox-B, GAox-C and GAox-D have more variable amino acid sites compared



**Fig. 4** Phylogenetic tree and gene structure of the GA oxidase in plants. **a** The maximum likelihood tree of 41 species' GA oxidase genes constructed on 854 deduced full-length peptide amino acid sequences and rooted with homologous sequences of algae. **b** The sketch map of GA oxidases structure: Exons and introns are indicated by blue boxes and yellow lines, respectively. For more details of the GA oxidase structures, see Additional file 4

**Table 1** Motif distribution in GA oxidases

Motif number	Length (aa)	Conserved sequence
motif1	30	PGAFVNVGDTLQALSNGRFKSVLHRVWN
motif2	18	HTDPILTILHQDQVGG
motif3	30	ARLWKACEEWGFFQVNVHGVPAELISRAE
motif4	19	FEESDILRLNHYPCCPEP
motif5	19	RLYRDFTWSEYLEFTQKHY
motif6	19	RLSMAYFLGPPLDKVISPL
motif7	30	GPPDPFGYGSKRIGPNDGVWLEYLLNTN
motif8	19	VVEEYCEAMKKLKLLLEL
motif9	14	RFSSKLPWKETLSF
motif10	30	DRFFALPLSQKQKAQRSPGECGYASAFIG
motif11	14	RADMNTLDAFSNWL
motif12	30	LSNGSYRWGTPATSLRQLSWSEAFHIPLT
motif13	19	SVKDYFRKTWGNDFEQFGK
motif14	19	SEAYREHPLHLKHIPLDF
motif15	10	LGDNRLGPFE

with the others. The subfamilies  $C_{19}$ -GA2ox,  $C_{20}$ -GA2ox, GA20ox and GA3ox have been studied previously, while GAox-A, GAox-B, GAox-C and GAox-D are proposed for the first time in this study; they include the newly named genes *OsGA2ox5*, *OsGA20ox5*, *OsGA20ox6*, *OsGA20ox7* and *OsGA20ox8* in rice, respectively [23].

Comparative gene structure analysis of 854 GAox gene sequences revealed that gene homologs of every GAox subfamily often have similar exon-intron structures, but the gene structure is more variable in subfamilies  $C_{19}$ -GA2ox and  $C_{20}$ -GA2ox, which had two kinds of these structures (Fig. 4b and Additional file 4: Figure S1). Most GAox gene subfamilies have three exons, but this is not the case in subfamilies GAox-C and GA3ox. The GA3ox genes have two exons, and the GAox-C genes generally have four exons. However, in monocots of subfamily GAox-C, three exons were also found in some GAox homologs.

#### Positive selection and functional divergence among GAox subfamilies

The phylogenetic trees were constructed based on the sequences of 46 monocots and two ACOs in order to investigate selection pressure among eight GAox subfamilies and selected amino acid sites in characteristic motifs. The LRTs for M2 vs. M1 ( $2\Delta L = 0$ ,  $p > 0.05$ ) suggested that the positive selection model (M2) was not significantly better than the nearly neutral model (M1). But models M3 and M8 fit the data significantly better than the null models M0 and M7 (for M3 vs. M0,  $2\Delta L = 1946.50$ ,  $p < 0.001$ ; for M8 vs. M7,  $2\Delta L = 244.99$ ,  $p < 0.001$ ), they identify several sites with an  $\omega$  value

significantly greater than 1. At the PP >95 % level, 13 amino acid sites were identified under positive selection by M8 (Table 2). As an exploratory analysis, the branch-site test suggested that all eight GAox subfamilies were under positive selection ( $p < 0.05$ ) (Fig. 5). Bayes Empirical Bayes (BEB) analysis showed that at the PP > 95 %, branch-site model A identified 2/13/29/8/1/12/2/29 sites as being potentially subjected to positive selection on the  $C_{19}$ -GA2ox/ $C_{20}$ -GA2ox/GA20ox/GA3ox/GAox-A/GAox-B/GAox-C/GAox-D branches, respectively.

The separate analyses of the eight subfamilies under the site models showed that the selective pressure could vary among amino acid sites in eight subfamilies by comparison of the M0 and M3 for all GAox subfamilies. The M3 detected no sites under positive selection in most of subfamilies, but both M1a against M2a, and M7 against M8 suggested the presence of sites under positive selection in each subfamily, except for the GAox-D. No sites were found in the GAox-D, possibly due to its small sample size (Additional file 1: Table S3).

The programs FunDi and GroupSim identified a number of sites with significant functional divergence scores for all subfamilies [60–62]. FunDi method predicted a wider range of site divergences, and the fractions of sites with posterior probability of functional divergence above 0.5 ranges from 0.798 to 0.987 across GAox subfamilies (Table 3). The GroupSim method detected 159 sites in all GA oxidases (Table 4).

In order to analyze the characteristics and functional divergence sites of each subfamily motif, we modeled the 3D structures of eight GA oxidases of *Oryza sativa* based on the higher reliability templates (Additional file 5: Table S1). According to the 3D protein structures, we found: 1) all the characteristic motifs locate in the crevice of GA oxidases; 2) each of the characteristic motifs consists of a loop and  $\beta$  sheet; 3) all the motifs detected have several positively selected or functional divergence sites by all or two programs (PAML 4.8, FunDi and GroupSim), and most of these sites are polar amino acids (Fig. 2). Those features suggested that the characteristic motif is an important component of GA oxidases and plays a significant role in the GA biosynthesis.

## Discussion

### Origin and evolution of GAox family

The present study showed that GAox genes, critical genes in regulating GA biosynthesis especially for the later steps of GA biosynthetic pathways, were ubiquitous in plants. Unlike the four previously well-studied subfamilies (i.e.,  $C_{19}$ -GA2ox,  $C_{20}$ -GA2ox, GA20ox and GA3ox), there were fewer homologs in the subfamilies GAox-A, GAox-B, GAox-C and GAox-D ranging from 30 to 63 (Additional file 2: Table S2). Homologs of subfamilies GAox-B and GAox-D genes were not found in the basal angiosperm *Amborella trichopoda*, but were



**Table 2** Summary statistics for detecting selection using branch-site and site models of PAML4.8

Model	<i>p</i>	<i>L</i>	Estimate of parameters	df	2Δ <i>L</i>	Positively Selected Sites <sup>a</sup>
Site-specific models						
M0: one-ratio	1	-33,700	$\omega$ : 0.160			None
M1a: nearly neutral	2	-33,148	$p_0 = 0.727, p_1 = 0.273, \omega_0 = 0.125, \omega_1 = 1.00$			Not allowed
M2a: positive selection	4	-33,148	$p_0 = 0.727, p_1 = 0.182, p_2 = 0.090; \omega_0 = 0.125, \omega_1 = 1.000, \omega_2 = 1.000$	2	0	
M3: discrete	5	-32,727	$p_0 = 0.212, p_1 = 0.594, p_2 = 0.194; \omega_0 = 0.038, \omega_1 = 0.180, \omega_2 = 1.934$	4	1946.50***	21,24,25,27,28 ~ 43,44,45,46,47,48,49,51 ~ 57, 60,103,105,106,134,136,137,138,243,279,325, 361,362,373,492,493,494,548,549,550 ~ 554, 556,557,558,560,561,562,563,566,567
M7: $\beta$	2	-32,705	$p = 0.642, q = 1.628$			Not allowed
M8: $\beta$ & $\omega$	4	-32,583	$p_0 = 0.858, p = 0.971, q = 4.418, p_1 = 0.142, \omega = 4.858$	2	244.99***	21,24,25,29,32,35,37,39,40,42,46,49,53,54, 56,57,103,134,362,493,550,551,563.
Branch-site model						
model A (C19-GA2ox)	4	-33,132	$p_0 = 0.509, p_1 = 0.189, p_2 = 0.302, \omega_2 = 8.762$	1	6.26*	393,520
model A (C20-GA2ox)	4	-33,129	$p_0 = 0.446, p_1 = 0.167, p_2 = 0.386, \omega_2 = 47.691$	1	15.83***	150,184,217,232,239,287,289,294,299, 383,519,531,535
model A (GA20ox)	4	-33,110	$p_0 = 0.442, p_1 = 0.163, p_2 = 0.395, \omega_2 = 999.000$	1	39.77***	65,71,81,114,148,149,159,161,172,190, 214,264,291,304,330,348,377,381,384, 385,417,437,447,471,477,507,515,519,524
model A (GA3ox)	4	-33,132	$p_0 = 0.466, p_1 = 0.175, p_2 = 0.358, \omega_2 = 7.881$	1	9.51**	156,158,163,232,288,339,505,525
model A (GAox-A)	4	-33,135	$p_0 = 0.556, p_1 = 0.205, p_2 = 0.238, \omega_2 = 13.489$	1	10.41**	340
model A (GAox-B)	4	-33,133	$p_0 = 0.464, p_1 = 0.174, p_2 = 0.363, \omega_2 = 39.904$	1	14.41***	197,231,343,352,353,354,416,444,469,471, 525,530
model A (GAox-C)	4	-33,136	$p_0 = 0.465, p_1 = 0.174, p_2 = 0.362, \omega_2 = 5.019$	1	3.97*	342,439
model A (GAox-D)	4	-33,124	$p_0 = 0.424, p_1 = 0.160, p_2 = 0.417, \omega_2 = 21.542$	1	17.55***	84,160,161,213,231,233,259,263,299,304, 310,335,339,341,347,352,364,366,423,427, 453,457,470,475,485,511,517,525,545

*p* is the number of free parameters in the  $\omega$  distribution

\*significant at *p* value < 0.05 level, \*\*significant at *p* value < 0.01 level, \*\*\*significant at *p* value < 0.001 level

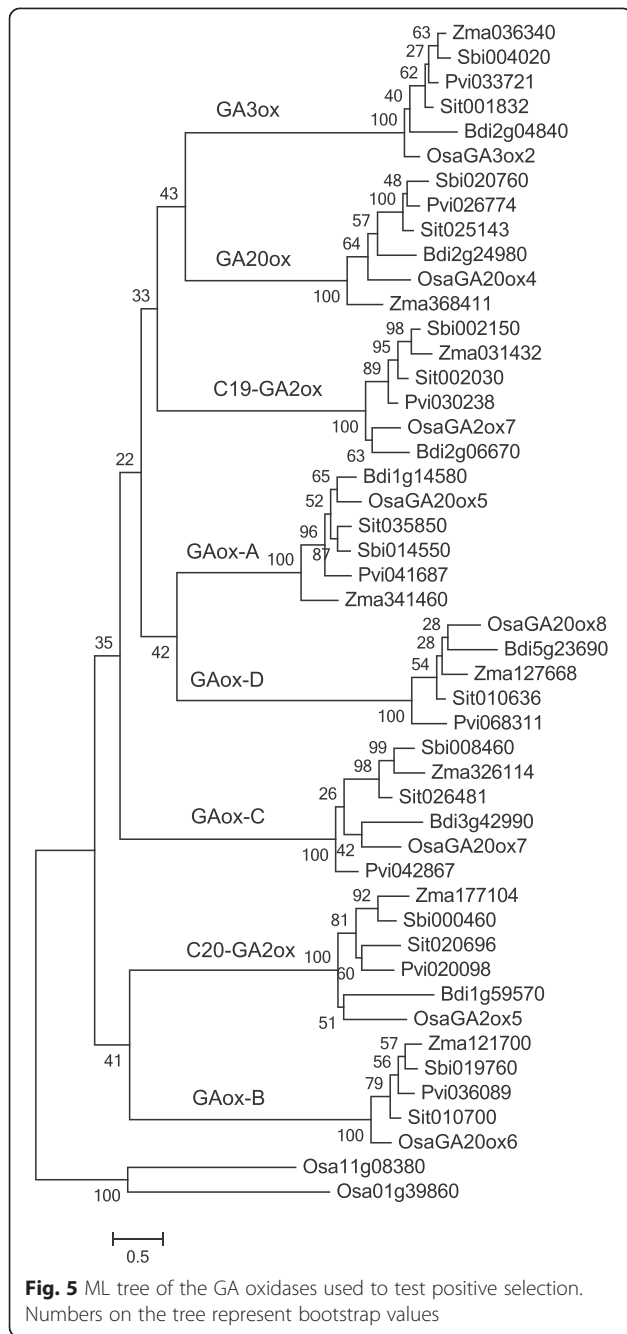
<sup>a</sup>positively selected sites identified under site and branch-site tests with PP  $\geq$  95 %

present in the gymnosperm *Picea abies*. Our phylogenetic analysis of the GAox genes showed that all six GAox homologs identified from green algae fell into a single clade, while those from land plants were clustered together and then divided into eight subclades (Fig. 4a). Previous studies found no bioactive GAs have been detected and chemically identified in mosses, nor could external GAs induce GA-specific developmental processes in these plants [27, 92, 93]. However, Anterola and Shanle (2008) found 15 putative GA oxidase homologs in *Physcomitrella patens*. This means that there are GA oxidases, but no bioactive GAs in moss species, which could be due to incompleteness of GA synthetic pathway. In this study, we found 18 putative GAox genes in *Physcomitrella patens*, belonging to three different subfamilies. These results indicate that diversification of GAox genes occurred before the divergence of land plants, then diversified and expanded further in the course of land plant evolution. Considering GAox genes in the main plant lineages, it is reasonable to believe that a rapid gene expansion occurred in the GAox family after the divergence of land plants from the common ancestor

of green plants; this may have been related to specialized metabolisms and adaptations to terrestrial environments.

According to our phylogenetic analysis, gene structure and the characteristic motifs of GAox genes, eight subfamilies were well defined under the GAox family of land plants, namely C<sub>19</sub>-GA2ox, C<sub>20</sub>-GA2ox, GA20ox, GA3ox, GAox-A, GAox-B, GAox-C and GAox-D. Of these subfamilies, C<sub>19</sub>-GA2ox, C<sub>20</sub>-GA2ox, GA20ox and GA3ox have been thoroughly studied previously, while new subfamilies GAox-A, GAox-B, GAox-C and GAox-D have not. According to our phylogenetic analysis, these four subfamilies resided in basal clades in the tree. All this information suggests that these subfamilies of genes were present in the common ancestral species of extant land plants, and these genes may have some unknown functions in land plants. Thus, the catalytic activity and biological function of these genes needs to be elucidated.

Although the structures and functional sites were predicted in our study, the function of GAox-A/B/C/D genes has not been examined yet. Therefore, it is unknown the location of these genes in GA biosynthesis



pathway. Combined with the findings of previous studies [11, 40, 92, 94], a proposed model of GAs biosynthesis pathway is presented in Fig. 6. Obviously, only very early steps of this pathway were present in *Physcomitrella patens*. Biochemically, P450s (cytochrome P450 monooxygenases) provide direct substrates or precursors of substrates for C<sub>20</sub>-GA2ox, so the presence of C<sub>20</sub>-GA2ox may be a prerequisite for the presence of the GAox family. As the GAox-A/B/C/D are basal clades of phylogenetic tree (Fig. 4), they may appear subsequently during the evolution of land plants. However, it is currently difficult to give a clear

**Table 3** Fraction of sites predicted to be functionally divergent by FunDi program

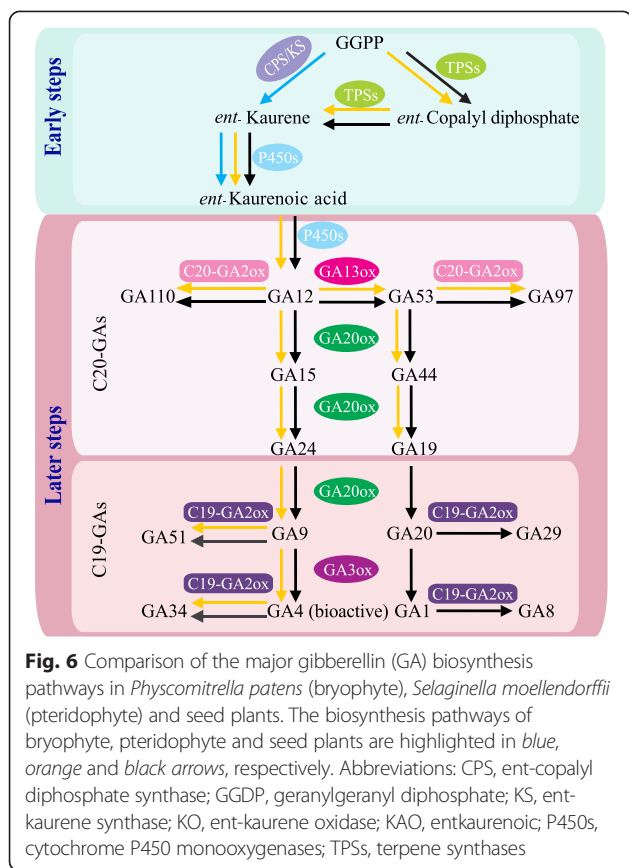
Data sets	FunDi (PP ≥ 95 %)
3ox Vs 20ox7	0.895522388
20ox5 Vs 20ox6	0.987641607
20ox Vs 20ox6	0.96789424
3ox Vs 20ox6	0.970978441
C19 Vs 20ox6	0.962149886
20ox6 Vs 20ox7	0.954333644
C20 Vs 20ox6	0.980555556
C19 Vs 3ox	0.941255908
C19 Vs 20ox7	0.914821124
C19 Vs 20ox	0.928341385
C19 Vs 20ox8	0.960144928
3ox Vs 20ox8	0.926456543
20ox8 Vs 20ox7	0.876175549
20ox Vs 20ox8	0.899563319
20ox8 Vs 20ox6	0.981299213
20ox5 Vs 20ox8	0.798290598
C20 Vs 20ox8	0.960727969
C20 Vs 3ox	0.94165536
C20 Vs 20ox	0.931743421
20ox Vs 20ox7	0.864295125
3ox Vs 20ox	0.897810219
C20 Vs 20ox5	0.952252252
C19 Vs 20ox5	0.930674264
3ox Vs 20ox5	0.926660915
20ox5 Vs 20ox7	0.808544304
20ox Vs 20ox5	0.879332478

explanation for the evolutionary history of GAox genes due to lack of information of causal relationship between active GAs and GAox-A/B/C/D.

Despite having similar catalytic cores of 2OGDs (Fig. 3), the GAox homologs of green algae exhibited some differences to those of the nonvascular plant *Physcomitrella patens* and vascular plants. They were grouped into one clade in the phylogenetic tree, and not classified into any of GAox subfamilies. In addition, there are no characteristic motifs of eight GAox subfamilies. These findings suggest that putative algal GAox genes have little diversification of sequences and functions.

**Table 4** Number of sites predicted to be functionally divergent by GroupSim program

Posterior probability	ln characteristic motifs	ln GAox family
0.99 ≤ PP	13	127
0.5 ≤ PP < 0.99	1	32



### Diversification of gene structure

Previous studies have shown that gene structural diversity is a possible mechanism for the evolution of multigene families [95–97]. The GAox family provides an example of exon-intron diversity in the evolution of gene family, although the protein sequences of its catalytic domain are well conserved. Comparing the phylogenetic tree with the exon-intron structure (Fig. 4 and Additional file 4: Figure S1), it is apparent that the most closely related members within subfamilies shared similar gene structure in terms of intron number and exon length, suggesting that the diversification of subfamilies could be under a strong selection and indicative of their functional conservativeness. Most of the GAox subfamilies have three exons, also suggesting that the diversification of the GAox family is under a strong selection. However, two recent diverged GAox subfamilies, GA3ox and GAox-C possess two and four exons, respectively; this supports the hypothesis that duplications of ancestral mosaic genes have been followed by more recent gains and losses of introns [98]. In view of the numbers and locations of introns, GAox genes may have experienced different evolutionary histories since the monocot-dicot split. Unlike the gene structure of dicots, that of monocots is usually simple, without or with one/two shorter introns, suggestive of the fact that these genes

probably arose from retrotransposon-based random insertions. This might indicate different evolutionary patterns for these two kinds of plants. Most homologs in the three GAox subfamilies (i.e. C<sub>19</sub>-GA2ox, C<sub>20</sub>-GA2ox and GAox-C subfamily) with relatively longer introns may have arisen from unequal crossing-over [23].

To elucidate the motif diversification of GAox subfamilies, putative motifs were predicted using MEME and alignments (Table 1). As expected, most of the closely related members had common motif compositions, and the length and location of GAox gene domains were highly conserved, suggesting similar functions among members of the same GAox subfamily and the same selection acting on them. All proteins of the eight subfamilies have very similar C-terminal structures of their catalytic domains (Fig. 3 and Additional file 3: Data S1), the function of most putative motifs remains unknown as they do not have homologs in the Pfam and SMART databases. Most of the motifs were shared by the majority of GAoxs, but the short amino acid stretches in the N-terminal may characterize each subfamily. Combined with the previous studies of GAox [35–39, 45, 46], we conclude that the alignment of amino acid sequence from 691 to 757 sites can be used to identify a putative protein of the GAox family (Fig. 1 and Additional file 3: Data S1). In addition, these motifs can be used to distinguish subfamily C<sub>20</sub>-GA2ox from C<sub>19</sub>-GA2ox [20]. According to the results of 3D modeling analysis, all characteristic motifs were located in the crevice of the protein surface (Fig. 2). In addition, the positive selection tests suggest that most of these motifs contain several sites that could have evolved in response to positive selection pressure. To date, no functional examination has been conducted on these conserved motifs, so our results provide evidence that these specific motifs may have some active sites and play critical roles in GA biosynthesis and catabolism.

Based on our findings, the similar gene structures and conserved motifs in each GAox subfamily provide an insight into the evolutionary history of the GAox family and its classification. In addition, the variation in gene structure and the difference in motif compositions among different subfamilies indicate that GAox subfamilies are functionally diversified. Furthermore, the eight conserved motifs are the diagnostic characters for the identification of GAox subfamilies, and they must be important for catalytic functions in the GA biosynthetic pathway.

### Function divergence and adaptation of the GAox family

Gene duplication is considered to be a major mechanism for the generation of evolutionary novelty and adaptation [50, 51]. In plants, gene duplication followed by functional divergence is particularly important for the diversification of biochemical metabolites [99–102]. Substitutions can change the function of duplicated genes,

and may be due to either a relaxation of purifying selection or the action of positive selection [56, 58, 59, 103]. To investigate evidence for positive selection in the GAox subfamilies, we analyzed the eight branches by site models and branch-site models (Fig. 5 and Additional file 1: Table S3). Further analyses on each of the subfamilies under the site models (M0 vs M3; M1a vs M2a; M7 vs M8) detected several positively selected sites for all the subfamilies except for the GAox-D, suggesting that the selective pressure varies among lineages and amino acid sites in GA oxidases. The biochemical context of substitutions that was under positive selection is consistent with a scenario involving the adaptive evolution of GAox genes. Under branch-site models, some selected sites (PP > 95 %) were found to be scattered throughout the primary sequences (data not shown), and they were folded in the large cavity of three-dimensional structures of the analyzed protein. Among these sites, one to eight positively selected sites (PP > 95 %) were located in or near the characteristic motifs, in all branches except C<sub>19</sub>-GA2ox and GAox-A. The mutations at these sites among the different subfamilies suggested their importance for the enzymatic activity in GA biosynthesis and catabolism. Further studies using site-directed mutagenesis are needed to determine whether these selected sites confer an ability on GAox genes to discriminate between different substrate types.

It should be noted that our branch-site model analyses might suffer from uncertainty because of limitations of the codon models, such as changes in equilibrium nucleotide and codon frequencies among clades with deep divergence (Additional file 6: Table S5) [85, 104, 105]. Therefore, we further performed FunDi and GroupSim to investigate the divergence of the distribution of site-rates. Because these methods are more appropriate for the high sequence divergences, they could provide an important cross-check of the codon-based results and confirmed that the GA oxidases underwent functional divergence. Despite this, the signal for positive selection contributing to the divergence of GA oxidases should be taken with caution considering that the codon and amino acid models can sometimes be negatively affected by model misspecification and alignments quality [83, 86, 106–110]. Taken together, our study provides new insights into the functional divergence of the GAox family and implicates the potential roles of positive selection in the evolution of GAox genes, which deserves further investigations.

## Conclusions

This study provides the first large-scale evolutionary analysis of GAox genes in plants, using an extensive whole-genome dataset of 41 species representing green

algae, bryophytes, pteridophytes and seed plants. Our study not only provides information about the classification of GAox genes, but also facilitates the further functional characterization and analysis of GA oxidases. Our results indicate: 1) Gibberellin oxidase (GAox) originated very early—before the divergence of bryophytes and the vascular plants, and diversified into eight subfamilies in the course of land plant evolution. Of these subfamilies, GAox-A, GAox-B, GAox-C and GAox-D are proposed for the first time; 2) the diversification of GAox genes could be attributed to functional divergence and such divergence is most likely to facilitate the completion of the GA synthesis pathway for plants to adapt to terrestrial environments; 3) each subfamily of GAox genes has its characteristic motif and a signature of positive selection have been detected in most subfamilies.

## Availability of supporting data

Additional file 3-Data S1 and Additional file 7-Data S2. Protein sequences used in the present analysis.

Additional file 8-Table S4. GenBank accession numbers for sequences used in this study.

## Additional files

**Additional file 1: Table S3.** MLEs of parameters and sites inferred to be under positive selection for the eight GAox subfamilies data set. (XLSX 15 kb)

**Additional file 2: Table S2.** GA oxidases in the species analyzed in this study. (XLSX 19 kb)

**Additional file 3: Data S1.** Alignment of putative GA oxidase sequences. (FASTA 1761 kb)

**Additional file 4: Figure S1.** The GA oxidase structure. (PDF 1521 kb)

**Additional file 5: Table S1.** List of templates used in this study. (XLSX 10 kb)

**Additional file 6: Table S5.** A. Frequencies of nucleotides; B. Frequencies of codons; C. Nucleotide frequencies by codon sites. (RAR 24 kb)

**Additional file 7: Data S2.** The aligned (protein-coding) DNA sequences (CDS). (RAR 238 kb)

**Additional file 8: Table S4.** The accession numbers of the protein sequences used in the present analysis. (XLSX 34 kb)

## Abbreviations

ACOs: 1-aminocyclopropane-1-carboxylate; CPS: Ent-copalyl diphosphate synthase; GA: Gibberellin; GAox: GA oxidase; GGDP: Geranylgeranyl diphosphate; KAO: Entkaurenoic; KO: Ent-kaurene oxidase; KS: Ent-kaurene synthase; LRT: Likelihood ratio test; P450s: Cytochrome P450 monooxygenases; TPSs: Terpene synthases; 2OGD: 2-oxoglutarate-dependent dioxygenase.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

GYR, SG, XW and YH designed the experiments. YH and XW performed the experiments. GYR, SG, YH and XW analyzed the data. YH and GYR wrote the manuscript. All authors read and approved the final manuscript.

## Acknowledgments

We are grateful to Dr. Ping-Li Liu of Beijing Forestry University for her technical assistance. We thank Dr. Hua Gao of Peking University for his helps in the structural modeling analysis. Our deep gratitude also goes to the editor and

anonymous reviewers for their valuable comments and constructive suggestions on the manuscript. This work is supported by the National Natural Science Foundation of China (Grant No. 30990240).

#### Author details

<sup>1</sup>College of Life Sciences, Peking University, Beijing 100871, China. <sup>2</sup>State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China.

Received: 4 May 2015 Accepted: 17 September 2015

Published online: 29 September 2015

#### References

- Jones RL, Varner JE. The bioassay of gibberellins. *Planta*. 1966;72:155–61.
- Pharis RP, King RW. Gibberellins and reproductive development in seed plants. *Annu Rev Plant Physiol*. 1985;36:517–68.
- Hookey R. Gibberellins: perception, transduction and responses. *Plant Mol Biol*. 1994;26:1529–55.
- Blázquez MA, Green R, Nilsson O, Sussman MR, Weigel D. Gibberellins promote flowering of *Arabidopsis* by activating the *LEAFY* promoter. *Plant Cell*. 1998;10:791–800.
- MacMillan J. Biosynthesis of the gibberellin plant hormones. *Nat Prod Rep*. 1997;14:221–43.
- MacMillan J. Occurrence of gibberellins in vascular plants, fungi, and bacteria. *J Plant Growth Regul*. 2001;20:387–442.
- Hedden P, Phillips AL. Gibberellin metabolism: new insights revealed by the genes. *Trends Plant Sci*. 2000;5:523–30.
- Graebe JE. Gibberellin biosynthesis and control. *Annu Rev Plant Physiol*. 1987;38:419–65.
- Olszewski N, Sun TP, Gubler F. Gibberellin signaling: biosynthesis, catabolism, and response pathways. *Plant Cell*. 2002;14:561–80.
- Sakamoto T, Miura K, Itoh H, Tatsumi T, Ueguchi-Tanaka M, Ishiyama K, et al. An overview of gibberellin metabolism enzyme genes and their related mutants in rice. *Plant Physiol*. 2004;134:1642–53.
- Yamaguchi S. Gibberellin metabolism and its regulation. *Annu Rev Plant Biol*. 2008;59:225–51.
- Phillips AL, Ward DA, Uknes S, Appleford NEJ, Lange T, Huttly AK, et al. Isolation and expression of three gibberellin 20-oxidase cDNA clones from *Arabidopsis*. *Plant Physiol*. 1995;108:1049–57.
- Kawai Y, Ono E, Mizutani M. Evolution and diversity of the 2-oxoglutarate-dependent dioxygenase superfamily in plants. *Plant J*. 2014;78:328–43.
- Carrera E, Jackson SD, Prat S. Feedback control and diurnal regulation of gibberellin 20-oxidase transcript levels in potato. *Plant Physiol*. 1999;119:765–74.
- Itoh H, Tanaka-Ueguchi M, Kawaide H, Chen X, Kamiya Y, Matsuoka M. The gene encoding tobacco gibberellins 3beta-hydroxylase is expressed at the site of GA action during stem elongation and flower organ development. *Plant J*. 1999;20:15–24.
- Oikawa T, Koshioka M, Kojima K, Yoshida H, Kawata M. A role of *OsGA20ox1*, encoding an isoform of gibberellin 20-oxidase, for regulation of plant stature in rice. *Plant Mol Biol*. 2004;55:687–700.
- Davidson SE, Swain SM, Reid JB. Regulation of the early GA biosynthesis pathway in pea. *Planta*. 2005;222:1010–9.
- Schomburg FM, Bizzell CM, Lee DJ, Zeevaert JA, Amasino RM. Overexpression of a novel class of gibberellin 2-oxidases decreases gibberellin levels and creates dwarf plants. *Plant Cell*. 2003;15:151–63.
- Thomas SG, Phillips AL, Hedden P. Molecular cloning and functional expression of gibberellin 2-oxidases, multifunctional enzymes involved in gibberellin deactivation. *Proc Natl Acad Sci U S A*. 1999;96:4698–703.
- Lee DJ, Zeevaert JAD. Molecular cloning of *GA2-oxidase3* from spinach and its ectopic expression in *Nicotianaglossyvestris*. *Plant Physiol*. 2005;138:243–54.
- Pimenta Lange MJ, Liebrandt A, Arnold L, Chmielewska SM, Felsberger A, Freier E, et al. Functional characterization of gibberellin oxidases from cucumber, *Cucumis sativus* L. *Phytochemistry*. 2013;90:62–9.
- Yang Y, Zhang F, Ge S. Evolutionary rate patterns of the Gibberellin pathway genes. *BMC Evol Biol*. 2009;9:206.
- Han F, Zhu B. Evolutionary analysis of three gibberellin oxidase genes in rice, *Arabidopsis*, and soybean. *Gene*. 2011;473:23–35.
- Giacomelli L, Rota-Stabelli O, Masuero D, Acheampong AK, Moretto M, Caputi L, et al. Gibberellin metabolism in *Vitis vinifera* L. during bloom and fruit-set: functional characterization and evolution of grapevine gibberellin oxidases. *J Exp Bot*. 2013;64:4403–19.
- Johri MM. Possible origin of hormonal regulation in green plants. *Proc Indian Natl Sci Acad*. 2004;70:335–65.
- Johri MM. Hormonal regulation in green plant lineage families. *Physiol Mol Biol Plants*. 2008;14:23–38.
- Yasumura Y, Crumpton-Taylor M, Fuentes S, Harberd NP. Step-by-step acquisition of the gibberellin-DELLA growth-regulatory mechanism during land-plant evolution. *Curr Biol*. 2007;17:1225–30.
- Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, et al. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science*. 2008;319:64–9.
- Hayashi K, Kawaide H, Notomi M, Sakigi Y, Matsuo A, Nozaki H. Identification and functional analysis of bifunctional-karene synthase from the moss *Physcomitrella patens*. *FEBS Lett*. 2006;580:6175–81.
- Anterola A, Shanle EK. Genomic insights in moss gibberellin biosynthesis. *The Bryologist*. 2008;111:218–30.
- Hedden P, Phillips AL, Rojas MC, Carrera E, Tudzynski B. Gibberellin biosynthesis in plants and fungi: a case of convergent evolution? *J Plant Growth Regul*. 2001;20:319–31.
- Tudzynski B. Gibberellin biosynthesis in fungi: genes, enzymes, evolution, and impact on biotechnology. *Appl Microbiol Biotechnol*. 2005;66:597–611.
- Pimenta Lange MJ, Lange T. Gibberellin biosynthesis and the regulation of plant development. *Plant Biol*. 2006;8:281–90.
- Aravind L, Koonin EV. The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol*. 2001;2:7.7–8.
- Chiang HH, Hwang I, Goodman HM. Isolation of the *Arabidopsis* GA4 locus. *Plant Cell*. 1995;7(2):195–201.
- Xu YL, Li L, Wu K, Peeterst Anton JM, Gage DA, Zeevaert JAD. The GA5 locus of *Arabidopsis thaliana* encodes a multifunctional gibberellin 20-oxidase: molecular cloning and functional expression. *Proc Natl Acad Sci U S A*. 1995;92(14):6640–4.
- Wu K, Li L, Gage DA, Zeevaert JAD. Molecular cloning and photoperiod-regulated expression of gibberellin 20-oxidase from the long-day plant spinach. *Plant Physiol*. 1996;110(2):547–54.
- Toyomasu T, Kawaide H, Sekimoto H, von Numer C, Phillips AL, Hedden P, et al. Cloning and characterization of a cDNA encoding gibberellin 20-oxidase from rice (*Oryza sativa*) seedlings. *Physiol Plant*. 1997;99(1):111–8.
- Kang HG, Jun SH, Kim J, Kawaide H, Kamiya Y, An G. Cloning and molecular analyses of a gibberellin 20-oxidase gene expressed specifically in developing seeds of watermelon. *Plant Physiol*. 1999;121(2):373–82.
- Lo SF, Yang SY, Chen KT, Hsing YI, Zeevaert JAD, Chen LJ, et al. A novel class of gibberellin 2-oxidases control semidwarfism, tillering, and root development in rice. *Plant Cell*. 2008;20:2603–18.
- Lange T, Hedden P, Graebe JE. Expression cloning of a gibberellin 20-oxidase, a multifunctional enzyme involved in gibberellin biosynthesis. *Proc Natl Acad Sci U S A*. 1994;91(18):8552–6.
- Huerta L, Garcia-Lor A, Garcia-Martinez JL. Characterization of gibberellin 20-oxidases in the citrus hybrid Carrizo citrange. *Tree Physiol*. 2009;29(4):569–77.
- Busov VB, Meilan R, Pearce DW, Ma C, Rood SB, Strauss SH. Activation tagging of a dominant gibberellin catabolism gene (*GA2-oxidase*) from poplar regulates tree stature. *Plant Physiol*. 2003;132:1283–91.
- Itoh H, Ueguchi-Tanaka M, Sentoku N, Kitano H, Matsuoka M, Kobayashi M. Cloning and functional analysis of two gibberellin 3β-hydroxylase genes that are differently expressed during the growth of rice. *Proc Natl Acad Sci U S A*. 2001;98(15):8909–14.
- Sakamoto T, Kobayashi M, Itoh H, Tagiri A, Kayano T, Tanaka H, et al. Expression of a gibberellin 2-oxidase gene around the shoot apex is related to phase transition in rice. *Plant Physiol*. 2001;125:1508–16.
- Lee DJ, Zeevaert JAD. Differential regulation of RNA levels of gibberellin dioxygenases by photoperiod in spinach. *Plant Physiol*. 2002;130:2085–94.
- Lester DR, Ross JJ, Davies PJ, Reid JB. Mendel's stem length gene (*Le*) encodes a gibberellin 3 beta-hydroxylase. *Plant Cell*. 1997;9:1435–43.
- Martin DN, Proebsting WM, Hedden P. Mendel's dwarfing gene: cDNAs from the *Le* alleles and function of the expressed proteins. *Proc Natl Acad Sci U S A*. 1997;94(16):8907–11.
- Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. *Nature Rev Genet*. 2010;11:97–108.

50. Semon M, Wolfe KH. Consequences of genome duplication. *Curr Opin Genetics Dev.* 2007;17:505–12.
51. Sun HZ, Ge S. Molecular evolution of the duplicated TFIIA genes in Oryzaeae and its relatives. *BMC Evol Biol.* 2010;10:128.
52. Ohno S. Evolution by gene duplication. London: George Allen & Unwin Ltd. Berlin, Heidelberg and New York: Springer; 1970.
53. Li W-H. Accelerated evolution following gene duplication and its implications for the neutralist-selectionist controversy. In: Otha T, Aoki K, editors. Population genetics and molecular evolution. Tokyo: Japan Scientific Press; 1985. p. 333–52.
54. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science.* 2000;290:1151–5.
55. Liu PL, Wan JN, Guo YP, Ge S, Rao GY. Adaptive evolution of the chrysanthemyl diphosphate synthase gene involved in irregular monoterpene metabolism. *BMC Evol Biol.* 2012;12:214.
56. Zhang J, Rosenberg HF, Nei M. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci U S A.* 1998;95:3708–13.
57. Shiu SH, Byrnes JK, Pan R, Zhang P, Li WH. Role of positive selection in the retention of duplicate genes in mammalian genomes. *Proc Natl Acad Sci U S A.* 2006;103:2232–6.
58. Dykhuizen D, Hartl DL. Selective neutrality of 6pgd allozymes in *E. COLI* and the effects of genetic background. *Genetics.* 1980;96(4):801–17.
59. Zhang J. Evolution by gene duplication: an update. *Trends Ecol Evol.* 2003;18:292–8.
60. Gaston D, Susko E, Roger AJ. A Phylogenetic Mixture Model for the Identification of Functionally Divergent Protein Residues. *Bioinformatics.* 2011;27:2655–63.
61. Capra JA, Singh M. Characterization and prediction of residues determining protein functional specificity. *Bioinformatics.* 2008;24:1473–80.
62. Gaston D, Roger AJ. Functional divergence and convergent evolution in the plastid-targeted glyceraldehyde-3-phosphate dehydrogenases of diverse eukaryotic algae. *PLoS One.* 2013;8(7):e70396.
63. Valegård K, van Scheltinga ACT, Lloyd MD, Hara T, Ramaswamy S, Perrakis A, et al. Structure of a cephalosporin synthase. *Nature.* 1998;394:805–9.
64. Altschul SF, Madden TL, Schäffer AA, Zhang JH, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 1997;25:3389–402.
65. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 1999;41:95–8.
66. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–7.
67. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 1997;25:4876–82.
68. Xia X, Xie Z. DAMBE: software package for data analysis in molecular biology and evolution. *J Hered.* 2001;92:371–3.
69. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 2006;22:2688–90.
70. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics.* 2011;27:1164–5.
71. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37:W202–8.
72. Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, et al. CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res.* 2005;33:D192–6.
73. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* 2003;31:3381–5.
74. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 2014;42:W252–8.
75. Benkert P, Kunzli M, Schwede T. QMEAN server for protein model quality estimation. *Nucleic Acids Res.* 2009;37:W510–4.
76. Seeliger D, de Groot BL. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J Comput Aided Mol Des.* 2010;24:417–22.
77. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, et al. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature.* 2007;450:203–18.
78. Löytynoja A, Goldman N. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A.* 2005;102:10557–62.
79. Löytynoja A, Goldman N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science.* 2008;320:1632–5.
80. Markova-Raina P, Petrov D. High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 *Drosophila* genomes. *Genome Res.* 2011;21:863–74.
81. Penn O, Privman E, Ashkenazy H, Landan G, Graur D, Pupko T. GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic Acids Res.* 2010;38 suppl 2:W23–8.
82. Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24:1586–91.
83. Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol.* 2000;17:32–43.
84. Bielawski JP, Yang Z. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J Mol Evol.* 2004;59:121–3.
85. Bay RA, Bielawski JP. Inference of functional divergence among proteins when the evolutionary process is non-stationary. *J Mol Evol.* 2013;76(4):205–15.
86. Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol.* 2005;22:2472–9.
87. Yang Z, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol.* 2002;19:908–17.
88. Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics.* 1998;148(3):929–36.
89. Yang Z, Swanson WJ. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol Biol Evol.* 2002;19(1):49–57.
90. Yang Z, Wong WSW, Nielsen R. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* 2005;22:1107–18.
91. Wang HC, Li K, Susko E, Roger AJ. A class frequency mixture model that adjusts for site-specific amino acid frequencies and improves inference of protein phylogeny. *BMC Evol Biol.* 2008;8(1):331.
92. Hirano K, Nakajima M, Asano K, Nishiyama T, Sakakibara H, Kojima M, et al. The GID1-mediated gibberellin perception mechanism is conserved in the lycophyte *Selaginella moellendorffii* but not in the bryophyte *Physcomitrella patens*. *Plant Cell.* 2007;19:3058–79.
93. Bomke C, Tudzynski B. Diversity, regulation, and evolution of the gibberellin biosynthetic pathway in fungi compared to plants and bacteria. *Phytochemistry.* 2009;70:1876–93.
94. Evans MMS, Poethig RS. Gibberellins promote vegetative phase change and reproductive maturity in maize. *Plant Physiol.* 1995;108(2):475–87.
95. Cho G, Doolittle RF. Intron distribution in ancient paralogs supports random insertion and not random loss. *J Mol Evol.* 1997;44:573–84.
96. Boudet N, Aubourg S, Toffano-Nioche C, Kreis M, Lecharny A. Evolution of intron/exon structure of DEAD helicase family genes in *Arabidopsis*, *Caenorhabditis*, and *Drosophila*. *Genome Res.* 2001;11:2101–14.
97. Babenko VN, Rogozin IB, Mekhedov SL, Koonin EV. Prevalence of intron gain over intron loss in the evolution of paralogous gene families. *Nucleic Acids Res.* 2004;32:3724–33.
98. Trotman CAN. Introns-early: Slipping lately? *Trends Genet.* 1998;14:132–4.
99. Pichersky E, Gang DR. Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci.* 2000;5(10):439–45.
100. Ober D. Seeing double: gene duplication and diversification in plant secondary metabolism. *Trends Plant Sci.* 2005;10(9):444–9.
101. Benderoth M, Textor S, Windsor AJ, Mitchell-Olds T, Gershenzon J, Kroymann J. Positive selection driving diversification in plant secondary metabolism. *Proc Natl Acad Sci U S A.* 2006;103(24):9118–23.
102. Kroymann J. Natural diversity and adaptation in plant secondary metabolism. *Curr Opin Plant Biol.* 2011;14(3):246–51.
103. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics.* 1999;151(4):1531–45.
104. Zhang J. Frequent false detection of positive selection by the likelihood method with branch-site models. *Mol Biol Evol.* 2004;21:1332–9.
105. Bielawski JP, Yang Z. Maximum likelihood methods for detecting adaptive evolution after gene duplication. *J Struct Funct Genomics.* 2003;3:201–12.

106. Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. *TREE*. 2000;15:496–503.
107. Wong WS, Yang Z, Goldman N, Nielsen R. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics*. 2004;168:1041–51.
108. Jordan G, Goldman N. The effects of alignment error and alignment filtering on the sitewise detection of positive selection. *Mol Biol Evol*. 2012;29:1125–39.
109. Privman E, Penn O, Pupko T. Improving the performance of positive selection inference by filtering unreliable alignment regions. *Mol Biol Evol*. 2012;29:1–5.
110. Levy Karin E, Susko E, Pupko T. Alignment errors strongly impact likelihood-based tests for comparing topologies. *Mol Biol Evol*. 2014;31:3057–67.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

