Dutartre *et al. BMC Evolutionary Biology* 2012, **12**:64
http://www.biomedcentral.com/1471-2148/12/64

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Phylogenomics of the benzoxazinoid biosynthetic pathway of Poaceae: gene duplications and origin of the *Bx* cluster

Leslie Dutartre, Frédérique Hilliou and René Feyereisen*

Abstract

Background: The benzoxazinoids 2,4-dihydroxy-1,4-benzoxazin-3-one (DIBOA) and 2,4-dihydroxy-7- methoxy-1, 4-benzoxazin-3-one (DIMBOA), are key defense compounds present in major agricultural crops such as maize and wheat. Their biosynthesis involves nine enzymes thought to form a linear pathway leading to the storage of DI(M) BOA as glucoside conjugates. Seven of the genes (*Bx1-Bx6* and *Bx8*) form a cluster at the tip of the short arm of maize chromosome 4 that includes four P450 genes (*Bx2-5*) belonging to the same *CYP71C* subfamily. The origin of this cluster is unknown.

Results: We show that the pathway appeared following several duplications of the *TSA* gene (α-*subunit of tryptophan synthase*) and of a Bx2-like ancestral *CYP71C* gene and the recruitment of *Bx8* before the radiation of Poaceae. The origins of *Bx6* and *Bx7* remain unclear. We demonstrate that the Bx2-like *CYP71C* ancestor was not committed to the benzoxazinoid pathway and that after duplications the *Bx2-Bx5* genes were under positive selection on a few sites and underwent functional divergence, leading to the current specific biochemical properties of the enzymes. The absence of synteny between available Poaceae genomes involving the *Bx* gene regions is in contrast with the conserved synteny in the *TSA* gene region.

Conclusions: These results demonstrate that rearrangements following duplications of an *IGL/TSA* gene and of a *CYP71C* gene probably resulted in the clustering of the new copies (*Bx1* and *Bx2*) at the tip of a chromosome in an ancestor of grasses. Clustering favored cosegregation and tip chromosomal location favored gene rearrangements that allowed the further recruitment of genes to the pathway. These events, a founding event and elongation events, may have been the key to the subsequent evolution of the benzoxazinoid biosynthetic cluster.

Keywords: Plants, Gene duplication, P450, Neofunctionalization, Gene cluster, Secondary metabolism

Background

Plants are sessile organisms which have evolved chemical and mechanical ways to defend against pathogens, herbivores and competitors. The synthesis of toxic compounds, generally arising from the so-called secondary metabolism is a hallmark of plant defense. Among the enzymes often involved in secondary metabolism and in particular in the synthesis of defense compounds and toxins in plants are the P450 enzymes [1]. These are heme-dependent oxidase enzymes that generally catalyze the insertion of one oxygen atom in a substrate after activation of molecular oxygen. The most common reaction catalyzed by this protein family is hydroxylation, but P450s are involved in a wide variety of catalyses such as dimerizations, isomerizations, dehydratations or reductions [2,3]. P450 proteins represent a large protein family very well represented in plants. For example 272 cytochrome P450 genes (CYP genes) are present in the Arabidopsis genome, including 26 pseudogenes [3]. This superfamily groups together proteins with as low as 20% sequence identity. Nevertheless secondary and tertiary structures are well conserved throughout the family. For instance P450 proteins share some conserved structures and sequences linked to properties such as oxygen or heme binding.

* Correspondence: rfeyer@sophia.inra.fr
Institut National de la Recherche Agronomique, UMR 1355 Institut Sophia Agrobiotech, Centre National de la Recherche Scientifique, UMR 7254, Université de Nice Sophia Antipolis, Sophia-Antipolis, France

In grasses, P450s of the CYP71C subfamily are involved in the biosynthesis of the cyclic hydroxamic acids 2,4-dihydroxy-1,4-benzoxazin-3-one (DIBOA) and 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA). These natural compounds are well described as natural pesticides and allelochemicals, as chemical defense against microbial diseases and herbivory [4,5]. Their occurrence depends on the plant species [6]. For example, the predominant cyclic hydroxamic acid in rye *Secale cereale* is DIBOA whereas the main cyclic hydroxamic acid in maize *Zea mays* is DIMBOA, that differs from DIBOA by an additional methoxy group [5]. The presence of cyclic hydroxamic acids in grain crops has been known for over 50 years [7]. The four P450 genes involved in DIBOA biosynthesis, *Bx 2* to *Bx5* (known as *CYP71C1* to *CYP71C4* in the P450 nomenclature), were first described in maize by Frey et al. [8,9]. The final reaction steps leading to DIMBOA-glucoside were also elucidated in maize [10-12]. The full pathway involves 9 enzymes (BX1 to BX9) thought to act sequentially in the synthesis of DIMBOA-glucoside from indole-3-glycerol phosphate (Figure 1). The *Bx2* to *Bx5* genes are clustered on the short arm of maize chromosome 4 [8,9]. Genetic analysis indicated that *Bx1*, *Bx6*, *Bx7* and *Bx8* are close to, or within this cluster, thus grouping genes of different families within a short chromosomal region [8-10,12]. Upon wounding, an additional *O*-methylation is activated, leading to HDMBOA [13,14] but the gene responsible for this reaction is still unknown. The same DI(M)BOA biosynthetic pathway has also been described in wheat *Triticum aestivum*, in rye *S. cereale* and in the wild barley *Hordeum lechleri*, the cultivated barley probably having lost the gene cluster [15-17]. The four *CYP71C* genes were cloned and characterized in diploid and hexaploid wheat and in wild barley [17-19].

A common evolutionary origin of this cluster of maize P450 genes by successive gene duplications has been proposed [8,16,20]. While it is no longer surprising to find biosynthetic gene clusters in bacteria and fungi [21], the nature and origin of such clusters in plants and animals is less studied. Large gene families such as the CYP family are often characterized by multiple gene duplications that leave a genomic trace as clusters of related genes [22,23]. However, these structural clusters such as the 13 *CYP71B* genes clustered on chromosome 3 of *Arabidopsis thaliana*, the 16 *CYP6* genes clustered in the mosquito, or the 22 *CYP2* loci clustered on mouse chromosome 7 [22,24,25] are not known to be co-regulated or to participate in a common pathway. Biosynthetic gene clusters are therefore different because they comprise non-homologous genes that function collectively. They have received increasing attention in plants [26], where known biosynthetic gene clusters serve in elaborating defense compounds such as phytoalexins from common precursors. They include the clusters for the biosynthesis of thalianol in *A. thaliana*, avenacin in *Avena strigosa*, momilactone and phytocassane in *Oryza sativa* [27] as well as the cyanogenic glucoside biosynthetic clusters in *Lotus japonicus*, *Manihot esculenta* and *Sorghum bicolor* [28].

The *Bx* gene cluster of maize is therefore of great interest, because it consists of an apparent structural cluster of four CYP71C genes in close genomic proximity with members of other gene families that, together, are known to direct the synthesis of DIMBOA glucoside from a common metabolic intermediate, indole 3-glycerol phosphate [20]. The sequence of the gene duplications, the nature of the ancestral genes and the mechanisms leading to the recruitment of several genes from different families into this biosynthetic cluster have not been determined in detail, yet these are the key questions in understanding the evolution of secondary metabolic gene clusters in plants [26,27].

Here, we take advantage of the information from newly sequenced genomes of Poaceae and of the known biochemical properties of the enzymes to describe the evolutionary origin of the DIMBOA biosynthetic pathway. We used a phylogenetic approach to establish the sequence of duplications of an ancestral *CYP71C* gene leading to the *Bx2-Bx5* cluster in maize. We delineated the involvement of critical amino acids in achieving the current biochemical specificities of the P450 enzymes. We studied the syntenic relationships of genes at the interface between primary metabolism and the DIMBOA pathway: *ZmBx1, ZmTSA* and *TSA- like, ZmIgl* and *Igl-like*. We also searched for synteny around the other genes of the *Bx* cluster to gain insights into the origin of the entire DIMBOA pathway.

Methods

Sequence data

The BX2-5 protein and transcript sequences from maize [8] were used for BLAST approaches (blastn, blastp and tblastn). Sequences with more than 50% amino acid sequence identity with one maize CYP71C were used to identify equivalent CYP71C sequences in other Poaceae. Searches were made on the BLAST server of NCBI [29] for all Poaceae, on the maize genome website [30], on the *Brachypodium distachyon* website [31] and on the Phytozome database [32]. P450 sequences were manually checked and their annotation corrected when necessary based on known P450 motifs [3]. Incomplete sequences, and pseudogenes were removed from further analysis, resulting in 75 sequences representing the CYP71C subfamily in Poaceae. Five *A. thaliana* sequences were chosen to root the tree (CYP76C3, CYP76C7, CYP76C4, CYP76C1 and CYP76G1). For the studies on *Bx1* (and *Igl, TSA, TSA_like), Bx6, Bx7, Bx8 and Bx9*, sequences

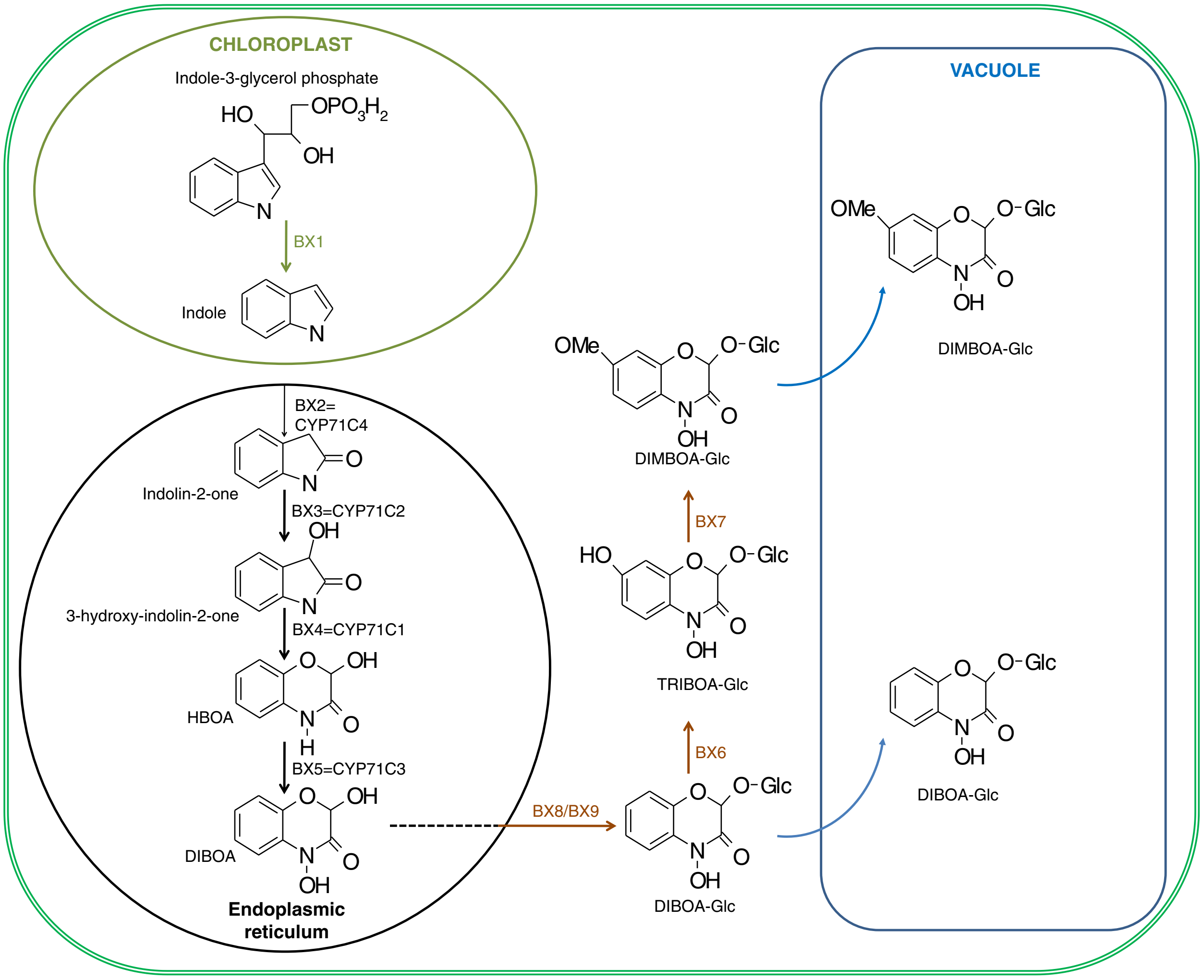

Figure 1 The DIMBOA biosynthetic pathway and cellular compartment localizations. Adapted from [8,12,20].

were searched by BLASTP based on maize sequences with the same criteria (id% > 50) and on the same databases as for the P450 study. Intron/exon structures were determined whenever the genomic sequence was available.

Phylogenetic tree reconstruction

The selection of the best-fit model of protein evolution for the CYP71C subfamily was done with ProTest software version 2.4 [33] based on protein alignment obtained with MUSCLE software (available on the website phylogeny.fr [34]) using the default parameters [35-37]. No curation was needed because of the high identity between sequences and the correct alignment of P450 motifs. The JTT model [38] was then chosen as the best fit model of protein evolution for our sequences. The tree of the CYP71C subfamily in Poaceae was generated on the phylogeny.fr website [34] using maximum likelihood as reconstruction method (PhyML program). Bootstrapping (100 iterations) was done to document branch support. These criteria were also used for the reconstruction of the other BX phylogenetic trees.

Positive selection in coding sequences

The codeml program of the PAML software package was used to detect positive selection in the CYP71C subfamily [39]. cDNA sequences were aligned based on protein alignment by using the RevTrans 1.4 server [40]. We used branch and site models which allow ω to vary among branches in the phylogeny or among sites on the alignment [41,42]. First, we compared the one-ratio model (M0, one ω estimated for all sites) to the free-ratio model (independent ω ratios for each branch) to test for the hypothesis of variable ω among branches. Then, the M7 (beta distribution of ω ratios, with $0 < \omega < 1$) *versus* M8 (extension of M7 with a supplementary site class with $\omega > 1$ estimated from the dataset) comparisons were done to test for positive selection among sites.

Finally, the models M8 and M8a (ω = 1) were compared to determine if for a small fraction of sites the ω estimated under M8 was significantly higher than 1 [43,44]. Likelihood-ratio tests (LRT) were used to compare models. Twice the log-likelihood difference 2ΔlnL was compared with a χ^2 distribution with degrees of freedom corresponding to the difference of free parameters numbers between the two models compared. The branch-site model MA was further applied to our data to detect positive selection affecting only a few sites on prespecified lineages [45]. Four independent branches were studied with this model and Likelihood-ratio tests (LRT) were used to compare models with Bonferroni's multiple testing corrections.

P450 secondary structure

P450 proteins generally present 13 conserved α-helices named from A to L [46]. The maize Bx2-5 proteins were analyzed using tools available on the web (Jpred 3 [47], Porter [48]) to define the consensus zones corresponding to these putative helices. Substrate recognition sites (SRSs) were localized by homology to the SRSs described by Gotoh [49].

Functional divergence analysis of BX2 to BX5 protein clades

The software DIVERGE2 (**D**etecti**n**g **V**ariability in **E**volutionary **R**ates among **Ge**nes) was used to identify critical amino acid residues involved in functional innovations after gene duplications [50]. The coefficients of type I and type II functional divergence (θI and θII, respectively) between two chosen clades were calculated. If these coefficients are superior to 1, it means that some amino acids were subject to altered selective constraints (type I functional divergence) or that a radical shift of amino acids physicochemical properties occurred after gene duplication and/or speciation (type II functional divergence) [50-53] . Thirty nine sequences comprising each BX clade and the sequences associated to the BX2 clade were aligned using MUSCLE (default parameters). This alignment and the equivalent portion of the tree were used as input parameters for analyses with DIVERGE2 software. We compared BX clades resulting from duplication events to each other. For each test, the posterior probability of each site to be under functional divergence was calculated. Sites with a posterior probability Qk > 0.67 (to select only radical cluster-specific sites [50-53]) were localized on the maize BX2-5 alignment.

BX2 molecular modeling and dockings

A model for BX2 was generated with the **i**terative **t**hreading **assem**bly **r**efinement server (I- TASSER [54]) [55,56] using the following templates: (pdb numbers) 3k9v, 3e6i, 3na0, 2hi4, 3czh and 1izo. We obtained a model with a C-score = −1.46, indicating a correct prediction. The top I-TASSER templates all presented a normalized Z-score > 1 with a large protein coverage (always superior to 76%), reflecting the high accuracy of the alignment. Predictions of binding sites were confident with a BS-score > 0.5 for all predictions. Dockings were assessed using the AutoDock4 software package [57,58]. Proteins were first prepared by removing water molecules, checking for missing atoms, adding of non-bonded hydrogens and computed Gasteiger charges under the AutoDockTool ADT 1.5.4 [59]. The protein model was then used to construct a grid box with a grid-point spacing of 0.375 Å. To dock the indole substrate, the input protein was the heme-containing Bx2. The grid centre position was positioned on top of the heme and included 40x40x40 = 64,000 points. Four independent Lamarckian genetic algorithm searches (250,000 and 2,500,000 evaluations) were run. For each analysis, the solutions were compared to determine if the results were reliable and reproducible. Interacting residues were also compared to determine the conserved residues in contact with indole.

Synteny of the Bx genes

Three approaches were used to analyze the synteny of maize *Bx* genes within available Poaceae genomes. We used the SyMAP v3.4 (**Sy**nteny **M**apping and **A**nalysis **P**rogram) to identify and display genome synteny alignments for *Z. mays, O. sativa, B. distachyon, S. bicolor* and *S. italica* genomes [60]. The Plant-Synteny site was also used to study synteny between *Z. mays*, *Triticum*, *O. sativa*, *S. bicolor* and *B. distachyon* [61]. Maizesequence.org was used to identify putative syntenies between maize and *O. sativa* and *S. bicolor.*

Results

Phylogenetic analysis of the genes related to Bx2-5 in the CYP71C subfamily

We aligned all available CYP71C sequences and reconstructed a phylogenetic tree to determine the evolutionary origin of the P450 genes of the *Bx* cluster in Poaceae (Figure 2 and Additional file 1). The tree we obtained showed that the maize BX2-5 sequences had orthologs distributed in four branches, boxed in Figure 2 as BX2 -Bx5 clades. Each of these four branches was strongly supported (bootstrap values 91–100%). The four clades contained sequences of biochemically characterized CYP71C enzymes such as those of maize [8,16] and wheat [19] as well as two sequences encoding P450s from rice that have not yet been biochemically characterized. Moreover, while the BX3, 4 and 5 clades were monophyletic, the BX2 clade was included in a parent clade that contained fourteen "BX2-like" sequences from

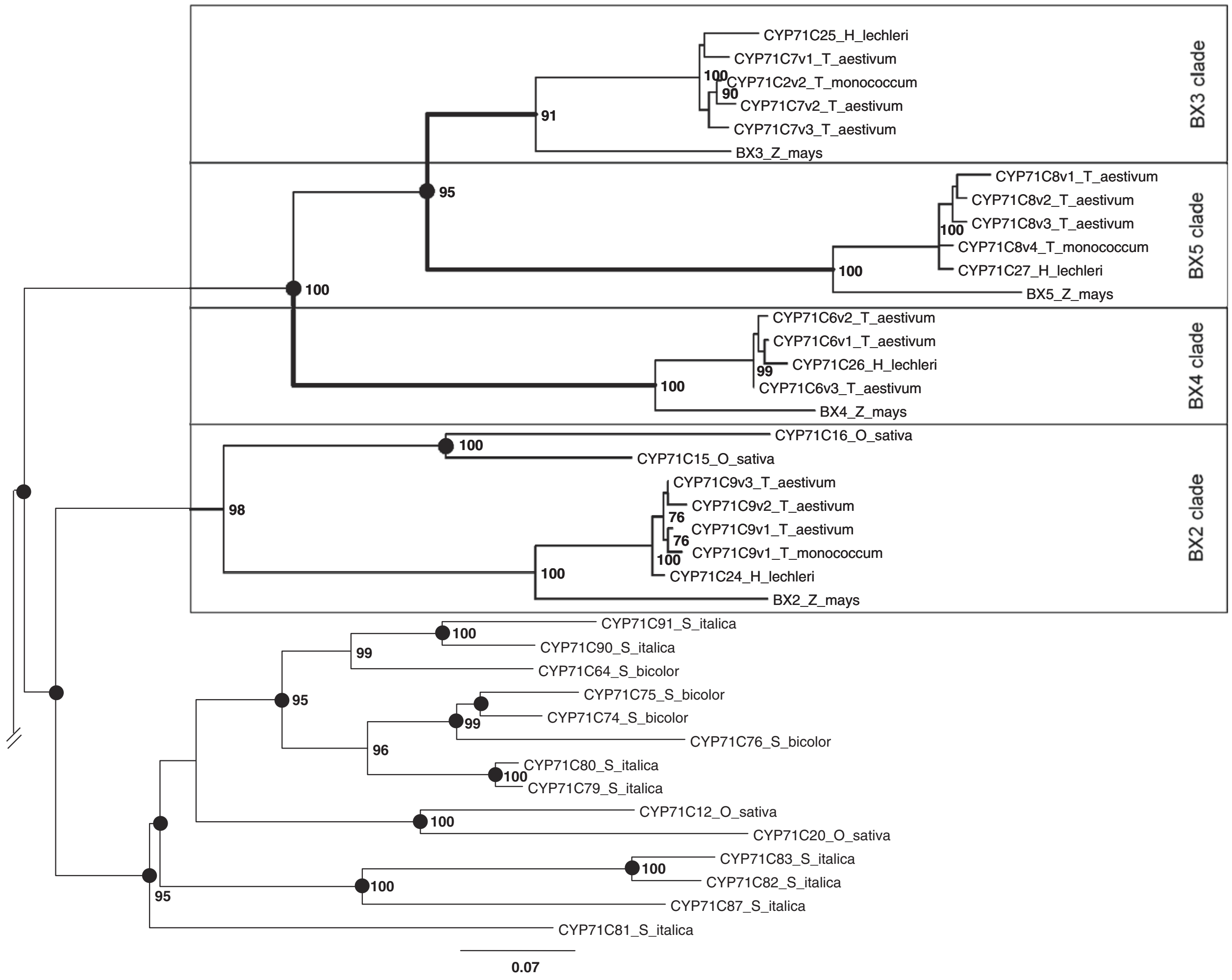

Figure 2 Phylogenetic tree of CYP71C sequences rooted with five CYP76 sequences of *Arabidopsis thaliana*. Maximum likelihood tree with branch lengths drawn to scale in terms of the number of substitutions per site. Bold branches indicate branches with ω > 1 according to the free ratio model of codeml. BX clades are boxed. Dots represent gene duplications. Only nodes with bootstrap values superior to 75% are shown. The full tree is shown in Additional file 1.

non-benzoxazinoid producers: *S. bicolor, S. italica* and *O. sativa*. The rest of the tree was composed of at least four branches with more distant sequences of the CYP71C subfamily from maize, *B. distachyon, S. italica, S. bicolor* and *O. sativa.*

Multiple intron losses in the CYP71C subfamily

Frey *et al*. in their landmark paper on the Bx cluster [8] suggested that the position of introns in the CYP71C genes was indicative of a common evolutionary origin. We therefore located the position of introns in all the available genomic CYP71C sequences (Additional file 1). Only two intron positions were found, both in phase zero, and located at conserved positions, 192 (intron 1) and 336 (intron 2) of BX2 (hereafter all amino acid positions are given in terms of equivalent positions of maize BX2). Thirty four genes had both common introns while twenty five sequences had only intron 2. All genes in the BX3, BX4 and BX5 clades had both introns, except for maize BX4 that had only intron 2. The three genomic sequences of the BX2 clade had only intron 2. Intron 2 is an ancestral intron found in most plant P450s [22], while the conserved position of intron 1 suggested that it had the same origin among all the sequences. The presence or absence of the introns did not follow a simple phylogeny. This indicated independent intron losses or gains in the different branches. Identification of paralog/ortholog relationships in the phylogeny allowed us to place duplication and speciation events on the tree. Accordingly, the most likely hypothesis is that intron 1 was introduced after the first duplication in the CYP71C subfamily and that more than a dozen independent intron losses have occurred. The alternative hypothesis of an equally high number of independent gains of intron 1 occurring at the same position and in the same phase is unlikely.

Positive selection in the CYP71C coding sequences

We analyzed the evolution of the *CYP71C* subfamily and in particular the evolution of the maize *Bx2-5* genes. We hypothesized that, after duplication and/or speciation events, the four *CYP71C* genes were subjected to selection. On one hand, the proteins maintained their P450 structure and overall catalytic activity (redox partner binding, dioxygen binding and activation), and on the other hand, they each acquired a distinct and high substrate specificity [8,16]. The conserved features of the four proteins i.e. helical structures, SRS regions, and conserved P450 motifs are illustrated in Figure 3. This balance of P450 structure conservation and substrate specificity would imply that genes are globally under purifying or neutral selection ($\omega < 1$ or $\omega = 1$ respectively), explaining the high sequence identity between proteins, and that some particular sites are under strong positive selection ($\omega > 1$), leading to the substrate specificity. This hypothesis was evaluated by testing for positive selection among specific lineages, among sites and among specific sites in specific lineages on our phylogeny. Branch and site models were first used to test the hypothesis of heterogeneous levels of selection among lineages and to test for positive selection among sites. The branch model would allow us to test if positive selection existed among the various branches. The site models permit ω to vary among sites. As the positively selected sites in proteins are generally very few, focusing on the overall sequence would fail to detect these selected sites as the ω value on the overall alignment would be inferior to 1. These tests would thus allow us to detect sites under positive selection in the CYP71C subfamily. The changes at these sites being favored during evolution would be potentially important for the specificity of the proteins. The comparison between the one-ratio model M0 (one ω ratio for all sites calculated from the data) and the free-ratio model (independent ω ratios for each branch) revealed a heterogeneous selection level among lineages ($2\Delta\ln L = 725.58$, $p < 0.01$). The three branches that define the BX3, 4 and 5 clades (Figure 2) showed ω values superior to 1 supporting the hypothesis of adaptive evolution while the branch defining the BX2 clade did not. The test of positive selection among sites, M7/M8 was statistically significant ($2\Delta\ln L = 680.30$, $p < 0.01$). This suggested that the ω ratio was variable among sites and that about 17% of the sites were under positive selection with $\omega = 1.69$. The M8/M8a test was also significant, meaning that the estimated ω in M8 was statistically different from 1. The M8 model identified 52 sites under positive selection with a posterior probability superior to 0.95 (Additional file 2). We also used the branch-site model MA to go further into the specific evolution of each group of the 4 P450s of the DIMBOA-biosynthetic pathway. The use of this model allows the detection of positively selected sites on prespecified lineages. The phylogenetic tree shows the duplicated origin of the four genes, represented by the four Bx clades on Figure 2. Our aim was to detect positive selection that affected sites in each of the four clades by specifying the foreground branches during the tests. Bonferroni's corrections for multiple testing were done. The test returned sites under positive selection (posterior probablility >0.5) in the BX2 and BX4 clades but the associated ω2 was only statistically superior to 1 ($\alpha = 0.10$) for BX2 sequences ($\omega 2 = 6.44$). Omitting sites in the membrane anchor not involved in substrate specificity, three sites were identified for BX2, including one site in the E/F loop and another one in SRS5 (Figure 3).

Functional divergence analyses of the BX P450 enzymes

Tests of functional divergence with the DIVERGE2 software were realized to identify critical amino acid residues in each of the four BX clades. Functionally divergent sites may explain the substrate specificity of each protein. Thirty nine protein sequences, including the sequences of the four BX clades and the BX2-like sequences were aligned and the equivalent portion of the tree were used for analyses of type I and type II functional divergences among BX clades. Clades of BX proteins originating from gene duplications were compared to each other. We also compared the BX2 clade *versus* the larger (BX4 (BX3/BX5) clade and the BX4 clade *versus* the (BX3/BX5) clade. Sites with significant posterior probability $Qk > 0.67$ were localized on the maize BX2-5 protein alignment (Figure 3).

Comparisons with proteins from the BX4 clade showed no type I or II functional divergence. The comparison of the BX2 clade with the (BX4(BX3/BX5)) clade returned a statistically significant type I functional divergence ($\theta I = 0.151 \pm 0.067$; LRT = 5.10; $p > 0.05$). Met 77 of maize BX2, localized within the first P450 motif, was returned with $Qk > 0.67$. The type I functional divergence test between the BX2 and BX3 clades was statistically significant ($\theta I = 0.353 \pm 0.084$; LRT = 17.610; $p > 0.05$) and identified 5 sites. One site was on P450 motif 1, one in SRS1, one in loop E/F, another one in SRS2, and the last one was after SRS5.

Functional divergence of type II was also found between these two clades ($\theta II = 0.140 \pm 0.055$). Twenty sites were identified, among which 14 amino acid residues differed between maize BX3 and BX2.

The type I functional divergence test between the BX2 and BX5 clades was statistically significant ($\theta I = 0.294 \pm$

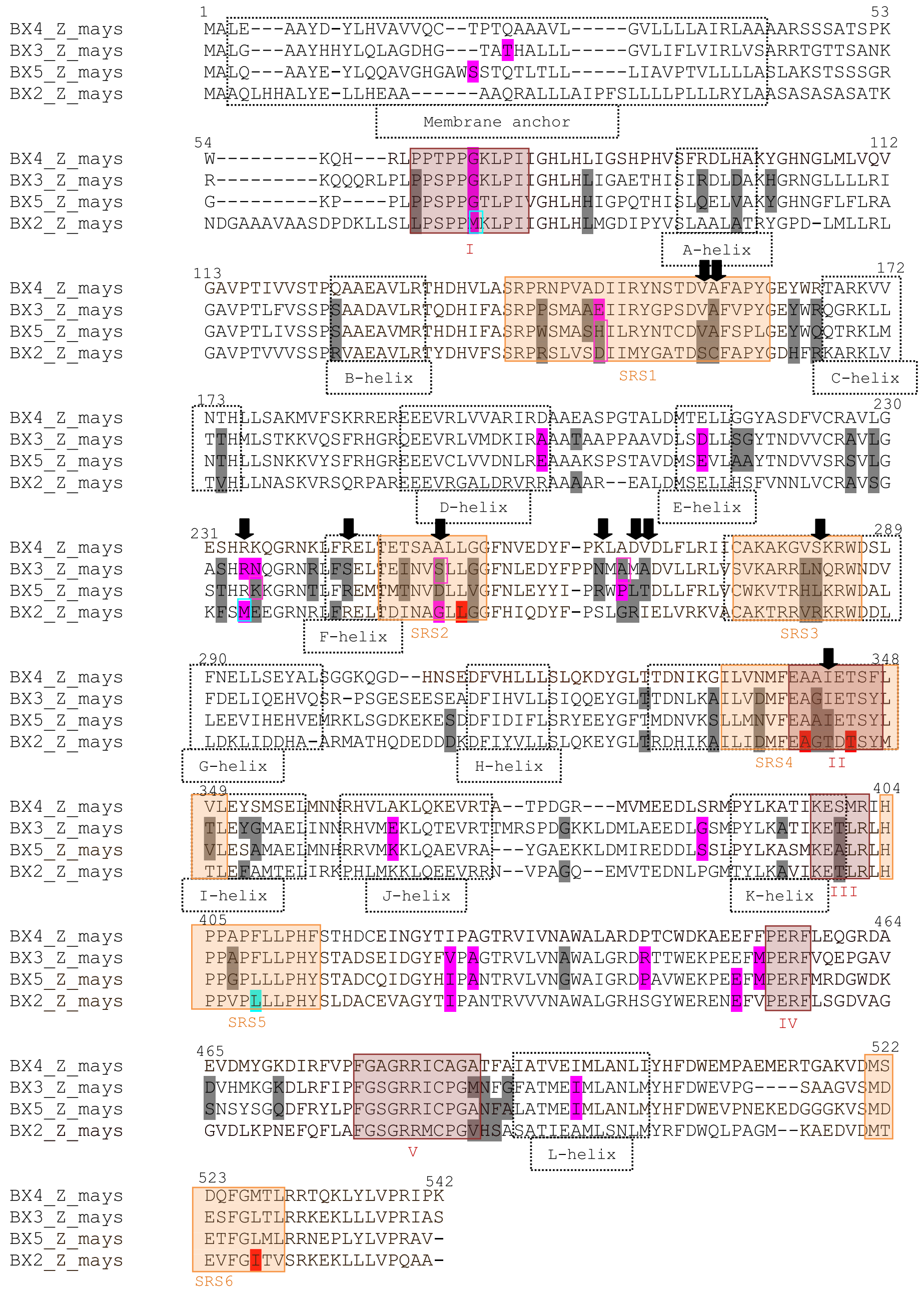

Figure 3 Maize BX2-5 protein alignment. P450 motifs are indicated in purple and the 6 SRSs localizations are indicated in orange. Sites potentially under positive selection according to the MA model from codeml analyses are in blue. Sites under type I functional divergence are in grey and sites under type II functional divergence are in pink. The 13 putative helices are marked in dotted boxes. Arrows indicate ten potentially important sites for the substrate specificity of each protein. Red highlighted sites represent the four residues (Leu253, Ala341, Thr345 and Ile527) in contact with indole according to docking results (AutoDock4).

0.097; LRT = 9.216; $p > 0.05$) with one site (Glu 451) at Qk > 0.67. The type II functional divergence test (θII = 0.214 ± 0.054) returned twenty five sites. In the comparison of the BX3/BX5 clades, functional divergence of type I (θI = 0.417 ± 0.066; LRT = 39.688; $p > 0.05$) was found on 7 sites. All these sites differed between maize BX3 and BX5, suggesting a radical shift in the rate of evolution for these amino acids. The type II functional divergence test returned a θII equal to 0.222 ± 0.0356 and identified 25 sites which all differed between maize BX3 and BX5. They were placed throughout the protein alignment with two sites in SRS1, three sites in SRS2, two sites in SRS3, three sites in SRS4, one in SRS5, two in the E/F loop and two in the F/G loop.

Comparisons of maize BX2-BX5 sites subjected to positive selection and to functional divergence

Positions on the maize BX2-5 alignment with at least one site under positive selection and/or functional divergence were checked. Among them, we focused on sites showing either amino acid conservation or four different amino acids. These sites could be involved in the specific evolution of the maize BX2-BX5 functions. Thirty two positions were extracted. Among them, 23 involved one or more radical biochemical changes, potentially leading to substrate specificity. Nine sites were identified when focusing on sites within SRSs or in the F and G helices or in the F/G loop (Figure 3). One site was added to these critical amino acids: the positively selected and under type II functional divergence Met 234 of BX2 just before the F- helix. This site seems to be very important from a selection point of view as it implies a radical biochemical shift from positively charged to non-polar.

Maize BX2 modeling and indole docking

Homology modeling was used to understand the relationship between sites under positive selection and/or functional divergence and protein substrate specificity. The model was also used in docking approaches to gain insight into the specific interactions of the protein with its substrate indole. Indole was docked into the heme-containing BX2 protein model and the residues in contact with indole in the 40 computed docked conformations were compared. Four protein residues were repeatedly found, namely Leu 253 on 35/40 conformations, Ile 527 on 36/40 conformations and Ala 341, Thr 345 and the heme molecule on all docked conformations. The relative position of the indole secondary amine was oriented towards Ala 341 and Ile 527 on 31/40 conformations. The four protein residues (Figures 3 and 4A) are all in SRS regions, and three were identical in the other BX P450s. The fourth residue, the apolar Ile 527 was replaced by a polar Thr in BX4 and BX3 and by an apolar Met in BX5. When looking at the spatial localization of all sites previously identified as under positive selection and/or under functional divergence, two sites, Ser 156 and Cys 157, were oriented toward the active site. Both sites were identified above among the ten sites potentially important for the substrate specificities (Figures 3 and 4B).

Synteny of the Bx2-5 genes and their homologs among Poaceae

We analyzed the synteny of the maize *Bx2-5* genes to determine whether it was conserved in the genomes of other Poaceae (Figure 5). Although synteny blocks were detected between maize chromosome 4 and *O. sativa*, *S. bicolor*, *B. distachyon* and *S. italica* genomes, none of them contained genes homologous to the *Bx2-5* genes. All the *CYP71C* homologs included in our phylogenetic analysis from those species were found in other genome locations. The cluster formed by *S. italica CYP71C88*, *C89* and *C92* for instance, was syntenic to the cluster of maize *CYP71C36*, *C56* and *C57* on chromosome 2 (Figure 5A) and phylogenetically distant from the *Bx* cluster. Moreover, non-*Bx* genes in the vicinity of the *Bx2-5* cluster did not show conservation of synteny. Therefore, in all currently available genomes, the position of the *Bx2-5* genes is unique to maize.

We also studied the phylogeny and syntenic relationships of the other maize *Bx* genes (*ZmBx*) involved in the DIMBOA-biosynthetic pathway. Their possible *Bx* orthologs in other Poaceae genomes were located on chromosomes (or scaffolds for *S. italica*) and their intron positions were mapped.

Phylogeny and synteny of the ZmBx1, ZmTSA, ZmIgl and ZmIgl_like genes

The *Bx1* and *Igl* genes are thought to result from duplications of the *TSA* gene [62]. The twenty nine most closely related sequences from *Z. mays, H. lechleri, T. aestivum, S. bicolor, S. italica, B. distachyon, O. sativa* and *Hordeum vulgare* were analyzed. On the phylogenetic tree, IGL/BX1 and TSA/TSA_like sequences formed two distinct clades. In the IGL/BX1 clade, ZmBX1 and *T. aestivum* BX1 were grouped together (Figure 6). The IGL sequences from Panicoideae were also closely related to BX1. A single branch included IGL from Pooideae and Ehrhartoideae. At first glance, these sequences followed the Poaceae phylogenetic history but this did not explain the origin of BX1 paralogs in *T. aestivum,* which belongs to the Pooideae (Figure 7). A likely explanation is that two duplication events of an *Igl* ancestor occurred before the separation of both Pooideae-Ehrhartoideae and Panicoideae (Figure 6), followed by reciprocal losses of paralogs in the Panicoideae and in the Pooideae- Ehrhartoideae lineages. The ancestral intron pattern of *ZmBx1* supports this view (Figure 6), but additional genome sequences from Pooideae are needed to validate it. TSA and TSA_like proteins form the second clade on our tree. They probably originated from a more basal duplication of a "TSA ancestor", leading to the IGL/BX1 and TSA/TSA_like lineages.

We did not detect blocks of synteny in *B. distachyon, S. italica, O. sativa* and *S. bicolor* that contained any genes homologous to *ZmBx1*. However, the maize *TSA_like/Igl* and *TSA* regions on chromosomes 1 and 7, respectively, showed syntenic blocks in common within each of the four other Poaceae genomes*.* The results imply that these genes arose early during the evolution of Poaceae and maintained their syntenic relationships. *Bx1* thus appears to be the only paralog of the *TSA/IGL* genes that lacks conserved synteny, and its location following

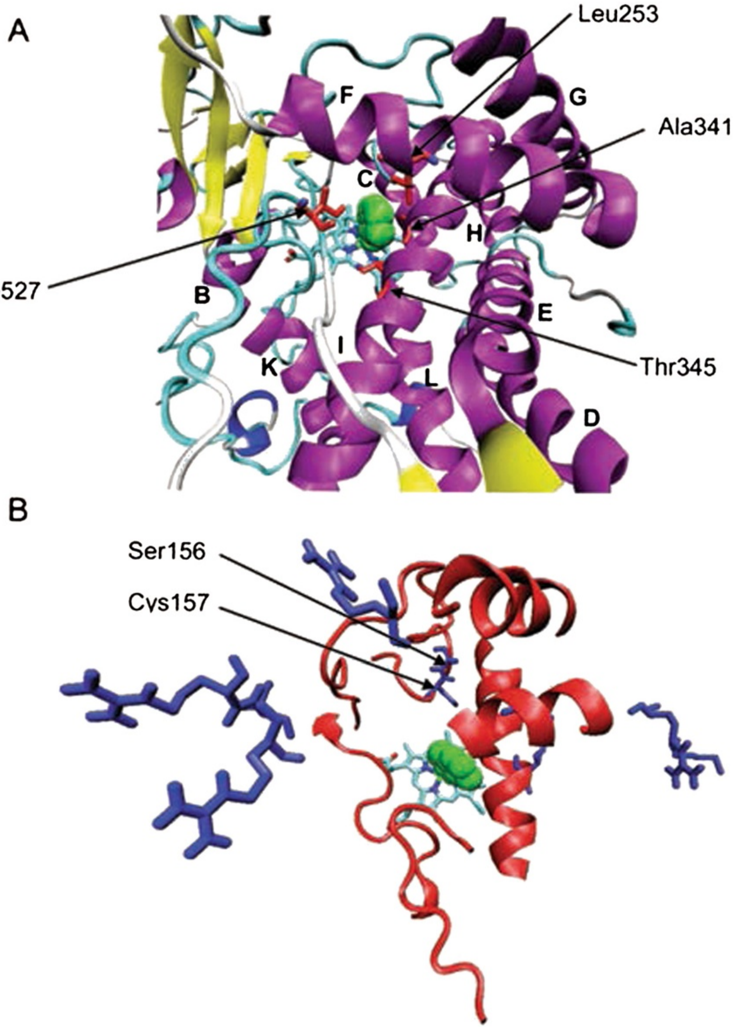

Figure 4 Indole docking in the maize BX2 model. Indole is colored in green. Heme is represented in licorice. VMD was used to create the pictures [88]. **A**. Overall fold and secondary structure contents of maize BX2. The model is represented as NewCartoons and colored as follows: α-helices in purple, 310-helices in dark blue, turns in light blue, β-strands in yellow and coils in white. The four red residues (Leu253, Ala341, Thr345 and Ile527) correspond to sites in contact with indole according to docking results (AutoDock4). **B**. Localization of the ten important sites (in dark blue) for maize CYP71C substrate specificity. The structures in red correspond to SRSs. Ser 156 and Cys 157 are inside SRSs and point towards the active site.

rearrangement separated its evolutionary fate from that of the other *TSA/IGL* genes.

Phylogeny and synteny of the ZmBx6, 7, 8 and 9 genes

ZmBX6, the 2-oxoglutarate-dependent dioxygenase which catalyses the hydroxylation reaction at C-7 of DIBOA, was compared to the 15 closest sequences from Poaceae to reconstruct a phylogenetic tree (Figure 8). The ZmBX6 protein was most closely related to Sb10g006910 of *S. bicolor* (59.9% identity) and to Si010340 of *S. italica* (65.7% identity). No sequences of *Bx6* orthologs are currently available from wheat or wild barley. Although genetic mapping studies place the *ZmBx6* gene in the *ZmBx* cluster on maize chromosome 4 [11,12], we did not find *ZmBx6* in the genomic sequence but found a close paralog (76.5% identity at the protein level) isolated on the long arm of chromosome 2. As the genomic sequence close to *Bx4* contains gaps of undetermined length, it is likely that *Bx6* is lacking in the current version of the maize genome. Therefore, synteny conservation around the *ZmBx6* gene cannot be studied presently.

ZmBX7 is a member of the large *O*-methyltransferase gene family, but paralogs were only found when lowering the BLASTP searches cutoff to 40% identity. The closest homologs were found in *S. italica* (Si022355, 62.9% identity and Si010415, 46.2% identity) and *B. distachyon* (Bradi1g47030, 49.9% identity). The corresponding genes

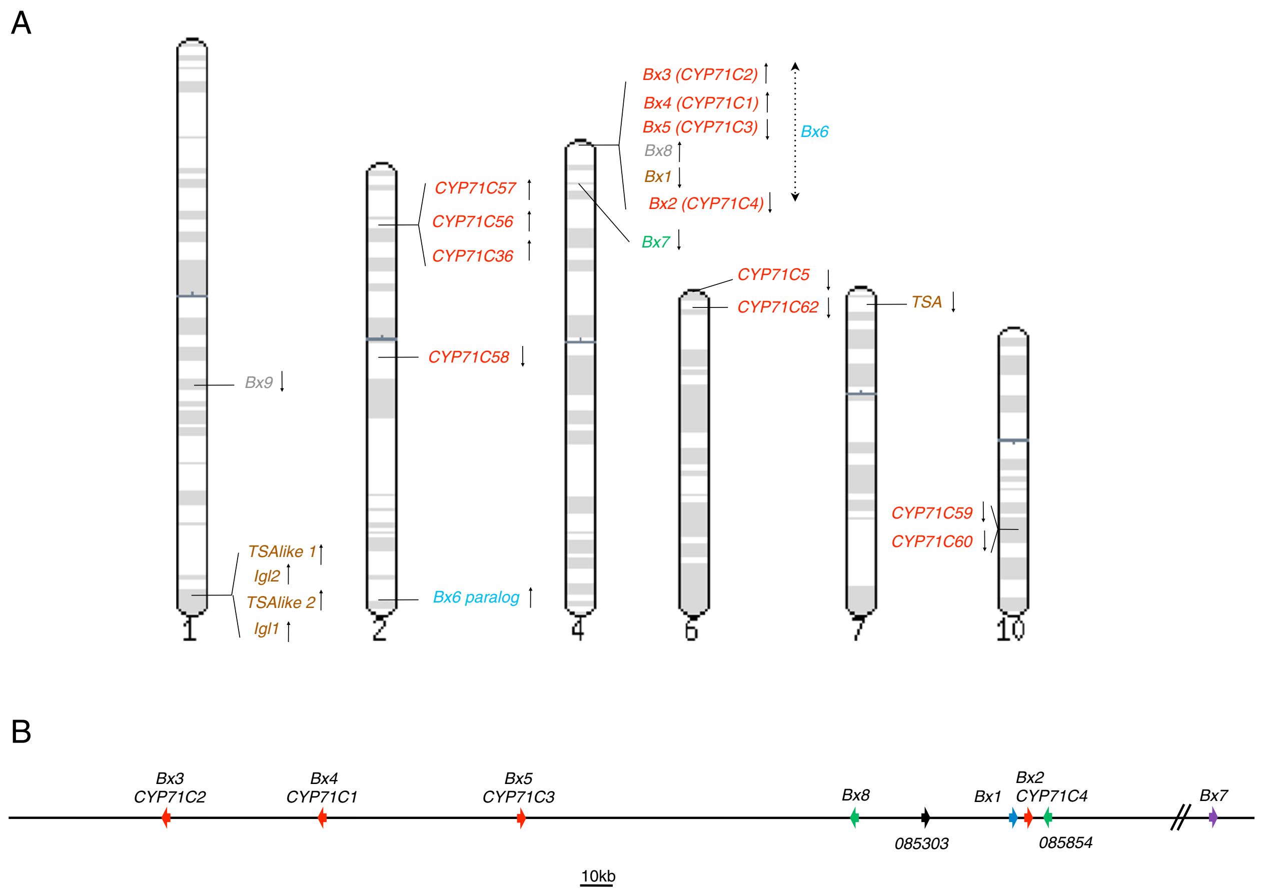

Figure 5 Chromosomal locations of maize Bx and related genes. A. Chromosomal locations of all maize genes used in this study. In brown are genes related to *Bx1*, *TSA*, *Igl* and *Igl_like*. In red are genes included in the phylogenetic analysis of the *CYP71C* subfamily. *Bx6* and *Bx6 paralog* are in blue, *Bx7* in green, *Bx8* and *Bx9* in grey. Arrows indicate the gene orientation. The chromosomes are taken from the maizesequence.org website [30]. **B**. The Bx gene cluster (~ 264 kb) on maize chromosome 4. Gene length is not to scale. 085303 is the GRMZM2G085303 gene of unknown function 085854 is the GRMZM2G085854 gene, an uncharacterized glucosyltransferase that has 27% amino acid identity with BX8. *Bx7* is at about 15 Mb towards the centromere (right), separated from the cluster by about 360 other genes. *Bx6* is not found in the current release of the maize genome. Genetic evidence places it within 7 cM of *Bx3* and *Bx4* [11,12].

had a common intron with *ZmBx7* in phase 0. *ZmBx7* was located on the short arm of maize chromosome 4, at about 15 Mb from *Bx2*, with over 360 annotated genes separating it from the *Bx* cluster. Its genetic proximity was previously reported [11,12]. The *ZmBx7* paralog of *B. distachyon* was on chromosome 1 and the two paralogs of *S. italica* were present on scaffolds 3 and 7. No single copy orthologs were common to maize and these two species in the vicinity of the *O*- methyltransferases genes.

The two glucosyltransferases ZmBX8 and ZmBX9 are also members of a multigene family and the twenty closest homologs from *Avena strigosa, O. sativa, S. cereale, T. aestivum, S. bicolor* and *S. italica* were analyzed. As *ZmBx8* was located between *Bx2* and *Bx5* where no synteny conservation was observed (see above), we focused on synteny blocks containing *ZmBx9*. However, the corresponding blocks in other genomes did not contain any genes homologous to *ZmBx9*. The reconstructed phylogenetic tree underlined the strong relationship between ZmBX8 and ZmBX9 (74.4% identity) and the presence of one close paralog in *S. cereale* (BAJ07107) and four in *T. aestivum* (BAJ07089, BAJ07091, BAJ07092 and BAJ07090) (Figure 9). These four proteins have been recently described as the glucosyltransferases involved in the DIMBOA-biosynthetic pathway in wheat [63]. The closest relatives to these glucosyltransferases were from *S. italica* (Si013725, Si015361 and Si013705), *S. bicolor* (Sb09g028320), *O. sativa* (Os11g0441500 and Os11g0444000) and *A. strigosa* (UGT710E5) (Figure 9). However none of these homologs were included in blocks of conserved synteny with the *ZmBx9* region.

Discussion

Our results provide the first phylogenomic analysis of a biosynthetic gene cluster in plants. We have mined the

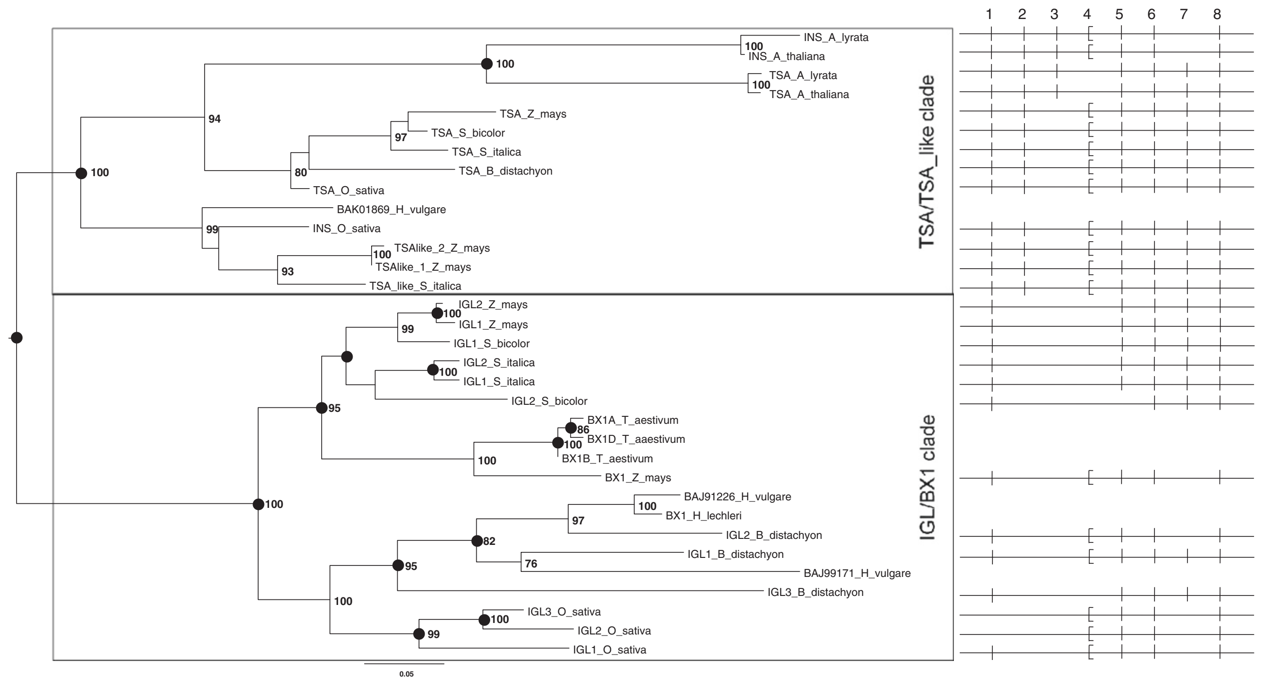

Figure 6 Phylogenetic tree and intron map of BX1, IGL, TSA and TSA_like sequences in Poaceae. Maximum likelihood tree with branch lengths drawn in scale in terms of the number of substitutions per site. Nodes with dots indicate duplication events. The locations of phase 0 introns (|) and phase 1 introns ([) are based on their relative place on the protein alignment and are numbered from 1 to 8. The intron map is not to scale. Only nodes with bootstrap values superior to 75% are shown. Two TSA and two indole synthase sequences of *A. thaliana* and *Arabidopsis lyrata* are closely related but are non-Poaceae sequences. Sequences were named according to their closest homolog in maize and IGL numbers were arbitrarily attributed.

complete genomes of Poaceae that are currently available, and all available sequences for genes related to the known *Bx* genes of maize. Our study included both species that produce benzoxazinoids and species that do not (Figure 7). Our results are relevant to the origin of benzoxazinoids, the evolution of the four key P450 genes of this cluster, and the chromosomal arrangement of this cluster.

Origin of benzoxazinoids

Our analysis supports the hypothesis of Frey et al. [20] that the pathway of benzoxazinoid biosynthesis in Poaceae is monophyletic, at least until DIBOA-glucoside. Indeed the phylogeny of the enzymes involved in its synthesis (BX1 to BX5, and the glucosyl transferase BX8 and BX9) is congruent in benzoxazinoid producers for which the genes encoding these enzymes are known. It is likely that a basal ancestor of Poaceae in the Late Cretaceous, i.e. nearly 70 MYA [64], produced DIBOA glucoside. The origin of the last two steps, hydroxylation (BX6) and methylation (BX7) is less clear, because sequences orthologous to the maize genes are not currently available. Benzoaxinoids have also been detected in single isolated species of two distant related orders of dicots, the Ranunculales and the Lamiales [6]. The dicots diverged from the monocots about 150 to 300 million years ago [65]. As in Poaceae, the first step involves an indole- glycerol phosphate lyase (IGL), separating the DI(M)BOA biosynthetic pathway from primary metabolism. However, duplications leading to *Igl* were independent events in monocots and dicots [66]. The second step of the pathway leading to the production of indolin-2-one from indole is a P450-dependent activity in the DIBOA-producing dicots [66] as in Poaceae, but as no *CYP71C* genes are found in dicots, the genes responsible cannot be orthologs. Independent evolution of benzoxazinoid biosynthesis in monocots and dicots is therefore most likely. This was also demonstrated recently for the biosynthesis of cyanogenic glucosides between monocots and dicots [28], and between plants and insects [67].

Duplications and neofunctionalization of the CYP genes

The phylogenetic relationship of each of the four BX-type P450s is robust and the *Bx* genes of the *CYP71C* subfamily genes in maize, wheat and wild barley are thus out-paralogs, their duplications occurring before speciation events. Interestingly, the intron-exon pattern of the *CYP71C* genes is not phylogenetically informative because of the many independent intron losses. The common ancestral origin of DIBOA glucoside biosynthesis in Panicoideae and Pooideae, and the availability

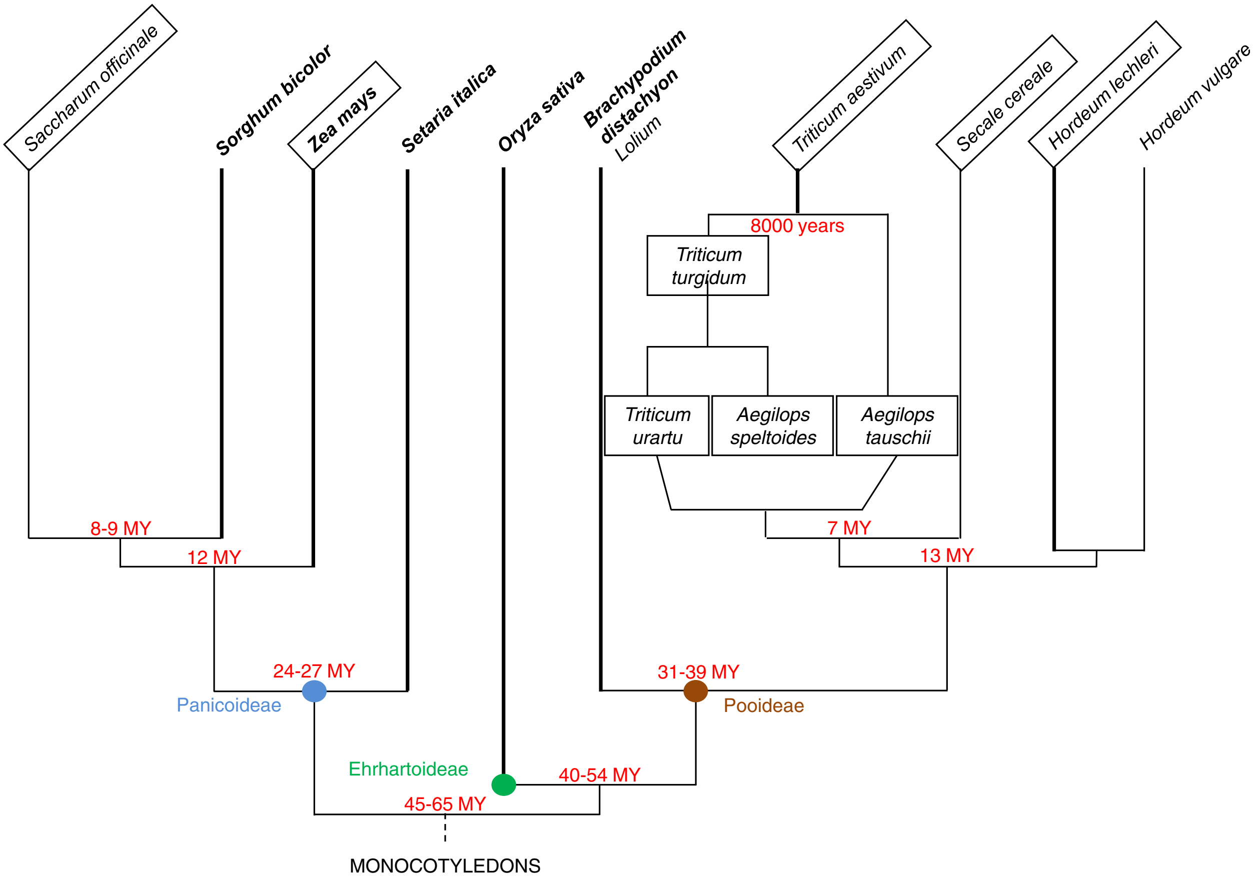

Figure 7 Phylogeny of Poaceae. The topology and estimated times of divergence in million years (MY**)** are based on [65,89-94]. Boxes indicate species accumulating benzoxazinoids [6,17,19,20,95]. In the *Hordeum* genus, benzoxazinoids are detected in some of the H-clade species (such as *H. flexuosum* and *H. lechleri*) and not in some species of the I-clade (such as *H. vulgare* and *H. spontaneum*) [17,63,96]. Bold branches indicate species included in our CYP71C phylogenetic study (only *B. distachyon* is included when considering the common branch *B. distachyon*/*Lolium*). Species / genus are in bold when a complete genome is available.

of genome sequences from the two lineages allowed us to examine in greater detail the origin and evolution of the four P450 genes, the *Bx2-Bx5* genes.

Although the similarity of the sequences and their clustering in maize would suggest that they are the product of a simple series of successive tandem duplications [27], our analysis shows a more complex evolutionary history. The first P450 gene, *Bx2*, is a member of a clade that contains many *CYP71C* sequences from a variety of Poaceae, including non-benzoxazinoid producers. Within this branch, all sequences from wheat and barley form a strongly supported monophyletic clade and all these sequences are biochemically characterized as encoding BX2 enzymes [8,16,19]. The other sequences, from rice, sorghum and millet have not been functionally characterized to date. These *Bx2-like* sequences might be active in other secondary metabolic pathway(s) in non-benzoxazinoid producers. We reconstructed with Codeml the sequence of the ancestor of the BX2/BX2-like clade, synthesized it and produced it in yeast to test our hypothesis of the original biochemical properties of BX2. However we were not able to express an enzymatically functional protein (results not shown). The function of the two close BX2 homologs of *O. sativa*, CYP71C15 and CYP71C16, is unknown. Moreover no benzoxazinoids have been found in rice ([5] and unpublished data in [20]). Maize BX2 catalyzes N-demethylation of *p*-chloro-N-methylaniline [16] in addition to the hydroxylation of indole, suggesting that the ancestral enzyme may have had some catalytic versatility as well. The original function of the BX2/BX2-like P450 ancestor thus remains hypothetical.

If we assume that a common ancestor encoded a BX2-like protein catalyzing indole hydroxylation, then its duplication could explain the expansion of the pathway from indole to DIBOA (Figure 10), by a series of non-successive duplications. This sequence of events is based on our phylogenetic analysis, and on the catalytic properties of the current representatives of each branch in maize, wheat and barley [8,16,17,19]. The first duplication led to “a

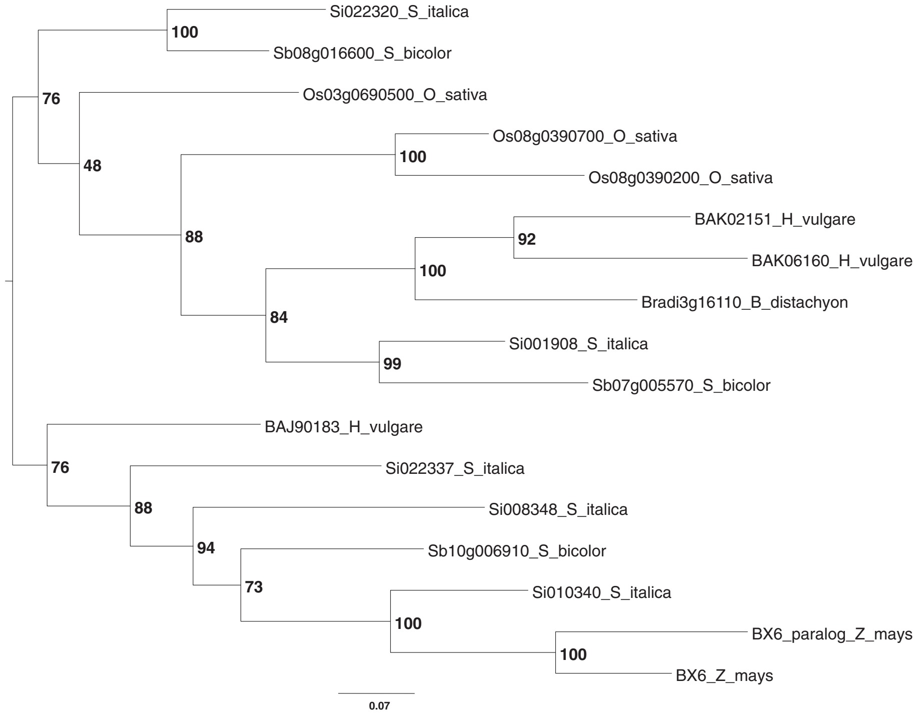

Figure 8 Phylogenetic tree of maize BX6 and related sequences in *Poaceae*. Maximum likelihood tree with branch lengths drawn in scale in terms of the number of substitutions per site.

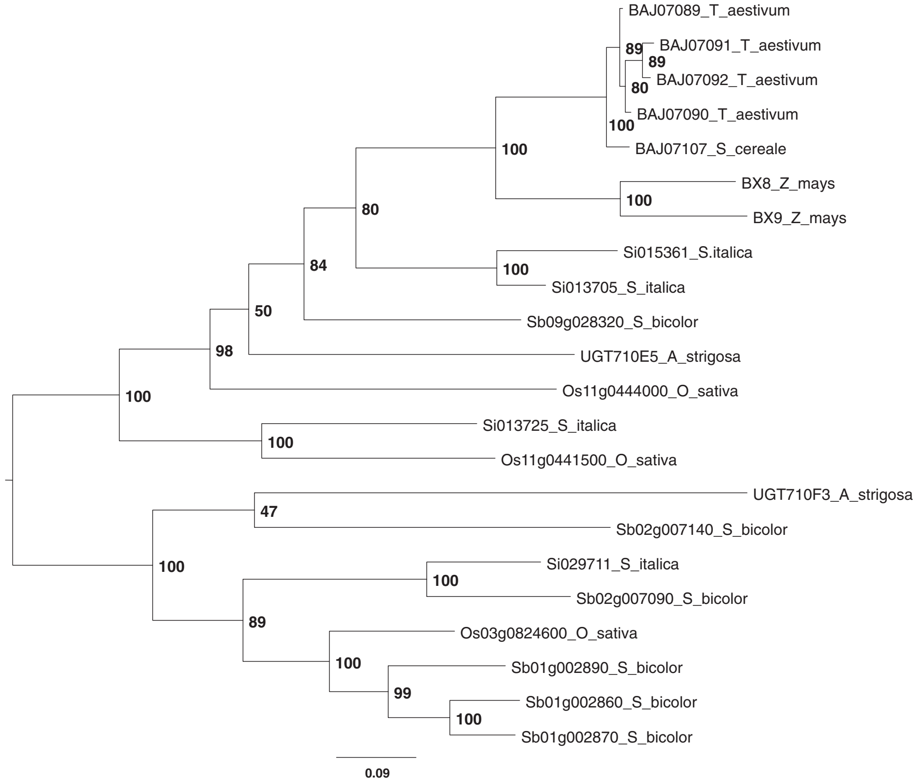

Figure 9 Phylogenetic tree of maize BX8/BX9 and related sequences in *Poaceae*. Maximum likelihood tree with branch lengths drawn in scale in terms of the number of substitutions per site.

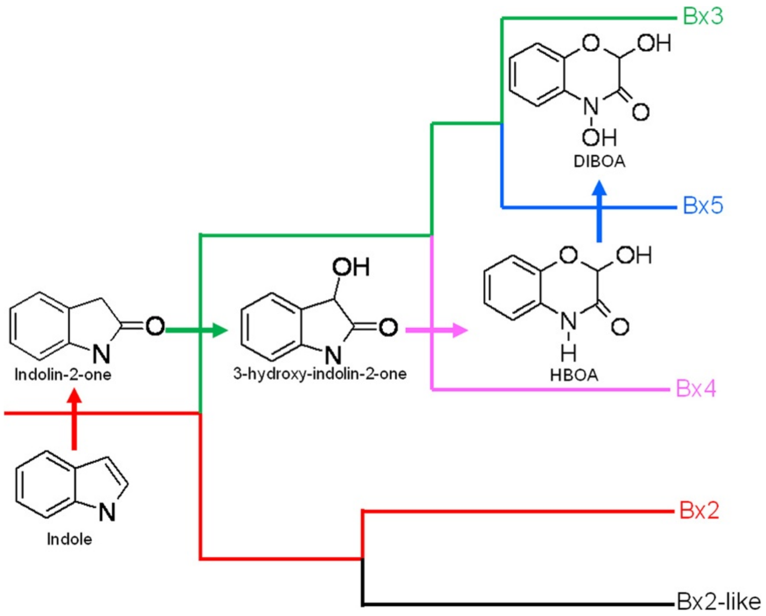

Figure 10 Simplified phylogenetic tree showing the catalytic properties of P450s of the DIMBOA-biosynthetic pathway in maize. The first common ancestor probably had the ancestral function of BX2, producing indolin-2- one from indole (in red). This reaction was initially shared with P450s that are not part of the DIMBOA pathway (BX2-like). The duplications leading to the present BX3, 4 and 5 are shown, with each color indicating the presumed conserved catalytic competence of each ancestral enzyme.

more recent BX2 ancestor" that maintained a C2 hydroxylase activity in all benzoxazinoid producers until the present time. Its duplicate was the common ancestor of BX3/4/5 whose neofunctionalization led to a new C3 hydroxylase activity that has been maintained in BX3. The BX3/4/5 ancestor was then duplicated first to give BX4 neofunctionalized to catalyze oxidative ring expansion and one more time leading to the present BX5 neofunctionalized to an N-hydroxylase. In this scheme (Figure 10) *Bx3* was duplicated twice, giving first *Bx4* then *Bx5*, so that the pathway was not elongated by successive tandem duplications of the gene encoding the last step. The relative orientation and distance between the *Bx2-5* genes in maize (Figure 5B) also suggest inversions and rearrangements with only *Bx3* and *Bx4* present as a tandem array. The neofunctionalization of each new duplicate subtly modified the active site resulting in the specific regioselectivity of oxidation on the indole-like substrate (Figure 10).

Neofunctionalization as understood here is restricted to the substrate specificity of each enzyme with the conservation of their overall P450 characteristics. Both maize and wheat BX3 enzymes hydroxylate indoline-2-one as well as 1,4-benzoxazin-3-one [15,16], suggesting that the enzyme does not discriminate between its natural substrate and its ring- expanded analog. The use of a ring-expanded substrate (HBOA) is a feature of BX5 substrate specificity that supports the origin of *Bx5* from *Bx3*.

Sites under selection in the P450 enzymes

The sequence identities at the amino acid level between BX2 to BX5 orthologs (e.g. from 76 to 79% identity between maize and wheat sequences, 76 to 81% between maize and wild barley sequences) and the BX enzyme substrate specificities are very high [8,15-17]. This suggests that the neofunctionalization following duplication in the ancestral, basal poaceous species was accompanied by selection on only a few sites. This is indeed what our results show. The BX3, BX5 and BX4 clades in our phylogeny are under strong adaptive evolution and some sites are under positive selection. We have identified the specific sites in the maize BX2-5 proteins that have undergone positive selection and functional divergence (type I and/ or type II). More than a half of these sites are localized between the SRS1 and the I-helix and, in particular, among SRS1, 2 and 4 and between the E and G helices. These regions seem to have a major impact on the catalytic properties and the evolution / divergence of the BX2-5 enzymes. Such a non homogenous repartition of sites along the protein was previously described in a phylogenetic analysis of the CYP3 genes family [68]. That study proposed that SRS1, 5 and 6 were performing a universal CYP3A function and that SRS2, 3 and 4 were responsible in part for the functional differences among the enzymes of this family. In contrast, functional divergence in the vertebrate CYP2 family is not clustered in SRSs regions but distributed all along the CYP2 alignment [69].

We identified ten residues potentially important for the substrate specificity of each of the four maize P450s (Figures 3 and 4B). Furthermore, the docking of indole in our model of the active site of maize BX2 identified four residues in close contact with the substrate. One was inside SRS6 and the three other sites were localized in SRS2 and SRS4. The four sites we identified in maize BX2, among which one (Ile 527) differs in the BX3-BX5 proteins, could explain the specificities of the enzyme. Our study thus points to Ser 156, Cys 157, Met 234 and Ile 527 as first candidates for mutagenesis approaches to test their impact on the biochemical properties of maize BX2.

The limited number of sites shown to be under selection or standing out as important in our modeling is consistent with experimental evidence obtained with other P450 enzymes. Few amino acid changes are needed to significantly change substrate specificity of these enzymes. For example, four residues at positions 117, 209, 365 and 481 of *Mus musculus* CYP2A5 are sufficient to determine the steroid substrate specificity [70]. The specificity of this P450 is influenced by the hydrophobicity and residues size [71,72]. In the human CYP2C19, three residues at positions 99, 220, and 221 are key residues that determine the hydroxylase activity for omeprazole [73]. In CYP2B11 of *Canis lupus* three sites in putative SRSs (residues 114, 290 and 363) are important for the enzyme substrate specificity and regio/stereoselectivity [74]. In *Papilio polyxenes,*residues 116, 117, 371 and 484 of CYP6B1 are critical for substrate binding affinity [75]. While the CYP2 and CYP6 enzymes are predominantly xenobiotic- metabolizing enzymes with diffuse substrate specificity, the CYP71C studied here are thought to have a tight biosynthetic function. For plant biosynthetic P450s, mutagenesis of the mint limonene hydroxylases from the CYP71D subfamily showed that a single amino acid change is sufficient to convert a C6- to a C3-hydroxylase [76]. The limited number of critial residues identified in our study is therefore reasonable to explain the subtle shifts in substrate regioselectivity that accompanied the evolution of the BX2-BX5 enzymes but this will require experimental confirmation.

Origin of the Bx biosynthetic cluster: founding event

The assembly of a biosynthetic gene cluster in plants was discussed by Frey *et al.* [20] with the benzoxazinoid pathway as a prototype. They saw three essential and sequential modules: a branchpoint reaction, chemical modification leading to a biological active compound, and detoxification. Osbourn [26,77] assigned the branchpoint reaction to a signature enzyme, and chemical modification to tailoring enzymes, but did not recognize the importance and integrality of detoxification. Instead, chromosomal clustering was seen as a way to prevent the accumulation of toxic intermediates in a pathway [77]. Although the description of pathway assembly by the juxtaposition of three modules is a useful guide, our analysis suggests that in the case of the benzoxazinoid pathway clustering of the first two genes, *Bx1* and *Bx2*, is the key event. Furthermore, we propose that both BX1 and BX2 are signature enzymes that only together constitute a branchpoint committing to benzoxazinoid biosynthesis. We propose to call their clustering the "founding event" of the biosynthetic cluster. The evidence for this view can be developed as follows:

Indole as a product of a branchpoint reaction is not a committed precursor of benzoxazinoids. Phylogenetic analysis shows that an initial duplication of *TSA* led to an *IGL* ancestor that was further duplicated to *Bx1*. TSA is a subunit of tryptophan synthase in "primary metabolism", and current IGL enzymes are involved in the generation of biologically active volatile indole [62,78]. Thus IGL and BX1 perform the same reaction, albeit with diverged catalytic properties [8,78,79]. Our study of *TSA, TSA_like, Igl* and *Bx1* demonstrated that *Bx1* originated before the radiation of Poaceae. Although Grun *et al.* proposed that independent TSA gene duplication events have created *Bx1*-function in maize and wheat on one hand and in barley on the other [17], our phylogenetic analysis clearly shows that this is not the case. The sequence of *H. lechleri* named as "BX1" by Grun et al. [17] clearly falls within the IGL clade, with strong bootstrap support. Furthermore, its catalytic properties are not characteristic of BX1 but rather of an IGL [17,79]. Its kcat/KM (31 mM^{-1} . s^{-1}) is much closer to that of *Z. mays* IGL (23 mM^{-1} . s^{-1}) than to *Z. mays* BX1 (215 mM^{-1} . s^{-1}) [17]. The sequence is therefore an ortholog of *H. vulgare* BAJ91226, and the *H. lechleri Bx1* remains to be discovered. The absence of synteny between *ZmBx1* and other Poaceae genome regions is quite surprising as we found synteny conservation for *TSA, TSA_like* and *Igl*. The synteny of *ZmBx1* and *Bx2* [8] is the only conserved feature in all benzoaxazinoid producers and points to the uniqueness of this clustering. The phylogenetic position of BX2 is similar to that of BX1, a sequence emerging from a background of several duplication events and remarkable only because it forms a monophyletic clade with the wheat and wild barley enzymes of identical function. Both BX1 and BX2 have close homologs that are not involved in benzoxazinoid biosynthesis, so that they are signature enzymes catalyzing branchpoint reactions only when considered together. Initial clustering of both genes enabled their subsequent coevolution and divergence from *Igl_like* and *Bx2-like* genes, respectively. In this view, genomic rearrangements that led to the random clustering of the newly duplicated ancestral *Bx1* and *Bx2* genes represents the "founding event" of the biosynthetic cluster. This key innovation is therefore a structural one, and it makes sense because it distinguishes a biosynthetic cluster of genes from an assemblage of genes (not necessarily clustered) that form a biosynthetic pathway. The terms signature/branchpoint and decoration/chemical modification can equally be applied to biosynthetic clusters as to genomically dispersed biosynthetic pathways, so a more specific nomenclature is required. What then would be the second step ? We propose to call it "elongation" in preference to recruitment, to emphasize the genomic feature over its functional aspect.

Origin of the Bx biosynthetic cluster: elongation

Conservation of the *Bx1 -Bx2* synteny from maize to wheat and rye [80] confirms that the founding event of the biosynthetic cluster occurred in an ancestor of Poaceae. Elongation of the cluster to *Bx5* by the P450 duplications and gene rearrangements described above led to cluster of 5 genes in maize. Is this the ancestral state or did the *Bx3-5* genes integrate the cluster together, as a separate event ? In both rye and wheat *Bx1-2* and *Bx3-5* form two distinct clusters. In rye, *ScBx1* and *ScBx2* are located on chromosome 7R and *ScBx3*, *ScBx4* and *ScBx5* are on chromosome 5R. In wheat, *TaBx1* and *TaBx2* are closely located on group-4 chromosomes and *TaBx3*, *TaBx4* and *TaBx5* are closely located on group-5 chromosomes [80]. Rye 5R and 7R chromosomes have high conserved synteny with group-5 and group-4 chromosomes of wheat [80-82]. Nomura et al. [80] proposed that the ancestral Bx cluster was split in a common ancestor of rye and wheat. Moreover, the presence of microlinearity and partial orthology has been demonstrated between wheat group-7 chromosomes (containing the glucosyltransferase of the DIMBOA-biosynthetic pathway) and maize chromosomes 1 and 4 (including respectively *ZmBx8* and *ZmBx9*). In rye, the glucosyltransferase is also found isolated on the 4R chromosome, consistent with the known synteny between rye and wheat [83]. The addition to the cluster of a glucosyltransferase gene necessary to convert DIBOA (the product of BX5) to DIBOA-glucoside resulted from an ancient gene rearrangement, and our phylogenetic analysis indicates the orthology of the rye, wheat and maize genes. We conclude that this cluster elongation was also an early event in an ancestral Poaceae species. It becomes difficult to distinguish detoxification as proposed by Frey *et al.* [20] and clustering (here of a glucosyltransferase) to prevent toxic intermediate accumulation as proposed by Osbourn [77]. Glucosyltransferases are integral parts of the cyanogenic glucoside biosynthetic clusters [27]. It has been proposed that physical disruption of the components of the cyanogenic glucoside metabolon can be a way to diversify the products of the pathway, as different intermediates are toxic to different targets [84]. There are therefore different ways to maintain integrity of a biosynthetic cluster: the genomic integrity that favors cosegregation of all components, and, at least for cyanogenic glucosides, the subcellular integrity as a metabolon. Is the glucosyltransferase the "last" enzyme in DIBOA-glucoside biosynthesis acting on the product of BX5? This is traditionally shown (Figure 1), and does not address the question of the earlier intermediates. Yet the products of BX3 and BX4 are observed as glucosylated metabolites in maize [85], Dutartre *et al.*, in preparation], so that the contribution of a glucosyltransferase to the biosynthetic cluster may have preceded the last two duplications of *Bx3*. The maize glucosyltransferases have significant activity towards HBOA, the product of BX4 [10]. Significantly, *Bx8* is closest to *Bx1* and *2* in the cluster (Figure 5B) and is the ortholog of the rye and wheat genes. The wheat sequences result from hexaploidization, with one duplication in the B genome [83]. *Bx9* is a recent duplicate of *Bx8* in the maize lineage as shown by our phylogenetic analysis and that of Sue *et al.* [83]. It is not located in the cluster, and its catalytic properties [10] indicate that it has lost considerable activity toward DIBOA. Following the *Bx8/Bx9* duplication, the sequence divergence of *Bx9* and its new location on another chromosome probably led to a new physiological role different from benzoxazinoid biosynthesis. The lack of QTL involved in DIMBOA synthesis associated with the *Bx9* region [86] supports this conclusion.

Further elongation of the cluster corresponds to the aromatic hydroxylation and methylation of DIBOA glucoside by BX6 and BX7 [12]. The evolutionary history of this elongation, and indeed of the further methylation to HDMBOA-glucoside is difficult to ascertain at present, because *Bx6* and *Bx7* have not been sequenced in other benzoxazinoid producers, and the last methyltransferase gene is still unknown. *Bx6* has a close paralog on chromosome 2, and *Bx7*, while close to the Bx cluster, is about 35 cM distant. Intriguingly, the closest homologs of *Bx6* and *Bx7* in *S. italica* are located on scaffold 7, in close proximity to four P450 genes, *CYP71C81*, and of the cluster of *CYP71C88, 89, 92*. The latter is orthologous to the maize *CYP71C36, 56, 57* cluster on chromosome 2. While the function of these genes is currently unknown, it is tempting to speculate that *Bx6* and *Bx7* are moonlighting in a different biosynthetic cluster. Their position on the outside of the Bx cluster may have prevented them from being lost when the *S. italica* ancestor lost the *Bx1-Bx5* genes. We note that the Km of BX6 and BX7 towards their benzoxazinoid substrates is the poorest of all BX enzymes [12], and as they take glucosylated substrates and not their aglycone, the aglycone contribution to substrate specificity may be weak, supporting the alternative function hypothesis.

There are several examples of plant genes clusters located close to the tip of the chromosome as is commonly found in actinomycetes and ascomycetes [26]. This position is particularly favorable to adaptive evolution and to coordinated regulation [26,87]. The presence of genes in a cluster would favor their co-segregation and thus favor the rapid evolution of the linked genes. As no *Bx* cluster is found in *S. bicolor*, *O. sativa* and *H. vulgare*, it is likely that the original, complete cluster was lost in a single evolutionary event [80]. The chromosomal position of the Bx cluster may have favored through cosegregation the subsequent rearrangements of the *Bx8*, then *Bx6* and *Bx7* genes in close proximity. Additional genome sequences from benzoxazinoid producers may provide additional evidence for this sequence of events. Our analysis suggests that the key factor in the origin of biosynthetic gene clusters in plants, and perhaps in other organisms, is a "founder event" where the first two genes originating from random duplications and rearrangements find themselves linked and commit to a new pathway. Whether co-regulation or co-segregation is a most important result of this clustering remains to be ascertained. The genetic arguments for the primacy of co-segregation have long been known. In higher eukaryotes, the evidence and mechanisms of co-regulation of recently rearranged genes are less well established.

Conclusions

Our phylogenomic analysis of the origin of the Bx cluster in maize shows that the first two closely linked genes of the benzoxazinoid pathway are located at a chromosomal region that has no synteny conservation with the genomes of other Poaceae beyond *Bx1/Bx2* themselves, and is therefore unique. Rearrangements following duplications of an *IGL/TSA* gene and of a *CYP71C* gene resulting in the clustering of the new copies (*Bx1* and *Bx2*) constitute the founding event of the biosynthetic cluster. This founding event is a genomic character, different and perhaps more general than "branchpoint reaction" [20] or "signature enzyme" [26,77] that denote biochemical characters that would not adequately describe the importance of the tight clustering of *Bx1* and *Bx2*. Elongation of the cluster involved duplications of a *Bx2*-like *CYP71C* gene and neofunctionalizations that involved positive selection at few distinct sites of these P450 enzymes. At least one glucosyltransferase gene was recruited into the pathway and rearranged into the cluster. Our data are consistent with our current understanding of biosynthetic clusters in plants [20,27,28,86], but highlight the importance of the founding event in seeding a biosynthetic cluster.

Additional files

Additional file 1: Phylogenetic tree and intron map of the CYP71C sequences of Poaceae.

Additional file 2: Maize BX2-BX5 alignment showing putative helices, P450 motifs, SRSs and sites under positive selection.

Abbreviations

Bx: Benzoxazinoid; DIBOA: 2,4-dihydroxy-1,4-benzoxazin-3-one; DIMBOA: 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA); IGL: Indole 3-glycerol phosphate lyase; SRS: Substrate Recognition Site; TSA: α-subunit of tryptophan synthase.

Authors' contributions

LD, FH and RF designed the study, LD and FH carried out the work, LD, FH and RF analyzed the results and wrote the paper. All authors read and approved the final manuscript.

Acknowledgements

We thank Drs Etienne Danchin and Jonathan Corbi for constructive discussions. This work was supported by the Agence Nationale de la Recherche Grant 06 BLAN0346 to RF and by INRA SPE grant 2010-1301-1 to FH. LD was supported by a doctoral grant co-funded by CNRS and PACA Region.

Received: 12 August 2011 Accepted: 17 April 2012
Published: 11 May 2012

References

1. Nelson D, Werck-Reichhart D: **A P450-centric view of plant evolution.** *Plant J* 2011, **66**:194–211.
2. Mansuy D: **The great diversity of reactions catalyzed by cytochromes P450.** *Comp Biochem Physiol C Pharmacol Toxicol Endocrinol* 1998, **121**:5–14.
3. Werck-Reichhart D, Bak S, Paquette S: *Cytochromes P450*.: The Arabidopsis book; 2002:1–29.
4. Sicker D, Frey M, Schulz M, Gierl A: **Role of natural benzoxazinones in the survival strategy of plants.** *Int Rev Cytol* 2000, **198**:319–346.
5. Niemeyer HM: **Hydroxamic acids (4-hydroxy-1,4-benzoxazin-3-ones), defense chemicals in the Gramineae.** *Phytochemistry* 1988, **27**:3349–3358.
6. Sicker D, Schulz M: **Benzoxazinones in plants: occurrence, synthetic access, and biological activity.** *Stud Nat Prod Chem* 2002, **27**:185–232.
7. Virtanen AI, Hietala PK, Wahlroos O: **Antimicrobial substances in cereals and fodder plants.** *Arch Biochem Biophys* 1957, **69**:486–500.
8. Frey M, Chomet P, Glawischnig E, Stettner C, Grun S, Winklmair A, Eisenreich W, Bacher A, Meeley RB, Briggs SP, *et al*: **Analysis of a chemical plant defense mechanism in grasses.** *Science* 1997, **277**:696–699.
9. Frey M, Kliem R, Saedler H, Gierl A: **Expression of a cytochrome P450 gene family in maize.** *Mol Gen Genet* 1995, **246**:100–109.
10. von Rad U, Huttl R, Lottspeich F, Gierl A, Frey M: **Two glucosyltransferases are involved in detoxification of benzoxazinoids in maize.** *Plant J* 2001, **28**:633–642.
11. Frey M, Huber K, Park WJ, Sicker D, Lindberg P, Meeley RB, Simmons CR, Yalpani N, Gierl A: **A 2-oxoglutarate-dependent dioxygenase is integrated in DIMBOA-biosynthesis.** *Phytochemistry* 2003, **62**:371–376.
12. Jonczyk R, Schmidt H, Osterrieder A, Fiesselmann A, Schullehner K, Haslbeck M, Sicker D, Hofmann D, Yalpani N, Simmons C, *et al*: **Elucidation of the final reactions of DIMBOA-glucoside biosynthesis in maize: characterization of Bx6 and Bx7.** *Plant Physiol* 2008, **146**:1053–1063.
13. Oikawa A, Ishihara A, Iwamura H: **Induction of HDMBOA-Glc accumulation and DIMBOA-Glc 4-O-methyltransferase by jasmonic acid in poaceous plants.** *Phytochemistry* 2002, **61**:331–337.
14. Oikawa A, Ishihara A, Tanaka C, Mori N, Tsuda M, Iwamura H: **Accumulation of HDMBOA-Glc is induced by biotic stresses prior to the release of MBOA in maize leaves.** *Phytochemistry* 2004, **65**:2995–3001.
15. Nomura T, Ishihara A, Imaishi H, Endo TR, Ohkawa H, Iwamura H: **Molecular characterization and chromosomal localization of cytochrome P450 genes involved in the biosynthesis of cyclic hydroxamic acids in hexaploid wheat.** *Mol Genet Genomics* 2002, **267**:210–217.

16. Glawischnig E, Grun S, Frey M, Gierl A: **Cytochrome P450 monooxygenases of DIBOA biosynthesis: specificity and conservation among grasses.** *Phytochemistry* 1999, **50**:925–930.
17. Grun S, Frey M, Gierl A: **Evolution of the indole alkaloid biosynthesis in the genus Hordeum: distribution of gramine and DIBOA and isolation of the benzoxazinoid biosynthesis genes from Hordeum lechleri.** *Phytochemistry* 2005, **66**:1264–1272.
18. Nomura T, Ishihara A, Iwamura H, Endo TR: **Molecular characterization of benzoxazinone-deficient mutation in diploid wheat.** *Phytochemistry* 2007, **68**:1008–1016.
19. Nomura T, Ishihara A, Yanagita RC, Endo TR, Iwamura H: **Three genomes differentially contribute to the biosynthesis of benzoxazinones in hexaploid wheat.** *Proc Natl Acad Sci U S A* 2005, **102**:16490–16495.
20. Frey M, Schullehner K, Dick R, Fiesselmann A, Gierl A: **Benzoxazinoid biosynthesis, a model for evolution of secondary metabolic pathways in plants.** *Phytochemistry* 2009, **70**:1645–1651.
21. Fischbach MA, Walsh CT, Clardy J: **The evolution of gene collectives: How natural selection drives chemical innovation.** *Proc Natl Acad Sci* 2008, **105**:4601–4608.
22. Paquette SM, Bak S, Feyereisen R: **Intron-exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of Arabidopsis thaliana.** *DNA Cell Biol* 2000, **19**:307–317.
23. Feyereisen R: **Arthropod CYPomes illustrate the tempo and mode in P450 evolution.** *Biochim Biophys Acta* 2011, **1814**:19–28.
24. Strode C, Wondji CS, David JP, Hawkes NJ, Lumjuan N, Nelson DR, Drane DR, Karunaratne SH, Hemingway J, Black WCt, Ranson H: **Genomic analysis of detoxification genes in the mosquito Aedes aegypti.** *Insect Biochem Mol Biol* 2008, **38**:113–123.
25. Wang H, Donley KM, Keeney DS, Hoffman SM: **Organization and evolution of the Cyp2 gene cluster on mouse chromosome 7, and comparison with the syntenic human cluster.** *Environ Health Perspect* 2003, **111**:1835–1842.
26. Osbourn A: **Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation.** *Trends Genet* 2010, **26**:449–457.
27. Chu HY, Wegel E, Osbourn A: **From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants.** *Plant J* 2011, **66**:66–79.
28. Takos AM, Knudsen C, Lai D, Kannangara R, Mikkelsen L, Motawia MS, Olsen CE, Sato S, Tabata S, Jorgensen K, *et al*: **Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in Lotus japonicus and suggests the repeated evolution of this chemical defence pathway.** *Plant J* 2011, **68**:273–286.
29. *BLAST: Basic Local Alignment Search Tool*. http://blast.ncbi.nlm.nih.gov/Blast.cgi.
30. *MaizeSequence 5b60: Home*. http://www.maizesequence.org/index.html.
31. BRACHYPODIUM.ORG: *The Brachypodium distachyon Information Resource - Home*. http://www.brachypodium.org/.
32. *Phytozome v7.0: Home*. http://www.phytozome.net/.
33. Abascal F, Zardoya R, Posada D: **ProtTest: selection of best-fit models of protein evolution.** *Bioinformatics* 2005, **21**:2104–2105.
34. *Phylogeny.fr: Home*. http://www.phylogeny.fr/version2_cgi/index.cgi.
35. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**:1792–1797.
36. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, *et al*: **Phylogeny.fr: robust phylogenetic analysis for the non-specialist.** *Nucleic Acids Res* 2008, **36**:465–469.
37. Dereeper A, Audic S, Claverie JM, Blanc G: **BLAST-EXPLORER helps you building datasets for phylogenetic analysis.** *BMC Evol Biol* 2010, **10**:8.
38. Jones DT, Taylor WR, Thornton JM: **The rapid generation of mutation data matrices from protein sequences.** *Comput Appl Biosci* 1992, **8**:275–282.
39. Goldman N, Yang Z: **A codon-based model of nucleotide substitution for protein-coding DNA sequences.** *Mol Biol Evol* 1994, **11**:725–736.
40. Wernersson R, Pedersen AG: **RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences.** *Nucleic Acids Res* 2003, **31**:3537–3539.
41. Yang Z: **Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution.** *Mol Biol Evol* 1998, **15**:568–573.
42. Yang Z, Nielsen R, Goldman N, Pedersen AM: **Codon-substitution models for heterogeneous selection pressure at amino acid sites.** *Genetics* 2000, **155**:431–449.
43. Bielawski JP, Yang Z: **Maximum likelihood methods for detecting adaptive evolution after gene duplication.** *J Struct Funct Genomics* 2003, **3**:201–212.
44. Low WY, Ng HL, Morton CJ, Parker MW, Batterham P, Robin C: **Molecular evolution of glutathione S-transferases in the genus Drosophila.** *Genetics* 2007, **177**:1363–1375.
45. Yang Z, Nielsen R: **Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages.** *Mol Biol Evol* 2002, **19**:908–917.
46. Poulos TL, Finzel BC, Howard AJ: **High-resolution crystal structure of cytochrome P450cam.** *J Mol Biol* 1987, **195**:687–700.
47. *Jpred 3*. http://www.compbio.dundee.ac.uk/www-jpred/advanced.html.
48. *PORTER*. http://distill.ucd.ie/porter/.
49. Gotoh O: **Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences.** *J Biol Chem* 1992, **267**:83–90.
50. Gu X: **A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences.** *Mol Biol Evol* 2006, **23**:1937–1945.
51. Gu X: **Maximum-likelihood approach for gene family evolution under functional divergence.** *Mol Biol Evol* 2001, **18**:453–464.
52. Gu X: **Statistical methods for testing functional divergence after gene duplication.** *Mol Biol Evol* 1999, **16**:1664–1674.
53. Gu X, Vander Velden K: **DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family.** *Bioinformatics* 2002, **18**:500–501.
54. *I-TASSER server for protein structure and function prediction*. http:// zhanglab.ccmb.med.umich.edu/I-TASSER/.
55. Zhang Y: **I-TASSER server for protein 3D structure prediction.** *BMC Bioinformatics* 2008, **9**:40.
56. Roy A, Kucukural A, Zhang Y: **I-TASSER: a unified platform for automated protein structure and function prediction.** *Nat Protoc* 2010, **5**:725–738.
57. *AutoDock - AutoDock*. http://autodock.scripps.edu/.
58. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ: **AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility.** *J Comput Chem* 2009, **30**:2785–2791.
59. Sanner MF: **Python: a programming language for software integration and development.** *J Mol Graph Model* 1999, **17**:57–61.
60. Soderlund C, Nelson W, Shoemaker A, Paterson A: **SyMAP: a system for discovering and viewing syntenic regions of FPC maps.** *Genome Res* 2006, **16**:1159–1168.
61. Murat F, Xu JH, Tannier E, Abrouk M, Guilhot N, Pont C, Messing J, Salse J: **Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution.** *Genome Res* 2010, **20**:1545–1557.
62. Gierl A, Frey M: **Evolution of benzoxazinone biosynthesis and indole production in maize.** *Planta* 2001, **213**:493–498.
63. Gianoli E, Niemeyer HM: **DIBOA in wild Poaceae: sources of resistance to the Russian wheat aphid (Diuraphis noxia) and the greenbug (Schizaphis graminum).** *Euphytica* 1998, **102**:317–321.
64. Prasad V, Stromberg CA, Leache AD, Samant B, Patnaik R, Tang L, Mohabey DM, Ge S, Sahni A: **Late Cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae.** *Nature Comm* 2011, **2**:480.
65. Abrouk M, Murat F, Pont C, Messing J, Jackson S, Faraut T, Tannier E, Plomion C, Cooke R, Feuillet C, Salse J: **Palaeogenomics of plants: synteny-based modelling of extinct ancestors.** *Trends Plant Sci* 2010, **15**:479–487.
66. Schullehner K, Dick R, Vitzthum F, Schwab W, Brandt W, Frey M, Gierl A: *Benzoxazinoid biosynthesis in dicot plants. Phytochemistry* 2008, **69**:2668–2677.
67. Jensen NB, Zagrobelny M, Hjerno K, Olsen CE, Houghton-Larsen J, Borch J, Moller BL, Bak S: **Convergent evolution in biosynthesis of cyanogenic defence compounds in plants and insects.** *Nature Comm* 2011, **2**:273.
68. McArthur AG, Hegelund T, Cox RL, Stegeman JJ, Liljenberg M, Olsson U, Sundberg P, Celander MC: **Phylogenetic analysis of the cytochrome P450 3 (CYP3) gene family.** *J Mol Evol* 2003, **57**:200–211.
69. Kirischian N, McArthur AG, Jesuthasan C, Krattenmacher B, Wilson JY: **Phylogenetic and functional analysis of the vertebrate cytochrome P450 2 family.** *J Mol Evol* 2010.
70. Negishi M, Iwasaki M, Juvonen RO, Sueyoshi T, Darden TA, Pedersen LG: **Structural flexibility and functional versatility of cytochrome P450 and rapid evolution.** *Mutat Res* 1996, **350**:43–50.
71. Iwasaki M, Juvonen R, Lindberg R, Negishi M: **Alteration of high and low spin equilibrium by a single mutation of amino acid 209 in mouse cytochromes P450.** *J Biol Chem* 1991, **266**:3380–3382.

72. Juvonen RO, Iwasaki M, Negishi M: Structural function of residue-209 in coumarin 7-hydroxylase (P450coh). *J Biol Chem* 1991, **266**:16431–16435.
73. Ibeanu GC, Ghanayem BI, Linko P, Li L, Pederson LG, Goldstein JA: Identification of residues 99, 220, and 221 of human cytochrome P450 2 C19 as key determinants of omeprazole activity. *J Biol Chem* 1996, **271**:12496–12501.
74. Hasler JA, Harlow GR, Szklarz GD, John GH, Kedzie KM, Burnett VL, He YA, Kaminsky LS, Halpert JR: Site-directed mutagenesis of putative substrate recognition sites in cytochrome P450 2B11: importance of amino acid residues 114, 290, and 363 for substrate specificity. *Mol Pharmacol* 1994, **46**:338–345.
75. Li W, Schuler MA, Berenbaum MR: Diversification of furanocoumarin-metabolizing cytochrome P450 monooxygenases in two papilionids: specificity and substrate encounter rate. *Proc Natl Acad Sci U S A* 2003, **100**:14593–14598.
76. Schalk M, Croteau R: A single amino acid substitution (F363I) converts the regiochemistry of the spearmint (−)-limonene hydroxylase from a C6- to a C3-hydroxylase. *Proc Natl Acad Sci* 2000, **97**:11948–11953.
77. Osbourn A: Gene clusters for secondary metabolic pathways: an emerging theme in plant biology. *Plant Physiol* 2010, **154**:531–535.
78. Frey M, Stettner C, Pare PW, Schmelz EA, Tumlinson JH, Gierl A: An herbivore elicitor activates the gene for indole emission in maize. *Proc Natl Acad Sci U S A* 2000, **97**:14801–14806.
79. Kriechbaumer V, Weigang L, Fiesselmann A, Letzel T, Frey M, Gierl A, Glawischnig E: Characterisation of the tryptophan synthase alpha subunit in maize. *BMC Plant Biol* 2008, **8**:44.
80. Nomura T, Ishihara A, Imaishi H, Ohkawa H, Endo TR, Iwamura H: Rearrangement of the genes for the biosynthesis of benzoxazinones in the evolution of Triticeae species. *Planta* 2003, **217**:776–782.
81. Liu CJ, Atkinson MD, Chinoy CN, Devos KM, Gale MD: Nonhomoeologous translocations between group 4, 5 and 7 chromosomes within wheat and rye. *Theor Appl Genet* 1993, **83**:305–312.
82. Devos KM, Atkinson MD, Chinoy CN, Francis HA, Harcourt RL, Koebner RMD, Liu CJ, Masojc P, Xie DX: Chromosomal rearrangements in the rye genome relative to that of wheat. *Theor Appl Genet* 1993, **85**:673–680.
83. Sue M, Nakamura C, Nomura T: Dispersed benzoxazinone gene cluster: molecular characterization and chromosomal localization of glucosyltransferase and glucosidase genes in wheat and rye. *Plant Physiol* 2011.
84. Moller BL: Plant science. *Dynamic metabolons. Science* 2010, **330**:1328–1329.
85. Glauser G, Marti G, Villard N, Doyen GA, Wolfender JL, Turlings TCJ, Erb M: Induction and detoxification of maize 1,4-benzoxazin-3-ones by insect herbivores. *Plant J* 2011.
86. Butron A, Chen YC, Rottinghaus GE, McMullen MD: Genetic variation at bx1 controls DIMBOA content in maize. *Theor Appl Genet* 2010, **120**:721–734.
87. Field B, Fiston-Lavier AS, Kemen A, Geisler K, Quesneville H, Osbourn AE: Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc Natl Acad Sci U S A* 2011, **108**:16116–16121.
88. Humphrey W, Dalke A, Schulten K: VMD: visual molecular dynamics. *J Mol Graph Model* 1996, **14**:33–38.
89. Huang S, Sirikhachornkit A, Faris JD, Su X, Gill BS, Haselkorn R, Gornicki P: Phylogenetic analysis of the acetyl-CoA carboxylase and 3- phosphoglycerate kinase loci in wheat and other grasses. *Plant Mol Biol* 2002, **48**:805–820.
90. Devos KM: Updating the ‘crop circle’. *Curr Opin Plant Biol* 2005, **8**:155–162.
91. Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, Calcagno T, Cooke R, Delseny M, Feuillet C: Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *Plant Cell* 2008, **20**:11–24.
92. Initiative IB: Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature* 2010, **463**:763–768.
93. Chalupska D, Lee HY, Faris JD, Evrard A, Chalhoub B, Haselkorn R, Gornicki P: Acc homoeoloci and the evolution of wheat genomes. *Proc Natl Acad Sci U S A* 2008, **105**:9691–9696.
94. Blattner FR: Phylogenetic analysis of Hordeum (Poaceae) as inferred by nuclear rDNA ITS sequences. *Mol Phylogen Evol* 2004, **33**:289–299.
95. Niemeyer HM, Copaja SV, Barria BN: The Triticeae as sources of hydroxamic acids, secondary metabolites in wheat conferring resistance against aphids. *Hereditas* 1992, **116**:295–299.
96. Jakob SS, Meister A, Blattner FR: The considerable genome size variation of Hordeum species (Poaceae) is linked to phylogeny, life form, ecology, and speciation rates. *Mol Biol Evol* 2004, **21**:860–869.

doi:10.1186/1471-2148-12-64
Cite this article as: Dutartre *et al.*: Phylogenomics of the benzoxazinoid biosynthetic pathway of Poaceae: gene duplications and origin of the *Bx* cluster. *BMC Evolutionary Biology* 2012 **12**:64.