TITLE:

Evolution and diversification of the momilactone biosynthetic gene cluster in the genus *Oryza*

AUTHORS:

Santiago Priego-Cubero[1], Tomonobu Toyomasu[2], Yuto Hasegawa[2], Youming Liu[3], Hideaki Nojiri[3], Kazunori Okada[3], Claude Becker[1,*]

[1] Faculty of Biology, Ludwig-Maximilians-Universität München, 82152 Martinsried, Germany

[2] Faculty of Agriculture, Yamagata University, Tsuruoka, Yamagata, 997-8555, Japan

[3] Agro-Biotechnology Research Center (AgTECH), Graduate School of Agricultural and Life Sciences (GSALS), The University of Tokyo, Tokyo, 113-8657, Japan

[*] To whom correspondence should be addressed: claude.becker@lmu.de

ORCIDs:

Santiago Priego Cubero: 0009-0006-5209-7942

Kazunori Okada: 0000-0001-7321-7911

Claude Becker: 0000-0003-3406-4670

ABSTRACT

Plants are master chemists and collectively are able to produce hundreds of thousands of different organic compounds. The genes underlying the biosynthesis of many specialized metabolites are organized in biosynthetic gene clusters (BGCs), which is hypothesized to ensure their faithful co-inheritance and to facilitate their coordinated expression. In rice, momilactones are diterpenoids that act in plant defence and various organismic interactions. Many of the genes essential for momilactone biosynthesis are grouped in a biosynthetic gene cluster. Here, we apply comparative genomics of diploid and allotetraploid *Oryza* species to reconstruct the species-specific architecture, evolutionary trajectory, and sub-functionalisation of the momilactone biosynthetic gene cluster (MBGC) in the *Oryza* genus. Our data show that the evolution of the MBGC is marked by lineage-specific rearrangements and gene copy number variation, as well as by pseudogenization and occasional cluster loss. We identified a distinct cluster architecture in *O. coarctata*, which represents the first instance of an alternative architecture of the MBGC in *Oryza* and strengthens the idea of a common origin of the cluster in *Oryza* and the distantly related genus *Echinochloa*. Our research illustrates the evolutionary and functional dynamics of a biosynthetic gene cluster within a plant genus.

INTRODUCTION

Biosynthetic gene clusters and their evolution

With more and more plant reference genome assemblies becoming available, biosynthetic gene clusters (BGCs), i.e., the co-localization of often phylogenetically unrelated genes that participate in the same biosynthetic cascade of specialized metabolites, have emerged as a common feature of genomic organization (reviewed in Polturak et al., 2022b). BGCs are postulated to confer evolutionary advantages because they facilitate coordinated gene expression, enable the reliable coinheritance of genes involved in the same metabolic pathway (thereby preventing the accumulation of toxic intermediates), or facilitate the formation of metabolons (Nützmann et al., 2016). However, the mechanisms by which such non-orthologous genes become localized in the same genomic region and act in the same biosynthetic pathway are still poorly understood. Currently, the most common model proposes that they have formed through a series of events that is driven by both positive and negative selection pressure, starting with gene duplication, followed by neo-functionalization, and ultimately relocation. In some cases, this process appears to have been mediated by transposable elements (Polturak et al., 2022b; Smit and Lichman, 2022).

Biological functions of rice phytoalexins, labdane-related diterpenoids and momilactones

Phytoalexins are low-molecular-mass specialized plant metabolites that are often produced under biotic and abiotic stress conditions (Ahuja et al., 2012). In rice (*Oryza sativa*), the major phytoalexins are a group of labdane-related diterpenoids (reviewed in Toyomasu et al., 2020); they derive from the cyclization of geranylgeranyl diphosphate (GGPP) into *ent*, *syn*, or *normal* stereoisomers of copalyl diphosphate (CDP) by the class II diterpene synthases Copalyl Diphosphate Synthases (CPSs). The biosynthesis of these metabolites has evolved from that of gibberellins (GAs), *ent* labdane-related diterpenoids themselves, through duplication and neofunctionalization of core biosynthetic enzymes (Zi et al., 2014). Several rice labdane-related diterpenoids have been identified, including momilactones A and B, phytocassanes A to F, and oryzalexins (A to F, and S) (Toyomasu et al., 2020). Notably, momilactone A and, more prominently, momilactone B have a strong allelopathic activity, i.e., they inhibit the germination and growth of nearby plants upon being released by the rice plants into the soil (Kato et al., 1973; Kato-Noguchi et al., 2010). Both compounds accumulate in rice husks but are also exuded from the roots (Kato-Noguchi et al., 2010; Kato-Noguchi and Ino, 2003).

Biosynthesis of momilactones and clustering of momilactone genes

Momilactone biosynthesis starts with the cyclization of GGPP into *syn*-copalyl diphosphate (*syn*-CDP), catalysed by the Copalyl Diphosphate Synthase 4 (CPS4) (Otomo et al., 2004b; Xu et al., 2004). *syn*-CDP is further cyclized into *syn*-pimara-7,15-diene by *ent*-kaurene synthase-like 4 (KSL4), a class I diterpene synthase (Otomo et al., 2004a; Wilderman et al., 2004). Because *syn*-CDP is also a substrate for oryzalexin S biosynthesis, the KSL4-mediated cyclization is the first truly dedicated step towards momilactone production (Tamogani et al., 1993). *syn*-pimara-7,15-diene undergoes several oxidation steps, catalysed first by cytochrome P450 (CYP) monooxygenases CYP76M8 and CYP99A3, followed by the short-chain alcohol dehydrogenase MOMILACTONE A SYNTHASE (MAS) and CYP701A8, to yield momilactone A (De La Peña and Sattely, 2021; Kitaoka et al., 2021). CYP76M14 catalyses the final hydroxylation of C20, leading to spontaneous closure of the hemi-acetal ring and forming momilactone B (De La Peña and Sattely, 2021). Among the biosynthetic genes, *CPS4*, *KSL4*, and the paralogs *CYP99A2/3* and *MAS1/2* are co-localized on chromosome 4 in a momilactone biosynthetic gene cluster (MBGC) (Miyamoto et al., 2016; Shimura et al., 2007). However, not all genes required for momilactone biosynthesis are contained in the MBGC: *CYP76M8* is located on chromosome 2 and is part of another BGC that is required for

phytocassane and oryzalexin production (Okada, 2011), while *CYP701A8* and *CYP76M14* are located on chromosomes 6 and 1, respectively (De La Peña and Sattely, 2021).

Evolution of the momilactone biosynthetic gene cluster in Oryza

The *Oryza* genus (Poaceae family) consists of 27 known species with 11 different genome types (classified based on cytogenetics and genetic hybridization studies): 6 diploids (n = 12; AA, BB, CC, EE, FF and GG) and 5 allotetraploids (n = 24; BBCC, CCDD, HHJJ, HHKK and KKLL) (Ge et al., 1999; Lu et al., 2009). The MBGC was established in *Oryza* before the domestication of rice (*O. sativa*; AA) (Miyamoto et al., 2016). It is highly conserved among AA and BB genome type *Oryza* species, while only a partial cluster exists in the early-divergent *Oryza* species *O. brachyantha* (FF), which harbours only two clustered *CYP99A2/3* paralogs (Miyamoto et al., 2016). Because the MBGC is incomplete in *O. brachyantha*, it likely evolved prior to the divergence of the BB lineage in *Oryza* (Miyamoto et al., 2016). Due to the limited availability of genome assemblies from species positioned between *O. brachyantha* and *O. punctata* in the phylogenetic tree, it is unclear whether the MBGC is restricted to species of the AA and BB lineages. Alternatively, the MBGC could have been lost specifically in the *O. brachyantha* lineage and might be present in other *Oryza* lineages. Lastly, the presence and configuration of the MBGC have not yet been studied in any of the allotetraploid *Oryza* species, and it therefore remains unknown if the cluster is present in both their sub-genomes.

Momilactone biosynthetic gene cluster in other species

Studies of the MBGC have not been limited to the *Oryza* genus; some have expanded into the wider Poaceae family. Poaceae encompass twelve subfamilies, of which nine belong to two core clades: BOP and PACMAD. The BOP clade comprises Bambusoideae, Oryzoideae (including *Oryza*), and Pooideae (including the *Triticeae* tribe); the PACMAD clade consists of Panicoideae, Aristidoideae, Chloridoideae, Micrairoideae, Arundinoideae, and Danthonioideae (Soreng et al., 2017). The MBGC is not restricted to the *Oryza* genus and has been identified through phylogenomic methods in species belonging to the PACMAD clade, specifically in the Panicoideae and Chloridoideae subfamilies (Wu et al., 2022a). However, it remains uncertain whether these species actually produce momilactones or other types of labdane-related diterpenoids. *Echinochloa crus-galli* (Panicoideae), a prevalent weed associated with rice cultivation, is the only species confirmed to produce momilactone A (Kraehmer et al., 2016; Wu et al., 2022a). Interestingly, *E. crus-galli* exhibits a distinct MBGC architecture, including an additional cytochrome P450 (*EcCYP71L11*) (Guo et al., 2017; Kitaoka et al., 2021). EcCYP71L11, similarly to CYP76M8 in *Oryza sativa*, can catalyse the

conversion of *syn*-pimaradiene into 6β-hydroxy-*syn*-pimaradiene (Kitaoka et al., 2021). Despite the shared origin of the MBGC in *Oryza* and *Echinochloa*, the *CYP76L11* gene has not been identified in the MBGC or the genome of any *Oryza* species (Wu et al., 2022a).

Here, we studied the presence, architecture, and evolution of the MBGC in cultivated rice and wild relatives from the *Oryza* genus by mining recently published assemblies of reference genomes. We identified MBGC-like clusters in *Oryza* CC species, in the C sub-genome of the allotetraploid *O. alta*, and in the early-divergent *O. coarctata* (KKLL), but not in *O. australiensis* (EE) and *O. granulata* (GG). In the *O. coarctata* MBGC, we identified an additional gene, *OcCYP76L11-like*, encoding a CYP that is different from the canonical *O. sativa* CYP99A2/3 and for which a corresponding ortholog could be found only in the MBGC of *E. crus-galli*. Our study shows how a biosynthetic gene cluster diversified within a plant genus while largely maintaining its biosynthetic function.

RESULTS

The momilactone biosynthetic gene cluster is prevalent throughout the Oryza genus

Among the 11 different genome types within the genus *Oryza*, the MBGC has been extensively studied in only 3 lineages. The cluster exhibits a high level of conservation among *Oryza* species of the AA and BB genome groups, whereas it is incomplete in the FF species *O. brachyantha* (Figure 1a) (Miyamoto et al., 2016). To better understand the evolution of MBGC in the different *Oryza* lineages, we mined newly published genome assemblies for the presence and location of the MBGC (see methods for details). Our analysis included genomes of the diploid species *O. officinalis*, *O. eichingeri*, *O. rhizomatis* (all CC), *O. australiensis* (EE), and *O. granulata* (GG), as well as of the allotetraploid species *O. alta* (CCDD) and *O. coarctata* (KKLL). We also included *Leersia perrieri* as an outgroup to *Oryza* in our analyses. We used OrthoFinder (Emms and Kelly, 2019) to identify orthogroups containing putative *O. sativa* momilactone biosynthetic genes. Based on phylogenetic relationships between the set of genes in each orthogroup, we considered as homologs those genes that belonged to the same clade as the *O. sativa* momilactone biosynthetic genes (Supplemental Figures 1-4). We further mined the genomic regions upstream and downstream of the identified homologs to confirm their organization in a genomic cluster.

We detected the MBGC in all three CC genome species: *O. officinalis*, *O. eichingeri*, and *O. rhizomatis* (CC). Within the CC lineage, gene copy numbers varied, and some genes showed duplication depending on the species. In *O. officinalis* (CC), the cluster (~385 kb in length)

was located on chromosome 4 and contained *CPS4*, *MAS*, two copies of *KSL4*, and two *CYP99A2/3* orthologs (Figure 1a). In *O. eichingeri* and *O. rhizomatis* (both CC), we identified orthologs of *CPS4*, *MAS*, two copies of *KSL4*, and *CYP99A2/3*; *O. eichingeri* carried two copies each for *CYP99A2/3* and *MAS*, with the second copy of the latter appearing pseudogenic (Figure 1a). However, in the scaffold-level assemblies of *O. eichingeri* and *O. rhizomatis*, we found these genes localized on different scaffolds. In *O. rhizomatis*, the *CYP99A2/3* and one of the *KSL4* paralogs co-occurred on Scaffold2280, *CPS4* and the other *KSL4* paralog on Scaffold1491 (Figure 1a). In agreement with the presence of the cluster in all CC species, we identified a ~520 kb MBGC cluster on chromosome 4 of sub-genome C of the allotetraploid *O. alta* (CCDD). This cluster contained orthologs of *MAS*, *KSL4*, *CYP99A2/3*, and three orthologs of *CPS4* (Figure 1a). In contrast, we did not find any orthologs of momilactone biosynthetic genes in sub-genome D of *O. alta*. This observation is in line with *O. australiensis* (EE), the closest relative of the *O. alta* D-type sub-genome (the ancestral diploid D-type donor is presumably extinct) lacking a MBGC (Bao and Ge, 2004; Ge et al., 1999). We identified a MBGC (~55 kb on Scaffold26 and containing one ortholog each of *CPS4*, *MAS*, *KSL4*, and *CYP99A2/3*) in the allotetraploid *O. coarctata* (KKLL). Because this lineage diverged earlier than the MBGC-lacking EE lineage (Figure 1a), we speculate that the cluster most likely became lost in the EE lineage. Lastly, we did not detect the cluster or homologs thereof in *O. granulata* (GG).

To further confirm that we had detected the MBGC and not another diterpene-related BGC, we analysed the synteny of the corresponding regions among the different species. We hypothesized that if the MBGC in the *Oryza* genus had a common origin, one should observe high synteny of the genomic region among all the species. We excluded *O. rhizomatis* and *O. eichingeri* from the synteny analysis, because in those genome assemblies the momilactone biosynthesis genes were located on short scaffolds that often carried only very few (~3-4) genes in total. We also excluded *O. australiensis* and *O. granulata*, since their fragmented scaffold-level assemblies hindered the detection of syntenic blocks. For the remaining seven species, the genomic region containing the MBGC was highly syntenic (Figure 1b), which further confirmed the presence of the MBGC in the different reference genomes and indicated that this region shares a common evolutionary origin. Taken together, these results suggest that the MBGC is prevalent at a basal position of the *Oryza* phylogeny. The evolution of the cluster in different lineages is characterized by differences in cluster size, lineage-specific rearrangements with an important variation in gene copy number, and lineage-specific loss of the cluster.

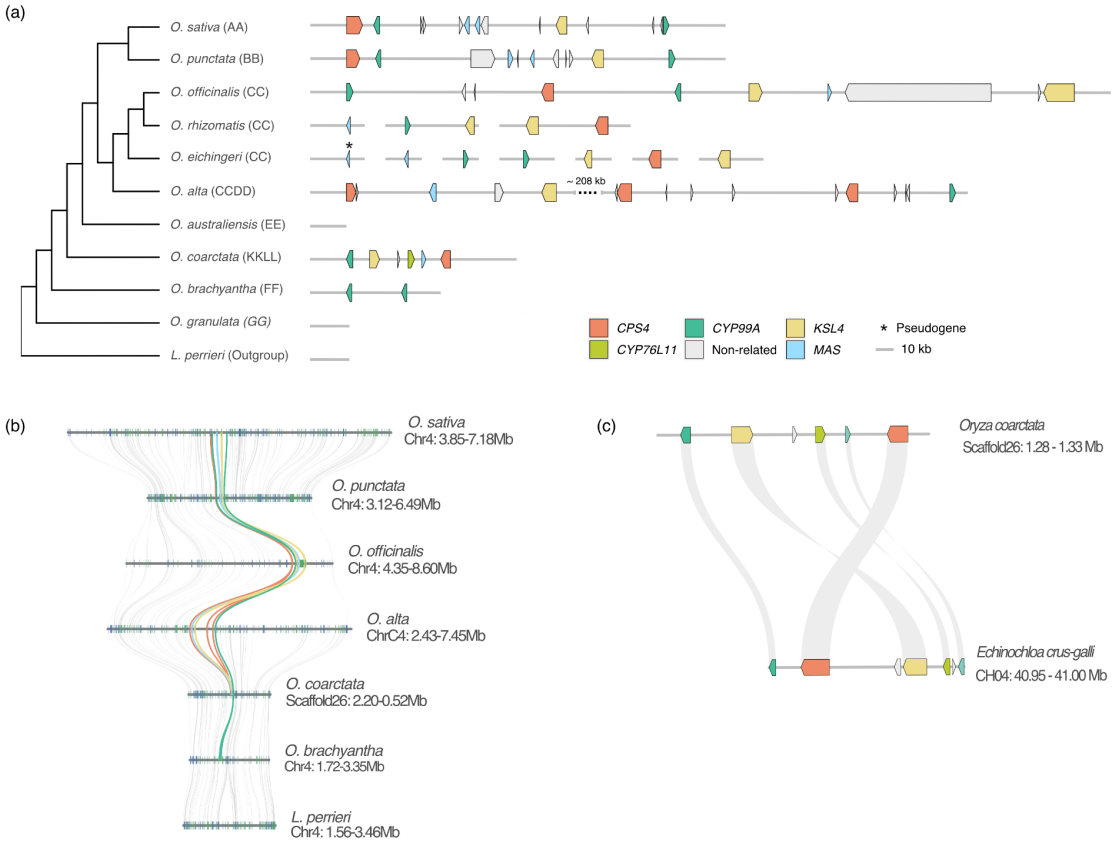


Figure 1. The momilactone biosynthetic gene cluster in the Oryza genus. a) Phylogenetic relationship between the different *Oryza* species included in this study (Ge et al., 1999; Shenton et al., 2020), and their respective MBGCs. Coloured block arrows represent genes. **b).** Microsynteny between the genomic regions containing the MBGC from *O. sativa*, *O. punctata*, *O. officinalis*, *O. alta*, *O. coarctata*, *O. brachyantha*, and *L. perrieri*. **c)** Microsynteny between MBGC in *O. coarctata* and *E. crus-galli*. In **b)** and **c)**, grey lines connect the respective orthologs. "Non-related" refers to genes that do not seem to be functionally related to the main biosynthetic cascade of the cluster.

Species-specific deviations in the MBGC architecture in the Oryza genus

We next investigated whether there were any noticeable species-specific deviations in the MBGC architecture. In *O. coarctata*, we identified an extra gene located between *MAS* and *KSL4* and encoding a CYP. Importantly, this enzyme does not belong to the CYP99A subfamily, which is the canonical CYP in the MBGC of the other *Oryza* species (Figure 1a). Instead, the MBGC in *O. coarctata* strikingly resembled that of *E. crus-galli* (Figure 1c), which also contains an additional CYP-encoding gene, *EcCYP76L11* (Guo et al., 2017; Kitaoka et al., 2021) that does not belong to the CYP99A subfamily either. EcCYP76L11 catalyses the same reaction as OsCYP76M8, converting *syn*-pimaradiene to 6β-hydroxy-*syn*-pimaradiene (Kitaoka et al., 2021). The corresponding homolog in rice, *OsCYP76M8*, is located outside of

the MBGC, which suggests that *E. crus-galli* possibly contains a more compacted version of the cluster (Kitaoka et al., 2021).

We further investigated the MBGC in *O. coarctata*. An initial microsynteny comparison of the MBGCs in *O. coarctata* and *E. crus-galli* revealed that the additional CYP-encoding gene could indeed be an ortholog of *EcCYP76L11* (Figure 1c). To verify their relationship, we conducted an amino-acid-sequence phylogenetic analysis including all *Oryza* species from our panel and different representative species of the main Poaceae subfamilies: *Olyra latifolia* (Bambusoideae), *Brachypodium distachyon* and *Hordeum vulgare* (Pooideae), *Setaria italica*, *Panicum hallii* and *E. crus-galli* (Panicoideae), and *Pharus latifolius* (outgroup to the core Poaceae). Our analysis placed the *O. coarctata* CYP in a clade with EcCYP76L11 and two CYPs from *P. hallii* (Pahal.1G039100) and *S. italica* (Seita.2G145700), respectively. Interestingly, the *O. coarctata* CYP did not group in a clade with the unique CYP76L member from *O. sativa* (CYP76L1) and the CYP76L1 orthologs in other *Oryza* species (Figure 2). Consequently, based on the similarity to EcCYP76L11, we named the newly identified *O. coarctata* CYP-encoding gene *OcCYP76L11*.

We observed a phylogenetic incongruence with regards to OcCYP76L11. Within its clade, this enzyme showed a closer similarity to homologs from the Panicoideae subfamily rather than to genes in the Pooideae subfamily (Figure 2). Based on the species tree, one would expect, however, that OcCYP76L11 would be more similar to its homologs in the Pooideae subfamily. Our analysis suggests that *OcCYP76L11* was present in the ancestral *Oryza* MBGC and was retained in the early-divergent *O. coarctata*, while it became lost in the later-diverging clades of the *Oryza* genus.

Momilactone production in different Oryza lineages

Among the investigated *Oryza* genome types (CC, CCDD, EE, KKLL and GG), only species and sub-genomes belonging to the lineages CC and KKLL include an orthologous MBGC. Notably, *O. coarctata* (KKLL), the earliest-divergent *Oryza* species with a MBGC, showed a distinct cluster architecture. To assess the functionality of the predicted clusters in the CC and KKLL lineages, we proceeded to evaluate the momilactone-producing potential of their respective representative species *O. officinalis* (CC) and *O. coarctata* (KKLL).

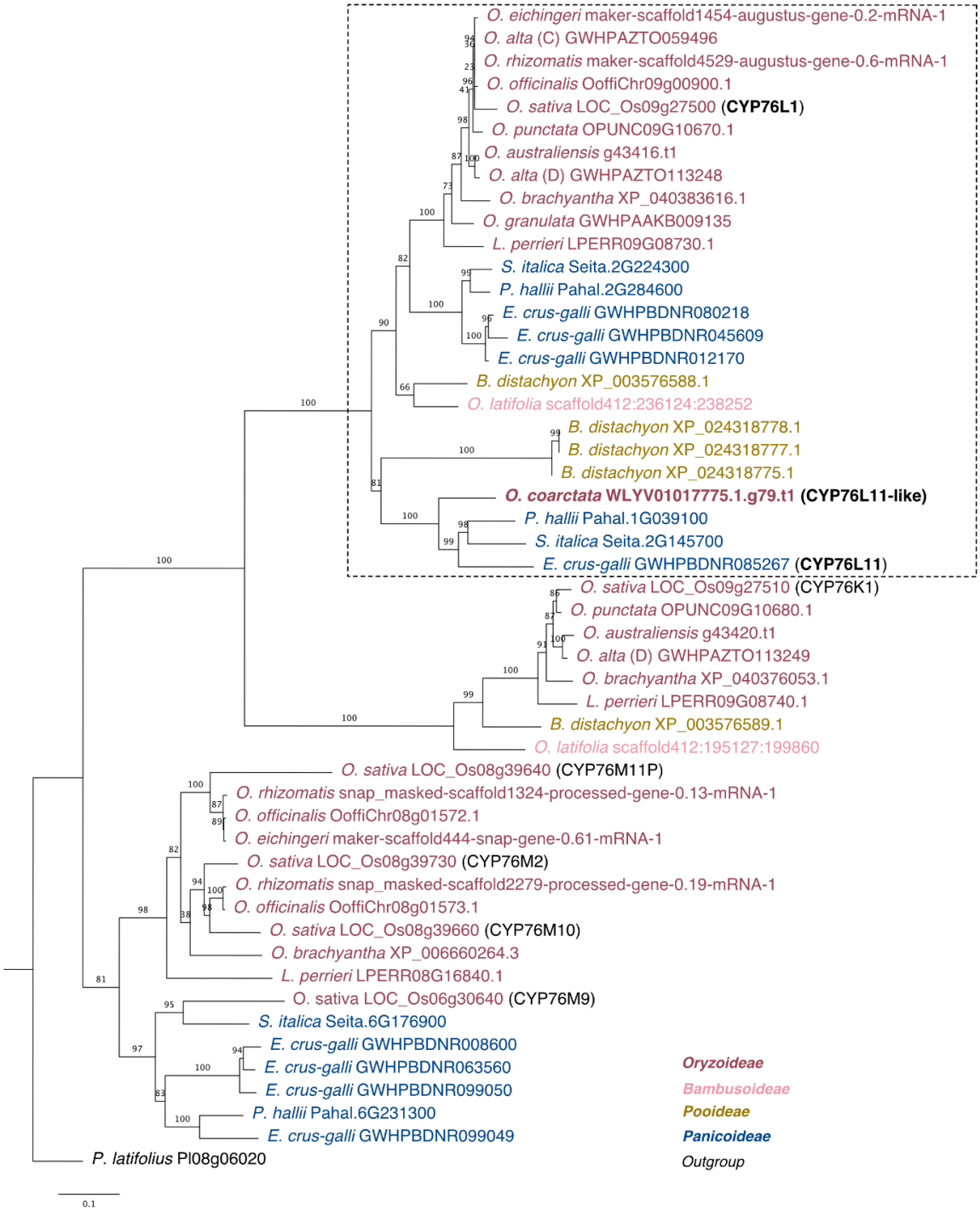


Figure 2. Phylogenetic analysis of a newly identified CYP ortholog in the MBGC. Maximum likelihood tree with the best-fit model selected by ModelFinder representing the relations between *O. coarctata* CYP76L11 and BLASTp best matches in *O. sativa*, *O. punctata*, *O. officinalis*, *O. eichingeri*, *O. rhizomatis*, *O. alta*, *O. australiensis*, *O. brachyantha*, *O. granulata*, *L. perrieri* (*Oryzoideae*), *Olyra latifolia* (*Bambusoideae*), *Brachypodium distachyon* and *Hordeum vulgare* (*Pooideae*), *Setaria italica*, *Panicum hallii* and *E. crus-galli* (*Panicoideae*) and *Pharus latifolius*. Numbers represent 1000 replicates of ultrafast bootstrapping. For *H. vulgare*, we did not identify an ortholog with >50% sequence identity, which we applied as a minimum filter in our analysis.

We conducted liquid chromatography-tandem mass spectrometry (LC-MS/MS) to detect momilactones A and B in leaf blades. Momilactone basal levels can be close to detection limits, which is why we induced momilactone production by treating the plants with $CuCl_2$ 3 days before extraction, a treatment known to increase momilactone levels (Miyamoto et al., 2016). Consistent with the presence of the MBGC, we detected both momilactone A and B in *O. officinalis* (Figure 3a). However, in samples from *O. coarctata*, we couldn't detect either momilactone A or B, even though we had detected an orthologous MBGC in this species. Furthermore, in *O. coarctata* leaves induced with $CuCl_2$, we couldn't detect the precursor of momilactones, 9βH-pimara-7,15-diene (Figure 3a,b). Instead, we observed an accumulation of *ent*-cassa-12,15-diene, the precursor of phytocassanes (Figure 3a,b). Taken together, these results provide evidence for the functionality of the identified MBGC in the CC species *O. officinalis*, while suggesting that the distinct orthologous MBGC in the KKLL species *O. coarctata* is likely responsible for the production of other labdane-related diterpenoids.

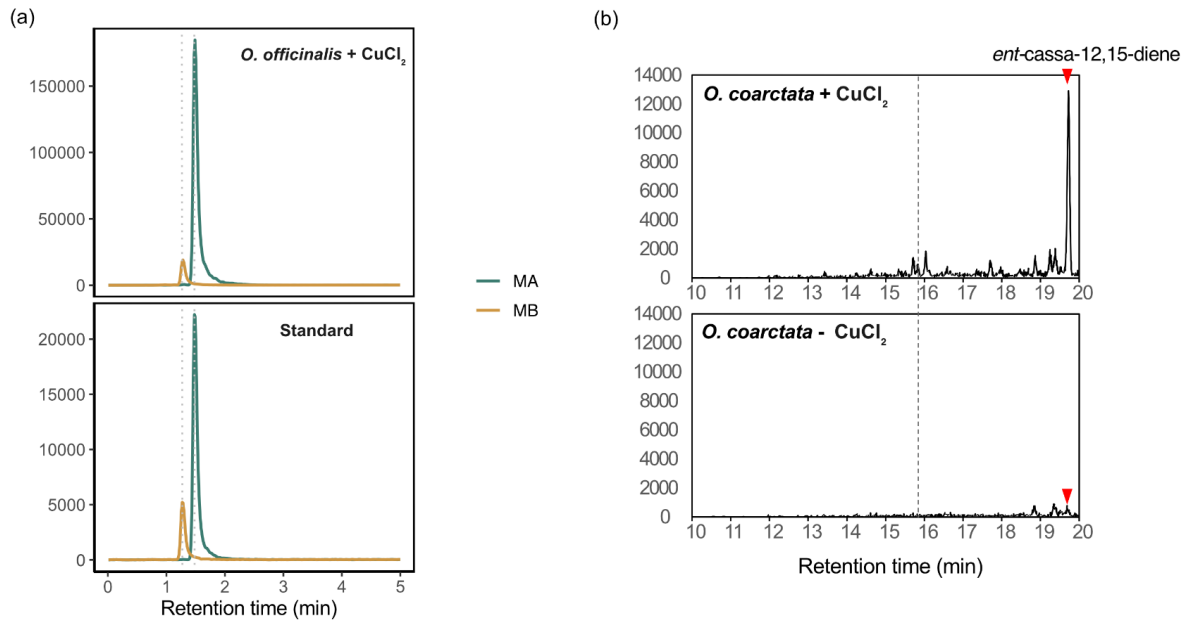


Figure 3. **Analysis of momilactones A and B in representative species of the *Oryza* genus.** Extracts from *O. officinalis* **(a)** and *O. coarctata* **(b)** $CuCl_2$-treated leaf blades were analysed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Five µl of the extract were subjected to LC-MS/MS under the conditions described in the Materials and Methods section. Momilactones were detected with combinations of m/z 315/271 for momilactone A and m/z 331/269 for momilactone B in the multiple reaction monitoring mode. MA, momilactone A; MB, momilactone B. The dashed line in panel b) indicates retention time of 9βH-pimara-7,15-diene.

Functional analysis of KSL4 in O. officinalis and O. coarctata

After establishing the existence of MBGC-like clusters in wild relatives of cultivated rice from the *Oryza* genus and confirming the presence of momilactone A and B in *O. officinalis* but not in *O. coarctata*, we next wanted to confirm the enzymatic functionality of the orthologs from those clusters. We focused on KSL4, which is known to catalyse the first dedicated step towards momilactone biosynthesis (*i.e.*, the conversion of *syn*-CDP to *syn*-pimaradiene). We tested the enzymatic activity of the KSL4 orthologs in representative species of the late-diverged CC and early-diverged KKLL lineages: OoKSL4-1 and OoKSL4-2 from *O. officinalis*, and OcKSL4 from *O. coarctata*.

The available annotation of the genes in the MBGC in *O. officinalis* showed inconsistencies, such as an extra sequence of amino acids in the C-terminal part of OoKSL4-1 and a missing exon in both OoKSL4-1 and OoKSL4-2, compared to the orthologs in *O. sativa* (*OsKSL4*) (Supplementary Figure 5). These discrepancies may be attributed to limitations in the genome annotation process, which relied on RNA-seq data from tissues in which the constitutive expression of momilactone biosynthetic genes is generally low (Shenton et al., 2020). In the case of *O. coarctata*, for which no gene annotation was available, we relied on *ab initio* prediction of open reading frames, which is less accurate and carries the risk of incorrectly annotating some exons.

Exposure to UV light is known to induce the accumulation of momilactones in rice leaves (Kodama et al., 1988). To accurately investigate the function of KSL4 in *O. officinalis* and *O. coarctata*, we used previously published mRNA-Seq data from UV-irradiated leaf sheaths of both species to assemble a de novo transcriptome (Itoh et al., 2021). Subsequently, we conducted a homology search for OsKSL4 homologs in *O. officinalis* and *O. coarctata*. Using the assembled contig sequences as a reference, we designed specific primers and successfully cloned the cDNAs of these three *KSL4* homologs (Supplementary Figure 5). Surprisingly, all cDNA clones of *OoKSL4-2*, obtained from independent plants of the same accession as the reference genome (W0002), exhibited a single nucleotide polymorphism (SNP) at position 7,882,960 that resulted in a G-to-A substitution, introducing a premature stop codon at amino acid position 349 (Supplementary Figure 5). Based on this observation, we hypothesized that *OoKSL4-1* likely encodes the functional KSL4 in the *O. officinalis* MBGC, while *OoKSL4-2* has undergone pseudogenization, resulting in the loss of its enzymatic function.

We proceeded to confirm the enzymatic activity of OoKSL4-1 and OcKSL4. We performed a syn-CDP conversion assay and analysed the products using gas chromatography coupled to mass spectrometry (GC-MS). The retention time of the product generated by OcKSL4 was identical to that of 9βH-pimara-7,15-diene, indicating that OcKSL4 has the same enzymatic activity as OsKSL4 (Figure 4c). The fact that we OcKSL4 is functional *in vitro* and produces 9βH-pimara-7,15-diene, differs from our *in planta* results in which we could not detect 9βH-pimara-7,15-diene.

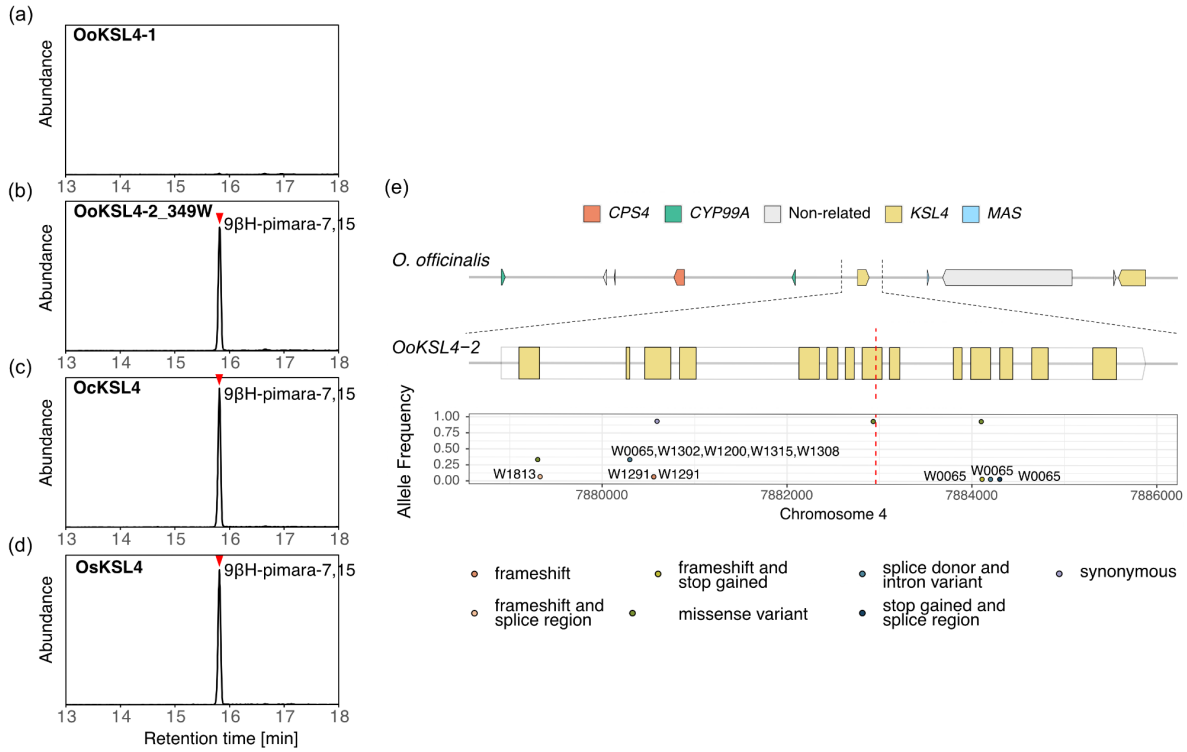


Figure 4. Functional characterization of KSL4 orthologs. a) GC-MS chromatograms of the products obtained from recombinant OoKSL4-1, **b)** OoKSL4-2 with the reverted Trp at position 349, **c)** OcKSL4, and **d)** OsKSL4 in a metabolic engineering system. Reversed red filled triangles indicate production of 9βH-pimara-7,15-diene from *syn*-copalyl diphosphate. **e)** The biosynthetic gene cluster (MBGC) from *O. officinalis*. Middle panel shows a magnification of the *OoKSL4-2* locus where the yellow filled rectangles represent exons of *OoKSL4-2*. Bottom panel shows the allele frequency of single nucleotide polymorphisms (SNPs) with a potential disruptive effect on *OoKSL4-2* from 15 *O. officinalis* accessions. The SNPs represented have allele frequency > 0.2 or a putative impact annotated as 'high'. The red dashed line indicates the position of the G-to-A transition in the original cDNA clones. The labels in the allele frequency panel indicate the name of the accessions with an SNP with a putative impact annotated as 'high'.

Surprisingly, we did not observe any enzymatic activity of OoKSL4-1 towards *syn*-CDP (Figure 4a). This result was intriguing, considering that we had detected momilactone in *O. officinalis*.

We wondered whether the nonsense G-to-A polymorphism in *OoKSL-2* might be a *de novo* mutation in the particular line that we used for cDNA cloning, particularly since it did not match with the corresponding genome reference sequence. To determine the prevalence of the SNP at the population level, we performed SNP calling in fifteen additional *O. officinalis* accessions, including W0002 (Kajiya-Kanegae et al., 2021). We did not observe the SNP in any of the fifteen accessions (Figure 4e). Notably, we did not detect any other deleterious mutations in W0002 either. Taken together, these results suggest that the observed mutation in our plants was likely a *de novo* mutation. We therefore decided to revert the G-to-A polymorphism in our cDNA clone to the consensus reference sequence of *OoKSL4-2* and to examine whether the resulting protein would regain the enzymatic activity. Indeed, when the mutation was reverted to produce the predicted Trp, OoKSL4-2 was able to produce *syn*-pimaradiene (Figure 4b). We therefore conclude that OoKSL4-2 is the enzyme responsible for catalysing the conversion of *syn*-CDP to *syn*-pimaradiene in *O. officinalis*.

DISCUSSION

The momilactone BGC is ancestral and has evolved dynamically in the Oryza genus

BGCs arise from the duplication of enzymes involved in primary metabolism, followed by neofunctionalization and/or relocation processes that are influenced by positive and negative selection pressures (Polturak et al., 2022b; Smit and Lichman, 2022). Ecological factors shape the direction and strength of the selection process, determining the evolution of the metabolism and ultimately affecting the architecture of BGCs. Here, we studied the composition and evolution of the momilactone biosynthetic gene cluster in previously unexplored lineages in the genus *Oryza*. We provide evidence that the MBGC is prevalent in the CC and KKLL lineages, while it is not present in other lineages (EE and GG) (Figure 1a). Our data show that the evolution of the MBGC is marked by lineage-specific rearrangements, such as gene copy number variation or the increase of the size of the cluster, on-going pseudogenization in certain paralogs, and occasional cluster loss. Moreover, we found a distinct MBGC architecture in *O. coarctata*, with the cluster in that genome harbouring a close ortholog of *EcCYP71L11*. This finding represents the first instance of an alternative architecture of the MBGC in *Oryza* and strengthens the idea of a common origin of the MBGC in *Echinochloa* and *Oryza*. Overall, our research sheds light on the evolutionary dynamics and distribution of the MBGC in the genus *Oryza*.

In our study, we found the MBGC at a basal position of the *Oryza* phylogeny in the species *O. coarctata* (KKLL). *O. coarctata* is a polyploid species resulting from the hybridization of the KK and LL diploid genomes, which are presumably extinct or remain to be discovered (Ge et al., 1999; Lu et al., 2009). The KK sub-genome occupies a basal position relative to the EE and DD lineages. On the other hand, the *O. coarctata* LL sub-genome (initially designated as HH), occupies a more basal position compared to KK and it is closer to the diploid FF genome species. Because we found the cluster in *O. coarctata*, a species that branched off before the divergence of the EE lineage, our results indicate that *O. australiensis* (EE) likely lost the MBGC. There are only a limited number of documented cases of interspecific losses of biosynthetic gene clusters (Smit and Lichman, 2022). Such a lineage-specific loss of a BGC was previously described in *Oryza* for the phytocassane biosynthetic gene cluster: *Oryza* AA species and the out-groups *L. perrieri* and *Zizania palustris* species possess the phytocassane cluster, whereas it was partially lost in *O. punctata* (BB) and *O. brachyantha* (FF) (Miyamoto et al., 2016; Yan et al., 2022). It is worth noting that the MBGC is found also in the *Panicoideae* and *Chloridoideae* subfamilies of *Poaceae* and evolved from a common ancestor. However, the distribution of the clusters is restricted to certain tribes, while it is either fragmented or completely lost in others (Wu et al., 2022a). In line with these observations, our results highlight the dynamic nature of biosynthetic gene clusters and suggest that BGC loss might be prevalent within *Poaceae*.

The MBGC and MBGC-like are widespread in the *Poaceae* family (Wu et al., 2022a). Species belonging to the *Triticeae* tribe, phylogenetically more closely related to *Oryza* than to *Echinochloa*, possess a functionally distinct diterpene cluster that shares similarities with the MBGC (Liu et al., 2021; Polturak et al., 2022a; Wu et al., 2022a). This MBGC-like is located on the homeologous chromosomes 2A and 2D in wheat (*Triticum aestivum*), homologs of rice chromosome 4, and exhibits synteny with the region containing the MBGC in rice (Polturak et al., 2022a). However, the cluster in *Triticeae* lacks certain important momilactone biosynthesis genes, including a *MAS* and a *syn-CPS* to produce *syn* labdane-related diterpenoids. Instead, the *Triticeae* cluster produces normal labdane-related diterpenoids in wheat (*Triticum aestivum*) and barley (*Hordeum vulgare*) (Liu et al., 2021; Polturak et al., 2022a). Owing to the shared synteny between the MBGC in rice and the MBGC-like in wheat, Polturak et al. suggested that the *Triticeae* diterpene cluster and the MBGC share a common evolutionary origin and hypothesized that both clusters could have originated from a common ancestral BGC before the divergence of the PACMAD and BOP clades (Polturak et al., 2022a). Wu and colleagues, who performed a more extensive survey of the MBGC in different Poaceae

genomes, alternatively attributed the evolution of the cluster in Poaceae to multiple instances of lateral gene transfer (LGT) (Wu et al., 2022a) and proposed that the MBGCs found in the PACMAD clade originated from the MBGC-like cluster in *Triticeae* through LGT. After being acquired by PACMAD species, the cluster would then have recruited *MAS*. Subsequently, another LGT event would have had to occur, resulting in the acquisition of the cluster by *Oryzoideae*, followed by the loss of the *CYP76L11* gene (Wu et al., 2022a). There are several phylogenomic studies that suggest that LGT is prevalent among grasses; however, the underlying mechanisms by which natural plant-plant LGT would occur are still largely unknown (Dunning et al., 2019; Hibdige et al., 2021). Our findings provide evidence that the *CYP76L11* gene has not been completely lost in the genus *Oryza*. The discovery of *CYP76L11* in *O. coarctata* implies an even tighter notion of a common origin of the MBGC between *Echinochloa* and *Oryza*. These results would align with the lateral gene transfer (LGT) hypothesis, as the phylogenetic analysis shows incongruences with the species tree and places *OcoCYP76L11* closer to PACMAD than to BOP genes. However, the findings from both Polturak et al. and our study (Supplementary Figure 6) indicate high synteny between regions containing the MBGC-like cluster in *Triticeae* and the MBGC in *Oryza*, and thus suggest a common evolutionary origin of this genomic region, including the MBGC. This suggests that there may have been a common ancestor of the MBGC that existed before *Oryza* and *Triticeae* diverged. However, it is difficult to conciliate the synteny results shown by us and Polturak *et al*. with the phylogenomic signatures of LGT shown by Wu et al. and our CYP76L11 analysis. To shed light to this question, it is important to consider the impact and prevalence of gene and MBGC losses in *Poaceae*. The presence or absence of MBGCs in different species can have significant implications for phylogenomic signals and evolutionary dynamics. The loss of MBGCs in certain lineages could potentially shape the phylogenomic patterns observed in Poaceae. Further investigations into the frequency and significance of MBGC losses across Poaceae will contribute to a better understanding of the evolutionary history of this cluster.

Allotetraploid rice species carry only one MBGC copy in their genome

In our study, we aimed to investigate whether allotetraploid species retain two copies of the MBGC. Previous research has shown that whole genome duplication (WGD) can lead to the silencing and progressive loss of one copy of a biosynthetic gene cluster, likely to balance the effects of increased gene dosage (Polturak et al., 2022a; Yang et al., 2021). We found that each of the two allotetraploid species we studied had retained one copy of the MBGC. In *Oryza alta* (CCDD), the cluster was present in the CC but not the DD sub-genome, which means that

the MBGC was either lost in the DD sub-genome after hybridization or that the DD ancestor had already lost it prior to the hybridisation. In the case of KKLL, we identified only one copy of the MBGC but - due to the fragmented genome assembly - we were unable to determine the specific sub-genome that harbours the cluster. To better understand the cluster evolution in allotetraploid rice, it would be valuable to generate haplotype-resolved assemblies of BBCC hybrid species, as both their ancestors harbour the MBGC.

O. coarctata does not produce momilactones A or B

At the interspecies level, the existence of conserved BGCs does not always align consistently with the production of identical metabolites, while the metabolites derived from orthologous BGCs can also be species-specific. The resulting divergence in chemotypes between species is associated with variations in biosynthetic genes external to the BGC, genes within the BGC, or transcription factors controlling the biosynthesis of these compounds (Liu et al., 2020; Zhou et al., 2016). In some cases, the function of the BGC is retained, and orthologous BGCs produce the same molecule. This is exemplified by *Oryza* species from the AA lineage and BB lineage, which produce momilactones A and B (Miyamoto et al., 2016).

In our study, we found that the CC lineage, represented by *O. officinalis*, also possesses an MBGC with slight differences in gene copy number, and that momilactones can be detected in this lineage. In contrast, the *O. coarctata* genome contains an orthologous and syntenic MBGC with a distinct architecture compared to *O. sativa*. However, and despite our phylogenomic analysis suggesting the potential for momilactone production in *O. coarctata*, we were unable to detect momilactones A or B in leaf blades induced with $CuCl_2$ or UV light - commonly used momilactone inducers (Kodama et al., 1988; Miyamoto et al., 2016). However, we did confirm the functionality of the first dedicated step in momilactone biosynthesis, catalysed by KSL4, in *O. coarctata*: OcKSL4 catalysed the conversion from syn-CDP to 9βH-pimara-7,15-diene. The apparent contradictions in our findings could be attributed to several factors. First, it is plausible that one or more genes within the cluster (*OcCPS4*, *OcCYP99A*, or *OcMAS*) are non-functional, which would halt the biosynthetic cascade. Alternatively, auxiliary genes outside the cluster, such as *CYP76M14*, which play a role in momilactone biosynthesis in *O. sativa*, may be non-functional or absent in *O. coarctata*. Notably, we could find orthologs to all known momilactone biosynthetic genes in *O. coarctata* by nucleotide BLAST analysis (data not shown), suggesting rather a loss than an absence of functionality. Moreover, the MBGC in *O. coarctata* could be under a distinct regulatory network compared to *O. sativa*, potentially being tightly repressed in leaves and/or not inducible by $CuCl_2$. This

scenario could impede the detection of momilactones. Such a hypothesis is in line with the possibility of the absence of key transcription factors crucial for regulating momilactone production under $CuCl_2$ induction. In future studies, evaluating the enzymatic activity of genes within the *O. coarctata* orthologous MBGC will provide valuable insights into whether these genes have undergone sub-functionalization or, conversely, play a role in momilactone biosynthesis. Additionally, exploring alternative inducers, such as jasmonic acid, and investigating potential momilactone biosynthesis in different tissues can help shed light onto the functionality of the *O. coarctata* MBGC.

Pseudogenization after duplication

One of the more widely accepted hypotheses to account for the formation and function of BGC is co-inheritance, where physical colocalization enables faithful transmission across generations, thus mitigating the risk of losing biosynthetic genes through recombination (Nützmann et al., 2016; Takos and Rook, 2012). This implies that the proximity of the genes reduces the likelihood of losing crucial genes through recombination. Our findings reveal that the MBGC in *O. alta* spans approximately 520 kb (and hence 340 kb more than the canonical MBGC in *O. sativa*), possibly indicating an evolutionary trend towards cluster loss. This notion aligns with the observation that the cluster in *O. officinalis* (~385 kb) features a more 'isolated' *OoKSL4-1* with a premature stop codon. *OoKSL4-1* has a softer linkage with the rest of the cluster, which - coupled with the presence of a paralog - might have allowed the fixation of the stop codon. Furthermore, we identified potentially disruptive mutations at a low allele frequency in *OoKSL4-2* in 6 out of 15 accessions. Among those 6 accessions, W0065 was the only one with premature stop codon mutations (Figure 4e). In W1291, we found two frameshift variants caused by two independent insertions that ultimately truncate OoKSL4-2. Lastly, we also found a potential frameshift variant caused by a splicing variant in W0065, W1302, W1200, W1315, and W1308 (Figure 4e). This might indicate that *OoKSL4-2* might be undergoing a pseudogenization process with the subsequent loss of the BGC function on momilactone production. A more comprehensive exploration of CC and CCDD lineage species' capacity to produce momilactones, along with investigations at the population level, could provide valuable insights into the evolutionary trajectory of the cluster within these lineages.

MATERIALS AND METHODS

Oryza and Poaceae datasets

O. sativa (v7.0) (Ouyang et al., 2007) protein sequences and gene annotations were downloaded from Phytozome (https://phytozome-next.jgi.doe.gov/); *O. punctata* (Oryza_punctata_v1.2) (Stein et al., 2018) protein sequences and gene annotations from EnsemblPlants; the *O. officinalis* (Shenton et al., 2020) genome sequence from NCBI (assembly GCA_008326285.1) and annotations from the original article repository; *O. eichingeri* and *O. rhizomatis* (Shenton et al., 2020) protein sequences and gene annotations from Cyverse Data Commons (https://doi.org/10.25739/awh3-dm39); *O. alta* (Yu et al., 2021) protein sequences and gene annotations from the Genome Sequence Archive in BIG Data Center (http://bigd.big.ac.cn) accession number GWHAZTO00000000; the *O. australiensis* (Phillips et al., 2022) genome sequence from NCBI (GCA_019925245.2) and annotations from figShare (https://figshare.com/collections/Oryza_australiensis_Keep_River_Genome_Assembly_V2/5875592/2); the *O. coarctata* (Mondal et al., 2018) genome assembly from NCBI (GCA_009761635.1); the *O. brachyantha* (Chen et al., 2013) (ObraRS2) genome assembly and gene annotations from NCBI (GCF_000231095.2); *O. granulata* (Shi et al., 2020) protein sequences and gene annotations from Genome Sequence Archive in BIG Data Center (GWHAAKB00000000); *L. perrieri* (Stein et al., 2018) protein sequences and gene annotations (V1.4) from Gramene (https://www.gramene.org/). *O. officinalis*, *O. australiensis* and *O. brachyantha* protein sequences were extracted from the gene annotation using gffread (v0.12.7) (Pertea and Pertea, 2020).

Brachypodium distachyon (V3.1) (Vogel et al., 2010) sequences were retrieved from Phytozome; *Echinochloa crus-galli* (V3.0) (Wu et al., 2022b) sequences from Genome Sequence Archive in BIG Data Center, accession number GWHBDNR00000000; *Hordeum vulgare* cv Morex (V3) (Mascher et al., 2021) sequences from Phytozome; *Olyra latifolia* (Guo et al., 2019) sequences from http://www.genobank.org/bamboo; *Panicum hallii* var. *filipes* (V3.2) (Lovell et al., 2018) sequences from Phytozome; *Pharus latifolius* (Ma et al., 2021) sequences from http://www.genobank.org/grass; and *Setaria italica* Yugu1 (Bennetzen et al., 2012) (V2.2) sequences from Phytozome.

Ortholog detection

OrthoFinder (v2.5.4) (Emms and Kelly, 2019, 2015) was used to infer orthogroups containing *O. sativa* momilactone biosynthesis genes (i.e., genes from multiple species descended from a single gene in the last common ancestor of these species). The input for OrthoFinder included primary transcript proteomes and a rooted species tree (Figure 1A) of the following species: *O. sativa*, *O. punctata*, *O. officinalis*, *O. eichingeri*, *O. rhizomatis*, *O. alta*, *O. australiensis*, *O. brachyantha*, *O. granulata*, and *L. perrieri* (included as the out-group).

Due to the unavailability of a suitable annotation for *O. coarctata*, putative momilactone genes in that species were inferred by homology search. The genomic locus for each momilactone gene from *O. sativa* was used as a query in the Basic Local Alignment Search Tool (BLAST) against the *O. coarctata* assembly. Subsequently, *ab initio* open-reading-frame prediction with AUGUSTUS (version 3.4.0) (version 3.4.0) (Stanke et al., 2008) was performed in the scaffold containing putative momilactones genes (scaffold26; WLYV01017775.1). Only predicted proteins longer than 100 amino acids were considered.

The predicted homologs of momilactone genes in *O. coarctata* were added to the corresponding identified orthogroups. Each orthogroup was aligned using MAFFT (v7.490) (Katoh et al., 2002; Katoh and Standley, 2013) and the resulting alignment trimmed using trimAl (v1.4.rev15) (Capella-Gutiérrez et al., 2009) with a heuristic selection of the automatic trimming method. IQTREE (v2.1.4-beta) (Minh et al., 2020) was used to infer the maximum likelihood (ML) tree with the best-fit model selected by ModelFinder (Kalyaanamoorthy et al., 2017) with 1000 replicates of ultrafast bootstrapping (Hoang et al., 2018).

ML rooted trees of each orthogroup were used to identify orthologs of *O. sativa* momilactone biosynthetic genes in *Oryza*. Genes within the same clade as *O. sativa* momilactone biosynthesis genes were considered as potentially clustered homologs. The genomic location upstream and downstream of these homologs was further analyzed to confirm clustering. Trees were represented with FigTree (v1.4.4) (https://github.com/rambaut/figtree).

Oryza coarctata CYP76L11 orthologs

Since a suitable annotation was not available for *O. coarctata*, it was not possible to use OrthoFinder to find *OcoCYP76L11* orthologs. Instead, protein-protein BLAST (v2.12.0+) search was performed using the predicted *OcCYP76L11* protein sequence against various *Oryza* species (including *O. sativa*, *O. punctata*, *O. officinalis*, *O. eichingeri*, *O. rhizomatis*, *O. alta*, *O. australiensis*, *O. brachyantha*, *O. granulata*), as well as other *Poaceae* species such

as *L. perrieri*, *Olyra latifolia*, *Brachypodium distachyon, Hordeum vulgare*, *Setaria italica*, *Panicum hallii*, *Echinochloa crus-galli*, and *Pharus latifolius* (serving as the outgroup to the core *Poaceae*). Only proteins with an identity higher than 50% and e-value < 1e-50 were considered. Protein sequences were processed as described above for the orthogroups. The names of the *O. sativa* P450 genes in the resultant tree were obtained from https://funricegenes.github.io/cytochrome_P450_monooxygenase_superfamily/.

Microsynteny Analysis

Microsynteny analysis and figures were done with MCScan implemented in Python (v3.9) with JCVI utility libraries (v1.1.11) following the package workflow: https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version). Scripts are detailed in https://github.com/spriego/Priego-Cubero_et_al_MBGC-Oryza. Briefly, protein FASTA files and GFF3 files of each species were used as input. When needed, the FASTA files were processed using Biopython's SeqIO package (Cock et al., 2009). Pairwise ortholog and synteny blocks search was performed with default settings using a lift-over distance of 5, a c-score = 0.70 and a maximum of 1 iterations. Multi-synteny blocks were constructed combining the syntenic blocks detected on every species with respect to *Oryza sativa*.

SNP calling

Short-read Illumina sequencing data for 15 *O. officinalis* accession were downloaded from OryzaGenome 2.1 (http://viewer.shigen.info/oryzagenome21detail/index.xhtml) (Kajiya-Kanegae et al., 2021). SNP calling was performed using the nf-core pipeline sarek v3.0.1 with default parameters in Nextflow 22.10.4 (Di Tommaso et al., 2017; Ewels et al., 2020; Garcia et al., 2020). Scripts are detailed in https://github.com/spriego/Priego-Cubero_et_al_MBGC-Oryza. Briefly, the fastq files underwent sequence quality assessment and pre-processing using fastp v0.23.2 (Chen et al., 2018). The reads were mapped to the *O. officinalis* reference genome using BWA mem v0.7.17-r1188 (Li, 2013). Duplicates were marked with GATK MarkDuplicates version 4.2.6.1 (McKenna et al., 2010). Variant calling was performed with HaplotypeCaller (Poplin et al., 2018). The BaseRecalibrator step was skipped as it requires a database of known polymorphic sites, which is unavailable for *O. officinalis.* The output vcf files were merged with bcftools v1.13 (Danecek et al., 2021) and SNPs effect were annotated with snpEff v5.0.1 (Cingolani et al., 2012).

Data analysis

RStudio with R version 4.3.1 was used for generating Figures 1a, 3a, and 4a. Scripts used for data analysis are available at https://github.com/spriego/Priego-Cubero_et_al_MBGC-Oryza. To generate the figures, the following R packages were used: Tidyverse (Wickham et al., 2019), ggtree (Yu et al., 2017), gggenes (https://wilkox.org/gggenes/), khroma (Frerebeau, 2023), ggrepel https://github.com/slowkow/ggrepel, readxl https://github.com/tidyverse/readxl, MetBrewer (https://github.com/BlakeRMills/MetBrewer), and MoMAColors (https://github.com/BlakeRMills/MoMAColors).

Momilactone measurements

Each harvested leaf blade was treated with 0.5 mM $CuCl_2$ for 72 h before the analysis. The treated sample (roughly 100 to 500 mg) was submerged in 4 ml of 70% methanol at 4°C for 24 h. For *O. coarctata*, methanol extract was concentrated to water and extracted by ethyl acetate, and evaporated to dryness, then re-dissolved in small amount of 70% methanol. 2 uL of the extract was subjected to LC-MS/MS analysis (Sciex QTRAP3200) with the following selected reaction monitoring transitions used as previously described (Miyamoto et al., 2016): momilactone A, m/z 315/271; momilactone B, m/z 331/269.

Cloning of OoKSL4 and OcKSL4 cDNAs

The cDNA encoding full-length OoKSL4-1 was amplified by RT-PCR using primers 5'-GGATCCATGGCGAATCCTGTGGAAG-3' (forward; BamHI site is underlined) and 5'-GTCGACTTAAGATGAACTTGTAGCACCTAG-3' (reverse; SalI site is underlined), and subcloned into pGEX-4T-3 (GE Healthcare) using BamHI and SalI. The cDNA encoding pseudo-mature OoKSL4-2 was amplified by RT-PCR using forward primer 5'-GGTTCCGCGTGGATCCGAGGCTAGAATAAGGAGGC-3' and reverse primer 5'-GGGAATTCGGGGATCCCTAGTAGCCTATAGTTTTC-3' (terminal sequences of pGEX-4T-3 digested by BamHI are underlined), and inserted into BamHI-digested pGEX-4T-3 by using In-Fusion Smart Assembly Cloning Kit (Takarabio-Clontech). Full-length *OcKSL4* cDNA was amplified by RT-PCR using primers 5'-GGATCCATGGCGAATTATCCCATGGAG-3' (forward; BamHI site is underlined) and 5'-TTAAGATGGACTAGTAGCCTCTAGTTTAAGTGGC-3' (reverse; NotI site is underlined), and subcloned into pGEM-T-Easy. *OcKSL4* cDNA was inserted into pGEX-4T-3 using BamHI and NotI. For RT-PCR, we used cDNA prepared from RNA from leaf blades of *O. officinalis* W0002 and *O. coarctata* W0551 24 h after UV treatment. OoKSL4-2_349W was generated by mutagenesis PCR using OoKSL4-2 in pGEX as template,

with primers 5'-gTCAGGTAGCTTATTGAAAC-3' (forward) and 5'-caACAACCCATTTTCTCTAGGATTG-3' (reverse; small characters indicate W codon for mutagenesis).

Activity measurement of OoKSL4

Functional analyses of OoKSL4-1, OoKSL4-2 and OcKSL4 were performed by a metabolic engineering system, as described previously (Toyomasu et al., 2016). The *n*-hexane extract from culture medium containing recombinant *E. coli* BL21-DE3 harbouring pRSF-PaGGS:OrCPS4ind and a respective construct to express one of the above proteins was subjected to GC-MS analysis. PaGGS is a geranylgeranyl synthase domain of fusicoccadiene synthase in *Phomopsis amygdali* (PaFS) (Toyomasu et al., 2007) and OrCPS4ind is a *syn*-CDP synthase from *O. rufipogon* W0106 (Miyamoto et al., 2016), to supply GGDP and *syn*-CDP, respectively. GC-MS was conducted using an Agilent 8860 GC-5977 MSD system, fitted with a fused silica, chemically bonded capillary column (HP-1MS; 0.25 mm in diameter, 60 m long, 0.25 μm film thickness). Ionization voltage was 70 eV. Each sample was injected onto the column at 250°C in the splitless mode. After a 2-min isothermal hold at 60℃, the column temperature was increased by 30℃/min to 150℃, 10℃/min to 180℃, 2℃/min to 210℃, and 30℃/min to 300℃. The flow rate of the helium carrier gas was 1 mL/min.

REFERENCES

Ahuja, I., Kissen, R., Bones, A.M., 2012. Phytoalexins in defense against pathogens. Trends Plant Sci. 17, 73–90. https://doi.org/10.1016/j.tplants.2011.11.002

Bao, Y., Ge, S., 2004. Origin and phylogeny of Oryza species with the CD genome based on multiple-gene sequence data. Plant Syst. Evol. 249, 55–66. https://doi.org/10.1007/s00606-004-0173-8

Bennetzen, J.L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A.C., Estep, M., Feng, L., Vaughn, J.N., Grimwood, J., Jenkins, J., Barry, K., Lindquist, E., Hellsten, U., Deshpande, S., Wang, X., Wu, X., Mitros, T., Triplett, J., Yang, X., Ye, C.-Y., Mauro-Herrera, M., Wang, L., Li, P., Sharma, M., Sharma, R., Ronald, P.C., Panaud, O., Kellogg, E.A., Brutnell, T.P., Doust, A.N., Tuskan, G.A., Rokhsar, D., Devos, K.M., 2012. Reference genome sequence of the model plant Setaria. Nat. Biotechnol. 30, 555–561. https://doi.org/10.1038/nbt.2196

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chen, J., Huang, Q., Gao, D., Wang, Junyi, Lang, Y., Liu, T., Li, B., Bai, Z., Luis Goicoechea, J., Liang, C., Chen, C., Zhang, W., Sun, S., Liao, Y., Zhang, X., Yang, L., Song, C., Wang, M., Shi, J., Liu, G., Liu, J., Zhou, H., Zhou, W., Yu, Q., An, N., Chen, Y., Cai, Q., Wang, B., Liu, B., Min, J., Huang, Y., Wu, H., Li, Z., Zhang, Y., Yin, Y., Song, W., Jiang, J., Jackson, S.A., Wing, R.A., Wang, Jun, Chen, M., 2013. Whole-genome sequencing of Oryza brachyantha reveals mechanisms underlying Oryza genome evolution. Nat. Commun. 4, 1595. https://doi.org/10.1038/ncomms2596

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor.

Bioinformatics 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin) 6, 80–92. https://doi.org/10.4161/fly.19695

Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., De Hoon, M.J.L., 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25, 1422–1423. https://doi.org/10.1093/bioinformatics/btp163

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. GigaScience 10, giab008. https://doi.org/10.1093/gigascience/giab008

De La Peña, R., Sattely, E.S., 2021. Rerouting plant terpene biosynthesis enables momilactone pathway elucidation. Nat. Chem. Biol. 17, 205–212. https://doi.org/10.1038/s41589-020-00669-3

Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E., Notredame, C., 2017. Nextflow enables reproducible computational workflows. Nat. Biotechnol. 35, 316–319. https://doi.org/10.1038/nbt.3820

Dunning, L.T., Olofsson, J.K., Parisod, C., Choudhury, R.R., Moreno-Villena, J.J., Yang, Y., Dionora, J., Quick, W.P., Park, M., Bennetzen, J.L., Besnard, G., Nosil, P., Osborne, C.P., Christin, P.-A., 2019. Lateral transfers of large DNA fragments spread functional genes among grasses. Proc. Natl. Acad. Sci. 116, 4416–4425. https://doi.org/10.1073/pnas.1810031116

Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20, 238. https://doi.org/10.1186/s13059-019-1832-y

Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157. https://doi.org/10.1186/s13059-015-0721-2

Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P., Nahnsen, S., 2020. The nf-core framework for community-curated bioinformatics pipelines. Nat. Biotechnol. 38, 276–278. https://doi.org/10.1038/s41587-020-0439-x

Frerebeau, N., 2023. khroma: Colour Schemes for Scientific Data Visualization. Zenodo. https://doi.org/10.5281/zenodo.8269306

Garcia, M., Juhos, S., Larsson, M., Olason, P.I., Martin, M., Eisfeldt, J., DiLorenzo, S., Sandgren, J., Ståhl, T.D.D., Ewels, P., Wirta, V., Nistér, M., Käller, M., Nystedt, B., 2020. Sarek: A portable workflow for whole-genome sequencing analysis of germline and somatic variants. F1000Research. https://doi.org/10.12688/f1000research.16665.2

Ge, S., Sang, T., Lu, B.-R., Hong, D.-Y., 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc. Natl. Acad. Sci. U. S. A. 96, 14400–14405.

Guo, L., Qiu, J., Ye, C., Jin, G., Mao, L., Zhang, H., Yang, X., Peng, Q., Wang, Yingying, Jia, L., Lin, Z., Li, G., Fu, F., Liu, C., Chen, L., Shen, E., Wang, W., Chu, Q., Wu, D., Wu, S., Xia, C., Zhang, Y., Zhou, X., Wang, L., Wu, L., Song, W., Wang, Yunfei, Shu, Q., Aoki, D., Yumoto, E., Yokota, T., Miyamoto, K., Okada, K., Kim, D.-S., Cai, D., Zhang, C., Lou, Y., Qian, Q., Yamaguchi, H., Yamane, H., Kong, C.-H., Timko, M.P., Bai, L., Fan, L., 2017. Echinochloa crus-galli genome analysis provides insight into its adaptation and invasiveness as a weed. Nat. Commun. 8, 1031. https://doi.org/10.1038/s41467-017-01067-5

Guo, Z.-H., Ma, P.-F., Yang, G.-Q., Hu, J.-Y., Liu, Y.-L., Xia, E.-H., Zhong, M.-C., Zhao, L., Sun, G.-L., Xu, Y.-X., Zhao, Y.-J., Zhang, Y.-C., Zhang, Y.-X., Zhang, X.-M., Zhou, M.-Y., Guo, Y., Guo, C., Liu, J.-X., Ye, X.-Y., Chen, Y.-M., Yang, Y., Han, B., Lin, C.-S., Lu, Y., Li, D.-Z., 2019. Genome Sequences Provide Insights into the Reticulate Origin and Unique Traits of Woody Bamboos. Mol. Plant 12, 1353–1365. https://doi.org/10.1016/j.molp.2019.05.009

Hibdige, S.G.S., Raimondeau, P., Christin, P.-A., Dunning, L.T., 2021. Widespread lateral gene transfer among grasses. New Phytol. 230, 2474–2486. https://doi.org/10.1111/nph.17328

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol. 35, 518–522. https://doi.org/10.1093/molbev/msx281

Itoh, A., Nakazato, S., Wakabayashi, H., Hamano, A., Shenton, M.R., Miyamoto, K., Mitsuhashi, W., Okada, K., Toyomasu, T., 2021. Functional kaurene-synthase-like diterpene synthases lacking a gamma domain are widely present in Oryza and related species. Biosci. Biotechnol. Biochem. 85, 1945–1952. https://doi.org/10.1093/bbb/zbab127

Kajiya-Kanegae, H., Ohyanagi, H., Ebata, T., Tanizawa, Y., Onogi, A., Sawada, Y., Hirai, M.Y., Wang, Z.-X., Han, B., Toyoda, A., Fujiyama, A., Iwata, H., Tsuda, K., Suzuki, T., Nosaka-Takahashi, M., Nonomura, K., Nakamura, Y., Kawamoto, S., Kurata, N., Sato, Y., 2021. OryzaGenome2.1: Database of Diverse Genotypes in Wild Oryza Species. Rice 14, 24. https://doi.org/10.1186/s12284-021-00468-x

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285

Kato, T., Kabuto, C., Sasaki, N., Tsunagawa, M., Aizawa, H., Fujita, K., Kato, Y., Kitahara, Y., Takahashi, N., 1973. Momilactones, growth inhibitors from rice, oryza sativa L. Tetrahedron Lett. 14, 3861–3864. https://doi.org/10.1016/S0040-4039(01)87058-1

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066. https://doi.org/10.1093/nar/gkf436

Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010

Kato-Noguchi, H., Hasegawa, M., Ino, T., Ota, K., Kujime, H., 2010. Contribution of momilactone A and B to rice allelopathy. J. Plant Physiol. 167, 787–791. https://doi.org/10.1016/j.jplph.2010.01.014

Kato-Noguchi, H., Ino, T., 2003. Rice seedlings release momilactone B into the environment. Phytochemistry 63, 551–554. https://doi.org/10.1016/S0031-9422(03)00194-8

Kitaoka, N., Zhang, J., Oyagbenro, R.K., Brown, B., Wu, Y., Yang, B., Li, Z., Peters, R.J., 2021. Interdependent evolution of biosynthetic gene clusters for momilactone production in rice. Plant Cell 33, 290–305. https://doi.org/10.1093/plcell/koaa023

Kodama, O., Suzuki, T., Miyakawa, J., Akatsuka, T., 1988. Ultraviolet-Induced Accumulation of Phytoalexins in Rice Leaves. Agric. Biol. Chem. 52, 2469–2473. https://doi.org/10.1080/00021369.1988.10869067

Kraehmer, H., Jabran, K., Mennan, H., Chauhan, B.S., 2016. Global distribution of rice weeds – A review. Crop Prot. 80, 73–86. https://doi.org/10.1016/j.cropro.2015.10.027

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. https://doi.org/10.48550/arXiv.1303.3997

Liu, Y., Balcke, G.U., Porzel, A., Mahdi, L., Scherr-Henning, A., Bathe, U., Zuccaro, A., Tissier, A., 2021. A barley gene cluster for the biosynthesis of diterpenoid phytoalexins. bioRxiv. https://doi.org/10.1101/2021.05.21.445084

Liu, Z., Cheema, J., Vigouroux, M., Hill, L., Reed, J., Paajanen, P., Yant, L., Osbourn, A., 2020. Formation and diversification of a paradigm biosynthetic gene cluster in plants. Nat. Commun. 11, 5354. https://doi.org/10.1038/s41467-020-19153-6

Lovell, J.T., Jenkins, J., Lowry, D.B., Mamidi, S., Sreedasyam, A., Weng, X., Barry, K., Bonnette, J., Campitelli, B., Daum, C., Gordon, S.P., Gould, B.A., Khasanova, A., Lipzen, A., MacQueen, A.,

Palacio-Mejía, J.D., Plott, C., Shakirov, E.V., Shu, S., Yoshinaga, Y., Zane, M., Kudrna, D., Talag, J.D., Rokhsar, D., Grimwood, J., Schmutz, J., Juenger, T.E., 2018. The genomic landscape of molecular responses to natural drought stress in Panicum hallii. Nat. Commun. 9, 5213. https://doi.org/10.1038/s41467-018-07669-x

Lu, F., Ammiraju, J.S.S., Sanyal, A., Zhang, S., Song, R., Chen, J., Li, G., Sui, Y., Song, X., Cheng, Z., Oliveira, A.C. de, Bennetzen, J.L., Jackson, S.A., Wing, R.A., Chen, M., 2009. Comparative sequence analysis of MONOCULM1-orthologous regions in 14 Oryza genomes. Proc. Natl. Acad. Sci. 106, 2071–2076. https://doi.org/10.1073/pnas.0812798106

Ma, P.-F., Liu, Y.-L., Jin, G.-H., Liu, J.-X., Wu, H., He, J., Guo, Z.-H., Li, D.-Z., 2021. The Pharus latifolius genome bridges the gap of early grass evolution. Plant Cell 33, 846–864. https://doi.org/10.1093/plcell/koab015

Mascher, M., Wicker, T., Jenkins, J., Plott, C., Lux, T., Koh, C.S., Ens, J., Gundlach, H., Boston, L.B., Tulpová, Z., Holden, S., Hernández-Pinzón, I., Scholz, U., Mayer, K.F.X., Spannagl, M., Pozniak, C.J., Sharpe, A.G., Šimková, H., Moscou, M.J., Grimwood, J., Schmutz, J., Stein, N., 2021. Long-read sequence assembly: a technical evaluation in barley. Plant Cell 33, 1888–1906. https://doi.org/10.1093/plcell/koab077

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. https://doi.org/10.1101/gr.107524.110

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015

Miyamoto, K., Fujita, M., Shenton, M.R., Akashi, S., Sugawara, C., Sakai, A., Horie, K., Hasegawa, M., Kawaide, H., Mitsuhashi, W., Nojiri, H., Yamane, H., Kurata, N., Okada, K., Toyomasu, T., 2016. Evolutionary trajectory of phytoalexin biosynthetic gene clusters in rice. Plant J. 87, 293–304. https://doi.org/10.1111/tpj.13200

Mondal, T.K., Rawal, H.C., Chowrasia, S., Varshney, D., Panda, A.K., Mazumdar, A., Kaur, H., Gaikwad, K., Sharma, T.R., Singh, N.K., 2018. Draft genome sequence of first monocot-halophytic species Oryza coarctata reveals stress-specific genes. Sci. Rep. 8, 13698. https://doi.org/10.1038/s41598-018-31518-y

Nützmann, H.-W., Huang, A., Osbourn, A., 2016. Plant metabolic clusters – from genetics to genomics. New Phytol. 211, 771–789. https://doi.org/10.1111/nph.13981

Okada, K., 2011. The Biosynthesis of Isoprenoids and the Mechanisms Regulating It in Plants. Biosci. Biotechnol. Biochem. 75, 1219–1225. https://doi.org/10.1271/bbb.110228

Otomo, K., Kanno, Y., Motegi, A., Kenmoku, H., Yamane, H., Mitsuhashi, W., Oikawa, H., Toshima, H., Itoh, H., Matsuoka, M., Sassa, T., Toyomasu, T., 2004a. Diterpene Cyclases Responsible for the Biosynthesis of Phytoalexins, Momilactones A, B, and Oryzalexins A–F in Rice. Biosci. Biotechnol. Biochem. 68, 2001–2006. https://doi.org/10.1271/bbb.68.2001

Otomo, K., Kenmoku, H., Oikawa, H., König, W.A., Toshima, H., Mitsuhashi, W., Yamane, H., Sassa, T., Toyomasu, T., 2004b. Biological functions of ent- and syn-copalyl diphosphate synthases in rice: key enzymes for the branch point of gibberellin and phytoalexin biosynthesis. Plant J. 39, 886–893. https://doi.org/10.1111/j.1365-313X.2004.02175.x

Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R.L., Lee, Y., Zheng, L., Orvis, J., Haas, B., Wortman, J., Buell, C.R., 2007. The TIGR Rice Genome Annotation Resource: improvements and new features. Nucleic Acids Res. 35, D883–D887. https://doi.org/10.1093/nar/gkl976

Pertea, G., Pertea, M., 2020. GFF Utilities: GffRead and GffCompare. F1000Research. https://doi.org/10.12688/f1000research.23297.1

Phillips, A.L., Ferguson, S., Watson-Haigh, N.S., Jones, A.W., Borevitz, J.O., Burton, R.A., Atwell, B.J., 2022. The first long-read nuclear genome assembly of Oryza australiensis, a wild rice from northern Australia. Sci. Rep. 12, 10823. https://doi.org/10.1038/s41598-022-14893-5

Polturak, G., Dippe, M., Stephenson, M.J., Chandra Misra, R., Owen, C., Ramirez-Gonzalez, R.H., Haidoulis, J.F., Schoonbeek, H.-J., Chartrain, L., Borrill, P., Nelson, D.R., Brown, J.K.M., Nicholson, P., Uauy, C., Osbourn, A., 2022a. Pathogen-induced biosynthetic pathways encode defense-related molecules in bread wheat. Proc. Natl. Acad. Sci. 119, e2123299119. https://doi.org/10.1073/pnas.2123299119

Polturak, G., Liu, Z., Osbourn, A., 2022b. New and emerging concepts in the evolution and function of plant biosynthetic gene clusters. Curr. Opin. Green Sustain. Chem. 33, 100568. https://doi.org/10.1016/j.cogsc.2021.100568

Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Auwera, G.A.V. der, Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M.J., Neale, B., MacArthur, D.G., Banks, E., 2018. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv 201178. https://doi.org/10.1101/201178

Shenton, M., Kobayashi, M., Terashima, S., Ohyanagi, H., Copetti, D., Hernández-Hernández, T., Zhang, J., Ohmido, N., Fujita, M., Toyoda, A., Ikawa, H., Fujiyama, A., Furuumi, H., Miyabayashi, T., Kubo, T., Kudrna, D., Wing, R., Yano, K., Nonomura, K.-I., Sato, Y., Kurata, N., 2020. Evolution and Diversity of the Wild Rice Oryza officinalis Complex, across Continents, Genome Types, and Ploidy Levels. Genome Biol. Evol. 12, 413–428. https://doi.org/10.1093/gbe/evaa037

Shi, C., Li, W., Zhang, Q.-J., Zhang, Y., Tong, Y., Li, K., Liu, Y.-L., Gao, L.-Z., 2020. The draft genome sequence of an upland wild rice species, Oryza granulata. Sci. Data 7, 131. https://doi.org/10.1038/s41597-020-0470-2

Shimura, K., Okada, A., Okada, K., Jikumaru, Y., Ko, K.-W., Toyomasu, T., Sassa, T., Hasegawa, M., Kodama, O., Shibuya, N., Koga, J., Nojiri, H., Yamane, H., 2007. Identification of a Biosynthetic Gene Cluster in Rice for Momilactones. J. Biol. Chem. 282, 34013–34018. https://doi.org/10.1074/jbc.M703344200

Smit, S.J., Lichman, B.R., 2022. Plant biosynthetic gene clusters in the context of metabolic evolution. Nat. Prod. Rep. https://doi.org/10.1039/D2NP00005A

Soreng, R.J., Peterson, P.M., Romaschenko, K., Davidse, G., Teisher, J.K., Clark, L.G., Barberá, P., Gillespie, L.J., Zuloaga, F.O., 2017. A worldwide phylogenetic classification of the Poaceae (Gramineae) II: An update and a comparison of two 2015 classifications. J. Syst. Evol. 55, 259–290. https://doi.org/10.1111/jse.12262

Stanke, M., Diekhans, M., Baertsch, R., Haussler, D., 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24, 637–644. https://doi.org/10.1093/bioinformatics/btn013

Stein, J.C., Yu, Y., Copetti, D., Zwickl, D.J., Zhang, L., Zhang, C., Chougule, K., Gao, D., Iwata, A., Goicoechea, J.L., Wei, S., Wang, J., Liao, Y., Wang, M., Jacquemin, J., Becker, C., Kudrna, D., Zhang, J., Londono, C.E.M., Song, X., Lee, S., Sanchez, P., Zuccolo, A., Ammiraju, J.S.S., Talag, J., Danowitz, A., Rivera, L.F., Gschwend, A.R., Noutsos, C., Wu, C., Kao, S., Zeng, J., Wei, F., Zhao, Q., Feng, Q., El Baidouri, M., Carpentier, M.-C., Lasserre, E., Cooke, R., Rosa Farias, D. da, da Maia, L.C., dos Santos, R.S., Nyberg, K.G., McNally, K.L., Mauleon, R., Alexandrov, N., Schmutz, J., Flowers, D., Fan, C., Weigel, D., Jena, K.K., Wicker, T., Chen, M., Han, B., Henry, R., Hsing, Y.C., Kurata, N., de Oliveira, A.C., Panaud, O., Jackson, S.A., Machado, C.A., Sanderson, M.J., Long, M., Ware, D., Wing, R.A., 2018. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus Oryza. Nat. Genet. 50, 285–296. https://doi.org/10.1038/s41588-018-0040-0

Takos, A.M., Rook, F., 2012. Why biosynthetic genes for chemical defense compounds cluster. Trends Plant Sci. 17, 383–388. https://doi.org/10.1016/j.tplants.2012.04.004

Tamogani, S., Mitani, M., Kodama, O., Akatsuka, T., 1993. Oryzalexin S structure: a new stemarane-type rice plant phytoalexin and its biogenesis. Tetrahedron 49, 2025–2032. https://doi.org/10.1016/S0040-4020(01)86302-X

Toyomasu, T., Miyamoto, K., Shenton, M.R., Sakai, A., Sugawara, C., Horie, K., Kawaide, H., Hasegawa, M., Chuba, M., Mitsuhashi, W., Yamane, H., Kurata, N., Okada, K., 2016. Characterization and evolutionary analysis of ent-kaurene synthase like genes from the wild rice species Oryza rufipogon. Biochem. Biophys. Res. Commun. 480, 402–408. https://doi.org/10.1016/j.bbrc.2016.10.062

Toyomasu, T., Shenton, M.R., Okada, K., 2020. Evolution of Labdane-Related Diterpene Synthases in Cereals. Plant Cell Physiol. 61, 1850–1859. https://doi.org/10.1093/pcp/pcaa106

Toyomasu, T., Tsukahara, M., Kaneko, A., Niida, R., Mitsuhashi, W., Dairi, T., Kato, N., Sassa, T., 2007. Fusicoccins are biosynthesized by an unusual chimera diterpene synthase in fungi. Proc. Natl. Acad. Sci. 104, 3084–3088. https://doi.org/10.1073/pnas.0608426104

Vogel, J.P., Garvin, D.F., Mockler, T.C., Schmutz, J., Rokhsar, D., Bevan, M.W., Barry, K., Lucas, S., Harmon-Smith, M., Lail, K., Tice, H., Schmutz (Leader), J., Grimwood, J., McKenzie, N., Bevan, M.W., Huo, N., Gu, Y.Q., Lazo, G.R., Anderson, O.D., Vogel (Leader), J.P., You, F.M., Luo, M.-C., Dvorak, J., Wright, J., Febrer, M., Bevan, M.W., Idziak, D., Hasterok, R., Garvin, D.F., Lindquist, E., Wang, M., Fox, S.E., Priest, H.D., Filichkin, S.A., Givan, S.A., Bryant, D.W., Chang, J.H., Mockler (Leader), T.C., Wu, H., Wu, W., Hsia, A.-P., Schnable, P.S., Kalyanaraman, A., Barbazuk, B., Michael, T.P., Hazen, S.P., Bragg, J.N., Laudencia-Chingcuanco, D., Vogel, J.P., Garvin, D.F., Weng, Y., McKenzie, N., Bevan, M.W., Haberer, G., Spannagl, M., Mayer (Leader), K., Rattei, T., Mitros, T., Rokhsar, D., Lee, S.-J., Rose, J.K.C., Mueller, L.A., York, T.L., Wicker (Leader), T., Buchmann, J.P., Tanskanen, J., Schulman (Leader), A.H., Gundlach, H., Wright, J., Bevan, M., Costa de Oliveira, A., da C. Maia, L., Belknap, W., Gu, Y.Q., Jiang, N., Lai, J., Zhu, L., Ma, J., Sun, C., Pritham, E., Salse (Leader), J., Murat, F., Abrouk, M., Haberer, G., Spannagl, M., Mayer, K., Bruggmann, R., Messing, J., You, F.M., Luo, M.-C., Dvorak, J., Fahlgren, N., Fox, S.E., Sullivan, C.M., Mockler, T.C., Carrington, J.C., Chapman, E.J., May, G.D., Zhai, J., Ganssmann, M., Guna Ranjan Gurazada, S., German, M., Meyers, B.C., Green (Leader), P.J., Bragg, J.N., Tyler, L., Wu, J., Gu, Y.Q., Lazo, G.R., Laudencia-Chingcuanco, D., Thomson, J., Vogel (Leader), J.P., Hazen, S.P., Chen, S., Scheller, H.V., Harholt, J., Ulvskov, P., Fox, S.E., Filichkin, S.A., Fahlgren, N., Kimbrel, J.A., Chang, J.H., Sullivan, C.M., Chapman, E.J., Carrington, J.C., Mockler, T.C., Bartley, L.E., Cao, P., Jung, K.-H., Sharma, M.K., Vega-Sanchez, M., Ronald, P., Dardick, C.D., De Bodt, S., Verelst, W., Inzé, D., Heese, M., Schnittger, A., Yang, X., Kalluri, U.C., Tuskan, G.A., Hua, Z., Vierstra, R.D., Garvin, D.F., Cui, Y., Ouyang, S., Sun, Q., Liu, Z., Yilmaz, A., Grotewold, E., Sibout, R., Hematy, K., Mouille, G., Höfte, H., Michael, T., Pelloux, J., O'Connor, D., Schnable, J., Rowe, S., Harmon, F., Cass, C.L., Sedbrook, J.C., Byrne, M.E., Walsh, S., Higgins, J., Bevan, M., Li, P., Brutnell, T., Unver, T., Budak, H., Belcram, H., Charles, M., Chalhoub, B., Baxter, I., The International Brachypodium Initiative, Principal investigators, DNA sequencing and assembly, Pseudomolecule assembly and BAC end sequencing, Transcriptome sequencing and analysis, Gene analysis and annotation, Repeats analysis, Comparative genomics, Small RNA analysis, Manual annotation and gene family analysis, 2010. Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 463, 763–768. https://doi.org/10.1038/nature08747

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. J. Open Source Softw. 4, 1686. https://doi.org/10.21105/joss.01686

Wilderman, P.R., Xu, M., Jin, Y., Coates, R.M., Peters, R.J., 2004. Identification of Syn-Pimara-7,15-Diene Synthase Reveals Functional Clustering of Terpene Synthases Involved in Rice Phytoalexin/Allelochemical Biosynthesis. Plant Physiol. 135, 2098–2105. https://doi.org/10.1104/pp.104.045971

Wu, D., Hu, Y., Akashi, S., Nojiri, H., Guo, L., Ye, C.-Y., Zhu, Q.-H., Okada, K., Fan, L., 2022a. Lateral transfers lead to the birth of momilactone biosynthetic gene clusters in grass. Plant J. 111, 1354–1367. https://doi.org/10.1111/tpj.15893

Wu, D., Shen, E., Jiang, B., Feng, Y., Tang, W., Lao, S., Jia, L., Lin, H.-Y., Xie, L., Weng, X., Dong, C., Qian, Q., Lin, F., Xu, H., Lu, H., Cutti, L., Chen, H., Deng, S., Guo, L., Chuah, T.-S., Song, B.-K., Scarabel, L., Qiu, J., Zhu, Q.-H., Yu, Q., Timko, M.P., Yamaguchi, H., Merotto, A., Qiu, Y., Olsen, K.M., Fan, L., Ye, C.-Y., 2022b. Genomic insights into the evolution of Echinochloa species as weed and orphan crop. Nat. Commun. 13, 689. https://doi.org/10.1038/s41467-022-28359-9

Xu, M., Hillwig, M.L., Prisic, S., Coates, R.M., Peters, R.J., 2004. Functional identification of rice syn-copalyl diphosphate synthase and its role in initiating biosynthesis of diterpenoid phytoalexin/allelopathic natural products. Plant J. 39, 309–318. https://doi.org/10.1111/j.1365-313X.2004.02137.x

Yan, N., Yang, T., Yu, X.-T., Shang, L.-G., Guo, D.-P., Zhang, Y., Meng, L., Qi, Q.-Q., Li, Y.-L., Du, Y.-M., Liu, X.-M., Yuan, X.-L., Qin, P., Qiu, J., Qian, Q., Zhang, Z.-F., 2022. Chromosome-level genome assembly of Zizania latifolia provides insights into its seed shattering and phytocassane biosynthesis. Commun. Biol. 5, 1–11. https://doi.org/10.1038/s42003-021-02993-3

Yang, X., Gao, S., Guo, L., Wang, B., Jia, Y., Zhou, J., Che, Y., Jia, P., Lin, J., Xu, T., Sun, J., Ye, K., 2021. Three chromosome-scale Papaver genomes reveal punctuated patchwork evolution of the morphinan and noscapine biosynthesis pathway. Nat. Commun. 12, 6030. https://doi.org/10.1038/s41467-021-26330-8

Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T.-Y., 2017. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. 8, 28–36. https://doi.org/10.1111/2041-210X.12628

Yu, H., Lin, T., Meng, X., Du, H., Zhang, J., Liu, G., Chen, Mingjiang, Jing, Y., Kou, L., Li, X., Gao, Q., Liang, Yan, Liu, X., Fan, Z., Liang, Yuntao, Cheng, Z., Chen, Mingsheng, Tian, Z., Wang, Y., Chu, C., Zuo, J., Wan, J., Qian, Q., Han, B., Zuccolo, A., Wing, R.A., Gao, C., Liang, C., Li, J., 2021. A route to de novo domestication of wild allotetraploid rice. Cell 184, 1156-1170.e14. https://doi.org/10.1016/j.cell.2021.01.013

Zhou, Y., Ma, Y., Zeng, J., Duan, L., Xue, X., Wang, H., Lin, T., Liu, Z., Zeng, K., Zhong, Y., Zhang, S., Hu, Q., Liu, M., Zhang, H., Reed, J., Moses, T., Liu, Xinyan, Huang, P., Qing, Z., Liu, Xiubin, Tu, P., Kuang, H., Zhang, Z., Osbourn, A., Ro, D.-K., Shang, Y., Huang, S., 2016. Convergence and divergence of bitterness biosynthesis and regulation in Cucurbitaceae. Nat. Plants 2, 1–8. https://doi.org/10.1038/nplants.2016.183

Zi, J., Mafu, S., Peters, R.J., 2014. To Gibberellins and Beyond! Surveying the Evolution of (Di)Terpenoid Metabolism. Annu. Rev. Plant Biol. 65, 259–286. https://doi.org/10.1146/annurev-arplant-050213-035705

AUTHOR CONTRIBUTIONS

S.P.C. and C.B. conceived the study. S.P.C. carried out the genomic and phylogenetic analyses. T.T. and Y.H. carried out in vitro enzymatic assays. the functional analysis of KSL4. Y.L. and K.O. carried out momilactone measurements. H.N., K.O., and C.B. supervised the work. S.P.C. and C.B. wrote the manuscript with contributions from all authors.

ACKNOWLEDGEMENTS

We are grateful to José Villaécija Aguilar and Niklas Schandry for critical reading of the manuscript and to Niklas Schandry for setting up and maintaining environments for computational analyses. We would like to thank Reshi Shanmuganathan for help with identifying orthologs. We acknowledge the Graduate School Life Science Munich (LSM) for their valuable support to S.P.C. We would also like to thank Prof. Yutaka Sato and Mr. Hiroyasu Furumi at the National Institute of Genetics, Japan, for providing wild rice material, and Chisato Goda (Yamagata University, Japan) for technical support.

FUNDING

This work was supported by the Deutsche Forschungsgemeinschaft (DFG; grant no. 457739273 (C.B.) and by JSPS KAKENHI Grant Numbers 20H02922 (K.O.) and 23H02148 (T.T.). For computational analyses, we used the BioHPC-Genomics compute cluster at the Leibniz Rechenzentrum (DFG; grant no. 450674345).

SUPPLEMENTAL MATERIAL

Supplementary Figure 1: Phylogenetic analysis based on CPS4 amino acid sequence.

Supplementary Figure 2: Phylogenetic analysis based on KSL4 amino acid sequence.

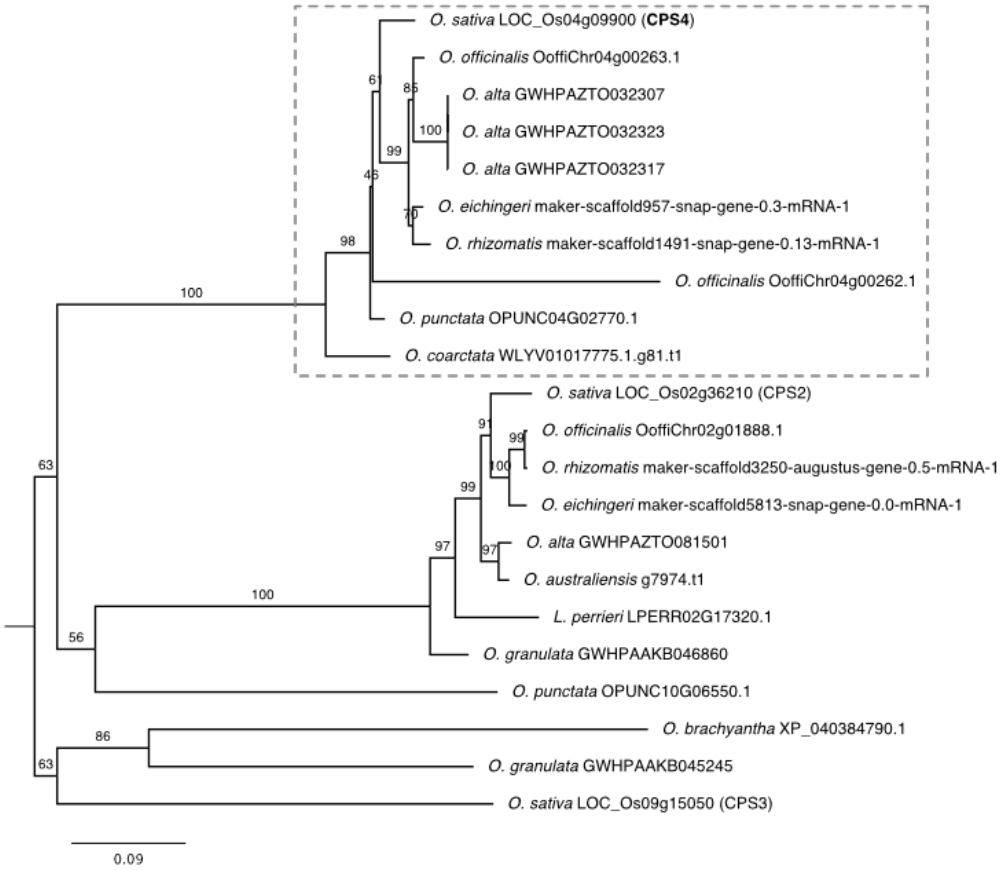
Supplementary Figure 3: Phylogenetic analysis based on MAS amino acid sequences.

Supplementary Figure 4: Phylogenetic analysis based on CYP99A amino acid sequences.

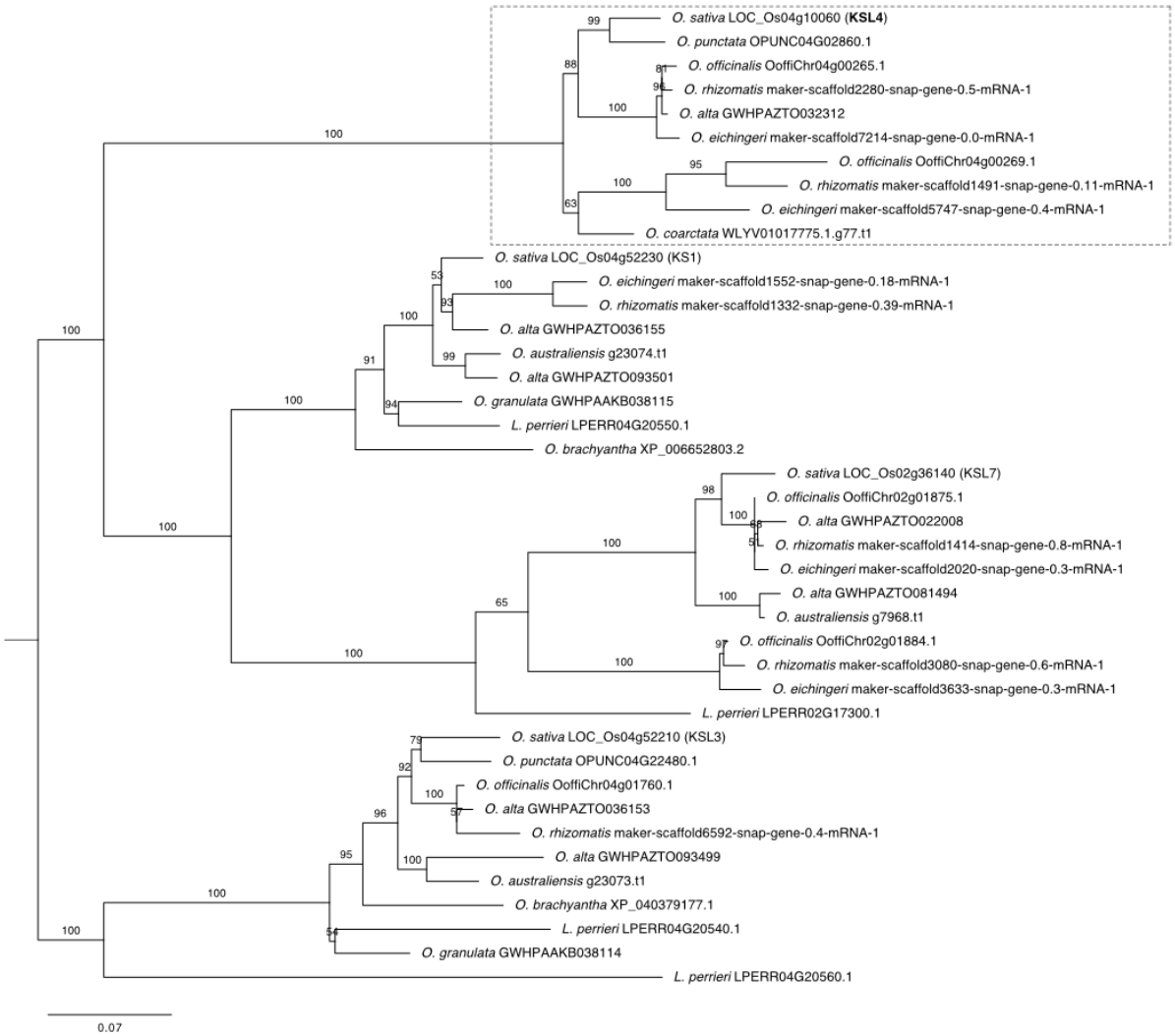
Supplementary Figure 5: Amino acid sequence alignment of KSL4 from *O. sativa* and its orthologs in *O. officinalis*.

Supplementary Figure 6: Microsynteny analysis of the genomic regions containing the MBGC-like.

SUPPLEMENTARY FIGURES

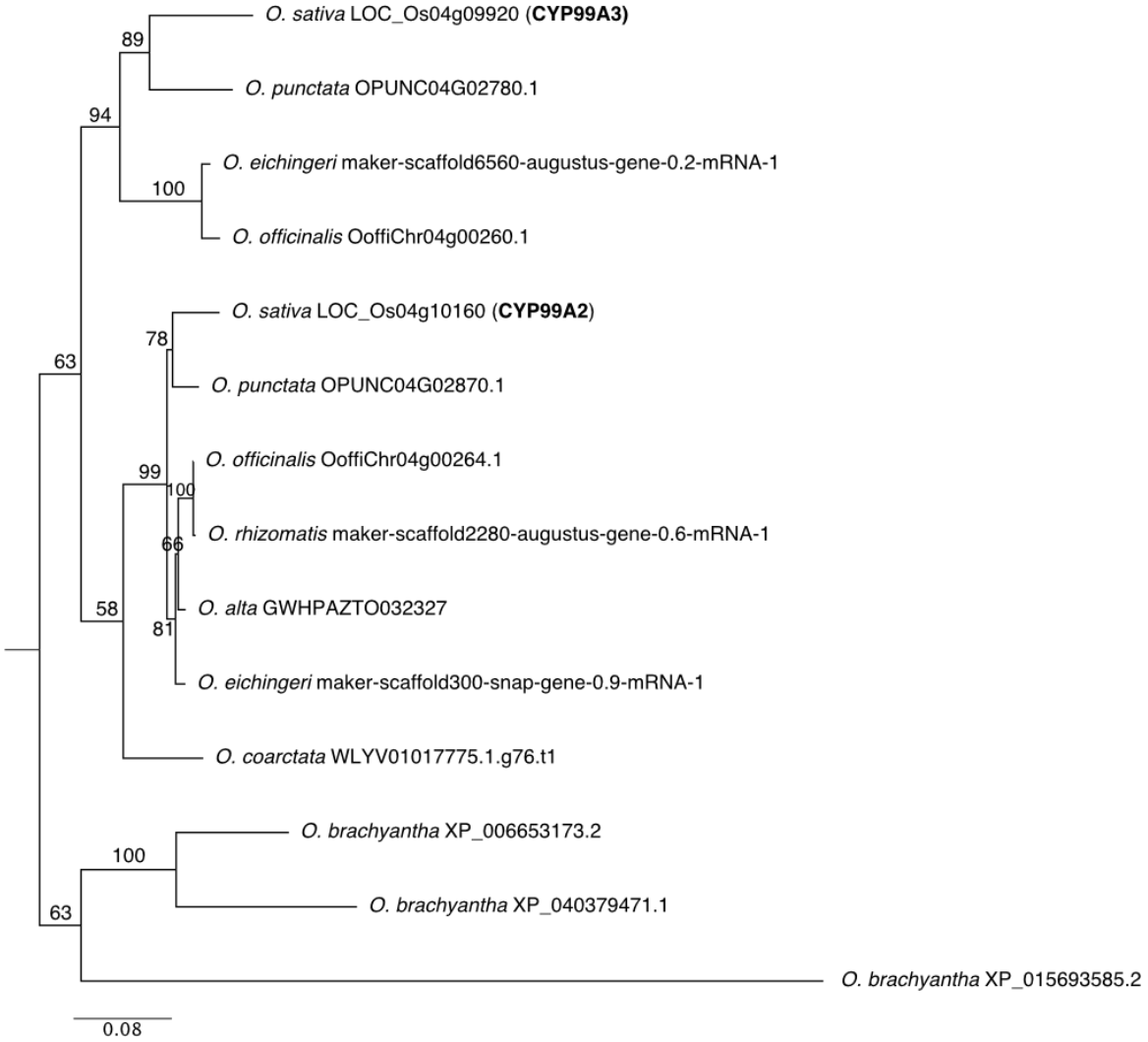


Supplementary Figure 1. Phylogenetic analysis based on CPS4 amino acid sequence. Maximum likelihood tree with the best-fit model selected by ModelFinder for the orthogroup containing *O. sativa* CPS4 (LOC_Os04g09900) in *O. sativa*, *O. punctata*, *O. officinalis*, *O. eichingeri*, *O. rhizomatis*, *O. alta*, *O. australiensis*, *O. brachyantha*, *O. granulata* and *L. perrieri*. Numbers represent 1000 replicates of ultrafast bootstrapping. Dashed rectangle highlights the CPS4 clade, which contains orthologs potentially involved in momilactones biosynthesis and clustered in *Oryza* species.



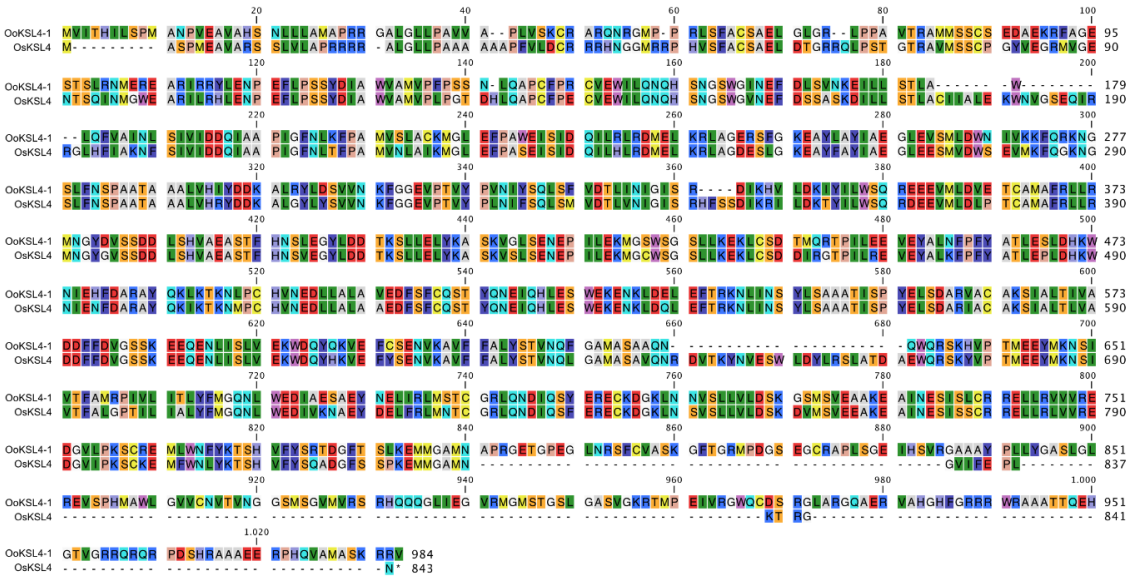
Supplementary Figure 2. Phylogenetic analysis based on KSL4 amino acid sequence. Maximum likelihood tree with the best-fit model selected by ModelFinder for the orthogroup containing *O. sativa* KSL4 (LOC_Os04g10060) in *O. sativa*, *O. punctata*, *O. officinalis*, *O. eichingeri*, *O. rhizomatis*, *O. alta*, *O. australiensis*, *O. brachyantha*, *O. granulata* and *L. perrieri*. Numbers represent 1000 replicates of ultrafast bootstrapping. Dashed rectangle highlights the KSL4 clade, which contains orthologs potentially involved in momilactones biosynthesis and clustered in *Oryza* species.

Supplementary Figure 3. Phylogenetic analysis based on MAS amino acid sequences. Maximum likelihood tree with the best-fit model selected by ModelFinder for the orthogroup containing *O. sativa* MAS1 and MAS2 (LOC_Os04g10010, LOC_Os04g10000) in *O. sativa*, *O. punctata*, *O. officinalis*, *O. eichingeri*, *O. rhizomatis*, *O. alta*, *O. australiensis*, *O. brachyantha*, *O. granulata* and *L. perrieri*. Numbers represent 1000 replicates of ultrafast bootstrapping. Since this orthogroup consisted of a small number of genes, we considered all of them as potentially involved in momilactone biosynthesis and clustered.

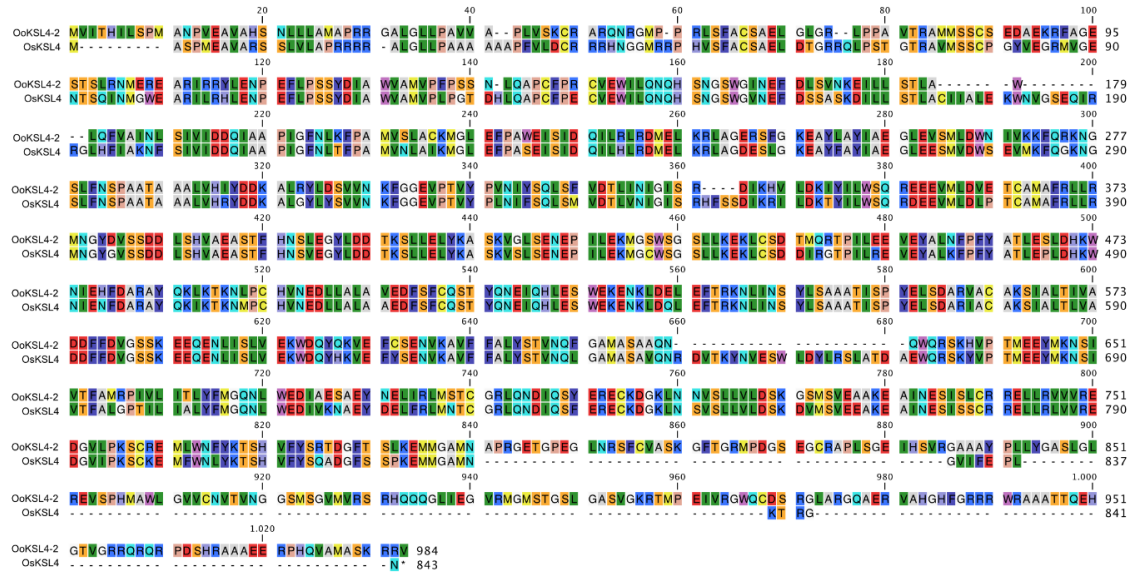


Supplementary Figure 4. Phylogenetic analysis based on CYP99A amino acid sequences. Maximum likelihood tree with the best-fit model selected by ModelFinder for the orthogroup containing *O. sativa* CYP99A2 and CYP99A3 (LOC_Os04g10160, LOC_Os04g09920) in *O. sativa*, *O. punctata*, *O. officinalis*, *O. eichingeri*, *O. rhizomatis*, *O. alta*, *O. australiensis*, *O. brachyantha*, *O. granulata* and *L. perrieri*. Numbers represent 1000 replicates of ultrafast bootstrapping. Since this orthogroup consisted of a small number of genes, we considered all of them as potentially involved in momilactone biosynthesis and clustered.

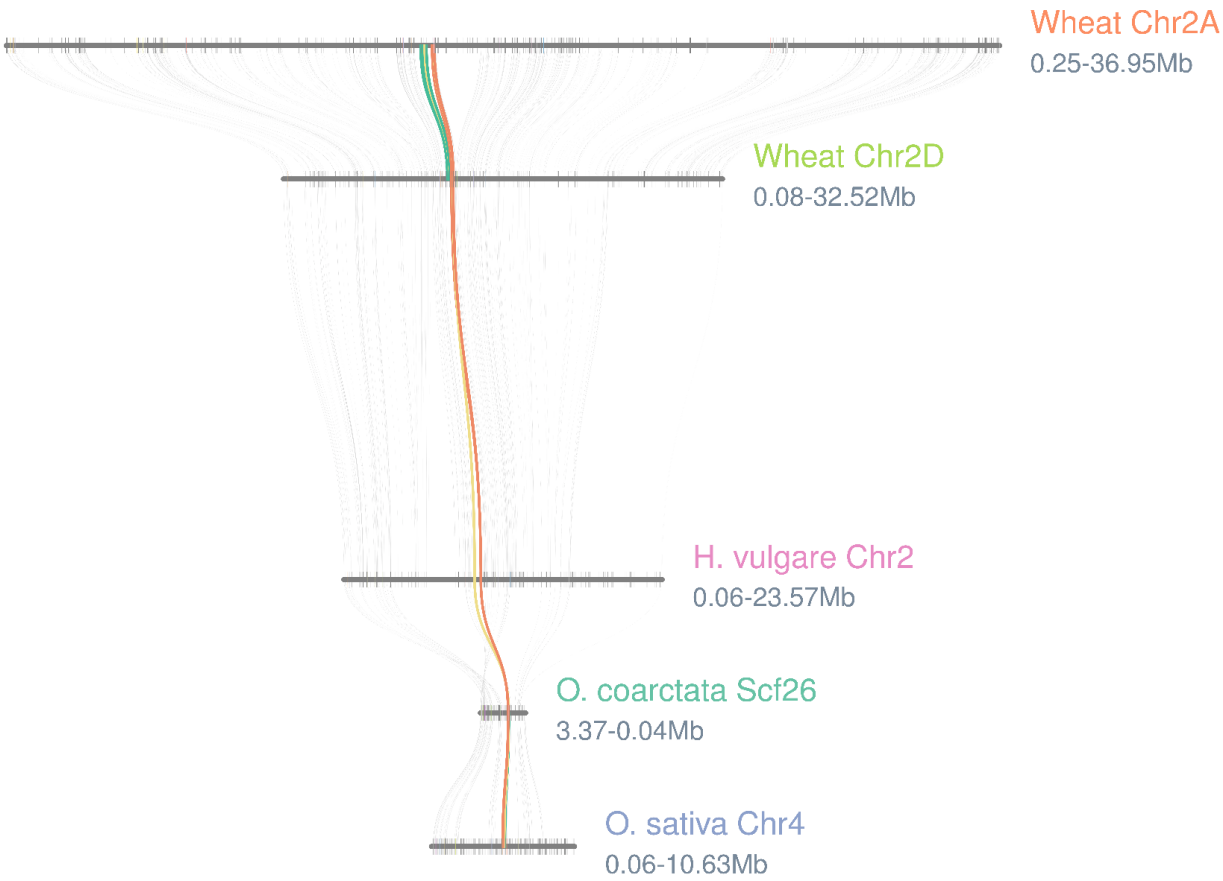
(a)



(b)



Supplementary Figure 5: Amino acid sequence alignment of KSL4 from *O. sativa* and its orthologs in *O. officinalis*. a) OoKSL4-1 and OsKSL4, and b) OoKSL4-2 and OsKSL4.



Supplementary Figure 6. Microsynteny analysis of the genomic regions containing the MBGC-like from *Triticum aestivum* (wheat), *Hordeum vulgare* and the MBGC from *O. coarctata* and *O. sativa*. Coloured lines represent MBGC-like orthologs and grey lines connect the respective orthologs.