

# SCIENTIFIC REPORTS



OPEN

## Multiple origins of BBCC allopolyploid species in the rice genus (*Oryza*)

Received: 02 July 2015

Accepted: 10 September 2015

Published: 13 October 2015

In the rice genus (*Oryza*), about one half of the species are allopolyploids. These species are not only important resources for rice breeding but also provide a unique opportunity for studying evolution of polyploid species. In the present study, we sequenced four biparentally inherited nuclear loci and three maternally inherited chloroplast fragments from all diploid and tetraploid species with the B- and C-genome types in this genus. We detected at least three independent origins of three BC-genome tetraploid species. Specifically, the diploid *O. punctata* (B-genome) and *O. officinalis* (C-genome) were the parental progenitors of *O. minuta* and *O. malampuzhaensis* with *O. punctata* being the maternal donors, whereas the diploid *O. punctata* and *O. eichingeri* (C-genome) were the progenitors of tetraploid *O. punctata* with *O. punctata* being the paternal donor. Our relaxed clock analyses suggest that all the BBCC species originated within the last one million years, which is coincident with the severe climate oscillations occurred during the last ice age, implying the potential impact of climate change on their formations and dispersals. In addition, our results support previous taxonomic arguments that the tetraploid *O. punctata* might be better treated as a separate species (*O. schweinfurthiana*).

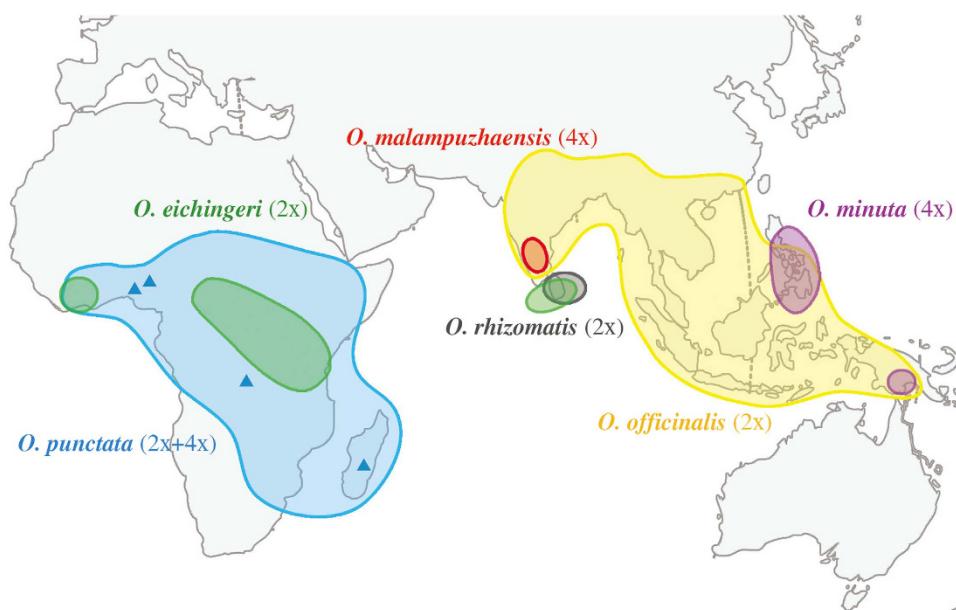
Hybridization between diploid species associated with genome doubling to produce an allopolyploid species is a prevalent phenomenon in plant evolution<sup>1,2</sup>. Recent studies have demonstrated that allopolyploidization (interspecific hybridization and genome doubling) is one of the major modes of diversification and speciation in plants, and the important source of morphological innovations<sup>3</sup>. Many crop plants such as wheat, cotton, tobacco, and *Brassica* spp. (cabbage/rape) are of allopolyploid origins<sup>4</sup>. Thus, studying the origin of allopolyploids is not only the key to the understanding of consequences and mechanisms of polyploidization and crop domestication, but will facilitate also the genetic improvements for important crops and utilization of genetic resources in wild relatives of crop plants<sup>1,4</sup>.

As one of the major crops, cultivated rice belongs to *Oryza*, a medium-size genus that consists of two cultivated and about 20 wild species<sup>5,6</sup>. Because of its economic importance and tremendous genetic and genomic resources available, rice along with its wild relatives has become a unique model for both theoretical studies and practical crop improvement<sup>7–9</sup>. Particularly, with the completion of genome sequencing of two rice subspecies, the initiation and implementation of the *Oryza* Map Alignment Project (OMAP) have laid an important foundation for a complete genomic interrogation of the wild relatives of rice<sup>7</sup>.

Of 10 distinct genome types recognized for *Oryza* species, six are diploid (A, B, C, E, F and G) ( $2n = 2x = 24$ ) and the other four are allotetraploid (BC, CD, HJ and HK) ( $2n = 4x = 48$ )<sup>10</sup>. Remarkably, about one half of the species in this genus are allotetraploids that originated through interspecific hybridization and genome doubling<sup>6,7,11</sup>. These allotetraploid species involved multiple independent

<sup>1</sup>State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, 100093, China. <sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China. <sup>3</sup>School of Integrative Plant Biology, Plant Breeding & Genetics Section, Cornell University, Ithaca, NY 14853, USA. <sup>†</sup>Present address: College of Horticulture and Landscape Architecture, Southwest University, Chongqing 400716, China.

<sup>‡</sup>Present address: Freshwater Ecological Field Station of Liangzi Lake, Wuhan University, Wuhan 430072, China. Correspondence and requests for materials should be addressed to S.G. (email: gesong@ibcas.ac.cn)



**Figure 1.** Geographical distribution of the *Oryza* species with the B-, C- and BC-genome types using software Adobe PhotoShop (Data from Vaughan 1994 and 1989)<sup>6,20</sup>. Blue, diploid *O. punctata* with the tetraploid *O. punctata* indicated by triangles; Yellow, *O. officinalis*; Green, *O. eichingeri*; Grey, *O. rhizomatis*; Purple, *O. minuta*; Red, *O. malampuzhaensis*.

polyploidizations, and include both ancient allotetraploids (HHJJ and HHKK) and recently formed ones (BBCC and CCDD)<sup>7,10,11</sup>, thus providing an increasingly attractive system for studies on the evolutionary dynamics of polyploid genomes organization and the mechanism of polyploidization in plants<sup>12–14</sup>. In addition, wild rice species including the polyploids have evolved under a wide range of climatic, geographic and ecological conditions over millions of years and conserved many useful alleles associated with yield-related traits and resistance to many abiotic and biotic stresses<sup>6</sup>. Previous studies showed that many valuable genes or alleles had been successfully transferred to cultivated rice from tetraploids, such as *O. minuta* (BBCC) and *O. grandiglumis* (CCDD), and have significantly increased rice yield and resistance to various diseases and pests<sup>15,16</sup>. Therefore, illustrating the origin and relationships of these tetraploids would also facilitate rice breeding and improvements.

To date, taxonomy and phylogeny of the rice genus have been extensively investigated and the evolutionary framework has been well established at the genus level<sup>7,10,17–19</sup>. Of the four tetraploid genomes, HJ and HK are ancient genomes and the diploid species that contributed the H, J, or K genomes have not been found (most likely extinct) despite great efforts. By contrast, two other tetraploid genomes (BC and CD) originated relatively recently and their parental diploid genomes/species are distributed widely across several continents<sup>6,20</sup> (Fig. 1). Multiple lines of evidence have demonstrated that the CD genome (including three species) originated through a single allopolyploidization event with the C genome as the maternal parent<sup>10,21</sup>. In contrast, questions concerning the origin of the BC-genome species, including where, when and how the tetraploids originated have remained unanswered, although early preliminary studies suggested that the BC-genome species exhibited multiple origins or experienced introgression from sympatric diploid C-genome species<sup>10,22,23</sup>. These uncertainties have in turn caused confusion in taxonomy of the B-, C-, and BC-genome species<sup>6,11,24,25</sup>.

Based on several decades of field expeditions, at least three tetraploid species with the BC-genome type have been found so far<sup>6,20</sup>, two in Asia (*O. minuta* J.S. Presl. et C.B. Presl. and *O. malampuzhaensis* Krish. et Chand.) and the third (*O. punctata* Kotchy ex Steud.) in Africa (Fig. 1). In Asia, *O. minuta* is distributed in Philippines and Papua New Guinea, and *O. malampuzhaensis* has a localized distribution in South India near the town of Malampuzha<sup>25</sup>. For the diploid species, only a B-genome species (*O. punctata*) was found, which is widely distributed from east to west Africa (Fig. 1). Of three C-genome species, *O. officinalis* Wall. ex Watt is the most common species and is widely distributed in south China, South and Southeast Asia, and Papua New Guinea; whereas *O. rhizomatis* Vaughan has only been reported from Sri Lanka. The third C-genome species, *O. eichingeri* A. Peter, is disjunctively distributed in Sri Lanka and West and East Africa (Fig. 1)<sup>20,26</sup>. The overlapping geographical distribution and similarities in gross morphology between the diploid and tetraploid species lead further to the complexity of taxonomy and phylogeny of this group of species<sup>6,11</sup>.

Despite substantial attempts in previous studies on the B-, C-, and BC-genome species, most of them did not include all the diploid and tetraploid species in their studies<sup>10,22,23,27</sup>. Based on SSR and PCR-RFLP analyses, Bao *et al.* investigated the genetic diversity and species relationships by sampling

Species	Genome types	Source	Accession	Abbreviation
<i>O. punctata</i>	BB	Tanzania	103903	pun_103903
<i>O. punctata</i>	BB	Cameroon	104071	pun_104071
<i>O. punctata</i>	BB	Chad	105607	pun_105607
<i>O. eichingeri</i>	CC	Uganda	101422	eic_101422
<i>O. eichingeri</i>	CC	Uganda	105159	eic_105159
<i>O. eichingeri</i>	CC	Sri Lanka	105407	eic_105407
<i>O. eichingeri</i>	CC	Sri Lanka	81803	eic_81803
<i>O. officinalis</i>	CC	Philippines	105085	off_105085
<i>O. officinalis</i>	CC	PNG	106519	off_106519
<i>O. officinalis</i>	CC	China	7904	off_7904
<i>O. rhizomatis</i>	CC	Sri Lanka	103410	rhi_103410
<i>O. rhizomatis</i>	CC	Sri Lanka	105448	rhi_105448
<i>O. malampuzhaensis</i>	BBCC	India	105223	MAL_105223
<i>O. malampuzhaensis</i>	BBCC	India	80764	MAL_80764
<i>O. minuta</i>	BBCC	Philippines	101082	MIN_101082
<i>O. minuta</i>	BBCC	Philippines	103874	MIN_103874
<i>O. punctata</i>	BBCC	Ghana	100937	PUN_100937
<i>O. punctata</i>	BBCC	Nigeria	104059	PUN_104059
<i>O. punctata</i>	BBCC	Zaire	105137	PUN_105137
<i>O. punctata</i>	BBCC	Kenya	105158	PUN_105158
* <i>O. punctata</i>	BBCC	Uganda	105160	*PUN_105160
* <i>O. punctata</i>	BBCC	Uganda	105182	*PUN_105182
<i>O. granulata</i>	GG	Vietnam	106469	<i>O. granulata</i>

**Table 1. Comprehensive list of the samples used in the present study including species name, genome type, source, abbreviation in figures, and accession numbers in the International Rice Research Institute (IRRI).** All accessions were obtained from leaf materials or seeds provided by IRRI at Los Banos, Philippines, except for 7904, which was collected by the authors. Two accessions marked with asterisks were originally labeled as *O. eichingeri* in the Germplasm Resource Center of IRRI but confirmed to be tetraploid *O. punctata*<sup>27,29,30</sup>.

all B-, C- and BC-genome species<sup>28,29</sup>. However, they were unable to obtain a fully resolved phylogeny due to limited genetic markers, and failed to identify the parental donors for the tetraploid species using only nuclear markers. In addition, the origin time of BBCC tetraploids in *Oryza* has been largely unknown with only *O. minuta* being dated previously<sup>12,13</sup>. Here, we sampled multiple populations from all the species involving the B-, C- and BC-genome types. Based on sequences of biparentally inherited nuclear genes and maternally inherited chloroplast regions, in conjunction with phylogenetic analyses and relaxed molecular dating, we fully resolved the phylogenetic relationships of these species, and particularly, determined the parental donors of all the BC-genome tetraploid species with their divergence times estimated. Specifically, we asked: (1) whether a single origin or multiple origins occurred for the BC-genome tetraploid species, (2) which diploid species were involved in the formation of the tetraploid species, and how many times did each contribute its genome, (3) when these allotetraploids originated. By illustrating the origin pattern of the BC-genome tetraploids in the rice genus, these investigations not only improve our understanding of evolutionary patterns of allopolyploid formation in the model system but also lay important foundation for utilization the wild rice germplasm in rice breeding and genetic improvement.

## Materials and Methods

**Plant Material and DNA sequencing.** Twenty-two accessions representing all species consisting of the B- and C-genome types were sampled, including four diploid species (one B-genome and three C-genome species), and three tetraploid species with the BC-genome type. Note, two tetraploid *O. punctata* accessions were originally labeled as *O. eichingeri* in the germplasm bank but confirmed to be tetraploid *O. punctata*<sup>27,29,30</sup>. One accession of *Oryza granulata* with the G genome was used as an outgroup. Information on all the materials is listed in Table 1. Total DNA was isolated from fresh or silica-gel dried leaves using the cetyltrimethylammonium bromide method<sup>31</sup>.

We sequenced fragments of four single-copy nuclear genes that previously used in our phylogenetic studies of rice tribe Oryzeae, i.e., alcohol dehydrogenase-1 (*Adh1*), alcohol dehydrogenase-2 (*Adh2*), *leafy hull sterile 1 (LHS1)*, and heterotrimeric G protein (*GPA1*)<sup>19,32</sup>. Primers specific to these loci were reported in previous studies<sup>10,32</sup>. In the case when we failed to obtain the two homoeologs of the tetraploids, additional homoeolog-specific primers were designed. In addition, we chose to sequence three fast-evolving regions of the chloroplast genome based on Tang *et al.* (2010), including one chloroplast gene (*matK*) and two intergenic spacers (*rps16-trnQ* and *trnT-trnD*). All the primers and their sequences are listed in Supplementary Table S1. PCR amplifications were carried out by standard methods. The amplified PCR products were sequenced directly for all chloroplast fragments and for nuclear genes in case of diploid individuals. For all accessions of tetraploids, cleaned PCR products were cloned into pGEM T-easy vectors (Promega, Madison, WI, USA) and 12 to 20 clones per accession were selected for sequencing. Two types of clones (putative homoeologs) could be visually identified for each nuclear gene by using phylogenetic analyses of accession-specific clone sequences. A consensus sequence with multiple clones for each sequence type was used in the following phylogenetic analyses to minimize the effect of PCR errors. Sequencing was carried out on an ABI 3730 automated DNA sequencer (Applied Biosystems, Foster City, CA, USA). All DNA sequences used for this study have been deposited in the GenBank database under accession numbers KP121693 - KP121896, and the sequence alignments have been uploaded to TreeBASE (Accession URL: <http://purl.org/phylo/treebase/phylows/study/TB2:S18041>).

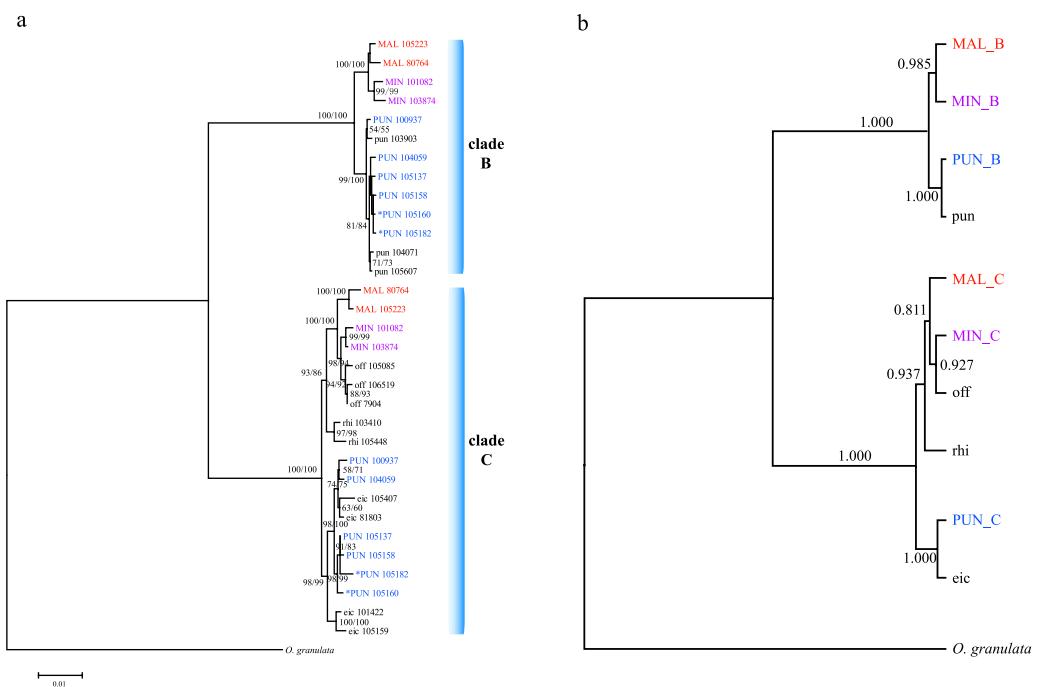
**Phylogenetic analyses.** Low-copy nuclear genes in combination with chloroplast sequences have proven to be a very effective way to address allopolyploidization event at the species level<sup>10,21,33</sup>. Because the homoeologous sequences of nuclear loci in allopolyploids are contributed by both of the diploid parents while chloroplast genome is maternally inherited in most angiosperms including species in the rice genus, the combined analyses of biparentally inherited nuclear genes and maternally inherited chloroplast regions enable us to identify parental donors of allotetraploids<sup>10,21,33</sup>. In this study, both chloroplast and nuclear sequences were aligned using MUSCLE<sup>34</sup> and then manually adjusted, with the regions of ambiguous alignment excluded. Prior to phylogenetic analyses, nucleotide frequencies were assessed for deviation from stationarity with the Chi-square test and no significant heterogeneity of base frequencies ( $P = 1.00$ ) was detected in our data. To identify potential intragenic recombinants, the RDP program were used to examine the alignments, with six recombination detection methods (RDP, GENECONV, Chimaera, MaxChi, BootScan and SiScan) implemented and the default settings used<sup>35</sup>. The recombinant sequences identified by RDP program were excluded from the phylogenetic analyses.

Phylogenetic trees were reconstructed using the maximum parsimony (MP) and maximum likelihood (ML) methods with PAUP\* 4.0b10<sup>36</sup>. For MP analysis, tree searches were performed using heuristic searches with tree bisection and reconnection (TBR) branch swapping and 1000 replicates of random addition sequence, with Multrees option on. Statistical reliability of topology was estimated by bootstrap analyses with 1000 replicates. For ML analysis, the best-fit model of DNA substitution was selected by the Akaike information criterion (AIC) in jModelTest 2<sup>37</sup>. Tree searches were performed using the heuristic algorithm and support for clades was assessed by bootstrap analyses with 500 replicates. The three chloroplast regions were concatenated for the phylogenetic analysis since they are genetically linked and form a single historical and phylogenetic unit. For the four nuclear loci, phylogenetic analyses were conducted on individual genes separately. After evaluating single gene trees, two distinct sequence types corresponding to two homoeologs could be identified for each tetraploid. Then, homoeologous sequences of four loci were concatenated and analyzed using ML and MP methods.

Besides the analyses of single-gene and concatenated sequences, we inferred the species trees using the four nuclear data sets using \*BEAST<sup>38</sup>, which is based on coalescent models. The 22 accessions of ingroup, including two homoeologues for each allotetraploids, were grouped into 10 operational taxonomic units (OTUs). The MCMC runs were set to 100 million generations, taking samples every 10000 generations, and convergence of MCMC chain was checked by running at least two independent analyses and by Tracer v1.4.

**Estimation of divergence time.** We estimate the divergence times in the relaxed-clock framework using two commonly used Bayesian Markov chain Monte Carlo (MCMC) programs, BEAST v1.7.0<sup>39</sup> and MCMCTREE in PAML v4.8<sup>40</sup>. These two relaxed-clock methods can account for the rate heterogeneity across lineages and accommodate multiple calibrations. Moreover, they can incorporate multiple loci into one analysis and deal with the different rates among loci appropriately.

In BEAST analyses, we used a relaxed clock model with the rate for each branch drawn from log-normal distribution. No topological constraints were used and we chose the coalescent as the tree prior since the phylogeny under scrutiny is a gene phylogeny rather than a species phylogeny<sup>39,41</sup>. The best-fit model of DNA substitution was used for each locus as selected by jModelTest 2<sup>37</sup>. Our previous estimates based on 106 single-copy nuclear genes<sup>18</sup> indicated that the G-genome species originated about 13.5 ~ 15.5 million years ago (Ma) and the C-genome species diverged from the lineage consisting of the A- and B-genome species around 5.5 ~ 6.5 Ma. Accordingly, we calibrated the divergence of G-genome with a normally distributed prior having a mean of 14.5 Ma and a standard deviation (SD) of 1.5 Ma, and the divergence of C-genome with a normally distributed prior having a mean of 6 Ma and a SD of 1 Ma.



**Figure 2.** (a) ML tree of the *Oryza* species with the B-, C- and BC-genome types based on the concatenated nuclear gene data. Two main clades represent homoeologous groups with the B and C genomes, respectively. Numbers besides nodes are ML/MP bootstrapping support over 50%. (b) Species tree inferred using \*BEAST with posterior probability indicated besides the branches. Sample abbreviation is indicated in Table 1, with tetraploids labeled with colored capital letters.

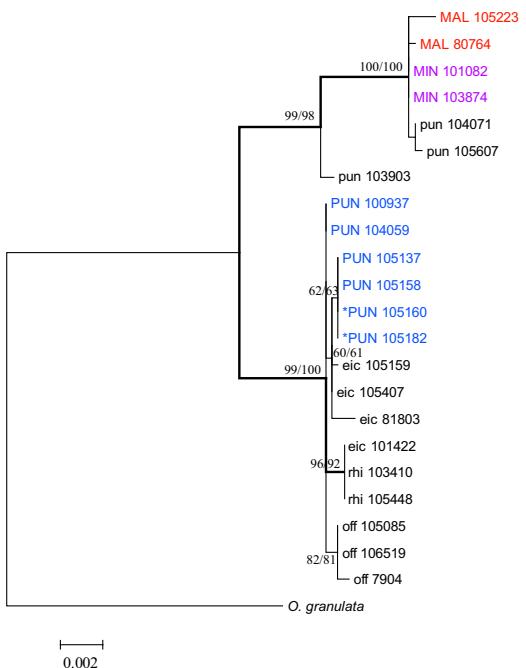
Final analyses consisted of 20 million generations of MCMC runs with sampling every 2000 generations and the initial 20% samples as burn-in.

In MCMCTREE analyses, soft bounds are imposed so that the minimum and maximum age constraint may be violated with a small probability (2.5%). We used the HKY85+ $\Gamma$  substitution model with different transition/transversion rate ration parameter ( $\kappa$ ) and different shape parameter ( $\alpha$ ) among loci, and the two calibration points were set similarly to the BEAST analyses. The ML tree obtained from the concatenated analyses was used as the input tree. A total of 100,000 generations was run with sampling every five generations after discarding the initial 10,000 samples as burn-in. In all dating analyses, each run was conducted at least twice to ensure consistency between different runs and convergence of the MCMC was evaluated by Tracer v1.4<sup>42</sup>.

## Results

Through PCR-amplifying, cloning and sequencing, we identified two types of distinct sequences at four single-copy nuclear loci for all allotetraploid accessions, corresponding to homoeologs of the B and C genomes, respectively. All four genes recovered similar gene trees in terms of topology, with two major clades (Supplementary Fig. S1). The first clade consisted of the accessions of all diploid B-genome species and the B-genome homoeologs of the tetraploid species, while the other clade included accessions of all diploid C-genome species and the C-genome homoeologs of the tetraploids (hereafter named as the clade B and clade C, respectively). The bootstrap values for the two major clades were all 100% except for *GPA1* in which the bootstrap value for clade C was over 90%.

Because the topology within the major clades was less resolved in individual gene trees due to insufficient informative sites, and given the substantial congruence among the four gene trees, we conducted a combined analysis based on the concatenated sequences from the four nuclear loci, with the variable sites being 16.1% and informative sites 7.3%. The resulting phylogenetic trees inferred from ML and MP methods were the same and well resolved (Fig. 2a). Within both major clades, two Asian tetraploids (*O. minuta* and *O. malampuzhaensis*) and the African tetraploid *O. punctata* were grouped into two clearly separate subclades with 93–100% ML bootstrap support. Within clade B, accessions from the two Asian tetraploids formed a monophyletic group with 100% support and those from African tetraploids grouped with the diploid *O. punctata* (BB) with 99–100% support. Within clade C, accessions from the two Asian tetraploids grouped with *O. officinalis* with 100% support, with this clade sister to *O. rhizomatis*, while accessions from African tetraploids grouped with *O. eichingeri* with 98–99% support. These results indicated that, of three C-genome diploid species, *O. officinalis* was most likely involved in the formation



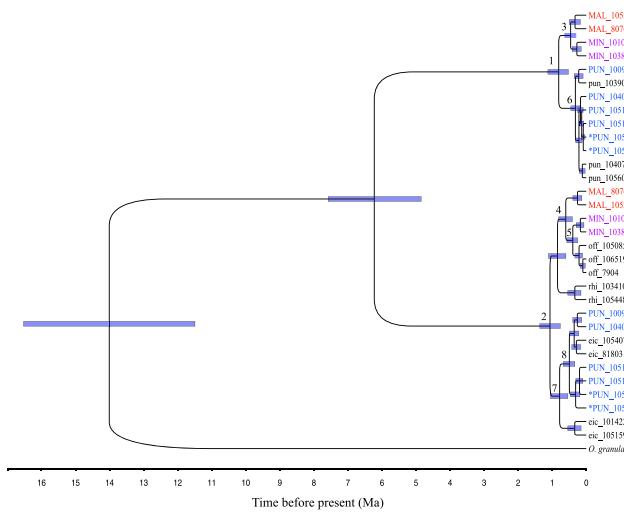
**Figure 3. ML tree based on the concatenated chloroplast sequences.** Numbers besides nodes are ML/MP bootstrapping supports over 50%. Branches having more than 90% ML and MP bootstrap supports are indicated by thickened lines.

of two Asian tetraploid species while *O. eichingeri* was most likely the C-genome donor of the African tetraploid species (Fig. 2a).

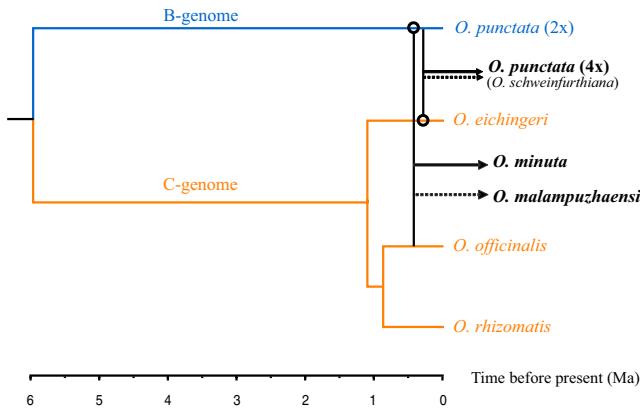
In addition to the individual gene analyses and concatenated analyses, we inferred the species trees using \*BEAST which utilized information contained in individual nuclear loci based on coalescent models (Fig. 2b). Each of the homoeologous genomes of each tetraploid species was grouped with its putative diploid progenitor and the overall topology corroborated the concatenated tree with good support. Based on these analyses, it could be inferred that diploid *O. punctata* and *O. officinalis* were most likely the parental progenitors of *O. minuta* and *O. malampuzhaensis*, whereas the diploid *O. punctata* and *O. eichingeri* were the progenitors of tetraploid *O. punctata*. Note that two tetraploid *O. punctata* accessions marked with asterisks in Fig. 2a and Supplementary Fig. S1 were always grouped with accessions from other tetraploid *O. punctata* in both concatenated and single-gene analyses, supporting the previous treatment of these two accessions as tetraploid *O. punctata* rather than *O. eichingeri*<sup>27,30</sup>.

The concatenated three chloroplast regions produced a haplotype dataset of 3713 bp in length, with 3.6% variable sites and 1.2% informative sites. The inferred phylogeny recovered a 100% supported monophyletic group consisting of haplotypes from tetraploid *O. minuta* and *O. malampuzhaensis* and diploid *O. punctata*, supporting the latter as the maternal donor of Asian tetraploids (Fig. 3). In the other clade, haplotypes from four out of six African tetraploid accessions grouped with diploid *O. eichingeri* haplotypes (with support of 60–61%), while another two were not resolved due to the lack of phylogenetic information. In combination with the results from the nuclear gene tree, this corroborated the hypothesis that *O. eichingeri* was most probably the maternal genome donor of the African tetraploids. Interestingly, alleles or haplotypes from individuals of tetraploid *O. punctata* did not form a monophyletic clade in all nuclear and chloroplast gene trees, implying recurrent origins of the tetraploid *O. punctata*.

We estimated the time of origin of BBCC tetraploids by utilizing relaxed clock methods based on sequences of nuclear markers. The Bayesian analyses using BEAST resulted in a completely congruent phylogeny to that from concatenated analyses. As shown in Fig. 4, nodes 6 and 8 are both potential nodes of origin of African tetraploids, with each being the divergence time of one of the homoeologs from their diploid progenitors respectively. Considering the standing variation of alleles in diploid progenitors, the youngest estimate, or the minimum distances between alleles of a tetraploid and that of its diploid progenitor are typically assumed to be the origin time of the tetraploid<sup>43</sup>. Therefore, the tetraploid *O. punctata* originated about 0.19–0.46 Ma (node 6) and *O. minuta* originated slightly earlier (about 0.25–0.57 Ma) (node 5). It is difficult to estimate the specific time of origin of *O. malampuzhaensis*, since all individuals did not closely group with any diploids. Because the C copy of *O. minuta* first grouped with its diploid donor rather than *O. malampuzhaensis*, it is most likely that the two Asian tetraploids originated independently. Therefore, we speculate that the origin of *O. malampuzhaensis* might be more recent than the date of the nodes 3 and 4, (later than 0.81 Ma). The estimates of the age of tetraploid origin using MCMCTREE resulted in similar results and the 95% HPD intervals of key nodes overlapped



**Figure 4. Chronogram of the *Oryza* species with the B-, C- and BC-genome types obtained based on four nuclear loci using relaxed clock implemented in BEAST.** Branch lengths indicate the posterior means of date estimates with the blue bars representing 95% highest posterior density (HPD) intervals for the divergence times estimates. Nodes of interest are numbered.



**Figure 5. Schematic of evolutionary relationships among the species with the B-, C-, and BC-genome types, with emphasis on the origins of the allotetraploid species.** Black lines represent the origins of the allotetraploids with the maternal donors indicated by circles. The solid and broken arrows indicate the confirmed and hypothetical allopolyploidization events, respectively.

considerably with those using BEAST (Supplementary Fig. S2 and Table S2). Given the estimates from the two relaxed clock methods, it appears that all BBCC species in *Oryza* originated within the last 1 million years.

## Discussion

The *Oryza* species with the B-, C- and BC-genome types have long been a subject of debate for both species delimitation and species relationships<sup>6,11,24,25</sup>. For example, both diploid and tetraploid forms of *O. punctata* were found in Africa, with distinct morphological differences between them<sup>44–46</sup>. Therefore, some taxonomists used *Oryza schweinfurthiana* Prod. to refer to the tetraploid *O. punctata*<sup>47</sup>. However, this treatment was not accepted by other authors. Similarly, the tetraploid *O. eichingeri* has been reported before<sup>44</sup>, but was considered to be a misidentification of tetraploid *O. punctata* when its morphology<sup>25,46</sup> and genetic markers<sup>27,28,30</sup> were analyzed. Another confounding issue is the identities of the exact parental donors of each allotetraploid, which have been inconsistent among previous studies. For example, in the case of *O. malampuzhaensis*, *O. officinalis* was indicated as its C-genome progenitor based on a study using AFLP markers<sup>27</sup> while either *O. rhizomatis* or *O. eichingeri* was regarded as the C-genome donors based on nuclear RFLP and SSR markers<sup>23,29</sup>. In the present study, we generated a well-resolved phylogenetic tree of all the diploid and tetraploid species with the B- and C-genome types based on sequences from biparentally inherited nuclear genes and maternally inherited chloroplast fragments. The main findings of this study are illustrated by Fig. 5 and have several important implications. First, we

demonstrated that at least three independent allotetraploid events gave rise to three extant BC-tetraploid species in this species complex, and the tetraploid *O. punctata* (*Oryza schweinfurthiana*) appears to have originated multiple times. Multiple origins were detected despite the fact that only a small number of accessions was sampled (Fig. 2). Interestingly, the single B-genome diploid species, *O. punctata*, served as the maternal progenitor for two Asian tetraploids (*O. minuta* and *O. malampuzhaensis*) and as the paternal progenitor for the African tetraploid (*O. punctata* or *O. schweinfurthiana*) (Fig. 5). Similar pattern of multiple origins or reticulate evolution of polyploidy species have been well documented in many model systems such as *Tragopogon*, *Mimulus*, *Rubus*, and *Glycine* (reviewed by Soltis *et al.*, 2014). With enormous genetic, genomic and functional resources available<sup>7,9</sup>, the *Oryza* species can become an additional model for investigations on polyploid speciation and evolution.

Second, we clearly identified the most likely parental diploid species that contributed to the formation of two allotetraploid species. Specifically, *O. minuta* appears to have originated from allopolyploidization of *O. punctata* (maternal) and *O. officinalis* (paternal); whereas the tetraploid *O. punctata* (*O. schweinfurthiana*) was derived from *O. eichingeri* (maternal) and *O. punctata* (paternal) (Fig. 5). Although it is evident that the maternal and paternal donors of *O. malampuzhaensis* are the B-genome and C-genome species, respectively, the exact diploid C-genome species involving the tetraploid formation remains to be identified (Figs 2 and 5). Of three C-genome diploid species, *O. rhizomatis* was not involved in the formation of extant polyploids. These findings raise some interesting questions regarding where and when the allopolyploidization events have taken place for the two Asian tetraploids, given that the distribution of extant diploid *O. punctata* is confined to Africa. Did the polyploidization occur in Africa and then disperse to Asia? Alternatively, the events could have taken place in Asia with the diploid B-genome donor later becoming extinct. The situation is reminiscent of *Gossypium*, where the New World allopolyploid *G. hirsutum* combines a New World D-genome with an exclusively African A-genome<sup>48</sup>. The clarification of the reticulate relationships of the *Oryza* species with the B- and C-genome types will inevitably facilitate the investigations of polyploid evolution using this group of species as working system given the fact that the *Oryza* polyploids have been increasingly used for studying polyploidization and genome evolution<sup>7,12,13</sup>.

Third, our phylogenetic results provided valuable information to guide taxonomic treatment of the tetraploids. Unlike *O. minuta* and *O. malampuzhaensis* that have been proved good species<sup>27,28,49,50</sup>, the debate over the delimitation of the African BC-genome tetraploids has continued for decades<sup>20,46</sup>. Earlier studies have claimed that both diploid and tetraploid forms were found for *O. punctata* and *O. eichingeri*<sup>44,46</sup> and the tetraploid *O. punctata* was distributed as widely as its diploid form<sup>28</sup>. Because the diploid and tetraploid forms of *O. punctata* differed by annual and perennial habits and in many morphological characteristics<sup>45</sup>, the name “*Oryza schweinfurthiana* Prod.” has been used by some taxonomists to refer to the tetraploid *O. punctata*<sup>47</sup>. Our study revealed that the tetraploid *O. punctata* arose by hybridization and polyploidization of two distinct diploid species (*O. punctata* and *O. eichingeri*); along with distinct gross morphology and life history, this strongly supports the species status of the tetraploid form of *O. punctata*, i.e., *O. schweinfurthiana*. With regard to the tetraploid *O. eichingeri*, Tateoka (1965) pointed out the possibility of misidentification. Several investigations based on molecular markers have also demonstrated that some of the materials designated as tetraploid *O. eichingeri* were instead tetraploid *O. punctata*<sup>27,29,30</sup>. In this study, homoeologues from two tetraploid *O. eichingeri* accessions grouped with tetraploid *O. punctata* rather than forming a monophyletic clade in all cases (Figs 2 and 3, Supplementary Fig. S1), providing further evidence that *O. eichingeri* is a diploid species and the materials previously identified as tetraploid *O. eichingeri* are most likely to be the tetraploid *O. punctata*.

Finally, through relaxed molecular clock analyses, we estimated that the BC-genome tetraploid species arose around 0.2–0.6 Ma (Figs 4 and 5), largely consistent with previous studies<sup>12,13</sup>. These estimates suggest that the BC-genome tetraploid species originated recently relative to the divergence of the A, B, and C genomes, around 5.5~6.5 Ma<sup>18</sup>. It is interesting to ask what factors may have contributed to the formation of the BC-genome tetraploids given the fact that no any polyploid species have been found among the eight A-genome species. It is well recognized that evolution of earth’s terrestrial biota is profoundly influenced by the global climate changes<sup>51,52</sup>. Evidence showed that there was a period of dramatic climate change within the series of ice ages from 0.9 Myr to the last ice age, during which severe climate oscillations happened followed a 100-Kyr cycle and average temperature would rise rapidly by about 7°C over just decades<sup>53</sup>. The origin of the BC-genome tetraploids occurred coincidentally with this period, implying the potential impact of climate change on the formation and dispersal of these tetraploids. Recent studies suggest that the polyploidy events may have generated sufficient novelty that increased tolerance toward the drastically changing global environment and thus enhanced the adaptability of species<sup>2,54</sup>. This is consistent with the observed correlation between polyploidy and invasiveness<sup>55</sup>.

It should be noted, nevertheless, that the materials we used in this study are limited in term of sample sizes per species, because of the difficulty in obtaining samples for this group of species. Therefore, the origin times of polyploids we estimated here is the maximum, and more recent dates could have been obtained by additional sampling<sup>43</sup>. Such sampling schemes also preclude an in-depth investigation of the potential diploid progenitors of some tetraploids (e.g., *O. malampuzhaensis*) and exact place of polyploid origin and their subsequent dispersals, as well as the potential populations that contributed to the multiple formations of specific tetraploid species (e.g., the tetraploid *O. punctata* or *O. schweinfurthiana*).

Further studies based on more extensive sampling across the entire distribution areas of the BC-genome species are needed to elucidate the evolutionary scenario of this species complex.

## References

- Doyle, J. J. *et al.* Evolutionary genetics of genome merger and doubling in plants. *Annu Rev Genet* **42**, 443–461 (2008).
- Soltis, D. E., Visger, C. J. & Soltis, P. S. The polyploidy revolution then...and now: Stebbins revisited. *Am J Bot* **101**, 1057–1078 (2014).
- Freeling, M. & Thomas, B. C. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res* **16**, 805–814 (2006).
- Renny-Byfield, S. & Wendel, J. F. Doubling down on genomes: polyploidy and crop plants. *Am J Bot* **101**, 1711–1725 (2014).
- Sang, T. & Ge, S. Genetics and phylogenetics of rice domestication. *Curr Opin Genet Dev* **17**, 533–538 (2007).
- Vaughan, D. A. The genus *Oryza* L.: current status of taxonomy. *IRRI Res Pap Ser* **138**, 1–21 (1989).
- Jacquemin, J., Bhatia, D., Singh, K. & Wing, R. A. The International *Oryza* Map Alignment Project: development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. *Curr Opin Plant Biol* **16**, 147–156 (2013).
- Sang, T. & Ge, S. Understanding rice domestication and implications for cultivar improvement. *Curr Opin Plant Biol* **16**, 139–146 (2013).
- Zhang, Q. & Wing, R. Genome studies and molecular genetics: understanding the functional genome based on the rice model. *Curr Opin Plant Biol* **16**, 129–132 (2013).
- Ge, S., Sang, T., Lu, B. R. & Hong, D. Y. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc Natl Acad Sci USA* **96**, 14400–14405 (1999).
- Bao, Y. & Ge, S. Historical retrospect and the perplexity on the studies of the *Oryza* polyploids. *J Syst Evol* **46**, 3–12 (2008).
- Ammiraju, J. S. S. *et al.* Spatio-temporal patterns of genome evolution in allotetraploid species of the genus *Oryza*. *Plant J* **63**, 430–442 (2010).
- Lu, F. *et al.* Comparative sequence analysis of MONOCULM1-orthologous regions in 14 *Oryza* genomes. *Proc Natl Acad Sci USA* **106**, 2071–2076 (2009).
- Sui, Y., Li, B., Shi, J. & Chen, M. Genomic, regulatory and epigenetic mechanisms underlying duplicated gene evolution in the natural allotetraploid *Oryza minuta*. *BMC genomics* **15**, 11 (2014).
- Ali, M. L., Sanchez, P. L., Yu, S. B., Lorieux, M. & Eizenga, G. C. Chromosome segment substitution lines: a powerful tool for the introgression of valuable genes from *Oryza* wild species into cultivated rice (*O. sativa*). *Rice* **3**, 218–234 (2010).
- Jena, K. K. The species of the genus *Oryza* and transfer of useful genes from wild species into cultivated rice, *O. sativa*. *Breed Sci* **60**, 518–523 (2010).
- Zhu, Q. & Ge, S. Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. *New Phytol* **167**, 249–265 (2005).
- Zou, X. H., Yang, Z., Doyle, J. J. & Ge, S. Multilocus estimation of divergence times and ancestral effective population sizes of *Oryza* species and implications for the rapid diversification of the genus. *New Phytol* **198**, 1155–1164 (2013).
- Zou, X. H. *et al.* Analysis of 142 genes resolves the rapid diversification of the rice genus. *Genome Biol* **9**, R49 (2008).
- Vaughan, D. A. *The Wild Relatives of Rice: A Genetic Resources Handbook*. (International Rice Research Institute, 1994).
- Bao, Y. & Ge, S. Origin and phylogeny of *Oryza* species with the CD genome based on multiple-gene sequence data. *Plant Syst Evol* **249**, 55–66 (2004).
- Dally, A. M. & Second, G. Chloroplast DNA diversity in wild and cultivated species of rice (Genus *Oryza*, section *Oryza*). Cladistic-mutation and genetic-distance analysis. *Theor Appl Genet* **80**, 209–222 (1990).
- Wang, Z. Y., Second, G. & Tanksley, S. D. Polymorphism and phylogenetic-relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. *Theor Appl Genet* **83**, 565–581 (1992).
- Lu, B. R., Ge, S., Sang, T., Chen, J. K. & Hong, D. Y. The current taxonomy and perplexity of the genus *Oryza* (Poaceae). *Acta Phytotax Sinica* **39**, 373–388 (2001).
- Nayar, N. M. Origin and cytogenetics of rice. *Adv Genet* **17**, 153–292 (1973).
- Zhang, L. B. & Ge, S. Multilocus analysis of nucleotide variation and speciation in *Oryza officinalis* and its close relatives. *Mol Biol Evol* **24**, 769–783 (2007).
- Aggarwal, R. K., Brar, D. S., Nandi, S., Huang, N. & Khush, G. S. Phylogenetic relationships among *Oryza* species revealed by AFLP markers. *Theor Appl Genet* **98**, 1320–1328 (1999).
- Bao, Y., Lu, B. R. & Ge, S. Identification of genomic constitutions of *Oryza* species with the B and C genomes by the PCR-RFLP method. *Genet Resour Crop Evol* **52**, 69–76 (2005).
- Bao, Y., Zhou, H. F., Hong, D. Y. & Ge, S. Genetic diversity and evolutionary relationships of *Oryza* species with the B- and C-genomes as revealed by SSR markers. *J Plant Biol* **49**, 339–347 (2006).
- Buso, G. S., Rangel, P. H. & Ferreira, M. E. Analysis of random and specific sequences of nuclear and cytoplasmic DNA in diploid and tetraploid American wild rice species (*Oryza* spp.). *Genome* **44**, 476–494 (2001).
- Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bulletin* **19**, 11–15 (1987).
- Tang, L. *et al.* Phylogeny and biogeography of the rice tribe (Oryzeae): evidence from combined analysis of 20 chloroplast fragments. *Mol Phylogenet Evol* **54**, 266–277 (2010).
- Doyle, J. J., Doyle, J. L., Brown, A. H. & Grace, J. P. Multiple origins of polyploids in the *Glycine tabacina* complex inferred from chloroplast DNA polymorphism. *Proc Natl Acad Sci USA* **87**, 714–717 (1990).
- Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).
- Martin, D. P. *et al.* RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**, 2462–2463 (2010).
- Swofford, D. L. *PAUP\*: Phylogenetic Analysis Using Parsimony (\*and Other Methods)*. (Sinauer Associates, 2002).
- Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**, 772 (2012).
- Heled, J. & Drummond, A. J. Bayesian inference of species trees from multilocus data. *Mol Biol Evol* **27**, 570–580 (2010).
- Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**, 1969–1973 (2012).
- Yang, Z. H. & Rannala, B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol* **23**, 212–226 (2006).
- Egan, A. N. & Doyle, J. A comparison of global, gene-specific, and relaxed clock methods in a comparative genomics framework: dating the polyploid history of soybean (*Glycine max*). *Syst Biol* **59**, 534–547 (2010).
- Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7** (2007).
- Doyle, J. J. & Egan, A. N. Dating the origins of polyploidy events. *New Phytol* **186**, 73–85 (2010).

44. Hu, C. H. Cytogenetic studies of *Oryza officinalis* complex. III. The genomic constitution of *O. punctata* and *O. eichingeri*. *Cytologia* **35**, 304–318 (1970).
45. Sano, Y. Adaptive strategies compared between the diploid and tetraploid forms of *Oryza punctata*. *Bot Mag Tokyo* **93**, 171–180 (1980).
46. Tateoka, T. Taxonomy and chromosome numbers of African representatives of the *Oryza officinalis* complex. *Bot Mag Tokyo* **78**, 198–201 (1965).
47. Sharma, S. D. & Sampath, S. In *Rice Research in India*. (ed P. L. Jaiswal), 21–43 (Indian Council of Agricultural Research, 1985).
48. Wendel, J. F. & Cronn, R. C. Polyploidy and the evolutionary history of cotton. *Adv Agron* **78**, 139–186 (2003).
49. Li, C. B., Zhang, D. M., Ge, S., Lu, B. R. & Hong, D. Y. Identification of genome constitution of *Oryza malampuzhaensis*, *O. minuta* and *O. punctata* by multicolor genomic *in situ* hybridization. *Theor Appl Genet* **103**, 204–211 (2001).
50. Vaughan, D. A. & Morishima, H. In *Rice: origin, history, technology, and production*. (eds C. W. Smith & R. H. Dilday) (John Wiley & Sons, Incorporation, 2003).
51. Stromberg, C. A. E. Evolution of grasses and grassland ecosystems. *Annu Rev Earth Planet Sci* **39**, 517–544 (2011).
52. Wing, S. L. et al. Transient floral change and rapid global warming at the Paleocene-Eocene boundary. *Science* **310**, 993–996 (2005).
53. Hewitt, G. The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913 (2000).
54. Vanneste, K., Baele, G., Maere, S. & Van de Peer, Y. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. *Genome Res* **24**, 1334–1347 (2014).
55. te Beest, M. et al. The more the better? The role of polyploidy in facilitating plant invasions. *Ann Bot* **109**, 19–45 (2012).

## Acknowledgements

We thank Ying Bao, Ang Li and other members in Ge's group for technical assistance. We are grateful to the International Rice Research Institute (Los Banos, Philippines) for providing seed and leaf samples. This work was supported by the National Natural Science Foundation of China (31300198; 30990240).

## Author Contributions

S.G. and T.S. designed the study. S.G. and X.H.Z. performed the research. X.H.Z. and S.G. analyzed the data. X.H.Z., S.G. and J.J.D. wrote the paper. Y.S.D., L.T., X.W.X. and X.H.Z. performed the experiments.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Zou, X.-H. et al. Multiple origins of BBCC allopolyploid species in the rice genus (*Oryza*). *Sci. Rep.* **5**, 14876; doi: 10.1038/srep14876 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>