

Molecular phylogeny and divergence of the rice tribe Oryzeae, with special reference to the origin of the genus *Oryza*

Song Ge, Ya-long Guo, and Qi-hui Zhu

The rice tribe Oryzeae consists of 12 genera and more than 70 species distributed in tropical and temperate regions worldwide (Tzvelev 1989, Vaughan 1994). Species in the genus *Oryza* and other genera closely related to *Oryza* have been extensively studied either because of their agronomically useful traits in rice genetic improvement (wild species in *Oryza* and *Porteresia*) or because of their economic value as part of cuisine (*Zizania*) and forage (*Leersia*) (Vaughan and Morishima 2003). However, phylogenetic relationships among genera in this tribe have not been well studied, and the circumscription and taxonomic position of some genera have remained controversial for decades. In addition, the origin and diversification of this tribe, in particular the origin of the genus *Oryza* and its divergence, remain largely unclear. Recently, sequences of low-copy nuclear genes have been successfully used for addressing phylogenetic questions and, in combination with chloroplast DNA fragments, have provided especially powerful markers in terms of phylogenetic reconstruction and biogeographic inference (Sang 2002).

The alcohol dehydrogenase (*Adh*) gene is the most widely used low-copy nuclear gene, whereas the chloroplast *matK* gene, *trnL* intron, and *trnL-trnF* spacer, as well as mitochondrial *nad1* intron, are widely used in phylogenetic studies. Nuclear *GPA1* that encodes a G protein α subunit is also a source of choice because it is a single copy in higher plants and well characterized in function and structure. To better understand the evolutionary and biogeographic history of the rice tribe, we sequenced two chloroplast (*matK* and *trnL-trnF*) fragments and one mitochondrial (*nad1* intron) fragment and portions of two nuclear genes (*Adh2* and *GPA1*) from 35 species representing 12 genera in the tribe, as well as *Ehrharta* and *Phyllostachys* (subfamily Bambusoideae) as the outgroups. In addition, to clarify the phylogenetic relationship among the A-genome species, which include cultivated rice (*O. sativa*), we have chosen to sequence introns of three anonymous nuclear single-copy genes (*OsRFCD001283*, *OsRFCD017357*, and *OsRFCD009971* on chromosomes 1, 4, and 2, respectively) that evolve much faster than commonly used ITS and cpDNA fragments. Based on sequences of these introns, we reconstructed the phylogeny and dated the origin and divergence of the A-genome species.

Molecular phylogeny based on multiple sequences from three genomes

Phylogenetic analyses of the aligned sequences of the above fragments from three genomes were conducted using maxi-

mum parsimony (MP), neighbor-joining (NJ), and Bayesian (BI) approaches. A comparison of the phylogenies inferred from two nuclear genes shows that they are largely congruent, with only one area of disagreement involving the placement of the monotypic genus *Hygroryza* (see below). The partition-homogeneity test (PHT) indicated that two data sets (*Adh2* and *GPA1*) were statistically incongruent ($P < 0.01$) when all the species were included, whereas the topological incongruence was not significant ($P = 0.12$) when the monotypic genus *Hygroryza* was excluded. Therefore, phylogenetic analysis was conducted based on the combined data sets, excluding *Hygroryza*, and the strict consensus tree using the MP method is presented in Figure 1. Essentially, the NJ and BI methods generate topologies almost identical to that of MP, except for slightly different statistical support for a few clades. As Figure 1 shows, the species of the rice tribe Oryzeae form a strongly supported monophyly and the tribe diverged early into two major clades with high bootstrap support (over 95%). These two clades correspond to the traditionally recognized subtribes Oryzinae and Zizaniinae (Pyrah 1969). One clade includes three genera, *Leersia*, *Oryza*, and *Porteresia*, with *Leersia* as the basal lineage. It is obvious that *Porteresia coarctata* is nested into *Oryza* species, suggesting that it should be treated as a member of the genus *Oryza* rather than as a separate monotypic genus (Ge et al 2002). The other clade consists of three highly supported monophyletic groups involving the remaining genera, except *Hygroryza*. The first group includes eastern Asian *Chikusichloa*, Australian monotypic *Potamophila*, and African *Prospytochloa*, whereas the second group contains the American *Zizaniopsis* and *Luziola*, which is a sister to the group including the American monotypic *Rhynchoryza* and the eastern Asian–North American disjunct *Zizania* (Fig. 1). The combined data sets of *matK*, *trnL-trnF*, and *nad1* provide similar topology but the resolution is slightly lower (not shown).

It is noteworthy that *Hygroryza aristata* forms a clade with the *Leersia/Oryza* group on the *Adh2* tree but clusters with the *Rhynchoryza/Zizania + Luziola/Zizaniopsis* group on the *GPA1* tree (Fig. 1). The combined *matK/trnL-F/nad1* phylogeny presents a topology similar to that of *GPA1* and the previous *matK* phylogeny (Ge et al 2002). However, because this monotypic genus is a diploid ($2n=24$) endemic to southern and southeastern Asia, the incongruence does not likely arise from hybridization and subsequent polyploidization, which may result in reticulate evolution. Whether this incongruence reflects “real” underlying biological phenomena or

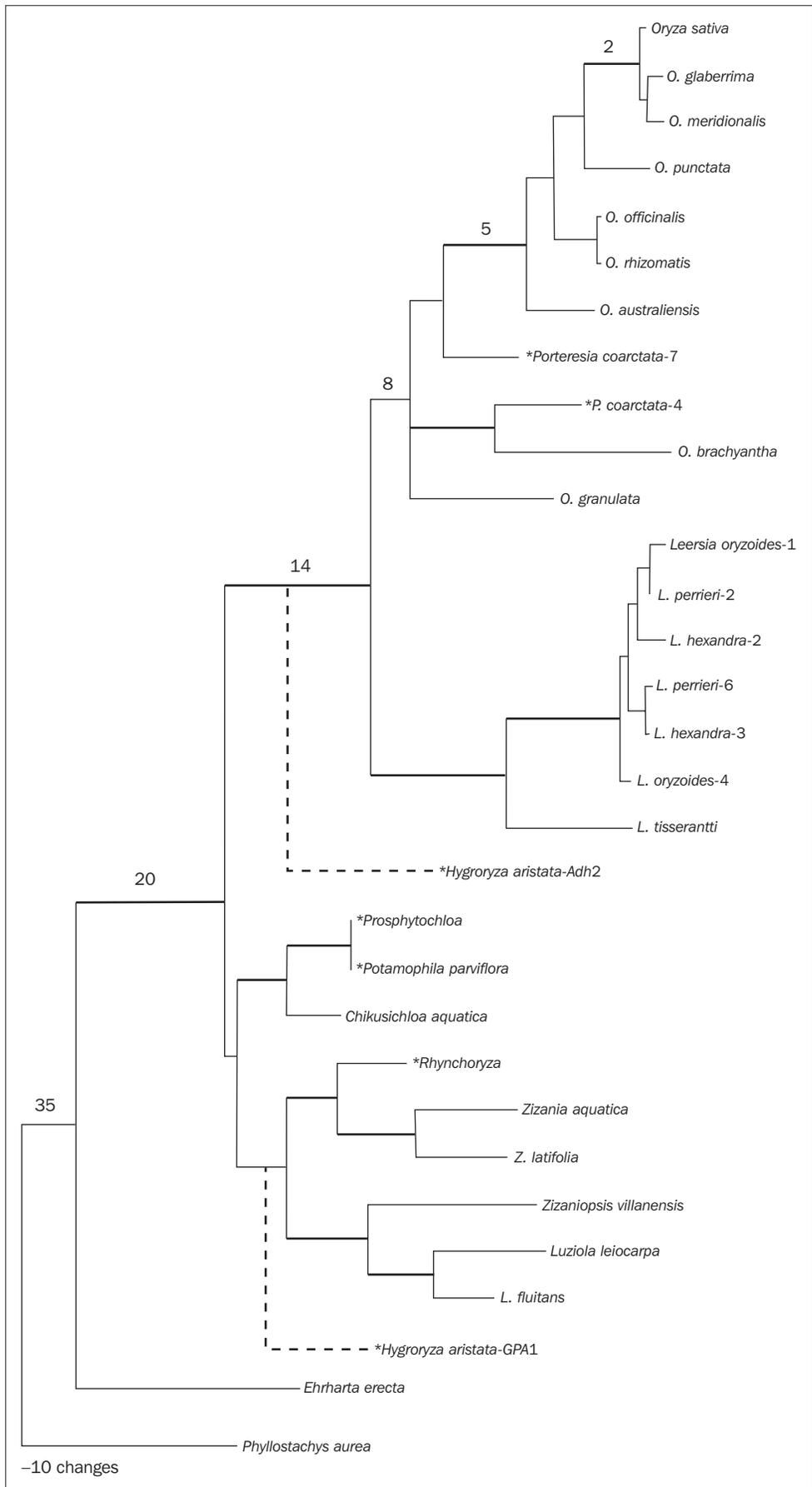


Fig. 1. Strict consensus tree of the rice tribe Oryzaceae generated from the combined *Adh2* and *GPA1* sequences. The thick lines indicate the nodes with the bootstrap support and Bayesian posterior probability over 95%, while the thin lines indicate the nodes with the bootstrap support and Bayesian posterior probability from 75% to 94%. The numbers on the node represent divergence times (approximate) in million years ago (MYA) as estimated by molecular clock approaches. Broken lines indicate the placement of *Hygroryza* on *Adh2* and *GPA1* trees, respectively. The numbers following the species name represent the clone sequenced. The asterisks indicate the monotypic genera.

just “spurious” insufficient data or some other unknown artifact is unclear and needs to be further explored.

Divergence of the rice tribe and the origin of *Oryza*

Although the phylogeny and divergence time of Poaceae have been revealed by molecular approaches and estimated based on fossil evidence (GPWG 2001, Gaut 2002), little in the fossil record reflects upon the origin or diversification of the genus *Oryza*. A few studies have been made to estimate the divergence time of *Oryza* and its related genera (Second 1991, Vaughan and Morishima 2003). However, estimation of the possible dates of divergent events related to the evolution of *Oryza* was largely based on isozyme studies (Second 1991) and many contradictions exist regarding the divergence event proposed (see Vaughan and Morishima 2003). Despite several limitations to the use of clock-based sequence data (Gaut 2002), they are useful for estimating divergence time when the clock can be calibrated with some confidence, in particular when fossil evidence is not available or is inadequate.

To test the molecular clock hypothesis in Oryzaceae, we performed relative rate tests for all the data sets. Results show that no rate heterogeneity exists at synonymous sites for both *Adh2* and *matK* data sets and so the molecular clock hypothesis cannot be rejected for the synonymous evolution at both loci. Using the synonymous molecular clocks of *matK* and *Adh2* and assuming that maize and rice diverged 50 million years ago (MYA) (Gaut 2002), we calculated approximate divergence times for various lineages within the tribe and within *Oryza* (Fig. 1). The estimates suggest that the rice tribe originated roughly 35 MYA, during approximately the transition from the Eocene to the Oligocene. Within the tribe, *Oryza* and *Leersia* separated from the rest of the tribe approximately 20 MYA, and from each other 14 MYA. Within *Oryza*, the age of the deepest split between the most basal G genome and remaining genomes was estimated at about 8 MYA.

Phylogeny, origin, and divergence of A-genome species in *Oryza*

Evidence showed that the A-genome group that consists of eight diploid species is one of the most recently diverged lineages within the genus *Oryza* (Ge et al 2001). Their evolutionary relationships have long been controversial and are still not well studied at the sequence level. This is mainly because the commonly used sequences such as ITS and cpDNA fragments lack sufficient resolution in groups that have undergone rapid or recent radiations. Accordingly, we sequenced introns of three nuclear genes located on different chromosomes from multiple accessions of eight A-genome species. As expected, the intron sequences provide much higher (more than two times higher) informative characters than those of ITS (data not shown). Phylogenetic relationships and divergence times of the main lineages of the A-genome species were inferred based on these intron sequences using the similar approaches men-

tioned above. Figure 2 presents a phylogeny and divergence times among eight species, along with the B- (*O. punctata*) and E-genome (*O. australiensis*) species as the outgroups. All the species except *O. rufipogon* and *O. nivara* form a monophyletic group and the Australian endemic *O. meridionalis* is the earliest divergent lineage. However, accessions from *O. rufipogon* and *O. nivara* entirely mix together, supporting previous opinion that the two species should be treated as a single one. Moreover, two ecogeographic races of *O. sativa* (indica and japonica) form monophyletic groups along with some *O. rufipogon* and *O. nivara* accessions.

Sequence divergence is low in all pair-wise comparisons among A-genome species, suggesting that the group radiated relatively recently. Using the similar molecular clock approaches mentioned above, we estimate that the A-genome species began to diverge roughly 2 MYA. The Asian cultivated rice (*O. sativa*) diverged from the African cultivated rice (*O. glaberrima*) at about 0.6 MYA, whereas two races of *O. sativa* (indica and japonica) separated approximately 0.4 MYA (Fig. 2).

To put the divergence of the rice tribe and rice genus into a temporal framework raises an interesting question: How can the present distribution patterns in the rice tribe be reconciled with the apparent pantropical distribution for the tribe and many genera such as *Oryza* and *Leersia*? In particular, many closely related species (e.g., the A-genome species) are currently geographically isolated from one another by thousands of kilometers of open ocean. This implies that oceanic dispersal would contribute to the evolution and divergence of the genus *Oryza* and its tribe.

References

- Gaut BS. 2002. Evolutionary dynamics of grass genomes. *New Phytol.* 154:15-28.
- Ge S, Li A, Lu BR, Zhang SZ, Hong DY. 2002. A phylogeny of the rice tribe Oryzaceae (Poaceae) based on *matK* sequence data. *Am. J. Bot.* 89:1967-1972.
- Ge S, Sang T, Lu BR, Hong DY. 2001. Phylogeny of the genus *Oryza* as revealed by molecular approaches. In: Khush GS, Brar DS, Hardy B, editors. *Rice genetics IV*. Enfield, N.H. (USA): Science Publishers, Inc. and Los Baños (Philippines): International Rice Research Institute. p 89-105.
- GPWG (Grass Phylogeny Working Group). 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). *Ann. Missouri Bot. Garden* 88:373-457.
- Pyrah GL. 1969. Taxonomic and distributional studies in *Leersia* (Gramineae). *Iowa State J. Sci.* 44:215-270.
- Sang T. 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. *Crit. Rev. Biochem. Mol. Biol.* 37:121-147.
- Second G. 1991. Molecular markers in rice systematics and the evaluation of genetic resources. In: Bajaj YPS, editor. *Rice biotechnology in agriculture and forestry*. Vol. 14. Berlin (Germany): Springer-Verlag. p 468-494.
- Tzvelev NN. 1989. The system of grasses (Poaceae) and their evolution. *Bot. Rev.* 55:141-204.

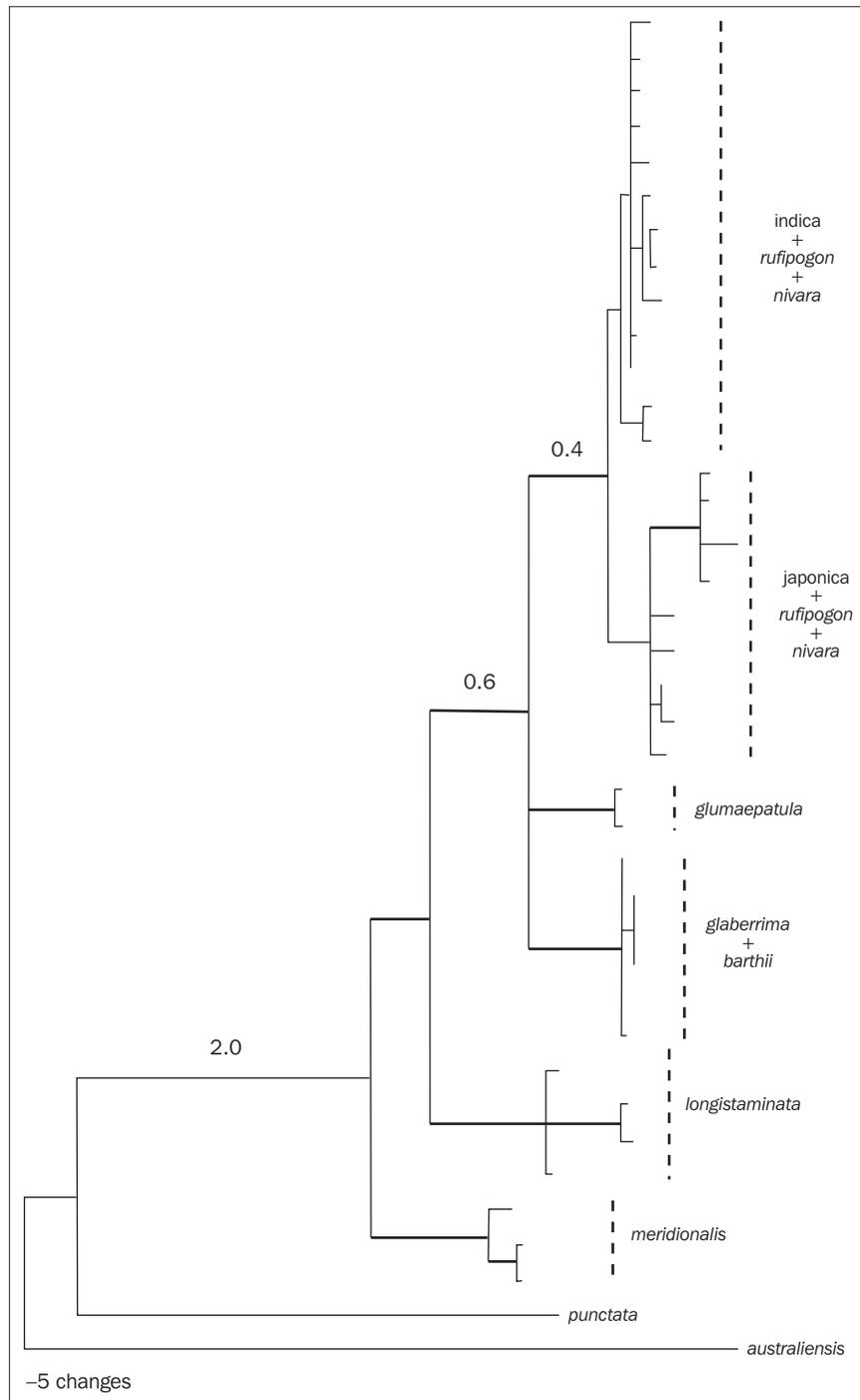


Fig. 2. Strict consensus tree of the A-genome species in *Oryza* generated from the combined intron sequences of three nuclear genes. The thick lines indicate the nodes with the bootstrap support and Bayesian posterior probability over 95%, while the thin lines indicate the nodes with the bootstrap support and Bayesian posterior probability from 75% to 94%. The numbers on the node represent divergence times in million years ago (MYA) as estimated by molecular clock approaches. Accessions were omitted from the figures.

- Vaughan DA, Morishima H. 2003. Biosystematics of the genus *Oryza*. In: Smith W, editor. Rice: origin, history, technology, and production. New York, N.Y. (USA): John Wiley & Sons, Inc. p 27-65.
- Vaughan DA. 1994. The wild relatives of rice: a genetic resources handbook. Manila (Philippines): International Rice Research Institute.

Notes

Authors' address: Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China, e-mail: gesong@ibcas.ac.cn.

Acknowledgments: We thank Tao Sang of Michigan State University (East Lansing, Mich., USA) and Bao-rong Lu of Fudan University (Shanghai, China) for their encouragement and assistance. We are also grateful to the International Rice Genebank at IRRI (Los Baños, Philippines) for providing seed samples. This study was supported by the Chinese Academy of Sciences (Kscxz-sw-101A), the National Natural Science Foundation of China (30025005), and the Program for Key International S & T Cooperation Project of China (2001CB711103).

Eco-genetic diversification in the genus *Oryza*: implications for sustainable rice production

Duncan Vaughan, Koh-ichi Kadowaki, Akito Kaga, and Norihiko Tomooka

Rice is a component of some remarkably productive and sustainable farming systems. However, many factors threaten sustainable rice farming, such as economic development leading to labor, land, and freshwater constraints to rice-farming systems. In addition, abiotic stresses, including the consequences of climate change, and ever-changing biotic stresses require diverse approaches to maintain and improve rice yields. In this paper, the contributions that diversity in the gene pools of rice can make to rice improvement are considered.

Primary gene pool

Since *Oryza* species having the AA genome can hybridize naturally with one another, they can all be considered part of the primary gene pool of rice. AA-genome *Oryza* species have astonishing eco-genetic diversification in relation to environments with different hydrological regimes. During domestication, particular ecotypes of wild rice were selected for increased yield. Subsequently, because of selection and accompanying human migrations, diversification of the cultivated rice gene pool extended the climatic and geographic range of rice to areas where its wild relatives do not grow (Vaughan et al 2004a,b). The wild relatives of rice in this gene pool have been the main source of useful genes for rice improvement from wild *Oryza*. Recent examples of the use of the wild species in this gene pool are given below.

Abiotic stress resistance in *O. rufipogon* from the Mekong Delta

In December 1990, a collaborative collecting mission for the wild relatives of rice was undertaken by the Vietnam Agricultural Sciences Institute and the International Rice Research Institute (IRRI) in the Mekong Delta. During that mission, areas of adverse soils where the wild rice *O. rufipogon* Griff.

grows were targeted for collection. Many samples of *O. rufipogon* were collected from different parts of the Mekong Delta, including Phup Hiep and Thap Muoi districts, where the soils were bright red and known to be highly acid sulfate. Passport data of the collections reported this. Consequently, germplasm from these locations could be rapidly evaluated in standardized conditions to determine the type and degree of abiotic stress resistance it had. Three accessions (106412, 106423, and 106424)¹ were found to be useful sources of resistance to aluminum toxicity. One accession (106424) was crossed to rice (cv. IR64) and recombinant inbred lines developed from this cross were analyzed. Using root length to evaluate the response to stress, five QTLs from *O. rufipogon* were found that explained aluminum tolerance and these were mapped to rice chromosomes 1, 3, 7, 8, and 9 (Nyugen et al 2003). Since this accession (106424) has useful traits and is being used in rice improvement programs, it is one of the accessions chosen for analysis by the *Oryza* Map Alignment Project (www.genome.arizona.edu/BAC_special_projects/).

Biotic stress resistance in *O. rufipogon* from the Central Plains of Thailand

Tungro is the most serious virus disease of rice in South and Southeast Asia (Azzam and Chancellor 2002). Sources of resistance to tungro viruses are found in cultivated rice but some of these have broken down as new forms of these viruses have emerged. Generally, field collection notes cannot help guide

¹Accession numbers mentioned are materials in the International Rice Genebank at IRRI and refer to germplasm collected directly by the first author with collaborating colleagues in national programs while a staff member at IRRI.