Complicated evolutionary history of wild and cultivated Asian rice disclosed by gene tree incongruence analysis

Ching-chia Yang1,2  Hiroaki Sakai2  Takeshi Itoh2
87358@ib.k.u-tokyo.ac.jp  hirsakai@affrc.go.jp  taitoh@affrc.go.jp

1 Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8561, Japan
2 Division of Genome and Biodiversity Research, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan

Keywords: gene tree incongruence, Oryza sativa, Oryza rufipogon, rice evolution

1 Introduction

The genus Oryza (rice) contains two cultivated and ~20 wild species with various genotypes and ecotypes. Here we attempt to elucidate the evolutionary history of rice from a genomic perspective. Oryza sativa is the Asia-originated cultivar, which is thought to be domesticated from its wild progenitor, O. rufipogon. O. sativa can be classified into diverse varietal groups including japonica and indica. In this study, to reconstruct phylogenetic trees, we utilized the genome sequences obtained from japonica and indica, and thousands of full-length cDNAs (FLcDNAs) from O. rufipogon. We expected that a large amount of high-throughput sequence data should be useful for accurate molecular evolutionary analysis.

To infer phylogenetic relationship, evolutionary biologists usually reconstruct gene trees from several gene sequences, and a possible species tree is deduced from the gene trees. It is straightforward to consider that with a large number of sequences, a reliable species tree will be obtained. However, previous studies have shown that with increasing amounts of molecular data, trees from different sets could result in conflicting branching patterns, which are known as gene tree incongruence [1]. Gene tree incongruence was found in several studies of closely-related species, such as hominids, pines and fruit flies. Taking gene tree incongruence into account, we can reconstruct precise evolutionary processes of closely-related rice, which have not been examined previously because of a paucity of sequence data.

2 Method and Results

The FLcDNA sequences of O. rufipogon Griff. W1943 were downloaded from http://202.127.18.228/ricd/dyn/ftp.php [2], and 2,026 sequences without repeats, vectors and poly A sequences were used for mapping to the japonica (IRGSP build 4) and indica (BGI-RIS assembly) genomes. About 92% of FLcDNAs (1,863 sequences) could be mapped to both genomes. A total of 1,030 representative sets of O. rufipogon FLcDNAs, and their corresponding japonica and indica genomic sequences were used for reconstructing gene trees.

To determine the tree topology of three species, the molecular clock was used. Under the assumption of the molecular clock hypothesis, the most distant species should contain the most differences among three species. Therefore, a multinomial test was performed to investigate whether or not one sequence has significantly more differences than the other two sequences in a set. Among 168 significant sets (p <0.05), 100 sets (60%) showed that O. rufipogon and japonica are closely-related (tree 1), 24 sets (14%) showed that O. rufipogon and indica are closely-related (tree 2), and 44 sets (26%) showed that indica and japonica are closely-related (tree 3) (Figure 1A). The fact that three different evolutionary relationships were supported suggested gene tree incongruence of these rice species. In addition, the p distances of the closely-related species in trees 1, 2, and 3 were 8.5 x 10^-4, 2.5 x 10^-3, and 1.4 x 10^-3, respectively (Figure 1B).

Paralogy, lineage sorting and hybridization are major factors that cause gene tree incongruence [3]. Since we examined speciation events during short-term, the effect of independent deletions of paralogs should be negligible. If lineage sorting occurred among three species, there should be three tree topologies: a major tree that is identical to the species tree, and two minor trees that are equal in occurrence frequency. A theory suggests that the distances between the closely-related species in two minor trees should be similar,
and that they are longer than that in the major tree. Our observation indicated that the major tree (tree 1) and a minor tree (tree 2) showed an expected result, but the ratio and distance of tree 3 were unexpectedly different from that of tree 2. Hence, lineage sorting does not seem to be the only factor. The shorter distance between indica and japonica in tree 3 implies that hybridization occurred between indica and japonica after the speciation of *O. rufipogon* and japonica (Figure 2). Therefore, tree 3 can be a composite of gene trees that have undergone lineage sorting and hybridization. This hypothesis accounts for our observation that the ratio of tree 3 was larger than that of tree 2 (Figure 1A) and the distance between closely-related species in tree 3 was shorter than that in tree 2 (Figure 1B).

Figure 1. (A) Numbers and ratios of the tree topologies. (B) d distances between closely-related species of each tree. (Or, *O. rufipogon*; J, japonica; I, indica)

Figure 2. Inferred evolutionary relationship of *O. rufipogon*, japonica and indica. It was suggested that (1) *O. rufipogon* and japonica are closely-related, (2) lineage sorting occurred and resulted in gene tree incongruence, and (3) hybridization happened between japonica and indica after the speciation of *O. rufipogon* and japonica, which lead to inequality in ratio and distance of minor trees.

3 Discussions

With a large amount of sequence information, we examined gene tree incongruence of wild and cultivated Asian rice. We found lineage sorting and hybridization events among the three rice species, and showed a possible evolutionary history. Thus, evolutionary events inferred from gene tree incongruence analysis should help resolve complicated evolutionary processes of closely-related species.

References