Combinatorial Evolution of a Terpene Synthase Gene Cluster Explains Terpene Variations in *Oryza*[1][OPEN]

Hao Chen,[a,2] Tobias G. Köllner,[b,2] Guanglin Li,[a] Guo Wei,[a] Xinlu Chen,[a] Dali Zeng,[c] Qian Qian,[c] and Feng Chen[a,3,4]

[a]Department of Plant Sciences, University of Tennessee, Knoxville, Tennessee 37996
[b]Department of Biochemistry, Max Planck Institute for Chemical Ecology, D-07745 Jena, Germany
[c]State Key Lab for Rice Biology, China National Rice Research Institute, Hangzhou 310006, China

ORCID IDs: 0000-0002-7037-904X (T.G.K.); 0000-0001-5977-0681 (G.W.); 0000-0002-7560-6125 (X.C.); 0000-0003-2349-8633 (D.Z.); 0000-0002-0349-4937 (Q.Q.); 0000-0002-3267-4646 (F.C.).

Terpenes are specialized metabolites ubiquitously produced by plants via the action of terpene synthases (TPSs). There are enormous variations in the types and amounts of terpenes produced by individual species. To understand the mechanisms responsible for such vast diversity, here we investigated the origin and evolution of a cluster of tandemly arrayed *TPS* genes in *Oryza*. In the *Oryza* species analyzed, *TPS* genes occur as a three-*TPS* cluster, a two-*TPS* cluster, and a single *TPS* gene in five, one, and one species, respectively. Phylogenetic analysis revealed the origins of the two-*TPS* and three-*TPS* clusters and the role of species-specific losses of *TPS* genes. Within the three-*TPS* clusters, one orthologous group exhibited conserved catalytic activities. The other two groups, both of which contained pseudogenes and/or nonfunctional genes, exhibited distinct profiles of terpene products. Sequence and structural analyses combined with functional validation identified several amino acids in the active site that are critical for catalytic activity divergence of the three orthologous groups. In the five *Oryza* species containing the three-*TPS* cluster, their functional *TPS* genes showed both conserved and species-specific expression patterns in insect-damaged and untreated plants. Emission patterns of volatile terpenes from each species were largely consistent with the expression of their respective *TPS* genes and the catalytic activities of the encoded enzymes. This study indicates the importance of combinatorial evolution of *TPS* genes in determining terpene variations among individual species, which includes gene duplication, retention/loss/degradation of duplicated genes, varying selection pressure, retention/divergence in catalytic activities, and divergence in expression regulation.

Due to their sessile nature, plants rely on chemicals to interact with the environment. Terpenes are often used by plants as their chemical languages in defenses against biotic and abiotic stresses or in the attraction of beneficial organisms (Gershenzon and Dudareva, 2007; Tholl, 2015). Plant terpenes are enormously diverse (Connolly and Hill, 1991). Critical for terpene structural diversity are typical plant terpene synthase (TPS) genes. Ubiquitous in land plants (Chen et al., 2011), *TPS* genes encode enzymes that catalyze the formation of monoterpenes, sesquiterpenes, and diterpenes using geranyl diphosphate (C_{10}), farnesyl diphosphate (FPP; C_{15}), and geranylgeranyl diphosphate (C_{20}) as substrate, respectively. Many TPSs produce multiple products, a feature mainly resulting from the carbocation-driven reaction mechanism catalyzed by these enzymes (Gao et al., 2012; Garms et al., 2012). Moreover, changes of a few or even a single amino acid in the active site of TPSs can have dramatic effects on catalytic activity and product specificities (Chen et al., 2014). Adding these features together, TPSs play an important role in the generation of the vast variations of terpene chemistry among land plants (Chen et al., 2003, 2004; Tholl et al., 2005; Matsuba et al., 2013).

There are variations in the numbers of *TPS* genes that different species of plants contain. For example, the moss *Physcomitrella patens* contains a single *TPS* gene (Chen et al., 2011). By contrast, most seed plants contain a midsized *TPS* gene family with approximately 20 to 100 members (Chen et al., 2011). Local gene duplication, which often results in tandem repeats, has been described as an important driver for the expansion of *TPS* gene families in plants (Chen et al., 2011). However, the number of tandem repeats varies among plant species. For example, none of the *TPS* genes in the

[1]This work was supported by the University of Tennessee, Institute of Agriculture (to F.C.), and the Max Planck Society (to T.G.K.).
[2]These authors contributed equally to the article.
[3]Author for contact: fengc@utk.edu.
[4]Senior author
The author responsible for distribution of materials integral to the findings presented in the Instructions for Authors (www.plantphysiol.org) is: Feng Chen (fengc@utk.edu).

H.C., T.G.K., Q.Q., and F.C. designed the research; H.C., T.G.K., G.-L.L., G.W., X.-L.C., and D.-L.Z. performed the experiments; H.C., T.G.K., G.-L.L., G.W., X.-L.C., D.-L.Z., F.C., Q.Q., and F.C. analyzed the data; H.C., T.G.K., and F.C. wrote the article with assistance from all the authors.

[OPEN]Articles can be viewed without a subscription.
www.plantphysiol.org/cgi/doi/10.1104/pp.19.00948

genome of seagrass (*Zostera marina*; Olsen et al., 2016) are localized in tandem repeats. By contrast, 13 of the 32 intact *TPS* genes in Arabidopsis (*Arabidopsis thaliana*) are arranged in tandem arrays (Aubourg et al., 2002), indicating a substantial role of tandem duplication in the expansion of the TPS family in this plant. The functions of some of these clustered *TPS* genes have been studied. For example, Arabidopsis contains a monoterpene synthase gene cluster formed by At3g25810, At3g25820, and At3g25830 (Chen et al., 2003, 2004). In this cluster, At3g25820 and At3g25830 are identical, indicating very recent gene duplication, Whereas At3g25820 and At3g25830, mainly expressed in roots, encode a multifunctional monoterpene synthase that produces mainly 1,8-cineole, two of the minor products of this enzyme, myrcene and (*E*)-*β*-ocimene, are the major products of At3g25810, which shows flower-specific expression (Chen et al., 2003), thus indicating functional divergence of the duplicated members. In another example, tomato (*Solanum lycopersicum*) possesses a sesquiterpene synthase gene cluster with six members (*TPS30*–*TPS35*; Falara et al., 2011). *TPS30* and *TPS34* in the tomato *TPS* gene cluster appear to be pseudogenes, whereas tomato *TPS31* and *TPS32* have been biochemically characterized to show divergent catalytic activities (Bleeker et al., 2011; Falara et al., 2011). The above examples illustrate that *TPS* genes in gene clusters undergo dynamic evolution leading to divergent functions. However, our understanding of the evolutionary processes leading to such *TPS* gene clusters is still limited (i.e. how such *TPS* gene clusters originated, and what the mechanisms that determine the fates of duplicated *TPS* genes are, including the divergence of their catalytic activities and biological functions). To gain insight into these questions, it is important to investigate *TPS* gene clusters in plant species with well-defined phylogeny.

In this study, we investigated the evolution of a *TPS* gene cluster in the genus *Oryza*. This gene cluster in the Asian rice *Oryza sativa* has three members: *Os08g07080*, *Os08g07100*, and *Os08g07120*. In our previous study, *Os08g07100* was shown to be involved in defense against insects (Yuan et al., 2008). When infested by the larvae of the generalist herbivore *Spodoptera frugiperda*, commonly known as fall armyworm, *O. sativa* plants emitted elevated levels of volatiles dominated by terpenes, which were demonstrated to attract the parasitoid *Cotesia marginiventris*. Os08g07100 is one of the major genes for synthesizing *S. frugiperda*-induced terpene volatiles in *O. sativa*. The functions of *Os08g07080* and *Os08g07120*, however, have not been characterized. Besides *O. sativa*, the genus *Oryza* contains *O. glaberrima*, known as African rice, and more than 20 wild relatives (Vaughan et al., 2003). Asian and African rice and their wild relatives have diversified across a broad range of ecological environments within the last 15 million years (Jacquemin et al., 2014), a relatively short evolutionary time scale. The well-defined phylogeny of these species as well as the availability of genome sequences for several members of this genus (Jacquemin et al., 2014; Stein et al., 2018) make *Oryza* an exceptional model for studying the short-term evolutionary dynamics of plants (Jacquemin et al., 2014). Here, we used the *Oryza* system to investigate the origin and evolution of the three-*TPS* gene cluster observed in *O. sativa*. For comparative functional analysis, we analyzed the emission of their volatile terpene products from respective *Oryza* plants in the context of plant-insect interactions. By uncovering the mechanisms underlying the evolution of the *Os08g07080-Os08g07100-Os08g07120* gene cluster in *Oryza*, it is also our hope that these mechanisms can be applicable for explaining the vast diversity of terpene chemistry among land plants.

RESULTS

Putative Orthologs of *Os08g07080*, *Os08g07100*, and *Os08g07120* in Other *Oryza* Species

To study the origin and evolution of the three-*TPS* gene cluster initially observed in *O. sativa*, six additional species of *Oryza* were selected for analysis. These included the cultivated rice *O. glaberrima* (African rice) and five wild rice species: *O. rufipogon* and *O. nivara* (the wild relatives of *O. sativa*), *O. barthii* (the wild relative of African rice), *O. punctata*, and *O. brachyantha*. There were three justifications for this selection. First, all these six species have sequenced nuclear genomes. Second, all six species are diploids, like *O. sativa*. Third, the six species together with *O. sativa* have well-defined phylogeny (Fig. 1). To identify putative orthologs of *Os08g07080*, *Os08g07100*, and *Os08g07120*, the scaffolds containing the top hits to *Os08g07080-Os08g07100-Os08g07120* from the six *Oryza* species were aligned and each collinear set of matching genes was drawn as a contiguously colored local collinear block. The alignment showed that the surrounding regions of *TPS* gene(s) were collinear

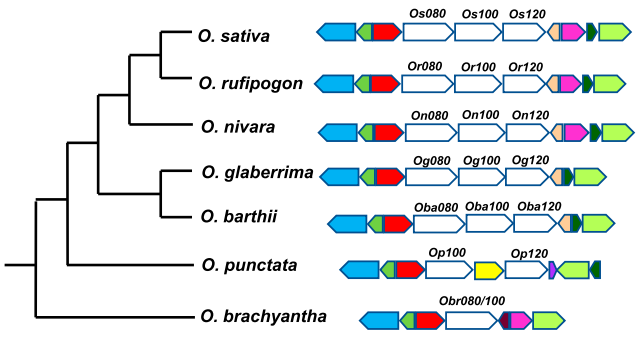

Figure 1. A *TPS* gene cluster in seven species of *Oryza*. The phylogeny of the seven *Oryza* species was adopted from Stein et al. (2018). The blocks depict the arrangement of tandemly arrayed *TPS* genes and their neighboring genes. Blocks of the same color represent apparent orthologous genes.

among all seven *Oryza* species (Fig. 1). Whereas five species contain three *TPS* genes in this region, two *TPS* genes were annotated for *O. punctata* and only one *TPS* gene was identified in *O. brachyantha* (Fig. 1). For simplicity, *Os08g07080*, *Os08g07100*, and *Os08g07120* were renamed as *Os080*, *Os100*, and *Os120*, respectively. The corresponding *TPS* genes in the other six *Oryza* species were renamed based on their sequence identities to *Os080*, *Os100*, and *Os120* (Supplemental Table S1). Genomic sequences from ATG to the stop codon covering introns for each of the 18 *TPS* genes were extracted and compared. *Or080*, *On080*, *Og080*, and *Oba080* share the highest sequence identities with *Os080*. *Or100*, *On100*, *Og100*, *Oba100*, and *Op100* share the highest sequence identities with *Os100*, whereas *Or120*, *On120*, *Og120*, *Oba120*, and *Op120* are most similar to *Os120*. *Obr080/100* shares a similar sequence identity with *Os100* and *Os080* (Supplemental Table S1).

Phylogenetic Analysis of *Oryza TPS* Genes and a Model for the Origin of the *TPS* Cluster

Phylogenetic analysis was performed for the 18 *TPS* genes identified from the seven species of *Oryza* using the maize (*Zea mays*) *TPS10* (*ZmTPS10*; Schnee et al., 2006) as outgroup. Phylogenetic reconstruction was first made with the sequences of genomic DNAs including introns. Two clades could be recognized: six "120"TPSs forming one clade and the rest forming the other clade (Fig. 2A). In the latter clade, six "100"TPSs grouped together and five "080"TPSs grouped together, with Obr080/100 being basal to the two groups

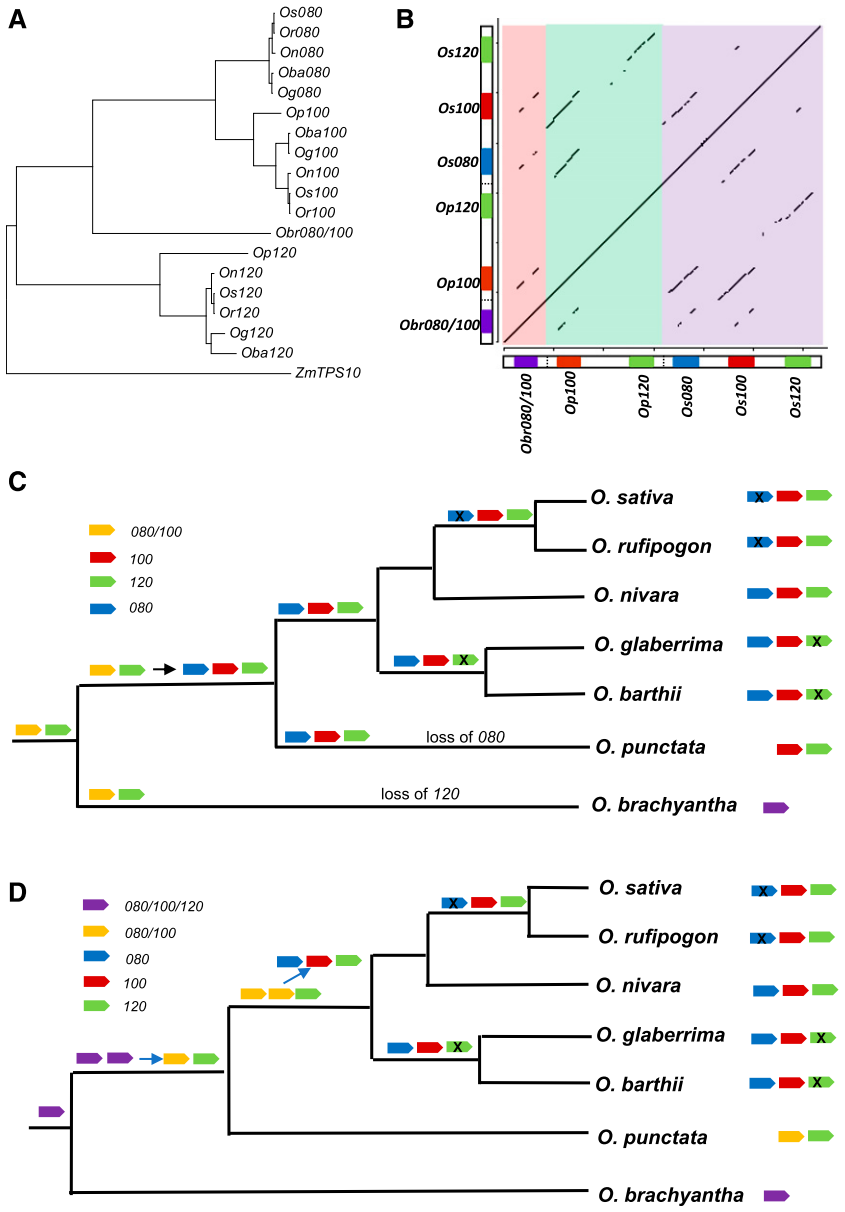

Figure 2. Evolutionary analysis of the *TPS* gene cluster. A, Phylogenetic tree of the *Oryza TPSs* based on DNA sequences of complete coding regions with *ZmTPS10* from maize as outgroup. B, Dot-plot analysis covering genomic DNAs for *Obr080/100* in *O. brachyantha*, *Op100+Op120* in *O. punctata*, and *Os080+Os100+Os120* in *O. sativa*. C and D, Two inferred models for the evolutionary trajectory of the *TPS* cluster in the seven species of *Oryza*. The blocks depict *TPS* genes. The same color indicates shared catalytic activities. Blocks with an X depict pseudogenes.

(Fig. 2A). Phylogenetic reconstruction was also made with annotated complementary DNA (cDNA) sequences of the 18 *TPS* genes, and a phylogenetic tree identical to the topology made with genomic DNA sequences was obtained (Supplemental Fig. S1).

Phylogenetic analysis implies that the evolution of the three-*TPS* cluster involved two rounds of local gene duplication, with the first round of gene duplication occurring before the divergence of the seven *Oryza* species and the duplication that leads to ancestral "100"*TPS* and "080"*TPS* occurring at a later time. Additional evidence to support this order of events was obtained from the block analysis of the genome sequences covering *Obr080/100* in *O. brachyantha*, *Op100* and *Op120* in *O. punctana*, and *Os080*, *Os100*, and *Os120* in *O. sativa* (Fig. 2B). *Obr080/100* showed comparable sequence similarity with *Os080* and *Os100* and a slightly higher similarity with *Op100*. *Op100* showed much higher sequence similarity with *Os080* and *Os100* than with *Obr080/100*, and its similarity with *Os080* and *Os100* was similar. *Op120* and *Os120* showed substantial sequence similarity to each other but not to others. These data support that "080"*TPS* and "100"*TPS* were derived from the second duplication event.

Next, a model was proposed to explain the possible origination and evolution of this *TPS* cluster in *Oryza* (Fig. 2C). In this model, the common ancestor of the seven *Oryza* species contains a two-*TPS* cluster consisting of the ancestral "080/100"*TPS* and "120"*TPS* genes. After the divergence of *O. brachyantha* from the common ancestor of the other six *Oryza* species, two events happened: one being the loss of ancestral "120"*TPS* in *O. brachyantha* and the other being the duplication of ancestral "080/100"*TPS*, giving rise to the ancestral three-*TPS* cluster in the common ancestor of the other six *Oryza* species. After the divergence of *O. punctata* from the common ancestor of the five *Oryza* species, ancestral "080"*TPS* was lost in *O. punctata* whereas the three-*TPS* clusters were retained in all subsequently evolved species (Fig. 2C). At the protein level, due to a premature stop codon, the cDNAs of *Os080* and *Or080* are translated into proteins of 45 amino acids, much shorter than functional TPSs. Similarly, the identical open reading frames of *Oba120* and *Og120* from the two African rice species possess three stop codons (Supplemental Fig. S2), which lead to truncated proteins. These four *TPS* genes were judged to be pseudogenes.

Since genes in different species may evolve at different rates (Wolf et al., 2009), a gene phylogeny may not accurately reflect the evolutionary relationships. An alternate model that did not involve gene loss may also explain the evolution of the *TPS* cluster (Fig. 2D). In this model, the common ancestor of the seven *Oryza* species contained a single *TPS* gene in this region. After the split from *O. brachyantha*, the ancestral *TPS* gene underwent a duplication. Then, another duplication occurred after the split from *O. punctata* in the common ancestor of the five species leading to a three-*TPS* cluster. The fact that no partial *TPS* genes were found in the region of the *TPS* cluster in *O. brachyantha* and *O. punctata* provides additional evidence for this alternate model.

Catalytic Activities of TPS Proteins

Full-length cDNAs for each of the 11 intact *TPS* genes from the five *Oryza* species within a three-*TPS* gene cluster were amplified from respective *Oryza* plants by reverse transcription (RT)-PCR. For *Op100* and *Op120* from *O. punctata* and *Obr080/100* from *O. brachyantha*, full-length cDNAs were synthesized. The obtained sequences were cloned into a protein expression vector and heterologously expressed in *Escherichia coli*. In our previous study, Os100 was demonstrated to be a sesquiterpene synthase and had no activity with geranylgeranyl diphosphate (Yuan et al., 2008). Phylogenetic analysis indicated that Os080 and Os120, like Os100, belong to the TPS-a subfamily (Supplemental Fig. S3), which is mainly associated with sesquiterpene synthases (Chen et al., 2011). At the protein sequence level, Os080, Os100, Os120, and their respective orthologs share high sequence similarity (Supplemental Fig. S4). Based on this evidence, it is sensible to assign all these TPSs as sesquiterpene synthases. As such, their respective recombinant proteins were tested for sesquiterpene synthase activity using FPP as substrate. The obtained data confirmed that all tested genes encode active sesquiterpene synthases with multiple sesquiterpene products, including 7-*epi*-sesquithujene, (*E*)-α-bergamotene, sesquisabinene A, (*E*)-β-farnesene, zingiberene, β-bisabolene, β-sesquiphellandrene, and (*E*)-γ-bisabolene (Fig. 3). Among all "100" orthologs, zingiberene was the most abundant product with the exceptions of On100, which produced β-sesquiphellandrene and zingiberene as codominant products, and Op100, which produced mainly β-bisabolene. Among the three intact "080" orthologs, Og080 did not show any activity with FPP. By contrast, On080 and Oba080 were able to convert FPP to sesquiterpenes (Fig. 3). On080 produced (*E*)-β-farnesene as the dominant product and β-bisabolene as a minor product. Oba080 produced a similar set of sesquiterpenes like "100" enzymes but with a clear distinction: (*E*)-β-farnesene, being a minor product of "100" enzymes, was the most abundant product of Oba080. All three intact "120"TPSs, Os120, Or120, and On120, encoded active sesquiterpene synthases. These all produced (*E*)-β-farnesene, β-bisabolene, and β-sesquiphellandrene, with β-bisabolene (Os120, Or120, and On120) or (*E*)-β-farnesene (Op120) being the major products.

Activity Restoration of Og080 and Os080

Comparative sequence analysis showed that the inactive enzyme Og080 possessed a unique Tyr at residue 527, whereas all active enzymes characterized in this study and some known sesquiterpene synthases from other species contained an Asp at the corresponding

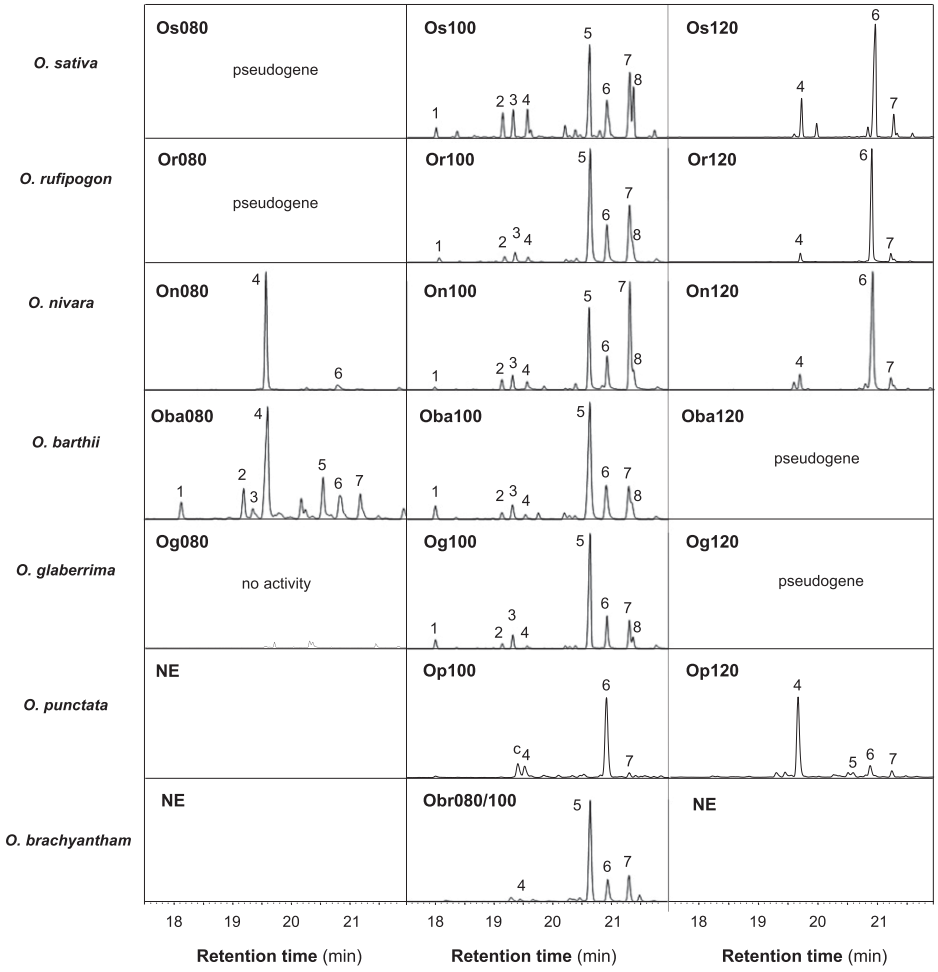

Figure 3. Catalytic activities of TPSs. Os080, Or080, Oba120, and Op120 are indicated as pseudogenes and were not analyzed. NE indicates that the respective gene is nonexistent. All other genes were expressed in *E. coli*, and partially purified recombinant proteins were assayed with the substrate FPP. The produced terpenes were separated and analyzed by gas chromatography-mass spectrometry (GC-MS). The traces of the MS detector are shown for the active enzymes. Products were identified as follows: 1, 7-*epi*-sesquithujene; 2, (*E*)-α-bergamotene; 3, sesquisabinene A; 4, (*E*)-β-farnesene; 5, zingiberene; 6, β-bisabolene; 7, β-sesquiphellandrene; 8, (*E*)-γ-bisabolene; C, nonterpene contamination.

position (Fig. 4A). To examine whether this amino acid change is responsible for the activity loss of Og080, we generated the mutant Og080-Y527D and subjected it to enzyme assays. With FPP as substrate, Og080-Y527D was active and produced a single product, (*E*)-β-farnesene (Fig. 4B).

By comparing the open reading frames of the two pseudogenes *Os080* and *Or080* with the functional "080"*TPSs*, it appeared that the premature translation of Os080 and Or080 was due to a four-nucleotide TAGC deletion (Fig. 4C). To verify this as the mechanism of pseudogenization, we inserted TAGC into the open reading frames of *Os080* and *Or080* by site-directed mutagenesis. Os080-TAGC exhibited similar activity to that of On080 and Og080-Y527D and produced (*E*)-β-farnesene as a single product (Fig. 4D). However, Or080-TAGC remained inactive, suggesting that additional mutations contribute to its activity loss.

Selection Pressure for the Lineages Derived from Gene Duplications

Within the five three-*TPS* clusters, all "100"*TPS* orthologs were functional genes with conserved catalytic activities (Fig. 3), whereas both "080"*TPS* and "120"*TPS* orthologous groups contained pseudogenes. This suggests that duplicated genes have undergone different selection pressures. We calculated the ω (dN/dS) values on the branch that leads to intact *TPS* genes. The ω value on the branch leading to "100"*TPS* orthologs was 0.22 (Fig. 5). By contrast, the ω values on branches leading to "080"*TPS* and "120"*TPS* orthologs were 0.46 and 0.25, respectively (Fig. 5). With "080"*TPSs* derived from the recent duplication, this suggests less selection pressure exerted on "080" genes, consistent with the recent loss of function of *Og080* (Fig. 4, A and B) and *Os080* (Fig. 4, C and D).

Identification of the Sites Responsible for Enzyme Functional Diversification

Although sharing a relatively high sequence similarity of 91%, On080 and Oba080 possessed different product specificities. Thus, they offered a useful system to pinpoint the key amino acids responsible for such functional divergence. Structure modeling of On080 and sequence comparisons showed that only four amino acids in the active site cavity differed between

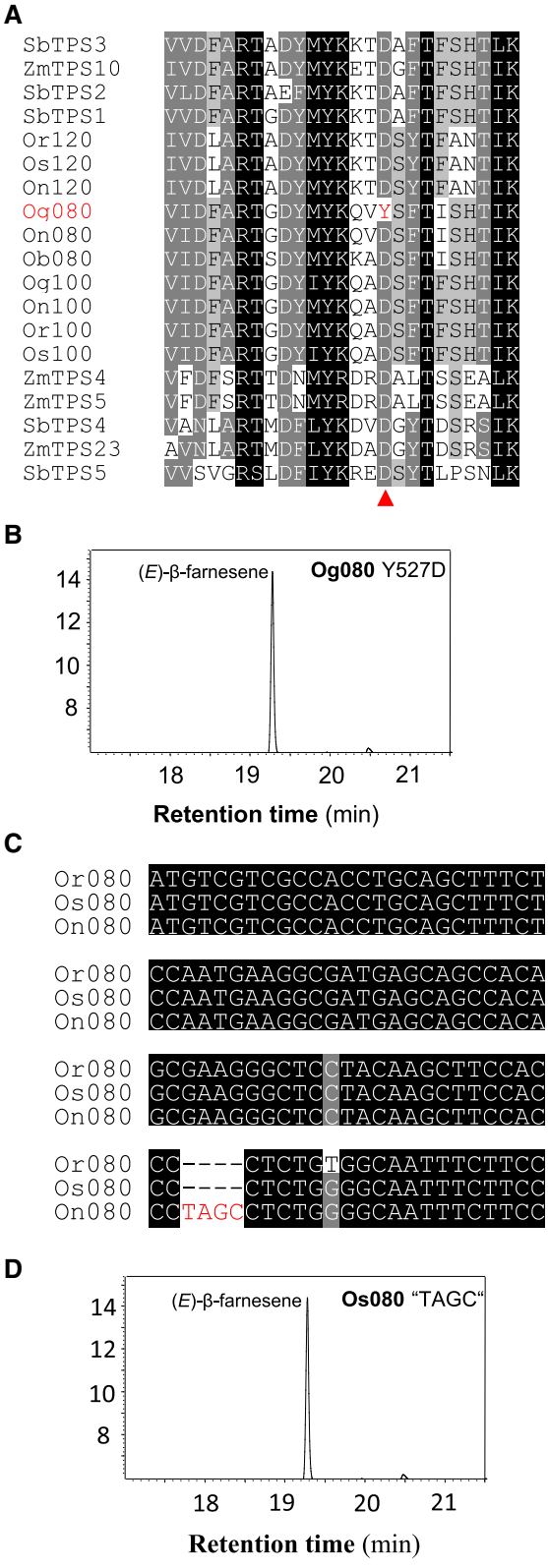

Figure 4. Activity restoration of Og080 and Os080. A, Comparative sequence analysis for residue 527 (referring to Ob080 and Og080) among the active enzymes characterized in this study and some known sesquiterpene TPS from other species. B, Activity of Og080-Y527D. C, Comparison of sequences of functional On080 and two nonfunctional Os080 and Or080 to indicate a deletion of TAGC in the two pseudogenes. D, Activity of Os080-TAGC.

the two enzymes (Fig. 6). The switches at these sites, I294T, F375S, E406Q, and G519S (referring to On080), may be responsible for the different product specificities of On080 and Oba080.

To verify the functions of the four amino acids, On080 was selected as target for the creation of a set of mutants by site-directed mutagenesis. A total of 16 mutants (2^4) covering all possible forms of conversion from On080 to Oba080 in the four amino acids were produced and subjected to sesquiterpene synthase enzyme assays. As shown in Figure 7A, besides (*E*)-β-farnesene, the single mutant On080-I294T also produced the minor product β-sesquiphellandrene, a monocyclic product of Ob080 and "100"TPSs, whereas the other three single mutants did not, indicating that this switch is important for the transition from acyclic (*E*)-β-farnesene to cyclic products. Excitingly, the triple mutant On080-I294T+F375S+G519S started producing (*E*)-α-bergamotene and sesquisabinene A, two bicyclic products of Ob080 and "100"TPSs. The tetra mutant On080-I294T+F375S+E406Q+G519S produced 7-*epi*-sesquithujene, another bicyclic product of Ob080 and "100"TPSs, indicating that the remaining three switches are important for the transition from monocyclic to bicyclic products.

Sequence comparison revealed that divergence at these positions also occurred among "100"TPSs, "120"TPSs, and "080"TPSs (Supplemental Fig. S3). We thus generated a series of Or100 mutants to attempt the conversions from Or100 to "120"TPSs and "080"TPSs. As shown in Figure 8A, the single mutant Or100-Q404E dominantly produced β-bisabolene, mimicking Or120 with the exception of the production of zingiberene. For the conversion from Or100 to "080"TPS, the proportion of (*E*)-β-farnesene, the dominant product of the "080"TPS, was increased in the products of either Or100-T292I or Or100-Y373F. The double mutant Or100-T292I+Y373F highly mimicked On080 (Fig. 8A). When a clustering analysis based on the proportion of each sesquiterpene was performed, Or100-T292I+Y373F clustered with On080. Or100-Q404E′, an artificial mutant derived from Or100-Q404E with zingiberene removed from its product profile, clustered with wild-type Or120 (Fig. 8B).

Expression Analysis of Functional *TPS* Genes in the Three-*TPS* Cluster-Containing Species of *Oryza* with and without Insect Herbivory

The existence of the three-*TPS* gene cluster in five closely related *Oryza* species raised an interesting question about the biological function for each of the *TPS* genes. One general function of TPSs is for plant

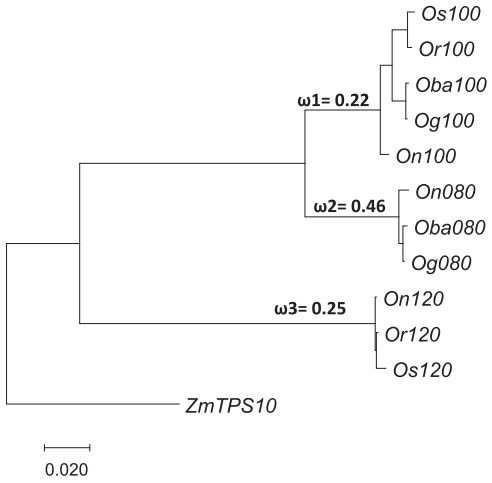

Figure 5. Analysis of selection pressure on three lineages leading to "100"TPSs, "080"TPSs, and "120"TPSs. These TPSs are intact TPSs from five *Oryza* species containing a three-TPS cluster.

defense against insects (Pichersky and Gershenzon, 2002; Dudareva et al., 2006). In our previous study with *O. sativa*, Os100 was shown to be involved in making insect-induced volatile terpenes for attracting the natural enemies of the attacking insects (Yuan et al., 2008). In this study, we analyzed the expression of the *TPS* genes in the aboveground tissues of 2-week-old seedlings of the five *Oryza* species in the context of plant-insect interactions. The expression of each of the functional *TPS* genes was compared in untreated plants and plants damaged by *S. frugiperda* feeding using reverse transcription quantitative PCR (RT-qPCR). The expression of "100"*TPS* genes in all analyzed species was significantly up-regulated by insect feeding (Fig. 9). The expression of "080"*TPS* was significantly up-regulated in *O. barthii* but down-regulated in *O. nivara* by insect feeding. In the three species in which the "120"*TPS* gene was analyzed, the expression of the "120"*TPS* gene was significantly up-regulated in *O. rufipogon* but remained unchanged in *O. sativa* and *O. nivara* upon herbivory (Fig. 9).

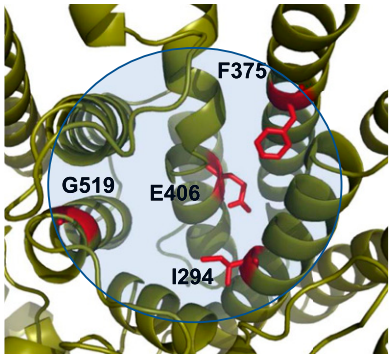

Figure 6. The active site cavity of On080. The model was created using the crystal structure of δ-cadinene synthase (Protein Data Bank identifier 3g4dA) from *Gossypium arboreum* (Gennadios et al., 2009). The image shows a view from the opening (illustrated as a blue circle) to the bottom of the active site cavity. Amino acids in the active site differing between On080 and Oba080 are shown in red.

Terpene Volatiles of *Oryza* Plants with and without Insect Herbivory

In addition to gene expression analysis, we performed chemical profiling of the five *Oryza* species for the emission of the eight sesquiterpenes as products of "080"TPS, "100"TPS, and "120"TPS. Whereas 7-*epi*-sesquithujene was not detected, all other seven sesquiterpenes, (*E*)-α-bergamotene, sesquisabinene A, (*E*)-β-farnesene, zingiberene, β-bisabolene, β-sesquiphellandrene, and (*E*)-γ-bisabolene, were detected from one or more species (Table 1). There were two observations. The first was that three species (*O. sativa*, *O. rufipogon*, and *O. glaberrima*) did not emit any of the sesquiterpenes without insect herbivory. The second was that for all five species, the emission of the sesquiterpenes was largely induced from insect-damaged plants compared with that from plants without insect herbivory.

DISCUSSION

The quality and quantity of terpenes that each species of plants produce are often unique to that species. TPSs, the pivotal enzymes of terpene biosynthesis, play a central role in determining the species-specific profile of terpenes. Apart from the ability of many TPSs to produce multiple products from a single substrate, continued *TPS* gene duplication followed by functional divergence has been regarded as critical for generating the chemical diversity of terpenes (Chen et al., 2011). Despite this general conception, concrete examples of the evolution of duplicated *TPS* genes across multiple plant species are scarce. Here, we systematically investigated the origin and dynamic evolution of a *TPS* gene cluster in seven *Oryza* species (Fig. 1). Our results revealed combinatorial evolution of duplicated *TPS* genes in a dynamic species/lineage-specific manner, which was largely responsible for the vast chemical diversity of terpenes among the *Oryza* plants in the context of plant-insect interactions.

Gain Versus Loss of *TPS* Genes: Dynamic Process in a Species/Lineage-Specific Manner

We proposed two models to explain the origin and evolutionary trajectory of the *TPS* clusters in *Oryza* (Fig. 2). One major difference between the two models is the timing of the very first gene duplication that leads to the two-TPS cluster. In one model, the two-*TPS* cluster evolved in the common ancestor of the seven *Oryza* species studied in this work (Fig. 2C). The subsequent evolution involved both gene duplication (in

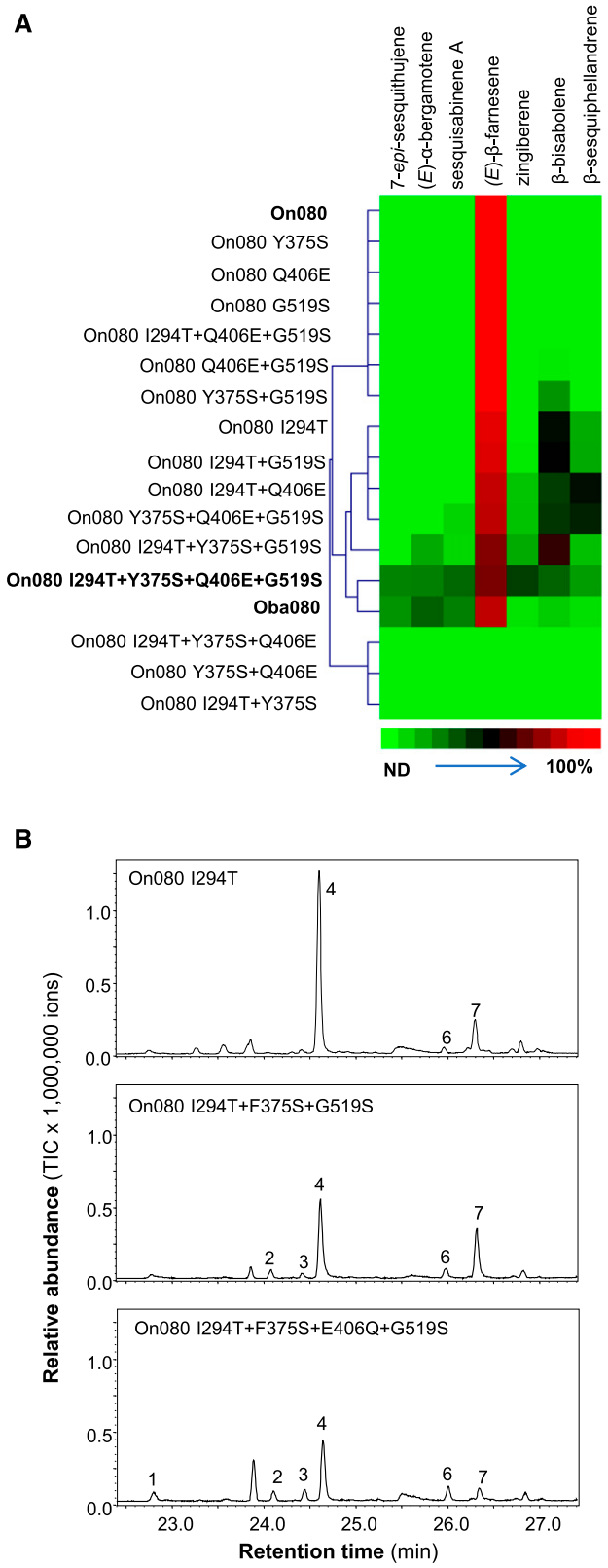

Figure 7. Activity analysis of On080 mutants. A, Clustering analysis of 16 On080 mutants together with On080 wild-type enzyme and Oba080 wild-type enzyme based on the proportions of their sesquiterpene products. B, Product profiles of three representative mutants On080-I294T, On080-I294T+F375S+G519S, and On080-I294T+F375S+E406Q+G519S. Peak numbers correspond to those in Figure 3: 1, 7-*epi*-sesquithujene; 2, (*E*)-α-bergamotene; 3, sesquisabinene A; 4, (*E*)-β-farnesene; 6, β-bisabolene; 7, β-sesquiphellandrene. TIC, Total ion chromatogram; ND, not detected or 0%.

the common ancestor of the six *Oryza* species after the split with *O. brachyantha*) and multiple gene losses (independent gene losses in *O. brachyantha* and *O. punctata*). In the alternate model, the two-*TPS* cluster evolved in the common ancestor of the six *Oryza* species after divergence from *O. brachyantha* (Fig. 2D). In a previous study (Zhuang et al., 2012), the *Os080-Os100-Os120* three-*TPS* cluster was inferred to have evolved after the split of rice and sorghum (*Sorghum bicolor*). Therefore, even if the first model reflects the real evolutionary history, the first duplication to form a two-TPS cluster must have occurred within *Oryza*. Therefore, regardless of the real trajectory, dynamic changes have occurred to the duplicated *TPS* genes in these seven *Oryza* species as revealed in this study. With or without physical gene losses, functional loss of *TPS* genes (i.e. pseudogenization) is apparent in all five species within a three-*TPS* cluster except in *O. nivara*.

Functional Diversification of the Duplicated *TPS* Genes: A Perspective of Reaction Mechanisms

With the three members of the *TPS* gene cluster demonstrated to encode enzymes with different catalytic activities, it is interesting to ask what the catalytic activity of the common ancestor of these three TPSs was. In all *Oryza* species examined, the "100"*TPS* orthologs encode functional enzymes (Fig. 3), and this *TPS* lineage was under more stringent selection. It is thus sensible to suggest that "100"TPSs carry the ancestral, critical activities. With this proposal, it is interesting to ask how the divergence in catalytic activities of the three types of TPSs has been achieved. We may look at this question from the perspective of reaction mechanisms. The biosynthetic pathways for the majority of the identified products of the three TPSs in the cluster were schemed based on the carbocationic mechanisms proposed for other sesquiterpene synthases (Fig. 10; Köllner et al., 2004). In a first step, the diphosphate moiety of FPP is cleaved off and the resulting farnesyl cation can be deprotonated to form acyclic (*E*)-β-farnesene, the dominant product of "080"TPS orthologs. Alternatively, the farnesyl cation can be isomerized via the formation of nerolidyl diphosphate and a subsequent C2-C3 bond rotation. Intramolecular attack on the proximal double bond then leads to the formation of a new C-C bond between carbons 1 and 6 (numbered as for FPP) and a deprotonation of the resulting bisabolyl cation yields β-bisabolene, the dominant product of "120" orthologs. The bisabolyl cation can also undergo a 4,7-hydride shift or a 1,7-hydride

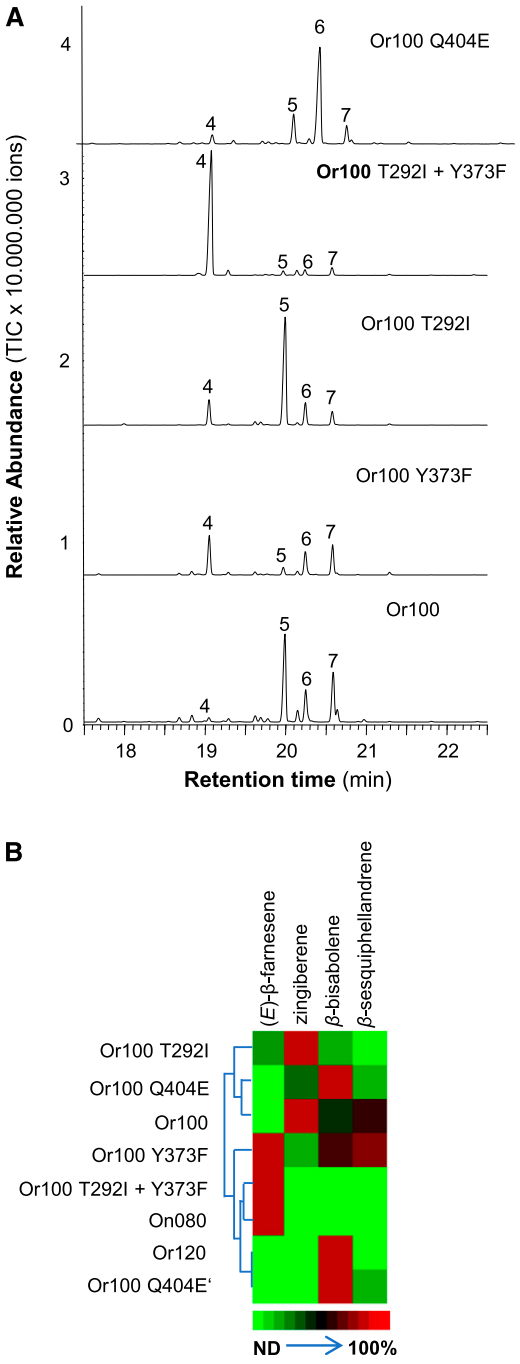

Figure 8. Activity analysis of Or100 mutants. A, Product profiles of four Or100 mutants (Or100-Y373F, Or100-T292I, Or100-T292I+Y373F, and Or100-Q404E) together with Or100 wild-type enzyme. Peak numbers correspond to those in Figure 3: 4, (*E*)-*β*-farnesene; 5, zingiberene; 6, *β*-bisabolene; 7, *β*-sesquiphellandrene. B, Clustering analysis of four Or100 mutants together with On080, Or100, and Or120 wild-type enzymes and one artificial mutant Or100-Q404E′ based on the proportions of their sesquiterpene products. Or100-Q404E′ was based on Or100-Q404E with zingiberene removed from the product profile. TIC, Total ion chromatogram; ND, not detected or 0%.

shift that, after deprotonation of the respective carbocations, leads to the formation of zingiberene or *β*-sesquiphellandrene, respectively. Both compounds are monocyclic sesquiterpenes and dominant products of the majority of the "100" orthologs. Besides, the bisabolyl cation can undergo other downstream

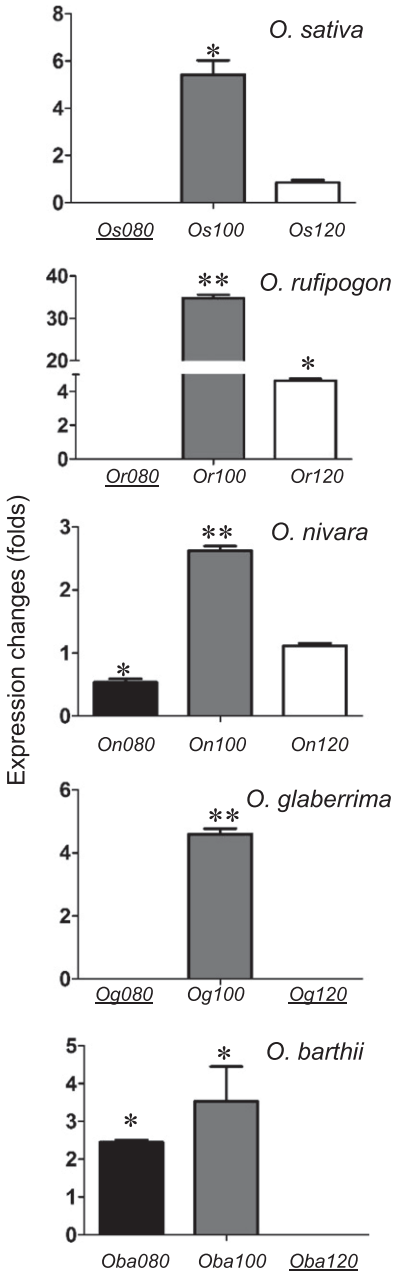

Figure 9. Expression analysis of functional *TPS* genes of the three-*TPS* cluster in the five species of *Oryza*. Genes that are underlined indicate nonfunctional genes, and the expression for such genes was not measured. The expression of individual functional *TPS* genes in each species was quantified as fold change in insect-damaged plants versus undamaged control plants using RT-qPCR. Data are presented as means ± SD ($n = 3$). Asterisks indicate statistical significance at $P \leq 0.05$ (*) and $P \leq 0.01$ (**).

Table 1. *Sesquiterpene volatiles as products of "080/100/120"TPS emitted from the plants of the five* Oryza *species with or without insect herbivory*

S1 to S5 are (*E*)-α-bergamotene, sesquisabinene A, (*E*)-β-farnesene, zingiberene, and β-bisabolene, respectively. S6/S7 represents two coeluted compounds, β-sesquiphellandrene and (*E*)-γ-bisabolene. ND, not detected

Species	Treatment	Sesquiterpenes[a]					
		S1	S2	S3	S4	S5	S6/S7
O. sativa	Control	ND	ND	ND	ND	ND	ND
	Herbivory	44 ± 1.0	47 ± 2.5	13 ± 1.7	123 ± 3.5	52 ± 2.7	108 ± 3.7
O. rufipogon	Control	ND	ND	ND	ND	ND	ND
	Herbivory	9.6 ± 0.2	17.4 ± 0.7	6.6 ± 0.1	155.6 ± 2.0	52.0 ± 0.8	81.0 ± 2.0
O. nivara	Control	ND	ND	3.8 ± 0.3	2.7 ± 0.1	ND	1.2 ± 0.1
	Herbivory	ND	2.2 ± 0.15	5.6 ± 0.3	7.0 ± 0.2	6.2 ± 0.1	5.5 ± 0.1
O. glaberrima	Control	ND	ND	ND	ND	ND	ND
	Herbivory	1.2 ± 0.2	1.9 ± 0.1	ND	13.4 ± 0.5	4.0 ± 0.1	4.0 ± 0.1
O. barthii	Control	140.8 ± 4.0	15.2 ± 0.2	377.0 ± 6.4	ND	22.0 ± 2.0	8.9 ± 0.5
	Herbivory	211.8 ± 0.9	36.7 ± 1.6	689.3 ± 2.7	10.7 ± 0.2	54.9 ± 1.1	25.3 ± 1.7

[a]Emission rates were calculated from three replicates as ng h^{-1} g^{-1} fresh weight. Control and Herbivory indicate untreated plants and *S. frugiperda*-damaged plants, respectively.

modifications, including a C2-C7 ring closure, a 6,7-hydride shift followed by a C2-C6 ring closure, and related deprotonations, that result in the formation of other minor products of the "100" orthologs. Based on this reaction mechanism and the phylogeny shown in Figure 2, it is tempting to speculate that the "100" ancestor was already able to catalyze various downstream modifications of the bisabolyl cation, whereas both the "120"TPS orthologs and the "080"TPS orthologs lost this ability during neofunctionalization. In addition, On080 even lost its ability to catalyze the initial isomerization or cyclization and thus produces only acyclic (*E*)-β-farnesene. Guided by homology-based structural modeling (Fig. 6), our data from site-directed mutagenesis (Figs. 7 and 8) indicate the importance of key amino acids in the active site in determining the functional divergence of the three TPS orthologous groups (Fig. 3).

Expression Regulation and Biological Function of the *TPS* Genes

At the level of gene expression, both conservation and divergence were observed for the functional three-*TPS* gene cluster in the five *Oryza* species. For the "100"*TPS* orthologs, all transcript levels were significantly induced by the feeding of fall armyworm (Fig. 9).

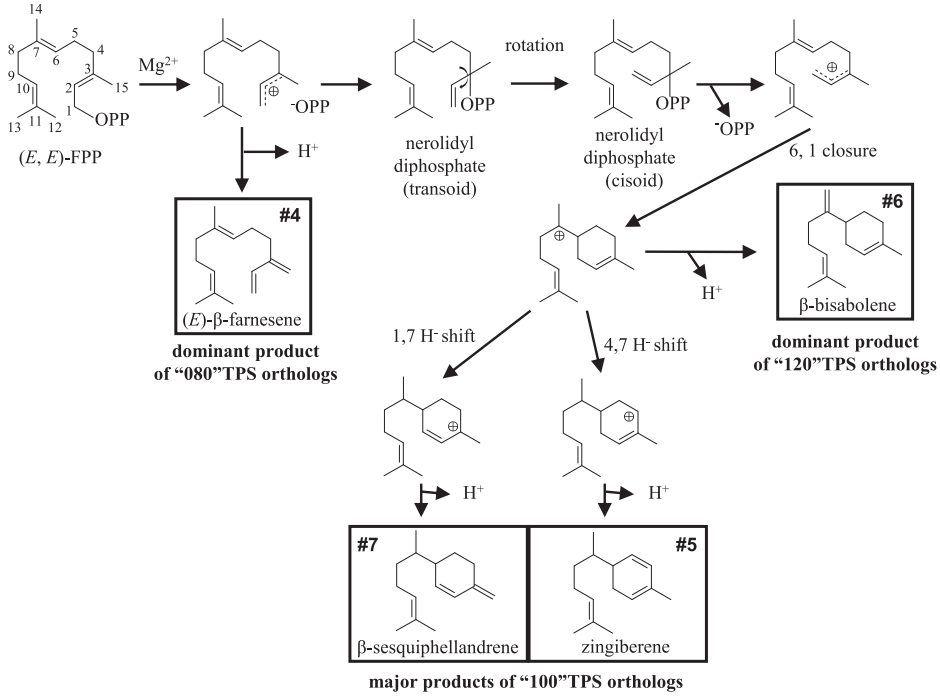

Figure 10. Scheme for the biochemical reaction mechanisms involving the functional divergence of three orthologous groups of TPSs ("080"TPSs, "100"TPSs, and "120"TPSs). The four labeled numbers (#4, #5, #6, and #7) correspond to the peak numbers in Figure 3.

In *O. sativa*, it has been demonstrated that the elevated emission of volatiles, including the products of Os100, after insect herbivory serves as a chemical signal for attracting the natural enemies of the herbivores (Yuan et al., 2008). This strongly suggests that the orthologs of Os100 have similar biological functions in the other four *Oryza* species. The emission patterns of the major sesquiterpene products of "080," "100," and "120" orthologs (Table 1) from each species were largely consistent with the expression patterns of their respective *TPS* genes (Fig. 9) and the catalytic activities of the encoded enzymes (Fig. 3). Among the five *Oryza* species, however, there are large differences in the types and amounts of sesquiterpenes emitted from control and insect-damaged plants (Table 1).

Such differences may reflect adaptations to different environments with varying herbivore pressures. Among the five *Oryza* species, three are wild species (*O. rufipogon*, *O. nivara*, and *O. barthii*). It is interesting that the gene expression patterns and terpene emission patterns of the Asian rice *O. sativa* were highly similar to those of its wild relative *O. rufipogon* (Fig. 9; Table 1). By contrast, the gene expression patterns and terpene emission patterns of African rice *O. glaberrima* were extremely different from those of its wild relative *O. barthii* (Fig. 9). This may reflect the divergence of terpene chemistry between *O. glaberrima* and *O. barthii*. At the gene level, both *O. sativa* and *O. rufipogon* contain "*080*" as a pseudogene. By contrast, both *O. glaberrima* and *O. barthii* contain "*120*" as a pseudogene. These observations are consistent with the proposition that *O. sativa* was domesticated from *O. rufipogon* whereas *O. glaberrima* was domesticated from *O. barthii* (Vaughan et al., 2003). It is interesting that such variations in *TPS* genes and terpene biosynthesis may have important agricultural implications. For example, many North American maize lines, unlike European lines, cannot produce (*E*)-β-caryophyllene in roots when damaged by western corn rootworm (*Diabrotica virgifera virgifera*). Consequently, they lack the ability to attract beneficial nematodes for the control of the western corn rootworm (Rasmann et al., 2005). Therefore, from an application perspective, the identification of diverse functional *TPS* genes in Asian and African rice may help guide breeding programs to develop new varieties for improved resistance against insects, which is a major constraint in rice production (Chen et al., 2012).

CONCLUSION

In summary, by linking the evolutionary trajectory of the *TPS* gene cluster, *TPS* gene expression, and the terpene chemistry of *Oryza* plants, it becomes evident that the diverse terpene chemical phenotypes among the five *Oryza* species examined in the context of plant-insect interactions in this study are determined by combinatorial evolution of *TPS* genes at multiple levels: gene duplication, retention/loss/degradation of duplicated genes, varying selection pressure, retention/divergence in catalytic activities, and divergence in gene expression regulation. In most sequenced land plants, *TPS* genes exist as midsized gene families that resulted from repeated gene duplications. The repeated processes of combinatorial evolution for *TPS* genes as revealed in this study can explain in large part the vast chemical diversity of terpenes among land plants.

MATERIALS AND METHODS

Seed Sources, Plant Growth, and Insect Treatment

Seeds of *Oryza sativa* 'Nipponbare' were kindly provided by the Dale Bumpers National Rice Research Center. Seeds of *Oryza rufipogon* (PI 590418), *Oryza nivara* (PI 590425), *Oryza barthii* (GSOR311687), and *Oryza glaberrima* (PI 450396) were obtained from the National Plant Germplasm System of the United States. The seeds were dehulled, placed on germination paper, and kept in an incubator at 28°C. After 3 d, seedlings were planted in 200-mL glass jars filled with soil with eight plants per jar. Plants were grown in a growth chamber at 28°C under a 14-h-light/10-h-dark cycle. For herbivory treatment, 2-week-old seedlings of each *Oryza* species were subject to feeding by *Spodoptera frugiperda*. For each treatment, two second-instar larvae of *S. frugiperda* were placed onto each seedling at 4 PM. After 18 h of feeding, the insects were removed and the damaged seedlings were used for either volatile profiling (whole plants in the jar) or gene expression analysis (aboveground tissues). Undamaged seedlings were used as control plants. Each treatment was conducted with three biological replicates.

Sequence Retrieval and Analysis

The sequences of the seven *Oryza* genomes analyzed in this study were obtained from EmsenblGenomes (http://www.ensembl.org). The genomic sequences of Os080, Os100, and Os120 were used as individual queries to search against the genomic sequences of *O. rufipogon*, *O. nivara*, *O. barthii*, *O. glaberrima*, *Oryza punctata*, and *Oryza brachyantha* using National Center for Biotechnology Information BLAST+ (Camacho et al., 2009). After the identification of a top hit for each query, the genomic sequences covering the putative *TPS* gene and its neighboring genes were extracted for each species and subject to manual annotation and synteny analysis. For phylogenetic analysis, either full-length cDNA or genomic sequences of *TPS* genes were used. Multiple sequence alignments were performed using MAFFT (v7.450; Katoh and Standley, 2013) in accurate strategy (L-INS-i) with 1,000 iterations of improvement. Phylogenetic trees were constructed using RAxML (Stamatakis, 2014) with 1,000 bootstrap replicates. For dot-blot analysis, genomic DNA sequences covering the respective *TPS* gene or *TPS* gene cluster from *O. brachyantha*, *O. punctata*, and *O. sativa* were extracted, combined as a single DNA fragment, and subject to dot-blot analysis using TBtools_JRE1.6 with default settings (Chen et al., 2018).

Gene Cloning and RT-qPCR

Total RNAs were isolated from appropriate plant tissues using the Plant RNeasy Mini Kit (Qiagen). An on-column DNase (Qiagen) treatment was applied to remove DNA contamination. First-strand cDNAs were synthesized from respective total RNAs using the High Capacity cDNA Reverse Transcription Kit (Thermo Fisher) and used for gene cloning and RT-qPCR analysis. A full-length cDNA for each *TPS* gene was amplified using RT-PCR, cloned into the vector pEXP5-CT/TOPO (Invitrogen), and fully sequenced. The primers used for cDNA cloning were 5′-ATGTCATCGACACCTGCAG-3′ (forward) and 5′-ATGATCAGCAGCCACCATCAC-3′ (reverse) for "100"*TPS* orthologs, 5′-ATGGATAGAATGGTAATCAAACC-3′ (forward) and 5′-AATGCTCAG AACTATATCGG-3′ (reverse) for "120"*TPS* orthologs, and 5′-ATGTCGTCG CCACCTGCAG-3′ (forward) and 5′-CCATGGAAACTTTCTAATGAGG-3′ (reverse) for "080"*TPS* orthologs.

For RT-qPCR, one pair of gene-specific primers was designed for each orthologous group of *TPS* genes (Supplemental Table S2) and tested with full-length cDNAs of three *TPS* genes of each species as template in PCR to verify the target sequence and gene specificity of each pair of primers. qPCR was performed on the QuantStudio6 Flex Real-Time PCR System (Thermo Fisher

Scientific) on a Microamp Fast 96-Well Reaction Plate (Thermo Fisher Scientific) using power SYBR Green PCR Master (Thermo Fisher Scientific). Relative levels of gene expression were calculated using the $2^{-\Delta\Delta CT}$ method (Yuan et al., 2006). The sequences of the primers for target *TPS* genes and the rice ubiquitin gene *OsUBQ5* as an internal control are listed in Supplemental Table S2. Three biological replicates were analyzed for each sample. Statistical analysis was conducted using Student's *t* test in SPSS version 25.

Protein Expression in *Escherichia coli* and TPS Assay

An *E. coli* BL21 Codon Plus strain (Invitrogen) was used for the expression of recombinant TPS enzymes. Expression was induced by the addition of isopropyl-1-thio-D-galactopyranoside to a final concentration of 1 mM. After this, cultures were grown for 17 h at 18°C. Cells were collected by centrifugation at 4,000*g* for 6 min and disrupted by a 4 × 30-s treatment with a sonicator in chilled extraction buffer (50 mM MOPSO, pH 7, 5 mM $MgCl_2$, 5 mM sodium ascorbate, 0.5 mM phenylmethylsulfonyl fluoride, 5 mM dithiothreitol, and 10% [v/v] glycerol). Cell fragments were removed by centrifugation at 14,000*g*, and the supernatant was desalted into assay buffer (10 mM MOPSO, pH 7, 1 mM dithiothreitol, and 10% [v/v] glycerol) by passage through an Econopac 10DG column (Bio-Rad). Enzyme assays were performed in a Teflon-sealed, screw-capped 1-mL GC glass vial containing 50 μL of the bacterial extract and 50 μL of assay buffer with 10 μM (*E*,*E*)-FPP, 10 mM $MgCl_2$, 0.2 mM $NaWO_4$, and 0.1 mM NaF. A solid-phase microextraction fiber consisting of 100-μm polydimethylsiloxane (Supelco) was placed into the headspace of the vial for 60 min of incubation at 30°C and then inserted into the injector of the gas chromatograph for analysis of the adsorbed reaction products. Terpenes were analyzed using an Agilent 6890 series gas chromatograph (Agilent Technologies) coupled to an Agilent 5973 quadrupole mass selective detector (interface temperature, 250°C; quadrupole temperature, 150°C; source temperature, 230°C; electron energy, 70 eV). Separation was performed on a Restek SHR5XLB column (30 m × 0.25 mm internal diameter × 0.25 μm thickness; Shimadzu). Helium was used as carrier gas (flow rate of 1 mL min^{-1}), a splitless injection (injector temperature, 250°C) was used, and a temperature gradient of 5°C min^{-1} from 40°C (3-min hold) to 240°C was applied. Sesquiterpenes were identified by comparing the retention times and mass spectra of those assay products with authentic compounds obtained from Fluka, Sigma, and Dr. Wilfried A. König.

Homology-Based Structural Modeling

Protein structure modeling was performed with the SWISS-MODEL service (https://swissmodel.expasy.org/) using the crystal structure of δ-cadinene synthase (Protein Data Bank identifier 3g4dA) from *Gossypium arboreum* (Gennadios et al., 2009) as a modeling template. Models were visualized and analyzed using the program PyMOL (Guindon et al., 2010).

Site-Directed Mutagenesis

The QuickChange site-directed mutagenesis kit (Agilent) was used to generate mutants of TPSs (for primer sequence information, see Supplemental Table S2). All mutant *TPS* genes were verified by sequencing.

Molecular Evolutionary Analysis

To perform phylogeny reconstruction, multiple protein sequence alignment was performed by using the ClustalX2 program (Larkin et al., 2007). The phylogenetic tree based on the alignment was constructed using the neighbor-joining method in MEGA7 using default parameters (Kumar et al., 2016). Based on the phylogenetic tree, *Oryza* TPSs were divided into three clades. To evaluate variations in selective pressure between them, branch models of CODEML in PAML (Yang, 2007) were used to estimate the ratio of nonsynonymous versus synonymous substitutions (ω = dN/dS) among the branches of three major clades.

Plant Volatile Collection and Identification

Volatiles emitted from the insect-damaged plants and control plants were collected using a headspace sampler device (Analytical Research Systems). A glass jar with control or insect-damaged *Oryza* plants was placed in a collection chamber, and volatiles were collected using a Super Q Volatile Collection Trap (Analytical Research System) for 4 h. Volatiles were eluted using 100 μL of methylene chloride containing 0.003% (v/v) 1-octanol as an internal standard for quantification. Plant volatiles were analyzed using a Shimadzu 17A gas chromatograph coupled to a Shimadzu QP5050A quadrupole mass selective detector under the conditions described above for GC-MS analysis of terpene products of TPS enzymes.

Accession Numbers

Sequence data from this article can be found in the GenBank/National Center for Biotechnology Information databases under the following accession numbers: MN607010 to MN607027.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Phylogenetic tree of *TPSs* based on cDNA sequences.

Supplemental Figure S2. *Oba120* and *Og120* as pseudogenes.

Supplemental Figure S3. A phylogenetic tree of all *TPS* genes from *O. sativa* and Arabidopsis.

Supplemental Figure S4. Multiple sequence alignment of *Oryza* TPSs.

Supplemental Table S1. Sequence identities of *TPS* genes.

Supplemental Table S2. Primers used for gene expression analysis by RT-qPCR.

ACKNOWLEDGMENTS

We thank the Dale Bumpers National Rice Research Center and the National Plant Germplasm System for providing us the seeds of *Oryza* species used in this study as well as Natascha Rauch for excellent technical assistance.

Received August 1, 2019; accepted October 28, 2019; published November 11, 2019.

LITERATURE CITED

Aubourg S, Lecharny A, Bohlmann J (2002) Genomic analysis of the terpenoid synthase (*AtTPS*) gene family of *Arabidopsis thaliana*. Mol Genet Genomics **267:** 730–745

Bleeker PM, Spyropoulou EA, Diergaarde PJ, Volpin H, De Both MTJ, Zerbe P, Bohlmann J, Falara V, Matsuba Y, Pichersky E, et al (2011) RNA-seq discovery, functional characterization, and comparison of sesquiterpene synthases from *Solanum lycopersicum* and *Solanum habrochaites* trichomes. Plant Mol Biol **77:** 323–336

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: Architecture and applications. BMC Bioinformatics **10:** 421

Chen C, Chen H, He Y, Xia R (2018) TBtools, a toolkit for biologists integrating various biological data handling tools with a user-friendly interface. bioRxiv 289660

Chen F, Ro DK, Petri J, Gershenzon J, Bohlmann J, Pichersky E, Tholl D (2004) Characterization of a root-specific Arabidopsis terpene synthase responsible for the formation of the volatile monoterpene 1,8-cineole. Plant Physiol **135:** 1956–1966

Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J **66:** 212–229

Chen F, Tholl D, D'Auria JC, Farooq A, Pichersky E, Gershenzon J (2003) Biosynthesis and emission of terpenoid volatiles from Arabidopsis flowers. Plant Cell **15:** 481–494

Chen H, Li G, Köllner TG, Jia Q, Gershenzon J, Chen F (2014) Positive Darwinian selection is a driving force for the diversification of terpenoid biosynthesis in the genus *Oryza*. BMC Plant Biol **14:** 239

Chen H, Stout MJ, Qian Q, Chen F (2012) Genetic, molecular and genomic basis of rice defense against insects. Crit Rev Plant Sci **31:** 74–91

Connolly JD, Hill RA (1991) Dictionary of Terpenoids. Chapman & Hall, London

Dudareva N, Negre F, Nagegowda DA, Orlova I (2006) Plant volatiles: Recent advances and future perspectives. Crit Rev Plant Sci **25:** 417–440

Falara V, Akhtar TA, Nguyen TTH, Spyropoulou EA, Bleeker PM, Schauvinhold I, Matsuba Y, Bonini ME, Schilmiller AL, Last RL, et al (2011) The tomato terpene synthase gene family. Plant Physiol **157:** 770–789

Gao Y, Honzatko RB, Peters RJ (2012) Terpenoid synthase structures: A so far incomplete view of complex catalysis. Nat Prod Rep **29:** 1153–1175

Garms S, Chen F, Boland W, Gershenzon J, Köllner TG (2012) A single amino acid determines the site of deprotonation in the active center of sesquiterpene synthases SbTPS1 and SbTPS2 from *Sorghum bicolor*. Phytochemistry **75:** 6–13

Gennadios HA, Gonzalez V, Di Costanzo L, Li A, Yu F, Miller DJ, Allemann RK, Christianson DW (2009) Crystal structure of (+)-delta-cadinene synthase from *Gossypium arboreum* and evolutionary divergence of metal binding motifs for catalysis. Biochemistry **48:** 6175–6183

Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. Nat Chem Biol **3:** 408–414

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst Biol **59:** 307–321

Jacquemin J, Ammiraju JSS, Haberer G, Billheimer DD, Yu Y, Liu LC, Rivera LF, Mayer K, Chen M, Wing RA (2014) Fifteen million years of evolution in the *Oryza* genus shows extensive gene family expansion. Mol Plant **7:** 642–656

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol **30:** 772–780

Köllner TG, Schnee C, Gershenzon J, Degenhardt J (2004) The variability of sesquiterpenes emitted from two *Zea mays* cultivars is controlled by allelic variation of two terpene synthase genes encoding stereoselective multiple product enzymes. Plant Cell **16:** 1115–1131

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol **33:** 1870–1874

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics **23:** 2947–2948

Matsuba Y, Nguyen TTH, Wiegert K, Falara V, Gonzales-Vigil E, Leong B, Schäfer P, Kudrna D, Wing RA, Bolger AM, et al (2013) Evolution of a complex locus for terpene biosynthesis in *Solanum*. Plant Cell **25:** 2022–2036

Olsen JL, Rouzé P, Verhelst B, Lin YC, Bayer T, Collen J, Dattolo E, De Paoli E, Dittami S, Maumus F, et al (2016) The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. Nature **530:** 331–335

Pichersky E, Gershenzon J (2002) The formation and function of plant volatiles: Perfumes for pollinator attraction and defense. Curr Opin Plant Biol **5:** 237–243

Rasmann S, Köllner TG, Degenhardt J, Hiltpold I, Toepfer S, Kuhlmann U, Gershenzon J, Turlings TCJ (2005) Recruitment of entomopathogenic nematodes by insect-damaged maize roots. Nature **434:** 732–737

Schnee C, Köllner TG, Held M, Turlings TCJ, Gershenzon J, Degenhardt J (2006) The products of a single maize sesquiterpene synthase form a volatile defense signal that attracts natural enemies of maize herbivores. Proc Natl Acad Sci USA **103:** 1129–1134

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

Stein JC, Yu Y, Copetti D, Zwickl DJ, Zhang L, Zhang C, Chougule K, Gao D, Iwata A, Goicoechea JL, et al (2018) Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. Nat Genet **50:** 285–296

Tholl D (2015) Biosynthesis and biological functions of terpenoids in plants. Adv Biochem Eng Biotechnol **148:** 63–106

Tholl D, Chen F, Petri J, Gershenzon J, Pichersky E (2005) Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from Arabidopsis flowers. Plant J **42:** 757–771

Vaughan DA, Morishima H, Kadowaki K (2003) Diversity in the *Oryza* genus. Curr Opin Plant Biol **6:** 139–146

Wolf YI, Novichkov PS, Karev GP, Koonin EV, Lipman DJ (2009) The universal distribution of evolutionary rates of genes and distinct characteristics of eukaryotic genes of different apparent ages. Proc Natl Acad Sci USA **106:** 7273–7280

Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol **24:** 1586–1591

Yuan JS, Köllner TG, Wiggins G, Grant J, Degenhardt J, Chen F (2008) Molecular and genomic basis of volatile-mediated indirect defense against insects in rice. Plant J **55:** 491–503

Yuan JS, Reed A, Chen F, Stewart CN Jr. (2006) Statistical analysis of real-time PCR data. BMC Bioinformatics **7:** 85

Zhuang X, Köllner TG, Zhao N, Li G, Jiang Y, Zhu L, Ma J, Degenhardt J, Chen F (2012) Dynamic evolution of herbivore-induced sesquiterpene biosynthesis in sorghum and related grass crops. Plant J **69:** 70–80