Characterization of genetic diversity and genetic structure of a mutational rice (B810S) from China by atpH gene sequence

zhongbo li[1], Congtian Wang[1], Kebao Song[1], Huanyan Wang[1], and Yaling Li[1]

[1]Affiliation not available

August 28, 2020

Abstract

In the present study, to the evaluate genetic diversity and genetic structure of a mutational rice (B810S) with a color marker from China, the entire sequence of the atpH gene of the chloroplast genome from B810S was amplified by polymerase chain reaction (PCR). The PCR products were sequenced, and sequence variations in the atpH gene were then examined. The entire sequence of the atpH gene was 245 bp in size. Compared it with 810S, whichis the male parent of B810S, the results showed a similarity of 51.22%. There are 2 haplotypes in their atpH gene sequences and the haplotype diversity and nucleotide diversity were 1.0000 and 0.41250, respectively. In addition, phylogenetic analysis of B810S and other normal rice strains was conducted and a phylogenetic tree was constructedusing maximum likelihood (ML) method based on the atpH gene sequence from the chloroplast genome. The results of phylogenetic analysis showed that B810S separated from other rice strains formed an independent branch in the phylogenetic tree, which suggests that B810S with a color marker is a new male sterile rice line. The molecular approach employed in this study provides a foundation for further study of the genetic diversity and structure of rices in different geographical regions of China and other regions in the world.

Introduction

Rice, which is a very important and common crop, is grown in many countries and regions of the world, especially in China. There is wide genetic variation amongall plant species, especially for rices. The genus *Oryzais* is more wildely grown among all rice species , and also shows higher genetic diversity, so it is often regarded as a valuable resource for rice improvement (Brar, 2003; Sun *et al* . 2001). Using wild rice in breeding programs can facilitate adaptation of rice to climate change and meet the demand for food security in the face of rapid world population growth (Henry, 2016; Moner *et al* . 2018). Some previous studies have revealed that members of the genus*Oryzais* can provide a rich repository of genes and alleles for potential utilization in rice improvement with the help of genomics-assisted breeding (Lam *et al* . 2018). For example, in recent years, several quantitative trait locus (QTL) studies have revealed that *O. rufpogon* has genetic potential for improving yield-related traits (Shishido *et al* . 2019). These results are in accordance with some previous studies which showed that *O. rufpogon* from the low-yielding had more beneficial effects than*O. sativa* in different elite cultivars (Fu *et al.* 2010, Xie *et al.* 2008). Some alleles which contributed to early spikelet opening times of wild rice were detected by Thanh *et al.,* (Thanh *et al.* 2010). Their study results showed that this trait is useful for changing cultivar flowering times to avoid pollen sterility at high temperatures. Moreover, three novel alleles which enhanced the phosphorus uptake efficiency of *O. rufpogon* were also reported by Neelam *et al.,* (Neelam*et al.* 2017). These above-mentioned alleles will promote a better understanding ecological diversity of rice, and provide useful gene markers for further studies of the genetic diversity and genetic structure of rice.

Generally, genetic variation is a very common phenomenon. There are many species of organisms which

exhibit greater genetic variations including animals and plants. Rice, as a more common of plant type in the natural environment, must exhibit genetic variations. Presently, for identification and differentiation of rice species, we often employ traditional methods that are based mainly on morphological features, such as the color of rice leaves, and their shapes . However, applying these criteria to identify and distinguish organisms which have similar morphology or are at different stages of development is sometimes difficult for non-specialists, especially for rices with very similar morphologies. The *Oryzais* are the largest genus in the rice family, and their members have similar morphologies, so we sometimes don't identify and distinguish them by traditional methods. In addition, we also can't clearly know the taxonomic status and evolutionary relationship of new rice species and wild rice using morphological methods. Molecular techniques not only overcome the limitations of traditional methods, but also provide alternative approaches for accurate identification and differentiation of many species of crops with very similar morphologies. Many gene markers are widely used as a powerful tool to study the genetic diversity, population structures and genetics of crop species (Gao et al, 2004; Song et al. 2003; Kaewcheenchai et al., 2018). In fact, some previous studies have used some genes from the chloroplast (cp) genome as powerful gene markers to monitor transition in population structures in nature (Orn et al., 2015; Wang et al., 2012). For example, Shishido et al., used two chloroplast DNA markers (e.g., ORF100 and ORF29-*trn* C) to study genetic diversity and genetic structure of wild rice populations in Myanmar (Shishido et al. 2019). China is a large agricultural country and large rice breeding country where rice is widely grown, but there is a paucity of information regarding chloroplast (cp) genome sequence variations, genetic diversity and genetic structures of rice.

The chloroplast (cp) genome contains three functional categories which include protein-coding genes, introns and intergenic spacers; the latter two do not encode proteins and are often referred to as non-coding regions but the genes of non-coding regions show that their nucleotide substitution rates are 2.3 times higher than those of protein-coding genes, so some gene sequences from the non-coding regions have been used to study genetic diversity, genetic structures and population structures of organisms. For example, the use of non-coding chloroplast DNA sequences to generate plant phylogenies began in the early 1990s (Taberlet et al. 1991; Clegg et al. 1994; Gielly and Taberlet. 1994). Shaw et al. used 21 non-coding region genes of chloroplast (cp) genomes to study the interspecific phylogenetic and intraspecific phylogeography of *Nicotiana* in 2007 (Shaw et al. 2007). Moreover, some research clearly indicates that non-coding cpDNA sequences have been used in molecular systematic research of plants for more than 15 years (Shaw et al. 2007). This evidence has suggested that non-coding gene sequences can provide powerful gene markers to study genetic diversity, genetic structures and population structures of organisms. According to the chloroplast map drawn by Shaw, the aptH gene is in a non-coding region and the aptH gene belongs to a non-coding gene. So, we used the atpH gene sequence as a gene marker to study the genetic diversity and genetic structure of B810S in this study.

The objectives of the present study were (a) to clarify the genetic diversity of B810S, and (b) reveal the genetic structure of B810S in China by using cytoplasmic markers. In addition, the results of the current study should provide a foundation for studying the genetic diversity and genetic structure of rice with different geographical origins in China and other regions of the world. Moreover, our study results also provide insight into mutational rice resources that could be useful for biological conservation as well as exploitation in rice breeding programs.

Materials and methods

Chloroplast DNA extraction

Mutational rice (B810S) with a color marker was obtained from a deserted paddy field in Huaihua, Hunan Province, China. After collection, all leaves from B810S and 810S were washed three times with physiological saline, and were then initially identified by their morphological characteristics using a light microscope. The genomic DNA contents of B810S and 810S were isolated from their leaves using a Plant DNA isolation Reagent kit (Tiangen, Beijing, China) following the manual instructions, and were then stored at -20°C until use.

PCR amplification and sequencing

The entire sequences of the *atp* H gene of B810S and 810S with specificity primers *atp* H-F(AAGGATGGAATGAAAGATNAG), and*atp* H-R (GCTTTTATTTGCGAACCCTTT

-TGTTTAA) was amplified by polymerase chain reaction (PCR). The PCR reactions contained ˜20 ng of genomic DNA and were carried out in 50-μL reaction volumes containing 25 μL of 2×Phusion EmwealdAmp MAX HS PCR Master Mix (TaKaRa, Dalian, China), 2 μL of each primer, 2 μL of DNA and 19 μL of ddH_2O and 2 μL of DNA sample in a thermocycler (Biometra, Göttingen, German). The PCR conditions were as follows: At the beginning of the preheat temperature 94°C for 5 min, followed by 35 cycles of denaturation at 94°C for 30 s; annealing at 50°C for 30 s, and extension temperature at 72°C for 2 min, with a final extension at 72°C for 5 min. All products of PCR amplification were detected by 1.5% agarose gel electrophoresis to validate the amplification efficiency. The PCR products were sent to BGI-Shenzhen (Shenzhen, China) for sequencing from both directions.

Sequences analysis

The entire sequence of the *atp* H genes of B810S and 810S were determined by comparisons with those of other previously reported rices (GenBank accession number NC_031333). The *atpH* gene sequences of B810S', 810S' and other rices (downloaded from Genebank)were aligned by the computer program Clustal X 1.81. To calculate the sequence differences between B810S and 810S, the formula $D = 1$ - (M/L) was used. In the formula, different capitals represent different means, where M is the number of alignment positions at which the two sequences have a common base , and L is the total number of alignment positions over which the two sequences are compared. The Megalign procedure within DNAStar 5.0 (Burland, 2000) was used to analyze sequence similarities of B810S and other rices which included 810S and other rice gene sequences which are downloaded from Genebank. Also, to align the variable sites of B810S and 810S, and calculate their haplotype (Hd) and nucleotide diversity (Pi), the computer program Dna SP 5.0 was used with default parameters.

Phylogenetic analysis

To infer the phylogenetic relationship of B810S, 810S and other rices, a phylogenetic tree was re-constructed using the maximum likelihood (ML) method. ML analyses were performed using PhyML 3.0 (Guindon et al. 2010), and the GTR+I model with its parameters determined by JModeltest (Posada, 2008). Bootstrap support (BS) for the ML tree was calculated using 100 bootstrap replicates. In the tree, B810S, 810S and other members of the *Oryzais* genus(e.g.,(*O.sativa* (MK348618),*O.coarctata* (MG383937), *O. longistaminata* (KF359907),*O.meridionalis* (KF359906), *O.barthii* (KF359904), *O. glaberrima*

(KF359903), *O. rufipogon* (KF359902), *O. glumipatula* (KR364803), *O. nivara* (AP006728),*O.rhizomatis* (KX085497), *O.eichingeri* (KX085496),*O.latifolia* (MF401451), *O. alta* (KF359913) and*O.officinalis* (KF359910)) were considered as the inside-group, and*Triticum aestivum* was considered as the out-group. Tree View program version 1.65 was used to draw the phylogenetic tree (Page, 1996).

Results and Discussion

The total genomic DNA of B810S, which was collected from a deserted paddy field, and that of 810S were extracted using a Plant DNA isolation Reagent kit (Tiangen, Beijing, China) following the manual specifications. The entire sequences of the *atp* H genes of B810S and 810S were amplified by PCR, and were then subjected to detection in 1.5% agarose gel electrophoresis, which appeared two single bands, and both are approximately 250 bp in length.

The entire sequences of the *atp* H genes of B810S and 810S were 245 bp in size. The *atp* H gene sequence of B810S has been deposited in GenBank under accession number: MN879320. A comparison of B810S with 810S, indicates that the similarity of both sequences is 51.22%, and there are 99 variation sites in their *atp* H gene sequences. The results of haplotype (Hd) and nucleotide diversity (Pi) analysis of the*atp* H gene sequences of B810S and 810S showed that there are 2 haplotypes in their *atp* H gene sequences and that

the haplotype diversity and nucleotide diversity were 1.0000 and 0.41250, respectively. Comparisons with other members of the *Oryzais*genus, indicated that the sequence differences were significantly higher, namely, 48.78-49.19% for their *atp* H gene sequences.

To test whether B810S with a color marker is a new male sterile rice line, the variation frequencies of allele genes between B810S and 810S were evaluated using Tajima's D statistical analysis.The result of Tajima's D statistical analysis showed that the Tajima's Dis was 2.32478. These results indicate that there is genetic differentiation between B810S and 810S. Also, the mismatch distribution analysis of the *atp* H gene sequence of B810S and 810S, showed that there were multi-peaks in their *atp* H gene sequences, which suggests that B810S is a new male sterile rice line.

In addition, to infer the phylogenetic relationship of B810S and other members of the *Oryzais* genus, phylogenetic analysis was performed using the maximum likelihood (ML) method with the GTR+I model in the study. The results of phylogenetic analysis showed that B810S evidently separated from 810S and other rice strains, and formed an independent branch in the phylogenetic tree (Fig. 1). This result is in accordance with the mismatch distribution analysis, which showed that there was obvious genetic differentiation among B810S and other members of *Oryzais*genus(including 810S and other rice sequences which were downloaded from Genebank), which suggest that B810S is a new male sterile rice line.

In conclusion, the genetic diversity and genetic structure of B810S were revealed by the *atp* H gene sequence of the cp genome from rice leaves, and this study also confirmed that B810S with a color marker is a new male sterile rice line. In addition, the molecular approach employed in this study provides a foundation for studying the genetic variations and structures of rice in different geographical regions of China and other regions of the world.

Acknowledgements

This work was supported in part by Scientific Research Fund of Hunan Provincial Education Department (No. 16A102).

Data Accessibility Statement

All data of our study have been deposited in GenBank, and the DNA sequence' Genbank accessions MN879320; NCBI SRA: SRX0110215.

References

Brar, D. S. (2003). Utilization of wild species of genus Oryza in rice improvement. In J. S. Nanda & S. D. Sharma (Eds.), *Monograph on genus Oryza* . Enfield, UK: Science Publishers, 283–309.

Burland TG. DNASTAR's Lasergene sequence analysis software. Methods Mol Biol 2000; 132:71-91.

Clegg MT, Gaut BS, Learn GHJ, Morton BR. (1994) Rates and patterns of chloroplast DNA evolution. Proceedings of the national Academy of Sciences, 91: 6795-6801.Fu Q, Zhang P, Tan L,Zhu Z, Ma D, Fu Y, Zhan X, Cai H, Sun C. Analysis of QTLs for yield-related traits in Yuanjiang common wild rice (Oryza rufpogon Griff.). J. Genet. Genomics 2010, 37: 147–157.

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

Gao LZ, Schaal BA, Zhang CH, Jia JZ, Dong YS. (2002) Assessment of population genetic structure in common wild rice Oryza rufpogon Griff. using microsatellite and allozyme markers. Theor Appl Genet. 106: 173–180.

Gielly L,Taberlet P. (1994) The use of chloroplast DNA to resolve plant phylogenies: non-coding versus rbcL sequences. Molecular Biology and Evolution 11: 769-777.

Henry, RJ. (2016). Genomics strategies for germplasm characterization and the development of climate resilient crops. In S. Kjmar (Ed.), Crop breeding: Bioinformatics and preparing for climate change (pp. 3–10). Waretown, NJ: Apple Academic Press. CRC Press.

Kaewcheenchai R, Promnart U, Soontrajarn K, Chotechuen S, Chitrakon S, Yuki H, Ishikawa R. (2018). Diverse genetic variation in maternal lineages with high heterogeneity among in-situconserved wild rice (Oryza rufipogon Griff.) developed in Thailand. Breeding Science, 68, 614–621

Lam DT, Buu BuiC, Lang Nguyen T, Toriyama Kinya, Nakamura Ikuo, Ishikawa Ryuji. Genetic diversity among perennial wild rice Oryza rufipogon Griff., in the Mekong Delta. Ecology and Evolution 2018,9(5):2964-2977.

Moner, AM., Furtado, A., Chivers, I., Fox, G., Crayn, D., & Henry, R.J. Diversity and evolution of rice progenitors in Australia. Ecology and Evolution 2018, 8:4360–4366.

Neelam K, Thakur S, Yadav NIS, Kumar K, Dhaliwal SS, Singh K. (2017) Novel alleles of phosphorus-starvation tolerance 1 gene (PSTOL1) from Oryza rufpogon confers high phosphorus uptake effciency. Front. Plant Sci. 8: 509.

Orn C, Shishido R, Akimoto M, Ishikawa R, Htun TM, Nonomura KI, Ishii I. (2015) Evaluation of genetic variation among wild rice populations in Cambodia. Breeding Science, 65, 430–437.

Posada D. JModelTest phylogenetic model averaging. Mol Biol Evol 2008; 25: 1253–1256.

Page RD. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 1996; 12: 357–358.

Sun, C. Q., Wang, X. K., Li, Z. C., Yoshimura, A., & Iwata, N. (2001) Comparison of the genetic diversity of common wild rice (Oryza rufipogon Griff.) and cultivated rice (O. sativa L.) using RFLP markers.Theoretical and Applied Genetics, 102, 157–162.

Shishido Rieko, Akimoto Masahiro, Htut Tin, Sato Tadashi, Ishii Takashige. Assessment of genetic diversity and genetic structure of wild rice populations in Myanmar. Breeding Science 2019,69: 471–477.

Song ZP, Xu X, Wang B, Chen JK, Lu BR. (2003) Genetic diversity in the northernmost Oryza rufpogon populations estimated by SSR markers. Theor. Appl. Genet. 107: 1492–1499.

Shaw J, Lickey EB, Schilling EE, Small RL. (2007) Comparison of whole chloroplast genome sequences to choose non-coding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. American Journal of Botany 94(3): 275–288.

Shishido R, Kikuchi M, Nomura K, Ikehashi H. (2006) Evaluation of genetic diversity of wild rice (Oryza rufpogon Griff.) in Myanmar using simple sequence repeats (SSRs). Genet. Resour. Crop Evol. 53: 179–186.

Taberlet P, Gielly L, Pautou G, Bouvet J. (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology, 17: 1105-1109.

Thanh PT, Phan PDT, Ishikawa R, Ishii T. (2010) QTL analysis for flowering time using backcross population between Oryza sativa Nipponbare and O. rufpogon. Genes Genet. Syst. 85: 273–279.

Wang YP, Bounphanousay C, Kanyavong K, Nakamura I, Sato YI, Sato T, Ishikawa R. (2012) Population structural analysis of an in-situ conservation site for wild rice in Laos. Genes & Genetic Systems,87, 311–322.

Xie X, Jin F, Song MH, Suh JP, Hwang HG, Kim YG, McCouch SR, Ahn SN. (2008) Fine mapping of a yield-enhancing QTL cluster associated with transgressive variation in an Oryza sativa x O. rufpogon cross. Theor. Appl. Genet. 116: 613–622.

Fig.1 Phylogenetic relationship of B810S and other rices was inferred by maximum likelihood analyses using the atpH, with *Triticum aestivum* as out-group.